Published in final edited form as:

Clin Cancer Res. 2011 May 1; 17(9): 2955–2966. doi:10.1158/1078-0432.CCR-10-2724.

Global gene expression profiling and validation in esophageal squamous cell carcinoma (ESCC) and its association with clinical phenotypes

Hua Su^{1,*}, Nan Hu^{1,*}, Howard H Yang², Chaoyu Wang¹, Mikiko Takikita³, Quan-Hong Wang⁴, Carol Giffen⁵, Robert Clifford², Stephen M Hewitt³, Jian-Zhong Shou⁶, Alisa M Goldstein¹, Maxwell P Lee^{2,**}, and Philip R Taylor^{1,**}

Hua Su: suhu@mail.nih.gov; Nan Hu: nh38k@nih.gov; Howard H Yang: hy43h@nih.gov; Chaoyu Wang: wangc@mail.nih.gov; Mikiko Takikita: mikitaki@belle.shiga-med.ac.jp; Quan-Hong Wang: zhongmeiket@yahoo.com.cn; Carol Giffen: giffenc@imsweb.com; Robert Clifford: clifforr@mail.nih.gov; Stephen M Hewitt: hewitts@mail.nih.gov; Jian-Zhong Shou: zhousaping@csc.org.cn; Alisa M Goldstein: ag26o@nih.gov

¹Genetic Epidemiology Branch, DCEG, NCI, NIH, Bethesda, MD 20892, USA

²Laboratory of Population Genetics, CCR, NCI, NIH, Bethesda, MD 20892, USA

³Laboratory of Pathology, CCR, NCI, NIH, Bethesda, MD 20892, USA

⁴Shanxi Cancer Hospital, Taiyuan, Shanxi 030013, PR China

⁵Information Management Services, Inc, Silver Spring, MD 20904, USA

⁶Cancer Institute and Hospital, Chinese Academy of Medical Sciences, Beijing, 100021, PR China

Abstract

Purpose—Esophageal squamous cell carcinoma (ESCC) is an aggressive tumor with poor prognosis. Understanding molecular changes in ESCC will enable identification of molecular subtypes and provide potential targets for early detection and therapy.

Experimental Design—We followed up a previous array study with additional discovery and confirmatory studies in new ESCC cases using alternative methods. We profiled global gene expression for discovery and confirmation, and validated selected dysregulated genes with additional RNA and protein studies.

Results—A total of 159 genes showed differences with extreme statistical significance ($P < E-15$) and ≥ 2 -fold differences in magnitude (tumor/normal RNA expression ratio, $N=53$ cases), including 116 up-regulated and 43 down-regulated genes. Of 41 genes dysregulated in our prior array study, all but one showed the same fold change directional pattern in new array studies, including 29 with ≥ 2 -fold changes. Alternative RNA expression methods validated array results: more than two-thirds of 51 new cases examined by RT-PCR showed ≥ 2 -fold differences for all seven genes assessed. Immunohistochemical protein expression results in 275 cases were concordant with RNA for five of six genes.

Conclusion—We identified an expanded panel of genes dysregulated in ESCC and confirmed previously identified differentially-expressed genes. Microarray-based gene expression results

**Correspondence should be sent to: (1) Philip R Taylor, Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, NCI, 6120 Executive Blvd, Rm 7006, MSC 7236 MD 20892-7236, Tel 301-594-2932, Fax 301-402-4489, ptaylor@mail.nih.gov or; (2) Maxwell P Lee, Laboratory of Population Genetics, Center for Cancer Research, NCI, Bldg 41, Rm D702, 41 Library Dr, Bethesda, MD 20892, Tel 301-435-8956, Fax 301-435-8963, leemax@mail.nih.gov.

*These authors contribute equally to this work

were confirmed by RT-PCR and protein expression studies. These dysregulated genes will facilitate molecular categorization of tumor subtypes and identification of their risk factors, and serve as potential targets for early detection, outcome prediction, and therapy.

Keywords

esophageal squamous cell carcinoma (ESCC); Affymetrix oligomicroarray; RT-PCR; tissue microarray (TMA)

Background

Esophageal cancer is the sixth most common fatal human cancer in the world (1) and the fourth most common new cancer in China (2). Shanxi Province, a region in north central China, has among the highest esophageal cancer rates in China and nearly all of these cases are esophageal squamous cell carcinoma (ESCC). ESCC is an aggressive tumor which is typically diagnosed only after the onset of symptoms when prognosis is very poor. The 19% 5-year survival rate is fourth worst among all cancers in the USA (3). One promising strategy to reduce ESCC mortality is early detection, and a better understanding of the molecular mechanisms underlying esophageal carcinogenesis and its molecular pathology will facilitate the development of biomarkers for early detection.

The application of microarray analysis is a promising method for finding clinical biomarkers in various cancers and has been successful in identifying subsets of tumors (including ESCC) that correlate with clinical parameters such as survival, histological grade, invasive status, and response to therapy (4-12). Gene expression changes that distinguish patient outcomes are subtle or variable and it is unlikely that individual genes will successfully predict clinical behavior. Taken together, however, gene expression profiles can be used to generate accurate predictors and could give us a better understanding of the molecular alterations during carcinogenesis.

In earlier studies, we documented genomic changes in ESCC, including widespread allelic loss and frequent mutations in certain putative tumor suppressor genes (13-17). Using a cDNA microarray with 7689 human cDNA clones, we previously tested expression in 19 ESCC patients and found 41 significant differentially-expressed genes. Patients with and without a positive family history for upper gastrointestinal (UGI) tract cancers were also distinguishable by their gene expression patterns (18). To confirm these original results, we expanded our expression studies of ESCC to evaluate more cases using alternative methods, including examination of more genes as well as validation/replication of selected dysregulated genes in additional patients using different methods. This confirmatory study was primarily based on global gene expression profiling of 53 ESCC patients with the Affymetrix Human Genome U133 Set (U133A and 133B). To further validate/replicate our findings we compared these data with: (i) RNA expression in 51 additional ESCC cases for seven differentially-expressed genes using quantitative real time RT-PCR; (ii) RNA expression from micro-dissected tumor and normal tissues in 17 ESCC cases for 41 dysregulated genes using the Affymetrix Human Genome U133A v2.0; and (iii) protein expression of six genes using immunohistochemistry (IHC) in 275 ESCCs on a tumor tissue microarray (TMA).

Methods

Patient selection

Four different groups of patients with ESCC were evaluated in this study, and all were enrolled in our UGI cancer genetic studies project, a single institution study using a common

research protocol. Patients enrolled in the project included consecutive cases of ESCC who presented to the Thoracic Surgery Department of the Shanxi Cancer Hospital in Taiyuan, Shanxi Province, People's Republic of China, between 1998-2001, who had no prior therapy for their cancer, and who underwent surgical resection of their tumor at the time of their hospitalization. Selection of patients for RNA studies was based solely on the availability of appropriate tissues for RNA testing (ie, consecutive testing of cases with available frozen tissue, tumor samples that were predominantly (>50%) tumor, and tissue RNA quality/quantity adequate for testing); patients without frozen tissues were included in the protein studies. After obtaining informed consent, patients were interviewed to obtain information on demographic and lifestyle cancer risk factors, and clinical data were collected. Selected demographic and clinical-pathologic features of the four different ESCC patient groups studied are shown in Table 1. In total, 396 different ESCC cases were evaluated. All cases were histologically confirmed as ESCC by pathologists at both the Shanxi Cancer Hospital and the NCI. This study was approved by the Institutional Review Boards of the Shanxi Cancer Hospital and the NCI.

