

SCIENTIFIC REPORTS



OPEN

Nuclear factor, erythroid 2-like 2-associated molecular signature predicts lung cancer survival

Zhongqing Qian^{1,2,*}, Tong Zhou^{2,*}, Christopher I. Gurguis², Xiaoyan Xu², Qing Wen^{3,4}, Jingzhu Lv⁵, Fang Fang¹, Louise Hecker^{2,6}, Anne E. Cress⁷, Viswanathan Natarajan^{8,9}, Jeffrey R. Jacobson⁹, Donna D. Zhang⁴, Joe G. N. Garcia² & Ting Wang²

Received: 30 March 2015
Accepted: 01 October 2015
Published: 24 November 2015

Nuclear factor, erythroid 2-like 2 (*NFE2L2*), a transcription factor also known as NF-E2-related factor 2 (Nrf2), is a key cytoprotective gene that regulates critical antioxidant and stress-responsive genes. Nrf2 has been demonstrated to be a promising therapeutic target and useful biomarker in malignant disease. We hypothesized that *NFE2L2*-mediated gene expression would reflect cancer severity and progression. We conducted a meta-analysis of microarray data for 240 *NFE2L2*-mediated genes that were enriched in tumor tissues. We then developed a risk scoring system based on *NFE2L2* gene expression profiling and designated 50 tumor-associated genes as the *NFE2L2*-associated molecular signature (NAMS). We tested the relationship between this gene expression signature and both recurrence-free survival and overall survival in lung cancer patients. We find that NAMS predicts clinical outcome in the training cohort and in 12 out of 20 validation cohorts. Cox proportional hazard regressions indicate that NAMS is a robust prognostic gene signature, independent of other clinical and pathological factors including patient age, gender, smoking, gene alteration, MYC level, and cancer stage. NAMS is an excellent predictor of recurrence-free survival and overall survival in human lung cancer. This gene signature represents a promising prognostic biomarker in human lung cancer.

Nuclear factor, erythroid 2-like 2 (*NFE2L2*), also known as NF-E2-Related Factor 2 (Nrf2), is a transcription factor encoded by the *NFE2L2* gene in humans¹. Nrf2 regulates the transcription of a wide array of genes that code for antioxidants and other proteins responsible for the detoxification of xenobiotics and reactive oxygen species, including a battery of phase II enzymes, such as NAD(P)H quinone oxidoreductase 1 (Nqo1), Heme oxygenase-1 (HO-1), and glutathione S-transferase (GST)². Under physiological conditions, cytosolic Nrf2 protein is maintained at very low levels by its selective negative regulator, Kelch-like ECH-associated protein 1 (KEAP1), a cytoplasmic protein which sequesters Nrf2 in the cytoplasm and directs it to CUL3 E3 ligase for ubiquitylation and subsequent degradation by the proteasome^{3,4}. Under conditions of oxidative stress or in the presence of Nrf2 activating compounds, E3 activity is compromised and Nrf2 is stabilized, with increasing the amount of Nrf2 relative to Keap1². Free Nrf2 moves to the nucleus, thereby activating expression of its downstream antioxidant genes^{3,5,6}.

¹Key Laboratory of Anhui Province for Infection and Immunology, Bengbu Medical College, Bengbu 233003, China. ²Arizona Respiratory Center and Department of Medicine, The University of Arizona, Tucson, Arizona, USA. ³Department of Pharmacy, Jinan Central Hospital, Jinan, Shandong, China. ⁴Department of Pharmacology and Toxicology, The University of Arizona, Tucson, Arizona, USA. ⁵Department of Biochemistry and Molecular Biology, Bengbu Medical College, Bengbu 233003, China. ⁶Southern Arizona VA Health Care System, Tucson, AZ 85723, USA. ⁷Arizona Cancer Center and Department of Cellular and Molecular Medicine, University of Arizona, Tucson, AZ, USA. ⁸Department of Pharmacology, The University of Illinois at Chicago, Chicago, IL, USA. ⁹Department of Medicine, The University of Illinois at Chicago, Chicago, IL, USA. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.W. (email: twang@email.arizona.edu)

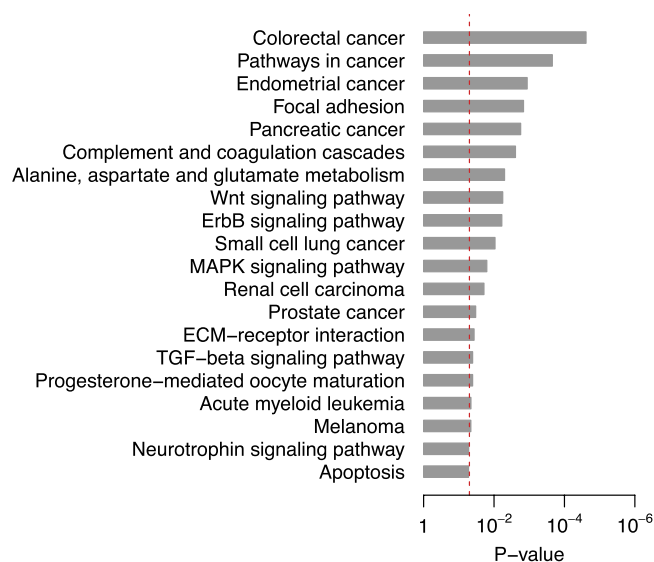


Figure 1. The top 20 KEGG pathways enriched in the *NFE2L2*-mediated genes in lung cancer cell. The *P*-values were calculated by Fisher's exact test. The red dash line denotes the significance level of $\alpha = 0.05$.

As a central regulator of antioxidant genes, *NFE2L2* has received great attention for its pivotal role in a number of pathologic conditions, including cancer^{7–10}. Increasingly, data have demonstrated that *NFE2L2* is over expressed in a great number of solid and hematologic tumors^{11–16}. Many cancer cells have been found to exhibit increased expression and activity of Nrf2¹⁷. In addition, *NFE2L2* is highly up-regulated in various types of tumors and the prognosis of patients with tumors expressing high levels of *NFE2L2* is poor¹⁸.

