Genetic Profiles of Gastroesophageal Cancer

combined analysis using expression array and tiling array-CGH

Ekstrand-Isinger. A[a*], Johansson J[b], Ohlsson M[c], Francis P[a], Staaf J[a], Jönsson M[a], Borg Å[a] and Nilbert M[a,d]

[a]Department of Oncology, Clinical Sciences, Lund University, Lund, Sweden
[b]Department of Surgery, Clinical Sciences, Lund University, Lund, Sweden
[c]Department of Theoretical Physics, Lund University, Lund, Sweden
[d]Clinical Research Centre, Hvidovre Hospital, Copenhagen University, Hvidovre, Denmark

[*] Corresponding author. Barngatan 2b, SE-22185 Lund, Sweden. Tel. +46 46 178524. Fax. +46 46 147327. E-mail. anna.ekstrand@med.lu.se.

Abstract

We aimed to characterize the genomic profiles of adenocarcinomas in the gastroesophageal junction in relation to cancers in the esophagus and the stomach. Profiles of gains/losses as well as gene expression profiles were obtained from 27 gastroesophageal adenocarcinomas using 32k high-resolution array-based comparative genomic hybridization (aCGH) and 27k oligo gene expression arrays and putative target genes were validated in an extended series. Adenocarcinomas in the distal esophagus and the gastroesophageal junction showed strong similarities with the most common gains at 20q13, 8q24, 1q21-q23, 5p15, 13q34, and 12q13, whereas different profiles with gains at 5p15, 7p22, 2q35, and 13q34 characterized gastric cancers. CDK6 and EGFR were identified as putative target genes in cancers of the esophagus and the gastroesophageal junction with upregulation in one quarter of the tumors. Gains/losses and gene expression profiles show strong similarity between cancers in the distal esophagus and the gastroesophageal junction with frequent upregulation of CDK6 and EGFR, whereas gastric cancer displays distinct genetic changes. These data suggest that molecular diagnostics and targeted therapies can be applied to adenocarcinomas of the distal esophagus and gastroesophageal junction alike.

1. Introduction

Gastroesophageal cancers are genetically complex and carry a serious prognosis with 5-year overall survival rates below 20%. Refined diagnosis and new treatment options, e.g. biological therapies directed at critical signaling pathways, are therefore needed. Worldwide, gastroesophageal cancer affects 1.5 million individuals annually [1]. Gastric adenocarcinomas (GA) show a decreasing incidence in the western world and are associated with *Helicobacter Pylori* infection, whereas adenocarcinomas in the distal esophagus (EA) constitutes one of the most rapidly increasing tumor types linked to overweight and gastroesophageal reflux [2]. Tumors that arise within the gastroesophageal junction (JA) may cause diagnostic dilemmas. Though correct classification has implications for the choice of treatment strategy, the current classification is based on the anatomical location and fails to account for differences in etiology and tumor biology [3-5].

Genetic changes have been closely linked to the metaplasia-dysplasia sequence of gastroesophageal cancer. EA as well as GA have demonstrated complex genetic alterations and gene expression changes [6, 7]. Though multiple genetic similarities have been recognized between EA and JA, differences with losses of 5q, 8p, 14q and gains of 2q, 6p, 12p and 20q to be discriminative have also been suggested [8, 9, 10-12]. With the aim to obtain a detailed genetic picture of gastroesophageal adenocarcinomas and to identify key genes and pathways herein, we applied a combined array-based gene expression and genomic profiling analysis to adenocarcinomas in the esophagus, the gastroesophageal junction, and the gastric body.

2. Materials & Methods

2.1. Materials

Tumor tissue from 27 primary gastroesophageal and gastric adenocarcinomas were snap frozen at -80°C. The mean age of the 18 males and 9 females was 65 (range 39-84) years and none of the patients had received neoadjuvant radiotherapy or chemotherapy. The tumors were classified according to the Siewert classification [5] and included 10 EA (located 1-5 cm above the gastroesophageal junction), 9 JA (located within 1 cm above and 2 cm below the gastroesophageal junction), and 8 GA (located in the gastric fundus-corpus with 5 being of the intestinal type and 3 being diffuse GA). Tumor stage was I in 2 tumors, II in 13, III in 5, and IV in 7. Barrett's esophagus was diagnosed in 7/10 tumors in the distal esophagus and in 1/9 tumors in the gastroesophageal junction and thus occurred at a higher frequency in the former subset ($p=0.02$). Ethical approval for this study was obtained from the Lund University ethics committee and the patients provided informed consent for participation.

2.2. Target preparation and hybridization

High-resolution tiling 32k BAC microarrays with complete genome coverage were produced by the Swegene DNA Microarray Center, Lund University, using the BAC Re-Array set Ver. 1.0 (BACPAC Resource Center, Children's Hospital Oakland Research Institute, Oakland, CA, USA (http://bacpac.chori.org/) [13]. BAC clones were mapped to the hg 17 build from the UCSC Genome Browser (http://genome.ucsc.edu/). Genomic DNA was extracted using proteinase K treatment followed by phenol chloroform purification. 2 µg of tumor DNA and 1.5 µg of a pool of male reference DNA (Promega, Madison, WI, USA), were differentially labeled with Cy3-dCTP and Cy5-dCTP (Amersham Biosciences, UK) using the Bioprime Genomic labeling system (Invitrogen Life Technologies, Carlsbad, CA, USA). Labeled DNA was pooled, mixed with 100 ug human Cot-1 DNA (Invitrogen Life Technologies), and dried in a Thermo Savant DNA SpeedVac™ (Thermo Electron Corporation, Waltham, MA, USA)

and thereafter re-suspended in 57 µl of hybridization buffer, denatured, re-annealed and applied to the arrays where after washing and scanning was performed [14].

