GENES, CHROMOSOMES & CANCER 46:914–928 (2007)

Transcriptional Oncogenomic Hot Spots in Barrett's Adenocarcinomas: Serial Analysis of Gene Expression

Mohammad H. Razvi,[1] Dunfa Peng,[1] Altaf A. Dar,[1] Steven M. Powell,[2] Henry F. Frierson Jr.,[3] Christopher A. Moskaluk,[3] Kay Washington,[4] and Wael El-Rifai[1,5]*

[1]Department of Surgery, Vanderbilt University Medical Center, Nashville, TN
[2]Division of Gastroenterology, Vanderbilt University Medical Center, Nashville, TN
[3]Department of Pathology, University of Virginia, Charlottesville, VA
[4]Department of Pathology, Vanderbilt University Medical Center, Nashville, TN
[5]Department of Cancer Biology, Vanderbilt University Medical Center, Nashville, TN

Serial analysis of gene expression (SAGE) provides quantitative and comprehensive expression profiling in a given cell population. In our efforts to define gene expression alterations in Barrett's-related adenocarcinomas (BA), we produced eight SAGE libraries and obtained a total of 457,894 expressed tags with 32,035 (6.9%) accounting for singleton tags. The tumor samples produced an average of 71,804 tags per library, whereas normal samples produced an average of 42,669 tags per library. Our libraries contained 67,200 unique tags representing 16,040 known gene symbols. Five hundred and sixty-eight unique tags were differentially expressed between BAs and normal tissue samples (at least twofold; $P \leq 0.05$), 395 of these matched to known genes. Interestingly, the distribution of altered genes was not uniform across the human genome. Overexpressed genes tended to cluster in well-defined hot spots located in certain chromosomes. For example, chromosome 19 had 26 overexpressed genes, of which 18 mapped to 19q13. Using the gene ontology approach for functional classification of genes, we identified several groups that are relevant to carcinogenesis. We validated the SAGE results of five representative genes (*ANPEP*, *ECGF1*, *PP1201*, *EIF5A1*, and *GKN1*) using quantitative real-time reverse-transcription PCR on 31 BA samples and 26 normal samples. In addition, we performed an immunohistochemistry analysis for ANPEP, which demonstrated overexpression of ANPEP in 6/7 (86%) Barrett's dysplasias and 35/65 (54%) BAs. ANPEP is a secreted protein that may have diagnostic and/or prognostic significance for Barrett's progression. The use of genomic approaches in this study provided useful information about the molecular pathobiology of BAs.

INTRODUCTION

Gastroesophageal reflux disease (GERD) is a major health problem in the United States with a prevalence of 5–7% in the general population and an increasing incidence rate (Serag, 2006). Approximately 10% of patients with chronic GERD develop a metaplastic condition known as Barrett's esophagus (BE) in which the normal squamous epithelium of the esophagus is replaced by a columnar epithelium with goblet cells. BE is a serious premalignant lesion that can ultimately progress from metaplasia to dysplasia and subsequently to Barrett's adenocarcinoma (BA) (Ferraris et al., 1997; O'Connor et al., 1999; Rana and Johnston, 2000). The incidence of BA has rapidly increased in the Western world over the past three decades (Hamilton et al., 1988; Phillips et al., 1991; Blot et al., 1993), and is comprised of aneuploid tumors characterized by complex molecular alterations (El-Rifai et al., 2001; El-Rifai and Powell, 2002). Several genetic abnormalities have been associated with Barrett's tumorigenesis, including microsatellite instability (Meltzer et al., 1994), loss of heterozygosity (Dolan et al., 1999), gene-promoter hypermethylation (Sato and Meltzer, 2006), as well as up- and down-regulation of various genes (Wu et al., 1993; Swami et al., 1995; Regalado et al., 1998; Brabender et al., 2002). Comprehensive molecular analyses of DNA amplifications and gene expression have revealed complex genetic alterations in gastroesophageal and lower esophageal adenocarcinomas (El-Rifai et al., 1998; Varis et al., 2002; van Dekken et al., 2004; Kuwano et al., 2005).

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, University of Virginia, or Vanderbilt University.

Supported by: National Cancer Institute; Grant numbers: R01CA106176 (WER), GI SPORE CA 95103.

*Correspondence to: Wael El-Rifai, MD, PhD, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, 1255 Light Hall, 2215 Garland Avenue, Nashville, TN 37232, USA.
E-mail: wael.el-rifai@vanderbilt.edu

Received 10 April 2007; Accepted 27 June 2007
DOI 10.1002/gcc.20479
Published online 17 July 2007 in Wiley InterScience (www.interscience.wiley.com).

Analyses of the human transcriptome map of normal tissues have shown clustering of highly expressed genes in chromosomal domains (Caron et al., 2001). 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 transcription and splicing (Cremer and Cremer, 2001). Recently, a few reports have suggested the presence of transcriptional hot spots in the cancer genome, (Wu et al., 2006) where overexpressed genes tend to cluster in defined chromosomal domains; however, similar information remains lacking for most cancer types. Serial analysis of gene expression (SAGE) provides unlimited, comprehensive, genome-wide analysis of gene expression in a given cell population (Velculescu et al., 1995, 2000). The major advantage in using SAGE is the quantitative ability to accurately evaluate transcript numbers without prior sequencing information. This method has proven invaluable in studies of several tumor types, including adenocarcinomas of the colon (Parle-McDermott et al., 2000; St Croix et al., 2000), prostate (Culp et al., 2001), pancreas (Argani et al., 2001), ovary (Hough et al., 2000), and breast (Seth et al., 2002). In this study, we explored the BA transcriptome using SAGE and mapped gene-expression changes to chromosomal positions, thereby generating a map of transcriptional oncogenomic hot spots of this deadly cancer.

MATERIALS AND METHODS

Serial Analyses of Gene Expression

High-quality total RNA (500 μg) was extracted from four intestinal-type, moderately to poorly differentiated, BA cases (three gastroesophageal junctional [GEJ] and one lower esophageal) using an RNeasy kit (QIAGEN, Hilden, Germany). In addition, four normal gastric mucosa pools were used as reference samples. Each of these pools consisted of four normal gastric mucosal biopsy samples from four different individuals. The tumors selected for SAGE analysis were estimated to consist of more than 70% tumor cells. All normal samples had histologically normal mucosae confirmed on review of hematoxylin- and eosin-stained sections. Importantly, histopathological examination confirmed that none of the normal samples had any areas of inflammation or necrosis. All samples were collected with consent in accordance with approved Institutional Review Board protocols. 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 the standard protocol (Velculescu et al., 1995). A detailed protocol and schematic of the method is available at (http://www.sagenet.org/protocol/index.htm). We sequenced 20,000 clones with an average of 2,500 clones per library, using the Cancer Genome Anatomy Project (CGAP). eSAGE 1.2a software was used 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-Tag-Mapping method (Margulies and Innis, 2000; Margulies et al., 2001). The resulting libraries' tags were compared with UniGene clusters and the SAGE tag "reliable" mapping database (http://www.sagenet.org/resources/genemaps.htm). Statistical analyses of these tags were then performed using eSAGE software.

Quantitative Real-Time Reverse-Transcription PCR

Quantitative real-time reverse-transcription PCR (qRT-PCR) was performed on 31 adenocarcinomas of Barrett's-related origin, 26 normal gastric epithelial tissues, and 6 Barrett's metaplasia tissue samples. All tissues were dissected to obtain ≥70% cell purity. All of the adenocarcinoma samples were collected from the GEJ or lower esophagus and ranged from well differentiated (WD) to poorly differentiated (PD), Stages I–IV, with a mix of intestinal- and diffuse-type tumors. RNA was purified from all samples using an RNeasy Kit. Single-stranded cDNA was generated using an Advantage™ RT-for-PCR Kit (Clontech, Palo Alto, CA). qRT-PCR was performed using an iCycler (BioRad, Hercules, CA) with SYBR Green technology, and the threshold cycle numbers were calculated using iCycler software v3.0. Reactions were performed in triplicate and threshold cycle numbers were averaged. For validation of SAGE results, we designed gene-specific primers for human *ANPEP*, *ECGF1*, *PP1201*, *EIF5A1*, *GKN1*, and *HPRT1*. These primers were obtained from Integrated DNA Technologies (IDT, Coralville, IA) and their sequences are available upon request. A single-melt curve peak was observed for each product, thus confirming the purity of all amplified cDNA products. The qRT-PCR results were normalized to *HPRT1*, which had minimal variation in all normal and neoplastic samples tested. Fold overexpression was calculated according to the formula, $2^{(R_t - E_t)}/2^{(R_n - E_n)}$, as described earlier (Buck-