As *NFE2L2* becomes dysregulated, cells may acquire several traits, including proliferation, apoptosis resistance, and a profound resistance to drugs and radiotherapy, which may promote tumor growth and cancer pathogenesis¹⁹. Nrf2 activators have been used in clinical trials for cancer therapy and the treatment of diseases associated with oxidative stress; on the other hand, constitutive activation of Nrf2 in many types of tumors contributes to the survival and growth of cancer cells, as well as resistance to anticancer therapy²⁰. Therefore, manipulation of Nrf2 has become of great interest with regard to its use in therapy and the diagnosis/prognosis of malignant diseases.

In this present study, we conducted secondary analysis of genome-wide expression data to identify *NFE2L2*-associated genes implicated in cancer pathobiology. Because of its strong association with cancer, we hypothesized that *NFE2L2*-mediated genes could be used to indicate cancer progression. Specifically, we aimed to assess whether genes associated with *NFE2L2* could be used as a prognostic tool for lung cancer patients. We found that an *NFE2L2*-mediated gene signature could effectively predict lung cancer survival.

Results

***NFE2L2* influences gene expression in human lung cancer.** We compared the gene expression patterns of control cells to those of *NFE2L2* knockdown (KD) human lung cancer cells to identify genes potentially regulated by *NFE2L2*. One microarray dataset (GSE38332) containing gene expression information from both control and *NFE2L2* KD A549 lung cancer cells²¹ was downloaded from the Gene Expression Omnibus (GEO) database²². At the specified significance level of false discovery rate (FDR) < 0.01 and fold change > 2 (see Methods for details), 1631 probe sets encoding 1172 genes were found to be up-regulated in *NFE2L2* KD cells, while 792 probe sets for 593 genes were down-regulated in *NFE2L2* KD cells (Supplementary Table S1). We next searched the enriched Kyoto Encyclopedia of Genes and Genomes (KEGG)²³ physiological pathways among the dysregulated genes revealing genes enriched in cancer-related KEGG terms (9 out of 20 significantly dysregulated pathways), such as “Colorectal cancer”, “Pathways in cancer”, “Endometrial cancer”, “Pancreatic cancer”, “Small cell Lung cancer”, “Renal cell carcinoma”, and “Prostate cancer” (Fig. 1). To address whether the *NFE2L2*-mediated genes (1765 differentially expressed genes) are statistically significantly associated with KEGG cancer pathways, we conducted a resampling test. We obtained 1,000 random gene sets by randomly selecting 1765 genes from human genome. For each random gene set, we counted the number of KEGG cancer pathways that are significantly associated with the random gene set. We found that the number of KEGG cancer pathways of the *NFE2L2*-mediated genes is significantly larger than that of the random gene sets ($P = 0.002$) (Supplementary Fig. S1). The results suggest that the *NFE2L2*-mediated genes are involved in human lung cancer pathology.

To explore the role of *NFE2L2*-mediated genes in human lung cancers, we compared the gene expression between the paired normal and tumor tissues in two lung cancer cohorts from Spain (GSE18842, ESP)²⁴ and Taiwan (GSE19804, TWN)²⁵. In total, 1695 probe sets encoding 1284 genes were found to be commonly differentially expressed (see Methods for details) between normal and tumor tissues in the two cohorts (Supplementary Table S2). Among these probe sets/genes, 299 probe sets encoding 240 genes were found to overlap with the *NFE2L2*-mediated genes (Supplementary Table S3), which is statistically significant ($P = 2.2 \times 10^{-14}$ by hypergeometric cumulative distribution function) and, thus, suggests that *NFE2L2*-mediated genes are significantly enriched among the lung cancer-related genes. The top KEGG pathways associated with the 240 overlapped genes includes “ECM-receptor interaction”, “Pathways in cancer”, “Complement and coagulation cascades”, “Jak-STAT signaling pathway”, “Focal adhesion”, “Small cell lung cancer”, and “TGF-beta signaling pathway” (Supplementary Fig. S2), which suggests the depth of involvement of *NFE2L2*-mediated genes in human cancer. Univariate Cox proportional hazards regression was applied to evaluate the relationship between lung cancer outcome and gene expression for the 240 genes. The association between recurrence-free survival and gene expression was computed in a lung cancer training cohort from Korea (GSE8894, KOR)²⁶. We also calculated the association between overall survival and gene expression in another training cohort from the United States (GSE3141, USA)²⁷. The Wald statistic (ratio of Cox regression coefficient to its standard error) of the 240 genes in the KOR cohort (Z_R) is positively correlated with that in the USA cohort (Z_D) (Supplementary Fig. S3). Among the 240 genes, we identified 50 genes with both $|Z_R| > 0.5$ and $|Z_D| > 0.5$. We hypothesized that the expression of the 50 genes would predict lung cancer outcome. We designated these genes as the *NFE2L2*-Associated Molecular Signature (NAMS) (Table 1 and Supplementary Fig. S4 and S5). The NAMS gene set (50 genes) is significantly associated with two KEGG pathways: ECM-receptor interaction and focal adhesion (Supplementary Fig. S6).

NAMS predicts recurrence-free survival in lung cancer. To confirm that the NAMS would be predictive of tumor outcome in lung cancer, we constructed a scoring system to assign each patient a recurrence-risk score, representing a linear combination of the NAMS gene expression values weighted by the coefficients obtained from the training cohort (KOR) (see Methods for details). NAMS_R^+ patients were defined as those having recurrence-risk scores greater than zero, and the other patients were assigned as NAMS_R^- . As expected, we found a significantly reduced recurrence-free survival for the NAMS_R^+ patients in the training cohort (Fig. 2 and Table 2).

We next tested the prognostic power of the NAMS based recurrence-risk score in three validation cohorts from Japan (GSE31210, JPN)²⁸, Sweden (GSE37745, SWE)²⁹, and Canada (GSE50081, CAN)³⁰. Kaplan-Meier survival curves demonstrated a significantly reduced recurrence-free survival for NAMS_R^+ patients in the three validation cohorts (log-rank test: $P = 1.5 \times 10^{-6}$ for JPN, $P = 2.5 \times 10^{-3}$ for SWE, and $P = 6.0 \times 10^{-3}$ for CAN) (Fig. 2). Univariate Cox proportional hazards regression indicated that NAMS_R^+ patients had an increased risk for recurrence of 3.59-, 2.65-, and 2.36-fold in the JPN, SWE, and CAN cohorts, respectively (Table 2). These findings suggest that the NAMS is predictive of recurrence-free survival in lung cancer.

