Cancer Genetics 248–249 (2020) 1–10

Contents lists available at ScienceDirect

Cancer Genetics

journal homepage: www.elsevier.com/locate/cancergen

Original Article

NRF2 metagene signature is a novel prognostic biomarker in colorectal cancer ☆

Séan M. O'Cathail[a,1,*], Chieh-Hsi Wu[b,1], Annabelle Lewis[c], Chris Holmes[b,d,e], Maria A Hawkins[f], Tim Maughan[g]

[a] Institute of Cancer Sciences, University of Glasgow, United Kingdom
[b] Department of Statistics, University of Southampton, United Kingdom
[c] Department of Life science, CHMLS, Brunel University London, United Kingdom
[d] Nuffield Department of Medicine, University of Oxford, United Kingdom
[e] Alan Turing Institute, London, United Kingdom
[f] University College London, United Kingdom
[g] Oxford Institute of Radiation Oncology, University of Oxford, United Kingdom

ARTICLE INFO

Article history:
Received 16 June 2020
Revised 14 August 2020
Accepted 17 August 2020

Keywords:
NRF2
Colorectal
Signature
Biomarker
Prognosis

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% C.I 1.2–2.004, $p = 0.0008$; GSE39582 HR=1.24, 95% C.I 1.086–1.416, $p = 0.001$; GSE87211 HR=1.431, 95% C.I 1.06–1.93, $p = 0.056$; MRC FOCUS trial HR=1.14, 95% C.I 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.

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 ration.

☆ All authors have read and approved the final manuscript and agree with submission.

* Corresponding author.
E-mail address: Sean.OCathail@glasgow.ac.uk (S.M. O'Cathail).

[1] Denotes equal contribution

https://doi.org/10.1016/j.cancergen.2020.08.006

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 *KEAP*1 in CRC silencing its ability to regulate NRF2 [19] and direct activation of NRF2 transcription via oncogenes $KRAS^{G12D}$, $BRAF^{V619E}$ and $c\text{-}MYC^{ERT2}$ 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 PCs[1] 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 'GEOquery' [26] and 'Biobase' [27] obtained from https://bioconductor.org/biocLite.R. All datasets are summarised in Table 1. The datasets were divided into a training 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

[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.

		GSE17536 (Training set)	GSE14333	GSE39582	MRC FOCUS trial	GSE87211
Patients		177	226	570	375 *(355 DNA mutant status)*	189
Tissue type		Fresh frozen	Fresh frozen	Fresh frozen	FFPE	Fresh frozen
Platform array		Affymetrix U133 v2.0	Affymetrix U133 v2.0	Affymetrix U133 v2.0	Affymetrix Xcel	Agilent Human 4 × 44k v2
Primary site		Colon	Colorectal	Colon	Colon	Rectum
Stage	I	24 (13.6%)	41 (18%)	37 (6.5%)		
	II	57 (32.2%)	95 (42%)	267 (47%)		70 (30.8%)
	III	57 (32.2%)	93 (41%)	206 (36%)		143 (63%)
	IV	39 (22%)		60 (10.5%)	375 (100%)	14 (6.2%)
Outcome variable		OS	DFS	DFS OS	OS	DFS OS
Covariates						
Chemo(radio)therapy			Yes 72 No 154	Yes 240 No 326	375 (100%)	189 (100%)
Site of primary				Prox. 232 Dist. 351	Left 203 Right 152	
BRAF V600E mutation				Mut 51	38 (10%)	
Mismatch Repair status				dMMR 77 pMMR 459	dMMR 15 pMMR 326	

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).

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, ANKRD29, ANKRD44, BCHE, C15orf48, COL3A1, COL5A1, EGLN3, LIFR, METTL7A, PCM1, PLAU, PLCB4, RECK, RGCC, RRM2,* SEC*14L4, SERPINH1, SFN, SLIT3, SPP1, TNS1, TOM1L2, TSPAN5, TTYH3, VSIG10, 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/II/III disease, higher NRF2 activity corresponded to worse DFS in GSE14333 (HR[2]=1.551, 95% C.I 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% C.I 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=1.140, 95% C.I 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% C.I 1.060–1.933, LRT $p = 0.056$) but not OS (HR=1.464, 95% C.I 0.955–2.245; LRT, $p = 0.197$). Fig. 3 shows that high activity corresponded to worse prognosis for DFS.

[2] 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.

