ORIGINAL ARTICLE

NRF2 metagene signature is a novel prognostic biomarker in colorectal cancer

Séan M. O'Cathail¹,², *, Chieh-Hsi Wu³,¹, Annabelle Lewis⁵, Chris Holmes⁴,⁵,⁶, Maria A Hawkins⁷, Tim Maughan⁸

¹ Institute of Cancer Sciences, University of Glasgow, United Kingdom; ² Department of Statistics, University of Southampton, United Kingdom; ³ Department of Life science, CHMLs, Brunel University London, United Kingdom; ⁴ Nuffield Department of Medicine, University of Oxford, United Kingdom; ⁵ Alan Turing Institute, London, United Kingdom; ⁶ University College London, United Kingdom; ⁷ Oxford Institute of Radiation Oncology, University of Oxford, United Kingdom

Abstract

We hypothesise that the NRF2 transcription factor would act a biomarker of poor prognosis in colorectal cancer. We derived and validated an mRNA based metagene signature of NRF2 signalling and validated it in 1360 patients from 4 different datasets as an independent biomarker of poor prognosis. This is a novel insight into the molecular signalling of colorectal cancer.

Background: NRF2 over activity confers poor prognosis in some cancers but its prognostic role in colorectal cancer (CRC) is unknown. As a transcription factor, we hypothesise a signature of NRF2 regulated genes could act as a prognostic biomarker in CRC and reveal novel biological insights.

Methods: Using known NRF2 regulated genes, differentially expressed in CRC, we defined a signature of NRF2 pathway activity using principal component analysis and Cox proportional hazard models and tested it in four independent mRNA datasets, profiled on three different mRNA platforms.

Results: 36 genes comprised the final NRF2 signature. 1360 patients were included in the validation. High NRF2 was associated with worse disease free survival (DFS) and/or overall survival (OS) in all datasets: (GSE14333 HR=1.55, 95% CI 1.2–2.004, p = 0.0008; GSE39582 HR=1.24, 95% CI 1.086–1.416, p = 0.001; GSE87211 HR=1.431, 95% CI 1.06–1.93, p = 0.056; MRC FOCUS trial HR=1.14, 95% CI 1.04–1.26, p = 0.008). In multivariate analyses, NRF2 remained significant when adjusted for stage and adjuvant chemotherapy in stage I-III disease, and BRAF V600E mutation and sidedness in stage IV disease. NRF2 activity was particularly enriched in Consensus Molecular Subtype (CMS) 4.

Conclusion: For the first time, NRF2 is shown to be a consistent, robust prognostic biomarker across all stages of colorectal cancer with additional clinical value to current known prognostic biomarkers. High NRF2 signalling in CMS 4 further refines the molecular taxonomy of CRC, a new biological insight, suggesting avenues of further study.

Keywords NRF2, Colorectal, Signature, Biomarker, Prognosis.

© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Abbreviations: CRC, Colorectal cancer; DFS, Disease Free Survival; OS, Overall survival; CMS, Consensus Molecular Subtypes; ARE, Antioxidant response element; PCA, Principal component analysis; S:CORT, Stratification in Colorectal Cancer; LRT, Likelihood ratio test; KM, Kaplan Meier; MMR, Mismatch Repair; HR, Hazard ratio.

*Corresponding author.
E-mail address: Sean.OCathail@glasgow.ac.uk

Received June 16, 2020; received in revised form August 14, 2020; accepted August 17, 2020

Denotes equal contribution
Background

Colorectal cancer (CRC) is the 4th most common cancer in the UK with 41,700 new cases per annum and only 57% of patients will live ten years or more [1]. An improved understanding of the biology of colorectal cancer may provide the basis for stratification of patients for differing treatment programmes, first by identifying prognostic effects. Current, known prognostic factors include ‘sidedness’ (left or right side of the colon) [2]—which is assumed to be surrogate of tumour [3] and patient biology [4]—RAS mutant status, BRAF status [5] and mismatch repair status [6,7]. The most widely used, RNA expression based classifier is the Consensus Molecular Subtype (CMS) [8]. It highlights that colorectal cancer is a significantly heterogeneous disease with different prognostic expectations among four subgroups. The taxonomy applied to each of the four subtypes indicate enriched pathways, but no subtype is defined by individual events, genetic aberrations or expression pathways. Therefore, there may be additional, heretofore unknown biological pathways that contribute to the prognostic differences between the subtypes.