Tissue collection

Paired esophageal cancer and normal tissue distant to the tumor were collected during surgery. Tissues for RNA analyses were snap-frozen in liquid nitrogen and stored at -130°C until used, while tissues for IHC analyses were fixed in 70% alcohol and processed to paraffin.

Total RNA preparation

RNA was extracted by two methods. For the confirmatory analysis of ESCC cases with the Affymetrix U133A/B chip set and the validation/replication in cases using real time RT-PCR, total RNA was extracted by the Trizol method following the protocol of the manufacturer. Only tumor samples with high purity ($\geq 50\%$ tumor cells) were selected for this extraction and subsequent analyses. A second method of RNA extraction was used for the micro-dissected tissue samples. For these samples, five to ten consecutive 8-micron sections were cut from frozen tumor tissues and the normal counterpart tissues, and tumor and/or normal cells were manually micro-dissected under light microscopy. RNA from tumor and matched normal tissue was extracted using the protocol from PureLink RNA mini kit (Catalog number 12183-018A, Invitrogen, Carlsbad, CA, 92008, USA). For both extraction methods, the quality and quantity of total RNA were determined on the RNA 6000 Labchip/Agilent 2100 Bioanalyzer (Agilent Technology, Inc, Germantown, MD).

Probe preparation and hybridization

Each microarray experiment was performed using eight micrograms of total RNA obtained from either the Trizol or PureLink extraction methods. Probes were prepared according to the protocol provided by the manufacturer (19). Procedures included first strand synthesis, second strand synthesis, double-strand cDNA clean up, in vitro transcription, cRNA purification, and fragmentation. Twenty micrograms of biotinylated cRNA were finally applied to each hybridization array, either onto the Affymetrix GeneChip Human Genome U133 Set (HG_U133A and HG_U133B, Affymetrix, Santa Clara, CA) or the Affymetrix GeneChip Human Genome U133A 2.0. After hybridization at 45°C overnight, arrays were developed with phycoerythrin-conjugated streptavidin using a fluidics station (Genechip Fluidics Station 450, Santa Clara, CA) and scanned (Genechip Scanner 3000, Santa Clara, CA) to obtain quantitative gene expression levels. Paired tumor and normal tissue specimens from each patient were processed simultaneously during the RNA extractions and hybridizations.

Quantitative RT-PCR

Confirmation by real-time RT-PCR analysis for seven genes was performed on an ABI 7000 Sequence Detection System using paired tumor/normal ESCC samples from 51 cases as previously described (20). Briefly, one to five micrograms of total RNA were first converted to cDNA using Superscript II (Invitrogen Corporation) in the presence of an oligo (dT)12-18 primer, and 100 ng of cDNA was applied for the subsequent PCR reaction (94°C × 10 min; 95°C × 15 seconds, 60°C × 1 min; 40 cycles). Results of the real-time RT-PCR data are presented as C_T values, where C_T is defined as the threshold PCR cycle number at which an amplified product is first detected. The average C_T was calculated for each gene evaluated and GAPDH, and the ΔC_T was determined as the mean of the triplicate C_T values for the evaluated gene minus the mean of the triplicate C_T values for GAPDH. The $\Delta\Delta C_T$ represents the difference between the paired tissue samples, as calculated by the formula $\Delta\Delta C_T = (\Delta C_T \text{ of tumor} - \Delta C_T \text{ of normal})$. The N-fold differential expression of the evaluated gene for a tumor sample compared with its normal epithelial counterpart was expressed as $2^{-\Delta\Delta C_T}$, which represents the fold change in the target gene expression in tumor normalized to an internal control gene (GAPDH) and relative to the normal control.

Immunohistochemistry (IHC) analysis of ESCC tumor microarray (TMA)

The details of patient selection and TMA construction were previously described (20). Six genes that were significantly over- or under-expressed on our previous 8K cDNA array were selected for IHC evaluation on the ESCC tumor TMA. These included *CDC25B*, *LAMC2*, *FADD*, *KRT14*, *FSCN1* (all over-expressed) and *KRT4* (under-expressed).

Slides were stained according to manufacturer's protocols for each of the seven gene proteins (for details, see Supplemental Table 1). In brief, five μm thickness deparaffinized sections were pretreated with 3% H_2O_2 in methanol for 10 minutes. Antigen retrieval included pressure cooker treatment for 5 or 25 minutes and 10% normal goat serum for one hour to block endogenous peroxidase activity, followed by incubation with primary antibodies at an appreciated dilution of 1:40 or 1:50 for overnight at 4°C. The next day the slides were treated using the secondary antibody (anti-mouse IgG (H+L), Vector Laboratories, Burlingame, CA, 1:500 dilution) for one hour at room temperature, followed by the ABC (Vector Laboratories, Burlingame, CA) solution for one hour at room temperature. Slides were developed with 0.02% 3', 3'-diaminobenzidine solution (DAB, Sigma), counterstained with hematoxylin, dehydrated in ethanol, and cleared in xylene. These procedures were performed for all antibodies studied.

Immunohistochemical assessment

For assessment of gene proteins, two scores were assigned to each core: (i) the cytoplasmic staining intensity [categorized as 0 (absent), 1 (weak), 2 (moderate), or 3 (strong)]; and (ii) the percentage of positively stained epithelial cells [scored as 0 (0%), 1 (1-25%), 2 (26-50%), 3 (51-75%), or 4 (>75%)]. An overall protein expression score was calculated by multiplying the intensity and positivity scores (overall score range, 0-12). This overall score for each patient was further simplified by dichotomizing it to negative (overall score of ≤ 3) or positive (score of ≥ 4). Stains were reviewed by two pathologists (MK and SMH) and discussed to determine an appropriate analytic approach. Following the establishment of criteria, all cores on both arrays were read by a single pathologist (MT) using the described criteria.

Statistical analyses

Formal statistical analyses were applied only to cases studied with the Affymetrix U133A/B set. Data from the other three groups studied here were limited to descriptive statistics. For

all the Affymetrix U133A/B array data, raw data sets (CEL files on all samples) after scanning were normalized using RMA, implemented in Bioconductor in R (21). The GEO accession numbers for these array data are GSE23400. Hierarchical clustering was performed to characterize RNA array expression patterns and distinguish differences between tumor and normal samples. Paired t-tests were used to identify differences in matched tumor/normal sample expression. Paired t-tests were all performed using the R package.

Results

Patient information

Characteristics of the 396 total patients evaluated here are shown in Table 1. The four separate study groups included cases evaluated by the Affymetrix U133A/B chip set (n=53), RT-PCR (n=51), the Affymetrix U133 V2 chip (n=17), and the tumor TMA (n=275). The median age for cases in the four study groups ranged from 53 to 58 years, males predominated in all but one of the groups, tobacco use (medians 13 to 60%) and alcohol use (medians 12 to 53%) were common as was a family history of UGI cancer (medians 26 to 47%). The vast majority of the tumors were grade 2, over three-fourths were stage III, and metastatic disease was evident for nearly half the cases.