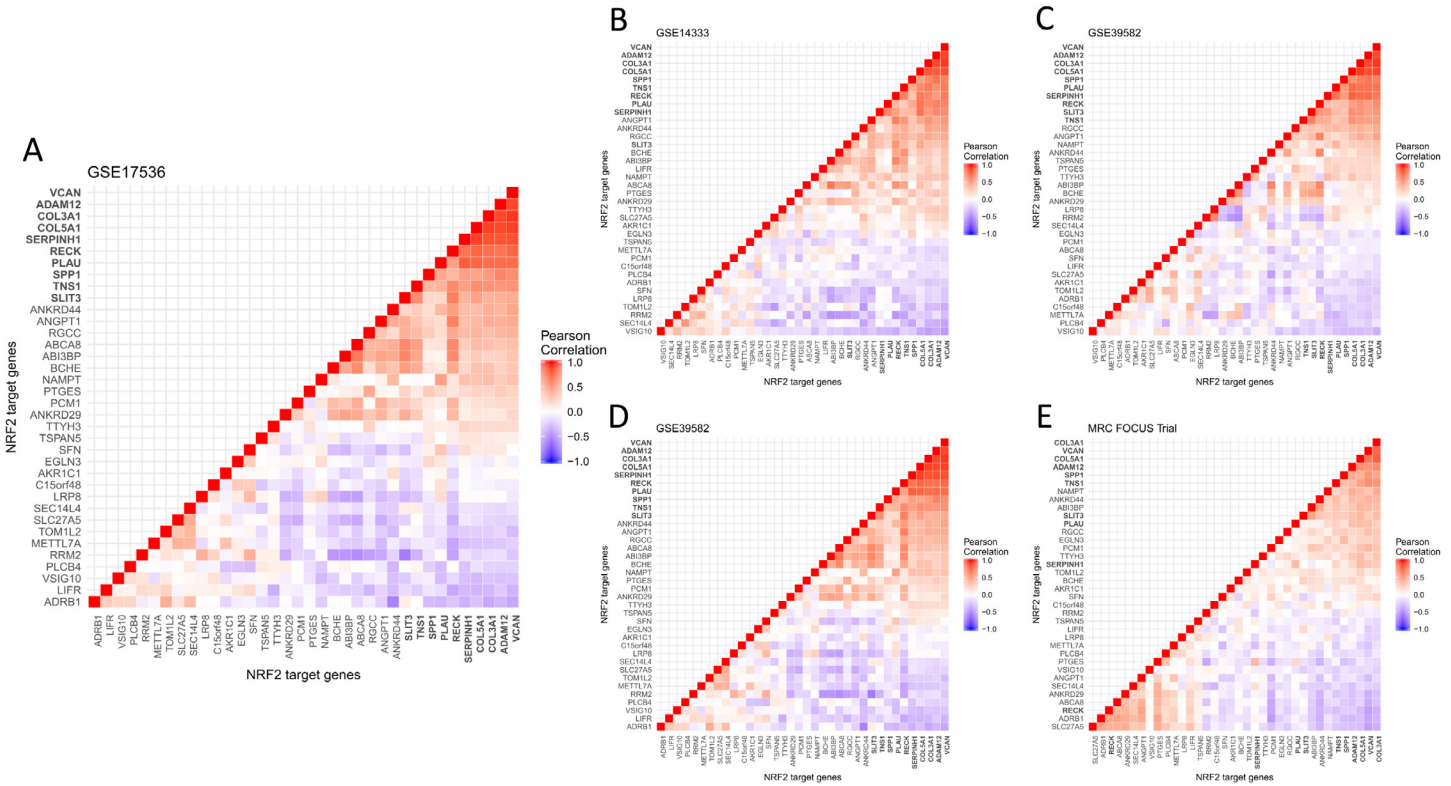

Fig. 1. A) Pairwise correlation heatmap showing the degree of positive (red) and negative (blue) Pearson correlation between the 36 genes of the NRF2 pathway in the training set. The high positive correlation between a subgroup of genes (bold text) indicate a high degree of co-expression. B) Pairwise correlation heatmaps Pearson correlation between the 36 genes of the NRF2 pathway in the four validations sets; *B* = GSE14333; *C* = GSE39582 RFS; *D* = GSE39582 OS; *E* = MRC FOCUS Trial. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

