AWARD NUMBER: W81XWH-16-1-0541

TITLE:
Fusion Genes Predict Prostate Cancer Recurrence

PRINCIPAL INVESTIGATOR:
Jianhua Luo

CONTRACTING ORGANIZATION: University of Pittsburgh,
3550 Terrace Street,
Pittsburgh, PA15261

REPORT DATE:
JANUARY 2021

TYPE OF REPORT:
Final report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JANUARY 2021	Final report	9/15/2016-9/14/2020

4. TITLE AND SUBTITLE
Fusion Genes Predict Prostate Cancer Recurrence

5a. CONTRACT NUMBER
W81XWH-16-1-0541

5b. GRANT NUMBER
PC150332

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jianhua Luo
James Brooks
David Jarrard

E-Mail: Luoj@upmc.edu

5d. PROJECT NUMBER
0010844952

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Pittsburgh,
3550 Terrace Street,
Pittsburgh, PA15261

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Prediction of the clinical outcomes of prostate cancer remains a challenge. Recently, we discovered a panel of 8 fusion genes that occurred in aggressive prostate cancer. In order to make the fusion gene test clinically ready as a predictor, we have modified to test into a semi-quantitative Taqman QRT-PCR. In the funding period, Two hundred seventy-one prostate cancer samples with clinical follow-up were collected from University of Pittsburgh Medical Center. In addition, 194 prostate cancer samples from University of Wisconsin, Madison and 108 prostate cancer samples from Stanford University were collected. Taqman QRT-PCRs were performed on these samples. Significant numbers of samples were found positive for some of these fusion genes. The expression of MAN2A1-FER, SLC45A2-AMACR, MTOR-TP53BP1 fusions are associated with prostate cancer recurrence in the UPMC cohort. Cross-validation showed that fusion gene model predicts up to 91% clinical outcomes of prostate cancer accurately. When cohorts of UPMC, Stanford and Wisconsin were combined, the accuracy is 74%. The combination of fusion with Gleason appeared to improve the overall accuracy from 77% (Gleason) to 92% (Gleason+fusion) in the UPMC cohort, and from 71% (Gleason) to 82% (Gleason+fusion) when all three cohorts are combined. When fusion combined with both pathology stage and Gleason, the accuracy was improved a little further: 93% accuracy in the UPMC cohort and 83% when all three cohorts are combined. In summary, fusion transcript prediction model may have a role in prostate cancer prognosis prediction and guiding the management of prostate cancer patients.

15. SUBJECT TERMS

NONE LISTED

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	Unclassified		**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified Unclassified U	Unclassified		40	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

Page No.

Major Task 1 : We will conduct analysis of MAN2A1-FER, SLC45A2-AMACR, TRMT11-GRIK2, MTOR-TP53BP1, LRRC59-FLJ60017, CCNH-C5orf30, KDM4-AC011523.2, TMEM135-CCDC67 on 5106 prostate cancer samples collected from University of Pittsburgh, Stanford University and University of Wisconsin Madison. We will first establish prostate cancer recurrence model and short PSADT prediction models either by fusion gene status alone or in combination with nomogram based on the cohort from 600 radical prostatectomy samples from UPMC. This model will be locked in and tested on cohorts from University of Pittsburgh, Stanford University and University of Wisconsin. The prediction accuracy, sensitivity and specificity within each cohort will be evaluated.

Subtask 1: In the first 3 months of the funded period, we plan to establish this test in the CLIA certified laboratory at the University of Pittsburgh Medical Center. Fifty-six FFPE samples that were shown to be positive for at least one fusion transcripts in the matched frozen tissues. These FFPE samples had been tested in non CLIA certified laboratory, and achieved 98.9% sensitivity and 100% specificity. We will repeat the same tests on these samples in CLIA certified laboratories. All PCR products will be analyzed through Sanger's sequencing to confirm the authenticity of the fusion products. In addition, all fusion minigene RNA templates will be serially diluted. TAQMAN QRT-PCR will be performed to evaluate the sensitivity of the test. Detection threshold will be obtained. Random selection of 600 prostate cancer samples with definitive clinical outcomes will be carried out in UPMC campus. TAQMAN QRT-PCR on -actin will be used as RNA quality control. For sites 2 and 3, all relevant institutional review board exempt protocols will be secured and approved.

Progress: We have procured the CLIA certified lab space in the beginning of the funded period. To accommodate the reality of formalin-fixed and paraffin-embedded tissues, we have designed a set of new primers and Taqman PCR probes for highly fragmented RNA species. These sets of primers and probes were subsequently tested and validated on synthetic mini-fusion genes of MAN2A1-FER, TRMT11-GRIK2, MTOR-TP53BP1, CCNH-C5orf30, KDM4-AC011523.2, SLC45A2-AMACR, TMEM135-CCDC67, and LRRC59-FLJ60017. The probe and primers for β-actin were also revised to accommodate a shorter RNA fragment. The analyses showed that these assays detect as low as 600-1000 molecules of these fusion transcripts. We then analyzed 56 FFPE samples whose frozen counterparts have been previous found to contain at least one fusion gene using these sets of probes and primers. All samples that were positive for these fusion genes were also positive in the new Taqman qRT-PCR assays. The positive match rate is 100%. All participating institutes, including University of Pittsburgh, Stanford University and University of Wisconsin Madison, had obtained the institutional approval for the exempt protocols.

Subtask 2: From month 4-9 of the first funded year, we will perform TAQMAN QRT-PCR and Sanger's sequencing on a randomly selected cohort of 600 samples from phase 1 that have at least 5 years clinical follow-up. These tests will be performed in CLIA certified laboratory of University of Pittsburgh. The prediction models of PCa recurrence and PSADT mentioned will be developed based on this large number of samples. For sites 2 and 3, the first 300 prostate cancer cases from each site will be selected and evaluated for sufficient materials for the assay.

Progress: To create a training set, we performed Taqman qRT-PCR using the primers and probes as mentioned from above on 271 samples from University of Pittsburgh, 155 samples from University of Wisconsin Madison, and 150 samples from Stanford University. The results show surprisingly high positive rate of SLC45A2-AMACR in Stanford and Wisconsin cohort, reaching 96% and 92.6% respectively. Among these fusion genes, the lowest frequent one is TMEM135-CCDC67: A total of 8 samples were found positive. In addition, high positive rate of CCNH-C5orf30 was also found in the prostate cohort from University of Wisconsin. In general, the rates of fusion gene positive samples are comparable among the 3 cohorts (table 1). Subsequent analyses showed that MAN2A1-FER (or normalized MANA1-FER), TRMT11-GRIK2, and mTOR-TP53BP1 gene fusions have the highest odd ratios for predicting the recurrence of prostate cancer for UPMC and University of Wisconsin cohorts (Table 2).

Table 1 Positive rate of fusion in prostate cancers

Cohort	MAN2A1/ FER	TRMT11/ GRIK2	MTOR/ TP53BP1	CCNH/ C5orf30	KDM4B/ AC011523.2	SLC45A2/ AMACR	TMEM135/ CCDC67	LRRC59/ FLJ60017
UPMC	13% (60)	25.8% (119)	2.8% (13)	33.4% (154)	0.4% (2)	50.1% (234)	1% (5)	3.4% (16)
Stanford	18% (9)	20% (10)	10% (5)	12% (6)	4% (2)	96% (48)	6% (3)	22% (11)
UWisc	19% (31)	12.9% (21)	4.3% (7)	76.7% (125)	9.2% (15)	92.6% (151)	0.6% (1)	26% (43)

Table 2
The cutoffs(and OR) of each fusion gene in each cohort

Cohort	MAN2A1/FER	MAN2A1/FER-actin	TRMT11/GRK2	MTOR/TP53BP1	CCNH/C5Orf30	KDM4/AC011523.2	SL45A2/AMACR	TMEM135/CCDC67
UPMC	32(26.3)	0 (25.9)	43(5.53)	42(inf)	39(0.12)	44(inf)	34(1.57)	47(1.54)
Wisconsin	35(14.81)	3(7.57)	42(inf)	40(23.6)	38(0.49)	40(1.5)	31(1.7)	N/A
Stanford	39(1.71)	0(0.34)	39(4.03)	39(inf)				

To establish a prediction model, we combined top 6 fusion genes that have prediction power to construct classification models to predict prostate cancer recurrence. As shown in table 3, all three models (Random Forest,

Table 3

Model	Fusion genes only					2x2 table		
	Sensitivity	Specificity	Youden	Accuracy	AUC		Recurrent (n=107)	Non-Recurrent (N=164)
RF	0.80	0.81	0.61	0.81	0.862	Positive	TP=86	FP=31
	Top 6, cutoff=0.2					Negative	FN=21	TN=133
SVM	0.71	0.87	0.58	0.81	0.77	Positive	TP=76	FP=21
	Top 4, cutoff=0.2					Negative	FN=31	TN=143
LDA	0.71	0.88	0.59	0.81	0.85	Positive	TP=76	FP=20
	Top 6, cutoff=0.4					Negative	FN=31	TN=144

Support vector machine and Linear discriminant analysis) yielded very similar accuracy: 81%, even though the specificity and sensitivity may vary. When combined with gleason's score and TNM pathology staging, the accuracy improves to 84-86%.

Table 4

Model	Gleason and TNM stage + fusion genes					2x2 table		
	Sensitivity	Specificity	Youden	Accuracy	AUC		Recurrent (n=208)	Non-Recurrent (N=368)
RF	0.79	0.82	0.61	0.81	0.86	Positive	TP=164	FP=67
	top 5 fusion genes, cutoff=0.2					Negative	FN=44	TN=301
SVM	0.75	0.83	0.59	0.81	0.84	Positive	TP=157	FP=61
	Top 6 fusion genes, cutoff=0.2					Negative	FN=51	TN=307
LDA	0.77	0.85	0.61	0.82	0.87	Positive	TP=160	FP=57
	Top 5 fusion genes, cutoff=0.4					Negative	FN=48	TN=311

When the same models were applied to the dataset from University of Wisconsin, the accuracy rate yielded 75-84%. Interestingly, when combined with Gleason's grade and pathology TNM staging, the accuracy rate improved to 88-90%. However, the same model fared worse in Stanford data set: 67-68% accuracy was found. Combination with fusion genes, Gleason's grade and pathology TNM staging improves the accuracy to 75%.

When all data were pooled, and 10 fold cross validation was performed. The accuracy of fusion gene prediction rate is 76-77%. When combined fusion genes, Gleason's grading and TNM staging, the accuracy is improved to 81-82% (table 4). In contrast, if prediction is only relied on Gleason's grade and TNM staging, the prediction accuracy is 74-76% across all three data set. As a result, we concluded that fusion contains independent prediction value and can assist in predicting the clinical outcomes of prostate cancer.

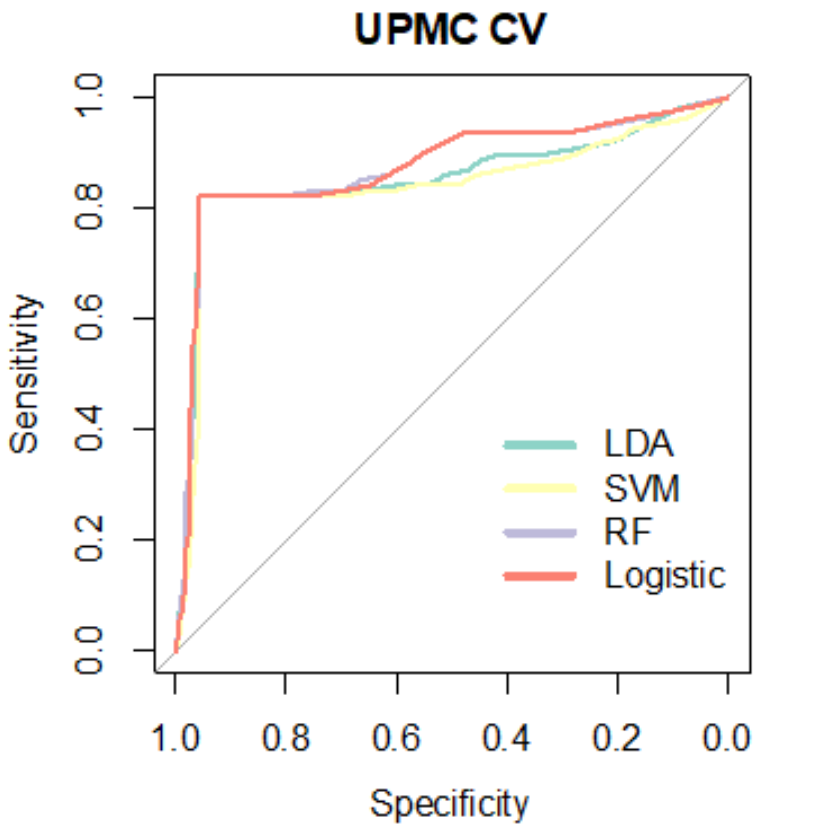

Figure 1. ROC analyses of fusion transcript model predicting prostate cancer recurrence of UPMC prostate cancer samples. The model utilized fusion transcripts only as prediction factor. LDA-Linear discriminant analysis; SVM-Support-vector machine; RF-Random forest.

Subtask 3: From month 10 of the first year to the end of year 3 of the funded period, we will validate predictive models based on the fusion transcript panel and clinical and pathological parameters on independent datasets from the University of Pittsburgh, University of Wisconsin and Stanford University.

Progress: After the collection of the results of all the qRT-PCR analyses for the prostate cancer samples from UPMC, we performed a leave-one-out cross-validation to examine whether fusion transcripts are predictive of prostate cancer recurrence. As shown in figure 1, 90.4% accuracy was achieved based the detection of 4 fusion transcripts (MAN2A1-FER, SLC45A2-AMACR,

mTOR-TP53BP1 and LRRC59-FLJ60017 in the cancer samples. The sensitivity is 82% and the specificity 95%. When all fusion transcripts were included, the accuracy only slightly improved to 91%. Among the fusion transcripts, MAN2A1-FER is most associated with prostate cancer recurrence, with a p-value <0.000001. Survival analysis showed that patients with positive recurrence prediction had only 15% PSA-free survival 5 years after the radical prostatectomy (figure 2), while patients with negative prediction had over 90% PSA-free survival. When the same leave-one-out approach was applied to Gleason grading prediction, it generates 77% accuracy of prediction with 57% sensitivity and 91% specificity. When fusion transcripts were considered along with Gleason grade, the accuracy improves to 92% with sensitivity of 89% and specificity of 93%. When fusion transcripts, Gleason grade and pathology stage are combined as a model for prediction, the accuracy improves to 94% with sensitivity and specificity 94.4% and 93.9%, respectively. When the analyses includes the cohorts from Stanford University (108 samples) and University of Wisconsin (194 samples), the same model prediction yielded an accuracy of 74% with sensitivity of 41% and specificity of 94%, very similar to the Gleason grade prediction (75% accuracy with sensitivity of 47% and specificity of 92%). When fusion transcripts were combined with Gleason grade, the accuracy improved to 77%. The highest accuracy (83%) was achieved when fusion transcripts, Gleason grade and pathology stage were combined as a model (figure 3). The PSA-free survival analysis on the combined prediction model showed that prostate cancer patients with positive prediction had less than 20% PSA-free survival in 5 years, while patients with negative prediction had PSA-free 5-year survival rate more than 85% (figure 4).