In addition, we checked prognostic power of NAMS on different types of lung cancer (including adenocarcinoma, squamous cell carcinoma, and large cell carcinoma). JPN cohort are all significantly reduced recurrence-free survival for NAMS_R^+ adenocarcinoma patients ($P = 4.7 \times 10^{-4}$), but not squamous cell carcinoma patients, was found in the CAN cohort (Supplementary Fig. S7). In the SWE cohort, significantly reduced recurrence-free survival for NAMS_R^+ squamous cell carcinoma patients ($P = 2.2 \times 10^{-2}$), but not the other two types, was observed (Supplementary Fig. S8). Regardless of statistical significance (mainly due to the limitation of the same size), all recurrence-free survival data in all types of cancers remain at the same trend that NAMS_R^+ patients have reduced survival.

NAMS predicts overall survival in lung cancer. We also tested the power of the NAMS in predicting overall survival in lung cancer. The USA cohort was used for training. Similarly, a scoring system was developed to assign each patient a death-risk score, calculated as a linear combination of the NAMS gene expression values weighted by the coefficients obtained from the USA cohort (see Methods for details). NAMS_D^+ patients were defined as those having death-risk scores greater than zero, while the other patients were NAMS_D^- . Kaplan-Meier survival curves demonstrated a significantly reduced overall survival for the NAMS_D^+ patients in both the training and validation cohorts (Fig. 3). Univariate Cox proportional hazards regression indicated that the NAMS_D^+ patients had an increased risk for death of 2.71-, 5.92-, 1.46-, and 1.76-fold in the USA, JPN, SWE, and CAN cohorts, respectively (Table 2). These results suggest that NAMS can be used to predict overall survival in lung cancer.

Same as the recurrence-free survival study, we examined the prognostic power of NAMS on subtypes of lung cancer on overall survival. Significantly reduced recurrence-free survival for NAMS_R^+ adenocarcinoma patients ($P = 7.7 \times 10^{-3}$), but not squamous cell carcinoma patients, was found in CAN cohort (Supplementary Fig. S9). In SWE cohort, NAMS fails to differentiate overall survival significantly in three subtypes of lung cancer (Supplementary Fig. S10). Same as recurrence-free survival study, all overall survival data in all lung cancer subtypes remain at the same trend that NAMS_R^+ patients have reduced survival, regardless of statistical significance.

Gene symbol	Gene title
<i>ABCA8</i>	ATP-binding cassette, sub-family A (ABC1), member 8
<i>ABI3BP</i>	ABI family, member 3 (NESH) binding protein
<i>ADAM12</i>	ADAM metallopeptidase domain 12
<i>ADRB1</i>	adrenoceptor beta 1
<i>ANGPT1</i>	angiopoietin 1
<i>ANKRD29</i>	ankyrin repeat domain 29
<i>ANKRD44</i>	ankyrin repeat domain 44
<i>ATL3</i>	atlastin GTPase 3
<i>BCHE</i>	butyrylcholinesterase
<i>C15orf48</i>	chromosome 15 open reading frame 48
<i>COL3A1</i>	collagen, type III, alpha 1
<i>COL5A1</i>	collagen, type V, alpha 1
<i>DEPDC7</i>	DEP domain containing 7
<i>EGLN3</i>	egl nine homolog 3 (<i>C. elegans</i>)
<i>EHF</i>	ets homologous factor
<i>GALNT7</i>	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7 (GalNAc-T7)
<i>GPX3</i>	glutathione peroxidase 3 (plasma)
<i>HLA-E</i>	major histocompatibility complex, class I, E
<i>IPO4</i>	importin 4
<i>ITGB4</i>	integrin, beta 4
<i>LIFR</i>	leukemia inhibitory factor receptor alpha
<i>MARC2</i>	mitochondrial amidoxime reducing component 2
<i>MED20</i>	mediator complex subunit 20
<i>METTL7A</i>	methyltransferase like 7A
<i>NCALD</i>	neurocalcin delta
<i>PCM1</i>	pericentriolar material 1
<i>PLAU</i>	plasminogen activator, urokinase
<i>PLCB4</i>	phospholipase C, beta 4
<i>PLEKHH2</i>	pleckstrin homology domain containing, family H (with MyTH4 domain) member 2
<i>RECK</i>	reversion-inducing-cysteine-rich protein with kazal motifs
<i>RGCC</i>	regulator of cell cycle
<i>RILPL2</i>	Rab interacting lysosomal protein-like 2
<i>RRM2</i>	ribonucleotide reductase M2
<i>SEC14L4</i>	SEC14-like 4 (<i>S. cerevisiae</i>)
<i>SERPINH1</i>	serpin peptidase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1)
<i>SFN</i>	stratifin
<i>SLIT3</i>	slit homolog 3 (<i>Drosophila</i>)
<i>SPP1</i>	secreted phosphoprotein 1
<i>TACCI</i>	transforming, acidic coiled-coil containing protein 1
<i>TBX2</i>	T-box 2
<i>TNS1</i>	tensin 1
<i>TOM1L2</i>	target of myb1-like 2 (chicken)
<i>TPPP</i>	tubulin polymerization promoting protein
<i>TSPAN5</i>	tetraspanin 5
<i>TTYH3</i>	tweety homolog 3 (<i>Drosophila</i>)
<i>TXNL1</i>	thioredoxin-like 1
Continued	

Gene symbol	Gene title
VARS	valyl-tRNA synthetase
VCAN	versican
VSIG10	V-set and immunoglobulin domain containing 10
ZNF25	zinc finger protein 25

Table 1. NAMS gene set.

The prognostic power of NAMS is non-random. A computational study indicated that the prognostic power of most published gene signatures in breast cancer are not significantly better than that of random signatures with identical size³¹. Here, we performed a resampling test to check whether the NAMS performed better than random gene signatures. We constructed 1,000 random gene signatures with identical size as the NAMS (50 genes). Cox proportional hazards regression of survival was conducted for each resampled gene signature. The association between each random gene signature and recurrence-free/overall survival was measured by the sum of Wald statistic in the three validation cohorts. Our alternative hypothesis was that the sum of the Wald statistic of the NAMS would be more positive than expected by chance if the prognostic power of the NAMS was significantly better than that of random gene signatures. We found that the sum of the Wald statistic of the NAMS was significantly larger than that of randomized gene signatures ($P < 0.001$ for recurrence-free survival and $P = 0.001$ for overall survival) (Fig. 4).