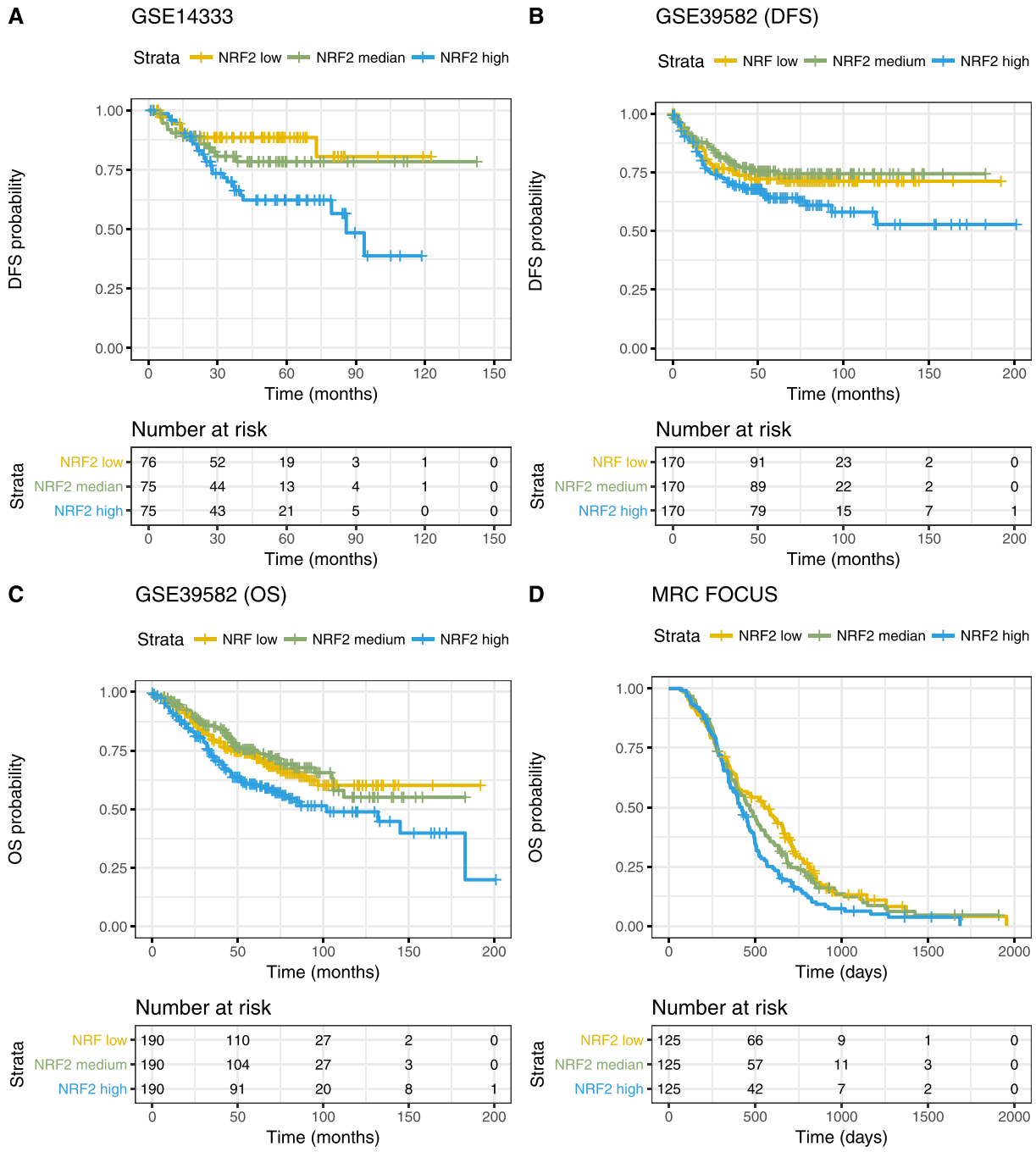

Fig. 2. Kaplan-Meier curves and associated risk tables for the primary analysis of three datasets. All show survival outcomes for patients with high, intermediate and low (induced by tertiles) of NRF2 metagene expression. A) GSE14333 and B) GSE39582 represents early stage I-III patients C) GSE39582 represents 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.

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[3]=1.365, 95% 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 *BRAF* ^{V600E} mutation. Again, high NRF2 activity corresponded to worse overall sur-

[3] See footnote 2.