Affymetrix U133A/B experimental quality control

In the present study, we used the Affymetrix Human U133 set (Chip A and Chip B) which contain 39,000 transcripts and variants, including approximately 33,000 well-substantiated human genes in greater than 45,000 probesets. We assayed hybridization quality using the Affymetrix GCOS software. The average MAS5 Present call of the 106 HG_U133A chips from the 53 ESCC patients was 50% (range 41 – 59%), average scale factor was 3.0 (range 1.7–8.7), average background was 58.5 (range 36.5–81.2), average noise was 2.5 (range 1.3–3.56), and ratio of 3'/5' signal of house keeping gene GAPDH was 0.9 (range 0.7–1.3). Averages for the 102 HG_U133B chips from 51 ESCC patients (two cases had no total RNA left) were: Present call 34% (range 22 – 42%), scale factor 7.7 (range 4.4 – 15.6), background 65.5 (range 36.8–145.3), noise 2.8 (range 1.5–5.9), and ratio of 3'/5' signal of housekeeping gene GAPDH 1.0 (range 0.8–1.6). Other sample quality control parameters built into the chips by Affymetrix were also consistent with high quality data. Expression signals for all probesets were used for the analysis.

Hierarchical clustering analysis of gene expression data

We used hierarchical clustering to characterize gene expression for all tumor/normal tissue pairs that had both U133 A and B array data (n=51 pairs). First, we selected the 10% of probesets (n=4498) that had the highest variation across all 102 samples examined (variance > 0.31). An unsupervised 2-way hierarchical clustering analysis with the 4498 probesets clearly separated tumors from normal samples (Supplemental Figure 1). Only two normal samples and three tumors were misclassified based on this structure of two clusters. Tumors were further separated into several sub-clusters, although no clinical data, such as grade, stage, and metastasis, were associated with these sub-clusters.

Identification of genes differentially expressed between tumors and normal samples

To identify genes whose expression levels were altered in tumors, we performed paired t-tests for 53 cases with the Affymetrix U133A/B chip data. We found 642 genes (854 probesets) that showed significant differences in gene expression between tumor and normal tissues; these genes showed 2-fold or greater changes and were statistically significant after Bonferroni correction (ie, *P*-values less than 1.12E-6) (Supplemental Table 2). To highlight

a shorter list of target genes, we also applied a more extreme p-value criterion (P -value $< E-15$) in conjunction with at least a 2-fold change, which identified 159 genes – 116 up-regulated genes (Table 2A) and 43 down-regulated genes (Table 2B).

Affymetrix U133 v2.0 micro-dissected tissue validation

In our initial RNA expression study (ie, 8K cDNA study) [18], we identified 41 differentially-expressed genes. As part of our validation efforts here, we also compared RNA expression for these 41 dysregulated genes by alternative methods, including different microarray platforms as well as different methods for RNA extraction and tissue procurement. These comparisons included results from three sets of analyses involving independent samples, consisting of the 8K cDNA study (N=19), the Affymetrix U133A/B chip set (N=53), and the Affymetrix U133 V2 chip (N=17). Both the 8K cDNA and the Affymetrix U133A/B chip studies used total RNA extracted with the Trizol method but without micro-dissection. The Affymetrix U133 V2 chip study employed micro-dissected tissues from which RNA was extracted with the PureLink protocol.

A cross-platform comparison between the 8K cDNA and Human U133A/B set showed that, of the 41 dysregulated genes from our previous study, 40 were evaluable on both platforms (one gene was not found in the Affymetrix probeset). Of these 40 genes, all but one (*CD3EAP*) showed the same gene expression pattern on both platforms (ie, both up- or both down-regulated) (Table 3). In addition to the directionality of the changes, the magnitude of the changes was also very similar: changes were 2-fold or greater for 10 of 13 (77%) up-regulated genes on both platforms, while 19 of 28 (68%) down-regulated genes showed fold changes of 0.50 or less on both platforms.

A comparison of different RNA extraction methods applied to the Affymetrix platforms showed that, of 38 genes examined on both Affymetrix platforms, only one had a different expression pattern (Table 3). *EGR1* was down-regulated on the Affymetrix U133A/B chip set (fold change 0.56), but up-regulated on the Affymetrix U133 v2 chip (fold change 1.19).

It is also apparent from inspection of the data in Table 3 that the magnitude of the fold changes among up-regulated genes appears to be highest in the micro-dissected tissue samples (ie, Affymetrix U133A). For example, among the 13 up-regulated genes, none tested on the 8K cDNA array showed a fold change of three or more, while six exceeded 3-fold changes on the Affymetrix U133A/B chip set, and eight were higher than 3-fold on the Affymetrix U133 V2 chip, including five cases which reached over 5-fold changes. Although less consistent, the magnitude of the fold changes among down-regulated genes also appeared to be more extreme in the micro-dissected tissue samples.

Quantitative real time RT-PCR validation

Seven genes were selected for validation (Table 4) in a new group of 51 ESCC cases as illustrative examples of the genes which showed the most prominent differences in either the current Affymetrix or the prior cDNA array evaluations. Among the seven selected genes, four were up-regulated (*COL1A2*, *COL3A1*, *MET*, and *KRT14*) and three were down-regulated (*SPINK7/ECG2*, *HPGD*, and *SASH1*). Briefly, in at least two-thirds or more of the 51 patients, all four up-regulated genes showed increased mRNA expression (≥ 2 -fold in tumor vs normal) while all three down-regulated genes showed decreased mRNA expression (≤ 0.5 -fold in tumor vs normal). Specifically, *KRT14* was increased in 67% of cases, *COL1A2* in 67%, *COL3A1* in 84%, and *MET* in 72%. Likewise, *ECG2* was decreased in 84% of cases, *HPGD* in 80%, and *SASH1* in 67%.

Protein expression validation

Tumor tissue samples from 313 ESCC cases were arrayed on the tumor TMA. After exclusion of cores with inadequate tissue following sectioning and tissue transfer, a total of 275 ESCC cases had IHC-based protein expression data available for at least one of the six markers evaluated as part of our validation here (Table 1). Protein expression positivity (number of evaluable ESCC cases) was: CDC25B 59% (N=275), LAMC2 82% (275), FADD 15% (248), KRT14 33% (249), FSCN1 56% (231) and KRT4 84% (171).

Discussion

In the present study we compared genome-wide gene expression in the tumors from 53 ESCC cases to their matched normal tissue samples. We found that tumors and normal tissues had intrinsically different expression patterns and were easily separated into two clusters based on unsupervised 2-way hierarchical clustering analysis. We identified 642 genes whose gene expressions differed between tumors and normal tissues using typical criteria (at least 2-fold change and P -value less than $1.12\text{E-}6$).

Several recent studies analyzed gene expression profiling for ESCC (22-25). One study also used Affymetrix chips but studied only 15 ESCC cases (24), while other studies applied cDNA microarrays with Cy3 or Cy5 labeling that examined more limited numbers of genes (22;23). The largest number of ESCC cases studied was a Japanese report of 54 cases examined with the Affymetrix Human U133 A chip; however, pair-matched normal tissues were not used in that study (26). Thus, among genome-wide expression studies employing the optimal design – pair-wise matched tumor/normal tissue comparisons – the present study is the largest (53 cases) and most comprehensive (33,000 genes) ESCC evaluation to date.

