Role of FRG1 in predicting the overall survivability in cancers using multivariate based optimal model

Rehan Khan
School of Biological Sciences, National Institute of Science Education and Research, Bhubaneswar, HBNI, P.O. Jatni, Khurda 752050, Odisha https://orcid.org/0000-0003-4579-0360

Ananya Palo
School of Biological Sciences, National Institute of Science Education and Research, Bhubaneswar, HBNI, P.O. Jatni, Khurda 752050, Odisha https://orcid.org/0000-0001-5276-2742

Manjusha Dixit (manjusha@niser.ac.in)
School of Biological Sciences, National Institute of Science Education and Research, Bhubaneswar, HBNI, P.O. Jatni, Khurda 752050, Odisha https://orcid.org/0000-0002-6997-1054

Research Article

Keywords: expression, survival, effect, difference, genes, cancer, multigene, role, analysis

Posted Date: July 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-745024/v1

Version of Record: A version of this preprint was published at Scientific Reports on November 18th, 2021. See the published version at https://doi.org/10.1038/s41598-021-01665-w.

Abstract

FRG1 has a role in tumorigenesis and angiogenesis. Our preliminary analysis showed FRG1 expression is associated with the overall survival (OS) in cancers the effect varies. In cervix and gastric cancers, we found a clear difference in the OS between the low and high FRG1 expression groups, but in breast, lung, and liver cancers the difference was not prominent. We hypothesized that the functionality of the genes correlated with FRG1 could be getting affected by FRG1 or vice versa, which might mask the effect of a single gene on the OS analysis in cancer patients.

We used the multivariate Cox regression, risk score, and Kaplan Meier analyses to determine OS in a multigene model. STRING, Cytoscape, and HIPPIE were used to deduce FRG1 associated pathways.

In breast, lung, and liver cancer we found a distinct difference in the OS, between the low and high FRG1 expression groups in the multigene model, suggesting an independent role of FRG1 in survival. Risk scores were calculated based upon regression coefficients in the multigene model. Low and high-risk score groups revealed a significant difference in the FRG1 expression and survival. HPF1, RPL34, and EXOSC9 were the most common genes present in FRG1 associated pathways across the cancer types. Validation of the effect of FRG1 expression on these genes by qRT-PCR, supports that FRG1 might be an upstream regulator of their expression. These genes may have multiple regulators which also affect their expression, leading to the masking effect in the survival analysis. In conclusion, our study highlights the role of FRG1 in the survivability of cancer patients in tissue-specific manner and the use of multigene models in prognosis.

Introduction

FSHD Region Gene 1 (FRG1) gene is present on human chromosome 4q35. Being the primary candidate gene of FSHD, studies related to FRG1 primarily focused on muscles [1]. While the exact function of FRG1 is yet to be deciphered, various studies have indicated its role in mRNA splicing [2]. The biochemical activity analysis of human FRG1 revealed RNA binding and actin-binding properties which have direct implications in RNA biogenesis, transport, and cytoplasmic localization [3]. The recent 3D cryo-EM structure of the human spliceosomal C complex has shown that the FRG1 gene is part of spliceosome machinery and it can have multiple prospects on gene expression regulation [4]. The first clue regarding the role of FRG1 in angiogenesis or tumorigenesis came from a study in *X. laevis* where an increase in branching and size of vasculature was observed by overexpressing FRG1 [5]. Our research group for the first time showed the reduced expression of FRG1 in cancer tissues. FRG1 affected the proliferation, migration, invasion, and angiogenic potential of cancer cell lines and the expression of G-CSF and MMP10 [6]. Reduced FRG1 expression in AR negative prostate cancer cell lines increased invasiveness and cell migratory properties [7].

Although our previous study showed that FRG1 affects EMT, yet its role in survival is not clear. Our preliminary analysis did not indicate the robust effect of FRG1 on overall cancer survival in all cancer

type. It is possible that FRG1 in conjunction with other genes affects the survival of cancer patients. Alternatively, other genes which are also altered in cancers conceal the analysis of the effect of FRG1 on the OS.

Based on this hypothesis, we first determined the genes positively correlated with FRG1 using multiple databases in different cancer types. Cox regression analysis was performed to come up with the model which predicts cancer survival significantly. Later we used these genes to determine the pathways in which the FRG1 is involved. Common genes which were part of the FRG1 related pathway among different cancers, were experimentally validated. Our study shows the importance of the use of multigene models in survival prediction.

Material And Method

Workflow

We chose the top seven cancers with the highest incidence based on Global Cancer Observatory data [8]. Co-expression dataset of FRG1 with the top 20 most correlated genes was obtained from cBioPortal [9, 10]. mRNA expression and clinical datasets for all the patients in each cancer type were downloaded from Genomic Data Commons (GDC) Data Portal [11]. Figure 1. shows the workflow for the entire analysis. Kaplan Meier survival analysis was performed in each cancer type to observe the effect of FRG1 expression on the survivability of patients [12]. Stratified multivariate cox regression [13] was performed to determine the association between overall patient survival and gene expression levels of FRG1 along with the 20 correlated genes (top correlated genes based on the spearman's correlation (r_s)) in all the cancer types. The model was optimized by removing the least correlated genes sequentially till the FRG1 was significant. The risk score was calculated for each patient, and then the patients were divided into low and high-risk groups based on the median risk score [14, 15]. Kaplan-Meier plots were created to identify the difference between the low and high-risk groups in different cancer types. Box plots were created to represent the low and high expression of FRG1 in both risk types.

Using STRING [16] and HIPPIE [17] web tools, we developed a network of FRG1 and the top correlated genes, to find the known interactions at various levels. From the cancer type-specific

pathways, a common pathway was identified. Effect of FRG1 expression on the common genes, present across the different cancer types in multigene survival prediction models and from the pathways, was validated via RT-PCR.

Data sources and processing

Co-expression data of FRG1 was obtained from cBioPortal (accessed on 20 Aug 2020) using TCGA, Firehose Legacy dataset by comparing RNA Seq V2 RSEM for all cancer types.

For survival analysis, data of expression profiles along with clinical data was downloaded from GDC Data Portal (up to 19 Dec 2020) for all the cancer types. Settings chosen to download the data from GDC TCGA were as follows; Data Category- Transcriptome Profiling, Data type- Gene Expression Quantification, Experimental Strategy- RNA-Seq, Primary site- Cancer Type, Program- TCGA, Workflow Types- Htseq-FPKM-UQ.

Overall survival analysis for single gene

To test the effect of FRG1 expression on the survivability of patients in our chosen cancer types, we performed the Kaplan Meir survival analysis using the TCGA data downloaded from GDC. We performed the analysis in R using the "survival" and "survminer" libraries. To determine the optimal cut-off point of FRG1 expression for KM plot we used the res.cut() function and plotted the final data using ggsurvplot() (Supplementary Fig. S1).

Survival analysis and identification of prognostic/signature genes

We analyzed the correlation between OS time and gene expression by using stratified multivariate Cox regression. The model was optimized by removing the least correlated genes until the FRG1 remained significant. A risk score was calculated for each patient based on the following equation,

$$Risk\ score = \sum_{i=1}^{n} exp_i \beta_i$$

where n was the number of prognostic genes, exp_i the expression value of gene i, and β_i the regression coefficient of gene i in the Cox regression analysis. Using the median risk score as a cutoff value, patients were classified into high- and low-risk groups. Box plots were generated to compare the FRG1 expression level between the low and high-risk groups. The Log-Rank test was used to determine the statistical significance of the difference in OS between the two groups.

Pathway analysis

Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is a biological database and web resource of known and predicted protein-protein interactions (PPI). For each cancer type, a model was created using STRING PPI network data. In the model, each node is represented by a protein and edges show physical interaction between the two proteins. The missing links between FRG1 and co-expressed genes were found using Human Integrated Protein-Protein Interaction rEference (HIPPIE), which is based on the earlier reports of FRG1 interacting proteins. Cytoscape (Version: 3.8.2) was used to visualize the network and to find the intersection of all the pathways using the merge tool, to get the final most common pathway across all cancer types [18].

qRT-PCR

HEK293 cells were transfected with the pLKO.1-FRG1sh vector (Sigma, USA) for FRG1 knockdown and with the pLKO.1-scrambled sequence vector to get the corresponding control. Total RNA was extracted using the RNeasy Mini kit (Qiagen, Germany). RNA was quantified using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA). RNA was converted into cDNA using oligo dT primer and random hexamer (in 1:3 ratio) (Verso cDNA Synthesis Kit, Thermo Scientific, USA). RT-PCR primers were designed for HPF1, RPL34, and EXOSC9 using Primer-BLAST [19] (Supplementary Table S1). Fast Start Universal SYBR Green Master Mix (Roche, Switzerland) was used to perform qRT-PCR in QuantStudio™ 3 Real-Time PCR System. The experiment was performed in triplicate for each sample and GAPDH was used as an internal control.