When cohort from Wisconsin was independently analyzed, the fusion gene predicted 84% recurrence accurately, comparable with 85% Gleason prediction rate. When Gleason grade was combined with fusion gene model, it improved to 86%. In contrast, Stanford cohort showed that fusion transcript model was less accurate, generating only 65% accuracy in predicting prostate cancer recurrence, comparing with 75%

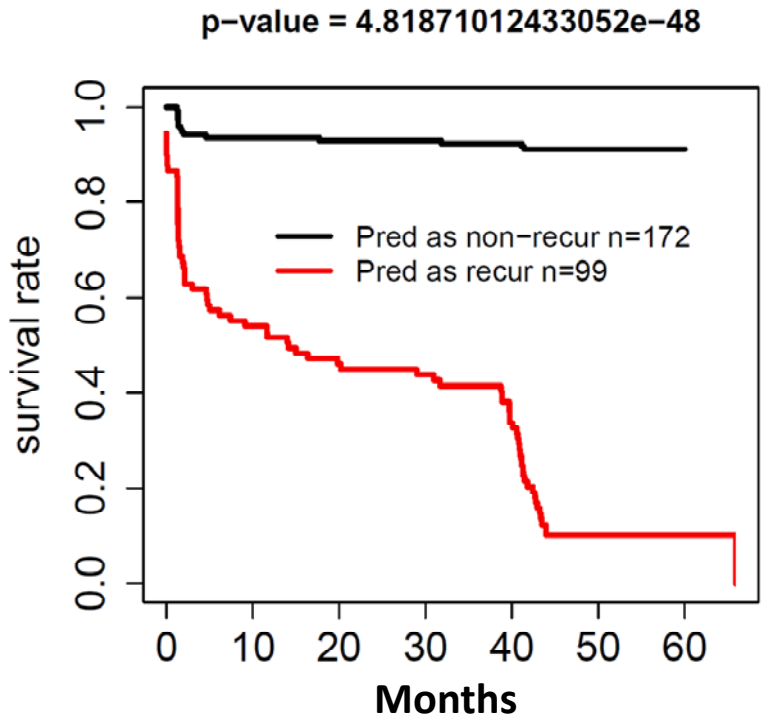

Figure 2. Kaplan-Mier analysis of PSA-free survival of prostate cancer based on fusion prediction model on prostate cancer samples from UPMC cohort. The prediction was based on the results from Random Forest method.

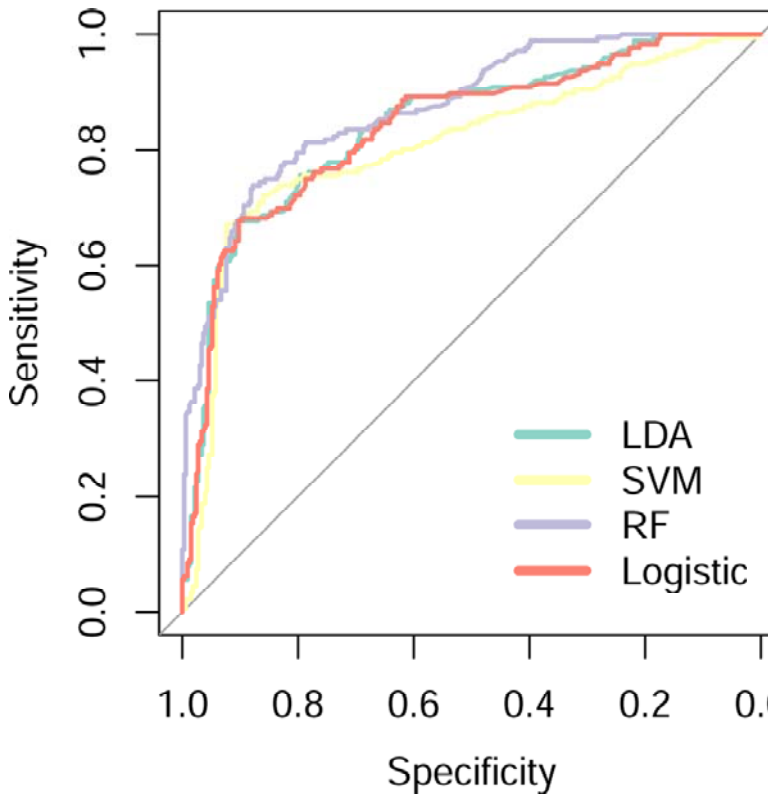

Figure 3. ROC analysis of prostate cancer recurrence prediction based on prostate cancer samples of combined cohorts of UPMC, Stanford and University of Wisconsin. The model utilized fusion transcripts, Gleason grade and pathology stage as prediction factors. LDA-Linear discriminant analysis; SVM-Support-vector machine; RF-Random forest.

accuracy by Gleason grade. However, when fusion model was combined with Gleason grade, the model prediction improved to 81% accuracy. When we combined cohorts from Wisconsin and Stanford into one validation cohort, the fusion prediction model predicts 72% recurrence accurately, slightly lower than Gleason grade prediction (73%). Combining Gleason with fusion transcripts improves the correct prediction rate to 75%. Fusion transcript also improve on Gleason plus pathology stage model (78% over 76%).

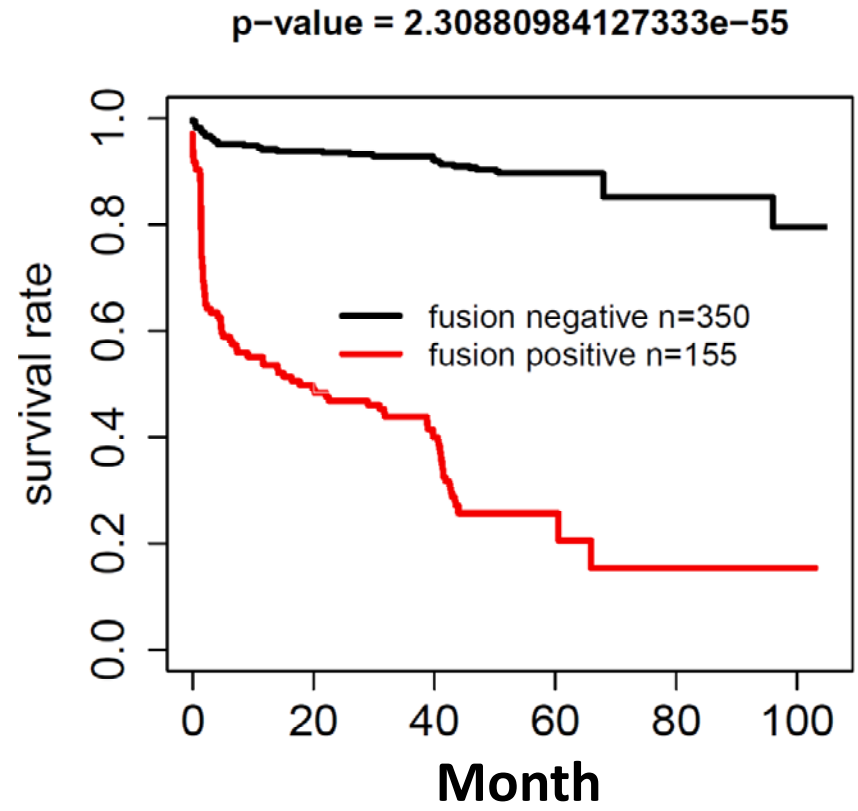

Figure 4. Kaplan-Mier analysis of PSA-free survival of prostate cancer based on fusion prediction model on prostate cancer samples from UPMC, Stanford and University of Wisconsin cohort. The prediction was based on the results from Random Forest method using fusion transcripts, Gleason grade and pathology stage as prediction factors.

Association analysis showed that MAN2A1-FER, TRMT11-GRIK2 and SLC45A2-AMACR are associated with higher pre-operational PSA, with p-value 0.0025, 0,016, and 0.008, respectively. mTOR-TP53BP1 and TRMT11-GRIK2 are associated with higher Gleason's grade (p=0.002 for mTOR-TP53BP1 and 0.0002 for TRMT11-GRIK2). mTOR-TP53BP1 is associated with larger tumor volume (p=0.049) and more advanced stage of cancer (p=0.007). Samples with lymph node metastasis tends to have mTOR-TP53BP1 and SLC45A2-AMACR present in the primary cancer samples (p=0.008 for mTOR-TP53BP1 and p=0.044 for SLC45A2-AMACR).

Overall, fusion transcripts were found widely present in prostate cancer samples, and were associated with prostate cancer recurrence and other pathological features of the cancer. Even though there were significant inter-cohort variations in terms of predictability of prostate cancer clinical outcomes, fusion transcript detection improves on the current clinical means in predicting the cancer outcomes, and thus, can be utilized as an important tool in predicting cancer outcomes in patient management.

Additional progress on fusion gene analysis supported by this grant:

Fusion transcripts are present in human prostate cancer and other cancer cell lines

In our previous studies, we have characterized eight fusion genes identified in aggressive prostate cancer samples. Additional analyses showed that one of the fusion genes called MAN2A1-FER is frequently present in 5 other types of human malignancies. To expand our analyses of other fusion genes in the panel, we analyzed TRMT11-GRIK2, MTOR-TP53BP1, CCNH-C5orf30, KDM4-AC011523.2, TMEM135-CCDC67, LRRC59-FLJ60017 in 20 cancer cell lines from six different human malignancies (figure 5). TRMT11-GRIK2 fusion transcript was identified in breast

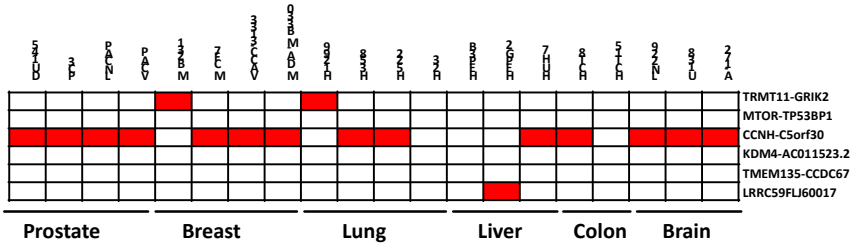

Figure 5. Fusion transcripts are present in human cancer cell lines. Representative cancer cell lines from lung, liver, prostate, breast, colon and brain were examined for the presence of fusion transcript TRMT11-GRIK2, CCNH-C5orf30, mTOR-TP53BP1, TMEM135-CCDC67, KDM4-AC011523.2 and LRRC59-FLJ60017 using quantitative TaqMan RT-PCR. Red denotes positive for the fusion transcript, while blank negative for the fusion transcript. All positive fusion samples were verified by Sanger sequencing.

cancer cell line MD-MB231 and lung cancer cell line H1299, while CCNH-C5orf30 was positive in 14 of 20 cancer cell lines, including all prostate cancer cell lines tested (PC3, DU145, LNCaP and VCaP), 3 of 4 breast cancer cell lines (MCF7, VACC-3133 and MDA-MB330), 2 of 4 lung cancer cell line (H358 and H522), 1 of 3 liver cancer cell lines (HepG2), 1 of 2 colon cancer cell lines (HCT8) and 3 of 3 GBM cell lines (LN229, U138 and A-172). Also, KDM4-AC011523.2 and LRRC59-FLJ60017 were present in liver cancer cell line HepG2. These results suggest that these fusion genes are not specific for prostate cancer. They may be present in the primary cancer samples of a variety of human malignancies.

Fusion transcripts are present in 7 other types of human malignancies:

To investigate whether any of the above-mentioned fusion genes has a role in human cancers, Quantitative TaqMan qRT-PCRs using primers and probe specific for each fusion gene were performed on primary human cancer samples representing seven different types of human malignancies. Our results showed that TRMT11-GRIK2 is present in all seven types of human malignancies (figure 6), including breast cancer (41/60, 68.33%), colon cancer (25/60, 41.7%), esophageal adenocarcinoma (9/34, 26.5%), hepatocellular carcinoma (9/70, 12.9%), ovarian adenocarcinoma (28/61, 45.9%), glioblastoma multiforme (26/150, 17.3%) and non-small cell lung cancer (39/141, 27.7%). CCNH-C5orf30 transcript was also detected in 7 different types of human cancers with relatively high frequencies: 85% (51/60) breast cancer, 43% (26/60) colon cancer, 50.8% (31/61) ovarian cancer, 67.6% (23/34) esophageal adenocarcinoma, 41.8% non-small cell lung cancer (59/141), 37% (26/70) liver cancer and 53% (80/150) glioblastoma multiforme. mTOR-TP53BP1, on the other hand, was detected in five different types of human malignancies with significantly lower frequencies: breast cancer (10/60, 16.7%), colon cancer (4/60, 6.7%), ovarian adenocarcinoma (4/61, 6.6%), glioblastoma multiforme (7/150, 4.7%) and lung cancer (8/141, 5.7%). LRRC59-FLJ60017 was found present in four different types of human cancers: esophageal adenocarcinoma (3/34, 8.8%), ovarian adenocarcinoma (4/61, 6.6%), glioblastoma multiforme (16/150, 10.7%), and non-small cell lung cancer (33/141, 23.4%). Only two glioblastoma multiforme and one lung cancer samples were found positive for TMEM135-CCDC67. KDM4-AC011523.2 fusion was only found in 3 breast cancer and 3 lung cancer samples.

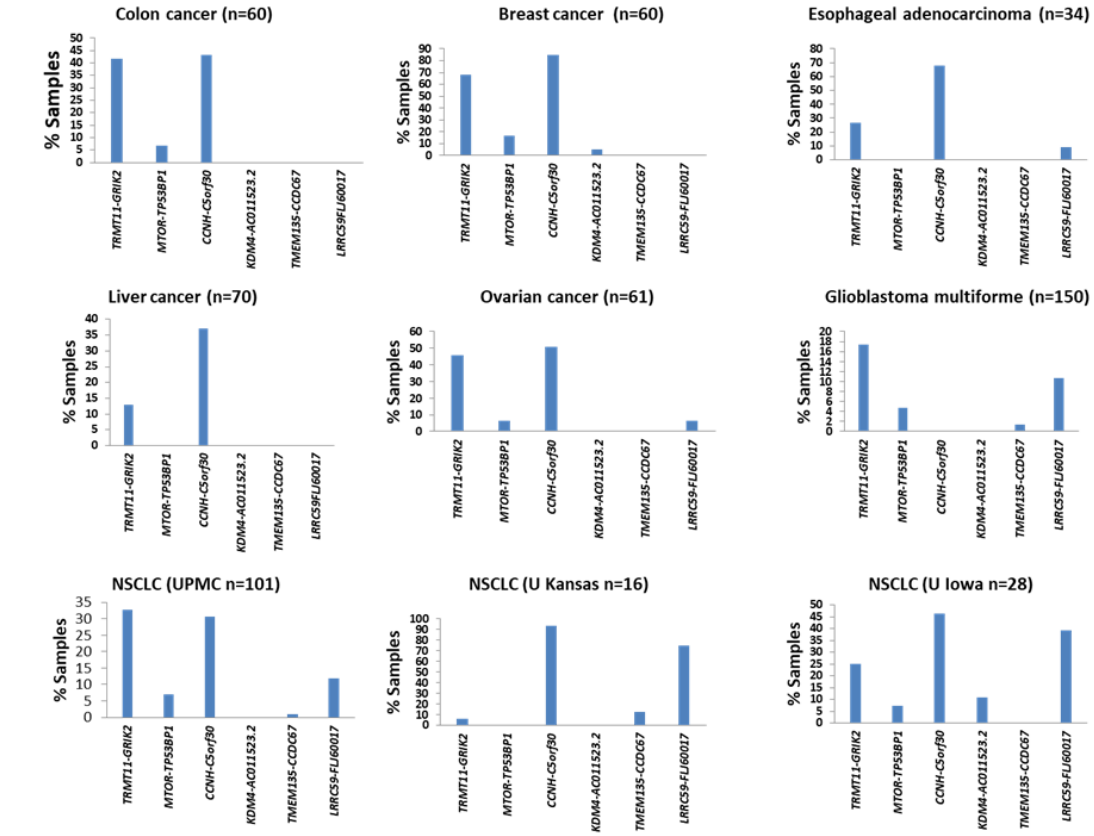

Figure 6. Frequency of gene fusion in primary human malignant samples. Real-time TaqMan qRT-PCRs were performed on formalin-fixed paraffin-embedded samples of breast cancer, colon cancer, esophageal adenocarcinoma, ovarian cancer, hepatocellular carcinoma, non-small cell lung cancer and glioblastoma multiforme. The numbers of samples of each cohort are indicated. Sanger sequencing was performed on 20% samples that were found positive for the fusion genes.

Interestingly, ductal type of breast cancers positive for TRMT11-GRIK2 was associated with less likely to develop local lymph node metastasis (46.7% versus 93.3%, p=0.014). Liver cancer positive for TRMT11-GRIK2 was also associated with a higher rate of overall survival (41.7% versus 6.9%, p=0.006). KDM4-AC011523.2 was only detected in lobular type breast cancer and adenocarcinoma of the lung. Patients with lobular breast cancers positive for mTOR-TP53BP1 were also less likely to have lymph node metastasis (0% versus 40%, p=0.017). CCNH-C5orf30 was more frequent in adenocarcinoma of

the lung cancer versus squamous type (67.7% versus 34.8%, p=0.001), and colon cancer with advanced stages at the time of diagnosis (52.3% versus 18.8%, p=0.037).