The KEAP1-NRF2 pathway is an canonical signalling pathway which has been implicated in all the Hallmarks of Cancer [9]. Though of prognostic importance in many tumours types, notably lung cancer [10], its significance in CRC is unknown. NRF2 is a potent transcriptional activator that plays a central role in cell protection against oxidative and electrophilic stress. NRF2 activity is tightly regulated by KEAP1. Under basal (unstressed) conditions KEAP1, part of the Cul3 ubiquitin ligase family, mediates polyubiquitination and proteasomal degradation of NRF2 protein [11]. Cellular stresses modify the structural integrity KEAP1-CUL3 ligase complex resulting in declining ubiquitination activity and an increase in cellular NRF2. Unbound NRF2 translocates to the nucleus and binds to antioxidant response element (ARE) sequences to regulate the transcription of suites of genes, including intracellular redox control, metabolic pathways, autophagy and drug transport [12]. Historically the NRF2 pathway was deemed to function in ‘tumour suppressor’ like capacity, allowing the cell to defend against stressors such as carcinogens [12]. Recent evidence shows that some tumours acquire constitutive activation of the pathway which allow it to function in an ‘oncogene’ like fashion, promoting cell survival, resisting radiation, chemotherapeutics and dysregulating metabolism [13,14].

There are a number of distinct mechanisms by which the NRF2 pathway can become constitutively activated in CRC [15]. Mutations in both KEAP1 and NFE2L2 have been described in a up to 7.8% of colorectal cancers [16], although the rate in TCGA is less than 2.4% and 0.9% respectively [17]. However, the level of NRF2 signalling in the TCGA dataset is higher than expected for the low somatic mutation rate observed [18], suggesting complex post transcriptional mechanisms of activation. Epigenetic modifications, methylation, of KEAP1 in CRC silencing its ability to regulate NRF2 [19] and direct activation of NRF2 transcription via oncogenes KRAS [20], BRAF [21] and c-MYC [22] have all been described [20].

It is unlikely that NFE2L2 (the gene encoding NRF2 protein) mutation or expression in isolation will capture the full effect of pathway activity, and prove a useful biomarker. However, as NRF2 functions as a transcription factor controlling a known suite of antioxidant response element (ARE) regulated genes, we hypothesise that a ‘signature’ of NRF2 activity, could be used to aggregate different mechanisms of pathway activity and act as a biomarker of prognosis in CRC. Here, we define a signature of NRF2 activity as ‘a metagene which is a set of known NRF2 targets with coordinated mRNA expression representing the component of NRF2 pathway potentially relevant to prognostic prediction’. We detail the derivation of an NRF2 signature from RNA expression data using a candidate gene approach [21] in colorectal cancer datasets and demonstrate, for the first time, that high NRF2 activity is a biomarker of poor prognosis across all stages of CRC.

Methods

Candidate gene selection

Known NRF2 regulated genes were selected from two published prognostic lung cancer signatures [22,23] and refined for differential expression using the Oncomine database [24]. Input parameters “Cancer Type” and “Analysis Type” were set to ‘colorectal cancer’ and ‘cancer versus normal’ respectively. Differential expression was determined by threshold values of fold change > 2, p-value < 0.0001 and gene rank of top 10%. The database normalises gene expression across all selected datasets to allow summative gene expression comparisons. The resulting median gene rank for the meta-analysis across all selected datasets was calculated with its associated p-value, which was corrected for multiple hypothesis testing using the false discovery rate method [25].

Construction of the signature of NRF2

Principal component analysis (PCA) was applied to all probes representing the candidate NRF2 target genes. This generated a new set of continuous variables, principal components (PCs), which were weighted averages of the RNA expressions across the probes considered.

Supervised variable selection was performed to decide how many major PCs1 would be useful for predicting prognosis in the training set. A Cox proportional hazard regression model was used to model the prognosis predicted by the PCs of NRF2 activity.

The NRF2 activity in each validation set was obtained by performing PCA on the corresponding probe sets (see supplementary figure 1 and supplementary information for further details).