27k oligonucleotide arrays printed from the Human Genome Oligo Set Version 2.1 (21, 329; 70-mer probes) and the Version 2.1 upgrade (5, 462 probes) were produced at the Swegene DNA Microarray Resource Centre [15]. RNA was extracted from 80-120 mg freshly frozen tumor tissue using TRizol (Invitrogen Life Technologies) and the RNeasy Midi Kit (Qiagen, Valencia, CA, USA) according to the manufactures' recommendation. Target preparation was performed using the Pronto!TM Plus System (Corning Incorporated, Corning, NY, USA), in which 5 µg of tumor RNA and reference RNA (Stratagene Universal Reference, Stratagene, La Jolla, CA, USA) was incubated with a mix of random primers and oligo(dT). cDNA was labeled with Cy3 dCTP (tumor) and Cy5 dCTP (reference). Purified labeled tumor cDNA and reference cDNA was combined and dried in a Thermo Savant DNA SpeedVac^{TM} and resuspended in Universal Hybridization Solution. Including 10 alien RNA sequences and 10 *Arabidopsis thaliana* derived RNA sequences without any homology validated the quality of the arrays (Stratagene, La Jolla, CA, USA).

RNA/DNA quality was assessed using an Agilent Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and RNA/DNA quantity using a NanoDrop Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Arrays were scanned using an Agilent G2565AA Microarray Scanner (Agilent Technologies).

2.3. Data analysis

GenePix^{TM}Pro 4.1 (Axon Instruments Inc., Union City, CA, USA) was used for identification of individual spots. Gene expression data were uploaded into the Bio Array Software Environment (http://base.thep.lu.se) for further analysis [16]. Correction of background intensities of Cy3 and Cy5 were calculated using a median-feature and median-local background intensities of the uploaded files. The intensity ratios for the individual probes were calculated as intensity of the background corrected tumor channel (Cy3) divided with the intensity of the background corrected reference channel (Cy5). Low quality spots flagged during image analysis and spots with a signal-to-noise ratio (SNR) <1.5 and >10% saturation were excluded from further analysis. To compensate for dye bias and local background effects, data were normalized using a lowess algorithm in BASE [17]. The uncertainty of a spot u was estimated using SNR1-2 + SNR2-2, where SNRi corresponds to the signal-to-noise ratio for channel i. Replicates assays were merged and represented by a weighted mean. Expression values were modified according to an error model in that merged genes with large uncertainties u_i were moved closer to a weighted mean m across assays for that gene, as described [18]. Genes missing in more than 20% of the samples or with a standard deviation of the modified log ratios ≤0.3 was excluded. After these steps, 93 ,777 spots representing 3,535 genes remained for further processing. Hierarchical cluster analysis was performed using Pearson correlation. Significant Analysis of Microarrays (SAM) was used to identify significant genes with 1,000 permutations (cut-off level of median 0 false significant genes, corresponding to a false discovery rate of 0) [19]. Cluster analysis and SAM analysis were performed using the MeV application from the TM_4 microarray software suite, freely available at www.tigr.org/software/tm4/ [20]. The Expression Analysis Systematic Explorer (EASE) software (http://david.niaid.nih.gov/david/ease.htm) was used for the interpretation of biological function and the genes were classified into groups according to the Gene Ontology

Consortium (GO Biological Process) and the KEGG pathway. Functional groups with EASE score less than 0.05 are included in the results.

Copy number data were processed and handled as described in the gene expression section with the following modifications. In BASE, individual spots with a SNR $\leq$ 5 were removed from further analysis and were handled as missing values. The X and the Y chromosomes were excluded from the normalization step using a pin-based lowess algorithm. After these steps, 645,690 spots representing 29,780 BAC-clones remained. Copy number data was loaded into Nexus software, v.3 (Biodiscovery) for visualization. Data was segmented using a rank based segmentation algorithm with at least 5 probes per segment. Gains and losses were defined as log2 (ratios) of $\geq$ 0.2 and $\leq$ 0.3 respectively, with high gains defined as log2 (ratio) of $\geq$ 0.8. Copy number variation frequencies and comparison between groups were estimated using Nexus software.

2.4. Correlation between copy number and gene expression

After filtering and normalization of gene expression data, the remaining 14,000 genes were used to create a "map-list" with start and end positions of each 70-mer oligo in relation to position of the BAC-clones. Since most genes were covered by more than one BAC-clone a weighted mean copy number, derived from the size of the overlap, was used. Genes with expression values that varied between assays with a standard deviation of the modified log ratios $\geq$0.2 were further used for correlation analysis. Relative gene expression and copy number correlations were calculated using Spearman rank correlation. Plots showing the overall relative correlation between expression and copy number were generated using the $\log_2$ of the mean of expression/copy number in each tumor group (figure 1). The data were

filtered using a sliding window to reduce noise. Each plot was normalized to a maximum absolute value of 1.

2.5. Immunohistochemistry

Immunohistochemical validation of EGFR and CDK6 was applied in an extended set of 47 tumors, which represented a consecutive series of patients with EA/JA that had undergone surgery. Immunohistochemical stainings for EGFR (clone E30, 1:50, DAKO A/S, Glostrup, Denmark) and CDK6 (clone DCS-83, 1:25, Progen Heidelberg, Germany) were performed using 4-µm sections of formalin-fixed, paraffin embedded tissue, which were mounted on DAKO ChemMate Capillary Gap Microscope Slides (DAKO). Tissue sections were deparaffinized in xylol, and rehydrated through descending concentrations of alcohol. Sections for EGFR staining were pretreated with Proteinase K for 10 minutes. For CDK6 staining, antigen retrieval was performed in DAKO Target Retrieval Solution, pH 9 (DAKO), in a pressure cooker (Histolab, Gothenburg, Sweden). Immunohistochemical staining was performed using LSAB+ (EGFR) or the Envision method (CDK6) (DAKO) in an automated immunostainer (TechMate 500 Plus, DAKO). Membranous EGFR staining was scored as 0 (no expression), 1 (expression in few cells), 2 (moderate expression), and 3 (strong expression). Nuclear CDK6 expression was scored according to the fraction of stained tumor cells; 0 (0% of the cells), 1 (5-30%), 2 (31-60%) and 3 (61-100%). The staining was independently evaluated by two of the coauthors and a score of 2 or 3 was defined as strong expression. Fisher's exact t-test was used for categorical analysis of gene expressions, copy number and protein expressions.