Fusion transcripts are present in metastatic lymph nodes:

To investigate whether fusion genes are also present in the metastatic lesion of human cancers, breast cancer, colon cancer and ovarian cancer samples with matched lymph node metastasis were analyzed (figure 7). Twenty-six of 30 metastatic breast cancers in the lymph nodes were positive for TRMT11-GRIK2, including seven metastatic cancers whose matched primary breast cancers were negative for the fusion. For colon cancers, however, the status of TRMT11-GRIK2 between primary cancer samples and lymph node metastases was matched by 78.5% (11/14): Eleven lymph node metastases were exactly matched with the status of the primary colon cancer samples, while two samples of lymph node metastases were found negative for TRMT11-GRIK2 fusion. One lymph node metastasis was found positive for the fusion gene while the matched primary sample was negative (figure 3). For ovarian adenocarcinoma, nine metastatic lesions were found to contain TRMT11-GRIK2 fusion gene, matching all primary samples. However, four lymph node metastases contained no TRMT11-GRIK2 while the matched primary cancer samples were positive. One lymph node metastasis gained the fusion of TRMT11-GRIK2 over the primary cancer sample. For CCNH-C5orf30, the matching rate of primary breast cancer with lymph node metastases was 62%, while the matching rates for ovarian cancer and colon cancer with their corresponding lymph node metastases were 72% and 73%, respectively. For mTOR-TP53BP1, two of 3 lymph node metastases retained the fusion in both colon and ovarian cancers. Two of 2 lymph node metastases in ovarian cancer retained LRRC59-FLJ60017 fusion. These results suggest significant heterogeneity of the cancer samples. However, most fusion genes were retained in metastatic lesions.

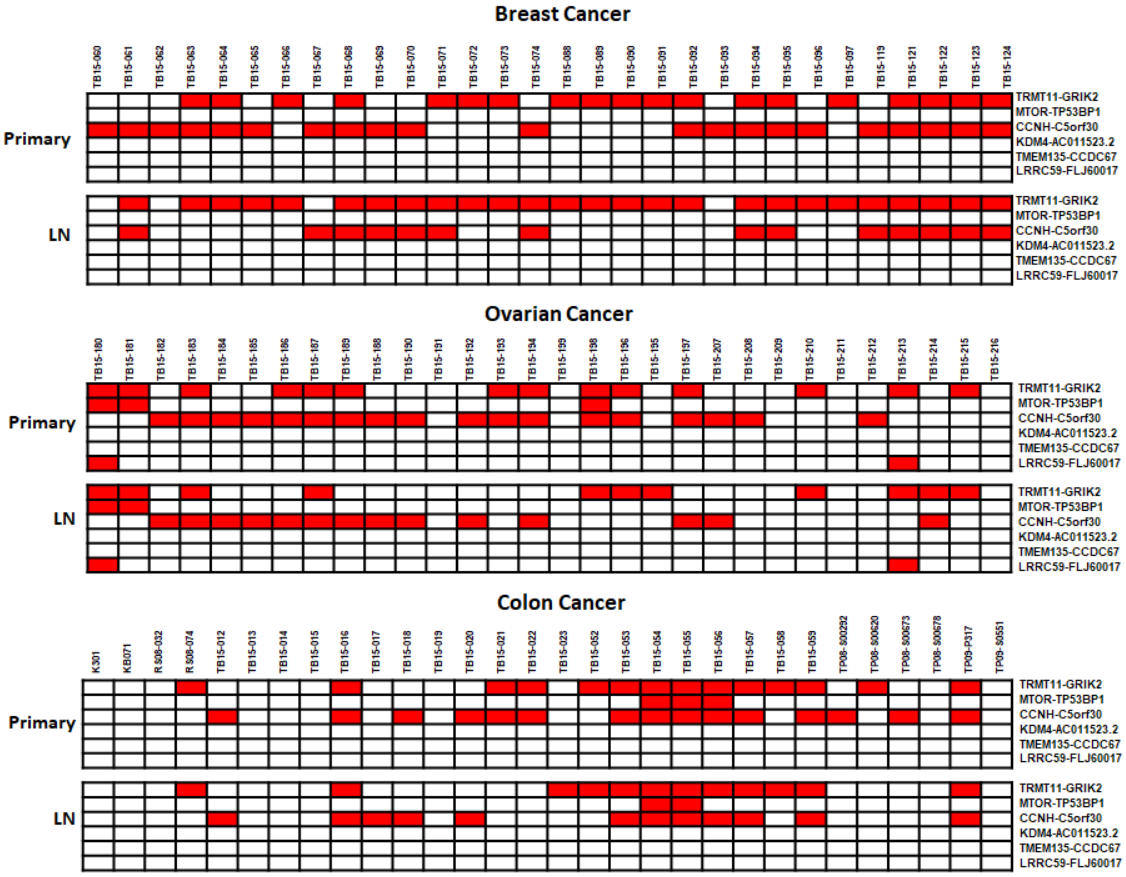

Figure 7. Fusion transcript in metastatic lymph node samples. Real-time TaqMan qRT-PCRs were performed on formalin-fixed paraffin-embedded samples of breast cancer (top), ovarian cancer (middle) and colon cancers (bottom) and their corresponding lymph node metastases. Both primary (top) and matched lymph node (bottom) cancer samples are shown. Red indicates positive detection of the fusion transcript, while blank indicates negative detection of the fusion transcript. Sanger sequencing was performed on 20% samples that were found positive for the fusion genes.

1. **INTRODUCTION:** *Narrative that briefly (one paragraph) describes the subject, purpose and scope of the research.*

Cancer specific fusion genes are the results of chromosome rearrangement in the cancer genomes. The detection of fusion transcripts in cancer cells may reflect the progression of human cancer. Previously, we have identified a panel of 8 fusion genes in prostate cancer. The presence of these fusion transcripts correlated with the aggressive behavior of prostate cancer. In this proposed study, we will conduct large scale analysis to evaluate whether the detection of these fusion transcripts is predictable for poor clinical outcomes.

2. **KEYWORDS:** *Provide a brief list of keywords (limit to 20 words).*

Fusion gene, RNA, Taqman RT-PCR, in situ hybridization, RNA, prostate cancer, cancer relapse, chromosome

3. **ACCOMPLISHMENTS:** *The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction.*

What were the major goals of the project?

List the major goals of the project as stated in the approved SOW. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

We will conduct analysis of MAN2A1-FER, SLC45A2-AMACR, TRMT11-GRIK2, MTOR-TP53BP1, LRRC59-FLJ60017, CCNH-C5orf30, KDM4-AC011523.2, TMEM135-CCDC67 on over 1000 prostate cancer samples collected from University of Pittsburgh, Stanford University and University of Wisconsin Madison. We will first establish prostate cancer recurrence model and short PSADT prediction models either by fusion gene status alone or in combination with nomogram based on the cohort from 600 radical prostatectomy samples from UPMC. This model will be locked in and tested on cohorts from University of Pittsburgh, Stanford University and University of Wisconsin. The prediction accuracy, sensitivity and specificity within each cohort will be evaluated.

1) In the first 3 months of the funded period, we plan to establish this test in the CLIA certified laboratory at the University of Pittsburgh Medical Center. Fifty-six FFPE samples that were shown to be positive for at least one fusion transcripts in the matched frozen tissues. Detection threshold will be obtained.

2) From month 4-9 of the first funded year, we will perform TAQMAN QRT-PCR and Sanger's sequencing on a randomly selected cohort of 600 samples from phase 1 that have at least 5 years clinical follow-up. These tests will be performed in CLIA certified laboratory of University of Pittsburgh. The prediction models of PCa recurrence and PSADT mentioned will be developed based on this large number of samples. For sites 2 and 3, the first 300 prostate cancer cases from each site will be selected and evaluated for sufficient materials for the assay.

3) TAQMAN QRT-PCR analysis for the fusion genes will be carried out at the CLIA certified lab at the University of Pittsburgh using approximately 200 samples provided by Drs. Brooks and Jarrard. In addition, validation of selected prostate cancer samples on specific fusion genes using FISH will be performed. Statistical analyses will be performed to evaluate whether the fusion gene status is predicative for the clinical outcomes of prostate cancers.

What was accomplished under these goals?

For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative); and/or 4) other achievements. Include a discussion of stated goals not met. Description shall include pertinent data and graphs in sufficient detail to explain any significant results achieved. A succinct description of the methodology used shall be provided. As the project progresses to completion, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments.

1. We have reaffirmed the detection of fusion transcripts in prostate cancer samples.
2. Using sensitive detection methods, we have found that fusion genes are widely present in prostate cancer samples.
3. We have established a preliminary training model for the prediction of the clinical outcomes of prostate cancer.
4. We have found the fusion genes identified in prostate cancer are also present in other human malignancies.
5. We have found that the fusion transcripts are present in the serum samples of human cancers in cell-free RNA form.

What opportunities for training and professional development has the project provided?

If the project was not intended to provide training and professional development opportunities or there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe opportunities for training and professional development provided to anyone who worked on the project or anyone who was involved in the activities supported by the project. "Training" activities are those in which individuals with advanced professional skills and experience assist others in attaining greater proficiency. Training activities may include, for example, courses or one-on-one work with a mentor. "Professional development" activities result in increased knowledge or skill in one's area of expertise and may include workshops, conferences, seminars, study groups, and individual study. Include participation in conferences, workshops, and seminars not listed under major activities.

Nothing to report

How were the results disseminated to communities of interest?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how the results were disseminated to communities of interest. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

Nothing to report

What do you plan to do during the next reporting period to accomplish the goals?
If this is the final report, state "Nothing to Report."

Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Nothing to report

4. **IMPACT:** D*escribe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project relative to:*

What was the impact on the development of the principal discipline(s) of the project?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that an intelligent lay audience can understand (Scientific American style).

MAN2A1-FER is the first tyrosine kinase fusion genes found to play critical roles in prostate cancer, it opens a new door for the treatment of prostate cancer using tyrosine kinase inhibitors. In addition, we developed a novel approach to treat human cancers that are positive for fusion gene by inserting a suicide gene into the chromosomal breakpoint of a fusion gene in the cancer genome. This could be a new way to treat prostate cancers that are refractory to other modes of the cancer treatment.

What was the impact on other disciplines?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Fusion transcripts mentioned in the proposed study was also present in ovarian cancer, breast cancer, colon cancer, lung cancer, liver cancer, GBM and esophageal adenocarcinoma, and may play roles in the developments of those cancers.

What was the impact on technology transfer?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe ways in which the project made an impact, or is likely to make an impact, on commercial technology or public use, including:

- *transfer of results to entities in government or industry;*
- *instances where the research has led to the initiation of a start-up company; or*
- *adoption of new practices.*

Nothing to report

What was the impact on society beyond science and technology?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

- *improving public knowledge, attitudes, skills, and abilities;*
- *changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or*
- *improving social, economic, civic, or environmental conditions.*

Nothing to report

5. CHANGES/PROBLEMS: *The PD/PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. If not previously reported in writing, provide the following additional information or state, "Nothing to Report," if applicable:*

Nothing to report

Changes in approach and reasons for change
Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

Actual or anticipated problems or delays and actions or plans to resolve them
Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

Nothing to report

Changes that had a significant impact on expenditures
Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to report

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Significant changes in use or care of human subjects

Nothing to report

Significant changes in use or care of vertebrate animals

Nothing to report

Significant changes in use of biohazards and/or select agents

Nothing to report

6. PRODUCTS: *List any products resulting from the project during the reporting period. If there is nothing to report under a particular item, state "Nothing to Report."*

- **Publications, conference papers, and presentations**
  Report only the major publication(s) resulting from the work under this award.

Journal publications. *List peer-reviewed articles or papers appearing in scientific, technical, or professional journals. Identify for each publication: Author(s); title; journal; volume: year; page numbers; status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).*

1. Yan-Ping Yu, Allan Tsung, Silvia Liu, Michael Nalesnik, David Geller, George Michalopoulos and **Jian-Hua Luo** (2019). Detection of fusion transcripts in the serum samples of patients with hepatocellular carcinoma. Oncotarget 10, 3352-3360.

2. Yan-Ping Yu, Peng Liu, Joel Nelson, Ronald L. Hamilton, Rohit Bhargava, George Michalopoulos, Qi Chen, Jun Zhang, Deqin Ma, Arjun Pennathur, Michael Nalesnik, George Tseng and **Jian-Hua Luo** (2019). Identification of recurrent fusion genes across multiple cancer types. Scientific Reports 9:1074. https://doi.org/10.1038/s41598-019-38550-6.

Books or other non-periodical, one-time publications. *Report any book, monograph, dissertation, abstract, or the like published as or in a separate publication, rather than a periodical or series. Include any significant publication in the proceedings of a one-time conference or in the report of a one-time study, commission, or the like. Identify for each one-time publication: author(s); title; editor; title of collection, if applicable; bibliographic information; year; type of publication (e.g., book, thesis or dissertation); status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).*

Nothing to report

Other publications, conference papers and presentations. *Identify any other publications, conference papers and/or presentations not reported above. Specify the status of the publication as noted above. List presentations made during the last year (international, national, local societies, military meetings, etc.). Use an asterisk (*) if presentation produced a manuscript.*

J Luo, G Michalopoulos, Y Yu. Identification of recurrent fusion genes across multiple cancer types. The FASEB Journal 33 (1_supplement), 802.32-802.32.

- **Website(s) or other Internet site(s)**
  List the URL for any Internet site(s) that disseminates the results of the research activities. A short description of each site should be provided. It is not necessary to include the publications already specified above in this section.

Nothing to report

- **Technologies or techniques**
  Identify technologies or techniques that resulted from the research activities. Describe the technologies or techniques were shared.

Nothing to report

- **Inventions, patent applications, and/or licenses**
  Identify inventions, patent applications with date, and/or licenses that have resulted from the research. Submission of this information as part of an interim research performance progress report is not a substitute for any other invention reporting required under the terms and conditions of an award.

Nothing to report

- **Other Products**
  Identify any other reportable outcomes that were developed under this project. Reportable outcomes are defined as a research result that is or relates to a product, scientific advance, or research tool that makes a meaningful contribution toward the understanding, prevention, diagnosis, prognosis, treatment and /or rehabilitation of a disease, injury or condition, or to improve the quality of life. Examples include:
  - *data or databases;*
  - *physical collections;*
  - *audio or video products;*
  - *software;*
  - *models;*
  - *educational aids or curricula;*
  - *instruments or equipment;*
  - *research material (e.g., Germplasm; cell lines, DNA probes, animal models);*
  - *clinical interventions;*
  - *new business creation; and*

- *other.*

Nothing to report

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Provide the following information for: (1) PDs/PIs; and (2) each person who has worked at least one person month per year on the project during the reporting period, regardless of the source of compensation (a person month equals approximately 160 hours of effort). If information is unchanged from a previous submission, provide the name only and indicate "no change".

Example:

Name: *Mary Smith*
Project Role: *Graduate Student*
Researcher Identifier (e.g. ORCID ID): *1234567*
Nearest person month worked: *5*

Contribution to Project: *Ms. Smith has performed work in the area of combined error-control and constrained coding.*
Funding Support: *The Ford Foundation (Complete only if the funding support is provided from other than this award.)*

Name: *Zehua Zuo*
Project Role: *Post-doctoral fellow*
Researcher Identifier (e.g. ORCID ID): *None*
Nearest person month worked: *28*
Contribution to Project: *Dr. Zuo did most work on characterizing Taqman RT-PCR*

Name: *Yanping Yu*
Project Role: *Associate Professor*
Researcher Identifier (e.g. ORCID ID): *none*
Nearest person month worked: *8*
Contribution to Project: *Dr. Yu designed the experiments for Taqman RT-PCR.*

Name: *Wei Sun*
Project Role: *Post-doctoral fellow*
Researcher Identifier (e.g. ORCID ID): *None*
Nearest person month worked: *24*
Contribution to Project: *Dr. Sun conduct most clinical sample microdissection and nucleic acid purification and Sanger's sequencing.*

Name:	*Jianhua Luo*
Project Role:	*Professor*
Researcher Identifier (e.g. ORCID ID):	*None*
Nearest person month worked:	*10.8*
Contribution to Project:	*Dr. Luo oversaw the project.*

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

If the active support has changed for the PD/PI(s) or senior/key personnel, then describe what the change has been. Changes may occur, for example, if a previously active grant has closed and/or if a previously pending grant is now active. Annotate this information so it is clear what has changed from the previous submission. Submission of other support information is not necessary for pending changes or for changes in the level of effort for active support reported previously. The awarding agency may require prior written approval if a change in active other support significantly impacts the effort on the project that is the subject of the project report.

