The discovery and validation of biomarkers for the diagnosis of esophageal squamous dysplasia and squamous cell carcinoma

George Couch[1*], James E Redman[1*], Lorenz Wernisch[2], Richard Newton[2], Shalini Malhotra[3], Sanford M Dawsey[4], Pierre Lao-Sirieix[1], Rebecca C Fitzgerald[1]

[1] MRC Cancer Unit, Hutchison/MRC research centre, University of Cambridge, Cambridge, UK

[2] MRC Biostatistics Unit, Robinson Way, Cambridge, UK

[3] Dept. Histopathology, Cambridge University Hospitals NHS Foundation Trust , Cambridge, UK

[4] Division of Cancer Epidemiology & Genetics, National Cancer Institute, Bethesda, USA

*These authors contributed equally to the work

Running title: Biomarkers for the diagnosis of esophageal squamous cell carcinoma

Keywords: screening, esophageal squamous cell, cancer, carcinoma, biomarker

Financial support:

The Addenbrooke's Hospital Human Research Tissue Bank, supported by the NIHR Cambridge Biomedical Research Centre, supported this study. R.C. Fitzgerald received funding from the NIHR and the Evelyn Trust. This study was also supported in part by the intramural research program of the National Cancer Institute. R.C. Fitzgerald has programmatic funding from the Medical Research Council and infrastructure support from the Biomedical Research Centre and the Experimental Medicine Centre.

Correspondence to:

Professor Rebecca Fitzgerald FMedSci

MRC Cancer Unit, Hutchison-MRC Research Centre, University of Cambridge

Hills Road, Cambridge, CB2 0XZ

Tel: 01223 763292

Email: rcf29@MRC-CU.cam.ac.uk

Conflict of Interest:

None to declare

Word count: ~3000; **Figure count:** 5

Abstract

Background: The 5-year survival rate of esophageal cancer is less than 10% in developing countries, where more than 90% of these cancers are squamous cell carcinomas (ESCC). Endoscopic screening is undertaken in high incidence areas. Biomarker analysis could reduce the subjectivity associated with histological assessment of dysplasia and thus improve diagnostic accuracy. The aims of this study were therefore to identify biomarkers for esophageal squamous dysplasia and carcinoma.

Methods: A publically available dataset was used to identify genes with differential expression in ESCC compared with normal esophagus (NE). Each gene was ranked by a support vector machine separation score. Expression profiles were examined, before validation by qPCR and immunohistochemistry.

Results: 800 genes were overexpressed in ESCC compared to NE ($p<10^{-5}$). Of the top 50 genes, 33 were expressed in ESCC epithelium and not in NE epithelium or stroma using the Protein Atlas website. These were taken to qPCR validation and 20 genes were significantly overexpressed in ESCC compared to NE ($p<0.05$). TNFAIP3 and CHN1 showed differential expression with immunohistochemistry. TNFAIP3 expression increased gradually through NE, mild, moderate and severe dysplasia, and SCC ($p<0.0001$). CHN1 staining was rarely present in the top third of NE epithelium and extended progressively towards the surface in mild, moderate, and severe dysplasia, and SCC ($p<0.0001$).

Conclusions: Two novel promising biomarkers for ESCC were identified, TNFAIP3 and CHN1.

Impact: CHN1 and TNFAIP3 may improve diagnostic accuracy of screening methods for ESCC.

Introduction

Cancer of the esophagus is the 6th most common cause of cancer death in the world [1]. While the squamous cell carcinoma (ESCC) subtype has been declining to around 30% of all esophageal cancers in the Western world, it remains the most common subtype in developing countries and can represent up to 90% of cases in the highest risk areas of Iran and China [1]. Patients with ESCC usually present late, with locally advanced disease or metastases, resulting in a 5-year survival rate in the USA of 18% [2]. However, survival can be as low as 10% in high risk populations, where the medical infrastructure is less well developed [3]. Early detection and treatment are associated with improved survival [4, 5]. Rapid advances in imaging [6], minimally invasive endoscopic therapies [7-9], and novel chemoradiotherapy regimes [10] provide the opportunity to improve patient outcomes when disease is diagnosed early.

In high incidence areas for ESCC, screening for cancer and dysplasia using Lugol's iodine chromoendoscopy is in use [11, 12] and has been demonstrated to significantly reduce in ESCC mortality in a recent 10-year prospective community assignment study in China [13]. Dysplasia diagnosis is difficult with intra- and inter-observer variation. There is currently a lack of suitable adjunctive diagnostic biomarkers for ESCC to facilitate the diagnosis of dysplasia [14, 15]. Recent work has begun to identify candidate genes for differentiating ESCC pre-malignant changes [16-22], however the studies have different designs and rarely examine the same genes, making cross comparisons difficult especially given that some studies are qualitative rather than quantitative [15].