3. Results

3.1. Gene expression profiling

Unsupervised cluster analysis based on the gene expression data (3,535 reporters) identified several sub-clusters, one of which contained 5/8 GA and the other containing a mixture of JA, EA and GA (supplemental data, figure 1). A similar pattern was observed when a less stringent variation filter (resulting in 7,451 genes) was applied (data not shown). The gene expression profiles in EA and JA showed strong similarity, whereas SAM-analysis identified 38 (FDR of 0%) and 164 (FDR of 5%) genes that were significantly upregulated in EA/JA compared to GA (supplemental data, table 1). Overexpressed genes were by EASE analysis found to be involved in cell cycle regulation, proliferation, MAP kinase pathway, stress response, biotic stimulus, and apoptotic response

3.2. Copy number analysis

High-quality copy number profiles were obtained from 23/27 tumors and showed multiple gains/losses and several recurrent alterations. Gains were more frequent than losses and similar to the expression profiles, the gains/losses identified in EA and JA were highly similar (figure 2) with the most common gains at 20q13.33 (88%), 8q24.3 (75%), 1q21.3-q23.1 (69%), 5p15.33 (69%), 13q34 (62%), 12q13.31 (56%), 12q13.13 (56%), 18q11.1-q11.2 (56%), 2q14.1 (50%), 1p36.33 (44%), 1p36.31 (44%), 6p21.1 (44%), 2q35 (38%) and 16q24.2-24.3 (38%). Genes located within these regions were preferentially involved in transcription, defense response and RNA elongation. The only significant loss (present in 69% of the tumors) was located on the Y-chromosome (Yq11.1-q11.23). Gains in GA affected 5p15.33 (86%), 7p22.3 (87%), 2q35 (57%) and 13q34 (57%) and these regions contained several genes involved in regulation of proteolysis, B cell proliferation and defense response. Multiple high-level amplifications (HLAs) with $\log_2$ ratios >0.8 were identified in

the majority of JA and EA but were rarely observed in GA. Candidate genes linked to these HLAs were *VEGFA* (6p21.1), *EGFR* (7p11.2), *CDK6* (7q21.2) and *ERBB2* (7q12).

Differential analysis between EA and GA identified 2,205 significantly different BAC clones (Mann-Whitney U-test, $p<0.05$). Supervised cluster analysis applied on all three sub-sets identified two sub-clusters with a close relation between EA and JA (data not shown). Application of the Nexus software identified only two significantly gained/lost regions between EA and JA with gain of 12q23.3-q24.21 and gain of 18p11.32, respectively.

3.3. Correlation between gene expression and copy number and protein expression

Overexpression and copy-number gains were highly correlated for the majority of the regions identified, which supports biologically relevant roles (figure 1). Based on concordant amplification and over expression in EA/JA, and previous links to gastroesophageal tumorigenesis and application of targeted therapies, CDK6 and EGFR were chosen for immunohistochemical validation in an extended sample set of 47 EA/JA. Gain/amplification and upregulated gene expression of CDK6 was identified in 4/14 (29%) and 2/14 (14%) tumors, respectively. Immunostaining of CDK6 showed strong expression in 11/47 (23%), weak in 18/47 (38%) and no expression in 19/47 (40%) (figure 3a). Amplification and upregulated gene expression of EGFR was found in 7/14 (50%) and in 1/14 (7%) EA/JA, respectively, and strong immunohistochemical expression in 11/46 (24%), weak in 11/46 (24%) and no expression was found in 24/46 (52%) EA/JA (figure 3b). No correlations with gene expression/copy number and protein expression was found.

4. Discussion

Combined gene expression and copy number analysis demonstrated extensive similarity between adenocarcinomas in the distal esophagus and the gastroesophageal junction, whereas adenocarcinomas in the gastric body revealed different profiles. All three tumor types showed multiple gains and losses, including gains at 20q13 and 8q, which have been recognized as recurrent in gastroesophageal tumors [9, 11, 12]. Gain of 20q13 has also been described in other tumor types and has been suggested to be of prognostic value in e.g. breast cancer and colorectal cancer [21, 22]. The genomic profiles of EA and JA showed extensive similarity, though gain of 12q was overrepresented in EA and gain of 18q in JA. A number of genetic regions have been reported to discriminate between EA and JA, but none of these have been validated and overall the similarities by far outnumber the differences, suggesting similar and to a large extent shared tumorigenic pathways [8-10, 12].

The gene expression profiles in EA/JA demonstrated upregulation of several genes linked to gastroesophageal tumorigenesis, e.g. *ECGF1, H19, S100A10, MYC, MMP11,* and *CTSB* [23-25]. In EA, both *ECGF1* and *CTSB,* which are linked to tumor development and suggested to represent a diagnostic marker, showed a high degree of over expression and amplification [26, 27]. Upregulation of MAP-kinase genes, i.e. *PLA2G2A, DUSP1, MYC, FOS* and *GADD45B* were identified in EA/JA and is intriguing since *in vitro* data suggest an anti-proliferative effect from MAPK inhibition in Barrett's adenocarcinoma and because of promising therapeutic strategies aimed at that downregulatating MAPK signaling [28].

Two markers of potential diagnostic, prognostic and treatment predictive impact were further evaluated using immunostaining. The cell cycle regulator CDK6, which has been linked to prognosis in medulloblastoma, has been suggested to represent the target gene in the 7q21

amplicon in EA [29, 30]. We identified gain/amplification of the *CDK6* locus in 29%, upregulated gene expression in 14%, and increased immunostaining in 23% of EA/JA. EGFR represents a major target for amplification and overexpression in several cancer types, which is exploited in targeted therapies using monoclonal antibodies as well as tyrosine kinase inhibitors. Amplification of the *EGFR* locus was found in 50% of EA/JA, upregulated gene expression in 7%, and immunohistochemical overexpression in 24%. This finding is in line with e.g. data from fluoresecence in situ hybridization showing *EGFR* amplification in 8-31% of EA [31-33] and upregulation of *EGFR* has been suggested to correlate with poor prognosis [34].

In summary, gastroesophageal cancers are characterized by genetic complexity but herein show concordant copy-number gains and upregulation of target genes suggesting that molecular diagnostics and targeted therapies can be applied to EA and JA alike. The correlations between genomic and expression-based signatures highlight involvement of central signaling pathways, including MAPK, *CDK6* and *EGFR* that may be of relevance for refined diagnostics, prognosis and targeted treatment in a tumor type with considerable resistence to the therapeutic options currently available.