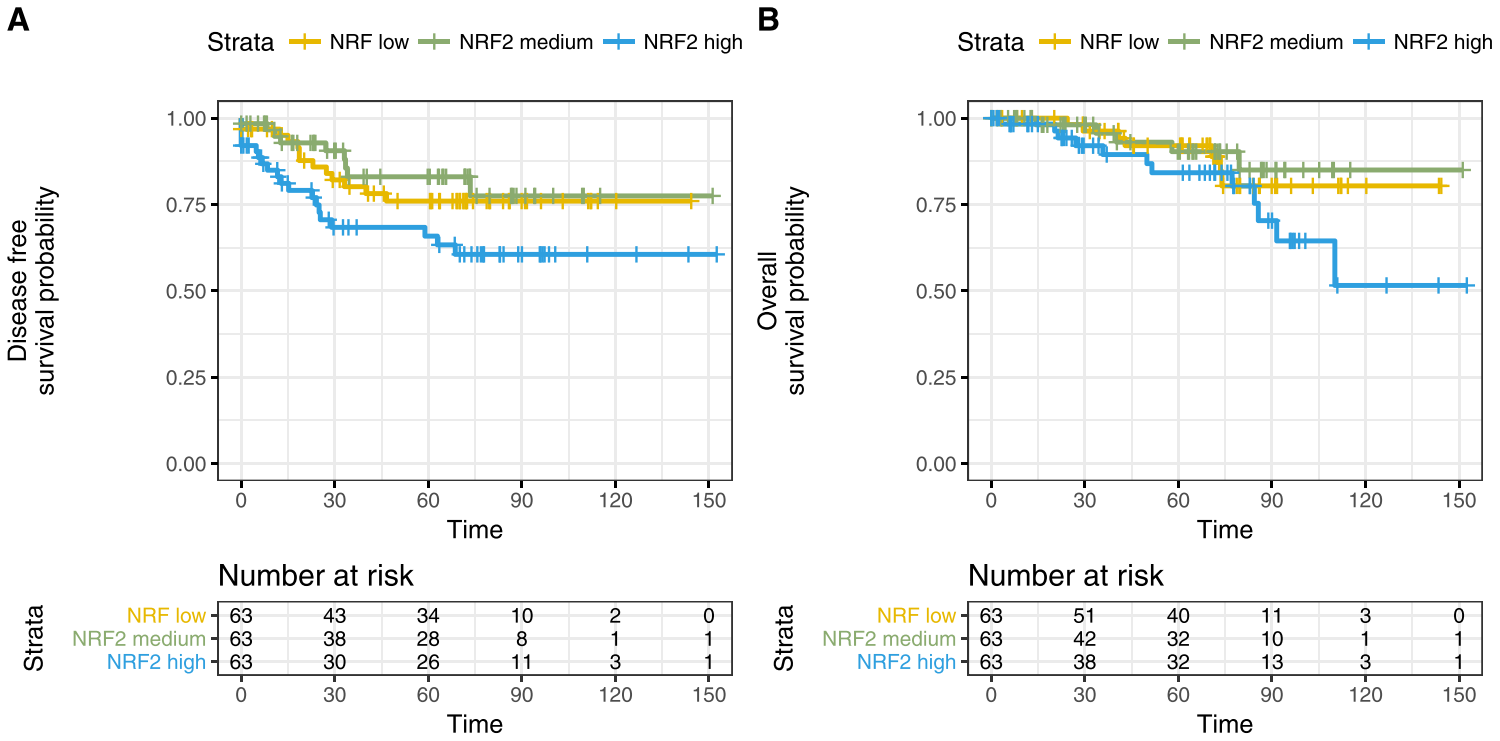

Fig. 3. Kaplan Meier curves and associated risk tables for the primary analyses of GSE87211. On the left, high NRF2 metagene expression is associated with worse disease free survival with persistent separation of the curves. On the right, there was no effect on overall survival. For the Kaplan Meier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

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

	N	Analysis	Variables	HR[a] (C.I)	LRT (p-value)
Data set					
GSE14333	226	Primary	N/A	1.113 (1.045–1.184)	0.0008
		Secondary	Stage Adjuvant chemotherapy		0.0201
MRC FOCUS	375	Primary	N/A	1.033[b] (1.009–1.245)	0.008
		Secondary	Sidedness[a] BRAF ^{V600E} mutation*		0.0185
GSE39582 (OS)	570	Primary	N/A	1.054 (1.02–1.089)	0.001
		Secondary	Stage		0.01
GSE39582 (RFS)	510	Primary	N/A	1.04 (1.002–1.08)	0.0383
		Secondary	Stage		0.049
			Mismatch repair (MMR)[a]		
GSE87211 (DFS)	189	Primary	N/A	1.067 (1.019–1.2445)	0.056
GSE87211 (OS)	189	Primary	N/A		0.197

It shows the numbers of patients included in each analysis, the adjust ing variables where used and the p-value for the Cox model comparison (LRT = Likelihood ratio test).
[a] 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.
[b] With imputation for missing variables.