We hypothesised that we could identify protein biomarkers for squamous cell dysplasia and ESCC that would be suitable for adjunctive use to pathology diagnosis and may inform us on the molecular event leading to carcinogenesis. The aims of this study were therefore to identify candidate genes that are upregulated in ESCC and squamous dysplasia compared to normal esophageal epithelium and then to validate the putative targets at both the RNA and protein levels on samples from a cohort of patients with ESCC and healthy controls.

Methods

Microarray analysis

A publically available cDNA microarray data set [23] was used to identify gene expression profiles from 65 samples (26 ESCC and 39 normal esophageal epithelium controls, Figure 1). A total of ~9,400 unique cDNA clones were available. The normalised test:reference hybridisation signal intensity ratios were converted to $\log_2$ ratios and clear outliers were excluded (1 from normal control, 2 from ESCC). A one-dimensional support

vector machine separation (SVM) score for each gene was calculated for high expression in ESCC compared to low expression in normal controls with a soft 1-norm margin with weight C=1000. For each gene, these scores were divided by the fold change between the geometric average of low and high expression and the geometric average fold change of high expression against control.

Genes were ranked using the SVM score, with a low score reflecting: i) consistent expression in each group of samples, ii) a good separation between normal and ESCC expression levels, and iii) a satisfactory level of expression in ESCC samples.

Protein Atlas Evaluation

The expected expression profile of the 50 genes with the lowest SVM scores were assessed using the Protein Atlas website (http://www.proteinatlas.org/) to ensure their suitability for paraffin embedded tissues. Genes were excluded based on reported protein expression in the epithelium of normal oesophagus or if they were known to be solely expressed in the stroma.

Human specimens

The putative genes were validated using real-time quantitative polymerase chain reaction (RT-qPCR) in 30 samples each of: normal esophagus from patients who were endoscopically normal (NN), ESCC from the tumour (T), and normal esophagus taken from the same patients as far from the tumour site as possible (NT). The NT and T groups were a 'matched cohort', as the corresponding NT and T samples were paired from each patient.

The protein expression of putative biomarkers validated by RT-qPCR was confirmed by immunohistochemistry on paraffin-embedded sections from the NT and T samples from the matched cohort, and on 34 paraffin-embedded biopsies of normal esophagus (NE), 31 mild dysplasia (Mild), 31 moderate dysplasia (Mod), and 31 severe dysplasia (Sev) samples from a 'dysplastic cohort'.

In this study, the NN samples were collected from patients attending endoscopy at Addenbrooke's Hospital, Cambridge UK for routine diagnostic procedures with endoscopically-normal esophagus; The NT and T samples from esophagectomy specimens used in a previous study in Linxian, China [24], and the biopsies of the "dysplastic cohort" from another previous study in Linxian, China [25]. The sample fixation and processing was all performed according to local, clinical standard operating procedures. All of the original studies, and the use of collected specimens for future evaluations, were approved by the appropriate IRBs.

RNA extraction and RT-qPCR

Total RNA was extracted from frozen samples using an AllPrep DNA/RNA Mini kit (QIAGEN Ltd, Manchester, UK) was then was reverse transcribed using the QuantiTect Reverse Transcription kit (QIAGEN Ltd, Manchester, UK). RT-qPCR was performed using the LightCycler 480 SYBR Green I Master mix according to manufacturer's instructions (Roche Diagnostics GmbH, Mannheim, Germany). PCR consisted of 45 cycles of 95°C denaturation (10s), 60°C annealing and extension (20s). Positive controls were identified for each primer pair. The cycle threshold (Ct) was determined for each sample and the average Ct of the triplicate samples was calculated. The expression of each gene relative to the geometric mean of the triplicate average Ct values for β-actin and 40S ribosomal protein S18 (RPS18) was determined as ΔCt. A melt curve was constructed for each primer.

Immunohistochemistry

Sections of 3.5µm each were stained using a Bond Max autostainer with the Bond Polymer Refine™ detection kit according to manufacturer's instructions (Leica Microsystems, Milton Keynes, UK). Origin of the primary antibodies and staining conditions are detailed in Table S1. A negative control was performed by omission of the primary antibody.