With extreme statistical criteria (P -values $< \text{E-}15$ and at least a 2-fold change) the number of dysregulated genes was reduced to 159 (Tables 2A and 2B). The functions of these 159 genes most prominently relate to biochemical enzymes (26 genes), protein transportation or binding (23 genes), DNA replication (20 genes), cell cycle regulation (19 genes), cell membrane proteins (16 genes), extracellular matrix (13 genes), and cell growth (11 genes). Some of these genes (eg, *MMP*, collagen families, keratins, *CDC25B*, calcium-binding S100 proteins, and Annexin families) have previously been shown to function in squamous cell differentiation, invasion, or proliferation (27-29).

Examination of mRNA expression by RT-PCR for seven array-dysregulated genes in an independent series of ESCC cases showed results that were highly comparable with both our current Affymetrix U133A/B chip data and the findings from our earlier cDNA microarray study (18). Taken together, these results indicate that gene expression profiles in ESCC are consistent across different platforms and that dysregulated gene expression is a reproducible biomarker discovery tool.

We previously found 41 differentially-expressed genes in ESCC cases using an 8K cDNA microarray (18). All 41 of these genes were evaluated in the current Affymetrix-based study and all showed the same tumor/normal expression ratio directionality, save for one gene (*FOSL2*). Both the cDNA and the Affymetrix arrays used total RNA extracted by the Trizol method. To minimize the impact of normal contamination of our tumor samples, we further evaluated these 41 genes using micro-dissected RNA procured from another set of 17 ESCC cases and tested with Affymetrix U133A v2.0 chips. Results showed that the tumor:normal expression ratio directionality was the same for most of the genes (88%), however, the magnitude of the fold changes was markedly higher in the micro-dissected as opposed to the non-micro-dissected samples (8K cDNA and U133A/B set chip studies). We presume that this reflected reduced heterogeneity of the tissue samples when micro-dissection was

employed, and a consequent increased signal-to-noise ratio. For example, *COLIA2* just reached the 2-fold change threshold in the 8K cDNA array study, was 6.6-fold increased in the Affymetrix U133A/B set array, but was nearly 12-fold increased when micro-dissected RNA was used with the Affymetrix U133A v2.0 array (Table 3). While *COLIA2* was the most extreme and clear-cut example of this increased signal-to-noise ratio, among 38 (of the 41) genes evaluated here, 28 showed their most extreme fold changes (either increased or decreased) in micro-dissected samples. Our comparison studies show that micro-dissection is a powerful approach. Results here are in agreement with other observations showing that micro-dissection provides relatively pure cell populations that are particularly useful for interrogating specific targets of interest (30).

Results from all three experiments reported here show broadly uniform findings for the expression patterns of the 41 genes emphasized, with the largest fold changes predominantly from the array that used micro-dissected RNA. To our knowledge, this is the first report confirming differential gene expression performed using two different microarray platforms and two different RNA extraction methods.

We chose six genes to evaluate at the protein level by applying IHC techniques to our ESCC tumor TMA. Three of the up-regulated genes (*CDC25B* 59%, *LAMC2* 82%, and *FSCN1* 56%) showed positive protein expression in the majority of ESCC cases studied, results that were highly concordant with RNA expression results. The other two up-regulated genes showed positive protein expression but in less than half of the ESCC cases studied (*KRT14* 33% and *FADD* 15%). The down-regulated gene, *KRT4*, was positive for protein expression in 84% of ESCCs, which did not correlate well with RNA results.

CDC25B has been shown to be a potential early biomarker as its protein expression increased with morphologic progression across the continuum of normal to dysplasia to invasive ESCC in our previous study (20). Patterns of *LAMC2* protein expression showed a strong relation to survival, suggesting a potential role in prognosis (20). The expression of *FSCN1* protein in epithelial neoplasms has been described (31-34), but its expression in ESCC is still unknown. The present study showed that *FSCN1* protein expression was observed in most ESCC tissue cores (68%), which is in accord with RNA expression findings. While *KRT14* protein expression was high in ESCCs in the current study, dysplastic and normal esophagus tissues were not evaluated. Xue et al did evaluate normal, dysplastic, and invasive ESCCs within esophagectomies from the same cases and observed that protein expression positivity increased across this morphologic progression from 13% to 41% to 62%, respectively, suggesting some discrimination between clinically and diagnostically important categories (35).

KRT4 protein expression has previously been reported in several tumors of the upper digestive tract, including esophageal adenocarcinoma (36). We found *KRT4* mRNA down-regulated in both our 8K and Affymetrix microarray studies, yet protein expression was positive in 84% of ESCCs in our tumor TMA. Chung et al (37) reported that *KRT4* protein expression decreased in the transition from normal to dysplasia to invasive tumor in a study of six carefully characterized ESCC cases. Of additional interest, ESCC cases with higher *KRT4* mRNA in the present study had longer survival.

In summary, we identified an expanded list of 642 dysregulated genes in ESCC. These genes provide potential new targets for early detection and treatment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by National Cancer Institute contract [N02-SC-66211] with the Shanxi Cancer Hospital and Institute; and by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics, and Center for Cancer Research.

References

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin.* 2005; 55:74–108. [PubMed: 15761078]
2. Yang L, Parkin DM, Ferlay J, et al. Estimates of cancer incidence in China for 2000 and projections for 2005. *Cancer Epidemiol Biomarkers Prev.* 2005; 14:243–50. [PubMed: 15668501]
3. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin.* 2010; 60:277–300. [PubMed: 20610543]
4. van d V, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med.* 2002; 347:1999–2009. [PubMed: 12490681]
5. Takahashi M, Rhodes DR, Furge KA, et al. Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. *Proc Natl Acad Sci U S A.* 2001; 98:9754–9. [PubMed: 11493696]
6. Tamoto E, Tada M, Murakawa K, et al. Gene-expression profile changes correlated with tumor progression and lymph node metastasis in esophageal cancer. *Clin Cancer Res.* 2004; 10:3629–38. [PubMed: 15173069]
7. Ishibashi Y, Hanyu N, Nakada K, et al. Profiling gene expression ratios of paired cancerous and normal tissue predicts relapse of esophageal squamous cell carcinoma. *Cancer Res.* 2003; 63:5159–64. [PubMed: 12941848]
8. Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A.* 2001; 98:13790–5. [PubMed: 11707567]
9. Selaru FM, Zou T, Xu Y, et al. Global gene expression profiling in Barrett's esophagus and esophageal cancer: a comparative analysis using cDNA microarrays. *Oncogene.* 2002; 21:475–8. [PubMed: 11821959]
10. Hu YC, Lam KY, Law S, et al. Profiling of differentially expressed cancer-related genes in esophageal squamous cell carcinoma (ESCC) using human cancer cDNA arrays: overexpression of oncogene MET correlates with tumor differentiation in ESCC. *Clin Cancer Res.* 2001; 7:3519–25. [PubMed: 11705871]
11. Kan T, Shimada Y, Sato F, et al. Gene expression profiling in human esophageal cancers using cDNA microarray. *Biochem Biophys Res Commun.* 2001; 286:792–801. [PubMed: 11520067]
12. Hu YC, Lam KY, Law S, et al. Identification of differentially expressed genes in esophageal squamous cell carcinoma (ESCC) by cDNA expression array: overexpression of Fra-1, Neogenin, Id-1, and CDC25B genes in ESCC. *Clin Cancer Res.* 2001; 7:2213–21. [PubMed: 11489794]
13. Wu, M.; Hu, N.; Wang, XQ. Genetic factor in the etiology of esophageal cancer and the strategy of its prevention in high-incidence areas of North China. In: Lynch, HT.; Hirayama, T., editors. *Genetic Epidemiology for Cancer.* CRC Press Inc; Boca Raton, FL: 1989. p. 187-200.
14. Hu N, Dawsey SM, Wu M, et al. Familial aggregation of esophageal cancer in Yangcheng County, Shanxi Province, China. *Int J Epidemiol.* 1992; 21:877–82. [PubMed: 1468848]
15. Hu N, Li WJ, Su H, et al. Common genetic variants of TP53 and BRCA2 in esophageal cancer patients and healthy individuals from low and high risk areas of northern China. *Cancer Detect Prev.* 2003; 27:132–8. [PubMed: 12670525]
16. Li WJ, Hu N, Su H, et al. Allelic loss on chromosome 13q14 and mutation in deleted in cancer 1 gene in esophageal squamous cell carcinoma. *Oncogene.* 2003; 22:314–8. [PubMed: 12527901]
17. Hu N, Wang C, Su H, et al. High frequency of CDKN2A alterations in esophageal squamous cell carcinoma from a high-risk Chinese population. *Genes Chromosomes Cancer.* 2004; 39:205–16. [PubMed: 14732922]