Nothing to report

What other organizations were involved as partners?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe partner organizations – academic institutions, other nonprofits, industrial or commercial firms, state or local governments, schools or school systems, or other organizations (foreign or domestic) – that were involved with the project. Partner organizations may have provided financial or in-kind support, supplied facilities or equipment, collaborated in the research, exchanged personnel, or otherwise contributed.

Provide the following information for each partnership:
Organization Name:
Location of Organization: (if foreign location list country)
Partner's contribution to the project (identify one or more)

- *Financial support;*
- *In-kind support (e.g., partner makes software, computers, equipment, etc., available to project staff);*
- *Facilities (e.g., project staff use the partner's facilities for project activities);*
- *Collaboration (e.g., partner's staff work with project staff on the project);*

- *Personnel exchanges (e.g., project staff and/or partner's staff use each other's facilities, work at each other's site); and*
- *Other.*

Nothing to report

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: *For collaborative awards, independent reports are required from BOTH the Initiating Principal Investigator (PI) and the Collaborating/Partnering PI. A duplicative report is acceptable; however, tasks shall be clearly marked with the responsible PI and research site. A report shall be submitted to* ***https://ers.amedd.army.mil*** *for each unique award.*

QUAD CHARTS: *If applicable, the Quad Chart (available on https://www.usamraa.army.mil) should be updated and submitted with attachments.*

9. APPENDICES: *Attach all appendices that contain information that supplements, clarifies or supports the text. Examples include original copies of journal articles, reprints of manuscripts and abstracts, a curriculum vitae, patent applications, study questionnaires, and surveys, etc.*

SCIENTIFIC REPORTS

OPEN

Identification of recurrent fusion genes across multiple cancer types

Yan-Ping Yu[1], Peng Liu[2], Joel Nelson[3], Ronald L. Hamilton[1], Rohit Bhargava[1], George Michalopoulos[1], Qi Chen[4], Jun Zhang[5], Deqin Ma[6], Arjun Pennathur[7], James Luketich[7], Michael Nalesnik[1], George Tseng[2] & Jian-Hua Luo[1]

Received: 12 June 2018
Accepted: 27 December 2018
Published online: 31 January 2019

Chromosome changes are one of the hallmarks of human malignancies. Chromosomal rearrangement is frequent in human cancers. One of the consequences of chromosomal rearrangement is gene fusions in the cancer genome. We have previously identified a panel of fusion genes in aggressive prostate cancers. In this study, we showed that 6 of these fusion genes are present in 7 different types of human malignancies with variable frequencies. Among them, the CCNH-C5orf30 and TRMT11-GRIK2 gene fusions were found in breast cancer, colon cancer, non-small cell lung cancer, esophageal adenocarcinoma, glioblastoma multiforme, ovarian cancer and liver cancer, with frequencies ranging from 12.9% to 85%. In contrast, four other gene fusions (mTOR-TP53BP1, TMEM135-CCDC67, KDM4-AC011523.2 and LRRC59-FLJ60017) are less frequent. Both TRMT11-GRIK2 and CCNH-C5orf30 are also frequently present in lymph node metastatic cancer samples from the breast, colon and ovary. Thus, detecting these fusion transcripts may have significant biological and clinical implications in cancer patient management.

In the last two decades, significant progress has been made in diagnosing and treating human cancers. However, cancers remain one of the most frequent causes of death in the United States. In 2017, 1,735,350 new cancer cases were diagnosed in the United States[1]. More than 600,000 cancer deaths are projected to occur in the United States in 2018: a death rate second only to cardiovascular diseases. Among human cancers, lung, prostate, breast, liver and colorectal cancers appear the most frequently, accounting for approximately 49% of all human cancers. These five types of cancers are projected to account for 305710 deaths in 2018 or over 50% of all cancer-related deaths in the US. Thus, understanding the mechanisms underlying the development of these cancers is crucial to reduce cancer mortality in the country.

Genome abnormalities are widely present in human cancers[2]. These abnormalities include single nucleotide mutations, copy number changes, chromosomal rearrangement, etc. Indeed, cancer genome abnormalities precede the development of cancer phenotypes[3–6]. Non-malignant tissues adjacent to cancers have been shown to contain similar genomic and transcriptomic changes as neighboring cancer tissues[3–11]. One of the salient abnormalities in the cancer genome is chromosomal rearrangements, which may result in the joining of 2 unrelated genes in the chromosome to produce a fusion gene. The most well-characterized example of fusion gene is the Philadelphia chromosome[12] that joins the N-terminus of BCR with the tyrosine kinase domain of ABL[13]. The resulting chimeric protein has constitutively activated tyrosine kinase activity and transforms benign tissue into malignant one[14]. Several cancer-specific fusion genes have been discovered in prostate cancer samples[15–17]. Some of these fusion genes appear to be transforming[18,19]. Interestingly, one of the fusion genes, MAN2A1-FER, was found in 5 other types of human malignancies, and has been shown to transform normal livers into hepatocellular carcinomas in a short period of time[19]. SLC45A2-AMACR was found independently in bladder cancer[20] and lung cancer cell lines[21]. These findings suggest that fusion genes may have wider implications than initially anticipated. To investigate whether fusion genes play a role in other human malignancies, we analyzed six fusion genes, including TRMT11-GRIK2, MTOR-TP53BP1, CCNH-C5orf30, KDM4-AC011523.2, TMEM135-CCDC67, and LRRC59-FLJ60017, in primary cancer samples from 7 different types of human malignancies and 20 cancer

[1]Departments of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA, 15261, USA. [2]Departments of Biostatistics, University of Pittsburgh School of Medicine, Pittsburgh, PA, 15261, USA. [3]Departments of Urology, University of Pittsburgh School of Medicine, Pittsburgh, PA, 15261, USA. [4]Department of Pharmacology, Toxicology & Therapeutics, University of Kansas, Kansas City, KS, 66160, USA. [5]Department of Medicine, University of Iowa, Iowa City, Iowa, 52242, USA. [6]Department of Pathology, University of Iowa, Iowa City, Iowa, 52242, USA. [7]Departments of Cardiothoracic Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA, 15261, USA. Correspondence and requests for materials should be addressed to J.-H.L. (email: luoj@upmc.edu)

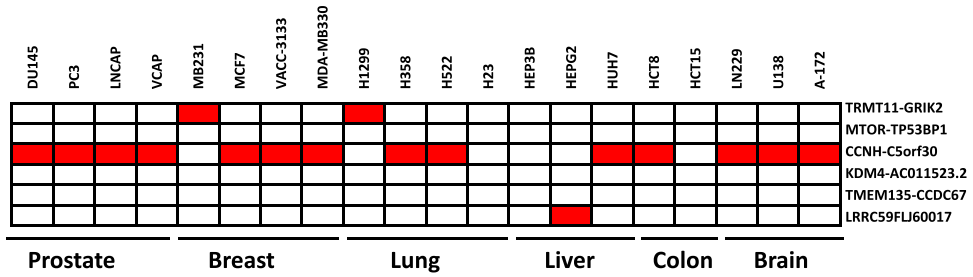

Figure 1. Fusion transcripts are present in human cancer cell lines. Representative cancer cell lines from the lung, liver, prostate, breast, colon and brain were examined for the presence of the TRMT11-GRIK2, CCNH-C5orf30, mTOR-TP53BP1, TMEM135-CCDC67, KDM4-AC011523.2 and LRRC59-FLJ60017 fusion transcripts using quantitative TaqMan RT-PCR. Red denotes positive detection of the fusion transcript, while blank indicates negative detection of the fusion transcript. All positive fusion samples were verified by Sanger sequencing (see Supplemental Figs 1–6).

cell lines originating from 6 human cancers. These fusion genes are present in human cancers with variable frequencies, suggesting a much wider role for these cancer-specific fusion genes in the development of human malignancies.

Materials and Methods

Tissue samples. The 536 tissue specimens used in the study consist of 101 non-small cell lung cancers, 61 ovarian cancers, 60 colon cancers, 70 liver cancers, 150 glioblastoma, 60 breast cancers, and 34 esophageal adenocarcinomas. These samples were obtained from the University of Pittsburgh Tissue Bank in compliance with institutional regulatory guidelines (Supplemental Table 1 through 7). The informed consent exemptions and protocol were approved by the Institution Review Board of University of Pittsburgh. Cancer cells were obtained by macro-dissection. Esophageal cancer specimens were from a prospective IRB approved protocol from University of Pittsburgh and were frozen tissue samples. Sixteen non-small cell lung cancer samples were obtained from the University of Kansas. Twenty-eight non-small cell lung cancer samples were obtained from the University of Iowa. All informed consent exemptions and protocols were approved by the Institution Review Board of the University of Kansas or University of Iowa. All 20 cell lines used in the study were purchased from the American Type Cell Culture (ATCC, Inc., Manassas, VA, USA) and were cultured and maintained following the manufacturer's recommendations.

RNA extraction, cDNA synthesis and detection of fusion genes. Formalin-fixed paraffin-embedded (FFPE) tissue blocks of each sample were cut for multiple unstained slides. One of these slides was stained with hematoxylin and eosin. The cancer regions were circled by pathologists and macrodissected. Total RNA was extracted using trizol to lyse the cancer tissues (Invitrogen, CA). First strand cDNA was synthesized using ~2 µg of RNA from each sample, random hexamers and Superscript II™ (Invitrogen, Inc, CA) at 42 °C for 2 hours. One microliter each cDNA sample was used for TaqMan PCR reactions with 50 heat cycles at 94 °C for 30 seconds, 61 °C for 30 seconds, and 72 °C for 30 seconds using primers and probes specific for CCNH-C5orf30 (AAAGTTATTTATCAGAGAGTCTGATGCTG/CTGTTCTACTCCAGGTATTTTCATTATATC; TaqMan probe, 5′-/56-FAM/ACAGGCAAG/ZEN/TTCTGTTCTCTTTCAGCA/3IABkFQ/-3′), mTOR-TP53BP1 (TGATAGACCAGTCCCGGGATG/CCACTGACATTCCCAGAACAAG; TaqMan probe, 5′-/56-FAM/TGTCAGCCT/ZEN/GTCAGAATCCAAGTCAAG/3IABkFQ/-3′), TRMT11-GRIK2 (GCGCTGTCGTGTACCCTTAAC/GAATGCAAGTTCCTCAGCTCC; TaqMan probe, 5′-/56-FAM/CGGAACTCC/ZEN/AGATGCTCCTGCG/3IABkFQ/-3′), LRRC59-FLJ60017 (GTGACTGCTTGGATGAGAAGC/CCCTCCTCTGGTTTGTTGTTG; TaqMan probe, 5′-/56-FAM/CAGTGTGCA/ZEN/AACAAGGTGACTGGAAG/3IABkFQ/-3′), TMEM135-CCDC67 (CAGCTGTCATGGAAGTTCAGAC/CCTCATTCTTTCCTGCTCAGAG; TaqMan probe, 5′-/56-FAM/AGTTCCTTT/ZEN/TAAGACTCACCAAGGGCAA/3IABkFQ/-3′), KDM4- AC011523.2 (AGACCACCTTCGCCTGGCAC/TCTCTCTCAGATCCAGGCTTG; TaqMan probe, 5′-/56-FAM/ACAGCATCA/ZEN/ACTACCTGCACTTTGGG/3IABkFQ/-3′), and β-actin (ACCCCACTTCTCTCTAAGGAG/GCAATGCTATCACCTCCCCTG; TaqMan probe, 5′-/56-FAM/CCAGTCCTC/ZEN/TCCCAAGTCCACAC/3IABkFQ/-3′) in a thermocycler (Eppendorf Realplex™ thermocycler). A negative control and synthetic positive control were included in each batch of reactions. The PCR products were gel purified and Sanger-sequenced on 20% positive samples.

Results

In our previous studies[16,17], we have characterized eight fusion genes identified in aggressive prostate cancer samples. Additional analyses showed that one of the fusion genes, MAN2A1-FER, is frequently present in 5 other types of human malignancies[19]. To expand our analyses of other fusion genes in the panel, we analyzed TRMT11-GRIK2, MTOR-TP53BP1, CCNH-C5orf30, KDM4-AC011523.2, TMEM135-CCDC67, and LRRC59-FLJ60017 in 20 cancer cell lines from six different human malignancies (Fig. 1). The TRMT11-GRIK2 fusion transcript was identified in the MD-MB-231breast cancer cell line and H1299 lung cancer cell line, while CCNH-C5orf30 was positive in 14 of 20 cancer cell lines, including all the prostate cancer cell lines tested (PC3,

DU145, LNCaP and VCaP), 3 of 4 breast cancer cell lines (MCF7, VACC-3133 and MDA-MB330), 2 of 4 lung cancer cell lines (H358 and H522), 1 of 3 liver cancer cell lines (HepG2), 1 of 2 colon cancer cell lines (HCT8) and 3 of 3 GBM cell lines (LN229, U138 and A-172). LRRC59-FLJ60017 was also present in the HepG2 liver cancer cell line. These results suggest that these fusion genes are not specific for prostate cancer and may be present in primary cancer samples from a variety of human malignancies.

To investigate whether any of the above-mentioned fusion genes has a role in human cancers, quantitative TaqMan qRT-PCR reactions using primers and probes specific for each fusion gene were performed on primary human cancer samples representing seven different types of human malignancies. Our results showed that TRMT11-GRIK2 is present in all seven types of human malignancies (Table 1, Fig. 2 and Supplemental Figs 1–6)), including breast cancer (41/60, 68.33%), colon cancer (25/60, 41.7%), esophageal adenocarcinoma (9/34, 26.5%), hepatocellular carcinoma (9/70, 12.9%), ovarian adenocarcinoma (28/61, 45.9%), glioblastoma multiforme (26/150, 17.3%) and non-small cell lung cancer (39/141, 27.7%). The CCNH-C5orf30 transcript was also detected in 7 different types of human cancers with the following relatively high frequencies: 85% (51/60) in breast cancer, 43% (26/60) in colon cancer, 50.8% (31/61) in ovarian cancer, 67.6% (23/34) in esophageal adenocarcinoma, 41.8% (59/141) in non-small cell lung cancer, 37% (26/70) in liver cancer and 53% (80/150) in glioblastoma multiforme. In contrast, mTOR-TP53BP1was only detected in five different types of human malignancies and with significantly lower frequencies, including breast cancer (10/60, 16.7%), colon cancer (4/60, 6.7%), ovarian adenocarcinoma (4/61, 6.6%), glioblastoma multiforme (7/150, 4.7%) and lung cancer (8/141, 5.7%). LRRC59-FLJ60017 was present in four different types of human cancers, esophageal adenocarcinoma (3/34, 8.8%), ovarian adenocarcinoma (4/61, 6.6%), glioblastoma multiforme (16/150, 10.7%), and non-small cell lung cancer (33/141, 23.4%). Only two glioblastoma multiforme and one lung cancer samples were positive for the TMEM135-CCDC67 fusion gene. The KDM4-AC011523.2 fusion was only found in 3 breast cancer and 3 lung cancer samples.