Datasets

Publically available colorectal datasets were downloaded from the Gene Expression Omnibus (GEO) database, accessed from the R programming environment using the packages ‘GOQuery’ [26] and ‘Biobase’ [27] obtained from https://bioconductor.org/biocLite.R. All datasets are summarised in Table 1. The datasets were divided into a train-

---

1 Here, we define the X major PCs as the X most variable PCs that explains 80% of the variation in the data.
Table 1  Cohort characteristics of the training and validation sets.

<table>
<thead>
<tr>
<th></th>
<th>GSE17536 (Training set)</th>
<th>GSE14333</th>
<th>GSE39582</th>
<th>MRC FOCUS trial</th>
<th>GSE87211</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>177</td>
<td>226</td>
<td>570</td>
<td>375 (355 DNA</td>
<td>189</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mutat status)</td>
<td></td>
</tr>
<tr>
<td>Tissue type</td>
<td>Fresh frozen</td>
<td>Fresh frozen</td>
<td>Fresh frozen</td>
<td>FFPE</td>
<td>Fresh frozen</td>
</tr>
<tr>
<td>Platform array</td>
<td>Affymetrix</td>
<td>Affymetrix</td>
<td>Affymetrix U133</td>
<td>Affymetrix Xcel</td>
<td>Agilent Human</td>
</tr>
<tr>
<td>Primary site</td>
<td>Colon</td>
<td>Colorectal</td>
<td>Colon</td>
<td>Colon</td>
<td>Rectum</td>
</tr>
<tr>
<td>Stage</td>
<td>I</td>
<td>24 (13.6%)</td>
<td>41 (18%)</td>
<td>37 (6.5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>57 (32.2%)</td>
<td>95 (42%)</td>
<td>267 (47%)</td>
<td>70 (30.8%)</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>57 (32.2%)</td>
<td>93 (41%)</td>
<td>206 (36%)</td>
<td>143 (63%)</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>39 (22%)</td>
<td>60 (10.5%)</td>
<td>375 (100%)</td>
<td>14 (6.2%)</td>
</tr>
<tr>
<td>Outcome variable</td>
<td>OS</td>
<td>DFS</td>
<td>DFS OS</td>
<td>OS</td>
<td>DFS OS</td>
</tr>
<tr>
<td>Covariates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemo(radio)therapy</td>
<td>Yes 72</td>
<td>No 240</td>
<td>Yes 240</td>
<td>375 (100%)</td>
<td>189 (100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No 326</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site of primary</td>
<td>Prox. 232</td>
<td>Dist. 351</td>
<td>Left 203</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Right 152</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRAF V600E mutation</td>
<td>Mut 51</td>
<td>dMMR 77</td>
<td>dMMR 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mismatch Repair status</td>
<td>pMMR 459</td>
<td>pMMR 326</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The numbers of cases, type of tissue, RNA expression platform, outcome variable and available covariates for adjusted analyses are indicated. (DFS = Disease Free Survival, OS = Overall survival, dMMR = deficient Mismatch repair, pMMR = proficient Mismatch Repair).

ing set, GSE17536 [28], and validation sets. Validation was carried out using datasets representative of non-metastatic, stage I–III disease (GSE14333 and GSE39582) [29,30], metastatic disease (MRC FOCUS trial) [31] and rectal only cancer (GSE87211) [32]. As part of the Stratification in Colorectal cancer (S:CORT) consortium, we had access to the MRC FOCUS trial data including the RNA expression profiles generated by S:CORT (See supplementary information). GSE17536, GSE14333, GSE39582 were profiled using the Affymetrix U133 Plus 2.0 array, GSE87211 used the Agilent-026652 Whole Human Genome Microarray 4 × 44 K v2 and the MRC FOCUS trial used the Affymetrix Xcel™ array.

Statistical analysis

Primary analysis

The primary analysis for each validation set was to determine whether the NRF2 activity has a prognostic effect on disease free survival (DFS) and/or overall survival (OS) by a Cox regression model with NRF2 activity as the only covariate. The likelihood ratio test (LRT) was used to quantify evidence against that NRF2 activity provides no explanatory power. For the construction of Kaplan-Meier (K-M) curves, NRF2 activity was subdivided by tertiles. Hazard ratios, and confidence intervals, presented for these curves are between the upper and lower tertiles.