Statistics

For multi-gene model based overall survival analysis multivariate Cox regression analysis was performed using SPSS (version 26) [20]. Risk score (generated via multigene model) based OS analysis was performed using Kaplan Meir analysis in SPSS. A log-rank test was used to find the statistical significance of the difference in survival between the groups. The prognostic value of the risk score was measured using a time-dependent receiver operating characteristic (ROC) curve in SPSS. Mean values were compared using Student's t-test (two-tailed, unpaired). For all the tests performed, a p-value ≤ 0.05 was considered significant.

Results

Effect of FRG1 alone on survival in different cancer types

Kaplan-Meir survival analysis was performed, to determine the effect of FRG1 expression on the OS across the seven most frequent cancer types. In cervix, stomach and prostate cancers there was a highly significant difference in the survival probability between high and low FRG1 expression groups (Fig. 2). In liver cancer the difference in survivability was marginally significant. In breast, lung, and colorectal cancers although the trend was there yet, the difference was not significant. Overall, this data suggests that FRG1 affects the survival in cancers but the extent of the effect is tissue specific. Analysis of FRG1 expression alone may not be enough to explain the contribution of other genes, which are affected by FRG1 directly or indirectly. Therefore, we did multigene model-based analysis in breast, lung, colorectal and liver cancer to get a clear idea about the effect of FRG1 expression on OS.

High FRG1 expression is associated with a good prognosis in the multigene model

To determine the contribution of FRG1 on survival, the effect of other genes correlated with FRG1, was neutralized using the multivariate Cox regression model.

Effect of FRG1 and correlated genes on survival in breast cancer

In breast carcinoma, initially we entered the top 20 genes (Supplementary Table S2) ($r_s \geq 0.353$) correlated with FRG1 to generate the multivariate cox regression model in the TCGA-BRCA dataset. Sequentially, lowest correlated genes were removed from the model till the FRG1 showed a maximum level of significant association with survival (Table 1). The hazard ratio of FRG1 was 0.133 (95% CI 0.029–0.599, p = 0.009) for breast cancer patient's death.

In order to analyze the combined effect of FRG1 and the correlated genes (genes present in the final model) on the OS, for each breast cancer patient risk score was calculated. The patients were stratified into low-risk (n = 612) and high-risk (n = 611) groups based on the median risk score value. A significant difference (p = 2.45E-13) was observed between the groups in OS (Fig. 3A). The AUC for this risk model was 0.645 (Supplementary Fig. S2). There was significantly higher (p = 0.0001) FRG1 expression in the low-risk group compared to the high-risk group (Fig. 3B).

Table 1
Covariates present in multivariate Cox regression model in breast cancer patients.

Genes	B	Sig.	Exp(B), 95.0% CI for Exp(B)
HPF1	1.233	0.134	3.433 (0.683,17.259)
ING2	-1.865	0.003	0.155 (0.045,0.535)
UFSP2	2.672	0	14.467 (3.49,59.965)
PFDN5	-1.501	0.089	0.223 (0.04,1.256)
EXOSC9	-1.376	0.08	0.253 (0.054,1.177)
SARNP	0.939	0.049	2.557 (1.005,6.504)
SRP19	-0.34	0.688	0.712 (0.135,3.743)
RPS3A	-0.76	0.39	0.468 (0.083,2.647)
NDUFC1	-0.472	0.514	0.624 (0.151,2.577)
NACA	0.959	0.302	2.609 (0.422,16.14)
RWDD4	-0.247	0.774	0.781 (0.145,4.201)
NSA2	-0.578	0.389	0.561 (0.15,2.092)
TBCA	1.132	0.136	3.102 (0.7,13.75)
MRPS18C	0.698	0.401	2.01 (0.395,10.239)
TRIM56	-0.226	0.663	0.797 (0.288,2.205)
TTC1	-0.236	0.792	0.79 (0.137,4.549)
PLRG1	0.758	0.368	2.134 (0.409,11.134)
MRPL1	0.536	0.498	1.71 (0.362,8.069)
RPL34	0.588	0.464	1.8 (0.373,8.688)
FRG1	-2.021	0.009	0.133 (0.029,0.599)
age_months	0.003	0	1.003 (1.002,1.004)

Effect of FRG1 and correlated genes on survival in Lung cancer

The top 20 genes (Supplementary Table S2) were added ($r_s \geq 0.535$) with FGR1 to generate the multivariate cox regression model using TCGA-MESO, TCGA-LUAD and TCGA-LUSC datasets. To investigate the prognostic effect of FRG1 on lung carcinoma patients, we applied the same strategy as described above. The final model had 17 genes where the hazard ratio of FRG1 was 0.235 (95% CI 0.074–0.742, p = 0.014) for lung cancer patient's death (Table 2).

All the patients were stratified into low-risk (n = 559) and high-risk (n = 572) groups based on the median value of the risk score. The AUC for this risk model was 0.569 (Supplementary Fig. S2). A significant difference (p = 1.0E-6) in OS was observed between the groups (Fig. 4A). There was significantly (p = 0.0007) high FRG1 expression in the low-risk group compared to the high-risk group (Fig. 4B).

Table 2
Covariates present in multivariate Cox regression model in lung cancer patients.

Genes	B	Sig.	Exp(B), 95.0% CI for Exp(B)
HPF1	0.801	0.09	2.227 (0.882,5.621)
MRPS18C	0.449	0.373	1.566 (0.584,4.2)
ANAPC10	-1.668	0.008	0.189 (0.055,0.652)
LSM6	0.725	0.149	2.064 (0.771,5.525)
ATP5PO	-0.004	0.993	0.996 (0.425,2.338)
UBE2B	0.018	0.966	1.019 (0.432,2.403)
THOC7	-0.65	0.229	0.522 (0.181,1.504)
NDUFS4	0.347	0.362	1.416 (0.671,2.988)
RWDD4	0.331	0.53	1.393 (0.495,3.917)
COX7B	-0.295	0.477	0.744 (0.33,1.681)
GSTO1	0.368	0.191	1.444 (0.833,2.505)
RPL24	0.143	0.691	1.154 (0.57,2.335)
UBE2D3	0.951	0.113	2.587 (0.799,8.377)
UXT	-0.246	0.617	0.782 (0.299,2.048)
LSM3	0.791	0.124	2.206 (0.805,6.041)
RPL34	-0.508	0.157	0.602 (0.298,1.216)
SHANK2	0.221	0.048	1.247 (1.002,1.552)
FRG1	-1.448	0.014	0.235 (0.074,0.742)
age (Months)	0.001	0.025	1.001 (1,1.002)

FRG1 and correlated genes do not predict survival in the colorectal cancer

To investigate the prognostic effect of FRG1 using colorectal cancer TCGA-READ and TCGA-COAD datasets, the top 20 correlated genes (Supplementary Table S2) were added ($r_s \geq 0.964$) in the multivariate cox regression model (Supplementary Table S3). Models with all the 20 genes, as well as any

other combination of genes didn't show significant effect of FRG1 on the OS of the colorectal cancer patients.

Effect of FRG1 and correlated genes on survival in liver cancer

The top 20 genes correlated (Supplementary Table S2) with FRG1, with r_s cutoff ≥ 0.539, were used to generate the multivariate cox regression model using TCGA-LIHC and TCGA-CHOL datasets. The final model had 16 genes (Table 3) where the hazard ratio of FRG1 was 0.18 (95% CI 0.034–0.948, p = 0.043) for liver cancer patient's death.