TABLE 1. TheTop 93 Deregulated Genes in Barrett's Adenocarcinomas

Tag sequence	UniGene cluster ID	Gene symbol	Title	Location	T4 tag count	N4 tag count	Ratio, T4/N4	*P* value
Upregulated genes								
GTGGCCACGG	Hs.112405	S100A9	S100 calcium binding protein A9	1q21	355	0	418	≤0.001
GAGCAGCGCC	Hs.112408	S100A7	S100 calcium binding protein A7	1q21	95	0	112	≤0.001
AAGATTGGTG	Hs.114286	CD9	CD9 antigen (p24)	12p13.3	112	7	10	≤0.001
GCACCTGTCG	Hs.1239	ANPEP	Alanyl (membrane) aminopeptidase	15q25-q26	76	0	89	≤0.001
GTGACAGAAG	Hs.129673	EIF4A1	Eukaryotic translation initiation factor 4A, isoform 1	17p13	92	4	14	≤0.001
TTTCCTGCTC	Hs.139322	SPRR3	Small proline-rich protein 3	1q21-q22	308	0	362	≤0.001
GTTCAAGTGA	Hs.186810	REPS2	RALBP1 associated Eps domain containing 2	Xp22.2	107	2	32	≤0.001
ACTGTATTTT	Hs.194691	Hs.194691	G protein-coupled receptor, family C, group 5, member A	12p13-p12.3	103	6	10	≤0.001
TGGATCCTGA	Hs.302145	HBG2	Hemoglobin, gamma G	11p15.5	75	0	88	≤0.001
CAGGAGGAGT	Hs.308709	GRP58	Protein disulfide isomerase family A, member 3	15q15	81	2	24	≤0.001
CTAGTCTTTG	Hs.353175	AGPAT4	1-acylglycerol-3-phosphate *O*-acyltransferase 4	6q26	85	0	100	≤0.001
TCACCCAGGG	Hs.391464	ABCC1	ATP-binding cassette, subfamily C member 1	16p13.1	52	0	61	≤0.001
CCTGGTCCCA	Hs.411501	KRT7	Keratin 7	12q12-q13	179	1	106	≤0.001
TTCTTTCTAA	Hs.411925	TMEM38B	Transmembrane protein 38B	9q31.2	58	1	34	≤0.001
TACCTGCAGA	Hs.416073	S100A8	S100 calcium binding protein A8	1q21	343	1	204	≤0.001
CAGCAGAAGC	Hs.424126	SERF2	Small EDRK-rich factor 2	15q15.3	79	4	12	≤0.001
GCGGCGGATG	Hs.445351	LGALS1	Lectin, galactoside-binding, soluble, 1	22q13.1	89	0	105	≤0.001
GAACATTGCA	Hs.447579	LOC339290	Hypothetical protein LOC339290	18p11.31	95	0	112	≤0.001
GTTTGGGTTG	Hs.459927	PTMA	Prothymosin, alpha (gene sequence 28)	2q35-q36	162	9	11	≤0.001
TCACCCACAC	Hs.462859	SCFD2	Short-chain dehydrogenase/reductase	17q12	337	31	6	≤0.001
CCCCCGCGGA	Hs.466507	LISCH7	Liver-specific bHLH-Zip transcription factor	19q13.12	48	0	56	≤0.001
CGGAGACCCT	Hs.473583	NSEP1	Y box binding protein 1	1p34	76	2	23	≤0.001
GCCGGGTGGG	Hs.501293	BSG	Basigin (OK blood group)	19p13.3	77	4	11	≤0.001
GATACTTGGA	Hs.501911	GALNTL4	Casein kinase 2, alpha 1 polypeptide	11p15.3	94	0	111	≤0.001
ACAGGCTACG	Hs.503998	TAGLN	Transgelin	11q23.2	71	3	14	≤0.001
GTGGCTCACA	Hs.504820	MGC14817	Hypothetical protein MGC14817	12q14.3	242	16	9	≤0.001
TAATTTTTGC	Hs.508113	OLFM4	Olfactomedin 4	13q14.3	228	1	136	≤0.001
GTGAGCCCAT	Hs.509736	HSPCB	Heat shock 90 kDa protein 1, beta	6p12	149	13	7	≤0.001
TGTCAGTCTG	Hs.512350	Hs.512350	LOC440676	1q21.1	108	1	64	≤0.001
AGTGCAGGGC	Hs.512488	Hs.512488	Similar to 60S ribosomal protein L10	12q21.2	98	1	58	≤0.001
GCGACCGTCA	Hs.513490	ALDOA	Aldolase A, fructose-bisphosphate	16q22-q24	206	4	31	≤0.001
ACCGCCGTGG	Hs.513803	CYBA	Cytochrome b-245, alpha polypeptide	16q24	77	0	91	≤0.001
AGCAGGAGCA	Hs.515714	S100A16	S100 calcium binding protein A16	1q21	61	0	72	≤0.001
GATCTCTTGG	Hs.516484	S100A2	S100 calcium binding protein A2	1q21	61	0	72	≤0.001
ATCGTGGCGG	Hs.520942	CLDN4	Claudin 4	7q11.23	62	0	73	≤0.001
CCCAAGCTAG	Hs.520973	HSPB1	Heat shock 27 kDa protein 1	7q11.23	175	7	15	≤0.001
AACATTCGCA	Hs.523302	PRDX3	Peroxiredoxin 3	10q25-q26	46	0	54	≤0.001
CTTCTCATCT	Hs.531719	ADCYAP1	Adenylate cyclase activating polypeptide 1	18p11	85	1	51	≤0.001
AACTGAGGGG	Hs.5333	KIAA0711	Kelch repeat and BTB (POZ) domain containing 11	8p23.3	94	0	111	≤0.001

(Continued)

TABLE 1. TheTop 93 Deregulated Genes in Barrett's Adenocarcinomas (Continued)

Tag sequence	UniGene cluster ID	Gene symbol	Title	Location	T4 tag count	N4 tag count	Ratio, T4/N4	*P* value
GACTCTTCAG	Hs.534293	SERPINA3	Serpin peptidase inhibitor, clade A member 3	14q32.1	125	1	74	≤0.001
CATCGCCAGT	Hs.54483	NMI	N-myc (and STAT) interactor	2p24.3-q21.3	285	0	335	≤0.001
GACGGCGCAG	Hs.546251	ECGF1	Endothelial cell growth factor 1	22q13	46	0	54	≤0.001
TAGCTTTAAA	Hs.554202	SVIL	Supervillin	10p11.2	210	0	247	≤0.001
TGGCCATCTG	Hs.555971	PP1201	Transmembrane BAX inhibitor motif containing 1	2p24.3-p24.1	90	1	54	≤0.001
CTATCCTCTC	Hs.75227	NDUFA9	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9, 39 kDa	12p13.3	51	0	60	≤0.001
ACTGCCCGCT	Hs.81071	ECM1	Extracellular matrix protein 1	1q21	77	1	46	≤0.001
Downregulated genes								
GAGAACCACT	Hs.110014	GIF	Gastric intrinsic factor (vitamin B synthesis)	11q13	0	87	0.010	≤0.001
TTGCCCCTAC	Hs.128814	CHIA	Chitinase, acidic	1p13.1-p21.3	7	185	0.020	≤0.001
ACACAGCAAG	Hs.131603	Hs.476965	EMI domain containing 2	7q22.1	44	250	0.100	≤0.001
ACCCTCCCCA	Hs.132087	FLJ46299	Kelch domain containing 6	3q21.3	0	35	0.024	≤0.001
AACCTCCCCG	Hs.132858	RAP1GDS1	RAP1, GTP-GDP dissociation stimulator 1	4q23-q25	0	33	0.026	≤0.001
CAGTGCCTCT	Hs.133539	MAST4	Microtubule associated serine/threonine kinase family member 4	5q12.3	1	51	0.010	≤0.001
AACCTCCCAC	Hs.134074	ARL2BP	Solute carrier family 35, member E1	19p13.11	1	42	0.010	≤0.001
CTGGCCCTCG	Hs.162807	TFF1	Trefoil factor 1	21q22.3	95	174	0.3	≤0.001
TTTAGGATGA	Hs.16757	GDDR	Down-regulated in gastric cancer GDDR	2p13.3	5	474	0.010	≤0.001
CACCCCTGAT	Hs.173724	CKB	Creatine kinase, brain	14q32	9	74	0.070	≤0.001
GACCTCCCCA	Hs.178728	MBD3	Methyl-CpG binding domain protein 3	19p13.3	2	64	0.020	≤0.001
AGTGCTCTTC	Hs.1867	PGC	Progastricsin (pepsinogen C)	6p21.3-p21.1	36	595	0.040	≤0.001
CCATTCTGAA	Hs.209217	ASTN2	Astrotactin 2	9q33.1	0	24	0.035	≤0.001
CAGTGCTTCC	Hs.220864	CHD2	Chromodomain helicase DNA binding protein 2	15q26	5	41	0.070	≤0.001
GCTGGAGGAA	Hs.2681	GAS	Gastrin	17q21	0	100	0.009	≤0.001
CACCTCCCCA	Hs.283739	BE614337	Ubiquilin 4	1q21	4	76	0.030	≤0.001
AGCCTCCCCA	Hs.2859	OPRL1	Opiate receptor-like 1	20q13.33	2	68	0.020	≤0.001
AAATCCTGGG	Hs.2979	TFF2	Trefoil factor 2 (spasmolytic protein 1)	21q22.3	62	1086	0.030	≤0.001
GCAGGCTCCA	Hs.302131	GHRL	Ghrelin precursor	3p26-p25	5	50	0.060	≤0.001
TGCCAATTAA	Hs.307835	PGM5	Phosphoglucomutase 5	9p12-q12	6	40	0.090	≤0.001
CCCTGGAAGC	Hs.309288	CUGBP2	CUG triplet repeat, RNA binding protein 2	10p13	1	33	0.020	≤0.001
CTGACTGTGC	Hs.36992	ATP4A	ATPase, H^+/K^+ exchanging, alpha polypeptide	19q13.1	10	384	0.020	≤0.001
GTTTGCTTGC	Hs.370480	ABCB7	ATP-binding cassette, sub-family B (MDR/TAP), member 7	Xq12-q13	1	26	0.020	≤0.001
AACCTCCTCA	Hs.386698	C10orf27	Chromosome 10 open reading frame 27	10q22.1	0	29	0.029	≤0.001
TATATCAGTG	Hs.388654	ATP6V1G1	ATPase, H+ transporting, lysosomal 13 kDa, V1 subunit G isoform 1	9q32	3	48	0.040	≤0.001
AACCTCCCCA	Hs.432854	PGA5	Porin, putative	11q13	365	6637	0.030	≤0.001
GGAACGCAAG	Hs.434202	ATP4B	ATPase, H+/K+ exchanging, beta polypeptide	13q34	4	138	0.020	≤0.001
TCTCCATACC	Hs.438454	FBXO25	F-box protein 25	8p23.3	12	376	0.020	≤0.001
TCCCTTTAAG	Hs.438824	CKIP-1	CK2 interacting protein 1	1q21.2	3	49	0.040	≤0.001
TTTTTCAAGA	Hs.445586	UNQ473	DMC	19q13.2	2	35	0.030	≤0.001
CAGTGCTCTT	Hs.445680	Hs.445680	Similar to anaphase promoting complex subunit 1	2q12.3	1	42	0.010	≤0.001
ACTGATCTGC	Hs.447547	VPS35	Hypothetical protein MGC34800	16q12	5	34	0.090	≤0.001