18. Su H, Hu N, Shih J, et al. Gene expression analysis of esophageal squamous cell carcinoma reveals consistent molecular profiles related to a family history of upper gastrointestinal cancer. *Cancer Res.* 2003; 63:3872–6. [PubMed: 12873975]
19. Affymetrix. Santa Clara, CA: Affymetrix; 2001. GeneChip Expression Analysis Technical Manual. Available from: URL: http://www.affymetrix.com/support/technical/manual/expression_manual.affx
20. Shou JZ, Hu N, Takikita M, et al. Overexpression of CDC25B and LAMC2 mRNA and protein in esophageal squamous cell carcinomas and premalignant lesions in subjects from a high-risk population in China. *Cancer Epidemiol Biomarkers Prev.* 2008; 17:1424–35. [PubMed: 18559558]
21. Fred Hutchinson Cancer Research Center. Bioconductor - Open Source Software for Bioinformatics. 2011. Available from: URL: <http://www.bioconductor.org/>
22. Sato T, Iizuka N, Hamamoto Y, et al. Esophageal squamous cell carcinomas with distinct invasive depth show different gene expression profiles associated with lymph node metastasis. *Int J Oncol.* 2006; 28:1043–55. [PubMed: 16596220]
23. Yamabuki T, Daigo Y, Kato T, et al. Genome-wide gene expression profile analysis of esophageal squamous cell carcinomas. *Int J Oncol.* 2006; 28:1375–84. [PubMed: 16685439]
24. Uchikado Y, Inoue H, Haraguchi N, et al. Gene expression profiling of lymph node metastasis by oligomicroarray analysis using laser microdissection in esophageal squamous cell carcinoma. *Int J Oncol.* 2006; 29:1337–47. [PubMed: 17088971]
25. Wong FH, Huang CY, Su LJ, et al. Combination of microarray profiling and protein-protein interaction databases delineates the minimal discriminators as a metastasis network for esophageal squamous cell carcinoma. *Int J Oncol.* 2009; 34:117–28. [PubMed: 19082484]
26. Kashyap MK, Marimuthu A, Kishore CJ, et al. Genomewide mRNA profiling of esophageal squamous cell carcinoma for identification of cancer biomarkers. *Cancer Biol Ther.* 2009; 8:36–46. [PubMed: 18981721]
27. Zhang X, Zhi HY, Zhang J, et al. Expression of annexin II in human esophageal squamous cell carcinoma. *Zhonghua Zhong Liu Za Zhi.* 2003; 25:353–5. [PubMed: 12921564]
28. Luo A, Kong J, Hu G, et al. Discovery of Ca²⁺-relevant and differentiation-associated genes downregulated in esophageal squamous cell carcinoma using cDNA microarray. *Oncogene.* 2004; 23:1291–9. [PubMed: 14647409]
29. Lee DG, Bell SP. ATPase switches controlling DNA replication initiation. *Curr Opin Cell Biol.* 2000; 12:280–5. [PubMed: 10801458]
30. Erickson HS, Gillespie JW, Emmert-Buck MR. Tissue microdissection. *Methods Mol Biol.* 2008; 424:433–48. [PubMed: 18369881]
31. Jawhari AU, Buda A, Jenkins M, et al. Fascin, an actin-bundling protein, modulates colonic epithelial cell invasiveness and differentiation in vitro. *Am J Pathol.* 2003; 162:69–80. [PubMed: 12507891]
32. Grothey A, Hashizume R, Sahin AA, McCrea PD. Fascin, an actin-bundling protein associated with cell motility, is upregulated in hormone receptor negative breast cancer. *Br J Cancer.* 2000; 83:870–3. [PubMed: 10970687]
33. Pelosi G, Pastorino U, Pasini F, et al. Independent prognostic value of fascin immunoreactivity in stage I nonsmall cell lung cancer. *Br J Cancer.* 2003; 88:537–47. [PubMed: 12592367]
34. Hashimoto Y, Shimada Y, Kawamura J, et al. The prognostic relevance of fascin expression in human gastric carcinoma. *Oncology.* 2004; 67:262–70. [PubMed: 15557788]
35. Xue LY, Hu N, Song YM, et al. Tissue microarray analysis reveals a tight correlation between protein expression pattern and progression of esophageal squamous cell carcinoma. *BMC Cancer.* 2006; 6:296. [PubMed: 17187659]
36. Tanieri P, Martel-Planche G, Maurici D, et al. Molecular and clinical differences between adenocarcinomas of the esophagus and of the gastric cardia. *Am J Pathol.* 2001; 158:33–40. [PubMed: 11141476]
37. Chung JY, Braunschweig T, Hu N, et al. A multiplex tissue immunoblotting assay for proteomic profiling: a pilot study of the normal to tumor transition of esophageal squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2006; 15:1403–8. [PubMed: 16835344]

Table 1

Summary of characteristics of patients in the 4 groups studied

	Study group by lab method applied			
	Cases studied with Affymetrix U133A/B set (N=53)	Cases studied with Affymetrix U133A V2 (N=17)	Cases studied with real time qRT-PCR (N=51)	Cases studied with tumor tissue microarray (N=275)
Gender (% male)	0.64	0.35	0.61	0.66
Age (years, median)	58	53	54	57
Tobacco use (% yes)	0.59	0.13	0.53	0.6
Alcohol use (% daily or weekly)	0.53	0.12	0.14	0.21
Family history of upper gastrointestinal cancer (% yes)	0.32	0.47	0.33	0.26
Tumor grade (%)				
1	0.09	0.00	0.06	0.17
2	0.72	0.88*	0.69	0.59
3	0.19	0.12	0.25	0.23
4	0.00	0.00	0.00	<0.01
Tumor stage (%)				
I	0.00	0.00	0.02	<0.01
II	0.25	0.24	0.24	0.12
III	0.74	0.76	0.75	0.87
IV	0.02	0.00	0.00	0.01
Lymph node metastasis (% yes)	0.43	0.47	0.51	0.44