The extent of staining on each slide was double scored. For all genes except CHN1, extent was scored based on percentage of stained epithelium: 0 if absent, 1 for one cell to 33%, 2 for 34-66%, and 3 for $\geq$67%. For CHN1, extent was scored based on staining from the basal membrane to the epithelial surface: 0 if absent, 1 for any staining in the basal third of the epithelium, 2 for staining in the basal two thirds of the epithelium, and 3 for staining in the superficial third of the epithelium. Intensity was scored as 0 if absent, 1 for weak, 2 for medium, and 3 for strong staining.

Statistical analysis

A one-way ANOVA analysis with Dunn's multiple comparisons test was performed to analyse difference in mRNA expression. A Kruskal-Wallis one-way analysis of variance by ranks was performed to analyse difference in IHC scoring between all sample groups. A Wilcoxon matched-pairs signed-rank test was performed to analyse difference in IHC scoring between NT and T samples from the matched cohort. A one-way ANOVA analysis with Dunn's multiple comparisons test was used to compare IHC scoring between sample groups. All statistics were performed using Prism (GraphPad Software).

Results

Identification of putative targets

The SVM score analysis of a publically available cDNA microarray data set (examples shown in figure S1) yielded 800 genes which were overexpressed in ESCC compared to normal esophagus ($p < 10^{-5}$, adjusted for

multiple comparison). Expected expression profiles were evaluated using the Protein Atlas website for the 50 most significant genes: 9 genes were found to be expressed in normal esophagus (BAP1, SERPINH1, GUCY1A3, LAMC2, MMP2, SQSTM1, MMP10, NT5E and SOCSE) and 8 were expressed only in the stroma and would therefore not be suitable for a biopsy or cytology screening test (POSTN, CSPG2, MFAP2, CDH11, COL4A2, PODXL, SPARC and D2S448), (figure 1). These 17 genes were excluded from validation but the 8 genes expressed solely in the stroma may be of importance in disease progression. A total of 33 genes were taken to mRNA validation.

mRNA validation

Altered expression in ESCC compared to normal esophagus was confirmed at the mRNA level by Rt-qPCR in 30 histopathologically verified tissues each from normal esophagus (NN), normal esophagus from ESCC patients (NT), and ESCC (T). The matched cohort of NT and T samples allowed analysis for the specificity of biomarkers compared to histologically confirmed ESCC within and between cancer patients.

A suitable primer pair could not be designed for E2Ig4. Validation of this target gene was therefore not taken any further. Eight genes (LUM, THY1, PLAU, TIA-2, PTGS2_2, UCHL1, PTSG2_1, LAP1B) were not detected in either normal (NN, NT) or ESCC (T) samples. Four genes (ALCAM, LAMA3, LTBR, ERBB2) had no statistical difference in expression between groups (figure 1 and 2A). These genes were excluded from further validation.

The expression of 20 genes was significantly higher in ESCC compared to normal samples (range $p = 0.0002$ to $p < 0.0001$). However, despite the significant difference in expression, 10 of these genes (FST, ITGA6, F2R, NELL2, DUSP6, SULF1, IL1B, FRP1, PLAT and IGFBP7) displayed marked overlap in expression between sample groups. It was therefore unlikely that the difference in mRNA expression would translate to a clear difference in protein expression and these genes were therefore not taken forward to validation at the protein level (figure 1 and 2B). It is very interesting to note that for 11 out of these 20 genes, their level of expression in NT was intermediate between NN and T (Figure 1, Figure 2 and Table S2) suggesting a strong field defect around ESCC.

The remaining 10 genes contained groups of analogous genes: tumour necrosis factor alpha-induced protein 3 and 6 (TNFAIP3, TNFAIP6), collagens type III alpha 1 and type I alpha 2 (COL3A1, COL1A2), and C-C motif chemokine ligands 18 and 3 and 3-like 1 (CCL18, CCL3, CCL3L1). The presence of multiple members of the same family suggests that these are biologically relevant in the disease pathogenesis. Some homologous genes validated at the mRNA level, TNFAIP3 and TNFAIP6, COL3A1 and COL1A2 as well as CCL18, CCL3 and CCL3L1. TNFAIP3, COL3A1 and CCL18 were selected over their homologues for protein validation, as they displayed a better separation in expression between sample groups (figure 1 and 2C).