Secondary analyses

To assess whether the prognostic effect of the NRF2 activity was confounded by other known prognostic variables we performed adjusted analyses using multivariate Cox PH regression models. Because adjusting variables varied across datasets, no adjusting variables were used in the training set. A LRT was performed to quantify the evidence against NRF2 activity provides no explanatory power in addition to the adjusting variables. The adjusting variables used for the secondary analyses are summarised in Table 1. All statistical analyses were conducted using R [33].

Results

NRF2 signature derivation and training

In total 62 candidate genes were analysed in 9 independent colorectal datasets for differential expression relative to normal tissue [17,34–39]. Some datasets were subsetted into different anatomical sites for the purposes of analysis resulting in 24 discrete sets of data (see supplementary Table 1). 40 were found to be differentially expressed in tumours, 21 of which were significantly over-expressed and 20 which were significantly under-expressed, in at least one or more of the datasets (supplementary Figure 1). One gene, COL3A1, was shared as it was over expressed in some datasets and under expressed in others. Of the 40 differentially expressed genes, four could not be matched between the training and validation dataset microarrays so were omitted from further analysis. The final group of 36 genes was: ABCA8, ABI3BP, ADAM12, ADRB1, ANGPT1, ANKR2D9, ANKR2D44, BCHE, C15orf48, COL3A1, COL5A1, EGLN3, LIFR, METTL7A, PCM1, PLAU, PLCB4, RECK, RGCC, RRM2, SEC14L4, SERPINH1, SFN, SLIT3, SPP1, TNS1, TOM1L2, TSPAN5, TTYH3, V1G10, VCAN, AKR1C1, LRP8, NAMPT, PTGES, SLC27A5. There was a very high level of co-ordinated expression between the 36 genes in the training dataset as evidenced by pairwise correlations (Fig. 1A).
Variable selection

Following PCA, PC1 was indicated to be useful for explaining the survival outcome by both Akaike and Bayesian information criteria. PC1 in the training set had absolute correlations $>0.5$ with probes that mapped to the following 10 genes: VCAN, ADAM12, COL3A1, COL5A1, SERPINH1, RECK, PLAU, SPPI, TNS1 and SLIT3. Due to the high correlation of these genes with PC1 in the training set, we hypothesised that they were of higher biological relevance for prognosis prediction than other NRF2 target genes. This expression pattern was detected in each of the validation sets (Fig. 1B-E). A summary overview of the process for signature derivation and training is provided (supplementary figure 2).

NRF2 activity a biomarker of worse survival

In stage I/I/II disease, higher NRF2 activity corresponded to worse DFS in GSE14333 (HR$^2$=1.551, 95% CI 1.200–2.004, LRT $p$ = 0.0008) and GSE39582 (HR=1.172, 95% CI 1.008–1.362, LRT $p$ = 0.0383). Including the 60 cases of stage IV disease also available in GSE39582, NRF2 activity was also associated with worse OS (HR=1.240, 95% CI 1.086–1.416, LRT $p$ = 0.001). In the MRC FOCUS trial, comprised of first line stage IV metastatic patients, NRF2 activity was again associated with a worse overall survival (HR$^3$=1.140, 95% CI 1.035–1.255, LRT $p$ = 0.008). Fig. 2 shows that high activity corresponded with worse prognosis for DFS in GSE14333 and GSE39582 (panels A and B), and for OS in GSE39582 and MRC FOCUS trial (panels C and D).

In order to assess the relevance of NRF2 activity in rectal cancer specifically, and the ability to migrate between RNA expression platforms, we performed the analysis on a rectal cancer only expression dataset, where all sampled patients received neoadjuvant chemoradiotherapy (GSE87211). Higher activity was associated with worse DFS (HR=1.431, 95% CI 1.060–1.933, LRT $p$ = 0.056) but not OS (HR=1.464, 95% CI 0.955–2.245; LRT, $p$ = 0.197). Fig. 3 shows that high activity corresponded to worse prognosis for DFS.

NRF2 activation provides additional explanatory power to known prognostic variables

Within the publicly available datasets there were additional variables that are known prognostic factors. The magnitude of their respective effects are summarised in the forest plot (supplementary figure 3). We used these in a multivariate analysis. In GSE14333, after adjusting for the effect of stage and adjuvant chemotherapy, NRF2 activity remained a significant predictor of worse DFS (HR$^4$=1.365, 95% CI 1.145–1.624, LRT $p$ = 0.0003).

---

1. As NRF2 expression is a continuous variable here, the HR reported in the text of this section is the HR between the upper and lower tertiles of the NRF2 expression.