Interestingly, ductal type breast cancers positive for TRMT11-GRIK2 were associated with a lower likelihood of developing local lymph node metastasis (46.7% versus 93.3%, $p = 0.014$). Liver cancers positive for TRMT11-GRIK2 were also associated with a higher rate of overall survival (41.7% versus 6.9%, $p = 0.006$). KDM4-AC011523.2 was only detected in lobular type breast cancer and adenocarcinoma of the lung. Patients with lobular breast cancers positive for mTOR-TP53BP1 were also less likely to have lymph node metastasis (0% versus 40%, $p = 0.017$). CCNH-C5orf30 was more frequent in lung cancer adenocarcinomas versus squamous type (67.7% versus 34.8%, $p = 0.001$) and colon cancer at advanced stages at the time of diagnosis (52.3% versus 18.8%, $p = 0.037$).

To investigate whether fusion genes are also present in human cancer metastatic lesions, breast cancer, colon cancer and ovarian cancer samples with matched lymph node metastases were analyzed (Fig. 3). Twenty-six of 30 metastatic breast cancers in lymph nodes were positive for TRMT11-GRIK2, including seven metastatic cancers whose matched primary breast cancers were negative for the fusion. For colon cancers, the matched status of TRMT11-GRIK2 between primary cancer samples and lymph node metastases was 78.5% (11/14Eleven lymph node metastases were exactly matched with the status of the primary colon cancer samples, while two samples of lymph node metastases were found negative for TRMT11-GRIK2 fusion. One lymph node metastasis was found positive for the fusion gene while the matched primary sample was negative (Fig. 3). For ovarian adenocarcinomas, nine metastatic lesions were found to contain the TRMT11-GRIK2 fusion gene, matching all the primary samples. However, four lymph node metastases contained no TRMT11-GRIK2 fusion gene while the matched primary cancer samples were positive. One lymph node metastasis gained the TRMT11-GRIK2 fusion over the primary cancer sample. For CCNH-C5orf30, the matching rate of primary breast cancer with lymph node metastases was 62%, while the matching rates for ovarian cancer and colon cancer with their corresponding lymph node metastases were 72% and 73%, respectively. For mTOR-TP53BP1, two of 3 lymph node metastases retained the fusion in colon and ovarian cancers. Additionally, two of 2 ovarian cancer lymph node metastases retained the LRRC59-FLJ60017 fusion. These results suggest significant heterogeneity among the cancer samples. However, most fusion genes were retained in metastatic lesions.

Discussion

Gene fusions are the result of recombination of two unrelated genes. Fusion events can also involve genes that have similar biological roles (for example, between two genes that both have roles in transcription, such as the ESR1-YAP1 driver fusion in breast cancer)[22]. Almost all cancer-specific gene fusions are the result of chromosomal rearrangements or translocations[17]. There is increasing evidence suggesting that gene fusions are some of the key drivers of human cancer development. New gene fusion events have been discovered in prostate cancer[15–17], breast cancer[23–25], NSCLC[19,26,27], colon cancer[28–30], glioblastoma multiforme[19,31,32], liver cancer[19,33–35] and ovarian cancer[19,36,37]. Some of these fusion genes appear to play driver roles in the aggressive behaviors of these cancers[19,23,26,31]. Gene fusions produce two possible outcomes for the genes involved. One outcome is a gain of function due to the loss of the regulatory domain in the protein, so the enzymatic domain of the same protein becomes hyper-activated. BCR-ABL and MAN2A1-FER are examples of gain of function fusion genes, leading to hyper-activation of the fusion partner. In-frame fusions can also contribute to disease pathogenesis by creating a fusion protein that contains not only complementary functions encoded by each partner gene, but also has neomorphic properties. ESR1-YAP1[22], a driver fusion found in advanced breast cancer, generates a hyperactive transcription factor through the combination of the ESR1 part that provides domains necessary for DNA binding, dimerization, and nuclear localization and the YAP1 part that provides components for transcriptional activation. ESR1-YAP1 is able to drive expression of genes that promote metastatic biology, a function that full length wild-type ESR1 lacks. The other outcome is a loss of function due to the truncation of the head gene and/or complete elimination of the open reading frame of the tail gene, such

Cancer	TRMT11-GRIK2 MTOR-TP53BP1 CCNH-C5orf30 KDM4-AC011523.2 TMEM135-CCDC67 LRRC59-FLJ60017					
Breast cancer						
Ductal Type	21/30	4/30	26/30	0/30	0/30	0/30
With LN met	7/15	0/15	11/15	0/15	0/15	0/15
Without LN met	14/15	4/15	15/15	0/15	0/15	0/15
Lobular	20/30	6/30	25/30	3/30	0/30	0/30
With LN met	12/15	0/15	10/15	0/15	0/15	0/15
Without LN met	8/15	6/15	15/15	3/15	0/15	0/15
Non-recurrence	36/53	10/53	47/53	3/53	0/53	0/53
Recurrence	5/7	0/7	4/7	0/7	0/7	0/7
Stage 1	21/33	8/33	30/33	3/33	0/33	0/33
Stage 2	11/16	2/16	12/16	0/16	0/16	0/16
Stage 3	6/9	0/9	7/9	0/9	0/9	0/9
Stage 4	1/2	0/2	2/2	0/2	0/2	0/2
Grade 1	5/8	2/8	8/8	0/8	0/8	0/8
Grade 2	26/38	8/38	33/38	3/38	0/38	0/38
Grade 3	10/14	0/14	10/14	0/14	0/14	0/14
ER positive	34/51	9/51	45/51	3/51	0/51	0/51
ER negative	7/9	1/9	6/9	0/9	0/9	0/9
PR positive	29/44	7/44	39/44	3/44	0/44	0/44
PR negative	12/16	3/16	12/16	0/16	0/16	0/16
HER2 amp+	2/3	0/3	3/3	0/3	0/3	0/3
HER2 amp−	39/57	10/57	48/57	3/57	0/57	0/57
Triple negative	7/9	1/9	6/9	0/9	0/9	0/9
Colon cancer						
With LN met	14/30	3/30	15/30	0/30	0/30	0/30
Without LN met	11/30	1/30	11/30	0/30	0/30	0/30
Recurrence 5-yr	3/5	1/5	4/5	0/5	0/5	0/5
Non-recurrence	22/55	3/55	22/55	0/55	0/55	0/55
Stage 1	1/3	0/3	1/3	0/3	0/3	0/3
Stage 2	3/13	0/13	2/13	0/13	0/13	0/13
Stage 3	20/42	3/42	21/42	0/42	0/42	0/42
Stage 4	1/2	1/2	2/2	0/2	0/2	0/2
Grade I	1/2	0/2	1/2	0/2	0/2	0/2
Grade II	19/51	3/51	21/51	0/51	0/51	0/51
Grade III	5/7	1/7	4/7	0/7	0/7	0/7
Ovarian cancer						
With LN met	14/30	3/30	18/30	0/30	0/30	2/30
Without LN met	14/31	1/31	13/31	0/31	0/31	2/31
Non-recurrence	16/33	1/33	15/33	0/33	0/33	2/33
Recurrence 5-yrs	12/28	3/28	16/28	0/28	0/28	2/28
Survive >5 yrs	16/34	1/34	18/34	0/34	0/34	2/34
Death from Ca	12/27	3/27	13/27	0/27	0/27	2/27
Stage 1	11/22	1/22	9/22	0/22	0/22	2/22
Stage 2	2/4	0/4	1/4	0/4	0/4	0/4
Stage 3	15/35	3/35	21/35	0/35	0/35	2/35
Grade I	5/7	1/7	2/7	0/7	0/7	0/7
Grade II	2/8	0/8	4/8	0/8	0/8	0/8
Grade III	16/34	2/34	16/34	0/34	0/34	3/34
Grade IV	5/12	1/12	9/12	0/12	0/12	1/12
Esophageal Adenocarcinoma						
	9/34	0/34	23/34	0/34	0/34	3/34
Non-small cell lung cancer						
UPMC cohort	33/101	7/101	31/101	0/101	1/101	12/101
Stage 1	10/24	3/24	8/24	0/24	0/24	3/24
Stage 2	19/62	4/62	21/62	0/62	1/62	7/62
Stage 3	2/5	0/5	0/5	0/5	0/5	0/5

Continued

Cancer	TRMT11-GRIK2	MTOR-TP53BP1	CCNH-C5orf30	KDM4-AC011523.2 TMEM135-CCDC67	LRRC59-FLJ60017	
Stage 4	2/10	0/10	2/10	0/10	0/10	2/10
With LN met	10/39	2/39	12/39	0/39	0/39	5/39
Without LN met	23/62	5/62	19/62	0/62	1/62	7/62
Survival 5-yr	9/34	2/34	10/34	0/34	0/34	6/34
Death 5 yrs	22/67	5/67	21/67	0/67	1/67	6/67
Ad Ca	0/0	0/0	0/0	0/0	0/0	0/0
Sq Ca	33/101	7/101	31/101	0/101	1/101	12/101
Kansas cohort	1/16	0/16	15/16	0/16	2/16	12/16
Stage I	0/5	0/5	5/5	0/5	0/5	4/5
Stage II	1/11	0/11	10/11	0/11	2/11	8/11
Ad Ca	1/8	0/8	8/8	0/8	0/8	6/8
Sq Ca	0/8	0/8	7/8	0/8	2/8	6/8
Iowa cohort	5/24	1/24	13/24	3/24	0/24	9/24
Stage I	2/8	0/8	4/8	0/8	0/8	3/8
Stage II	2/7	1/7	5/7	2/7	0/7	5/7
Stage III	0/6	0/6	3/6	1/6	0/6	1/6
Stage IV	1/3	0/3	1/3	0/3	0/3	0/3
With met	3/16	1/16	7/16	2/16	0/16	6/16
Without met	2/8	0/8	6/8	1/8	0/8	3/8
Ad Ca	5/23	1/23	13/23	3/23	0/23	9/23
Sq Ca	0/0	0/0	0/0	0/0	0/0	0/0
Sarcomatoid	0/1	0/1	0/1	0/1	0/1	0/1
Liver cancer						
Total	9/70	0/70	26/70	0/70	0/70	0/70
Survive 5 yrs	5/12	0/12	7/12	0/12	0/12	0/12
Death	4/58	0/58	19/58	0/58	0/58	0/58
Non-recurrence	5/34	0/34	13/34	0/34	0/34	0/34
Recurrence	4/36	0/36	13/36	0/36	0/36	0/36
Glioblastoma multiforme						
	26/150	7/150	80/150	0/150	2/150	16/150

Table 1. Distribution of fusion genes in colon cancer, breast cancer, esophageal adenocarcinoma, liver cancer, ovarian adenocarcinoma, glioblastoma multiforme and non-small cell lung cancer.

as with TRMT11-GRIK2 and mTOR-TP53BP1. For both these instances, tumor suppressor activities of GRIK2[38] and TP53BP1[39] are lost due to the loss of protein translation.

Among the prostate cancer fusion genes identified in our previous study[16], two of these fusion genes (SLC45A2-AMACR and MAN2A1-FER) were also found in other types of human cancers[19–21], suggesting that these gene fusions are not specific to prostate cancer but may be widely present in human cancers. MAN2A1-FER was found to be the driver for liver cancer in mouse[19]. The expression of this fusion gene induced spontaneous liver cancer in mice by ectopically phosphorylating the EGFR extracellular domain and activating its signaling pathways[19]. Although the biological roles of the other fusion genes remain unknown, five of these gene fusion events (TMEM135-CCDC67, mTOR-TP53BP1, LRRC59-FLJ10067, KDM4-AC011523.2 and TRMT11-GRIK2) eliminate the open-reading frames in the tail genes and thus produce gene knockouts. CCDC67[40], TP53BP1[39,41,42] and GRIK2[38] are tail genes that contain tumor suppressor activity. The TMEM135-CCDC67, mTOR-TP53BP1 and TRMT11-GRIK2 gene fusions are equivalent to the functional deletion of CCDC67, TP53BP1 and GRIK2, respectively. Deletions of these genes have been shown to promote the aggressive behaviors of cancers[38–42]. These gene fusion events may have significant biological implications in the development of cancer.

Based on the current analyses, TRMT11-GRIK2 and CCNH-C5orf30 are probably some of the most widely distributed gene fusions in human malignancies, being present in at least eight types of human malignancies. TRMT11-GRIK2 has frequencies ranging from 12.9% in liver cancer to 68.3% in breast cancer. This fusion gene is also found in a breast cancer cell line and a lung cancer cell line. CCNH-C5orf30 is also very frequent among different types of cancers and their corresponding cancer cell lines. In contrast, the positive rates of the other four fusion genes are much less frequent. The mechanism underlying the disparity of frequencies of these fusion genes is not clear. Although there is some correlation between the distance between the partner genes and the frequencies (both mTOR-TP53BP1 and LRRC59-FLJ60017 gene fusions have their gene partners located in different chromosomes, while both TRMT11-GRIK2 and CCNH-C5orf30 are located in the same chromosome with distances less than 24 MB), TMEM135 and CCDC67 are only separated by 6 MB, and the TMEM135-CCDC67

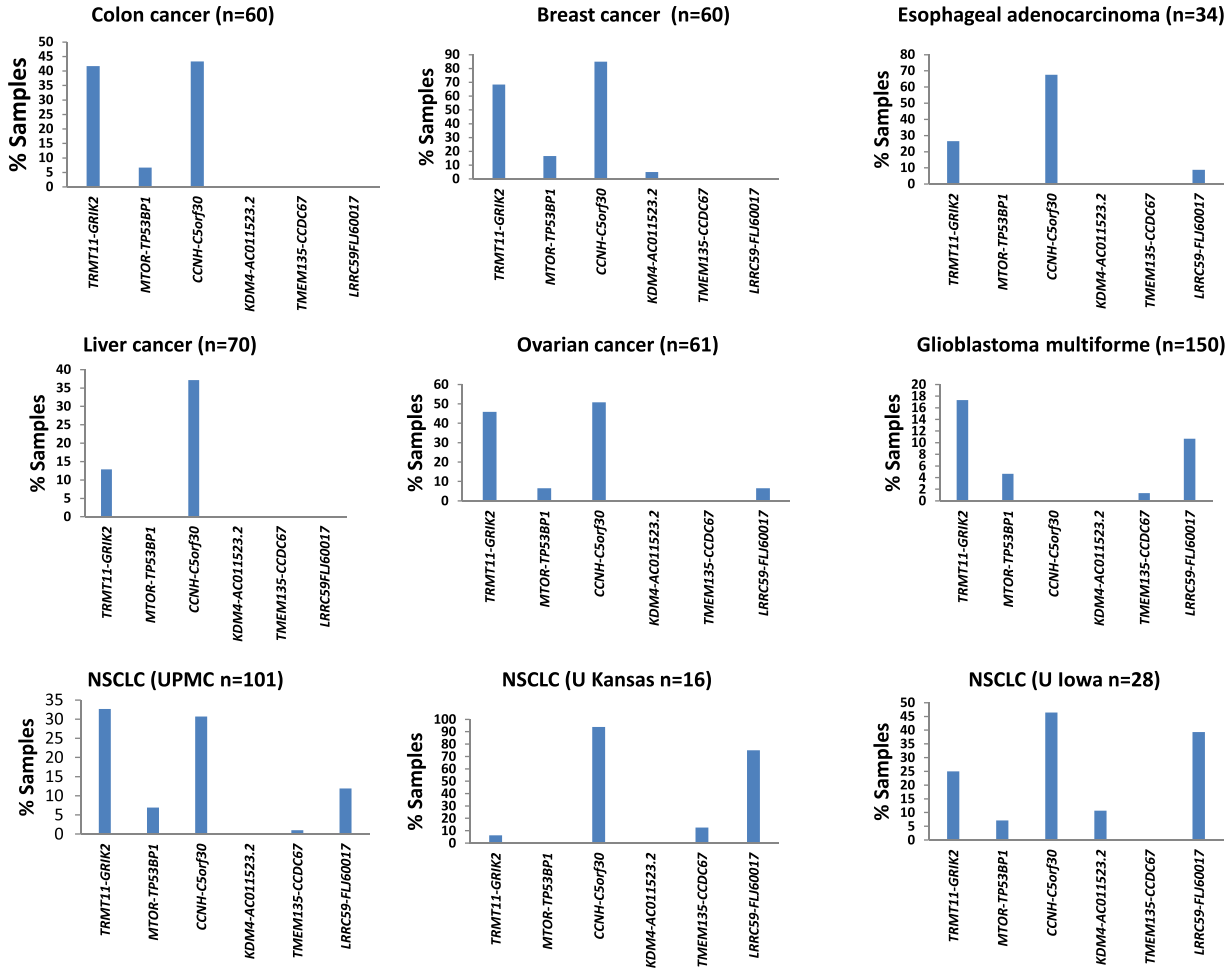

Figure 2. Frequency of gene fusion in primary human malignant samples. Real-time TaqMan qRT-PCR reactions were performed on formalin-fixed paraffin-embedded samples from breast cancer, colon cancer, esophageal adenocarcinoma, ovarian cancer, hepatocellular carcinoma, non-small cell lung cancer and glioblastoma multiforme. The number of samples from each cohort are indicated. Sanger sequencing was performed on 20–100% of samples that were found to be positive for the fusion genes (see Supplemental Figs 1–6).

fusion is exceeding rare in cancers. Nevertheless, the presence of these gene fusions suggests that chromosomal recombination and translocation are probably some of the most frequent events in human cancers.