We also compared the prognostic power of the NAMS against the other cancer related genes. Here, we defined the cancer related genes as those being differentially expressed between normal and tumor lung tissues (listed in Supplementary Table S2) and with $|Z_R| > 0.5$ or $|Z_D| > 0.5$. We performed a resampling test to check whether the prognostic power of NAMS was statistically better than the other cancer related genes. For each round of randomization, 50 genes were picked up from the pool of the cancer related genes. The performance of the random gene set was quantified by the sum of the Wald statistic of the validation cohorts. The prognostic power of the NAMS was significantly better than that of 1000 random gene signatures consisting of cancer related genes ($P = 0.002$ for recurrence-free survival and $P = 0.046$ for overall survival) (Fig. 4).

Independence of NAMS from the traditional clinical and pathological factors. To confirm the strength of the NAMS as an independent predictor, we investigated the performance of the NAMS in comparison with the traditional clinical and pathological variables associated with prognosis in lung cancers. For the JPN cohort, we considered factors including age, gender, smoking history, stage, *EGFR/KRAS/ALK* mutation status (Gene alteration), and *MYC* level. For the SWE cohort, we included age, gender, and stage as covariate. For the CAN cohort, age, gender, smoking history, and stage were taken into account. Multivariate Cox proportional hazards regression of recurrence-free survival indicated that NAMS remained a significant covariate in relation to the clinical and pathological factors in all the validation cohorts (Table 3). Interestingly, NAMS was the most significant (lowest P -value) covariate in each cohort for recurrence-free survival (Table 3). We also conducted multivariate Cox proportional hazards regression against overall survival. We found that NAMS was still a significant covariate in the JPN and SWE cohorts (Table 3). However, the P -value of NAMS was slightly above the α -level of 0.05 ($P = 0.075$) in the CAN cohort (Table 3). Because the covariates differed among each cohort, we also repeated these analyses using the same three covariates (age, gender, and stage) for the three cohorts. These results did not dramatically change our interpretation (reported in Supplementary Table S4). Taken together, the NAMS is a survival predictor in cancer patients, independent of the traditional clinical and pathological factors.

Discussion

Nrf2 is a transcription factor that acts as a master regulator for the expression of a wide array of anti-oxidant genes³². Up-regulation of *NFE2L2* by chemopreventive compounds confer protection against cancer initiation^{15,16}. Several studies have indicated that dysregulation of *NFE2L2* is strongly associated with human cancer^{7,8}. Though the full extent to which *NFE2L2* is involved in tumorigenesis is not known, genetic analyses have shown that *NFE2L2* has increased mutations or has been deregulated in human cancers^{12,33,34}. The role of *NFE2L2* in pulmonary neoplasia, a diverse disease for which few biomarkers exist, is complicated and appears to depend on several factors, including the existence of activating mutations in *NFE2L2* and/or loss of function mutations in *KEAP1*³⁵. Through a computational genomics approach, our current study confirmed a central role for *NFE2L2* in lung cancer. The complex functions of Nrf2 in carcinogenesis and chemotherapy resistance require more detailed characterization and mechanistic analyses, however, our data reveals some interesting patterns, in light of current knowledge about lung malignancy.

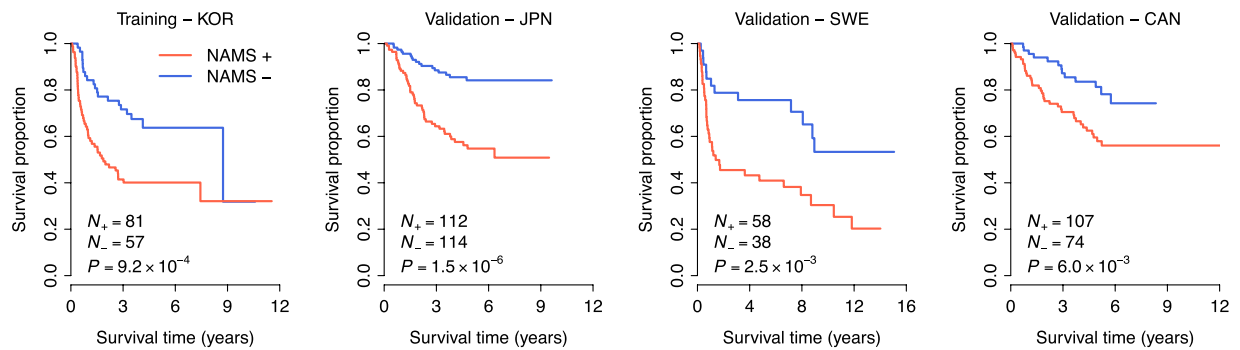


Figure 2. Kaplan-Meier curves of recurrence-free survival. The expression of the NAMS predicts poor recurrence-free survival in the discovery (KOR) and validation (JPN, SWE, and CAN) cohorts. The red curves are for the NAMS positive patients while the blue curves are for the NAMS negative patients. The NAMS positive patients were defined as those having a recurrence risk score greater than zero. P-values were calculated by log-rank tests for the differences in survival between the NAMS positive and negative groups.

	Recurrence-free survival				Overall survival			
	Cohort	HR	95% CI	P-value	Cohort	HR	95% CI	P-value
Training	KOR	2.35	(1.40, 3.96)	1.3×10^{-3}	USA	2.71	(1.52, 4.85)	7.8×10^{-4}
Validation	JPN	3.59	(2.06, 6.25)	6.7×10^{-6}	JPN	5.92	(2.46, 14.28)	7.5×10^{-5}
	SWE	2.65	(1.38, 5.12)	3.6×10^{-3}	SWE	1.46	(1.05, 2.03)	2.6×10^{-2}
	CAN	2.36	(1.25, 4.42)	7.7×10^{-3}	CAN	1.76	(1.08, 2.87)	2.5×10^{-2}

Table 2. Cox proportional hazards regression of survival by NAMS status. Note – HR: hazard ratio; CI: confidence interval.