Next, to determine the effect in the multigene model, the patients were divided into the low-risk group (n = 231) and high-risk group (n = 231) based on the median risk score. The AUC for this risk model was 0.616 (S2 Fig.). A significant (p = 0.0001) difference in OS was observed between the two groups (Fig. 5A). Comparison of FRG1 expression between the high-risk group and low-risk group (Fig. 5B) showed significantly (p < 0.0001) higher expression in the low-risk group.

Table 3
Covariates present in multivariate Cox regression model in liver cancer patients.

Genes	B	Sig.	Exp(B), 95.0% CI for Exp(B)
HPF1	0.791	0.238	2.206 (0.593,8.209)
POMP	0.885	0.174	2.424 (0.676,8.691)
UXT	-0.222	0.772	0.801 (0.179,3.591)
RREB1	-1.001	0.089	0.367 (0.116,1.164)
LMTK2	-0.551	0.289	0.576 (0.208,1.596)
NDUFC1	-1.594	0.009	0.203 (0.061,0.676)
EP300	0.786	0.335	2.195 (0.445,10.844)
NCOA2	-0.613	0.146	0.542 (0.237,1.239)
MRPL54	-1.53	0.009	0.217 (0.069,0.683)
KMT2C	-0.615	0.259	0.541 (0.186,1.574)
PRR14L	-0.073	0.914	0.93 (0.251,3.449)
UFSP2	1.583	0.022	4.871 (1.254,18.923)
HACD2	0.919	0.033	2.507 (1.075,5.846)
CELF1	0.741	0.421	2.097 (0.346,12.73)
NCOA6	-0.044	0.954	0.957 (0.216,4.245)
NDUFS5	1.581	0.004	4.86 (1.681,14.049)
FRG1	-1.716	0.043	0.18 (0.034,0.948)
age (Months)	0.001	0.055	1.001 (1,1.002)

FRG1 knockdown in HEK293T reduces expression of HPF1, RPL34 and EXOSC9

From the top 20 genes correlated with FRG1 across cancer types, we found that three genes (HPF1, RPL34 and EXOSC9) were common. We hypothesized that these genes could be part of pathway/pathways in which FRG1 has a role and could affect their expression. To validate this, the expression level of these three genes was analyzed in response to FRG1 depletion in the HEK293 cell line by quantitative real-time PCR. We observed that knockdown of FRG1 led to a significant decrease in expression of HPF1 (0.68-fold, p-value = 0.011), RPL34 (0.65-fold, p-value = 0.025) and EXOSC9 (0.54-fold, p-value = 0.012) (Fig. 6). These findings confirm the effect of FRG1 in transcriptional regulation of HPF1, RPL34, and EXOSC9, which could be direct or indirect.

Frg1 May Have Role In Multiple Pathways

To figure out the pathway/s where FRG1 may have a role, we used genes that show correlation with FRG1 expression and the genes, which interact with FRG1 (HIPPIE database) as input in the STRING database. Individual networks for each cancer type are shown in Fig. 7. Thereafter all the networks were merged and the intersection was obtained using the Merge tool of Cytoscape, giving us the most common pathway (Fig. 7). The merged pathway had 17 nodes (MEPCE, LARP7, SUMO2, UBE2O, HECW2, RBPMS, JUN, ESR2, SART3, EXOSC8, FRG1, PARP2, C4orf27 (HPF1), EFTUD2, SNRPD3, CWC22 and AQR) and 21 edges. All the networks formed are statistically significant with protein-protein interaction (PPI) enrichment p-value < 0.05.

Discussion

FRG1 protein is part of human spliceosomal complex C [21]. Earlier studies primarily focused on the role of FRG1 in facioscapulohumeral muscular dystrophy (FSHD). However, recent studies have demonstrated the role of FRG1 in tooth germ development and angiogenesis [6, 22]. Our group has revealed its role in tumor development [7]. Being part of the spliceosomal C complex, FRG1's downregulation might lead to instability and disruption in downstream processes affecting the normal mRNA levels. In concordance, recent studies have shown that the expression of splicing factors is frequently deregulated in different cancer types [23].

We have elucidated the role of FRG1 in the OS across cancer types. High FRG1 expression correlated with better survival in the cervix and gastric cancer patients. In cancer types such as breast, lung, liver, difference in FRG1 expression level did not affect OS significantly. We observed that the patients with low FRG1 expression were more frequent in cervix and gastric cancers. On the contrary, in liver, colorectal, lung, and prostate cancers just the opposite trend was observed. In breast cancer distribution was approximately equal. Expression of genes can correlate if one of them regulates the transcription of another, directly or indirectly. Upstream regulator genes may have mutation/s resulting in the masking of independent effects of mutation/s in the downstream target, in cancers. We used multi-gene models to nullify the effect of other genes on OS that correlate with FRG1. As expected, we observed a clear effect of FRG1 expression in breast, lung and liver cancers also after multivariate cox regression analysis.

Segregation of the patients based on the risk score (calculated based on the multigene model) showed that low-risk patients had better OS than high-risk patients. We also observed that low-risk patients had high FRG1 levels, which confirms the role of FRG1 expression in survival. Many previous studies support our observation directly, where increased FRG1 expression affected in-vitro cell migration, invasion, and angiogenesis inversely [6].

To further elucidate the molecular mechanism of the role of FRG1 in cells, we generated pathways. We came up with a final model (Fig. 8) which shows four types of functions where FRG1 might be involved namely, pre-mRNA processing (CWC22), mitochondrial functioning (MRPS18C, MRPL1, MRPL54, and NDUFC1), ribosomal functioning (RPL34, RPL24), and in DNA damage/repair pathway (HPF1, PARP1,

SUMO2). FRG1 with CWC22 interact [24] and they both are also part of the spliceosomal C complex [21]. Deregulation of these genes may have a direct effect on spliceosome complex functioning. Previous literature has shown the importance of CWC22 in pre-mRNA splicing [25]. CWC22 expression levels were found to be associated with colon cancer and its silencing led to increased p53 levels [26, 27]. SNRPD3 is also part of the spliceosome complex [28]. It has been found to have a regulatory effect on p53 expression in non-small cell lung cancer. It also has a role in triple-negative breast cancer cell proliferation [26, 29]. In our model we found EXOC9 to be highly correlated with FRG1 in multiple cancer types. Protein-protein interaction between FRG1 and EXOSC8 has been observed in previous studies [30]. EXOSC8 and EXOSC9 (both present in our model) are non-catalytic parts of the RNA exosome complex [31]. EXOSC8 and EXOSC9 are associated with many diseases [32, 33], but their role in cancer has recently been uncovered. EXOSC8 was found to be promoting tumor and cancer cell growth in colorectal carcinoma [34]. Reduction in EXOSC9 was found to be associated with reduction of p-body formation in cancer cells [35]. From all these studies, we can infer that FRG1 along with EXOSC8 and EXOSC9 might play a major role in controlling RNA processing and its depletion can affect functional RNAs.

Another very interesting observation was the mitochondria related genes in our model. Mitochondrial ribosomal proteins (MRPS18C, MRPL1 and MRPL54) and NDUFC1, which is a component of the C1 complex, are related to FRG1. All these genes play a major role in mitochondrial functioning. MRPS18C is downregulated in esophageal cancer [36]. Recent studies have shown that MRPL1 is a part of the gene signature for low-grade gliomas prognosis [37]. In malignant mesothelioma (MM) and lung cancer MRPL1 was found to be mutated [38]. Similarly, in HCC, high expression of MRPL54 was associated with better survival [39]. NDUFC1 may affect the production of ROS which has been observed in many cancer types [40]. Similarly, RPL24 and RPL34 that are part of the cytoplasmic ribosomal complex can affect protein production. Alteration in RPL34 expression affects non-small cell lung cancer cell proliferation [41, 42]. Depletion of RPL24 inhibits cacner cell growth, which makes RPL24 a potential therapeutic target [43]. Another gene in our model, RBPMS interacts with FRG1 at protein level [44]. RBPMS has been shown as a coactivator of transcriptional activity of many genes [45]. Multiple myeloma shows drug resistance when RBPMS is silenced [46]. These observations suggest that FRG1 might control the protein synthesis as well.