(Continued)

TABLE 1. The Top 93 Deregulated Genes in Barrett's Adenocarcinomas (Continued)

Tag sequence	UniGene cluster ID	Gene symbol	Title	Location	T4 tag count	N4 tag count	Ratio, T4/N4	*P* value
TCATTTTGAA	Hs.464472	MRCL3	Myosin regulatory light chain MRLC2	18p11.31	0	27	0.031	≤0.001
CAATGCTTCT	Hs.474751	MYH9	Myosin, heavy polypeptide 9, nonmuscle	22q13.1	2	70	0.020	≤0.001
TGCGAGACCA	Hs.490038	CPA2	Carboxypeptidase A2 (pancreatic)	7q32	0	24	0.035	≤0.001
CATTGCTTCT	Hs.516297	TCF7L1	Transcription factor 7-like 1 (T-cell specific, HMG-box)	2p11.2	0	82	0.010	≤0.001
CAGTGTTCTT	Hs.518611	TBC1D14	TBC1 domain family, member 14	4p16.1	2	29	0.040	≤0.001
AATGTACCAA	Hs.523130	LIPF	Lipase, gastric	10q23.31	1	51	0.010	≤0.001
CAGTGCTTCT	Hs.527922	DLEU1	Deleted in lymphocytic leukemia, 1	13q14.3	349	8046	0.020	≤0.001
ACCTCCCCAC	Hs.529117	CYP2B7P1	Cytochrome P450, family 2, subfamily B, polypeptide 7 pseudogene 1	19q13.2	1	41	0.010	≤0.001
CAGTGCTTTT	Hs.551178	Hs.551178	CDNA FLJ46627 fis, clone TRACH2010272		1	60	0.010	≤0.001
GAGATTATGT	Hs.551521	KCNE2	Potassium voltage-gated channel, Isk-related family, member 2	21q22.12	5	55	0.050	≤0.001
TGTACCTCAG	Hs.558365	ORM2	Orosomucoid 2	9q32	1	25	0.020	≤0.001
TCATTCTGAA	Hs.69319	GKN1	Gastrokine 1	2p13.3	51	3592	0.010	≤0.001
AATGTCCCCA	Hs.76253	ATXN2	Ataxin 2	12q24.1	2	37	0.030	≤0.001
TTAACCCCTC	Hs.78224	RNASE1	Ribonuclease, RNase A family, 1 (pancreatic)	14q11.2	26	219	0.070	≤0.001

T4, tag number in all tumor samples tested; N4, tag number in all normal samples. The expression of all genes was significantly altered in at least three tumor samples ($P \leq 0.05$), as compared to all normal samples. At least two tumors showed more than fivefold change ($P \leq 0.01$). Tags with "0" value were replaced with arbitrary 0.5 values for relative calculation of fold expression. The ratio was calculated after normalization to total tag numbers.

Figure 1. Chromosomal localization of deregulated genes. Chromosomal regions that contain up-regulated genes are shown in red, whereas those that contain down-regulated genes are displayed in green. Regions which contain both up- and down-regulated genes are colored in yellow. The distribution of these genes did not follow a random distribution pattern and several genomic regions contain clusters of deregulated genes. Some of the more significant "hot spots" can be seen here on chromosomes 1 ($P = 0.01$), 3 ($P = 0.02$), 12 ($P = 0.01$), 15 ($P = 0.01$), and 19 ($P = 0.01$).

haults et al., 2001; El-Rifai et al., 2002) 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 E_n is the threshold cycle number for the experimental gene observed in the normal sample. R_n and E_n values were averages of the corresponding normal analyzed samples. The relative fold expression with standard error of mean (±SEM) is shown in Figure 2.

Immunohistochemistry

Immunohistochemical (IHC) analysis of ANPEP protein expression was performed on a tumor tissue microarray (TMA) that contained 65 adenocarcinomas. Samples from adjacent normal and dysplastic tissues were included when available. All tissue samples were histologically verified, and representative regions were selected for inclusion in the TMA. All of the adenocarcinoma samples were collected from either the GEJ or lower esophagus and ranged from WD to PD, Stages I–IV, with a mix of intestinal- and diffuse-type tumors. Tissue cores with a diameter of 0.5 mm were retrieved

TABLE 2. Chromosomal Minimal Common Overlapping Regions of Transcription Hot Spots

Minimal common overlapping regions	Number of genes	Gene symbols
Overexpressed genes		
1q21	13	S100A16, S100A2, S100A7, S100A9, S100A8, ECM1, S100A10, S100A6, LMNA, SPRR3, HDGF, HIST2H2BE, TAGLN2
6p21	6	HSPA1A, HLA-A, HSPA1B, HLA-C, RPL10A, CLIC1
8q24-qter	4	AW103351, LY6D, LY6E, FLJ32440
11q13	4	FTH1, CCND1, DKFZP761E198, TNCRNA
12p13	9	GAPD, C1R, C1S, PHB2, MLF2, PTMS, FLJ22662, NDUFA9, CD9
14q32.3	4	CRIP2, C14ORF173, CRIP1, IGHG1
17q21	4	KRT17, PPP1R1B, GRN, COL1A1
17q25	4	LGALS3BP, MRPL12, ACTG1, NT5C
19q13.4	5	RPS9, RPS5, LENG8, CDC42EP5, Hs.534672
20q13	5	PI3, PPGB, TMEPAI, C20ORF149, GATA5
22q13	7	RPL3, Hs.102336, CDC42EP1, LGALS1, ATXN10, PLXNB2, ECGF1
Downregulated genes		
4q21	4	IGJ, CCNI, SEC31L1, CDS1
19q13.1	4	UNQ473, CYP2B7P1, FCGBP, ATP4A
21q22	4	KCNE2, CLIC6, TFF1, TFF2

TABLE 3. Chromosomal Location of Frequent Gene Alterations in Barrett's Adenocarcinomas

	Upregulated transcripts = 242			Downregulated transcripts = 153			
Chromosome	p arm	q arm	Total	p arm	q arm	Total	Grand total
1	15	20	35 (0.01)[a]	10	11	21 (0.35)	56
2	7	10	17 (0.2)	4	8	12 (0.39)	29
3	3	4	7 (0.13)	1	2	3 (0.06)	10
4	1	4	5 (0.1)	3	8	11 (0.02)	16
5	0	8	8 (0.26)	2	4	6 (0.4)	14
6	8	2	10 (0.38)	3	1	4 (0.2)	14
7	3	3	6 (0.08)	3	5	8 (0.12)	14
8	2	6	8 (0.27)	2	3	5 (0.37)	13
9	1	7	8 (0.46)	0	8	8 (0.29)	16
10	5	7	12 (0.27)	3	6	9 (0.28)	21
11	5	9	14 (0.3)	1	5	6 (0.11)	20
12	10	11	21 (0.01)	1	8	9 (0.04)	30
13	NA	3	3 (0.36)	NA	2	2 (0.24)	5
14	NA	10	10 (0.27)	NA	4	4 (0.17)	14
15	NA	8	8 (0.01)	NA	5	5 (0.19)	13
16	3	3	6 (0.11)	2	4	6 (0.07)	12
17	4	8	12 (0.3)	1	5	6 (0.22)	18
18	4	0	4 (0.3)	1	0	1 (0.44)	5
19	8	18	26 (0.01)	3	4	7 (0.37)	33
20	1	8	9 (0.26)	2	3	5 (0.41)	14
21	NA	2	2 (0.23)	NA	4	4 (0.05)	6
22	NA	8	8 (0.45)	NA	2	2 (0.2)	10
X	2	1	3 (0.07)	4	5	9 (0.08)	12
Y	0	0	NA	NA	0	NA	0

A total of 568 transcripts were up- or down-regulated with statistical significance in which 395 known gene symbols were identified. In order to investigate and find statistically significant hot spots, the location of altered genes was compared with the list of all genes that are transcribed in both tumor and normal samples. The analysis was performed using Onto-Express online software (http://vortex.cs.wayne.edu/index.htm).
[a]Values in parentheses are *P* values.

from the selected regions of the donor blocks and punched to the recipient block using a manual tissue array instrument (Beecher Instruments, Silver Spring, MD). Each tissue sample was represented by four tissue cores on the TMA. Sections (5 μm) were transferred to polylysine-coated slides (SuperFrostPlus, Menzel-Gläser, Braunschweig, Germany) and incubated at 37°C for 2 hr. The resulting TMA was used for IHC analysis utilizing a 1:50 dilution of ANPEP antibody (CD13/aminopepti-

dase-N Ab-3 mouse monoclonal antibody; Lab Vision Corporation, Fremont, CA). Sections were deparaffinized and rehydrated. TMA slides were treated in a microwave with citrate buffer for 20 min and incubated with the antibody at room temperature. Detection was performed using an avidin–biotin immunoperoxidase assay. Cores with no evidence of staining, or only rare scattered positive cells less than 3%, were recorded as negative. The overall intensity of staining was recorded as that for the core with the strongest intensity. IHC results were evaluated for intensity and frequency of staining. The intensity of staining was graded as 0 (negative), 1 (weak), 2 (moderate), and 3 (strong). The frequency was graded from 0 to 4 by percentage of positive cells as follows: Grade 0, <3%; Grade 1, 3–25%; Grade 2, 25–50%; Grade 3, 50–75%; Grade 4, >75%. The index score was the product of multiplication of the intensity and frequency grades, which was then classified into a 4-point scale: index score 0 = product of 0, index score 1 = products 1 and 2, index score 2 = products 3 and 4, index score 3 = products 6 through 12.