Lung cancer is the most frequently diagnosed cancers and leading cause of cancer death in males, comprising 17% of the total new cancer cases and 23% of the total cancer deaths³⁶. With the rapid advances taking place in molecular testing and associated technologies, the landscape of targetable genomic alterations in lung cancer is beginning to uncover the true complexity of the disease³⁷. In this study, we explored the prognostic value of those gene sets regulated by *NFE2L2* in lung cancer. First, we compared genes that were differentially expressed between lung cancer and *NFE2L2*-knockdown A549 cells. Over 2425 genes were found to be commonly differentially expressed between WT and *NFE2L2*-knockdown cells. We confirmed the critical role of *NFE2L2* in carcinogenesis by the gene ontology analysis of all *NFE2L2*-mediated genes: 12 of 20 significantly deregulated pathways are direct cancer pathways (Fig. 1). Second, we generated the *NFE2L2*-associated Molecular Signature (NAMS) by filtering gene express data sets through different cohorts. Third, we validated the NAMS as a powerful tool which provides important prognostic predictions in lung cancer, as we demonstrated that the NAMS was a significant and independent predictor of recurrence-free cancer survival. Although limited by the availability of the existing lung cancer microarray datasets deposited in GEO, the consistent findings from the population with diverse genetic background in discovery and validation cohorts (e.g., USA versus KOR) strongly suggest the NAMS in lung cancer is powerful and highly conserved molecular signature among populations. Finally, these NAMS predictions were independent of other known clinical pathological covariates in this study. This last result is especially intriguing given recent developments in the way that clinicians and cancer biologists have begun to think about the cancer disease process.

Although the origin of cancer is still hotly debated³⁸, cancer is unequivocally a quantitative trait³⁹. Our findings are consistent with a modern understanding of carcinogenesis as involving not only a somatic evolutionary process^{40,41}, but also contributions from the tumor microenvironment^{42,43}, as well as general dysregulation at the tissue level⁴⁴. In fact, the 50 gene set that comprised our NAMS itself highlights that cancer is a complex trait with many gene pathways involved (a fact that seems to be of broad significance, see also^{45,46}). These pathway-of-cancer enriched genes provide a set of *NFE2L2*-associated genes that might play critical roles in cancer pathogenesis. For example, RRM2 plays an important role in regulating expression of the anti-apoptotic protein Bcl-2 and reveal a critical link between RRM2 and Bcl-2 in apoptosis signaling and tumor developing⁴⁷. Recent studies showed that concomitant low expression levels of RRM2 was predictive of a better outcome, and low expression of RRM2 could be used to predict the treatment response to platinum-based chemotherapy and survival in lung cancer^{48,49}.

This study reinforces the value of re-examining available genomic/genetic data in the “big data” era with a novel translational approach. *NFE2L2* is confirmed to be a novel “oncogene” (broadly construed)

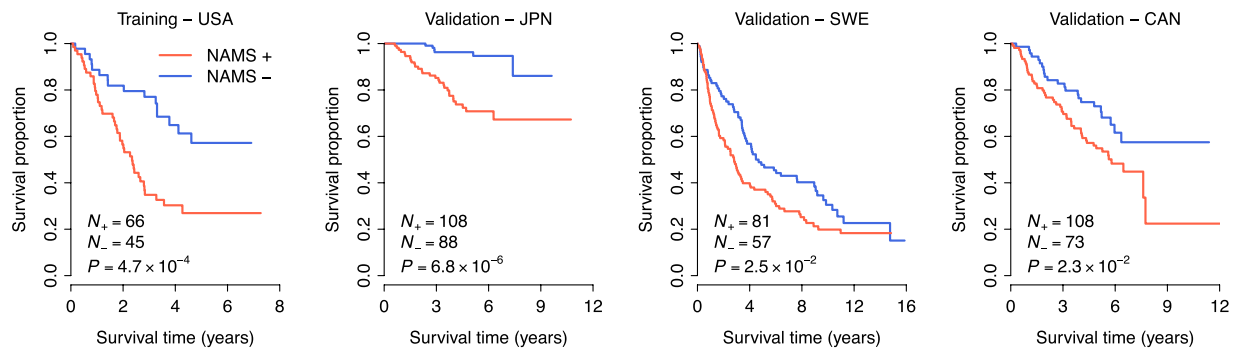


Figure 3. Kaplan-Meier curves of overall survival. The expression of the NAMS predicts poor overall survival in the discovery (USA) and validation (JPN, SWE, and CAN) cohorts. The red curves are for the NAMS positive patients while the blue curves are for the NAMS negative patients. The NAMS positive patients were defined as those having a death risk score greater than zero. P-values were calculated by log-rank tests for the differences in survival between the NAMS positive and negative groups.

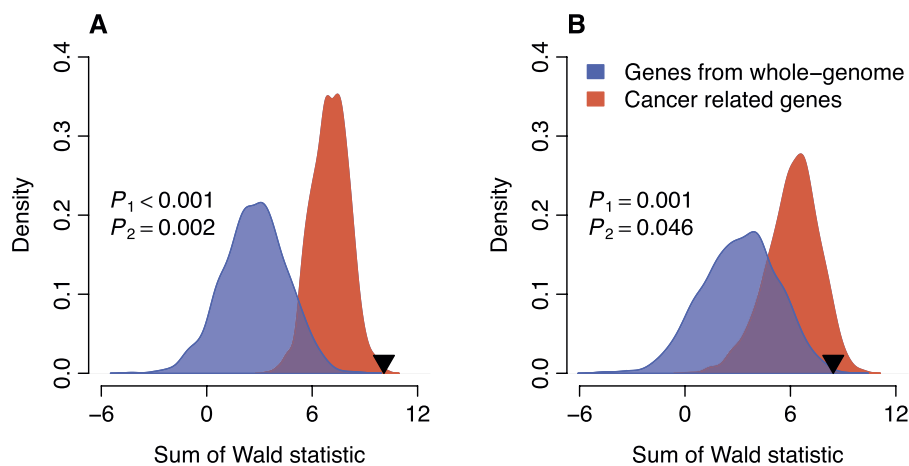


Figure 4. Non-random prognostic power of NAMS in lung cancer. The blue areas show the distributions of the sum of Wald statistic for the 1,000 resampled gene signatures picked up from whole human genome with identical size as NAMS. The red areas show the distributions of the sum of Wald statistic for the 1,000 resampled gene signatures picked up from the cancer related genes with identical size as NAMS. The black triangles stand for the sum of Wald statistic of NAMS. (A) Resampling pattern for recurrence-free survival; (B) Resampling pattern for overall survival. P_1 : the right-tailed P-values for the resampling test when the resampled gene signatures were picked up from whole human genome; P_2 : the right-tailed P-values for the resampling test when the resampled gene signatures were picked up from the cancer related genes. Figure S10. Kaplan-Meier curves of overall survival on lung cancer subtypes (SWE).

with a central role in carcinogenesis. More work on *NFE2L2* will be necessary to fully elucidate the mechanisms underlying this pattern. In addition to cancer prognosis, now well validated in the current study, *NFE2L2* holds promise for the management of multiple cancers. We confirm that the NAMS represents a promising prognostic biomarker in human lung cancer predicting recurrence-free survival and validation survival.