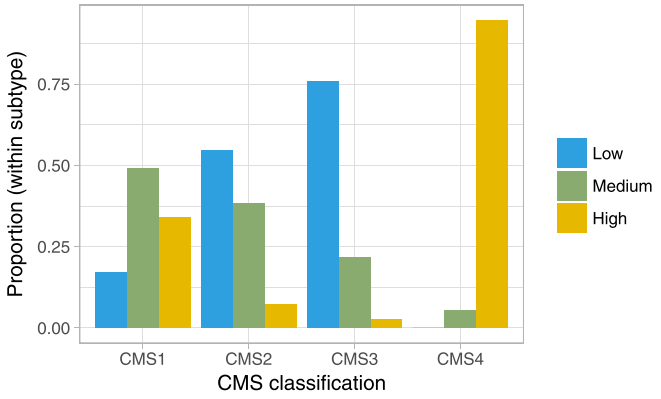

Fig. 4. CMS classification was derived for the MRC FOCUS trial dataset. The barplot shows the proportion of high and low NRF2 metagene expression in each of the four CMS subtypes. CMS 4 was substantially enriched for high NRF2 expression.

vival (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).

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 'metagene' 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 transcriptomic 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 ac-

count 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 pro-tumorigenic 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-hypermutated 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

Séan M. O'Cathail: Conceptualization, Methodology, Resources, Writing - original draft, Writing - review & editing. **Chieh-Hsi Wu:** Conceptualization, Methodology, Data curation, Resources, Writing - original draft, Writing - review & editing. **Annabelle Lewis:** Conceptualization, Resources, Writing - review & editing. **Chris Holmes:** Resources, Writing - review & editing, Supervision. **Maria A Hawkins:** Conceptualization, Resources, Writing - review & editing. **Tim Maughan:** Conceptualization, Resources, Writing - review & editing, Supervision.

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. SMO'C 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