TRMT11 is a tRNA methyltransferase. The protein is essential for m^2G formation at position 10 in tRNA[43]. This methylation event is required for tRNA stability and translation activity[44]. In contrast, GRIK2 encodes a glutamate receptor[45] and was shown to possess tumor suppressor activity[38]. The process of chromosomal recombination between TRMT11 and GRIK2 to create the TRMT11-GRIK2 gene fusion destroys the open-reading frames of both genes and produces functional knockouts of these two proteins. The absence of TRMT11 may produce less efficient and unstable translation of mRNA into protein in cancer cells due to tRNA defects. Alternatively, protein translation may be repressed. The lack of GRIK2 may accelerate cell cycle progression and promote cell migration. Thus, cells with the TRMT11-GRIK2 gene fusion may be unstable and tumorigenic.

CCNH is an important member of the cyclin family. It complexes with cdk7-MAT1 and is a component in the TFIIH and RNA polymerase complexes[46,47]. Thus, CCNH is a critical regulator for the processes of transcription and cell cycle progression. The CCNH-C5orf30 gene fusion produces a truncated cyclin H protein with the deletion of its H5′ and HC domains. A study showed that CCNH mutants lacking the HC domain do not activate cdk7[47]. As a result, the truncated CCNH from the gene fusion may have a negative impact on the functions of RNA polymerase and TFIIH. C5orf30 was also shown to inhibit the generation of cytokines involved in inflammation, such as TNF and IL1, and promote the expression of anti-inflammatory cytokines, such as IL10, in rheumatoid arthritis synovial fibroblasts[48,49]. The CCNH-C5orf30 fusion transcript contains an intact C5orf30 opening reading frame. The CCNH-C5orf30 gene fusion places C5orf30 expression under the CCNH promoter and may promote its expression. Over-expression of C5orf30 in cancer cells may help to fend off immune responses targeting the cancers.

The wide presence of the 6 fusion genes in a variety of human malignancies may provide significant utility for clinical cancer diagnosis and therapeutic targeting. The presence of these fusion genes in metastatic cancer

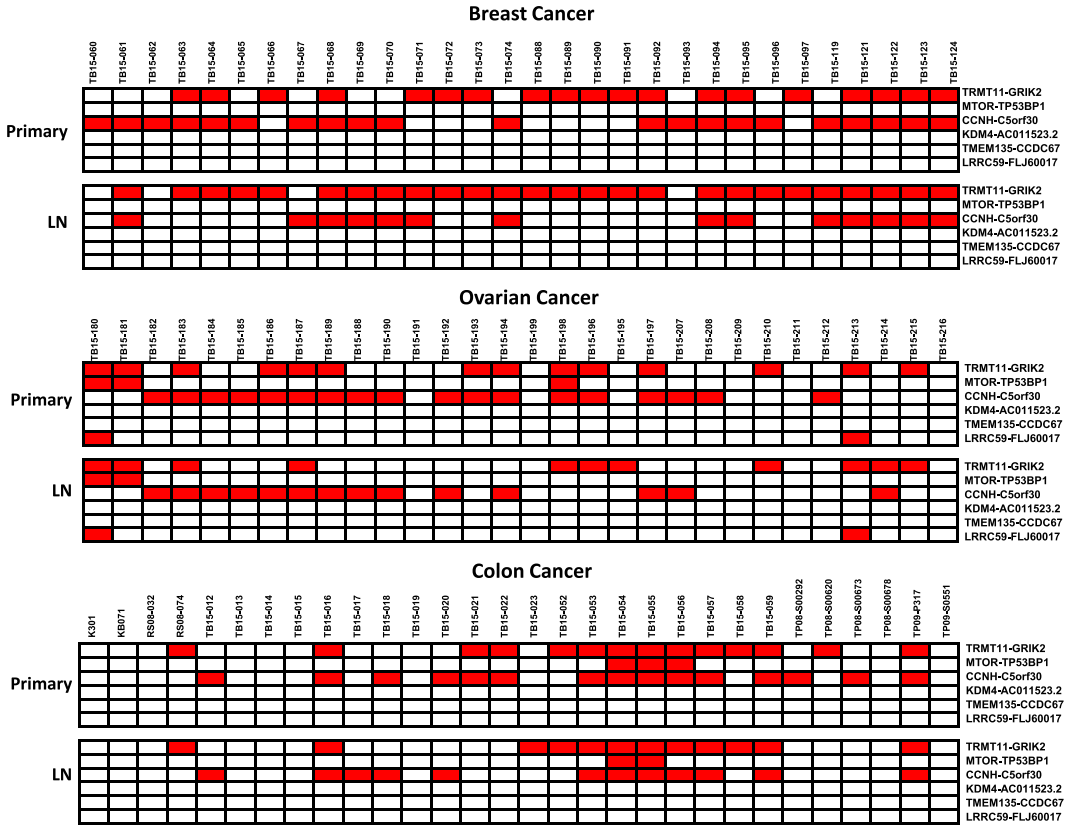

Figure 3. Fusion transcripts in metastatic lymph node samples. Real-time TaqMan qRT-PCR reactions were performed on formalin-fixed paraffin-embedded samples from breast cancer (top), ovarian cancer (middle) and colon cancer (bottom) and their corresponding lymph node metastases. Both primary (top) and matched lymph node (bottom) cancer samples are shown. Red indicates positive detection of the fusion transcript, while blank indicates negative detection of the fusion transcript. Sanger sequencing was performed on 20–100% of samples that were found to be positive for the fusion genes (see Supplemental Figs 1–6).

samples can be used in clinical follow-up studies to analyze the recurrence of human cancers. If these fusion genes are present in grey zone biopsy samples, it may also help to confirm or to make a correct diagnosis. Furthermore, the chromosomal breakpoints of significant numbers of these fusion genes have been identified. These chromosome breakpoints not only serve as cancer markers but also provide unique opportunities to treat human cancers using genome editing technologies[50]. When multiple fusion gene breakpoints are present in the same cancer cells, multi-targeting at these chromosomal breakpoints may significantly enhance the efficiency of genome editing treatments targeting the cancer cells.

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA: a cancer journal for clinicians* **68**, 7–30, https://doi.org/10.3322/caac.21442 (2018).
2. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
3. Yu, Y. P. *et al.* Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J Clin Oncol* **22**, 2790–2799 (2004).
4. Luo, J. H. *et al.* Transcriptomic and genomic analysis of human hepatocellular carcinomas and hepatoblastomas. *Hepatology (Baltimore, Md* **44**, 1012–1024 (2006).
5. Yu, Y. P. *et al.* Genome abnormalities precede prostate cancer and predict clinical relapse. *The American journal of pathology* **180**, 2240–2248 (2012).
6. Luo, J. H. Gene expression alterations in human prostate cancer. *Drugs Today (Barc)* **38**, 713–719 (2002).
7. Luo, J. H. & Yu, Y. P. Genetic factors underlying prostate cancer. *Expert reviews in molecular medicine* **5**, 1–26 (2003).
8. Luo, J. H. *et al.* Genome-wide methylation analysis of prostate tissues reveals global methylation patterns of prostate cancer. *The American journal of pathology* **182**, 2028–2036 (2013).
9. Yu, Y. P. *et al.* Genomic Copy Number Variations in the Genomes of Leukocytes Predict Prostate Cancer Clinical Outcomes. *PloS one* **10**, e0135982 (2015).
10. Yu, Y. P. & Luo, J. H. Pathological factors evaluating prostate cancer. *Histology and histopathology* **22**, 1291–1300 (2007).
11. Yu, Y. P. *et al.* Whole-Genome Methylation Sequencing Reveals Distinct Impact of Differential Methylations on Gene Transcription in Prostate Cancer. *The American journal of pathology* (2013).
12. Benson, E. S. Leukemia and the Philadelphia chromosome. *Postgrad Med* **30**, A22–A28 (1961).
13. Davis, R. L., Konopka, J. B. & Witte, O. N. Activation of the c-abl oncogene by viral transduction or chromosomal translocation generates altered c-abl proteins with similar *in vitro* kinase properties. *Molecular and cellular biology* **5**, 204–213 (1985).
14. Prywes, R., Foulkes, J. G. & Baltimore, D. The minimum transforming region of v-abl is the segment encoding protein-tyrosine kinase. *Journal of virology* **54**, 114–122 (1985).

15. Tomlins, S. A. *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science New York, N.Y* **310**, 644–648 (2005).
16. Yu, Y. P. *et al.* Novel fusion transcripts associate with progressive prostate cancer. *The American journal of pathology* **184**, 2840–2849 (2014).
17. Luo, J. H. *et al.* Discovery and Classification of Fusion Transcripts in Prostate Cancer and Normal Prostate Tissue. *The American journal of pathology* (2015).
18. Zong, Y. *et al.* ETS family transcription factors collaborate with alternative signaling pathways to induce carcinoma from adult murine prostate cells. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 12465–12470 (2009).
19. Chen, Z. H. *et al.* MAN2A1-FER Fusion Gene Is Expressed by Human Liver and Other Tumor Types and Has Oncogenic Activity in Mice. *Gastroenterology*, https://doi.org/10.1053/j.gastro.2016.12.036 (2017).
20. Yoshihara, K. *et al.* The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* **34**, 4845–4854, https://doi.org/10.1038/onc.2014.406 (2015).
21. Klijn, C. *et al.* A comprehensive transcriptional portrait of human cancer cell lines. *Nature biotechnology* **33**, 306–312, https://doi.org/10.1038/nbt.3080 (2015).
22. Li, S. *et al.* Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell reports* **4**, 1116–1130, https://doi.org/10.1016/j.celrep.2013.08.022 (2013).
23. Matissek, K. J. *et al.* Expressed Gene Fusions as Frequent Drivers of Poor Outcomes in Hormone Receptor-Positive Breast Cancer. *Cancer discovery* **8**, 336–353, https://doi.org/10.1158/2159-8290.CD-17-0535 (2018).
24. Lei, J. T. *et al.* Functional Annotation of ESR1 Gene Fusions in Estrogen Receptor-PositiveBreast Cancer. *Cell reports* **24**, 1434–1444e1437, https://doi.org/10.1016/j.celrep.2018.07.009 (2018).
25. Kim, J. *et al.* Recurrent fusion transcripts detected by whole-transcriptome sequencing of 120 primary breast cancer samples. *Genes, chromosomes & cancer* **54**, 681–691, https://doi.org/10.1002/gcc.22279 (2015).
26. Soda, M. *et al.* Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566, https://doi.org/10.1038/nature05945 (2007).
27. Takeuchi, K. *et al.* RET, ROS1 and ALK fusions in lung cancer. *Nature medicine* **18**, 378–381, https://doi.org/10.1038/nm.2658 (2012).
28. Davidsen, J. *et al.* The VTI1A-TCF4 colon cancer fusion protein is a dominant negative regulator of Wnt signaling and is transcriptionally regulated by intestinal homeodomain factor CDX2. *PloS one* **13**, e0200215, https://doi.org/10.1371/journal.pone.0200215 (2018).
29. Sun, Z. *et al.* The novel fusion transcript NR5A2-KLHL29FT is generated by an insertion at the KLHL29 locus. *Cancer* **123**, 1507–1515, https://doi.org/10.1002/cncr.30510 (2017).
30. Nome, T. *et al.* High frequency of fusion transcripts involving TCF7L2 in colorectal cancer: novel fusion partner and splice variants. *PloS one* **9**, e91264, https://doi.org/10.1371/journal.pone.0091264 (2014).
31. Singh, D. *et al.* Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science New York, N.Y* **337**, 1231–1235, https://doi.org/10.1126/science.1220834 (2012).
32. Bao, Z. S. *et al.* RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas. *Genome research* **24**, 1765–1773, https://doi.org/10.1101/gr.165126.113 (2014).
33. Kastenhuber, E. R. *et al.* DNAJB1-PRKACA fusion kinase interacts with beta-catenin and the liver regenerative response to drive fibrolamellar hepatocellular carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 13076–13084, https://doi.org/10.1073/pnas.1716483114 (2017).
34. Honeyman, J. N. *et al.* Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. *Science (New York, N.Y* **343**, 1010–1014, https://doi.org/10.1126/science.1249484 (2014).
35. Huang, Y. *et al.* Transcriptome profiling identifies a recurrent CRYL1-IFT88 chimeric transcript in hepatocellular carcinoma. *Oncotarget* **8**, 40693–40704, https://doi.org/10.18632/oncotarget.17244 (2017).
36. Earp, M. A. *et al.* Characterization of fusion genes in common and rare epithelial ovarian cancer histologic subtypes. *Oncotarget* **8**, 46891–46899, https://doi.org/10.18632/oncotarget.16781 (2017).
37. Kannan, K. *et al.* Recurrent BCAM-AKT2 fusion gene leads to a constitutively activated AKT2 fusion kinase in high-grade serous ovarian carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E1272–1277, https://doi.org/10.1073/pnas.1501735112 (2015).
38. Wu, C. S. *et al.* Glutamate receptor, ionotropic, kainate 2 silencing by DNA hypermethylation possesses tumor suppressor function in gastric cancer. *International journal of cancer* **126**, 2542–2552, https://doi.org/10.1002/ijc.24958 (2010).
39. Wang, B., Matsuoka, S., Carpenter, P. B. & Elledge, S. J. 53BP1, a mediator of the DNA damage checkpoint. *Science (New York, N.Y* **298**, 1435–1438, https://doi.org/10.1126/science.1076182 (2002).
40. Park, S. J. *et al.* Epigenetic alteration of CCDC67 and its tumor suppressor function in gastric cancer. *Carcinogenesis* **33**, 1494–1501, https://doi.org/10.1093/carcin/bgs178 (2012).
41. Schultz, L. B., Chehab, N. H., Malikzay, A. & Halazonetis, T. D. p53 binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. *The Journal of cell biology* **151**, 1381–1390 (2000).
42. Ward, I. M., Minn, K., van Deursen, J. & Chen, J. p53 Binding protein 53BP1 is required for DNA damage responses and tumor suppression in mice. *Molecular and cellular biology* **23**, 2556–2563 (2003).
43. Bourgeois, G., Marcoux, J., Saliou, J. M., Cianferani, S. & Graille, M. Activation mode of the eukaryotic m2G10 tRNA methyltransferase Trm11 by its partner protein Trm112. *Nucleic acids research* **45**, 1971–1982, https://doi.org/10.1093/nar/gkw1271 (2017).
44. El Yacoubi, B., Bailly, M. & de Crecy-Lagard, V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu Rev Genet* **46**, 69–95, https://doi.org/10.1146/annurev-genet-110711-155641 (2012).
45. Paschen, W., Blackstone, C. D., Huganir, R. L. & Ross, C. A. Human GluR6 kainate receptor (GRIK2): molecular cloning, expression, polymorphism, and chromosomal assignment. *Genomics* **20**, 435–440, https://doi.org/10.1006/geno.1994.1198 (1994).
46. Schneider, E., Kartarius, S., Schuster, N. & Montenarh, M. The cyclin H/cdk7/Mat1 kinase activity is regulated by CK2 phosphorylation of cyclin H. *Oncogene* **21**, 5031–5037, https://doi.org/10.1038/sj.onc.1205690 (2002).
47. Andersen, G. *et al.* The structure of cyclin H: common mode of kinase activation and specific features. *The EMBO journal* **16**, 958–967, https://doi.org/10.1093/emboj/16.5.958 (1997).
48. Muthana, M. *et al.* C5orf30 is a negative regulator of tissue damage in rheumatoid arthritis. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 11618–11623, https://doi.org/10.1073/pnas.1501947112 (2015).
49. Onuora, S. Rheumatoid arthritis: C5orf30 regulates severity of tissue destruction in RA. *Nat Rev Rheumatol* **11**, 622, https://doi.org/10.1038/nrrheum.2015.139 (2015).
50. Chen, Z. H. *et al.* Targeting genomic rearrangements in tumor cells through Cas9-mediated insertion of a suicide gene. *Nature biotechnology* **35**, 543–550, https://doi.org/10.1038/nbt.3843 (2017).