Methods

Gene expression profiling. We obtained the gene expression data in control and *NFE2L2* KD A549 lung cancer cells from the Gene Expression Omnibus (GEO) database (GSE38332)²¹. Seven independent microarray lung cancer datasets from Spain (GSE18842, ESP)²⁴, Taiwan (GSE19804, TWN)²⁵, Korea (GSE8894, KOR)²⁶, the United States (GSE3141, USA)²⁷, Japan (GSE31210, JPN)²⁸, Sweden (GSE37745, SWE)²⁹, and Canada (GSE50081, CAN)³⁰, were also downloaded from the GEO database (Supplementary Table S5). Expression data of paired normal and tumor tissues from lung cancer patients were available in the ESP and TWN cohorts. The information on recurrence-free survival was available in the KOR, JPN, SWE, and CAN cohorts. The information on overall survival was available for the USA, JPN, SWE, and CAN cohorts. The ESP and TWN cohorts were used to identify the differentially expressed genes

Cohort	Covariate	Recurrence-free survival			Overall survival		
		HR	95% CI	P-value	HR	95% CI	P-value
JPN	NAMS + vs. -	2.84	(1.58, 5.12)	5.1×10^{-4}	4.45	(1.76, 11.25)	1.6×10^{-3}
	Age (per year)	1.04	(1.01, 1.08)	2.3×10^{-2}	1.04	(0.99, 1.09)	1.5×10^{-1}
	Gender male vs. female	0.90	(0.45, 1.79)	7.6×10^{-1}	0.92	(0.36, 2.34)	8.6×10^{-1}
	Smoking + vs. -	1.02	(0.51, 2.03)	9.6×10^{-1}	1.06	(0.42, 2.69)	9.0×10^{-1}
	Stage (I and II)	2.43	(1.42, 4.14)	1.1×10^{-3}	2.79	(1.38, 5.64)	4.3×10^{-3}
	Gene alteration + vs. -	0.61	(0.37, 1.03)	6.4×10^{-2}	0.53	(0.27, 1.07)	7.5×10^{-2}
SWE	MYC level high vs. low	1.06	(0.42, 2.71)	9.0×10^{-1}	0.68	(0.16, 2.87)	6.0×10^{-1}
	NAMS + vs. -	2.72	(1.32, 5.61)	6.6×10^{-3}	1.43	(1.01, 2.04)	4.5×10^{-2}
	Age (per year)	1.00	(0.96, 1.03)	8.7×10^{-1}	1.03	(1.01, 1.05)	5.7×10^{-3}
	Gender male vs. female	0.80	(0.43, 1.48)	4.8×10^{-1}	0.98	(0.69, 1.38)	9.1×10^{-1}
	Stage (I-IV)	1.11	(0.75, 1.63)	6.1×10^{-1}	1.24	(1.02, 1.50)	3.3×10^{-2}
	CAN	NAMS + vs. -	2.60	(1.33, 5.07)	5.1×10^{-3}	1.65	(0.95, 2.85)
Age (per year)		1.00	(0.97, 1.03)	9.6×10^{-1}	1.01	(0.99, 1.04)	2.9×10^{-1}
Gender male vs. female		1.71	(0.94, 3.13)	8.1×10^{-2}	1.89	(1.11, 3.24)	2.0×10^{-2}
Smoking + vs. -		0.43	(0.21, 0.90)	2.6×10^{-2}	0.90	(0.42, 1.95)	7.9×10^{-1}
Stage (I and II)		1.90	(1.06, 3.39)	3.0×10^{-2}	1.93	(1.15, 3.23)	1.2×10^{-2}

Table 3. Multivariate Cox proportional hazards regression of survival in the validation cohorts. Note – HR: hazard ratio; CI: confidence interval.

between normal and tumor lung tissues. The KOR and USA cohorts were used for training purpose for recurrence-free and overall survival, respectively. The JPN, SWE, and CAN datasets were used as validation cohorts.

Microarray data analysis. We used the GCRMA algorithm in Bioconductor to normalize the expression level of each probe set for the microarray data of the control and *NFE2L2* KD lung cancer cells and of the paired normal and tumor tissues. Only the probe sets with unique annotations were included in this study. The genes on chromosomes X and Y were excluded to avoid the potentially confounding factor. The significance analysis of microarrays (SAM) algorithm⁵⁰, implemented in the samr library of the R Statistical Package, was used to compare log₂-transformed gene expression levels between the control and *NFE2L2* KD cells. FDR was controlled using the q-value method⁵¹. The probe sets with a fold-change >2 and FDR <0.01 were deemed differentially expressed between the control and *NFE2L2* KD cells. For the ESP and TWN cohorts, paired t-test was used to identify the genes differentially expressed between normal and tumor tissues. The P-values were adjusted by Benjamini-Hochberg correction. Only the probe sets with a fold-change >2 and adjusted P < 0.01 were deemed differentially expressed between normal and tumor tissues.

Risk score. For the two training cohorts (KOR and USA), univariate Cox proportional hazards regression was used to evaluate the association between recurrence-free survival/overall survival and gene expression. A recurrence-risk score and a death-risk score were then calculated for each patient, respectively, using a linear combination of gene expression weighted by the Wald statistic (ratio of regression coefficient to its standard error) as shown below:

$$S_R = \sum_{i=1}^n Z_{Ri}(e_i - \mu_i)/\tau_i \quad (1)$$

$$S_D = \sum_{i=1}^n Z_{Di}(e_i - \mu_i)/\tau_i \quad (2)$$

Here, S_R (Formula 1) and S_D (Formula 2) are the recurrence-risk and death-risk scores, respectively; n is the number of differentially expressed genes; Z_{Ri} denotes the Wald statistic of recurrence-free survival for the i^{th} gene, which was derived from the KOR cohort; Z_{Di} denotes the Wald statistic of overall survival for the i^{th} gene, which was derived from the USA cohort; e_i denotes the expression level of gene i ; and μ_i and τ_i are the mean and standard deviation of the gene expression values for gene i across all samples, respectively. A higher risk score implies a poor outcome.