2. Refer to footnote 2.

3. See footnote 2.

4. Refer to footnote 4.
C.I. 1.049–1.776, LRT \( p = 0.02 \). Similarly in GSE39582, the effect of high NRF2 activity corresponds to worse DFS (HR=1.168, 95% C.I 1.000–1.363, LRT \( p = 0.049 \)) after adjusting for the effect of stage and mismatch repair status (MMR). In the latter dataset, NRF2 activity was also significantly associated with worse OS when adjusting for stage alone (HR=1.185, 95% C.I 1.040–1.350, LRT \( p = 0.01 \)). No adjusted analysis was carried out for MMR with NRF2 activity on OS due to the known contrasting effects MMR status has on prognosis in early stage and metastatic disease, which could lead to model misspecification.

In the MRC FOCUS trial, prognostic factors within the dataset were site of the primary tumour (sidedness) and \( \text{BRAF}^{\text{V600E}} \) mutation. Again, high NRF2 activity corresponded to worse overall survival (HR=1.123, 95% C.I 1.020–1.237, LRT \( p = 0.0185 \)). In summary, there was systematic evidence that NRF2 signalling had an effect on DFS and/or OS in all available datasets (Table 2).

Fig. 2  Kaplan-Meier curves and associated risk tables for the primary analysis of three datasets. All show survival outcomes for patients with high, intermediate and low (induced by tertiles) of NRF2 metagene expression. A) GSE14333 and B) GSE39582 represents early stage I-III patients C) GSE39582 represents stage I-IV patients and C) MRC FOCUS represents stage IV first line metastatic patients. For the Kaplan Meier curves, a median cut point was used to binarise NRF2 metagene expression.
S.M. O’Cathail, C.-H. Wu and A. Lewis et al.

Fig. 3 Kaplan Meier curves and associated risk tables for the primary analyses of GSE87211. On the left, high NRF2 metagene expression is associated with worse disease free survival with persistent separation of the curves. On the right, there was no effect on overall survival. For the Kaplan Meier curves, high, intermediate and low (tertiles) of NRF2 expression was used.

Table 2 Summary table of the Cox proportional hazard model analyses.

<table>
<thead>
<tr>
<th>Data set</th>
<th>N</th>
<th>Analysis</th>
<th>Variables</th>
<th>HR (C.I)</th>
<th>LRT (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE14333</td>
<td>226</td>
<td>Primary</td>
<td>N/A</td>
<td>1.113 (1.045–1.184)</td>
<td>0.0008</td>
</tr>
<tr>
<td>MRC FOCUS</td>
<td>375</td>
<td>Secondary</td>
<td>Stage Adjuvant chemotherapy</td>
<td>1.033 (1.009–1.245)</td>
<td>0.008</td>
</tr>
<tr>
<td>GSE39582 (OS)</td>
<td>570</td>
<td>Secondary</td>
<td>Primary</td>
<td>N/A Stage</td>
<td>1.054 (1.02–1.089)</td>
</tr>
<tr>
<td>GSE39582 (RFS)</td>
<td>510</td>
<td>Secondary</td>
<td>Primary</td>
<td>N/A Stage</td>
<td>1.04 (1.002–1.08)</td>
</tr>
<tr>
<td>GSE87211 (DFS)</td>
<td>189</td>
<td>Primary</td>
<td>N/A</td>
<td>1.067 (1.019–1.2445)</td>
<td>0.056</td>
</tr>
<tr>
<td>GSE87211 (OS)</td>
<td>189</td>
<td>Primary</td>
<td>N/A</td>
<td>1.073</td>
<td>0.197</td>
</tr>
</tbody>
</table>

It shows the numbers of patients included in each analysis, the adjusting variables where used and the p-value for the Cox model comparison (LRT = Likelihood ratio test).

* Hazard ratio for event per unit increase in expression of NRF2 signature (continuous variable) which is different to HR between the upper and lower tertiles.