[1] Bowel cancer incidence statistics. Cancer Res UK 2015. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/incidence (accessed November 17, 2016).
[2] Petrelli F, Tomasello G, Borgonovo K, Ghidini M, Turati L, Dallera P, et al. Prognostic survival associated with left-sided vs right-sided colon cancer: a systematic review and meta-analysis. JAMA Oncol 2017;3:211–19. doi:10.1001/jamaoncol.2016.4227.
[3] Stintzing S, Miller-Phillips L, Modest DP, Fischer von Weikersthal L, Decker T, Kiani A, et al. Impact of BRAF and RAS mutations on first-line efficacy of FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab: analysis of the FIRE-3 (AIO KRK-0306) study. Eur J Cancer 2017;79:50–60. doi:10.1016/j.ejca.2017.03.023.
[4] Flemer B, Lynch DB, Brown JMR, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. GUT 2017;66:633–43. doi:10.1136/gutjnl-2015-309595.
[5] Sjoquist KM, Renfro LA, Simes RJ, Tebbutt NC, Clarke S, Seymour MT, et al. Personalizing survival predictions in advanced colorectal cancer: the ARCAD nomogram project. JNCI J Natl Cancer Inst 2018;110:638–48. doi:10.1093/jnci/djx253.
[6] Popat S, Hubner R, Houlston R s. Systematic review of microsatellite instability and colorectal cancer prognosis. J Clin Oncol 2005;23:609–18. doi:10.1200/JCO.2005.01.086.
[7] Guastadisegni C, Colafranceschi M, Ottini L, Dogliotti E. Microsatellite instability as a marker of prognosis and response to therapy: a meta-analysis of colorectal cancer survival data. Eur J Cancer 2010;46:2788–98. doi:10.1016/j.ejca.2010.05.009.
[8] Guinney J, Dienstmann R, Wang X, Reyniès A, de, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350–6. doi:10.1038/nm.3967.
[9] Rojo de la Vega M, Chapman E, Zhang DD. NRF2 and the hallmarks of cancer. Cancer Cell 2018;34:21–43. doi:10.1016/j.ccell.2018.03.022.
[10] Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic signaling pathways in the cancer genome atlas. Cell 2018;173:321–37 e10. doi:10.1016/j.cell.2018.03.035.
[11] Taguchi K, Motohashi H, Yamamoto M. Molecular mechanisms of the Keap1–Nrf2 pathway in stress response and cancer evolution. Genes Cells 2011;16:123–40. doi:10.1111/j.1365-2443.2010.01473.x.
[12] Menegon S, Columbano A, Giordano S. The dual roles of NRF2 in Cancer. Trends Mol Med 2016;22:578–93. doi:10.1016/j.molmed.2016.05.002.
[13] Sporn MB, Liby KT. NRF2 and cancer: the good, the bad and the importance of context. Nat Rev Cancer 2012;12:564–71. doi:10.1038/nrc3278.
[14] Jaramillo MC, Zhang DD. The emerging role of the Nrf2–Keap1 signaling pathway in cancer. Genes Dev 2013;27:2179–91. doi:10.1101/gad.225680.113.
[15] Hiroshi Kitamura, Hozumi Motohashi. NRF2 addiction in cancer cells. Cancer Sci 2018. doi:10.1111/cas.13537.
[16] Yoo NJ, Kim HR, Kim YR, An CH, Lee SH. Somatic mutations of the KEAP1 gene in common solid cancers. Histopathology 2012;60:943–52. doi:10.1111/j.1365-2559.2012.04178.x.
[17] Network TCGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330–7. doi:10.1038/nature11252.
[18] Levings DC, Wang X, Kohlhase D, Bell DA, Slattery M. A distinct class of antioxidant response elements is consistently activated in tumors with NRF2 mutations. Redox Biol 2018;19:235–49. doi:10.1016/j.redox.2018.07.026.
[19] Hanada N, Takahata T, Zhou Q, Ye X, Sun R, Itoh J, et al. Methylation of the KEAP1 gene promoter region in human colorectal cancer. BMC Cancer 2012;12:66. doi:10.1186/1471-2407-12-66.
[20] DeNicola GM, Karreth FA, Humpton TJ, Gopinathan A, Wei C, Frese K, et al. Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. Nature 2011;475:106–9. doi:10.1038/nature10189.
[21] Chibon F. Cancer gene expression signatures – The rise and fall? Eur J Cancer 2013;49:2000–9. doi:10.1016/j.ejca.2013.02.021.
[22] Qian Z, Zhou T, Gurguis CI, Xu X, Wen Q, Lv J, et al. Nuclear factor, erythroid 2-like 2-associated molecular signature predicts lung cancer survival. Sci Rep 2015:5. doi:10.1038/srep16889.
[23] Namani A, Cui QQ, Wu Y, Wang H, Wang XJ, Tang X, et al. NRF2-regulated metabolic gene signature as a prognostic biomarker in non-small cell lung cancer. Oncotarget 2017:5. doi:10.18632/oncotarget.19349.
[24] Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia N Y N 2004;6:1–6.
[25] Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci 2003;100:9440–5. doi:10.1073/pnas.1530509100.
[26] Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007;23:1846–7. doi:10.1093/bioinformatics/btm254.