Therefore, a total of 6 genes, CCL18, COL3A1, TNFAIP3 together with CHN1, CTSL, TNC, ,all overexpressed in ESCC compared to normal samples ($p < 0.0001$) and with less than 20% overlap between these groups, were selected to be taken to protein validation (figure 1 and 3). It is interesting to note that the expression of 5 of these genes (CCL18 $p = 0.0233$, COL3A1 $p = 0.0008$, CHN1 $p = 0.0004$, CTSL $p < 0.0001$, TNC $p < 0.0001$) in normal esophagus from normal patients was lower than in normal esophagus from cancer patients. Again, this suggests some degree of field defect in pathologically normal epithelium adjacent to the cancer.

Protein validation

Increased expression in ESCC compared to normal esophagus was validated at the protein level by immunohistochemistry in histopathologically verified sections from the 'matched' and 'dysplastic' cohorts of samples, containing normal esophagus from ESCC patients (NT), ESCC (T), normal esophagus (NE), and mild, moderate and severe dysplasia. CCL18 was not statistically overexpressed in ESCC compared to matched normal (figure 1 and figure S2). While TNC was statistically upregulated in ESCC compared to normal esophagus ($p<0.0002$), TNC was not expressed at all in nearly 38% of cases. Furthermore, when it was expressed, its expression was limited to small foci of tumour cells in 45% of samples (figure 1, figure S2 and S3). Therefore both CCL18 and TNC were excluded from further IHC (figure 1, figure S2 and S3). COL3A1 (figure 4A) showed a significant difference ($p = 0.0001$) in staining across the normal, dysplastic and cancer groups, Dunn's multiple comparisons test demonstrated that COL3A1 was overexpressed in all groups compared to normal esophagus but no statistical differences were seen with increased severity of dysplasia. COL3A1 was expressed mainly in the stromal compartment with very limited epithelial staining (figure S3) and was therefore unlikely to be a suitable diagnostic biomarker. CTSL was expressed in both stroma and epithelium in normal and cancer samples (figures 4B) but no significant difference ($p = 0.3586$) was seen across the normal, dysplastic and cancer groups. In contrast, staining of CHN1 (figures 4C and 5) showed a progressive extent of staining towards the superficial layers of the esophageal epithelium, with staining in the superficial third of the epithelium (i.e. a score of 3) seen in 25% of normal esophagus, 34% of mild dysplasia, 65% of moderate dysplasia, 84% of severe dysplasia, and 97% of ESCC samples ($p < 0.0001$) with a statistically higher expression in moderate dysplasia, severe dysplasia and ESCC compared to NE ($p<0.0001$ for each comparison) as well as for severe dysplasia and ESCC compared to mild dysplasia ($p<0.0001$ for each comparison). TNFAIP3 (figure 4D and 5) also demonstrated stronger staining with increasing dysplasia ($p < 0.0001$), with staining scores of ≥1 (i.e. at least 1 cell staining positively) in 16% of normal esophagus, 18% of mild dysplasia, 30% of moderate dysplasia, 55% of severe dysplasia, and 63% of ESCC samples with expression being significantly higher in ESCC compared to normal esophagus, mild and moderate dysplasia ($p<0.001$, $p<0.01$ and $p<0.05$ respectively).

Discussion

We have demonstrated using microarray data analysis followed by subsequent validation at the mRNA and protein level that CHN1 and TNFAIP3 are candidate biomarkers for ESCC to aid in the diagnosis of dysplasia and carcinoma. Furthermore, a number of genes which may play a role in the progression to ESCC were also identified andthe functional role of these genes would be interesting to explore in the future.

While assessment of mRNA expression by RT-qPCR is notably cheaper and less labour intensive than immunohistochemistry (IHC) thus allowing parallel throughput of multiple prospective biomarkers, RNA species are relatively unstable, and therefore although some progress has been made using such approaches to understand the pathogenesis of squamous cell carcinoma (for example,[26] https://www.jstage.jst.go.jp/article/tjem/226/4/226_4_301/_pdf) they rarely survive the paraffin embedding process. We have therefore aimed at identifying protein biomarkers that could be of clinical use using a standard technique such as IHC, [27] in keeping with other protein biomarker approaches (http://onlinelibrary.wiley.com/doi/10.1002/prca.201500079/pdf). However, we do acknowledge that one drawback of immunohistochemistry is that it does rely on subjective interpretation.

Comment [RF1]: Pierre please add in this ref to endnote/ref list

Out of the 50 markers selected for validation, only two were validated at the protein level using very stringent criteria. This may appear like a low validation rate especially given that 20 out of 33 genes validated at the mRNA level. This is an example of the difficulties faced when trying to identify biomarkers for a particular cancer. While significant differences can be seen in mRNA expression, they do not necessarily translate to levels of protein expression. Furthermore, it is possible that the changes in mRNA expression only equate to protein level changes that are too subtle to detect by IHC. While the expression level or staining extent of TNFAIP3 or CHN1 respectively increased along the progression from normal esophagus to SCC, neither marker is perfect at defining the dysplastic or cancer states. Combining both markers might however offer a specific and sensitive test for esophageal dysplasia and early squamous cell cancer of the esophagus.