FRG1 is also related to the DNA repair pathway (HPF1, PARP1, and SUMO2). HPF1 protects the DNA from damage by limiting hyper auto modification of PARP1 which is required for repair [47]. PARP2 shows a direct protein-protein interaction with FRG1 but its function is unknown. SUMO2, which plays an important role in post-translational modification and affects multiple cellular processes including DNA repair and replication [48, 49], has also been implicated in cancers [50, 51]. FRG1 may affect DNA repair by acting as a transcriptional regulator of these genes.

Overall, our analysis indicates two possibilities of FRG1's role, first being a part of the spliceosome complex and the other is by acting as a transcriptional regulator of other genes which are involved in various functions. To check the latter possibility, we performed qRT-PCR and found that FRG1 knockdown led to a reduction in expression levels of HPF1, EXOSC9, and RPL34. Further in-depth experiments are

needed to figure out the exact role of FRG1 in tumorigenesis via the first possibility. This study has additional limitations such as more number genes can be incorporated in our model. We chose the top seven cancer types with highest incident rates, studies in other cancers can give a more in-depth understanding of FRG1 pathway.

In conclusion, this study has clearly shown the role of FRG1 in predicting the survivability of cancer patients. The higher expression of the FRG1 gene has a protective effect. The use of the multi genes models can be useful in elucidating the effect of specific gene in a biologically complex background.

Declarations

Funding: This work was supported by intramural funding from the National Institute of

Science Education and Research (NISER), Department of Atomic Energy (DAE),

Government of India (GOI). RK and AP received fellowship from NISER, DAE, GOI.

Disclosure of Potential Conflicts of Interest

The authors declare no potential conflicts of interest

References

1. Grewal, P. K. *et al.* FRG1, a gene in the FSH muscular dystrophy region on human chromosome 4q35, is highly conserved in vertebrates and invertebrates., **216** (1), 13–19 (1998).
2. Sancisi, V. *et al.* Altered Tnnt3 characterizes selective weakness of fast fibers in mice overexpressing FSHD region gene 1 (FRG1). *Am J Physiol Regul Integr Comp Physiol*, **306** (2), 124–137 (2014).
3. Sun, C. Y. *et al.* Facioscapulohumeral muscular dystrophy region gene 1 is a dynamic RNA-associated and actin-bundling protein. *J Mol Biol*, **411** (2), 397–416 (2011).
4. Ratzke, C., Barrere, J. & Gore, J. Strength of species interactions determines biodiversity and stability in microbial communities. *Nature Ecology & Evolution*, **4** (3), 376–383 (2020).
5. Wuebbles, R. D., Hanel, M. L. & Jones, P. L. FSHD region gene 1 (FRG1) is crucial for angiogenesis linking FRG1 to facioscapulohumeral muscular dystrophy-associated vasculopathy. *Dis Model Mech*, **2** (5–6), 267–274 (2009).
6. Tiwari, A. *et al. Increased FSHD region gene1 expression reduces in vitro cell migration, invasion, and angiogenesis, ex vivo supported by reduced expression in tumors.*Bioscience Reports, 2017. **37**(5).
7. Tiwari, A. *et al.* Reduced FRG1 expression promotes prostate cancer progression and affects prostate cancer cell migration and invasion. *BMC Cancer*, **19** (1), 346 (2019).
8. *Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available from*: https://gco.iarc.fr/*today, accessed on 10 Aug* 2020

9. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*, **6** (269), 1 (2013).
10. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*, **2** (5), 401–404 (2012).
11. Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med*, **375** (12), 1109–1112 (2016).
12. Rich, J. T. *et al. A practical guide to understanding Kaplan-Meier curves.* Otolaryngology–head and neck surgery: official journal of American Academy of Otolaryngology-Head and Neck Surgery, 2010. **143**(3): p. 331–336.
13. Cox, D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34** (2), 187–220 (1972).
14. Kim, S. K. *et al.* A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. *Mol Oncol*, **8** (8), 1653–1666 (2014).
15. Liu, Q. *et al.* Risk score based on three mRNA expression predicts the survival of bladder cancer. *Oncotarget*, **8** (37), 61583–61591 (2017).
16. Jensen, L. J. *et al.* STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*, **37** (Database issue), 412–416 (2009).
17. Alanis-Lobato, G., Andrade-Navarro, M. A. & Schaefer, M. H. HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res*, **45** (D1), 408–414 (2017).
18. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, **13** (11), 2498–2504 (2003).
19. Ye, J. *et al.* Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, **13** (1), 134 (2012).
20. *IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk,NY: IBM Corp*.
21. Bertram, K. *et al.* Structural Insights into the Roles of Metazoan-Specific Splicing Factors in the Human Step 1 Spliceosome. *Mol. Cell*, **80** (1), 127–1396 (2020).
22. Hasegawa, K. *et al.* Facioscapulohumeral muscular dystrophy (FSHD) region gene 1 (FRG1) expression and possible function in mouse tooth germ development. *J Mol Histol*, **47** (4), 375–387 (2016).
23. Cerasuolo, A. *et al.* The Role of RNA Splicing Factors in Cancer: Regulation of Viral and Human Gene Expression in Human Papillomavirus-Related Cervical Cancer. *Frontiers in cell and developmental biology*, **8**, 474–474 (2020).
24. Hegele, A. *et al.* Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell*, **45** (4), 567–580 (2012).
25. Zhou, Z. *et al.* Comprehensive proteomic analysis of the human spliceosome. *Nature*, **419** (6903), 182–185 (2002).

26. Siebring-van Olst, E. *et al.* A genome-wide siRNA screen for regulators of tumor suppressor p53 activity in human non-small cell lung cancer cells identifies components of the RNA splicing machinery as targets for anticancer treatment. *Mol Oncol*, **11** (5), 534–551 (2017).
27. Steckelberg, A. L. *et al.* CWC22 Connects Pre-mRNA Splicing and Exon Junction Complex Assembly. *Cell Reports*, **2** (3), 454–461 (2012).
28. Will, C. L. *et al.* The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome., **10** (6), 929–941 (2004).
29. Koedoot, E. *et al.* Splicing factors control triple-negative breast cancer cell mitosis through SUN2 interaction and sororin intron retention. *Journal of Experimental & Clinical Cancer Research*, **40** (1), 82 (2021).
30. Rolland, T. *et al.* A proteome-scale map of the human interactome network., **159** (5), 1212–1226 (2014).
31. Kilchert, C., Wittmann, S. & Vasiljeva, L. The regulation and functions of the nuclear RNA exosome complex. *Nat Rev Mol Cell Biol*, **17** (4), 227–239 (2016).
32. Boczonadi, V. *et al.* EXOSC8 mutations alter mRNA metabolism and cause hypomyelination with spinal muscular atrophy and cerebellar hypoplasia. *Nat Commun*, **5**, 4287 (2014).
33. Burns, D. T. *et al.* Variants in EXOSC9 Disrupt the RNA Exosome and Result in Cerebellar Atrophy with Spinal Motor Neuronopathy. *The American Journal of Human Genetics*, **102** (5), 858–873 (2018).
34. Cui, K. *et al.* Comprehensive characterization of the rRNA metabolism-related genes in human cancer., **39** (4), 786–800 (2020).
35. Yoshino, S. *et al.* EXOSC9 depletion attenuates P-body formation, stress resistance, and tumorigenicity of cancer cells. *Sci. Rep*, **10** (1), 9275 (2020).
36. Gopisetty, G. & Thangarajan, R. Mammalian mitochondrial ribosomal small subunit (MRPS) genes: A putative role in human disease., **589** (1), 27–35 (2016).
37. Islam, T. *et al.* Drug repositioning and biomarkers in low-grade glioma via bioinformatics approach. *Informatics in Medicine Unlocked*, **17**, 100250 (2019).
38. Mäki-Nevala, S. *et al.* Driver Gene and Novel Mutations in Asbestos-Exposed Lung Adenocarcinoma and Malignant Mesothelioma Detected by Exome Sequencing., **194** (1), 125–135 (2016).
39. Huang, Y. *et al.* A Novel RNA Binding Protein-Related Prognostic Signature for Hepatocellular Carcinoma. *Frontiers in oncology*, **10**, 580513–580513 (2020).
40. Lin, L. L. *et al.* Downregulation of c-Myc is involved in TLR3-mediated tumor death of neuroblastoma xenografts. *Lab Invest*, **96** (7), 719–730 (2016).
41. Liu, H. *et al.* RNAi-mediated RPL34 knockdown suppresses the growth of human gastric cancer cells. *Oncol Rep*, **34** (5), 2267–2272 (2015).
42. Yang, S. *et al. Over-expressed RPL34 promotes malignant proliferation of non-small cell lung cancer cells.* Gene, 2016. **576**(1, Part 3): p. 421–428.