References

1. Biswas, C. *et al.* Nuclear heme oxygenase-1 (HO-1) modulates subcellular distribution and activation of Nrf2, impacting metabolic and anti-oxidant defenses. *The Journal of biological chemistry* **289**, 26882–26894, doi: 10.1074/jbc.M114.567685 (2014).
2. Jaramillo, M. C. & Zhang, D. D. The emerging role of the Nrf2-Keap1 signaling pathway in cancer. *Genes & development* **27**, 2179–2191, doi: 10.1101/gad.225680.113 (2013).
3. Rotblat, B., Melino, G. & Knight, R. A. Nrf2 and p53: Januses in cancer? *Oncotarget* **3**, 1272–1283 (2012).
4. Sporn, M. B. & Liby, K. T. Nrf2 and cancer: the good, the bad and the importance of context. *Nature reviews. Cancer* **12**, 564–571, doi: 10.1038/nrc3278 (2012).
5. Dinkova-Kostova, A. T. *et al.* Direct evidence that sulfhydryl groups of Keap1 are the sensors regulating induction of phase 2 enzymes that protect against carcinogens and oxidants. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11908–11913, doi: 10.1073/pnas.172398899 (2002).
6. Levenon, A. L. *et al.* Cellular mechanisms of redox cell signalling: role of cysteine modification in controlling antioxidant defences in response to electrophilic lipid oxidation products. *The Biochemical journal* **378**, 373–382, doi: 10.1042/BJ20031049 (2004).
7. Jaiswal, A. K. Nrf2 signaling in coordinated activation of antioxidant gene expression. *Free radical biology & medicine* **36**, 1199–1207, doi: 10.1016/j.freeradbiomed.2004.02.074 (2004).
8. Nguyen, T., Nioi, P. & Pickett, C. B. The Nrf2-antioxidant response element signaling pathway and its activation by oxidative stress. *The Journal of biological chemistry* **284**, 13291–13295, doi: 10.1074/jbc.R900010200 (2009).
9. Venugopal, R. & Jaiswal, A. K. Nrf1 and Nrf2 positively and c-Fos and Fra1 negatively regulate the human antioxidant response element-mediated expression of NAD(P)H:quinone oxidoreductase1 gene. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 14960–14965 (1996).
10. Venugopal, R. & Jaiswal, A. K. Nrf2 and Nrf1 in association with Jun proteins regulate antioxidant response element-mediated expression and coordinated induction of genes encoding detoxifying enzymes. *Oncogene* **17**, 3145–3156, doi: 10.1038/sj.onc.1202237 (1998).
11. DeNicola, G. M. *et al.* Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. *Nature* **475**, 106–109, doi: 10.1038/nature10189 (2011).
12. Hayes, J. D. & McMahon, M. Nrf2 and KEAP1 mutations: permanent activation of an adaptive response in cancer. *Trends in biochemical sciences* **34**, 176–188, doi: 10.1016/j.tibs.2008.12.008 (2009).
13. Moon, E. J. & Giaccia, A. Dual roles of Nrf2 in tumor prevention and progression: Possible implications in cancer treatment. *Free radical biology & medicine*, doi: 10.1016/j.freeradbiomed.2014.11.009 (2014).
14. Padmanabhan, B. *et al.* Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Molecular cell* **21**, 689–700, doi: 10.1016/j.molcel.2006.01.013 (2006).
15. Ramos-Gomez, M. *et al.* Sensitivity to carcinogenesis is increased and chemoprotective efficacy of enzyme inducers is lost in nrf2 transcription factor-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 3410–3415, doi: 10.1073/pnas.051618798 (2001).
16. Thimmulappa, R. K. *et al.* Identification of Nrf2-regulated genes induced by the chemopreventive agent sulforaphane by oligonucleotide microarray. *Cancer research* **62**, 5196–5203 (2002).
17. Shelton, P. & Jaiswal, A. K. The transcription factor NF-E2-related factor 2 (Nrf2): a protooncogene? *FASEB journal: official publication of the Federation of American Societies for Experimental Biology* **27**, 414–423, doi: 10.1096/fj.12-217257 (2013).
18. Taguchi, K., Motohashi, H. & Yamamoto, M. Molecular mechanisms of the Keap1-Nrf2 pathway in stress response and cancer evolution. *Genes to cells: devoted to molecular & cellular mechanisms* **16**, 123–140, doi: 10.1111/j.1365-2443.2010.01473.x (2011).
19. Geismann, C., Arlt, A., Sebens, S. & Schafer, H. Cytoprotection “gone astray”: Nrf2 and its role in cancer. *OncoTargets and therapy* **7**, 1497–1518, doi: 10.2147/OTT.S36624 (2014).
20. Namani, A., Li, Y., Wang, X. J. & Tang, X. Modulation of Nrf2 signaling pathway by nuclear receptors: implications for cancer. *Biochimica et biophysica acta* **1843**, 1875–1885, doi: 10.1016/j.bbamcr.2014.05.003 (2014).
21. Mitsuishi, Y. *et al.* Nrf2 redirects glucose and glutamine into anabolic pathways in metabolic reprogramming. *Cancer cell* **22**, 66–79, doi: 10.1016/j.ccr.2012.05.016 (2012).
22. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research* **30**, 207–210 (2002).
23. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic acids research* **32**, D277–280, doi: 10.1093/nar/gkh063 (2004).
24. Sanchez-Palencia, A. *et al.* Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer. *International journal of cancer. Journal international du cancer* **129**, 355–364, doi: 10.1002/ijc.25704 (2011).
25. Lu, T. P. *et al.* Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **19**, 2590–2597, doi: 10.1158/1055-9965.EPI-10-0332 (2010).
26. Lee, E. S. *et al.* Prediction of recurrence-free survival in postoperative non-small cell lung cancer patients by using an integrated model of clinical information and gene expression. *Clinical cancer research: an official journal of the American Association for Cancer Research* **14**, 7397–7404, doi: 10.1158/1078-0432.CCR-07-4937 (2008).
27. Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357, doi: 10.1038/nature04296 (2006).
28. Yamauchi, M. *et al.* Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. *PLoS one* **7**, e43923, doi: 10.1371/journal.pone.0043923 (2012).
29. Botling, J. *et al.* Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clinical cancer research: an official journal of the American Association for Cancer Research* **19**, 194–204, doi: 10.1158/1078-0432.CCR-12-1139 (2013).
30. Der, S. D. *et al.* Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer* **9**, 59–64, doi: 10.1097/JTO.0000000000000042 (2014).
31. Venet, D., Dumont, J. E. & Detours, V. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS computational biology* **7**, e1002240, doi: 10.1371/journal.pcbi.1002240 (2011).
32. Xiang, M., Namani, A., Wu, S. & Wang, X. Nrf2: bane or blessing in cancer? *Journal of cancer research and clinical oncology* **140**, 1251–1259, doi: 10.1007/s00432-014-1627-1 (2014).
33. Lu, H., Ouyang, W. & Huang, C. Inflammation, a key event in cancer development. *Molecular cancer research: MCR* **4**, 221–233, doi: 10.1158/1541-7786.MCR-05-0261 (2006).
34. Singh, A. *et al.* Dysfunctional KEAP1-Nrf2 interaction in non-small-cell lung cancer. *PLoS medicine* **3**, e420, doi: 10.1371/journal.pmed.0030420 (2006).
35. Bauer, A. K., Hill, T., 3rd & Alexander, C. M. The involvement of Nrf2 in lung cancer. *Oxidative medicine and cellular longevity* **2013**, 746432, doi: 10.1155/2013/746432 (2013).