Acknowledgements

The study was supported by grants from the Knut and Alice Wallenberg Foundation via the Swegene program, the Swedish Cancer Society, the Nilsson Cancer Research Fund, the Kamprad Research Fund, the Region Skåne Research Funds and the Inga-Britt and Arne Lundberg Foundation.

We thank Anna Laurell for technical assistance.

References

[1] Lambert R, Hainaut P. The multidisciplinary management of gastrointestinal cancer. Epidemiology of oesophagogastric cancer. Best Pract Res Clin Gastroenterol 2007;21:921-45.

[2] Forman D. Review article: oesophago-gastric adenocarcinoma -- an epidemiological perspective. Aliment Pharmacol Ther 2004;20 Suppl 5:55-60; discussion 1-2.

[3] Kalish RJ, Clancy PE, Orringer MB, Appelman HD. Clinical, epidemiologic, and morphologic comparison between adenocarcinomas arising in Barrett's esophageal mucosa and in the gastric cardia. Gastroenterology 1984;86:461-7.

[4] Ruol A, Parenti A, Zaninotto G, Merigliano S, Costantini M, Cagol M, Alfieri R, Bonavina L, Peracchia A, Ancona E. Intestinal metaplasia is the probable common precursor of adenocarcinoma in barrett esophagus and adenocarcinoma of the gastric cardia. Cancer 2000;88:2520-8.

[5] Siewert JR, Stein HJ. Classification of adenocarcinoma of the oesophagogastric junction. Br J Surg 1998;85:1457-9.

[6] Riegman PH, Vissers KJ, Alers JC, Geelen E, Hop WC, Tilanus HW, van Dekken H. Genomic alterations in malignant transformation of Barrett's esophagus. Cancer Res 2001;61:3164-70.

[7] Wu TT, Watanabe T, Heitmiller R, Zahurak M, Forastiere AA, Hamilton SR. Genetic alterations in Barrett esophagus and adenocarcinomas of the esophagus and esophagogastric junction region. Am J Pathol 1998;153:287-94.

[8] El-Rifai W, Frierson HF, Jr., Moskaluk CA, Harper JC, Petroni GR, Bissonette EA, Jones DR, Knuutila S, Powell SM. Genetic differences between adenocarcinomas arising in Barrett's esophagus and gastric mucosa. Gastroenterology 2001;121:592-8.

[9] van Dekken H, Geelen E, Dinjens WN, Wijnhoven BP, Tilanus HW, Tanke HJ, Rosenberg C. 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 1999;59:748-52.

[10] Menke-Pluymers MB, van Drunen E, Vissers KJ, Mulder AH, Tilanus HW, Hagemeijer A. Cytogenetic analysis of Barrett's mucosa and adenocarcinoma of the distal esophagus and cardia. Cancer Genet Cytogenet 1996;90:109-17.

[11] Moskaluk CA, Hu J, Perlman EJ. Comparative genomic hybridization of esophageal and gastroesophageal adenocarcinomas shows consensus areas of DNA gain and loss. Genes Chromosomes Cancer 1998;22:305-11.

[12] Varis A, Puolakkainen P, Savolainen H, Kokkola A, Salo J, Nieminen O, Nordling S, Knuutila S. DNA copy number profiling in esophageal Barrett adenocarcinoma: comparison with gastric adenocarcinoma and esophageal squamous cell carcinoma. Cancer Genet Cytogenet 2001;127:53-8.

[13] Jonsson G, Staaf J, Olsson E, Heidenblad M, Vallon-Christersson J, Osoegawa K, de Jong P, Oredsson S, Ringner M, Hoglund M, Borg A. High-resolution genomic profiles of breast cancer cell lines assessed by tiling BAC array comparative genomic hybridization. Genes Chromosomes Cancer 2007;46:543-58.

[14] Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D, Albertson DG. Assembly of microarrays for genome-wide measurement of DNA copy number. Nat Genet 2001;29:263-4.

[15] Jonsson G, Naylor TL, Vallon-Christersson J, Staaf J, Huang J, Ward MR, Greshock JD, Luts L, Olsson H, Rahman N, Stratton M, Ringner M, Borg A, Weber BL. Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. Cancer Res 2005;65:7612-21.

[16] Saal LH, Troein C, Vallon-Christersson J, Gruvberger S, Borg A, Peterson C. BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. Genome Biol 2002;3:SOFTWARE0003.

[17] Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 2002;30:e15.

[18] Fernebro J, Francis P, Eden P, Borg A, Panagopoulos I, Mertens F, Vallon-Christersson J, Akerman M, Rydholm A, Bauer HC, Mandahl N, Nilbert M. Gene expression profiles relate to SS18/SSX fusion type in synovial sarcoma. Int J Cancer 2006;118:1165-72.

[19] Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98:5116-21.

[20] Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. TM4: a free, open-source system for microarray data management and analysis. Biotechniques 2003;34:374-8.

[21] Aust DE, Muders M, Kohler A, Schmidt M, Diebold J, Muller C, Lohrs U, Waldman FM, Baretton GB. Prognostic relevance of 20q13 gains in sporadic colorectal cancers: a FISH analysis. Scand J Gastroenterol 2004;39:766-72.

[22] Isola JJ, Kallioniemi OP, Chu LW, Fuqua SA, Hilsenbeck SG, Osborne CK, Waldman FM. Genetic aberrations detected by comparative genomic hybridization predict outcome in node-negative breast cancer. Am J Pathol 1995;147:905-11.

[23] Hibi K, Nakamura H, Hirai A, Fujikake Y, Kasai Y, Akiyama S, Ito K, Takagi H. Loss of H19 imprinting in esophageal cancer. Cancer Res 1996;56:480-2.

[24] Hourihan RN, O'Sullivan GC, Morgan JG. Transcriptional gene expression profiles of oesophageal adenocarcinoma and normal oesophageal tissues. Anticancer Res 2003;23:161-5.