There were some limitations to the microarray experiments conducted, however these did not detract from the results obtained. The microarray experiments were not designed specifically to identify markers distinguishing between ESCC and normal esophagus, but rigorous statistical measures were employed to reduce the effect of this shortfall which also reduced the number of putative genes. It is interesting to note that only 20 of the 33 targets were validated by qPCR. Eight of the excluded genes were due to undetectable expression levels in biopsy samples. This is most likely due to the amplification of the RNA extracted from samples in the microarray protocol [23] which could account for the observed difference in base expression values.

The expression level of 13 out of the 20 genes increased gradually between normal oesophagus from normal patients, normal oesophagus from cancer patients and cancer samples (figures 2 and 3). Since the pathology of all samples was confirmed by an expert pathologist, it is unlikely that dysplasia or cancer was present in the normal samples from cancer patients. This intermediate level of SCC biomarkers suggests that a field defect exists in SCC patients. This field defect could be utilised diagnostically. Even in the event of biopsy that misses the area of cancer and/or dysplasia, an abnormal biomarker could still be detected and patients could be recalled for further investigation.

The biological reason for alterations in the expression of CHN1 and TNFAIP3 is worthy of further study. Chimerin 1 (CHN1) is expressed in neurones and is predominantly found in the cerebral cortex. CHN1 is a Rho GTPase-activating protein who plays a role in dendritic morphology [28] and axon guidance [29]. Missense mutations in *CHN1* have been associated with variants of Duane's retraction syndrome [30, 31] and cranial nerve abnormalities [32]. CHN1 may therefore play a role in cellular remodelling in dysplastic and cancer cells. Tumour necrosis factor, alpha-induced protein 3 (TNFAIP3) is a ubiquitin editing enzyme which inhibits NFκB and TNF-mediated apoptosis. It is associated with many autoimmune conditions [33-36] and has been noted to have tumour suppressor functions in lymphomas and colorectal cancer [37-40]. However, TNFAIP3 also has oncogenic properties, with implication in tamoxifen resistance in breast cancer and developing resistance to apoptosis to promote cancer cell survival [41]. It would be interesting to understand the role of TNFAIP3 in squamous cell cancer and its link with possible resistance to chemotherapy.

These biomarkers could also have applicability to non-endoscopic cytological screening methods since on a population wide scale, endoscopy based methods are not logistically or economically feasible due to their high cost and requirement for expertise [12, 15, 42, 43]. Non-endoscopic cell-sampling techniques are less invasive and costly, though the sensitivity and specificity of cytological assessment have been disappointing [12, 25, 44, 45]. Coupling a pan-esophageal non-endoscopic cell-collection device with analysis of biomarkers could improve diagnostic accuracy. Equally, biomarker analysis of endoscopic specimens could reduce both the requirement for histopathological expertise and the risk of sampling bias because of the molecular field defect, thus potentially reducing both the procedure length and the number of samples required. Hence the biomarker assisted analysis could reduce the cost of endoscopic diagnosis to a level where it could be considered for screening high risk populations [15, 46].

This work is complimentary to other work to identify biomarkers for the diagnosis of ESCC using a variety of techniques including methylation, array CGH, expression arrays and proteomics [26, 27, 47, 48]. (http://clincancerres.aacrjournals.org/content/19/21/5867.long; http://www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0103162&representation=PDF; https://www.jstage.jst.go.jp/article/tjem/226/4/226_4_301/_pdf;

http://onlinelibrary.wiley.com/doi/10.1002/prca.201500079/pdf). Most of these have focussed on invasive cancer whereas our focus is on the detection of dysplasia with a clinically applicable method.

Comment [RF2]: Pierre please add in these refs.

In summary, the biomarker discovery/validation pipeline successfully identified markers for esophageal squamous dysplasia and squamous cell carcinoma of the esophagus. A clinical study assessing the value of CHN1 and TNFAIP3 as diagnostic biomarkers in high incidence areas for ESCC would be the next step. In the context of ESCC screening, it is envisaged that these biomarkers could help the identification of patients with moderate or severe dysplasia that would benefit most from endoscopic treatments to prevent the development of invasive squamous cell cancers.