RESULTS

Sequence Analyses of SAGE Libraries

Sequence analyses of 20,000 clones from eight SAGE libraries produced 457,894 expressed tags, with 32,035 tags (6.9%) accounting for singleton tags. The four tumor SAGE libraries (GSM758, GSM757, HG7, and HS29) produced 287,219 tags with an average of 71,804 tags per library. The normal samples (GSM14780, GSM784, 13S, and 14S) produced 170,675 tags with an average of 42,669 tags per library. The comparison of expressed tags to the UniGene cluster release of May 2005 identified 67,200 unique SAGE tags. These tags represented 16,040 known gene symbols according to UniGene information. Of these, 568 unique tags were differentially expressed between BAs and normal tissue samples (at least twofolds and $P \leq 0.05$). These unique tags matched 395 known genes (242 upregulated and 153 downregulated) that regulate diverse cellular functions and signaling pathways, which may prove to be quite significant in the detection and prevention of cancer. Ninety-three genes were significantly altered, showing a greater than fivefold expression change in at least two tumor libraries as compared to all four normal libraries ($P \leq 0.01$) (Table 1). Forty-eight genes showed up-regulation, whereas 45 were down-regulated. The group of over-expressed genes contained several with known cancer-related functions, including members of S100A calcium-binding proteins, heat-shock protein 27 kDa (*HSB1*), heat-shock 90 kDa protein beta (*HSPCB)*, prothymosin (*PTMA*), transmembrane bax inhibitor motif containing-1 (*PP1201*), peroxiredoxin-3 (*PRDX3*), and endothelial growth factor-1 (*ECGF1*). Down-regulated transcripts included genes such as gastrokine (*GKN1*), down-regulated in gastric cancer (*GDDR*), gastric intrinsic factor (*GIF*), methyl-CpG binding domain protein 3 (*MBD3*), and trefoil factor 2 (*TFF2*). CGAP maintains the public SAGE database for gene expression in human cancer (Lal et al., 1999), and sequence data are publicly available at http://www.ncbi.nih.gov/geo and http://cgap.nci.nih.gov/SAGE/.

Transcriptional Oncogenomic Hot Spots and Functional Classification of Genes

Onto-Express online software (http://vortex.cs.wayne.edu/index.htm) (Khatri et al., 2002; Draghici et al., 2003) was used to identify potential transcriptional oncogenomic hot spots in the genome and obtain the functional classification of the deregulated genes. We mapped all SAGE unique transcripts (16,040 gene symbols) to their corresponding cytogenetic locations. The altered transcripts (395 known gene symbols) were analyzed against all transcripts to generate an expression ideogram and identify transcription hotspots (Fig. 1). Interestingly, the distribution of altered genes was not uniform along the human chromosomes. Overexpressed genes tended to cluster in well-defined hot spots across the human genome (Table 2). For example, 26 overexpressed genes mapped to chromosome 19, of which 18 mapped to the single chromosome band 19q13. Similarly, 35 genes mapped to chromosome 1, of which 13 mapped to the chromosome band 1q21. Table 3 and Figure 1 summarize these data and map the genes to their corresponding cytogenetic locations.

Gene ontology (GO) terms are organized in three general categories: biological process, cellular role, and molecular function; terms within each GO category are linked in defined parent–child relationships that reflect current biological knowledge (Ashburner et al., 2000). Among the 395 differentially expressed genes, the number corresponding to each category was tallied and compared with the number expected for each GO category based on its representation on the reference gene list, which contained all of the unique 16,040 known gene symbols detected by analysis of the eight SAGE libraries. Significant differences

TABLE 4. Functional Classification of Deregulated Genes in Barrett's Related Adenocarcinomas Using Gene Ontology (GO)

Gene symbol	Ratio	Gene symbol	Ratio	Gene symbol	Ratio	Gene symbol	Ratio
Cell cycle regulation[a]							
ALS2CR19	0.13	DUSP6	27.38	IGFBP7	3.14	PTMA	10.71
AURKAIP1	27.38	EMP1	10.27	ILK	27.38	PTMS	6.19
CRIP1	4.17	GKN1	0.01	LGALS1	105.95	S100A6	3.83
BTG1	0.31	GRN	4.63	MACF1	6.07	SFN	42.86
CCND1	32.14	HDGF	33.33	MDK	10.12	TIMP1	9.97
CDKN2A	27.38	HIF3A	5.21	MTSS1	0.17	TM4SF4	11.31
CHEK1	4.03	IFITM1	23.21	PPP2R1B	23.21	TSPAN1	0.01
DNA binding and replication[b]							
ABCB7	0.02	CTGF	22.62	HIST2H2BE	28.57	PTMS	6.19
ABCC1	61.9	CUGBP2	0.02	HSPA1B	11.61	RAB40C	71.43
ACTA1	20.24	DUT	0.04	ILK	27.38	RBM17	0.09
ACTB	4.5	ECGF1	54.76	MAST4	0.01	RHOD	26.19
ACTG1	3.06	EEF2K	0.03	MBD3	0.02	ROD1	28.57
ARF1	28.57	EIF5A	8.52	MYH9	0.02	SERPINA3	74.4
ATP1A1	14.05	ELF3	38.1	NCL	25	SET	0.29
ATP4A	0.02	ENO1	9.23	NT5C	2.52	WNK1	0.02
PTBP1	0.23	EPHA4	0.03	OBFC2A	0.23	YBX1	22.62
CDKN2A	27.38	GNAI2	15.18	PFKP	8.23	ZFHX1B	0.26
CHD2	0.07	GNAS	0.02	PPP2R1B	23.21	ZNF480	30.95
CHEK1	4.03	HDLBP	28.57				
RNA binding[c]							
CUGBP2	0.02	NCL	25	RNASE1	0.07	RPS5	3.07
EIF1AX	0.16	PTBP1	0.23	ROD1	28.57	SERBP1	4.32
HDLBP	28.57	RBM17	0.09	RPL18	5.7	SNRPB	9.33
MRPL12	15.48	RBM19	0.03	RPL3	21.73	YBX1	22.62
Transcription[d]							
ZFHX1B	0.26	FOXA2	0.11	NT5C	2.52	RPLP0	19.05
ZFP36L1	41.67	FOXD4L1	32.14	CDKN2A	27.38	EIF3S1	28.57
ELF3	38.1	LASS6	0.16	NMI	339.29	HSPB1	14.88
EEF1B2	0.37	RAI17	25	PTBP1	0.23	BTG1	0.31
AES	3.79	TCF7L1	0	ROD1	28.57	PPP2R1B	23.21
ENO1	9.23	TIMELESS	0.36	SNRPB	9.33	ESRRG	0.05
HIF3A	5.21	YBX1	22.62	HSPA1B	11.61	PCBD2	0.36
MBD3	0.02	ZNF480	30.95	EIF1AX	0.16	GATA5	48.81
PHB2	9.33	CHD2	0.07	EIF5A	8.52		
PTMA	10.71	JUND	12.2	EEF2K	0.03		
Receptor related[e]							
ANPEP	90.48	F3	19.05	INTS6		PHB2	9.33
ANXA1	4.6	GNB2L1	34.52	ITGB1	4.84	PLXNB2	8.81
ARF1	28.57	GPR68	0.16	LGALS3BP	47.62	SLAMF7	46.43
OPRL1	0.02	HSPA1A	55.95	LRP1B	38.1		
DRD5	0.02	IFITM1	23.21	MTSS1	0.17		
EPHA4	0.03	IL6ST	4.06				
Calcium ion binding[f]							
ACTN4	10	EEF2K	0.03	MRLC2	3.71	S100A7	113.1
ANXA1	4.6	EFHD2	11.31	PADI1	42.86	S100A8	204.17
ANXA10	0.24	ITGB1	4.84	PRKCSH	29.76	S100A9	422.62
ANXA11	16.67	ITPR3	0.22	REPS2	31.85	SPARC	4.31
C1R	24.4	LRP1B	38.1	S100A10	4.16	SVIL	250
C1S	19.05	MACF1	6.07	S100A16	72.62	TKT	35.71
CLTB	10.32	MMP11	14.58	S100A2	72.62	VMD2L3	27.38
CSPG2	27.38	MRCL3	4.76	S100A6	3.83		
Zinc ion binding[g]							
ALPPL2	34.52	CRIP2	25	MMP11	14.58	S100A7	113.1
ANPEP	90.48	ESRRG	0.05	MT1F	0.17	TRIM2	0.18
RAI17	25	GATA5	48.81	PARK2	0.02	ZFHX1B	0.26
CA2	0.26	GIT2	27.38	PDLIM1	15.48	ZFP36L1	41.67
CPA2	0.01	HERC2	36.9	PDLIM7	46.43	ZNF480	30.95
CRIP1	4.17	HINT1	24.4				

(Continued)

TABLE 4. Functional Classification of Deregulated Genes in Barrett's Related Adenocarcinomas Using Gene Ontology (GO) (Continued)

Gene symbol	Ratio	Gene symbol	Ratio	Gene symbol	Ratio	Gene symbol	Ratio
Cell signaling[h]							
ADCYAP1	50.6	EPHA4	0.03	IL6ST	4.06	PDIA3	24.12
ANXA1	4.6	FKBP8	41.67	ILK	27.38	PPP1R1B	40.48
ARF1	28.57	FMOD	0.17	ITGB1	4.84	PRKCSH	29.76
WNT4	0.03	GAST	0	ITPR3	0.22	PRMT1	30.95
BSG	11.46	GHRL	0.06	LGALS3BP	47.62	PYCR2	47.62
BTRC	7.54	GNAS	0.02	LY6E	7.29	RAB40C	71.43
C1S	19.05	GNB2L1	34.52	MDK	10.12	REPS2	31.85
C9orf86	25	GPR68	0.164	MKLN1	6.45	RHOD	26.19
CDS1	0.01	GRN	4.63	MTSS1	0.17	SFN	42.86
CEACAM6	8.57	HDGF	33.33	MYH9	0.02	SNX6	34.52
DRD5	0.02	HINT1	24.4	NMI	339.29	SPARC	4.31
ECGF1	54.76	IFITM1	23.21	OPRL1	0.02		
Inflammation[i]							
ANXA1	4.6	LGALS3BP	47.62	PDLIM1	15.48	SERPINA3	74.4
CYBB	0.018	LY6E	7.29	PRMT1	30.95	TFF1	0.32
GPR68	0.164	MLF2	6.94	PTMS	6.19	TFF2	0.03
GPX1	9.92	NMI	339.29	S100A8	204.17		
IL1RN	7.94	ORM2	0.024	S100A9	422.62		
Cell environment interaction[j]							
ACTN4	10	ECGF1	54.76	LY6D	45.83	S100A6	3.83
ADCYAP1	50.6	EMILIN1	26.19	MDK	10.12	S100A9	422.62
ANPEP	90.48	ENAH	0.01	MKLN1	6.45	SLAMF7	46.43
ANXA1	4.6	FCGBP	0.18	MTSS1	0.17	SPON2	6.67
BTG1	0.31	GRN	4.63	PGM5	0.09	TSPAN1	0.01
CD9	9.52	IL32	17.86	PPFIBP2	0.05	WNT4	0.03
CEACAM6	8.57	KLK6	35.71	PPP2R1B	23.21		
CTGF	22.62	LGALS3BP	47.62	PYCR2	47.62		