[25] Tselepis C, Morris CD, Wakelin D, Hardy R, Perry I, Luong QT, Harper E, Harrison R, Attwood SE, Jankowski JA. Upregulation of the oncogene c-myc in Barrett's adenocarcinoma: induction of c-myc by acidified bile acid in vitro. Gut 2003;52:174-80.

[26] Hughes SJ, Glover TW, Zhu XX, Kuick R, Thoraval D, Orringer MB, Beer DG, Hanash S. A novel amplicon at 8p22-23 results in overexpression of cathepsin B in esophageal adenocarcinoma. Proc Natl Acad Sci U S A 1998;95:12410-5.

[27] Mitas M, Almeida JS, Mikhitarian K, Gillanders WE, Lewin DN, Spyropoulos DD, Hoover L, Graham A, Glenn T, King P, Cole DJ, Hawes R, Reed CE, Hoffman BJ. Accurate discrimination of Barrett's esophagus and esophageal adenocarcinoma using a quantitative three-tiered algorithm and multimarker real-time reverse transcription-PCR. Clin Cancer Res 2005;11:2205-14.

[28] Holcombe RF, Marsh JL, Waterman ML, Lin F, Milovanovic T, Truong T. Expression of Wnt ligands and Frizzled receptors in colonic mucosa and in colon carcinoma. Mol Pathol 2002;55:220-6.

[29] Mendrzyk F, Radlwimmer B, Joos S, Kokocinski F, Benner A, Stange DE, Neben K, Fiegler H, Carter NP, Reifenberger G, Korshunov A, Lichter P. Genomic and protein expression profiling identifies CDK6 as novel independent prognostic marker in medulloblastoma. J Clin Oncol 2005;23:8853-62.

[30] van Dekken H, van Marion R, Vissers KJ, Hop WC, Dinjens WN, Tilanus HW, Wink JC, van Duin M. Molecular dissection of the chromosome band 7q21 amplicon in gastroesophageal junction adenocarcinomas identifies cyclin-dependent kinase 6 at both genomic and protein expression levels. Genes Chromosomes Cancer 2008;47:649-56.

[31] al-Kasspooles M, Moore JH, Orringer MB, Beer DG. Amplification and over-expression of the EGFR and erbB-2 genes in human esophageal adenocarcinomas. Int J Cancer 1993;54:213-9.

[32] Miller CT, Moy JR, Lin L, Schipper M, Normolle D, Brenner DE, Iannettoni MD, Orringer MB, Beer DG. Gene amplification in esophageal adenocarcinomas and Barrett's with high-grade dysplasia. Clin Cancer Res 2003;9:4819-25.

[33] Rygiel AM, Milano F, Ten Kate FJ, Schaap A, Wang KK, Peppelenbosch MP, Bergman JJ, Krishnadath KK. Gains and amplifications of c-myc, EGFR, and 20.q13 loci in the no dysplasia-dysplasia-adenocarcinoma sequence of Barrett's esophagus. Cancer Epidemiol Biomarkers Prev 2008;17:1380-5.

[34] Wang KL, Wu TT, Choi IS, Wang H, Resetkova E, Correa AM, Hofstetter WL, Swisher SG, Ajani JA, Rashid A, Albarracin CT. Expression of epidermal growth factor receptor in esophageal and esophagogastric junction adenocarcinomas: association with poor outcome. Cancer 2007;109:658-67.

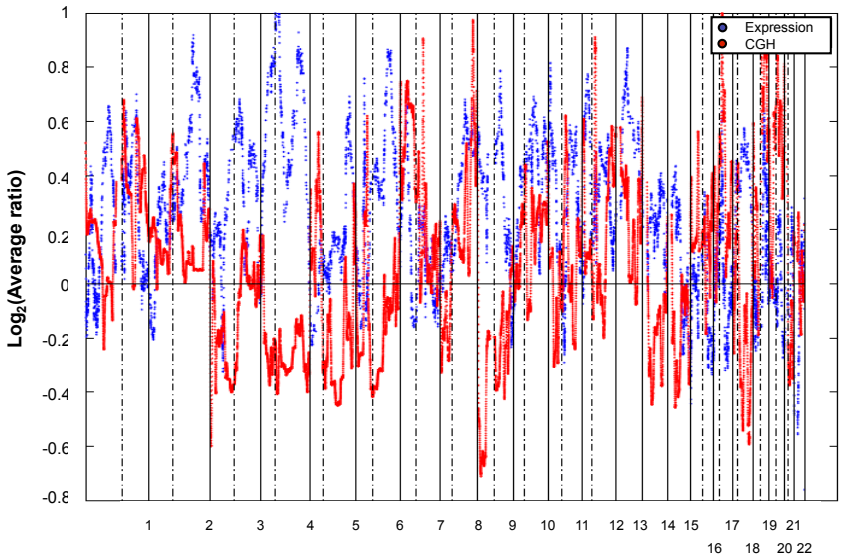
Titles and legends to figures

Figure 1 - Correlation between the relative mean expression and copy number data (log2 values and normalized to a maximum absolute value of 1) in the different tumor types; A, esophageal adenocarcinoma (EA); B, esophagogastric junction (JA) and C, gastric adenocarcinoma (GA). The p- and q-arms of the chromosomes are separated with dotted lines.

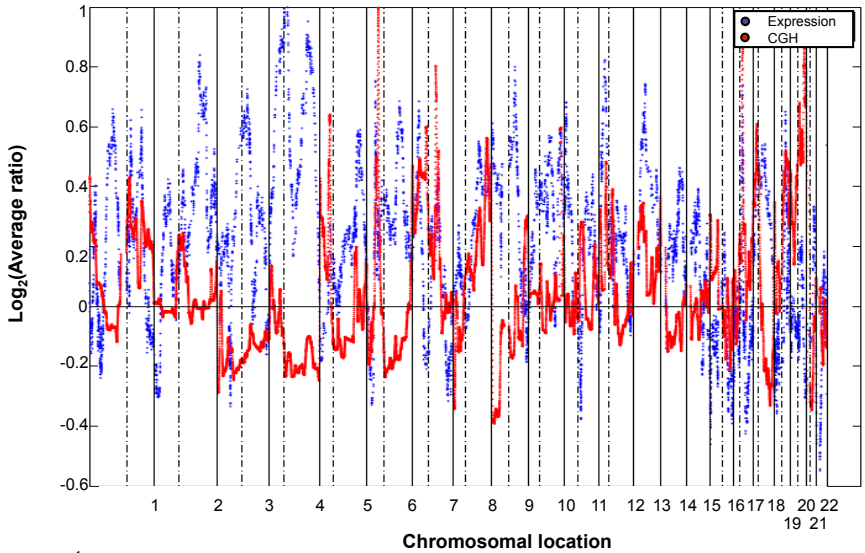
Figure 2 - Frequency plots demonstrating genomic gains and losses in distal esophageal tumors (EA), gastric tumors (GA) and esophagogastric junction tumors (JA) respectively. The plot at the top summarizes the frequencies of gains/losses for all tumors.