43. Wilson-Edell, K. A. *et al.* RPL24: a potential therapeutic target whose depletion or acetylation inhibits polysome assembly and cancer cell growth. *Oncotarget*, **5** (13), 5165–5176 (2014).
44. Yu, H. *et al.* Next-generation sequencing to generate interactome datasets. *Nat Methods*, **8** (6), 478–480 (2011).
45. Sun, Y. *et al. Potentiation of Smad-mediated transcriptional activation by the RNA-binding protein RBPMS*. 2006. **34**: p.6314–6326.
46. Rastgoo, N. *et al.* Dysregulation of EZH2/miR-138 axis contributes to drug resistance in multiple myeloma by downregulating RBPMS., **32** (11), 2471–2482 (2018).
47. Gibbs-Seymour, I. *et al.* HPF1/C4orf27 Is a PARP-1-Interacting Protein that Regulates PARP-1 ADP-Ribosylation Activity. *Mol. Cell*, **62** (3), 432–442 (2016).
48. Golebiowski, F. *et al.* System-wide changes to SUMO modifications in response to heat shock. *Sci Signal*, **2** (72), 24 (2009).
49. Han, Z. J. *et al.* The post-translational modification, SUMOylation, and cancer (Review). *Int J Oncol*, **52** (4), 1081–1094 (2018).
50. Subramonian, D. *et al.* Analysis of changes in SUMO-2/3 modification during breast cancer progression and metastasis. *J Proteome Res*, **13** (9), 3905–3918 (2014).
51. Hu, C. & Jiang, X. The SUMO-specific protease family regulates cancer cell radiosensitivity. *Biomed. Pharmacother*, **109**, 66–70 (2019).

Figures

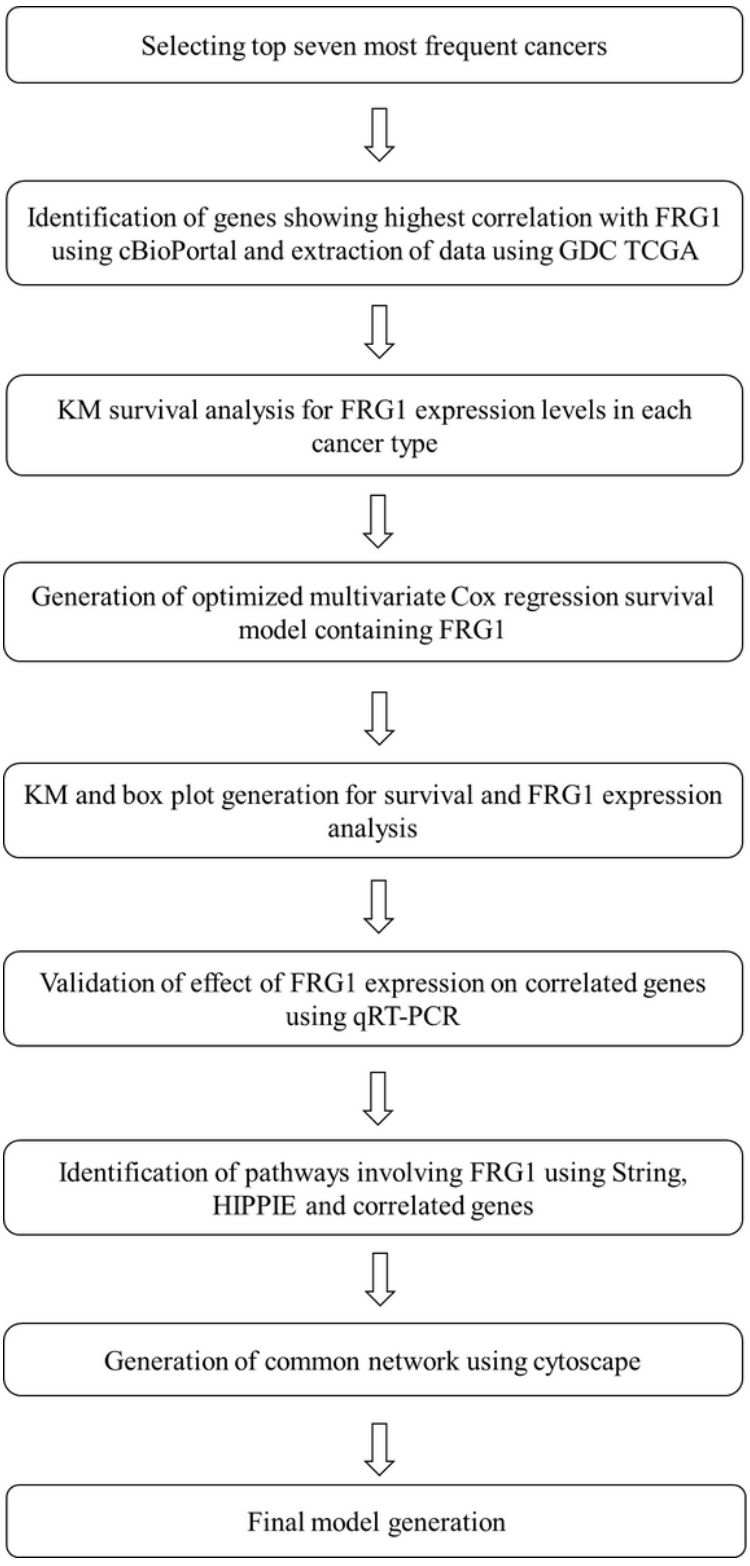

Figure 1

Flow chart of the study design.