Acknowledgements

We thank Songyang Zheng and Fenxia Li for their technical support. This work was partly supported by grants from the National Cancer Institute (RO1 CA098249 and 1R56CA229262-01 to J.H.L.), Department of Defense (W81XWH-16-1-0541 to J.H.L.) and UPMC enterprise (to J.H.L., J.N. and G.M.).

Author Contributions

Y.P.Y. performed most of the experimental analysis. P.L. and G.T. performed the statistical analysis. J.N., R.L.H., R.B., G.M., Q.C., J.Z., D.M., A.P. and M.N. provided materials and expertise. J.H.L. wrote the manuscript and oversaw the project.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-38550-6.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

www.oncotarget.com Oncotarget, 2019, Vol. 10, (No. 36), pp: 3352-3360

Research Paper

Detection of fusion transcripts in the serum samples of patients with hepatocellular carcinoma

Yan-Ping Yu[1], Allan Tsung[1,2], Silvia Liu[1], Michael Nalesnick[1], David Geller[1], George Michalopoulos[1] and Jian-Hua Luo[1]

[1]Department of Pathology and Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA

[2]Current address: Department of Surgery, Ohio State University School of Medicine, Columbus, Ohio 43210, USA

Correspondence to: *Jian-Hua Luo,* ***email:*** *luoj@upmc.edu*

Keywords: *serum detection; fusion transcript; cell-free RNA; HCC*

Received: *February 25, 2019* ***Accepted:*** *April 04, 2019* ***Published:*** *May 21, 2019*

ABSTRACT

Hepatocellular carcinoma is one of the most lethal cancers in the United States. Early detection of the disease is crucial for reducing the mortality of this malignancy. Recently, we identified a panel of fusion genes present in several types of human cancers, including hepatocellular carcinoma. Among 8 fusion genes, MAN2A1-FER, TRMT11-GRIK2 and CCNH-C5orf30 appear most frequently in hepatocellular carcinoma samples. In this study, we showed that the fusion transcripts of MAN2A1-FER, CCNH-C5orf30 and SLC45A2-AMACR were detected in the serum samples of liver cancer patients as circulating cell-free RNA. The distributions of these gene fusion RNA fragments largely matched those of the primary HCC samples. In contrast, the sera of all healthy individuals free of human malignancies were shown to be negative for these fusion genes. These results suggest that gene fusion RNA is frequently shed from liver cancer cells. The detection of serum cell-free fusion transcripts may provide a new approach to aid in the diagnosis, follow-up or therapy of liver cancers.

INTRODUCTION

Human cancer is one of the most frequent causes of death in the United States. In 2018, the mortality rate of cancer reached 606,880 in the US [1], making cancer the second most lethal cause of death after cardiovascular diseases [2]. Hepatocellular carcinoma (HCC) is one of the most lethal malignancies, accounting for more than 31,000 deaths in the US alone [1]. The five-year survival rate for HCC is approximately 18%. Only pancreatic adenocarcinoma and glioblastoma multiforme have lower survival rates [3]. The development of early detection methods and effective treatment for HCC is urgently needed to reduce the mortality of this disease.

Treating liver cancer in the early clinical stages offers a significant advantage for therapeutic options and a better prognosis [4]. Currently, surgical resection, ablation and liver transplant are the most effective approaches to treating early-stage HCC [5, 6]. HCC patients treated with these approaches typically survive long-term and can even be considered cured of the disease. However, patients with late clinical stage HCC without similar options of surgical intervention usually survive less than a year. Thus, early detection of HCC is crucial for reducing the mortality of liver cancer. Recently, we identified a panel of 8 fusion genes in human cancers [7–9]. Some of these fusion genes were shown to be present in a large proportion of HCC cancer samples [7, 10]. The mechanisms underlying these gene fusions are chromosomal translocation and rearrangement [8–10]. The presence of these fusion transcripts in liver cancer samples indicates that translocation and chromosomal rearrangement are common in liver cancer cells. To investigate the utility of these fusion transcripts in detecting liver cancer, we performed TaqMan qRT-PCR on the RNA extracted from cell-free serum. The results suggest that many of these fusion genes are detectable as cell-free circulating RNA.

RESULTS

One of the hallmarks of genomes of human cancer is chromosomal rearrangement and translocation [11, 12]. Previously, we identified a panel of fusion genes in prostate cancer samples from patients who experienced poor clinical outcomes. Subsequent analyses showed that many of these fusion genes are present in a variety of human cancers, including liver cancer [7, 10]. To investigate whether these fusion transcripts are detectable in the sera of HCC patients, we analyzed the presence or absence of 8 fusion genes in 118 serum samples from HCC patients and individuals free of malignant tumors. As shown in Table 1, all serum samples from individuals free of malignancies were negative (0/14) for the fusion transcripts of all the fusion genes. In contrast, 83.7% (87/104) of the serum samples from HCC patients were positive for at least one fusion transcript. Interestingly, all serum samples obtained from HCC patients with non-alcoholic steatohepatitis etiology (n=20) were positive for at least one fusion gene (p=0.019): 100% (20/20) versus 77.6% (59/76). The fusion gene is also more likely present in the serum samples from HCC patients with steatohepatitis background (p=0.02): 100% (20/20) versus 78.5% (62/79). Interestingly, multiple fusion transcripts (at least 2) detected in the serum are associated with moderate differentiation of HCC (42.4% [14/33] versus 9.5% [2/21] for all other category, p=0.014).

Upon analyzing individual fusion transcripts, we determined that MAN2A1-FER was frequently detected in the sera of HCC patients, reaching 78.8% (82/104, Table 2). SLC45A2-AMAMCR occurred at a frequency of 31.7% (33/104), while CCNH-C5orf30 occurred at a frequency of 10.6% (11/104). All HCC patients with NASH were positive for MAN2A1-FER in their serum samples (20/20), indicating a strong association between MAN2A1-FER and the etiology of NASH (p=0.005).

To determine whether the HCC samples from the same patients were positive for these fusion genes, six liver cancer samples from these HCC patients were analyzed. As shown in Table 3 and Figure 1, MAN2A1-FER and CCNH-C5orf30 were positive in all six HCC samples,

Figure 1: Detection of MAN2A1-FER, CCNH-C5orf30 and SLC45A2-AMACR in HCC samples and the corresponding serum samples. Six cases of HCC and matched serum samples were analyzed for the presence of transcripts of MAN2A1-FER, CCNH-C5orf30 and SLC45A2-AMACR using TaqMan qRT-PCR. The results of β-actin were used as normalization controls. Assays were performed twice independently. Sanger sequencing was performed on 20% of all positive samples.

Table 1: Fusion transcripts detection in the sera of HCC patients

Clinical characteristics	Fusion gene positive	Fusion gene negative
HCC patients		
Age:		
40s	5	0
50s	22	2
60s	33	11
70s	20	1
80s	4	3
Etiology:		
HCV	44	8
HBV	6	2
Ethanol	34	9
NASH	20	0
Other	4	4
Background liver:		
Cirrhosis/fibrosis	81	17
Steatosis	12	4
Steatohepatitis	20	0
Recurrent status:		
Recurrent	20	2
Non-recurrent	50	11
Response to therapy:		
Responsive	31	8
Progressive	42	6
Pathology grade:		
Poorly differentiated	5	0
Moderately differentiated	28	5
Well differentiated	18	6
Death	40	7
Alive	42	9
Healthy individuals	0	14

while 4 of 6 HCC samples were positive for SLC45A2-AMACR. All positive serum samples corresponded to the matched positive HCC samples, indicating that the source of the seral fusion transcripts was the liver cancer. Two serum samples were negative for MAN2A1-FER, while the matched HCC samples were positive for the fusion gene. Similarly, two HCC samples positive for SLC45A2-AMACR had matched serum samples that were negative for the same fusion gene. Even though TRMT11-GRIK2 and CCNH-C5orf30 were present in all six HCC samples, the transcripts of these fusion genes were undetectable in the sera of the same patients. Half of the HCC samples were also positive for LRRC59-FLJ60017, but the fusion transcript was not detected in the matched serum samples. Based on these results, MAN2A1-FER appears to be the most sensitive marker for serum detection of HCC (4/6 or 67%). SLC45A2-AMACR ranks second (2/4 or 50%), while TRMT11-GRIK2 and CCNH-C5orf30 are the most insensitive markers (0/6 or 0%). These results suggest that these fusion transcripts have different levels

Table 2: Frequency of individual fusion transcript detected in the sera of HCC patients

Clinical features	*MAN2A1-FER*	*TRMT11-GRIK2*	*MTOR-TP53BP1*	*CCNH-C5orf30*	*KDM4B-AC011523.2*	*SLC45A2-AMACR*	*TMEM135-CCDC67*	*LRRC59-FLJ60017*
All HCC patients	83/104	0/104	0/104	11/104	0/104	33/104	0/104	0/104
Ages:								
>80s	11/15	0/15	0/15	0/15	0/15	6/15	0/15	0/15
70s	21/22	0/22	0/22	2/22	0/22	7/22	0/22	0/22
60s	36/48	0/48	0/48	8/48	0/48	11/48	0/48	0/48
50s	12/14	0/14	0/14	1/14	0/14	5/14	0/14	0/14
40s	2/2	0/2	0/2	0/2	0/2	2/2	0/2	0/2
Etiology:								
HCV	42/54	0/54	0/54	8/34	0/54	19/54	0/54	0/54
HBV	7/8	0/8	0/8	1/8	0/8	1/8	0/8	0/8
Ethanol	31/43	0/43	0/43	2/43	0/43	11/43	0/43	0/43
NASH	21/21	0/21	0/21	4/21	0/21	7/21	0/21	0/21
Other	2/5	0/5	0/5	0/5	0/5	0/5	0/5	0/5
Background liver:								
Cirrhosis/fibrosis	81/101	0/101	0/101	12/101	0/101	28/101	0/101	0/101
Steatosis	11/16	0/16	0/16	0/16	0/16	7/16	0/16	0/16
Steatohepatitis	20/21	0/21	0/21	3/21	0/21	7/21	0/21	0/21
Recurrent status:								
Recurrent	20/22	0/22	0/22	2/22	0/22	5/22	0/22	0/22
Non-recurrent	50/63	0/63	0/63	10/63	0/63	20/63	0/63	0/63
Response to therapy:								
Responsive	29/38	0/38	0/38	5/38	0/38	12/38	0/38	0/38
Progressive	42/49	0/49	0/49	7/49	0/49	16/49	0/49	0/49
Pathology differentiation grade:								
Poor	4/5	0/5	0/5	0/5	0/5	3/5	0/5	0/5
Moderate	26/34	0/34	0/34	5/34	0/34	5/34	0/34	0/34
Well	18/24	0/24	0/24	2/24	0/24	3/24	0/24	0/24
Death	39/48	0/48	0/48	5/48	0/48	17/48	0/48	0/48
Alive	41/51	0/51	0/51	5/51	0/51	15/51	0/51	0/51
Healthy person	0/14	0/14	0/14	0/14	0/14	0/14	0/14	0/14

of detectability in the blood, probably due to differences in the sensitivity of these RNA sequences to circulating RNAses.

DISCUSSION

Cancer-specific gene fusions are the result of chromosomal rearrangement and translocation [8]. Many gene fusion events are not specific to one type of cancer. Indeed, most of the fusion genes we discovered in prostate cancer were later detected in a variety of human malignancies, including HCC [7, 10]. MAN2A1-FER and SLC45A2-AMACR belong to a class of fusion genes with gain of function. The MAN2A1-FER gene fusion generates a new chimera protein in which the C-terminal glycoside hydrolase of mannosidase alpha, class 2A,

Table 3: Fusion gene detection in serum versus matched HCC tissue

Case No.	*MAN2A1-FER*	*TRMT11-GRIK2*	*MTOR-TP53BP1*	*CCNH-C5orf30*	*KDM4B-AC011523.2*	*SLC45A2-AMACR*	*TMEM135-CCDC67*	*LRRC59-FLJ60017*
2274								
HCC tissue	+	+	-	+	-	+	-	-
Serum	+	-	-	-	-	+	-	-
2298								
HCC tissue	+	+	-	+	-	+	-	-
Serum	+	-	-	-	-	+	-	-
2209								
HCC tissue	+	+	-	+	-	-	-	+
Serum	-	-	-	-	-	-	-	-
2128								
HCC tissue	+	+	+	+	-	+	-	+
Serum	-	-	-	-	-	-	-	-
2218								
HCC tissue	+	+	-	+	-	-	-	-
Serum	+	-	-	-	-	-	-	-
2172								
HCC tissue	+	+	-	+	+	+	-	+
Serum	+	-	-	-	-	-	-	-

member 1 (MAN2A1) is replaced with an intact tyrosine kinase domain from FER [9, 10]. The new chimera protein exhibits an almost 4-fold increase in tyrosine kinase activity compared with that of the native FER, and the chimera protein is translocated to the Golgi apparatus [10]. The resulting chimera protein transforms immortalized cells into cancer through the ectopic phosphorylation of growth factor receptor [10]. MAN2A1-FER was detected in 15% of human HCC samples, and this fusion protein was shown to be a driver of mouse liver cancer [10]. SLC45A2-AMACR was also detected in a lung cancer cell line [13] and urothelial carcinoma [14], in addition to prostate cancer [9, 15]. To our knowledge, this is the first report showing that SLC45A2-AMACR is present in HCC and the corresponding serum samples. AMACR is a racemase responsible for branch fatty acid metabolism, while SLC45A2 is a solute transporter. The fusion generates a chimera protein such that 5 transmembrane domains of SLC45A2 are removed from its C-terminus and replaced with an intact racemase domain from AMACR. The overexpression of AMACR was shown to be associated with the aggressive behavior of multiple human cancers [16–25]. In vitro, AMACR was shown to increase cell growth and proliferation [26].

CCNH-C5orf30 and TRMT11-GRIK2 belong to a class of fusion genes with loss of function. CCNH-C5orf30 was detected in 37% of HCC samples [10], while TRMT11-GRIK2 was detected in 12.9% of HCC samples [7]. The gene fusion of CCNH-C5orf30 produces a truncation of the H5′ and HC domains of cyclin H (CCNH), which is an important cell cycle regulator for the progression to mitosis [27, 28], and an independent C5orf30 protein. The truncated CCNH in the fusion gene is defective in binding with cdk7 [29] and may be defective in its transcriptional activity and promotion of the cell cycle. A more dramatic loss of function is identified in the TRMT11-GRIK2 gene fusion; the TRMT11-GRIK2 gene fusion eliminates the open-reading frame of GRIK2, which is a potential tumor suppressor [30, 31], and produces a large truncation of TRMT11, which is a tRNA methyltransferase [32, 33]. Thus, the fusion event is equivalent to the structural deletion of both TRMT11 and GRIK2 genes. The deletion of TRMT11 reduces the stability of tRNA and may therefore adversely impact the protein translation of cancer cells, while the loss of GRIK2 may promote the growth of cancer cells. All these gene fusion events may play important roles in the development of human liver cancer.

The abnormal chromosomal recombination that generates these fusion genes is cancer-specific and is absent in normal tissues [8]. The frequent presence of these fusion transcripts in the serum samples of HCC

patients suggests that these RNAs are derived from liver cancer cells. Three lines of evidence support the hypothesis that the fusion RNA fragments detected in the serum are shed from liver cancer cells. First, the pattern of fusion transcript distribution in the serum completely overlaps that of the corresponding liver cancer samples, i.e., there is no fusion transcript that is positive in the serum but negative in the matched liver cancer sample. Second, all serum samples from healthy individuals are negative for these fusion transcripts. Normal individuals do not produce these fusion genes. Third, the normalized quantities of the detected fusion transcripts in the serum are generally 4- to 16-fold lower than those detected in the corresponding HCC samples, suggesting the shedding of RNAs from a fraction of cancer cells.