* One case missing information

Table 2

Table 2A: Summary of over-expressed genes ($P < E-15$)^{*,†}

No.	Symbol	Gene name and related function	Locus ID	P-value	Fold-change
Extracellular matrix					
1	<i>MMP1</i>	matrix metalloproteinase 1	4312	8.61E-22	21.74
2	<i>CTHRC1</i>	collagen triple helix repeat containing 1	115908	8.63E-26	14.71
3	<i>SPP1</i>	secreted phosphoprotein 1	6696	5.56E-21	9.11
4	<i>COL1A1</i>	collagen, type I, alpha 1	1277	3.07E-25	7.94
5	<i>COL1A2</i>	collagen, type I, alpha 2	1278	6.15E-21	6.49
6	<i>COL11A1</i>	collagen, type XI, alpha 1	1301	3.40E-16	5.30
7	<i>COL3A1</i>	collagen, type III, alpha 1	1281	1.16E-19	4.48
8	<i>COL5A2</i>	collagen, type V, alpha 2	1290	1.34E-16	4.25
9	<i>ANLN</i>	anillin, actin binding protein	54443	8.78E-17	4.18
10	<i>PLAU</i>	plasminogen activator, urokinase	5328	5.52E-18	3.98
11	<i>SPARC</i>	secreted protein, acidic, cysteine-rich	6678	7.78E-17	3.45
12	<i>MFAP2</i>	microfibrillar-associated protein 2	4237	1.73E-21	3.23
13	<i>COL7A1</i>	collagen, type VII, alpha 1	1294	2.34E-18	2.65
Cell adhesion					
14	<i>POSTN</i>	periostin, osteoblast specific factor	10631	2.01E-18	8.37
15	<i>CDH11</i>	cadherin 11, type 2, OB-cadherin	1009	1.33E-18	5.17
16	<i>CSPG2</i>	chondroitin sulfate proteoglycan 2	1462	1.30E-17	4.61
17	<i>LAMB3</i>	laminin, beta 3	3914	4.81E-17	3.49
18	<i>THBS2</i>	thrombospondin 2	7058	8.35E-17	2.86
19	<i>PTK7</i>	PTK7 protein tyrosine kinase 7	5754	2.02E-18	2.02
DNA replication/repair/transcription					
20	<i>SNAI2</i>	snail homolog 2	6591	1.35E-19	4.60
21	<i>TOP2A</i>	topoisomerase (DNA) II alpha 170kDa	7153	1.14E-17	4.01
22	<i>SOX4</i>	SRY (sex determining region Y)-box 4	6659	6.78E-19	3.45
23	<i>RFC4</i>	replication factor C (activator 1) 4, 37kDa	5984	1.92E-23	3.29
24	<i>GIN51</i>	GIN5 complex subunit 1	9837	7.69E-20	2.94

Table 2A: Summary of over-expressed genes ($P < E-15$)*,†

No.	Symbol	Gene name and related function	Locus ID	P-value	Fold-change
25	<i>HMGCB3</i>	high-mobility group box 3	3149	1.73E-16	2.94
26	<i>MCM2</i>	MCM2 minichromosome maintenance deficient 2, mitotin	4171	2.89E-20	2.70
27	<i>GMNN</i>	geminin, DNA replication inhibitor	51053	4.69E-17	2.49
28	<i>UHRF1</i>	ubiquitin-like, containing PHD and RING finger domains, 1	29128	1.32E-16	2.42
29	<i>MCM6</i>	MCM6 minichromosome maintenance deficient 6	4175	1.58E-20	2.41
30	<i>PCNA</i>	proliferating cell nuclear antigen	5111	4.42E-17	2.39
31	<i>MCM5</i>	MCM5 minichromosome maintenance deficient 5	4174	5.36E-16	2.34
32	<i>MCM4</i>	MCM4 minichromosome maintenance deficient 4	4173	4.32E-16	2.20
33	<i>FEN1</i>	flap structure-specific endonuclease 1	2237	2.96E-17	2.11
34	<i>MSH6</i>	mutS homolog 6 (E. coli)	2956	1.83E-16	2.10
35	<i>TOPBP1</i>	topoisomerase (DNA) II binding protein 1	11073	8.45E-18	2.10
36	<i>FANCI</i>	Fanconi anemia, complementation group I	55215	3.80E-18	2.10
37	<i>RAD51API</i>	RAD51 associated protein 1	10635	1.70E-16	2.09
38	<i>ZNF281</i>	zinc finger protein 281	23528	3.47E-16	2.08
Cell growth/proliferation/differentiation factor					
39	<i>ECT2</i>	epithelial cell transforming sequence 2 oncogene	1894	2.17E-22	4.32
40	<i>ASPM</i>	asp (abnormal spindle)-like, microcephaly associated	259266	5.15E-18	3.23
41	<i>PRCI</i>	protein regulator of cytokinesis 1	9055	3.25E-16	3.12
42	<i>FSCN1</i>	fascin homolog 1, actin-bundling protein	6624	4.16E-19	3.06
43	<i>NUSAP1</i>	nucleolar and spindle associated protein 1	51203	3.14E-16	2.92
44	<i>MET</i>	met proto-oncogene (hepatocyte growth factor receptor)	4233	5.76E-18	2.76
45	<i>CDKN3</i>	cyclin-dependent kinase inhibitor 3	1033	1.98E-17	2.47
46	<i>TPX2</i>	TPX2, microtubule-associated protein homolog	22974	5.17E-17	2.40
47	<i>KIF20A</i>	kinesin family member 20A	10112	6.32E-18	2.19
48	<i>AURKB</i>	aurora kinase B	9212	3.93E-16	2.09
Cell cycle regulators					
49	<i>CXS2</i>	CDC28 protein kinase regulatory subunit 2	1164	2.78E-18	3.91
50	<i>CEP55</i>	centrosomal protein 55KDa	55165	1.53E-19	3.59
51	<i>CDC20</i>	CDC20 cell division cycle 20 homolog	991	2.79E-17	2.92

Table 2A: Summary of over-expressed genes ($P < E-15$)*,†

No.	Symbol	Gene name and related function	Locus ID	P-value	Fold-change
52	<i>CDC2</i>	cell division cycle 2	983	1.42E-19	2.91
53	<i>FOXMI</i>	forkhead box M1	2305	4.02E-18	2.90
54	<i>CKS1B</i>	CDC28 protein kinase regulatory subunit 1B BUB1 budding uninhibited by benzimidazoles 1 homolog	1163	6.99E-21	2.89
55	<i>BUB1B</i>	beta	701	5.45E-20	2.82
56	<i>CCNB1</i>	cyclin B1	891	1.45E-16	2.71
57	<i>NUF2</i>	cell division cycle associated 1	83540	6.23E-16	2.67
58	<i>DLGAP5</i>	discs, large homolog-associated 5	9787	7.79E-19	2.65
59	<i>BIRC5</i>	baculoviral IAP repeat-containing 5	332	3.86E-16	2.62
60	<i>MARCKSL1</i>	<i>MARCKS-like 1</i>	65108	2.86E-17	2.51
61	<i>NEK2</i>	NIMA (never in mitosis gene a)-related kinase 2	4751	3.88E-17	2.44
62	<i>CENPF</i>	centromere protein F, 350/400ka (mitosin)	1063	1.57E-18	2.41
63	<i>AURKA</i>	serine/threonine kinase 6	6790	3.67E-17	2.40
64	<i>MAD2L1</i>	MAD2 mitotic arrest deficient-like 1	4085	5.76E-16	2.36
65	<i>CENPA</i>	centromere protein A	1058	5.93E-18	2.17
66	<i>BUB1</i>	BUB1 budding uninhibited by benzimidazoles 1 homolog	699	2.34E-17	2.09
67	<i>ATR</i>	ataxia telangiectasia and Rad3 related	545	2.69E-18	2.06
Cell membrane protein/signal transduction					
68	<i>FZD6</i>	frizzled homolog 6 (Drosophila)	8323	2.29E-18	2.99
69	<i>NETO2</i>	neuropilin (NRP) and tolloid (TLL)-like 2	81831	1.06E-17	2.68
70	<i>LAPTM4B</i>	lysosomal associated protein transmembrane 4 beta	55353	1.26E-17	2.55
71	<i>PLXNA1</i>	plexin A1	5361	3.73E-19	2.50
72	<i>TBL1XR1</i>	transducin (beta)-like 1X-linked receptor 1	79718	8.86E-17	2.17
73	<i>EFNA1</i>	ephrin-A1	1942	5.74E-18	2.15
74	<i>PPIL5</i>	peptidylprolyl isomerase (cyclophilin)-like 5	122769	9.88E-17	2.10
75	<i>RANBP1</i>	RAN binding protein 1	5902	1.61E-18	2.10
76	<i>DPY19L4</i>	dpy-19-like 4	286148	5.74E-16	2.09
77	<i>LRRC8D</i>	leucine rich repeat containing 8 family, member D	55144	6.80E-16	2.09
78	<i>ATP6V1C1</i>	ATPase, H+ transporting	528	3.35E-16	2.08
79	<i>TMEM185B</i>	transmembrane protein 185 B	79134	4.69E-16	2.07