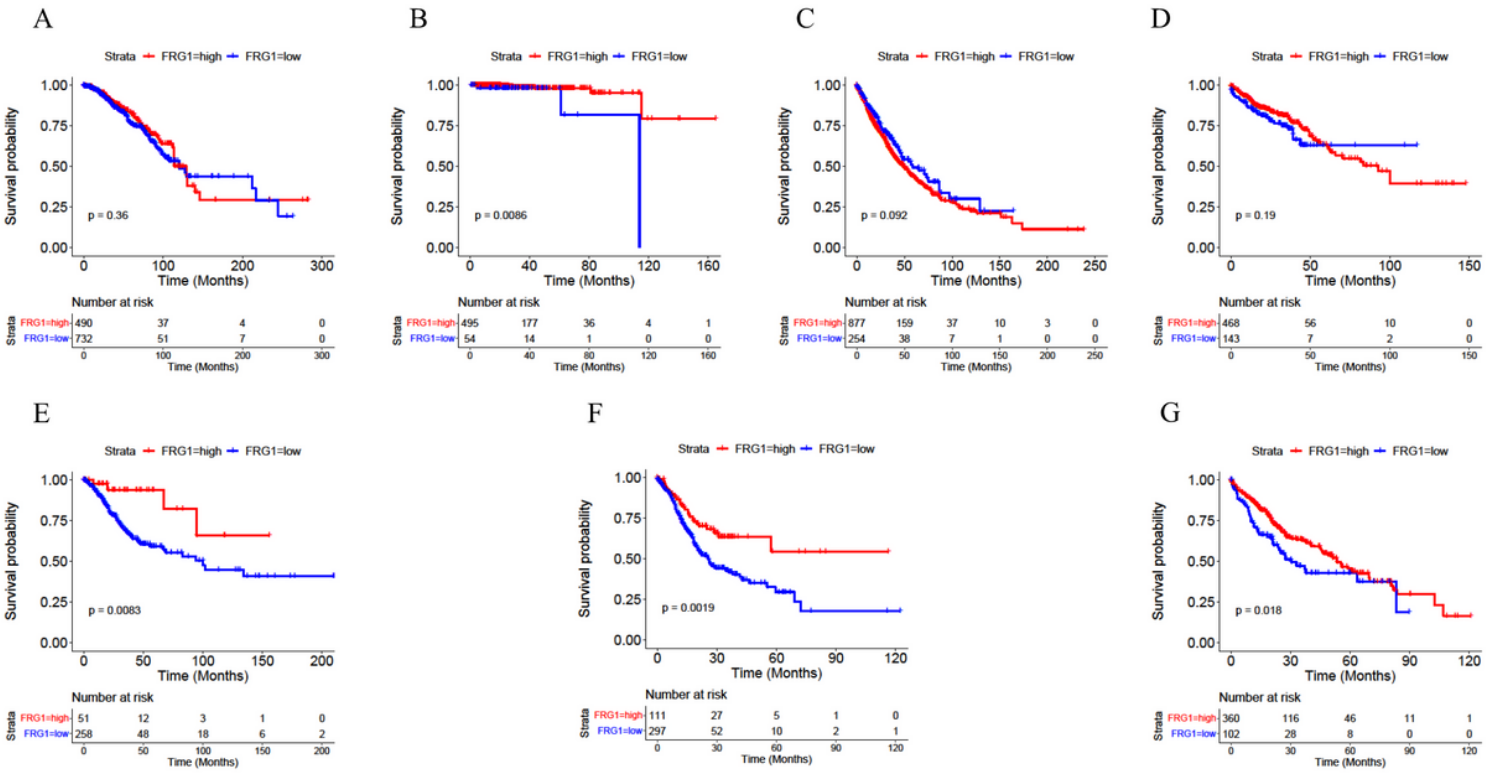

Figure 2

Kaplan-Meier plots showing overall survival with respect to FRG1 expression levels in different cancer types. Survival curves are shown for (A) Breast cancer (logrank P = 0.36), (B) Prostate cancer (logrank P = 0.0086), (C) Lung cancer (logrank P = 0.092) (D) Colorectal cancer (logrank P = 0.19) (E) Cervix uteri cancer (logrank P = 0.0083), (F) Stomach cancer (logrank P = 0.0019) and (G) Liver cancer (logrank P = 0.018). The X-axis represents the number of patients at risk at specific time (in months) and Y-axis shows the probability of survival. Red lines indicate FRG1-high expression group and blue lines indicate FRG1-low expression group divided based on logrank P test.

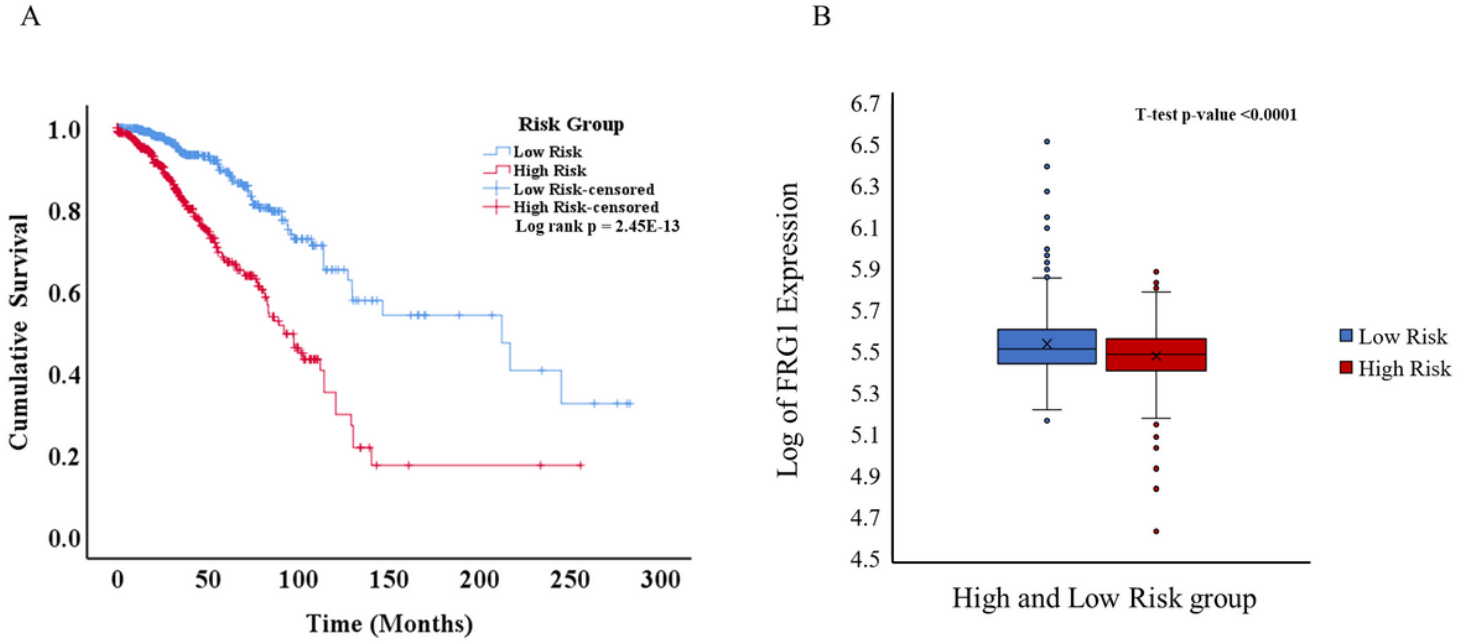

Figure 3

Kaplan-Meier plot and box plot of breast cancer patients risk groups based on the multigene model. (A) KM plot showing overall survival in low-risk and high-risk patient groups (Log-rank test p-value = 2.45E-13). Blue line shows the low-risk group and the red line shows the high-risk group (B) Box plot showing log of FRG1 expression level in low and high-risk groups. The Y-axis represents the log of FRG1 expression and the X-axis shows the group.

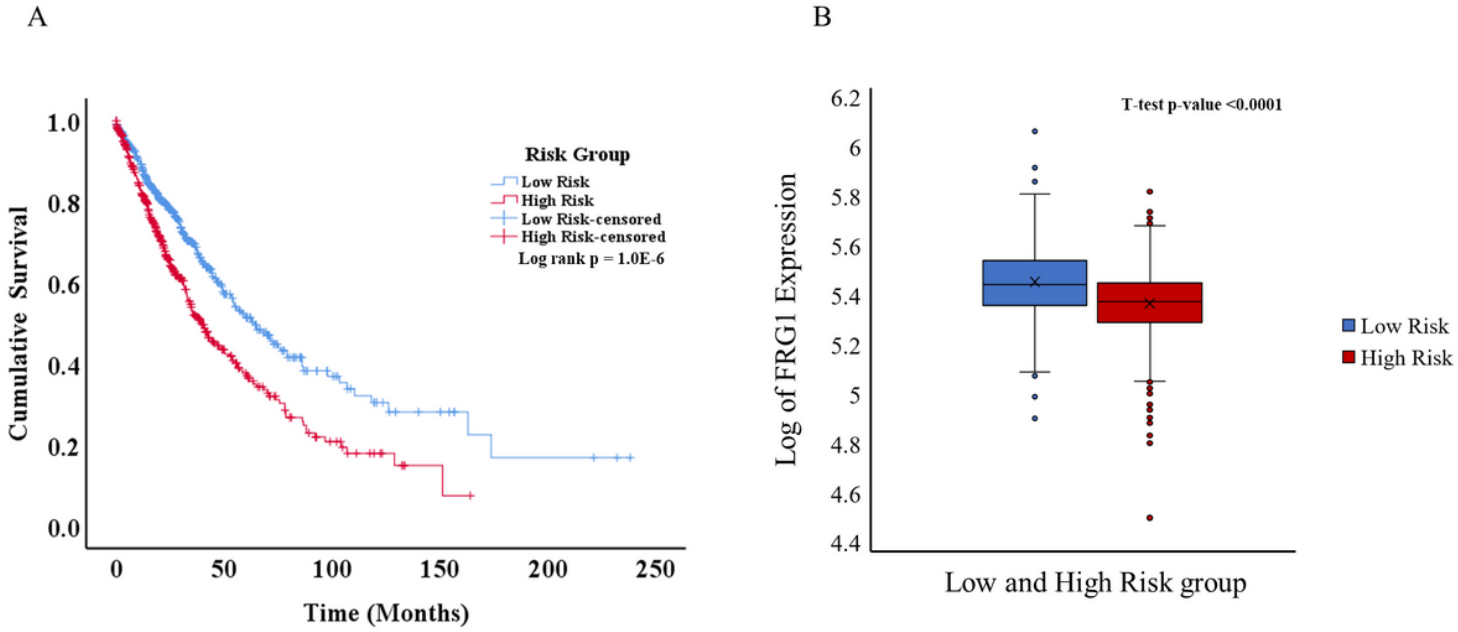

Figure 4

Kaplan-Meier plot and box plot of lung cancer patients risk groups based on the multigene model. (A) KM plot showing overall survival in low-risk and high-risk patient groups (Log-rank test p-value = 1.0E-6). Blue line shows the low-risk group and the red line shows the high-risk group (B) Box plot showing log of FRG1 expression level in low and high-risk groups. The Y-axis represents the log of FRG1 expression and the X-axis shows the group.