** With imputation for missing variables.

**NRF2 activation and consensus molecular subtypes (CMS)**

In order to understand how NRF2 activity aligns with the current transcriptomic landscape of colorectal cancer, we examined the distribution of the three groups of NRF2 activity level across the four CMS subtypes in the MRC FOCUS trial (Fig. 4). While high NRF2 activity can be seen across all subtypes, strikingly CMS 4 showed substantially higher NRF2 activity with no patients in the category of low NRF2 activity. By contrast, the majority of patients in CMS 2 or 3 had low and intermediate NRF2 activity.
Discussion

We derived an NRF2 signature to measure activation of the pathway in colorectal cancer and independently validated it as a biomarker of poor prognosis across all stages of colorectal cancer. This is an entirely new biological insight into colorectal cancer for several reasons.

To the authors’ knowledge, this is the first demonstration of NRF2 regulated genes behaving in a co-ordinated network fashion in CRC, in vivo, and supports our ‘meta-gene’ approach to represent KEAP1/NRF2 pathway activity. The high degree of co-expression seen between key genes VCAN, ADAM12, SPP1, COL3A1, COL5A1, TNS1, SLIT3, RECK, PLAU and SERPINH1 was consistent with the predicted behaviour of these genes on the STRING database [40] (http://string-db.org/; supplementary figure 4). Having chosen the candidate genes on the basis of NRF2 regulation in lung cancer strengthens the unbiased nature of the analysis. With the exception of COL3A1, all of the key genes contain ARE within their promoter region emphasising NRF2’s ability to directly regulate transcription (supplementary information, supplementary figure 5).

Secondly, NRF2 has been shown as a biomarker of poor prognosis across all stages of colorectal cancer in several large, independent datasets comprising 1360 patients making it one of the largest validation analyses of a transcriptionic biomarker in CRC to date. To place the analysis size in context, the OS and DFS cohorts used to validate CMS as a prognostic biomarker were 2129 and 1785 cases respectively [8]. The prognostic effect was maintained when adjusting for known prognostic clinical and molecular factors including stage, adjuvant chemotherapy and mismatch repair status in the non-metastatic setting, and, BRAF V600E mutation and tumour sidedness in the relapsed setting. The 375 patients with stage IV metastatic disease have been selected from a large randomised controlled phase III trial, which makes the findings more robust against unknown sampling biases. No other CRC biomarkers other than CMS 4 are consistent across early and late disease [41]. The enrichment of CMS 4 for NRF2 activity may explain this finding.

Thirdly, the effect was consistent and robust across the four datasets in spite of the technical differences in RNA profiling (three different expression platforms from two different manufacturers), a combination of FFPE and fresh frozen tissue and biopsy sizes.

In spite of the accumulating evidence that NRF2 plays a significant role in cancer [9,42,43] there is remarkably little information on the prognostic contribution of NRF2 in colorectal cancer. High levels of NRF2 activity within resected tumours have been found to be significantly correlated with p53 expression, Duke’s stage and poor clinical outcomes [44] as well as tumour size, TNM stage and metastases [45]. Whilst the latter investigated the association between NRF2 and survival status, their analysis did not take into account of the various aspects of time-to-event data, including data censoring or the concept of time. In addition, it was impossible to directly measure the prognostic effect of NRF2 in presence of potential confounders in their analysis framework.

CMS has defined the current RNA taxonomy of CRC. The resulting classifications (CMS 1–4) and their respective prognostic outlooks describe the landscape in which any novel expression biomarker must be evaluated, especially as CMS was derived using a network-based clustering approach, agnostic of underlying biological mechanisms. We have shown that, although distributed across all four subtypes, high NRF2 activity is significantly enriched in CMS 4. This was unexpected. A priori, given that NRF2 is primarily known as a metabolic pathway and previous data demonstrating co-occurrence and co-activation with PIK3CA [10] and KRAS mutation [46], one could have expected enrichment within CMS 3. We would argue therefore that this adds further biological insight into the CMS classification and the exact role of NRF2 in interacting with the CMS 4 subtype warrants further study. For example, one of the most enriched biological processes seen in our NRF2 signature is that of extracellular matrix (ECM) signalling. Recent proteogenomic work from squamous cell lung cancer [47] noted that somatic mutations of NFE2L2/KEAP1 are enriched for transcriptional programs in ECM similar to the data presented here. Open questions include whether the cancer cell exerts an outward influence on the ECM by increasing NRF2 signalling and creating a protumorigenic environment, or is it an inward effect, with the ECM altering the biology of the cancer cell. Certainly NRF2 is increasingly recognised to have a more far reaching effect on the cancer cell than traditional oxidative stress management [43].
The difference in prognosis between those with high and low NRF2 activity could, in part, be due to therapeutic resistance. Fluorouracil is the main chemotherapy drug used in both the adjuvant and metastatic setting, and was the backbone of therapy used in the FOCUS trial. Silencing of NRF2 signalling has been shown to overcome 5-FU resistance in colorectal cancer models in both an in vitro and in vivo setting [48]. Quantifying the effect of NRF2 activity in therapeutic resistance in relation to radiation and chemotherapy in colorectal cancer is ongoing in our laboratory. However, given the pluripotent nature of NRF2 it would seem plausible that it mediates poor prognosis by influencing multiple mechanisms [9,42].