36. Jemal, A. *et al.* Global cancer statistics. *CA: a cancer journal for clinicians* **61**, 69–90, doi: 10.3322/caac.20107 (2011).
37. Roh, M. S. Molecular pathology of lung cancer: current status and future directions. *Tuberculosis and respiratory diseases* **77**, 49–54, doi: 10.4046/trd.2014.77.2.49 (2014).
38. Sonnenschein, C., Soto, A. M., Rangarajan, A. & Kulkarni, P. Competing views on cancer. *Journal of biosciences* **39**, 281–302 (2014).
39. Plomin, R., Haworth, C. M. & Davis, O. S. Common disorders are quantitative traits. *Nature reviews. Genetics* **10**, 872–878, doi: 10.1038/nrg2670 (2009).
40. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158, doi: 10.1038/nature05610 (2007).
41. McFarland, C. D., Mirny, L. A. & Korolev, K. S. Tug-of-war between driver and passenger mutations in cancer and other adaptive processes. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 15138–15143, doi: 10.1073/pnas.1404341111 (2014).
42. Fidler, I. J. The organ microenvironment and cancer metastasis. *Differentiation; research in biological diversity* **70**, 498–505, doi: 10.1046/j.1432-0436.2002.700904.x (2002).
43. Sleeman, J. P. The metastatic niche and stromal progression. *Cancer metastasis reviews* **31**, 429–440, doi: 10.1007/s10555-012-9373-9 (2012).
44. Soto, A. M. & Sonnenschein, C. The tissue organization field theory of cancer: a testable replacement for the somatic mutation theory. *BioEssays: news and reviews in molecular, cellular and developmental biology* **33**, 332–340, doi: 10.1002/bies.201100025 (2011).
45. Zhou, T., Wang, T. & Garcia, J. G. Expression of nicotinamide phosphoribosyltransferase-influenced genes predicts recurrence-free survival in lung and breast cancers. *Scientific reports* **4**, 6107, doi: 10.1038/srep06107 (2014).
46. Zhou, T., Wang, T. & Garcia, J. G. Genes influenced by the non-muscle isoform of Myosin light chain kinase impact human cancer prognosis. *PLoS one* **9**, e94325, doi: 10.1371/journal.pone.0094325 (2014).
47. Rahman, M. A. *et al.* RRM2 regulates Bcl-2 in head and neck and lung cancers: a potential target for cancer therapy. *Clinical cancer research: an official journal of the American Association for Cancer Research* **19**, 3416–3428, doi: 10.1158/1078-0432.CCR-13-0073 (2013).
48. Wang, L., Meng, L., Wang, X. W., Ma, G. Y. & Chen, J. H. Expression of RRM1 and RRM2 as a novel prognostic marker in advanced non-small cell lung cancer receiving chemotherapy. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine* **35**, 1899–1906, doi: 10.1007/s13277-013-1255-4 (2014).
49. Zhao, H., Zhang, H., Du, Y. & Gu, X. Prognostic significance of BRCA1, ERCC1, RRM1, and RRM2 in patients with advanced non-small cell lung cancer receiving chemotherapy. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine* **35**, 12679–12688, doi: 10.1007/s13277-014-2592-7 (2014).
50. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5116–5121, doi: 10.1073/pnas.091062498 (2001).
51. Taylor, J., Tibshirani, R. & Efron, B. The ‘miss rate’ for the analysis of gene expression data. *Biostatistics* **6**, 111–117, doi: 10.1093/biostatistics/kxh021 (2005).

Acknowledgements

This work was supported by Anhui provincial natural science research project of university (KJ2013A188 to ZQ and KJ2014A164 to JL). This work was also supported by National Institutes of Health Grants HL58064 (JGNG), HL91899 (JGNG and TW), and CA154377 (DDZ). This work is also supported by Veterans Administration Health System grant, 1IK2BX001477 (LH) and by the Parker B. Francis Fellowship (TW).

Author Contributions

Z.Q., T.Z. and T.W. conceived of the study. Z.Q., T.Z., J.L., F.F. and T.W. participated in the design of the study. Z.Q., T.Z., J.L. and F.F. collected the microarray data. Z.Q. and T.Z. processed the microarray data. Z.Q., T.Z., C.I.G. and T.W. performed the statistical analysis. Q.W., L.H., A.E.C., V.N., J.R.J., D.D.Z., J.G.N.G. and T.W. helped to interpret the results. Z.Q., T.Z., C.I.G., X.X., Q.W., L.H., J.L., F.F., A.E.C., V.N., J.R.J., D.D.Z., J.G.N.G. and T.W. drafted the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Qian, Z. *et al.* Nuclear factor, erythroid 2-like 2-associated molecular signature predicts lung cancer survival. *Sci. Rep.* **5**, 16889; doi: 10.1038/srep16889 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>