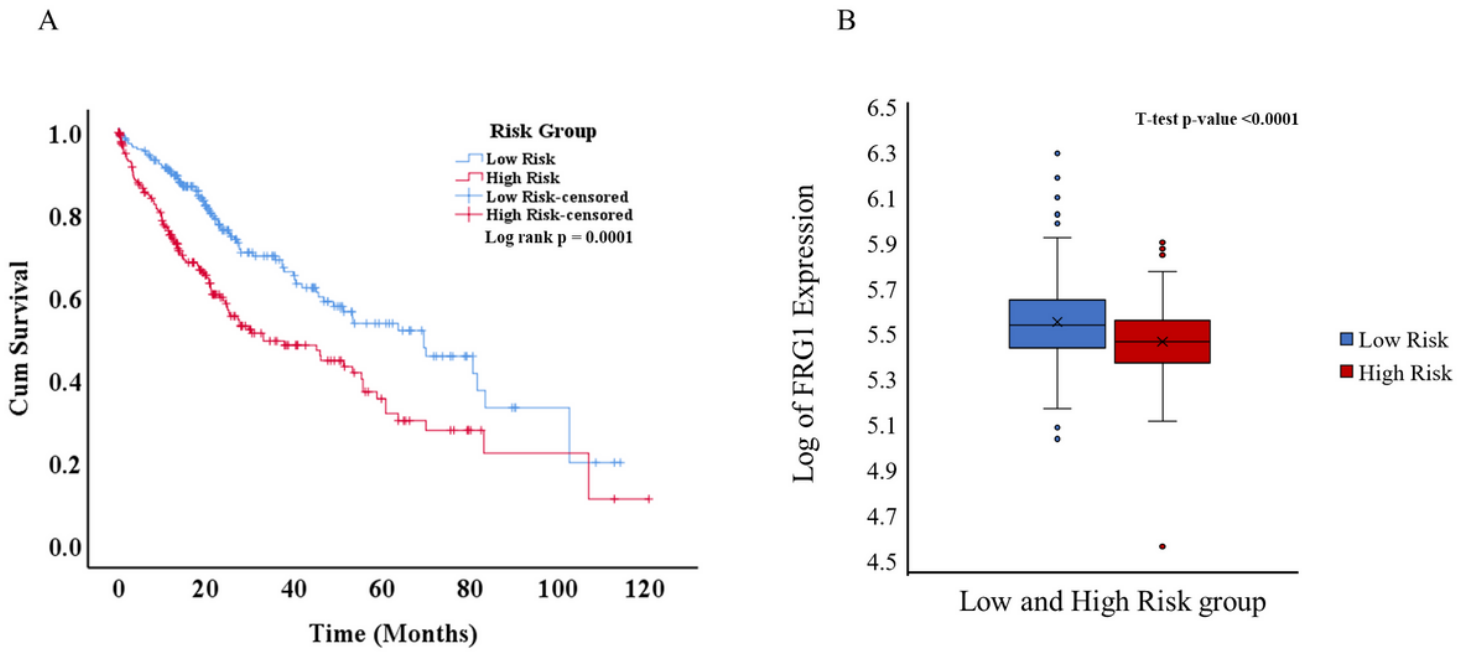

Figure 5

Kaplan-Meier plot and box plot of liver cancer patients risk groups based on the multigene model. (A) KM plot showing overall survival in low-risk and high-risk patient groups (Log-rank test p-value = 0.0001) Blue line shows the low-risk group and the red line shows the high-risk group. (B) Box plot showing log of FRG1 expression level in low and high-risk groups. The Y-axis represents the log of FRG1 expression and the X-axis shows the group.

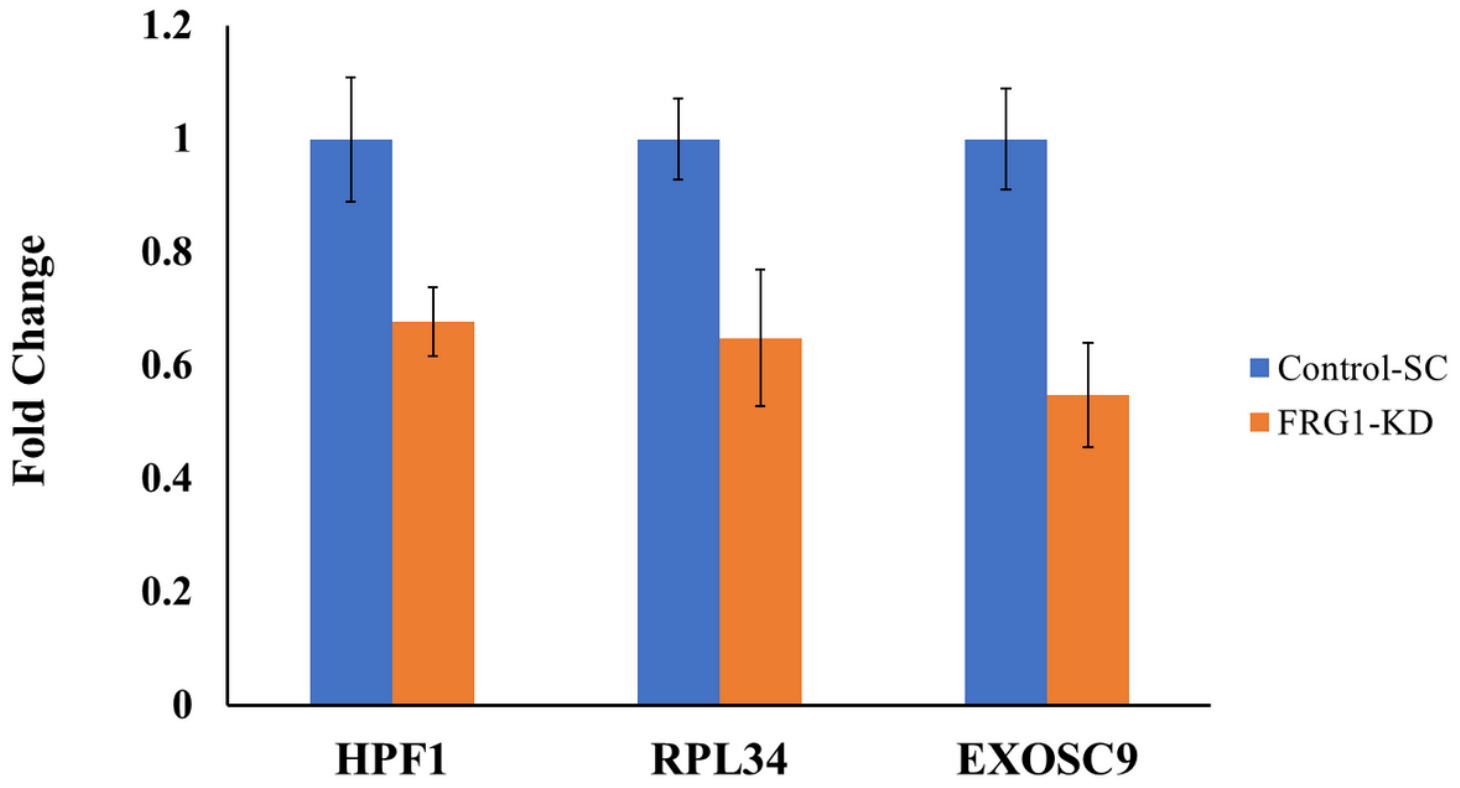

Figure 6

qRT-PCR analysis of the effect of FRG1 expression depletion on the correlated genes. Bar graph shows the expression levels of HPF1, RPL34 and EXOSC9 in HEK293 cells with FRG1 knockdown compared to the control. Y-axis shows fold change in expression using GAPDH as internal control.