Figure 3 - Genomic profiles of two gastroesophageal tumors, with high level amplification peaks corresponding to CDK6 (a) and EGFR (b) respectively, and with corresponding immunohistochemical photo of protein expression.

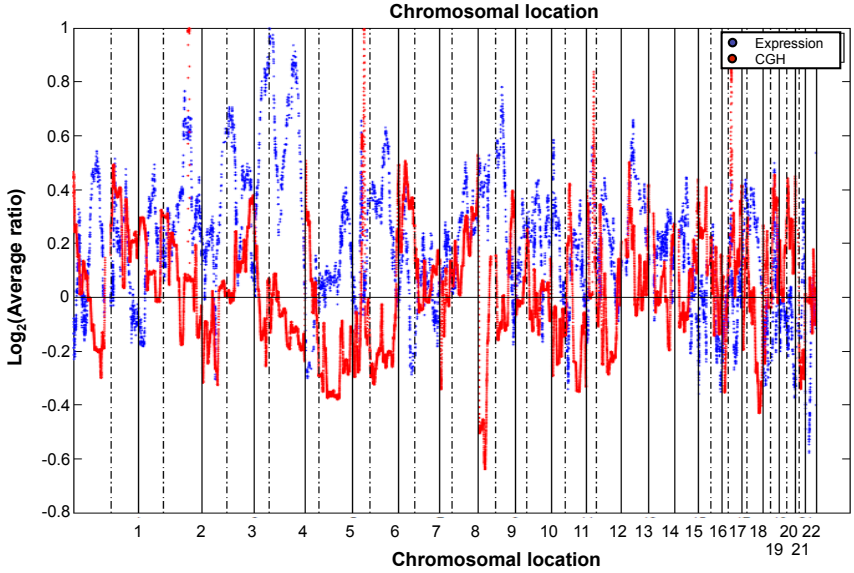
A


Expression
CGH
Log2(Average ratio)
Chromosomal location

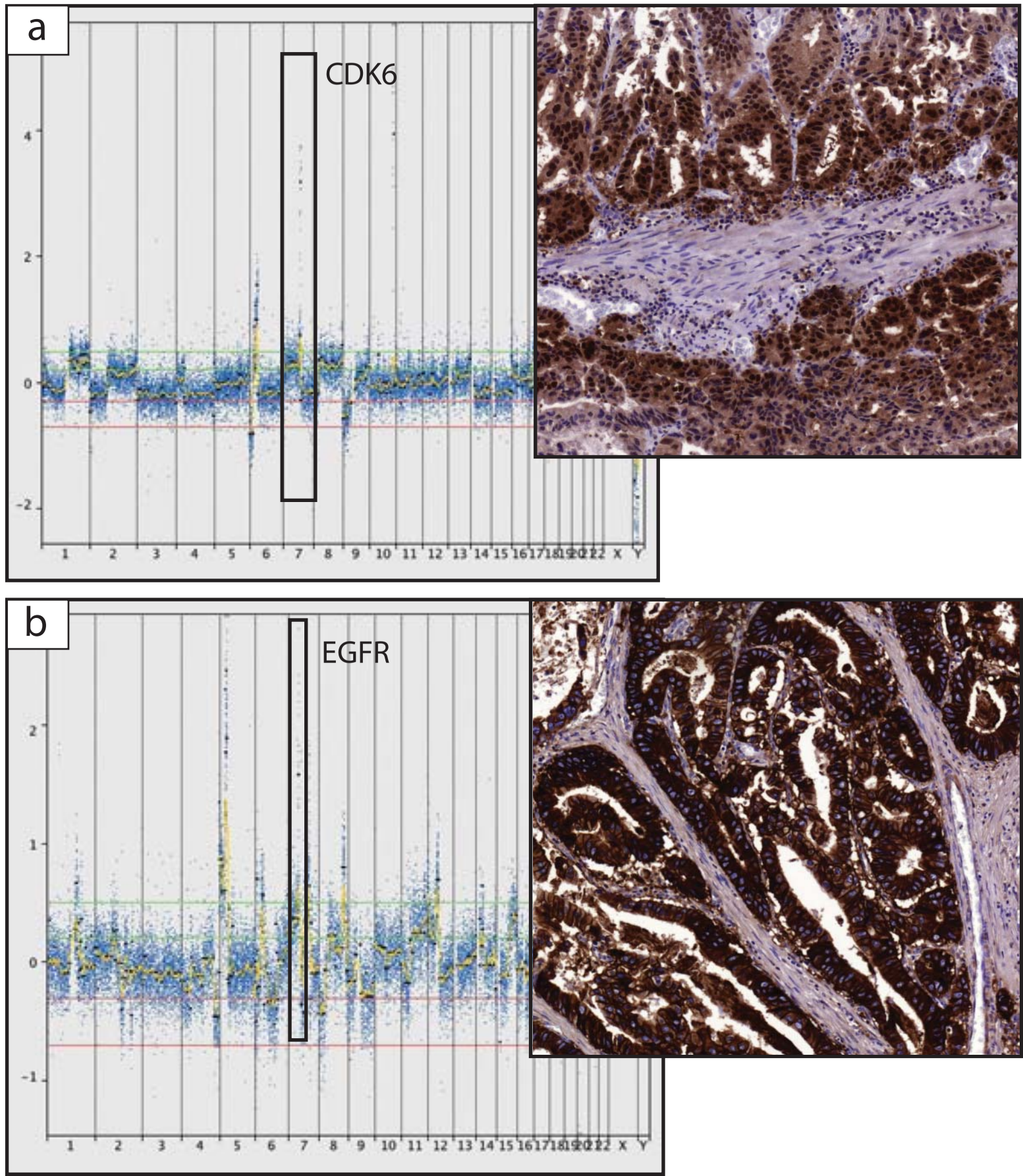
B

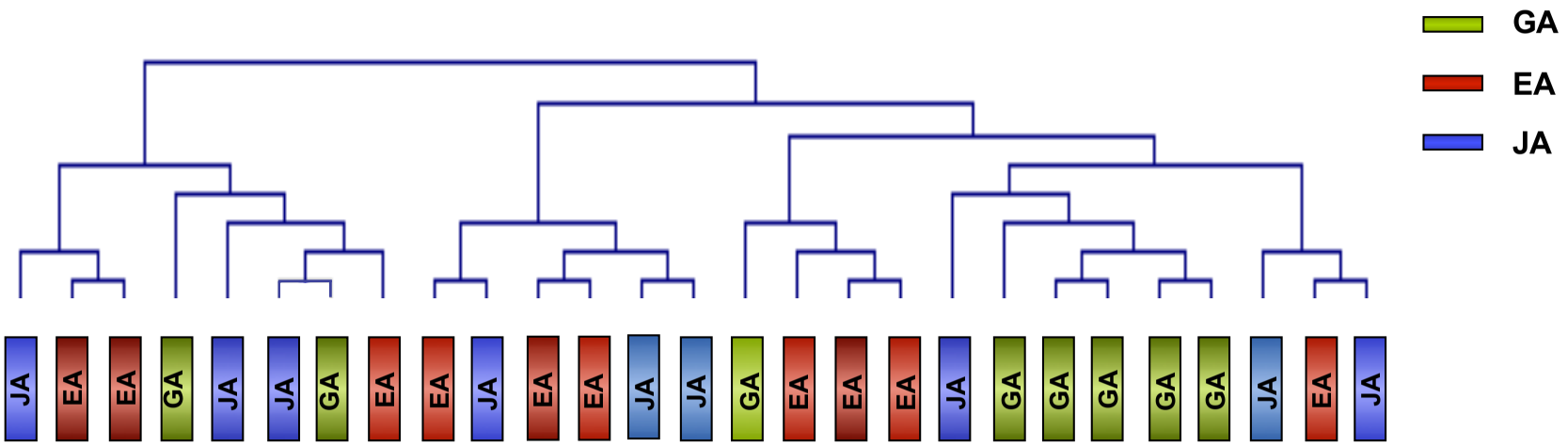

Expression
CGH
Log2(Average ratio)
Chromosomal location

C


Expression
CGH
Log2(Average ratio)
Chromosomal location


1p36.33 - Yq12
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
100%
50%
0%
50%
100%
Genes
Exons
CNVs
miRNA
BAC_32k
(7) EA
(7 GA
(9) JA
P-Value cut-off
0.05
Aggregate % cut-off
35.0
Peaks only


a
CDK6
4
2
0
-2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
b
EGFR
2
1
0
-1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y



Supplemental figure 1 - Unsupervised hierarchical clustering of esophagogastric adenocarcinomas based on ~3,500 genes in which 5/8 GA cluster together, whereas the 10 EA and the 9 JA fall into several sub-groups.

Supplemental table 1 - Upregulated genes in the distal esophageal adenocarcinomas (EA) and the gastroesophageal junction adenocarcinomas (JA) compared to the gastric adenocarcinomas (GA). The table includes 164 genes with a FDR of 5%, identified by two-paired SAM-analysis whereas genes showed in bold have an FDR of zero.