The average ratio is shown. This ratio was calculated by comparing the total number of tags in tumor samples and normal samples.
[a]Examples: GO: 0007049 cell cycle, GO: 0008283 cell proliferation, and GO: 0006915 apoptosis.
[b]Examples: GO: 0000166 nucleotide binding, GO: 0003677 DNA binding, and GO: 0006260 DNA replication.
[c]Examples: GO: 0003723 RNA binding and GO: 0003730 mRNA 3′-UTR binding.
[d]Examples: GO: 0003700 transcription factor activity, GO: 0006350 transcription, and GO: 0006355 DNA dependent regulation of transcription.
[e]Examples: GO: 0004872 receptor activity, GO: 0005102 receptor binding, and GO: 0005057 receptor signaling protein activity.
[f]Examples: GO: 0005509 calcium ion binding.
[g]Examples: GO: 0008270 zinc ion binding.
[h]Examples: GO: 0007165 signal transduction, GO: 0007166 cell surface receptor linked signal transduction, and GO: 0007186 G-protein coupled receptor protein signaling pathway.
[i]Examples: GO: 0006952 defense response and GO: 0006954 inflammatory response.
[j]Examples: GO: 0006928 cell motility, GO: 0007155 cell adhesion, and GO: 0007267 cell–cell signaling.

from the expected were calculated with a two-sided binomial distribution. False discovery rates (Benjamini et al., 2001) and Bonferroni adjustments were also calculated. The biological meaning of the *P* values obtained depends upon the list of genes that are submitted; as our gene list is from a comparison of BA samples, it can be inferred that this cancer stimulates the processes involved within the functional groups that were most highly represented in the results of the GO classification. In our set of differentially expressed genes, the functional groups demonstrating the most significant representation appear under the biological-process ontology and map to the cell-cycle regulation, DNA binding and regulation, cell–environment interaction, and cell-signaling categories. Table 4 summarizes several important GO functional classes.

Validation of Transcriptional Targets

To evaluate further the SAGE data, we selected five novel genes (*ANPEP*, *ECGF1*, *PP1201*, *EIF5A1*, and *GKN1*, all of which have important cellular or biological features) for validation with qRT-PCR. We confirmed over-expression of *ANPEP*, *ECGF1*, *PP1201*, and *EIF5A1* and down-regulation of *GKN1* in primary GEJ and lower esophageal adenocarcinoma samples (Table 5, Fig. 2). Interestingly, GKN1 was not expressed in normal esophageal mucosa samples but showed a transient expression in BE samples where 4/6 of these samples demonstrated expression levels com-

TABLE 5. Summary of qRT-PCR Results

	Overexpressed genes				Downregulated gene
	EIF51	ECGF1	ANPEP	PP1201	GKN1
All cases	9/31 (29)[a]	15/31 (48)	14/31 (45)	15/31 (48)	30/31 (97)
Gender					
Male	4/19 (21)	8/19 (42)	10/19 (53)	14/19 (74)	19/19 (100)
Female	2/4 (50)	3/4 (75)	1/4 (25)	1/4 (25)	4/4 (100
	3/8 (38)	4/8 (50)	3/8 (38)	0/8 (0)	7/8 (88)
Site					
GEJ	4/10 (40)	7/16 (44)	7/16 (44)	10/16 (63)	16/16 (100)
ESO	3/10 (30)	4/10 (40)	4/10 (40)	5/10 (50)	10/10 (100)
NA	2/5 (40)	4/5 (80)	3/5 (60)	0/5 (0)	4/5 (80)
Stage					
T1–T2	2/8 (25)	3/8 (37)	5/8 (62)	6/8 (75)	8/8 (100)
T3–T4	5/14 (36)	7/14 (50)	5/14 (36)	8/14 (57)	14/14 (100)
NA	3/9 (33)	5/9 (55)	4/9 (44)	1/9 (11)	8/9 (89)
Grade					
WD-MD	3/10 (30)	5/10 (50)	5/10 (50)	8/10 (80)	10/10 (100)
PD	2/9 (22)	4/9 (44)	5/9 (56)	6/9 (67)	9/9 (100)
NA	4/12 (33)	6/12 (50)	4/12 (33)	1/12 (8)	11/12 (92)
Node					
N0	2/8 (25)	2/8 (25)	5/8 (63)	6/8 (75)	8/8 (100)
N1–N2	4/13 (31)	7/13 (54)	4/13 (31)	7/13 (54)	13/13 (100)
N3–N4	0/0 (0)	0/0 (0)	0/0 (0)	0/0 (0)	0/0 (0)
NA	3/10 (30)	6/10 (60)	5/10 (50)	2/10 (20)	9/10 (90)

[a]Values in parentheses are percentages.
NA, information not available; GEJ, gastroesophageal junction; ESO, esophageal; WD, well-differentiated; MD, moderately-differentiated; PD, poorly differentiated. We did not observe statistical significance with any of the correlates due to small sample size.

parable to those observed in normal gastric mucosae. We did not have samples with Barrett's dysplasia for qRT-PCR. The *GKN1* expression was lost in almost all adenocarcinoma samples (Fig. 2). The qRT-PCR products were run on 1.2% agarose gels for visual confirmation of these results (Fig. 3). RT-PCR results for all five genes were also compared in each individual primary tissue sample to determine any correlations in combined gene expression levels; however, we were unable to find any correlations of statistical significance.

Expression of ANPEP in Tumor TMA

The IHC analysis demonstrated a lack of immunostaining for ANPEP in normal esophageal and gastric epithelial tissues. On the other hand, BAs showed overexpression of ANPEP (Score +1 to +3) in 35/65 (54%) tumors. A weak to moderate expression of ANPEP (Score +1 to +2) was observed in 6/7 (86%) high-grade Barrett's dysplasia samples. The immunostaining pattern of ANPEP was cytoplasmic with strong extracellular and luminal expression (Fig. 4). The immunostaining for ANPEP was observed in tumors with intestinal and diffuse histological subtypes and in all stages (Table 6). However, the relatively small sample size did not provide a sufficient statistical power to detect significant correlations between the IHC staining patterns and clinicopathological factors such as tumor histology, grade, or stage.

DISCUSSION

In this study, we performed a comprehensive analysis of the transcriptome of BAs using SAGE. The major advantage to using SAGE is the quantitative ability to evaluate accurately transcript numbers without prior sequence information. The SAGE analysis produced a great deal of information about transcripts and candidate cancer genes, and we have interpreted these data in terms of possible genomic and functional organization of candidate cancer genes.

SAGE analysis requires laborious and extensive sequencing that often limits the number of samples that are subjected to analysis. We obtained a total of 457,894 expressed tags from eight SAGE libraries with minimal singleton tags (32,035; 6.9%). The qRT-PCR analysis on a larger sample size confirmed the SAGE results and validated the overexpression of *ANPEP*, *ECGF1*, *PP1201*, and *EIF5A1* and downregulation of *GKN1*. *ECGF1* (thymidine phosphorylase) expression has been shown to correlate with the angiogenic activity of some tumors (Mazurek et al., 2006). *ECGF1* expression may be a sign of tumor-stromal interac-