Figure legend:

Figure 1: Overview of discovery-validation pipeline and results at each stage. Samples used at each stage are displayed on the left, with the sample cohort shown in brackets. NN = Normal esophagus from endoscopically normal patients (normal cohort); T = Tumour, NT = Normal Esophagus from ESCC patients (matched cohort); NE = Normal Esophagus, Mild = mild dysplasia, Mod = moderate dysplasia, Sev = severe dysplasia (dysplastic cohort).

Figure 2: Relative expression of putative biomarker in normal esophagus (NN), normal esophagus from ESCC patients (NT) and ESCC (T) samples. The mean and 95% confidence intervals are displayed. **A:** ALCAM, LAMA3, LTBR and ERBB2 showed no statistical difference in expression between sample groups. **B:** TNFAIP6, COL1A2, CCL3 and CCL3L1 showed less clear separation in expression between sample groups compared to their analogous gene counterparts (TNFAIP3, COL3A1, CCL18). **C:** FST, ITGA6, F2R, NELL2, DUSP6, SULF1, IL1B, FRP1, PLAT and IGFBP7 showed significant overlap in expression between sample groups.

Figure 3: Relative expression of putative biomarker in normal esophagus (NN), normal esophagus from ESCC patients (NT) and ESCC (T) samples. CCL18, COL3A1, TNFAIP3, CHN1, CTSL and TNC showed significant difference in expression between sample groups relative to housekeepers β-actin and RPS18. The mean and 95% confidence intervals are displayed. Adjusted p values are as follows: **** $p<0.0001$, *** $p=0.0003$, ** $p=0.0025$, * $p=0.01$, NS $p>0.05$.

Figure 4: IHC staining extent scoring for putative genes across matched and dysplastic cohorts of samples. Extent score percentage for each gene (**A:** COL3A1, **B:** CTSL, **C:** CHN1, and **D:** TNFAIP3) has been plotted against each sample group's pathology. The number of samples in each pathology group is noted above each column. The 'normal' group is comprised of NT samples from the matched cohort and NE samples from the dysplastic cohort.

Figure 5: Representative images of immunohistochemistry staining for normal esophagus from ESCC patients (NT), mild, moderate, and severe dysplastic, and ESCC samples for genes CHN1 and TNFAIP3 at 100x and 400x magnification.

Supplementary information:

Table S1: Antibodies used and the conditions for immunohistochemistry.

Table S2: Summary of known the functions and roles in cancer of the 20 genes overexpressed in tumour compared to normal tissue from normal patients. The p value indicates the level of significance for the difference in expression between normal from normal patients (NN) and normal from normal tissue from patients with esophageal cancer.

Figure S1: Representative examples of SVM separation score data selected from the 100 genes with the highest SVM scores using a random number generator. The x-axis shows the sample number; normal samples are on the left of each graph, tumour samples on the right separated by a red bar. The horizontal lines represent the means of the expression scores for the normal and tumour samples respectively.

Figure S2: Cumulative IHC scoring for normal esophagus from ESCC patients (NT) and ESCC (T) matched samples. CCL18 showed no statistical difference in staining between sample groups but TNC was significantly over-expressed in tumour samples ($p<0.0002$).

Figure S3: Representative images of immunohistochemistry staining for normal esophagus from ESCC patients (NT) and ESCC samples (T) for genes CTSL, TNC and COL3A1 at 100x magnification.

References:

1. Torre, L.A., et al., *Global cancer statistics, 2012.* CA Cancer J Clin, 2015. **65**(2): p. 87-108.
2. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2015.* CA Cancer J Clin, 2015. **65**(1): p. 5-29.
3. Aghcheli, K., et al., *Prognostic factors for esophageal squamous cell carcinoma--a population-based study in Golestan Province, Iran, a high incidence area.* PLoS One, 2011. **6**(7): p. e22152.
4. Chen, L.Q., et al., *Early detection of esophageal squamous cell carcinoma and its effects on therapy: an overview.* Dis Esophagus, 1999. **12**(3): p. 161-7.
5. Wang, G.Q., et al., *Long-term results of operation for 420 patients with early squamous cell esophageal carcinoma discovered by screening.* Ann Thorac Surg, 2004. **77**(5): p. 1740-4.
6. Reddymasu, S.C. and P. Sharma, *Advances in Endoscopic Imaging of the Esophagus.* Gastroenterology Clinics of North America, 2008. **37**(4): p. 763-+.
7. Inoue, H., et al., *Endoscopic mucosal resection and endoscopic submucosal dissection for esophageal dysplasia and carcinoma.* Gastrointest Endosc Clin N Am, 2010. **20**(1): p. 25-34, v-vi.
8. Bergman, J.J., et al., *Outcomes from a prospective trial of endoscopic radiofrequency ablation of early squamous cell neoplasia of the esophagus.* Gastrointest Endosc, 2011. **74**(6): p. 1181-90.
9. Haidry, R.J., et al., *Radiofrequency ablation for early oesophageal squamous neoplasia: outcomes form United Kingdom registry.* World J Gastroenterol, 2013. **19**(36): p. 6011-9.
10. Bass, G.A., et al., *Chemoradiotherapy, with adjuvant surgery for local control, confers a durable survival advantage in adenocarcinoma and squamous cell carcinoma of the oesophagus.* Eur J Cancer, 2014. **50**(6): p. 1065-75.
11. Dawsey, S.M., et al., *Mucosal iodine staining improves endoscopic visualization of squamous dysplasia and squamous cell carcinoma of the esophagus in Linxian, China.* Cancer, 1998. **83**(2): p. 220-231.
12. Lao-Sirieix, P. and R.C. Fitzgerald, *Screening for oesophageal cancer.* Nat Rev Clin Oncol, 2012. **9**(5): p. 278-87.
13. Wei, W.Q., et al., *Long-Term Follow-Up of a Community Assignment, One-Time Endoscopic Screening Study of Esophageal Cancer in China.* J Clin Oncol, 2015. **33**(17): p. 1951-7.
14. Sawyers, C.L., *The cancer biomarker problem.* Nature, 2008. **452**(7187): p. 548-52.
15. Taylor, P.R., C.C. Abnet, and S.M. Dawsey, *Squamous dysplasia--the precursor lesion for esophageal squamous cell carcinoma.* Cancer Epidemiol Biomarkers Prev, 2013. **22**(4): p. 540-52.
16. Takikita, M., et al., *Fascin and CK4 as Biomarkers for Esophageal Squamous Cell Carcinoma.* Anticancer Research, 2011. **31**(3): p. 945-952.
17. Xue, L.Y., et al., *Tissue microarray analysis reveals a tight correlation between protein expression pattern and progression of esophageal squamous cell carcinoma.* BMC Cancer, 2006. **6**: p. 296.
18. Kashyap, M.K., et al., *Genomewide mRNA profiling of esophageal squamous cell carcinoma for identification of cancer biomarkers.* Cancer Biol Ther, 2009. **8**(1): p. 36-46.
19. Zhou, J.H., et al., *Autoantibodies against MMP-7 as a novel diagnostic biomarker in esophageal squamous cell carcinoma.* World Journal of Gastroenterology, 2011. **17**(10): p. 1373-1378.
20. Su, H., et al., *Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes.* Clin Cancer Res, 2011. **17**(9): p. 2955-66.
21. Wang, W.C., et al., *Ki-67 and ProExC are useful immunohistochemical markers in esophageal squamous intraepithelial neoplasia.* Hum Pathol, 2011. **42**(10): p. 1430-7.
22. Takala, H., et al., *Toll-Like Receptor 9 Is a Novel Biomarker for Esophageal Squamous Cell Dysplasia and Squamous Cell Carcinoma Progression.* Journal of Innate Immunity, 2011. **3**(6): p. 631-638.
23. Greenawalt, D.M., et al., *Gene expression profiling of esophageal cancer: Comparative analysis of Barrett's esophagus, adenocarcinoma, and squamous cell carcinoma.* International Journal of Cancer, 2007. **120**(9): p. 1914-1921.
24. Koshiol, J. and A.R. Kreimer, *Lessons from Australia: Human Papillomavirus Is Not a Major Risk Factor for Esophageal Squamous Cell Carcinoma.* Cancer Epidemiology Biomarkers & Prevention, 2010. **19**(8): p. 1889-1892.