Gene symbol	Chromosome band	Locus link	Fold change
CD53	1p13	963	0.5950888
DNAJB4	1p31.1	11080	0.6175997
LAPTM5	1p34	7805	0.5276802
PLK3	1p34.1	1263	0.7473697
PLA2G2A	1p35	5320	0.30144104
CA6	1p36.2	765	0.74980944
CDA	1p36.2-p35	978	0.58441144
TNFRSF1B	1p36.3-p36.2	7133	0.75916666
S100A10	1q21	6281	0.45890749
S100A2	1q21	6273	0.14996576
S100A9	1q21	6280	0.28657454
ADAMTS4	1q21-q23	9507	0.6469072
FCGR3A	1q23	2214	0.5993047
FCGR2B	1q23	2213	0.3723505
IER5	1q25.3	51278	0.63266563
RGS2	1q31	5997	0.5641359
CFH	1q32	3075	0.68747014
NUAK2	1q32.1	81788	0.7528812
CAPG	2p11.2	822	0.67054206
DUSP2	2q11	1844	0.6577075
ARID5A	2q11.2	10865	0.67046505
AFF3	2q11.2-q12	3899	0.65619636
MGAT4A	2q12	11320	0.6667216
DPP4	2q24.3	1803	0.64388
TFPI	2q32	7035	0.7909449
HSPD1	2q33.1	3329	0.61402315
FN1	2q34	2335	0.56064487
FN1	2q34	2335	0.76720315
COL6A3	2q37	1293	0.6679635
ARL4C	2q37.1	10123	0.5192106
CMKOR1	2q37.3	57007	0.64132357
ITIH1	3p21.2-p21.1	3697	0.7313921
FBLN2	3p25.1	2199	0.62631726
COPG	3q21.3	22820	0.3697191
PLSCR1	3q23	5359	0.7631207
HOP	4q11-q12	84525	0.4049343
TRIO	5p15.1-p14	7204	0.7274474
NKD2	5p15.3	85409	0.7764414
F2R	5q13	2149	0.6378974
ARRDC3	5q14.3	57561	0.78025544
SEPP1	5q31	6414	0.5310646
EGR1	5q31.1	1958	0.5816082
GM2A	5q31.3-q33.1	2760	0.73493916
PDGFRB	5q31-q32	5159	0.7776797