Table 2A: Summary of over-expressed genes ($P < E-15$)*,†

No.	Symbol	Gene name and related function	Locus ID	P-value	Fold-change
Protein binding/modification/transportation/protein chaperon					
80	<i>DTL</i>	denticless homolog (Drosophila)	51514	1.78E-18	2.86
81	<i>IGF2BP2</i>	insulin-like growth factor 2 mRNA binding protein 2	10644	4.70E-16	2.84
82	<i>SERPINH1</i>	serpin peptidase inhibitor, clade H, member 1,	871	1.88E-16	2.81
83	<i>SMC2</i>	structural maintenance of chromosomes 2	10592	5.63E-16	2.55
84	<i>SLC12A4</i>	solute carrier family 12, member 4	6560	1.39E-16	2.54
85	<i>TFRC</i>	transferrin receptor (p90, CD71)	7037	6.63E-18	2.53
86	<i>ATAD2</i>	ATPase family, AAA domain containing 2	29028	6.46E-16	2.33
87	<i>SLC16A1</i>	solute carrier family 16, member 1	6566	6.79E-16	2.28
88	<i>HSPH1</i>	heat shock 105kDa/110kDa protein	10808	4.07E-16	2.28
89	<i>UBE2T</i>	ubiquitin-conjugating enzyme E2T	29089	5.30E-16	2.24
90	<i>HOMER3</i>	homer homolog 3 (Drosophila)	9454	3.13E-16	2.31
91	<i>AGRIN</i>	agrin	375790	9.27E-21	2.13
92	<i>SLC33A1</i>	solute carrier family 33 (acetyl-CoA transporter), member 1	9197	2.82E-16	2.12
93	<i>NUP107</i>	nucleoporin 107kDa	57122	7.19E-16	2.12
94	<i>HSP90AA1</i>	heat shock 90kDa protein 1, alpha	3320	1.97E-16	2.12
95	<i>HSPE1</i>	heat shock 10kDa protein 1 (chaperonin 10)	3336	1.13E-16	2.10
Biochemical enzymes activity					
96	<i>SULF1</i>	sulfatase 1	23213	9.88E-20	4.36
97	<i>MEST</i>	mesoderm specific transcript homolog (mouse)	4232	3.27E-18	2.92
98	<i>MTHFD1L</i>	formyltetrahydrofolate synthetase domain containing 1	25902	2.66E-17	2.78
99	<i>TTK</i>	TTK protein kinase	7272	4.79E-17	2.77
100	<i>MTHFD2</i>	methylene tetrahydrofolate dehydrogenase	10797	7.75E-16	2.68
101	<i>PRKDC</i>	protein kinase, DNA-activated, catalytic polypeptide	5591	5.67E-18	2.40
102	<i>MELK</i>	maternal embryonic leucine zipper kinase	9833	5.94E-16	2.39
103	<i>HPRT1</i>	hypoxanthine phosphoribosyltransferase 1	3251	6.76E-18	2.24
104	<i>C20orf3</i>	chromosome 20 open reading frame 3	57136	6.35E-20	2.16
105	<i>DNMT1</i>	DNA (cytosine-5-)methyltransferase 1	1786	2.09E-18	2.15
106	<i>GMPS</i>	guanine monophosphate synthetase	8833	3.14E-16	2.09

Table 2A: Summary of over-expressed genes ($P < E-15$)*,†

No.	Symbol	Gene name and related function	Locus ID	P-value	Fold-change
107	<i>PTDSSI</i>	phosphatidylserine synthase 1	9791	3.30E-17	2.07
Calcium ion binding/transport					
108	<i>ITPR3</i>	<i>inositol 1,4,5-trisphosphate receptor, type 3</i>	3710	7.09E-19	2.19
Others					
109	<i>KIF4A</i>	kinesin family member 4A	24137	2.87E-21	2.62
110	<i>KIF14</i>	kinesin family member 14	9928	9.05E-20	2.46
111	<i>SR140</i>	U2-associated SR140 protein	23350	1.64E-18	2.38
112	<i>VOPPI</i>	vesicular, overexpressed in cancer, prosurvival protein 1	81552	7.83E-16	2.36
113	<i>ACTL6A</i>	actin-like 6A	86	4.17E-19	2.36
114	<i>TRIP13</i>	thyroid hormone receptor interactor 13	9319	2.16E-16	2.28
115	<i>CBX3</i>	chromobox homolog 3 (HP1 gamma homolog, Drosophila)	11335	1.47E-22	2.18
116	<i>SLC20A1</i>	solute carrier family 20 (phosphate transporter), member 1	6574	3.75E-17	2.13

Table 2B: Summary of under-expressed genes ($P < E-15$)*,†

No.	Symbol	Genes name and related function	Locus ID	P-value	Fold-change
DNA replication/transcription					
1	<i>KLF4</i>	Kruppel-like factor 4 (gut)	9314	5.08E-16	0.43
Cell growth/proliferation/differentiation factor					
2	<i>NDRG2</i>	NDRG family member 2	57447	2.93E-19	0.38
Cell membrane protein/signal transduction					
3	<i>MAL</i>	mal, T-cell differentiation protein	4118	2.78E-16	0.06
4	<i>SIM2</i>	single-minded homolog 2 (Drosophila)	6493	7.36E-18	0.35
5	<i>EPS8L2</i>	EPS8-like 2	64787	3.91E-16	0.38
6	<i>UBL3</i>	ubiquitin-like 3	5412	9.89E-19	0.40
Protein transportation/protein binding					
7	<i>SHROOM3</i>	Shroom family member 3	57619	1.43E-19	0.29
8	<i>SASH1</i>	SAM and SH3 domain containing 1	23328	2.48E-16	0.38