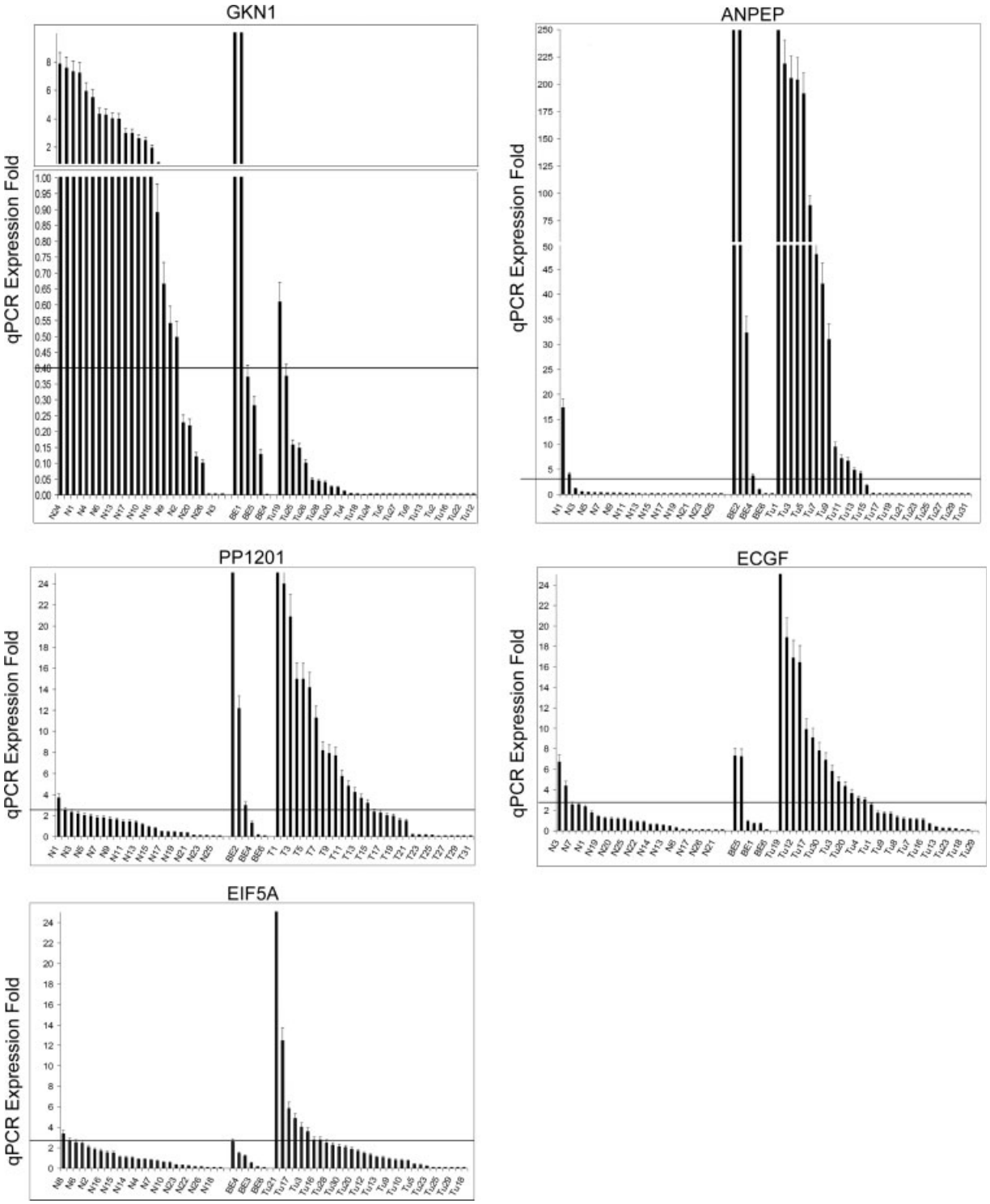

Figure 2. Quantitative real-time reverse-transcription PCR showing fold expression changes at the mRNA level of five representative genes. qRT-PCR analysis was performed using iCycler on 31 lower esophageal and GEJ adenocarcinoma samples (Tu) and 6 Barrett's esophagus (BE) samples in comparison with 26 normal glandular mucosa samples (N). The horizontal axis shows sample numbers, whereas the fold expression in tumor samples compared with that in normal samples is shown on the vertical axis. The fold expression was calculated according to the formula: $2^{(R_t-E_t)}/2^{(R_n-E_n)}$ as detailed in the "Materials and Methods" section. Each bar represents one sample. The displayed mean fold expression for each sample is calculated in comparison with the expression average of the 26 normal samples. The expression of each gene was normalized to the expression of *HPRT1*, which showed minimal variation in all normal and neoplastic samples tested. *GKN1* shows downregulation ($\leq$0.4-fold expression) whereas *ANPEP*, *PP1201*, *EIF5A1*, and *ECGF1* demonstrate overexpression ($\geq$2.5 fold expression) in primary tumors as compared to normal tissue samples.

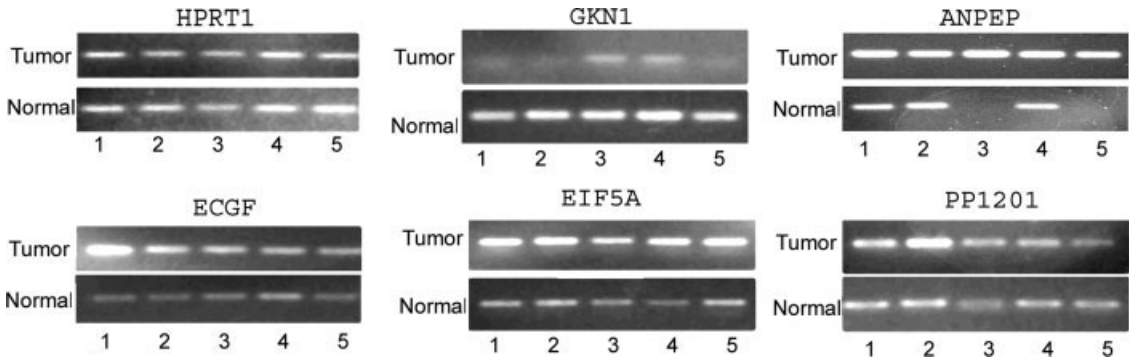

Figure 3. Visualization of RT-PCR products on gel electrophoresis. Five matched tumor and normal samples that were analyzed using qRT-PCR were subjected to 1.2% agarose gel electrophoresis and ethidium bromide staining. The intensity of bands confirms the PCR results, indicating higher mRNA expression levels of *ANPEP*, *PP1201*, *EIF5A1*, and *ECGF*, as well as lower expression of *GKN1* in most of the tumor samples as compared with their matched normal control samples. *HPRT1* was used as a control to show similar levels in each matched normal and tumor samples.

tion promoting greater vascularization around the cancer lesion and has also been found to protect cells from DNA-damaging agents and related apoptosis (Jeung et al., 2006). *EIF5A1* (eukaryotic translation factor 1) has been shown to be involved in cell proliferation through the action of polyamines (Nishimura et al., 2002, 2005), and plays a role in the regulation of TP53-related apoptosis (Li et al., 2004). PP1201, also known as transmembrane Bax inhibitor motif-containing 1 (*TMBIM1*), is a novel gene of cancer cells. Although very little is known regarding *GKN1*, it has been previously reported as highly expressed in normal gastric epithelium (Martin et al., 2003) and down-regulated in gastric carcinomas (Oien et al., 2004). We have detected strong expression of *GKN1* in BE that was followed with loss of its expression in adenocarcinomas. This transient expression of *GKN1* may be a protective response to acid-induced reflux-disease injury that is the lost with cellular progression to cancer. ANPEP, also known as CD13, is of a particular clinical interest since it is a secreted protein that may be used as a potential biomarker. Using IHC, analysis of ANPEP expression demonstrated protein expression at the outer cell membrane layers with significant secretion into the lumen of 6/7 Barrett's high-grade dysplasia samples and generally greater expression in 35/65 adenocarcinomas, suggesting that ANPEP overexpression may be an early event in carcinogenesis. *ANPEP* expression plays a role in angiogenesis where a reduction in expression has been shown to cause reduced capillary formation (Fukasawa et al., 2006), cell motility (Chang et al., 2005), and adhesion (Fukasawa et al., 2006). Inhibition of ANPEP decreases the invasive potential of metastatic tumor cells in vitro (Saiki et al., 1993). Interestingly, ANPEP is also a cell-surface metalloproteinase that acts as a receptor for human coronavirus (Yeager et al., 1992) and is considered to be a marker for epithelial–mesenchymal interaction (Sorrell et al., 2003).

The combination of transcriptional analysis together with cytogenetic information provided a powerful tool to align altered transcripts across the human genome. Interestingly, the distribution of deregulated genes did not follow a uniform pattern across the genome. Instead, we found a remarkable pattern of distribution with the presence of transcriptional hot spots along chromosomal domains. From this pattern, we were able to identify novel, transcriptionally active, and oncogenomic hot spots. One of our surprising findings was the clustering of 26 overexpressed genes in one of the smallest human chromosomes, 19. We also identified a number of other hot spots, such as 1q21 (13 genes), 12p13 (9 genes), and 6p21.2 (6 genes) (Table 2) in a recent analysis of amplification-based clustering demonstrated that cancers with similar etiology, cell-of-origin, or topographical location have a tendency to obtain convergent amplification profiles (Myllykangas et al., 2006). In line with this observation, Vogel et al. (2005) reported that genes expressed in concert are organized in a linear arrangement for coordinated regulation. The present evidence suggests organization of a large proportion of the human transcriptome into gene clusters throughout the genome, which are partly regulated by the same transcription factors, share biological functions, and are characterized by non-housekeeping genes (Vogel et al., 2005). Taken together, our results further highlight the complex organization of the cancer genome and suggest that integrated analysis of the transcriptome may reveal similar findings in other tumors as well.

Each cancer candidate gene was assigned to a functional group based on GO information (Table 4).