[27] Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 2015;12:115–21. doi:10.1038/nmeth.3252.
[28] Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang A, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology 2010;138:958. doi:10.1053/j.gastro.2009.11.005.
[29] Jorissen RN, Gibbs P, Christie M, Prakash S, Lipton L, Desai J, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with Dukes' stage B and C colorectal cancer. Clin Cancer Res Off J Am Assoc Cancer Res 2009;15:7642–51. doi:10.1158/1078-0432.CCR-09-1431.
[30] Marisa L, Reyniès A de, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLOS Med 2013;10:e1001453. doi:10.1371/journal.pmed.1001453.
[31] Seymour MT, Maughan TS, Ledermann JA, Topham C, James R, Gwyther SJ, et al. Different strategies of sequential and combination chemotherapy for patients with poor prognosis advanced colorectal cancer (MRC FOCUS): a randomised controlled trial. Lancet 2007;370:143–52. doi:10.1016/S0140-6736(07)61087-3.
[32] Hu Y, Gaedcke J, Emons G, Beissbarth T, Grade M, Jo P, et al. Colorectal cancer susceptibility loci as predictive markers of rectal cancer prognosis after surgery. Genes Chromosomes Cancer 2018;57:140–9. doi:10.1002/gcc.22512.
[33] R Core Team R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
[34] Gaedcke J., Grade M., Jung K., Camps J., Jo P., Emons G., et al. Mutated KRAS results in overexpression of DUSP4, a MAP-kinase phosphatase, and SMYD3, a histone methyltransferase, in rectal carcinomas. Genes Chromosomes Cancer n.d.;49:1024–34. 10.1002/gcc.20811.
[35] Graudens E, Boulanger V, Mollard C, Mariage-Samson R, Barlet X, Grémy G, et al. Deciphering cellular states of innate tumor drug responses. Genome Biol 2006;7:R19. doi:10.1186/gb-2006-7-3-r19.
[36] Hong Y, Downey T, Eu KW, Koh PK, Cheah PY. A 'metastasis-prone' signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics. Clin Exp Metastasis 2010;27:83–90. doi:10.1007/s10585-010-9305-4.
[37] Kaiser S, Park Y-K, Franklin JL, Halberg RB, Yu M, Jessen WJ, et al. Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer. Genome Biol 2007;8:R131. doi:10.1186/gb-2007-8-7-r131.
[38] Sabates-Bellver J, der Flier LGV, de Palo M, Cattaneo E, Maake C, Rehrauer H, et al. Transcriptome profile of human colorectal adenomas. Mol Cancer Res 2007;5:1263–75. doi:10.1158/1541-7786.MCR-07-0267.
[39] Skrzypczak M, Goryca K, Rubel T, Paziewska A, Mikula M, Jarosz D, et al. Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. PLoS One 2010;5:e13091. doi:10.1371/journal.pone.0013091.
[40] Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucl Acids Res 2017;45:D362–8. doi:10.1093/nar/gkw937.
[41] Sveen A, Kopetz S, Lothe RA. Biomarker-guided therapy for colorectal cancer: strength in complexity. Nat Rev Clin Oncol 2019:1. doi:10.1038/s41571-019-0241-1.
[42] Cloer EW, Goldfarb D, Schrank TP, Weissman BE, Major MB. NRF2 activation in cancer: from DNA to protein. Cancer Res 2019. doi:10.1158/0008-5472.CAN-18-2723.
[43] Hayes JD, Dinkova-Kostova AT, Tew KD. Oxidative stress in cancer. Cancer Cell 2020;38:167–97. doi:10.1016/j.ccell.2020.06.001.
[44] Ji L, Wei Y, Jiang T, Wang S. Correlation of NRF2, NQO1, MRP1, CMYC and p53 in colorectal cancer and their relationships to clinicopathologic features and survival. Int J Clin Exp Pathol 2014;7:1124–31.
[45] Hu T, Yao Y, Yu S, Guo H, Han L, Wang W, et al. Clinicopathologic significance of CXCR4 and NRF2 in colorectal cancer. J Biomed Res 2013;27:283–90. doi:10.7555/JBR.27.20130069.
[46] Romero R, Sayin VI, Davidson SM, Bauer MR, Singh SX, LeBoeuf SE, et al. *KEAP1* loss promotes *KRAS*-driven lung cancer and results in dependence on glutaminolysis. Nat Med 2017;23:1362–8. doi:10.1038/nm.4407.

[47] Stewart PA, Welsh EA, Slebos RJC, Fang B, Izumi V, Chambers M, et al. Proteogenomic landscape of squamous cell lung cancer. Nat Commun 2019;10:1–17. doi:10.1038/s41467-019-11452-x.

[48] Kang KA, Piao MJ, Kim KC, Kang HK, Chang WY, Park IC, et al. Epigenetic modification of NRF2 in 5-fluorouracil-resistant colon cancer cells: involvement of TET-dependent DNA demethylation. Cell Death Dis 2014;5:e1183. doi:10.1038/cddis.2014.149.

[49] Malhotra D, Portales-Casamar E, Singh A, Srivastava S, Arenillas D, Happel C, et al. Global mapping of binding sites for NRF2 identifies novel targets in cell survival response through CHIP-SEQ profiling and network analysis. Nucleic Acids Res 2010;38:5718–34. doi:10.1093/nar/gkq212.