Table 2B: Summary of under-expressed genes ($P < E-15$)*,†

No.	Symbol	Genes name and related function	Locus ID	P-value	Fold-change
9	<i>SORBS2</i>	Arg/Abl-interacting protein ArgBP2	8470	6.01E-19	0.40
10	<i>CAB39L</i>	calcium binding protein 39-like	81617	1.50E-17	0.45
11	<i>SH3GLB2</i>	SH3-domain GRB2-like endophilin B2	56904	2.99E-16	0.46
12	<i>MPP7</i>	membrane protein, palmitoylated 7	143098	8.38E-18	0.47
13	<i>SORT1</i>	sortilin 1	6272	4.98E-16	0.48
Biochemical enzymes activity					
14	<i>PPP1R3C</i>	protein phosphatase 1, regulatory (inhibitor) subunit 3C	5507	2.88E-17	0.15
15	<i>HPGD</i>	hydroxyprostaglandin dehydrogenase 15-(NAD)	3248	2.96E-18	0.16
16	<i>ADH1B</i>	alcohol dehydrogenase IB (class I), beta polypeptide	125	1.44E-18	0.20
17	<i>GPX3</i>	glutathione peroxidase 3 (plasma)	2878	7.40E-16	0.23
18	<i>FUT6</i>	fucosyltransferase 6 (alpha (1,3) fucosyltransferase)	2528	2.93E-16	0.23
19	<i>CFD</i>	complement factor D	1675	2.13E-18	0.23
20	<i>MGLL</i>	monoglyceride lipase	11343	1.06E-16	0.28
21	<i>PCAF</i>	p300/CBP-associated factor glycerophosphodiester phosphodiesterase domain containing	8850	7.04E-18	0.29
22	<i>GDPD3</i>	3	79153	3.31E-16	0.39
23	<i>GPD1L</i>	glycerol-3-phosphate dehydrogenase 1-like	23171	4.34E-19	0.41
24	<i>THSD4</i>	thrombospondin, type I	79875	8.17E-16	0.41
25	<i>HLCS</i>	holocarboxylase synthetase	3141	1.18E-17	0.42
26	<i>ECHDC2</i>	enoyl CoA hydratase domain containing 2	55268	6.38E-16	0.43
27	<i>PAD11</i>	peptidyl arginine deiminase, type I	29943	6.21E-16	0.45
Calcium ion binding					
28	<i>NUCB2</i>	nucleobindin 2	4925	9.48E-17	0.25
Others					
29	<i>CRISP3</i>	cysteine-rich secretory protein 3	10321	2.92E-17	0.05
30	<i>ENDOU</i>	endonuclease, ployU-specific	8909	6.17E-16	0.15
31	<i>GCOM1</i>	GRINL1A complex upstream protein	145781	1.16E-18	0.23
32	<i>LPIN1</i>	lipin 1	23175	7.90E-17	0.28
33	<i>SH3BGR2</i>	SH3 domain binding glutamic acid-rich protein like 2	83699	1.76E-19	0.30
34	<i>CGNLI</i>	cingulin-like 1	84952	4.74E-17	0.30

Table 2B: Summary of under-expressed genes ($P < E-15$)*,†

No.	Symbol	Genes name and related function	Locus ID	P-value	Fold-change
35	<i>TP53/NP2</i>	tumor protein p53 inducible nuclear protein 2	58476	1.47E-16	0.37
36	<i>KIAA0232</i>	KIAA0232 gene product	9778	3.54E-18	0.42
37	<i>UACA</i>	uveal autoantigen with coiled-coil domains	55075	8.51E-16	0.48
Unknown					
38	<i>C10orf116</i>	chromosome 10 open reading frame 116	10974	1.98E-16	0.27
39	<i>FAM46B</i>	family with sequence similarity 46, member B	115572	1.66E-17	0.28
40	<i>C21orf81</i>	chromosome 21 open reading frame 81	114035	2.46E-16	0.37
41	<i>RMND5B</i>	required for meiotic nuclear division 5 homolog B	64777	4.82E-16	0.40
42	<i>C15orf52</i>	chromosome 15 open reading frame 52	388115	6.21E-17	0.48
43	<i>COBL</i>	cordon-bleu homolog (mouse)	23242	6.89E-16	0.50

* Genes ordered by magnitude of fold change within each sub-category

† For genes with more than one probeset, only the most significant probeset in the gene is shown.

Table 3
Comparison of 41 dysregulated genes from previous 8K cDNA microarray with 2 Affymetrix microarrays

		Fold change by microarray type ^{*,†}		
No.	Gene	8K cDNA array (N=19)	Affymetrix U133A/B array (N=53)	Affymetrix U133A v2.0 array (N=17)
Over-expressed genes				
1	<i>COL3A1</i>	2.91	4.48	6.96
2	<i>COL7A1</i>	2.31	2.65	2.36
3	<i>KRT14</i>	2.31	2.50	2.60
4	<i>FSCN1</i>	2.22	3.06	2.96
5	<i>SPARC</i>	2.08	3.45	4.40
6	<i>LAMC2</i>	2.06	3.28	5.67
7	<i>TAGLN2</i>	2.06	1.54	1.30
8	<i>FADD</i>	2.04	2.30	5.44
9	<i>CST1</i>	2.04	3.47	7.40
10	<i>HLA-B</i>	2.03	1.43	1.56
11	<i>CXCL10</i>	2.03	1.88	6.69
12	<i>CDC25B</i>	2.01	2.22	3.49
13	<i>COL1A2</i>	2.01	6.49	11.82
Under-expressed genes				
14	<i>KRT4</i>	0.16	0.17	0.13
15	<i>TGM3</i>	0.21	0.10	0.07
16	<i>CSTA</i>	0.26	0.53	0.53
17	<i>CRCT1</i>	0.27	0.11	0.06
18	<i>SPRR1A</i>	0.28	0.38	0.35
19	<i>FOSL2</i>	0.29	NA	0.64
20	<i>UPK1A</i>	0.29	0.32	0.17
21	<i>EMP1</i>	0.31	0.24	0.15
22	<i>SPINK5</i>	0.31	0.14	0.08
23	<i>CSTB</i>	0.34	0.28	0.39
24	<i>C10orf116</i>	0.34	0.27	NA
25	<i>PRR4</i>	0.35	0.42	0.92
26	<i>SLURP1</i>	0.36	0.13	0.06
27	<i>KLK13</i>	0.38	0.19	0.21
28	<i>S100A9</i>	0.38	0.45	0.35
29	<i>CNN3</i>	0.38	0.47	0.29
30	<i>BTC</i>	0.38	0.56	0.80
31	<i>HEMGN</i>	0.40	0.97	NA
32	<i>PPL</i>	0.41	0.24	0.17
33	<i>EGR1</i>	0.43	0.56	1.19

No.	Gene	Fold change by microarray type ^{*,†}		
		8K cDNA array (N=19)	Affymetrix U133A/B array (N=53)	Affymetrix U133A v2.0 array (N=17)
34	<i>APC2</i>	0.44	0.92	0.99
35	<i>KLK11</i>	0.45	0.33	0.24
36	<i>HPGD</i>	0.45	0.16	0.19
37	<i>CD3EAP</i>	0.47	1.33	1.61
38	<i>DUSP5</i>	0.47	0.27	0.21
39	<i>EVPL</i>	0.48	0.37	0.21
40	<i>RARB</i>	0.48	0.86	NA
41	<i>CD48</i>	0.50	0.91	0.78

* NA = not available

[†] Gene order based on 8K cDNA array values

Table 4
Summary of 7 genes validated with quantitative RT-PCR and comparison with microarray results

RNA expression analysis method							

* NA = not available