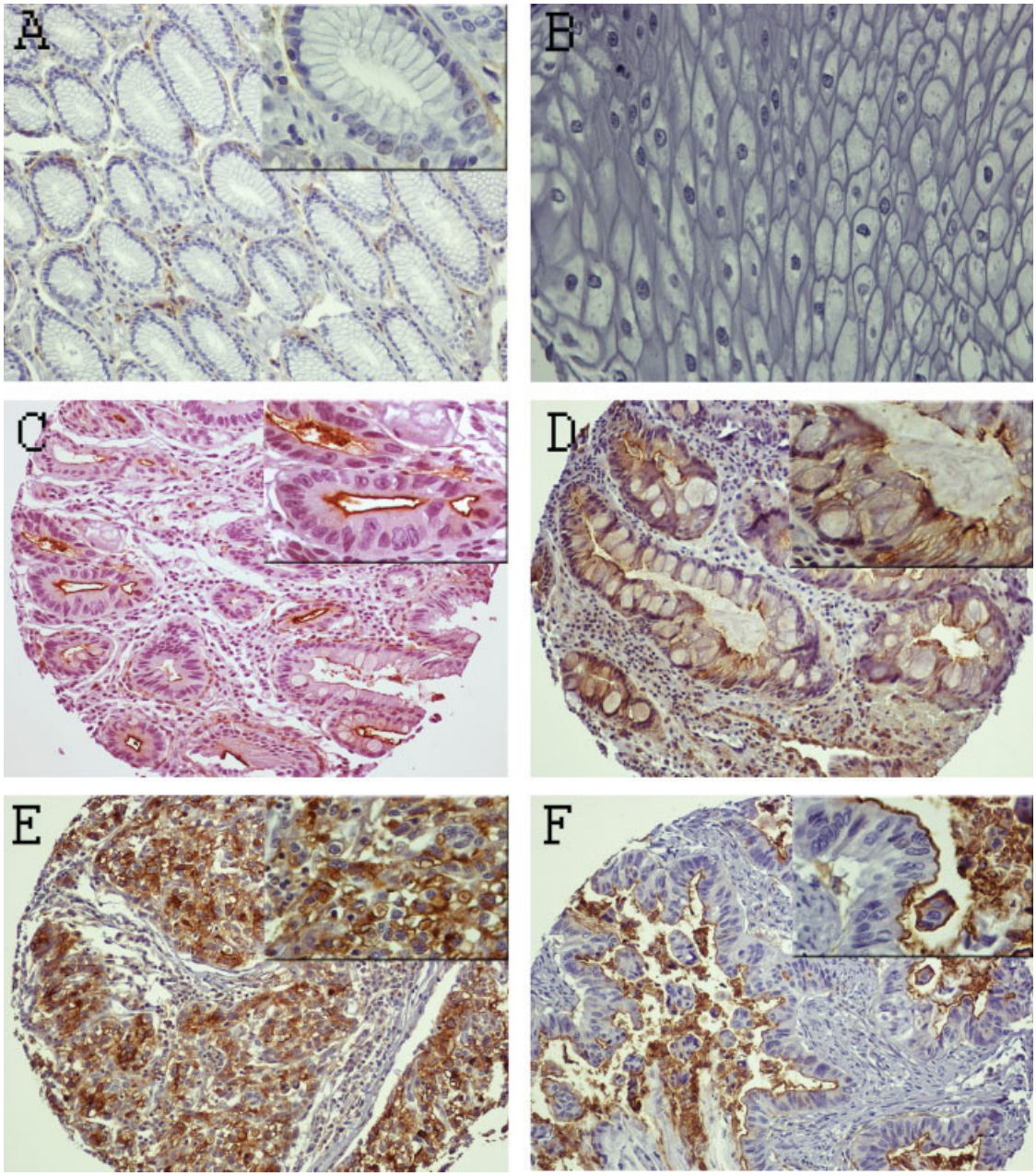

Figure 4. Immunohistochemical staining for ANPEP. (A, B) Normal gastric tissue glands (A) and normal esophageal squamous tissues (B) are negative for ANPEP immunostaining (Score 0). (C) Barrett's dysplastic tissue demonstrates immunostaining for ANPEP that is secreted in the lumen (Score +2). (D) Barrett's metaplasia tissue shows glandular staining (Score +2). (E) Diffuse-type esophageal adenocarcinoma tissue shows staining for ANPEP in the cell cytoplasm with significant localization along the cell membranes (Score +3). (F) Intestinal-type esophageal adenocarcinoma tissue showing high levels of ANPEP along the cell membranes as well as luminal secretion (Score +3). All photos (insets at upper-right quadrant) are taken at 200× and 400× magnification.

Using this approach, several groups that are highly interesting and relevant to carcinogenesis were identified including transcriptional regulators (38 genes) and zinc finger transcription factors (23 genes). Similarly, several candidate genes were found to be involved in the notable functional groups of cell-environment interaction and signal transduction. Subsets of these groups were of interest and included metalloproteinases and G proteins and their regulators. Among the interesting groups, we also observed deregulation of 31 genes that regulate cell calcium homeostasis. The role of calcium-binding proteins in carcinogenesis has drawn a complex picture showing downregulation or overexpression depending upon the tumor type and location (Kao et al., 1990; Mueller et al., 1999;

TABLE 6. Summary of Immunohistochemistry Analysis of ANPEP on Tissue Microarrays

	IHC score				
	0	1	2	3	Total
All cases	30 (46)[a]	21 (32)	6 (9)	8 (12)	65 (100)
Gender					
Male	22 (73)	16 (76)	6 (100)	7 (88)	51 (78)
Female	2 (7)	2 (10)	0 (0)	1 (13)	5 (8)
NA	5 (17)	3 (14)	0 (0)	0 (0)	8 (13)
Site					
GEJ	11 (37)	8 (38)	3 (50)	6 (75)	28 (43)
ESO	15 (50)	11 (52)	3 (50)	2 (25)	31 (48)
NA	3 (10)	2 (10)	0 (0)	0 (0)	5 (8)
Histology					
Diffuse	10 (33)	7 (33)	0 (0)	2 (25)	19 (29)
Intestinal	19 (63)	14 (67)	6 (100)	6 (75)	45 (69)
Stage					
T1–T2	6 (20)	10 (48)	2 (33)	1 (13)	19 (29)
T3–T4	15 (50)	6 (29)	3 (50)	4 (50)	28 (43)
NA	8 (27)	5 (24)	1 (17)	3 (38)	17 (26)
Grade					
WD	3 (10)	3 (14)	1 (17)	0 (0)	7 (11)
MD	4 (13)	5 (24)	2 (33)	2 (25)	13 (20)
PD	19 (63)	13 (62)	3 (50)	6 (75)	41 (63)
Node					
N0	18 (60)	10 (48)	4 (67)	2 (25)	34 (52)
N1–N2	3 (10)	8 (38)	1 (17)	4 (50)	16 (25)
N3–N4	1 (3)	0 (0)	0 (0)	0 (0)	1 (2)
NA	7 (23)	3 (14)	1 (17)	2 (25)	13 (20)

NA, information not available; GEJ, gastroesophageal junction; ESO, esophageal; WD, well-differentiated; MD, moderately-differentiated; PD, poorly differentiated. We did not observe statistical significance with any of the correlates due to small sample size.
[a]Values in parentheses are percentages.

Heighway et al., 2002; Heizmann et al., 2002; Imazawa et al., 2005). The SAGE data also indicated up-regulation of several members of the protein phosphatases such as *PPAP2B*, *HIF3A*, and *PPP2R1B* that are known to regulate and activate several cellular kinases (Parsons, 1998; Nigg, 2001; Bakkenist and Kastan, 2004; Ventura and Nebreda, 2006). We have recently shown that over-expression of *PPP1R1B* in gastrointestinal cancers is associated with several oncogenic properties including the resistance of cancer cells to drug-induced apoptosis (Belkhiri et al., 2005). Taken together, our data suggest a genomic organization of cancer genes, which are involved in the deregulation of specific cellular processes important for the tumorigenesis cascade.

In conclusion, our findings indicate the presence of transcriptionally active oncogenomic hot spots in the cancer genome of BAs. We have detected deregulation of several important cancer genes and identified novel targets for carcinogenesis. The biological functions and clinical significance of these genes will be elucidated in future studies.

ACKNOWLEDGMENTS

We thank Mr. Frank Revetta for his technical assistance and Mrs. Sheryl Mroz for editing this manuscript.

REFERENCES

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, Hruban RH. 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.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29.

Bakkenist CJ, Kastan MB. 2004. Phosphatases join kinases in DNA-damage response pathways. Trends Cell Biol 14:339–341.

Belkhiri A, Zaika A, Pidkovka N, Knuutila S, Moskaluk C, El-Rifai W. 2005. Darpp-32: A novel antiapoptotic gene in upper gastrointestinal carcinomas. Cancer Res 65:6583–6592.

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. 2001. Controlling the false discovery rate in behavior genetics research. Behav Brain Res 125:279–284.

Blot WJ, Devesa SS, Fraumeni JF, Jr. 1993. Continuing climb in rates of esophageal adenocarcinoma: An update. JAMA 270:1320.

Brabender J, Lord RV, Wickramasinghe K, Metzger R, Schneider PM, Park JM, Holscher AH, DeMeester TR, Danenberg KD, Danenberg PV. 2002. Glutathione *S*-transferase-pi expression is downregulated in patients with Barrett's esophagus and esophageal adenocarcinoma. J Gastrointest Surg 6:359–367.

Buckhaults P, Rago C, St Croix B, Romans KE, Saha S, Zhang L, Vogelstein B, Kinzler KW. 2001. Secreted and cell surface genes expressed in benign and malignant colorectal tumors. Cancer Res 61:6996–7001.

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, Versteeg R. 2001. The human transcriptome map: Clustering of highly expressed genes in chromosomal domains. Science 291:1289–1292.

Chang YW, Chen SC, Cheng EC, Ko YP, Lin YC, Kao YR, Tsay YG, Yang PC, Wu CW, Roffler SR. 2005. CD13 (aminopeptidase N) can associate with tumor-associated antigen L6 and enhance the motility of human lung cancer cells. Int J Cancer 116:243–252.

Cremer T, Cremer C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2:292–301.

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

Dolan K, Garde J, Walker SJ, Sutton R, Gosney J, Field JK. 1999. LOH at the sites of the DCC, APC, and TP53 tumor suppressor genes occurs in Barrett's metaplasia and dysplasia adjacent to adenocarcinoma of the esophagus. Hum Pathol 30:1508–1514.

Draghici S, Khatri P, Martins RP, Ostermeier GC, Krawetz SA. 2003. Global functional profiling of gene expression. Genomics 81:98–104.

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

El-Rifai W, Harper JC, Cummings OW, Hyytinen ER, Frierson HF, Jr., Knuutila S, Powell SM. 1998. Consistent genetic alterations in xenografts of proximal stomach and gastro-esophageal junction adenocarcinomas. Cancer Res 58:34–37.

El-Rifai W, Frierson HF, Jr., Moskaluk CA, Harper JC, Petroni GR, Bissonette EA, Jones DR, Knuutila S, Powell SM. 2001. Genetic differences between adenocarcinomas arising in Barrett's esophagus and gastric mucosa. Gastroenterology 121:592–598.

El-Rifai W, Smith MF, Jr., Li G, Beckler A, Carl VS, Montgomery E, Knuutila S, Moskaluk CA, Frierson HF, Jr., Powell SM. 2002. Gastric Cancers Overexpress DARPP-32 and a Novel Isoform, t-DARPP. Cancer Res 62:4061–4064.