CDX1	5q31-q33	1044	0.45628867
DUSP1	5q34	1843	0.56651855
TREM2	6p21.1	54209	0.7435078
NFKBIE	6p21.1	4794	0.6824272
CDKN1A	6p21.2	1026	0.69051194
HSPA1B	6p21.3	3303	0.34192654
HSPA1A	6p21.3	3303	0.35465878
LSM2	6p21.3	57819	0.7099511
MICA	6p21.3	4276	0.602144
C6orf149	6p25.1	57128	0.7735012
CD109	6q13	135228	0.5931137
C6orf150	6q13	115004	0.6841717
GJA1	6q21-q23.2	2697	0.72992593
MTRF1L	6q25	54516	0.7133811
SOD2	6q25.3	6648	0.6384444
UPP1	7p12.3	7378	0.4879369
GPNMB	7p15	10457	0.47042316
RALA	7p15-p13	5898	0.731414
SCAP2	7p21-p15	8935	0.6974611
YWHAG	7q11.23	7532	0.7440039
GTF2I	7q11.23	2969	0.63876134
SERPINE1	7q21.3-q22	5054	0.49055758
PODXL	7q32-q33	5420	0.70612633
CTSB	8p22	1508	0.536777
ATP6V1B2	8p22-p21	526	0.7464556
GEM	8q13-q21	2669	0.74662995
CTHRC1	8q22.3	115908	0.6243986
TNFRSF11B	8q24	4982	0.4656863
MYC	8q24.12-q24.13	4609	0.6064123
TATDN1	8q24.13	83940	0.74204445
GRINA	8q24.3	2907	0.6829245
BNC2	9p22.3-p22.2	54796	0.7179303
ANXA1	9q12-q21.2|9q12-q21.2	301	0.18238491
CTSL	9q21-q22	1514	0.5519466
GADD45G	9q22.1-q22.2	10912	0.76934904
ZNF33A	10p11.2	7581	0.52114326
ZNF438	10p11.23	220929	0.8107101
BAMBI	10p12.3-p11.2	25805	0.5222711
SH3PXD2A	10q24.33	9644	0.7740955
KIAA1754	10q25.1	85450	0.749174
GSTO1	10q25.1	9446	0.75374824
BAG3	10q25.2-q26.2	9531	0.5450204
H19	11p15.5	283120	0.20831887
MUC2	11p15.5	4583	0.37577808
STIP1	11q13	10963	0.7423256
FTH1	11q13	2495	0.6611367
YAP1	11q13	10413	0.7424274
CENTD2	11q13.4	116985	0.7427585
SERPINH1	11q13.5	871	0.5462724
PRCP	11q14	5547	0.6991295
CHORDC1	11q14.3	26973	0.5871089
ENDOD1	11q21	23052	0.6716178
REXO2	11q23.1-q23.2	25996	0.7803356
ETS1	11q23.3	2113	0.6304317
EMP1	12p12.3	2012	0.26516217

FLJ22662	12p13.1	79887	0.58745563
CSDA	12p13.1	8531	0.6497142
MGP	12p13.1-p12.3	4256	0.92134434
CD9	12p13.3	928	0.63136715
A2M	12p13.3	2	0.6739991
FKBP4	12p13.33	2288	0.52674276
CLEC2B	12p13-p12	9976	0.660021
KRT5	12q12-q13	3852	0.15863033
RND1	12q12-q13	27289	0.5358515
KRT8	12q13	3856	0.06145612
LOC144501	12q13.13	144501	0.50792146
PHLDA1	12q15	22822	0.70168537
NID2	14q21-q22	22795	0.72342676
PYGL	14q21-q22	5836	0.70001274
AHSA1	14q23.3-31	10598	0.7698342
LTBP2	14q24	4053	0.58221406
FOS	14q24.3	2353	0.41947952
HSP90AA1	14q32.33	3320	0.5715347
ITPKA	15q14-q21	3706	0.49442717
THBS1	15q15	7057	0.59211385
NOD27	16q13	84166	0.74626994
IRF8	16q24.1	3394	0.6988805
CLDN7	17p13	1366	0.5413264
KRT17	17q12-q21	3872	0.058498725
COL1A1	17q21.33	1277	0.4859523
MRC2	17q23.2	9902	0.754025
C1QTNF1	17q25.3	114897	0.75775164
TUBB6	18p11.21	84617	0.58213055
RAB31	18p11.3	11031	0.68956786
IFI30	19p13.1	10437	0.61414695
GADD45B	19p13.3	4616	0.6593149
WDR18	19p13.3	57418	0.76591265
ICAM1	19p13.3-p13.2	3383	0.5397866
ACP5	19p13.3-p13.2	54	0.72082096
PEPD	19q12-q13.2	5184	0.69930196
TOMM40	19q13	10452	0.27576482
TOMM40	19q13	10452	0.57135195
ZFP36	19q13.1	7538	0.51085496
TYROBP	19q13.1	7305	0.6816054
ITPKC	19q13.1	80271	0.75168335
PPP1R15A	19q13.2	23645	0.4729918
APOC2	19q13.2	344	0.43337035
KLK1	19q13.3	3816	0.38010582
FLJ20512	19q13.32	54958	0.7608311
NOL5A	20p13	10528	0.74574405
C20orf96	20p13	140680	0.7309621
CPXM	20p13-p12.3	56265	0.7354351
PROCR	20q11.2	10544	0.5665474
CPNE1	20q11.22	8904	0.7290372
MAFB	20q11.2-q13.1	9935	0.6669973
TGM2	20q12	7052	0.5686818
PLTP	20q12-q13.1	5360	0.6612585
CEBPB	20q13.1	1051	0.75960183
TMEPAI	20q13.31-q13.33	56937	0.62337506
UCKL1	20q13.33	54963	0.78323305

SAMSN1	21q11	64092	0.76850903
ADAMTS1	21q21.2	9510	0.6576979
ITGB2	21q22.3	3689	0.6448108
BID	22q11.1	637	0.7016404
MMP11	22q11.2|22q11.23	4320	0.616754
HMOX1	22q12|22q13.1	3162	0.44399348
LGALS2	22q12-q13|22q13.1	3957	0.58973575
ECGF1	22q13|22q13.33	1890	0.5001053
TIMP1	Xp11.3-p11.23	7076	0.51190704
BGN	Xq28	633	0.52467614