25. Pan, Q.J., et al., *Cytologic detection of esophageal squamous cell carcinoma and its precursor lesions using balloon samplers and liquid-based cytology in asymptomatic adults in Llinxian, China.* Acta Cytol, 2008. **52**(1): p. 14-23.
26. Tao, Y., et al., *Identification of distinct gene expression profiles between esophageal squamous cell carcinoma and adjacent normal epithelial tissues.* Tohoku J Exp Med, 2012. **226**(4): p. 301-11.
27. Coghlin, C. and G.I. Murray, *Progress in the development of protein biomarkers of oesophageal and gastric cancers.* Proteomics Clin Appl, 2015.
28. Buttery, P., et al., *The diacylglycerol-binding protein alpha1-chimaerin regulates dendritic morphology.* Proc Natl Acad Sci U S A, 2006. **103**(6): p. 1924-9.
29. Clark, C., et al., *alpha2-Chimaerin regulates a key axon guidance transition during development of the oculomotor projection.* J Neurosci, 2013. **33**(42): p. 16540-51.
30. Miyake, N., et al., *Expansion of the CHN1 strabismus phenotype.* Invest Ophthalmol Vis Sci, 2011. **52**(9): p. 6321-8.
31. Miyake, N., et al., *Human CHN1 mutations hyperactivate alpha2-chimaerin and cause Duane's retraction syndrome.* Science, 2008. **321**(5890): p. 839-43.
32. Hwang, J.M., et al., *Absence of CHN1 in two patients with a bilateral absence of cranial nerves IV and VI.* Graefes Arch Clin Exp Ophthalmol, 2014.
33. Lee, Y.H. and G.G. Song, *Associations between TNFAIP3 gene polymorphisms and systemic lupus erythematosus: a meta-analysis.* Genet Test Mol Biomarkers, 2012. **16**(9): p. 1105-10.
34. Lee, Y.H., et al., *Associations between TNFAIP3 gene polymorphisms and rheumatoid arthritis: a meta-analysis.* Inflamm Res, 2012. **61**(6): p. 635-41.
35. Nititham, J., et al., *Meta-analysis of the TNFAIP3 region in psoriasis reveals a risk haplotype that is distinct from other autoimmune diseases.* Genes Immun, 2014.
36. Song, R.H., et al., *Polymorphisms of the TNFAIP3 region and Graves' disease.* Autoimmunity, 2014. **47**(7): p. 459-65.
37. Honma, K., et al., *TNFAIP3/A20 functions as a novel tumor suppressor gene in several subtypes of non-Hodgkin lymphomas.* Blood, 2009. **114**(12): p. 2467-75.
38. Novak, U., et al., *The NF-{kappa}B negative regulator TNFAIP3 (A20) is inactivated by somatic mutations and genomic deletions in marginal zone lymphomas.* Blood, 2009. **113**(20): p. 4918-21.
39. Schmitz, R., et al., *TNFAIP3 (A20) is a tumor suppressor gene in Hodgkin lymphoma and primary mediastinal B cell lymphoma.* J Exp Med, 2009. **206**(5): p. 981-9.
40. Ungerback, J., et al., *Genetic variation and alterations of genes involved in NFkappaB/TNFAIP3- and NLRP3-inflammasome signaling affect susceptibility and outcome of colorectal cancer.* Carcinogenesis, 2012. **33**(11): p. 2126-34.
41. Vendrell, J.A., et al., *A20/TNFAIP3, a new estrogen-regulated gene that confers tamoxifen resistance in breast cancer cells.* Oncogene, 2007. **26**(32): p. 4656-67.
42. Gerson, L.B., P.W. Groeneveld, and G. Triadafilopoulos, *Cost-Effectiveness Model of Endoscopic Screening and Surveillance in Patients With Gastroesophageal Reflux Disease.* Clinical Gastroenterology and Hepatology, 2004. **2**(10): p. 868-879.
43. Roshandel, G., et al., *Endoscopic screening for precancerous lesions of the esophagus in a high risk area in Northern Iran.* Arch Iran Med, 2014. **17**(4): p. 246-52.
44. Boller, D., et al., *Lugol chromoendoscopy combined with brush cytology in patients at risk for esophageal squamous cell carcinoma.* Surg Endosc, 2009. **23**(12): p. 2748-54.
45. Roth, M.J., et al., *Cytologic detection of esophageal squamous cell carcinoma and precursor lesions using balloon and sponge samplers in asymptomatic adults in Linxian, China.* Cancer, 1997. **80**(11): p. 2047-59.
46. di Pietro, M., et al., *The combination of autofluorescence endoscopy and molecular biomarkers is a novel diagnostic tool for dysplasia in Barrett's oesophagus.* Gut, 2015. **64**(1): p. 49-56.
47. Shi, Z.Z., et al., *Consistent and differential genetic aberrations between esophageal dysplasia and squamous cell carcinoma detected by array comparative genomic hybridization.* Clin Cancer Res, 2013. **19**(21): p. 5867-78.

48. Li, X., et al., *Identification of a DNA methylome profile of esophageal squamous cell carcinoma and potential plasma epigenetic biomarkers for early diagnosis.* PLoS One, 2014. **9**(7): p. e103162.