Ferraris R, Bonelli L, Conio M, Fracchia M, Lapertosa G, Aste H. 1997. Incidence of Barrett's adenocarcinoma in an Italian popula-

tion: An endoscopic surveillance programme. Gruppo Operativo per lo Studio delle Precancerosi Esofagee (GOSPE). Eur J Gastroenterol Hepatol 9:881–885.
Fukasawa K, Fujii H, Saitoh Y, Koizumi K, Aozuka Y, Sekine K, Yamada M, Saiki I, Nishikawa K. 2006. Aminopeptidase N (APN/CD13) is selectively expressed in vascular endothelial cells and plays multiple roles in angiogenesis. Cancer Lett 8:135–143.
Hamilton SR, Smith RR, Cameron JL. 1988. Prevalence and characteristics of Barrett esophagus in patients with adenocarcinoma of the esophagus or esophagogastric junction. Hum Pathol 19:942–948.
Heighway J, Knapp T, Boyce L, Brennand S, Field JK, Betticher DC, Ratschiller D, Gugger M, Donovan M, Lasek A, Rickert P. 2002. Expression profiling of primary non-small cell lung cancer for target identification. Oncogene 21:7749–7763.
Heizmann CW, Fritz G, Schafer BW. 2002. S100 proteins: Structure, functions and pathology. Front Biosci 7:d1356–d1368.
Hough CD, Sherman-Baust CA, Pizer ES, Montz FJ, Im DD, Rosenshein NB, Cho KR, Riggins GJ, Morin PJ. 2000. Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer. Cancer Res 60:6281–6287.
Imazawa M, Hibi K, Fujitake S, Kodera Y, Ito K, Akiyama S, Nakao A. 2005. S100A2 overexpression is frequently observed in esophageal squamous cell carcinoma. Anticancer Res 25:1247–1250.
Jeung HC, Che XF, Haraguchi M, Zhao HY, Furukawa T, Gotanda T, Zheng CL, Tsuneyoshi K, Sumizawa T, Roh JK, Akiyama S. 2006. Protection against DNA damage-induced apoptosis by the angiogenic factor thymidine phosphorylase. FEBS Lett 580:1294–1302.
Kao JP, Alderton JM, Tsien RY, Steinhardt RA. 1990. Active involvement of $Ca2^{+}$ in mitotic progression of Swiss 3T3 fibroblasts. J Cell Biol 111:183–196.
Khatri P, Draghici S, Ostermeier GC, Krawetz SA. 2002. Profiling gene expression using onto-express. Genomics 79:266–270.
Kuwano H, Kato H, Miyazaki T, Fukuchi M, Masuda N, Nakajima M, Fukai Y, Sohda M, Kimura H, Faried A. 2005. Genetic alterations in esophageal cancer. Surg Today 35:7–18.
Lal A, Lash AE, Altschul SF, Velculescu V, Zhang L, McLendon RE, Marra MA, Prange C, Morin PJ, Polyak K, Papadopoulos N, Vogelstein B, Kinzler KW, Strausberg RL, Riggins GJ. 1999. A public database for gene expression in human cancers. Cancer Res 59:5403–5407.
Li AL, Li HY, Jin BF, Ye QN, Zhou T, Yu XD, Pan X, Man JH, He K, Yu M, Hu MR, Wang J, Yang SC, Shen BF, Zhang XM. 2004. A novel eIF5A complex functions as a regulator of p53 and p53-dependent apoptosis. J Biol Chem 279:49251–49258.
Margulies EH, Innis JW. 2000. eSAGE: Managing and analysing data generated with serial analysis of gene expression (SAGE). Bioinformatics 16:650–651.
Margulies EH, Kardia SL, 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.
Martin TE, Powell CT, Wang Z, Bhattacharyya S, Walsh-Reitz MM, Agarwal K, Toback FG. 2003. A novel mitogenic protein that is highly expressed in cells of the gastric antrum mucosa. Am J Physiol Gastrointest Liver Physiol 285:G332–G343.
Mazurek A, Kuc P, Terlikowski S, Laudanski T. 2006. Evaluation of tumor angiogenesis and thymidine phosphorylase tissue expression in patients with endometrial cancer. Neoplasma 53:242–246.
Meltzer SJ, Yin J, Manin B, Rhyu MG, Cottrell J, Hudson E, Redd JL, Krasna MJ, Abraham JM, Reid BJ. 1994. Microsatellite instability occurs frequently and in both diploid and aneuploid cell populations of Barrett's-associated esophageal adenocarcinomas. Cancer Res 54:3379–3382.
Mueller A, Bachi T, Hochli M, Schafer BW, Heizmann CW. 1999. Subcellular distribution of S100 proteins in tumor cells and their relocation in response to calcium activation. Histochem Cell Biol 111:453–459.
Myllykangas S, Himberg J, Bohling T, Nagy B, Hollmen J, Knuutila S. 2006. DNA copy number amplification profiling of human neoplasms. Oncogene 25:7324–7332.
Nigg EA. 2001. Cell cycle regulation by protein kinases and phosphatases. Ernst Schering Res Found Workshop 34:19–46.
Nishimura K, Ohki Y, Fukuchi-Shimogori T, Sakata K, Saiga K, Beppu T, Shirahata A, Kashiwagi K, Igarashi K. 2002. Inhibition of cell growth through inactivation of eukaryotic translation initiation factor 5A (eIF5A) by deoxyspergualin. Biochem J 363:761–768.
Nishimura K, Murozumi K, Shirahata A, Park MH, Kashiwagi K, Igarashi K. 2005. Independent roles of eIF5A and polyamines in cell proliferation. Biochem J 385:779–785.
O'Connor JB, Falk GW, Richter JE. 1999. The incidence of adenocarcinoma and dysplasia in Barrett's esophagus: Report on the Cleveland Clinic Barrett's Esophagus Registry. Am J Gastroenterol 94:2037–2042.
Oien KA, McGregor F, Butler S, Ferrier RK, Downie I, Bryce S, Burns S, Keith WN. 2004. Gastrokine 1 is abundantly and specifically expressed in superficial gastric epithelium, down-regulated in gastric carcinoma, and shows high evolutionary conservation. J Pathol 203:789–797.
Parle-McDermott A, McWilliam P, Tighe O, Dunican D, Croke DT. 2000. Serial analysis of gene expression identifies putative metastasis-associated transcripts in colon tumour cell lines. Br J Cancer 83:725–728.
Parsons R. 1998. Phosphatases and tumorigenesis. Curr Opin Oncol 10:88–91.
Phillips GL, Reece DE, Shepherd JD, Barnett MJ, Brown RA, Frei Lahr DA, Klingemann HG, Bolwell BJ, Spinelli JJ, Herzig RH, Herzig GP. 1991. High-dose cytarabine and daunorubicin induction and postremission chemotherapy for the treatment of acute myelogenous leukemia in adults. Blood 77:1429–1435.
Rana PS, Johnston DA. 2000. Incidence of adenocarcinoma and mortality in patients with Barrett's oesophagus diagnosed between 1976 and 1986: Implications for endoscopic surveillance. Dis Esophagus 13:28–31.
Regalado SP, Nambu Y, Iannettoni MD, Orringer MB, Beer DG. 1998. Abundant expression of the intestinal protein villin in Barrett's metaplasia and esophageal adenocarcinomas. Mol Carcinog 22:182–189.
Saiki I, Fujii H, Yoneda J, Abe F, Nakajima M, Tsuruo T, Azuma I. 1993. Role of aminopeptidase N (CD13) in tumor-cell invasion and extracellular matrix degradation. Int J Cancer 54:137–143.
Sato F, Meltzer SJ. 2006. CpG island hypermethylation in progression of esophageal and gastric cancer. Cancer 106:483–493.
Serag HB. 2006. Time trends of gastroesophageal reflux disease: A systematic review. Clin Gastroenterol Hepatol 5:17–26.
Seth P, Krop I, Porter D, Polyak K. 2002. Novel estrogen and tamoxifen induced genes identified by SAGE (Serial Analysis of Gene Expression). Oncogene 21:836–843.
Sorrell JM, Baber MA, Brinon L, Carrino DA, Seavolt M, Asselineau D, Caplan AI. 2003. Production of a monoclonal antibody, DF-5, that identifies cells at the epithelial-mesenchymal interface in normal human skin. APN/CD13 is an epithelial-mesenchymal marker in skin. Exp Dermatol 12:315–323.
St Croix B, Rago C, Velculescu V, Traverso G, Romans KE, Montgomery E, Lal A, Riggins GJ, Lengauer C, Vogelstein B, Kinzler KW. 2000. Genes expressed in human tumor endothelium. Science 289:1197–1202.
Swami S, Kumble S, Triadafilopoulos G. 1995. E-cadherin expression in gastroesophageal reflux disease, Barrett's esophagus, and esophageal adenocarcinoma: An immunohistochemical and immunoblot study. Am J Gastroenterol 90:1808–1813.
van Dekken H, Paris PL, Albertson DG, Alers JC, Andaya A, Kowbel D, van der Kwast TH, Pinkel D, Schroder FH, Vissers KJ, Wildhagen MF, Collins C. 2004. Evaluation of genetic patterns in different tumor areas of intermediate-grade prostatic adenocarcinomas by high-resolution genomic array analysis. Genes Chromosomes Cancer 39:249–256.
Varis A, Wolf M, Monni O, Vakkari ML, Kokkola A, Moskaluk C, Frierson H, Jr., Powell SM, Knuutila S, Kallioniemi A, El-Rifai W. 2002. Targets of gene amplification and overexpression at 17q in gastric cancer. Cancer Res 62:2625–2629.
Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. 1995. Serial analysis of gene expression. Science 270:484–487.
Velculescu VE, Vogelstein B, Kinzler KW. 2000. Analysing uncharted transcriptomes with SAGE. Trends Genet 16:423–425.
Ventura JJ, Nebreda AR. 2006. Protein kinases and phosphatases as therapeutic targets in cancer. Clin Transl Oncol 8:153–160.
Vogel JH, von Heydebreck A, Purmann A, Sperling S. 2005. Chromosomal clustering of a human transcriptome reveals regulatory background. BMC Bioinformatics 6:230.
Wu GD, Beer DG, Moore JH, Orringer MB, Appelman HD, Traber PG. 1993. Sucrase-isomaltase gene expression in Barrett's esophagus and adenocarcinoma. Gastroenterology 105:837–844.
Wu Z, Siadaty MS, Riddick G, Frierson HF, Jr., Lee JK, Golden W, Knuutila S, Hampton GM, El-Rifai W, Theodorescu D. 2006. A novel method for gene expression mapping of metastatic competence in human bladder cancer. Neoplasia 8:181–189.
Yeager CL, Ashmun RA, Williams RK, Cardellichio CB, Shapiro LH, Look AT, Holmes KV. 1992. Human aminopeptidase N is a receptor for human coronavirus 229E. Nature 357:420–422.