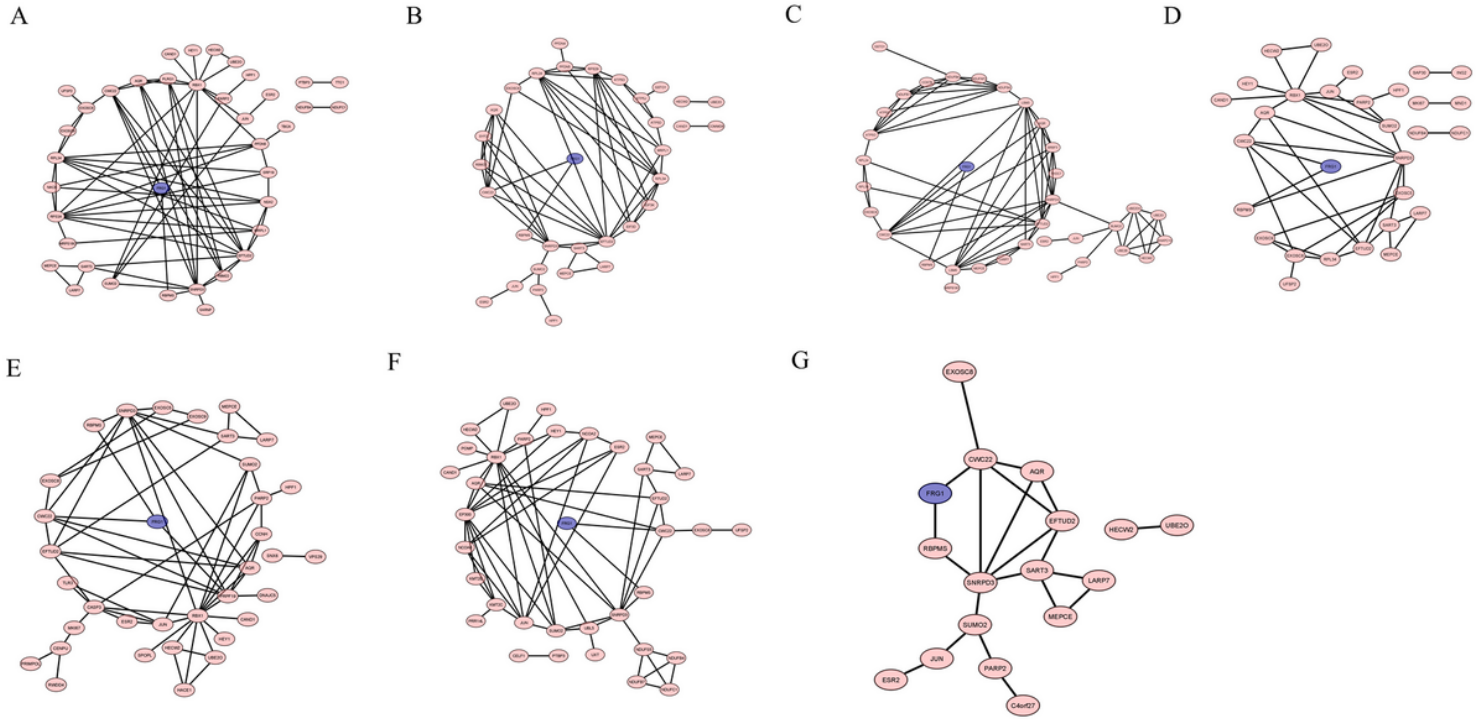

Figure 7

Co-expression and protein-protein interaction network analysis of FRG1 in different cancer types. Networks show FRG1 in blue) at the center and other genes with pink for (A) Breast Cancer (node = 38, edge = 72), (B) Prostate cancer (node = 33, edge = 62), (C) Lung Cancer (node = 35, edge = 78, (D) Cervix-Uteri cancer (node = 30, edge = 42) (E) Stomach cancer (node = 35, edge = 61) and (F) Liver Cancer (node = 36, edge = 64) (G) Common network across cancer types (node = 17, edge = 21). Nodes represent the number of genes and edges define interaction between genes.

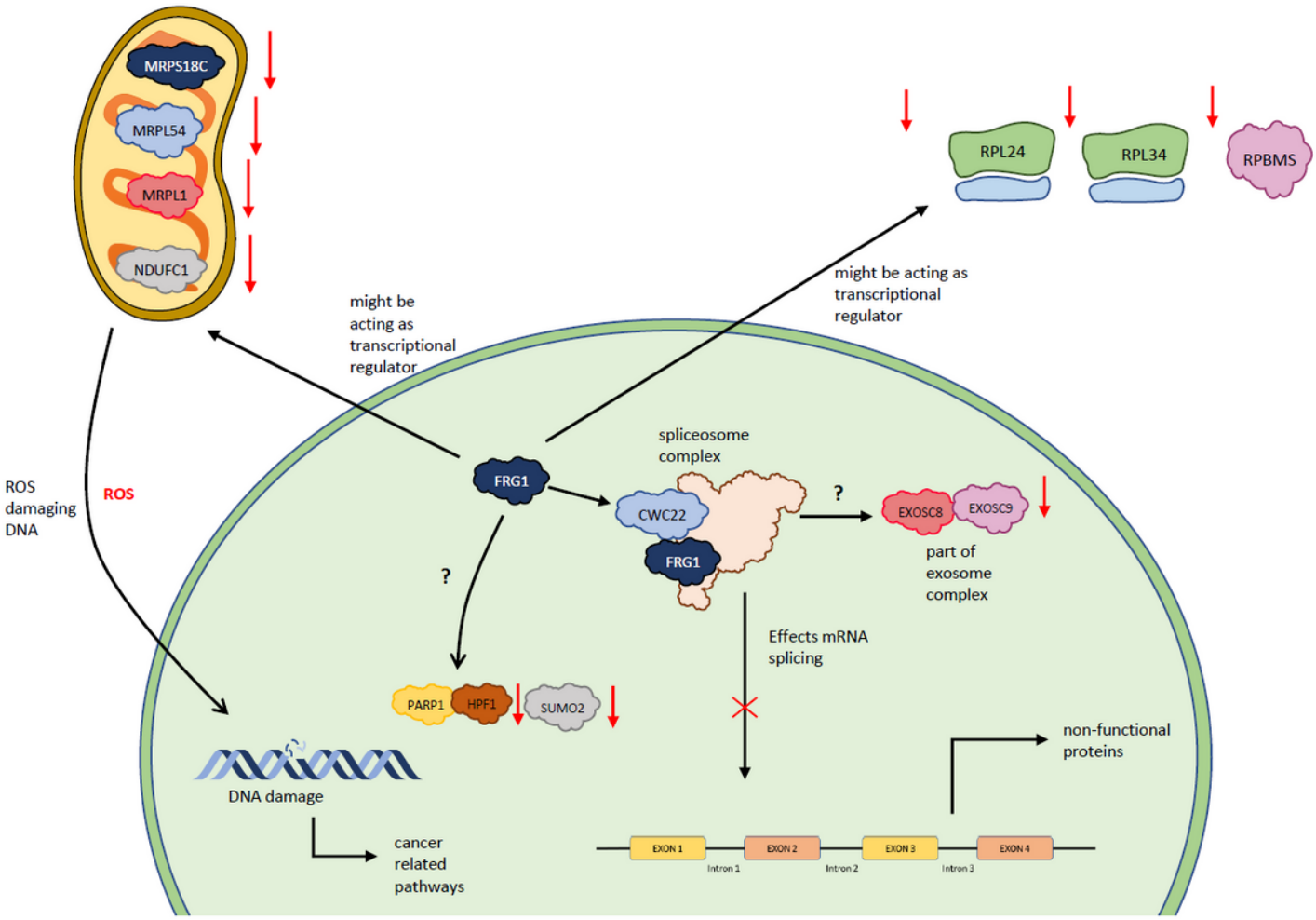

Figure 8

Hypothetical model showing functions of FRG1, based upon expression correlation and protein-protein interactions. Red downward arrows show the downregulation of expression and red cross shows inhibition.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.docx
- SupplementaryTables.xlsx