Some limitations should be addressed. A low number of rectal tumours were analysed and the effect appears statistically less marked in the rectal only dataset GSE87211. However the relatively smaller number of events (22% and 13% for DFS and OS respectively) probably mitigated the statistical power. It may also due to an artefact of migrating between RNA expression platforms, although the signal persisted. Non-hypermethylated colon cancer and rectal cancer are not distinguishable at the genomic level [17] so we argue that the biology is consistent across all anatomical sites. Having accounted for sidedness and MSI where available, high NRF2 pathway activity remained a poor prognostic feature. Although the effect size (HR) may be considered small, the additional prognostic and biological information offers genuine clinical utility. We used a 36 genes to represent NRF2 pathway activation but NRF2 is known to regulate a large number of gene targets [49]. There may be an alternative group NRF2 targets which could better represent pathway activity in colorectal cancer. Fundamentally, the primary purpose of our analysis was to represent NRF2 signalling at a transcriptomic level so as to assess and understand its biological relevance to colorectal cancer. This is the first rigorous demonstration that NRF2 signalling pathway is a biomarker of poor clinical outcomes in CRC.

Conclusions

For the first time, NRF2 is shown to be a consistent, robust prognostic biomarker across all stages of colorectal cancer with additional clinical value to current known prognostic biomarkers. Better characterisation of its role and relationship to other biological factors in colorectal cancer is needed where high activity in CMS 4 refines the molecular profile of subgroup. The small number of genes needed to quantify NRF2 activation make it potentially suitable for development as a prognostic tool from routine clinical samples.

Declaration of Competing Interest

Dr. Maughan reports grants from Medical Research Council, grants from Cancer Research UK, during the conduct of the study; non-financial support and other from Astra Zeneca, personal fees from Pierre Fabre, outside the submitted work. Dr Holmes reports funding from Novartis for research collaboration outside the submitted work.

CRediT authorship contribution statement

: Conceptualization, Methodology, Resources, Writing - original draft, Writing - review & editing. : Conceptualization, Methodology, Data curation, Resources, Writing - original draft, Writing - review & editing. : Conceptualization, Resources, Writing - review & editing. : Resources, Writing - review & editing. : Conceptualization, Resources, Writing - review & editing. : Conceptualization, Resources, Writing - review & editing. : Conceptualization, Resources, Writing - review & editing.

Ethics approval and consent to participate

The use of patient material for the S:CORT program which provided the data from the MRC FOCUS cohort was approved by the East of England- Cambridge South Research Ethics Committee [NHS Health Research Authority] (REC reference 15/EE/0241)

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the Gene Expression Omnibus database repository [https://www.ncbi.nlm.nih.gov/geo/] using the accession numbers: GSE17536, GSE14333, GSE39582, GSE87211. The data from the MRC FOCUS are available from the corresponding author upon reasonable request.

Funding

The authors acknowledge the support of the UK Medical Research Council and Cancer Research UK stratified medicine consortium for colorectal cancer (S:CORT) in relation to collection, processing and quality control of the FOCUS trial samples, funding of CHW, and also the patients who participated in the trial. SMOC is supported by CRUK (grant numbers H3R00390.H376; CAN-RES-UK (C7932/A25142)). MAH is funded by Medical Research Council (grant number MC/PC/12001/2). Annabelle Lewis is supported by MRC (MR/P000738/1). CCH is supported by the Medical Research Council, the EPSRC, the Alan Turing Institute and the Li Ka Shing Centre for Health Innovation and Discovery. Aspects of this work have been presented in abstract form at ASCO GI 2019 symposium.

Acknowledgements

Not applicable
Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cancergen.2020.08.006.

References