Early detection of HCC is probably the most effective way to reduce the mortality of the disease, due to the availability of many effective surgical treatments. Unfortunately, at the time of diagnosis, the HCCs of many patients are at the advanced stages, eliminating many options for curing the disease [4]. Currently, the primary means of diagnosing HCC relies on radiology imaging of liver cancers. The screening of HCC for patients with cirrhosis and chronic liver diseases involves biannual ultrasonography [4]. This screening method may be combined with a seral test for α-fetoprotein or other imaging analyses, such as MRI, CT and contrast-enhanced ultrasound, if a suspicious nodule appears. The presence of fusion transcripts from HCC cancer cells in the serum may represent a new approach for detecting liver cancer. Since the test is minimally invasive, it can be employed regularly in conjunction with ultrasound screening for patients with chronic liver disease or cirrhosis. When a suspicious nodule is detected, the presence of a fusion transcript may help to confirm the diagnosis. All 14 healthy individuals negative for the fusion transcript had no known liver disease and were cancer-free. The high frequency of fusion genes in liver cancer implies that gene fusion in the cancer genome is an early event for liver cancer development. It is of interest to investigate whether these fusion genes are also present in some of the HCC precursor lesions such as NASH. So far, the fusion transcripts appear cancer-specific. The serum detection of these fusion genes may provide a sensitive follow-up test for patients undergoing surgical resection or liver transplant to monitor the recurrence of liver cancer. A new therapeutic approach was recently developed to target the chromosomal breakpoints of these fusion genes using the CRISPR-cas9 gene editing system [34]. This approach led to the partial remission of xenografted human liver cancers when the animals were treated with reagents targeting the breakpoint of the MAN2A1-FER fusion gene [34]. As a result, the detection of these fusion genes in HCC may have significant therapeutic implications. The utilization of cell-free fusion RNA in the circulation as tumor markers may provide an important means for early detection, follow-up and therapeutic guidance for the management of HCC patients.

MATERIALS AND METHODS

Tissue samples

The 124 tissue specimens and serum samples used in this study consisted of 6 hepatocellular carcinoma samples from HCC patients, 104 serum samples from HCC patients and 14 serum samples from non-cancerous patients. These samples were obtained from the University of Pittsburgh Tissue Bank in compliance with institutional regulatory guidelines. The informed consent exemptions and protocol were approved by the Institution Review Board of the University of Pittsburgh. All serum samples and hepatocellular carcinoma samples were fresh-frozen and stored at -80°C. Some cases have multiple etiologies, pathological features, and backgrounds. They were classified multiple times. When statistical analyses were performed, however, the duplication was excluded from the analyses. HCC cases that do not contain the specific clinical information were also excluded from the analyses.

RNA extraction, cDNA synthesis and detection of fusion genes

The procedures for RNA extraction, cDNA synthesis and the detection of fusion genes were described previously [35–49]. Briefly, total RNA was extracted using Trizol to lyse the cells in the cancer tissues (Invitrogen, CA, USA). First strand cDNA was synthesized using ~2 μg of RNA from each sample, random hexamers and Superscript II™ (Invitrogen, Inc, CA, USA) at 42°C for 2 hours. One microliter of each cDNA sample was used for the TaqMan PCR reactions with 50 heat cycles, as follows: 94°C for 30 seconds, 61°C for 30 seconds, and 72°C for 30 seconds. The following primers and probes were used: MAN2A1-FER (AGCGCAGTTTGGGATACAGCA/ CTTTAATGTGCCCTTATATACTTCACC; TaqMan probe, 5'/56-FAM/TCAGAAAC A/ZEN/GCCTATGAGG GAAATT/3IABkFQ/3'), SLC45A2-AMACR (TTGAT GTCTGCTCCCATCAGG/CAGCTGGAGTTTCTCCAT GAC; TaqMan probe, 5'-/56-FAM/AAGAGGGCA/ZEN/ TGGTAGTGGAGGC/3IABkFQ/-3'), CCNH-C5orf30 (AAAGTTATTTATCAGAGAGTCTGATGCTG/CTGTT CTACTCCAGGTATTTTCATTATATC; TaqMan probe, 5'-/56-FAM/ACAGGCAAG/ZEN/TTCTGTTCTCTTTC AGCA/3IABkFQ/-3'), mTOR-TP53BP1 (TGATAGA CCAGTCCCGGGATG / CCACTGACATTCCCAGA ACAAG; TaqMan probe, 5'-/56-FAM/ TGTCAGCCT/ ZEN/GTCAGAATCCAAGTCAAG/3IABkFQ/-3'), TRM T11-GRIK2 (GCGCTGTCGTGTACCCTTAAC / GAAT GCAAGTTCCTCAGCTCC; TaqMan probe, 5'-/56-FAM/ CGGAACTCC/ZEN/AGATGCTCCTGCG/3IABkFQ/

-3'), LRRC59-FLJ60017 (GTGACTGCTTGGATG AGAAGC / CCCTCCTCTGGTTTGTTGTTG; TaqMan probe, 5'-/56-FAM/CAGTGTGCA/ZEN/AACAAGGT GACTGGAAG/3IABkFQ/-3'), TMEM135-CCDC67 (CAGCTGTCATGGAAGTTCAGAC / CCTCATTCT TTCCTGCTCAGAG; TaqMan probe, 5'-/56-FAM/ AGTTCCTTT/ZEN/TAAGACTCACCAAGGGCAA/3IA BkFQ/-3'), KDM4- AC011523.2 (AGACC ACCTTCGCCTGGCAC / TCTCTCTCAGATCCAG GCTTG; TaqMan probe, 5'-/56-FAM/ACAGCATCA/ZEN/ ACTACCTGCACTTTGGG/3IABkFQ/-3'), and β-actin (ACCCCACTTCTCTCTAAGGAG / GCAATGCTATC ACCTCCCCTG; TaqMan probe, 5'-/56-FAM/CCA GTCCTC/ZEN/TCCCAAGTCCACAC/3IABkFQ/-3'). The PCR reactions were performed in a thermocycler (Eppendorf Realplex™ thermocycler). A negative control and synthetic positive control were included in each batch of reactions. The PCR products were gel-purified and Sanger-sequenced for 20% of the positive samples.

Abbreviations

AC011523.2, Homo sapiens chromosome 19 clone LLNLF-214C7; **AMACR,** Alpha-methylacyl-CoA racemase; **C5orf30,** Chromosome 5 open-reading frame 30; **CCNH,** Cyclin H; **FER,** Fez related tyrosine kinase; **FLJ60017,** Homo sapiens cDNA FLJ60017; **GRIK2,** Glutamate ionotropic receptor kainate type subunit 2; **HCC,** Hepatocellular carcinoma; **KDM4B,** Lysine demethylase 4B; **LRRC59,** Leucine rich repeat containing 59; **MAN2A1,** Mannosidase class 2 member A1; **mTOR,** Mechanistic target of rapamycin kinase; **qRT-PCR,** quantitative reversed transcription polymerase chain reaction; **SLC45A2,** Solute carrier family 45 member 2; **TP53BP1,** Tumor protein p53 binding protein 1; **TRMT11,** tRNA methyltransferase 11 homolog

Author contributions

YY designed and supervised the experiment. SL performed the statistical analyses. AT, DG and MN provided the samples. MG provided experimental and clinical advices. JHL conceived the idea and oversaw the project.

ACKNOWLEDGMENTS AND FUNDING

We thank Songyang Zheng for the technical support. This work wa s partly supported by grants from the National Cancer Institute (RO1 CA098249 and 1R56CA229262-01 to JHL) and the Department of Defense (W81XWH-16-1-0541 to JHL).

Ethics statement

All the patients provided written informed consent and patient anonymity has been preserved. Investigation was conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the institutional review board of University of Pittsburgh.

CONFLICTS OF INTEREST

All authors declare no conflicts of interest related to the study.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019; 69:7–34. https://doi.org/10.3322/caac.21551. [PubMed]
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018; 68:7–30. https://doi.org/10.3322/caac.21442. [PubMed]
3. Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M, Gores G. Hepatocellular carcinoma. Nat Rev Dis Primers. 2016; 2:16018. https://doi.org/10.1038/nrdp.2016.18. [PubMed]
4. Plentz RR, Malek NP. Early Detection of Hepatocellular Carcinoma: How to Screen and Follow up Patients with Liver Cirrhosis According to the GERMAN S3 Guideline? Diagnostics (Basel). 2015; 5:497–503. https://doi.org/10.3390/diagnostics5040497. [PubMed]
5. She WH, Chan AC, Cheung TT, Lo CM, Chok KS. Survival outcomes of liver transplantation for hepatocellular carcinoma in patients with normal, high and very high preoperative alpha-fetoprotein levels. World J Hepatol. 2018; 10:308–18. https://doi.org/10.4254/wjh.v10.i2.308. [PubMed]
6. Bozorgzadeh A, Orloff M, Abt P, Tsoulfas G, Younan D, Kashyap R, Jain A, Mantry P, Maliakkal B, Khorana A, Schwartz S. Survival outcomes in liver transplantation for hepatocellular carcinoma, comparing impact of hepatitis C versus other etiology of cirrhosis. Liver Transpl. 2007; 13:807–13. https://doi.org/10.1002/lt.21054. [PubMed]
7. Yu YP, Liu P, Nelson J, Hamilton RL, Bhargava R, Michalopoulos G, Chen Q, Zhang J, Ma D, Pennathur A, Luketich J, Nalesnik M, Tseng G, Luo JH. Identification of recurrent fusion genes across multiple cancer types. Sci Rep. 2019; 9:1074. https://doi.org/10.1038/s41598-019-38550-6. [PubMed]
8. Luo JH, Liu S, Zuo ZH, Chen R, Tseng GC, Yu YP. Discovery and classification of fusion transcripts in prostate cancer and normal prostate tissue. Am J Pathol. 2015; 185:1834–45. https://doi.org/10.1016/j.ajpath.2015.03.008. [PubMed]

9. Yu YP, Ding Y, Chen Z, Liu S, Michalopoulos A, Chen R, Gulzar ZG, Yang B, Cieply KM, Luvison A, Ren BG, Brooks JD, Jarrard D, et al. Novel fusion transcripts associate with progressive prostate cancer. Am J Pathol. 2014; 184:2840–49. https://doi.org/10.1016/j.ajpath.2014.06.025. [PubMed]
10. Chen ZH, Yu YP, Tao J, Liu S, Tseng G, Nalesnik M, Hamilton R, Bhargava R, Nelson JB, Pennathur A, Monga SP, Luketich JD, Michalopoulos GK, Luo JH. MAN2A1-FER Fusion Gene Is Expressed by Human Liver and Other Tumor Types and Has Oncogenic Activity in Mice. Gastroenterology. 2017; 153:1120–1132.e15. https://doi.org/10.1053/j.gastro.2016.12.036. [PubMed]
11. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144:646–74. https://doi.org/10.1016/j.cell.2011.02.013. [PubMed]
12. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000; 100:57–70. https://doi.org/10.1016/S0092-8674(00)81683-9. [PubMed]
13. Klijn C, Durinck S, Stawiski EW, Haverty PM, Jiang Z, Liu H, Degenhardt J, Mayba O, Gnad F, Liu J, Pau G, Reeder J, Cao Y, et al. A comprehensive transcriptional portrait of human cancer cell lines. Nat Biotechnol. 2015; 33:306–12. https://doi.org/10.1038/nbt.3080. [PubMed]
14. Yoshihara K, Wang Q, Torres-Garcia W, Zheng S, Vegesna R, Kim H, Verhaak RG. The landscape and therapeutic relevance of cancer-associated transcript fusions. Oncogene. 2015; 34:4845–54. https://doi.org/10.1038/onc.2014.406. [PubMed]
15. Yang J, Chen Y, Lu J, Wang X, Wang L, Liang J, Sun ZS. Identification and characterization of novel fusion genes in prostate cancer by targeted RNA capture and next-generation sequencing. Acta Biochim Biophys Sin (Shanghai). 2018; 50:1166–72. https://doi.org/10.1093/abbs/gmy112. [PubMed]
16. Li W, Cagle PT, Botero RC, Liang JJ, Zhang Z, Tan D. Significance of overexpression of alpha methylacyl-coenzyme A racemase in hepatocellular carcinoma. J Exp Clin Cancer Res. 2008; 27:2. https://doi.org/10.1186/1756-9966-27-2. [PubMed]
17. Huang W, Zhao J, Li L, Huang Y, Yang X, Wang J, Zhang T. a-Methylacyl coenzyme A racemase is highly expressed in the intestinal-type adenocarcinoma and high-grade dysplasia lesions of the stomach. Histol Histopathol. 2008; 23:1315–20. https://doi.org/10.14670/HH-23.1315. [PubMed]
18. Shilo K, Dracheva T, Mani H, Fukuoka J, Sesterhenn IA, Chu WS, Shih JH, Jen J, Travis WD, Franks TJ. Alpha-methylacyl CoA racemase in pulmonary adenocarcinoma, squamous cell carcinoma, and neuroendocrine tumors: expression and survival analysis. Arch Pathol Lab Med. 2007; 131:1555–60. [PubMed]
19. Lee WA. Alpha-methylacyl-CoA-racemase expression in adenocarcinoma, dysplasia and non-neoplastic epithelium of the stomach. Oncology. 2006; 71:246–50. https://doi.org/10.1159/000106428. [PubMed]
20. Zha S, Ferdinandusse S, Denis S, Wanders RJ, Ewing CM, Luo J, De Marzo AM, Isaacs WB. Alpha-methylacyl-CoA racemase as an androgen-independent growth modifier in prostate cancer. Cancer Res. 2003; 63:7365–76. [PubMed]
21. Zhou M, Chinnaiyan AM, Kleer CG, Lucas PC, Rubin MA. Alpha-Methylacyl-CoA racemase: a novel tumor marker over-expressed in several human cancers and their precursor lesions. Am J Surg Pathol. 2002; 26:926–31. https://doi.org/10.1097/00000478-200207000-00012. [PubMed]
22. Rubin MA, Zhou M, Dhanasekaran SM, Varambally S, Barrette TR, Sanda MG, Pienta KJ, Ghosh D, Chinnaiyan AM. alpha-Methylacyl coenzyme A racemase as a tissue biomarker for prostate cancer. JAMA. 2002; 287:1662–70. https://doi.org/10.1001/jama.287.13.1662. [PubMed]
23. Luo J, Zha S, Gage WR, Dunn TA, Hicks JL, Bennett CJ, Ewing CM, Platz EA, Ferdinandusse S, Wanders RJ, Trent JM, Isaacs WB, De Marzo AM. Alpha-methylacyl-CoA racemase: a new molecular marker for prostate cancer. Cancer Res. 2002; 62:2220–26. [PubMed]
24. Kuefer R, Varambally S, Zhou M, Lucas PC, Loeffler M, Wolter H, Mattfeldt T, Hautmann RE, Gschwend JE, Barrette TR, Dunn RL, Chinnaiyan AM, Rubin MA. alpha-Methylacyl-CoA racemase: expression levels of this novel cancer biomarker depend on tumor differentiation. Am J Pathol. 2002; 161:841–48. https://doi.org/10.1016/S0002-9440(10)64244-7. [PubMed]
25. Kastelein F, Biermann K, Steyerberg EW, Verheij J, Kalisvaart M, Looijenga LH, Stoop HA, Walter L, Kuipers EJ, Spaander MC, Bruno MJ, and ProBar study group. Value of α-methylacyl-CoA racemase immunochemistry for predicting neoplastic progression in Barrett's oesophagus. Histopathology. 2013; 63:630–39. https://doi.org/10.1111/his.12216. [PubMed]
26. Festuccia C, Gravina GL, Mancini A, Muzi P, Cesare ED, Kirk R, Smith M, Hughes S, Gibson R, Lian LY, Ricevuto E, Carnell AJ. Trifluoroibuprofen inhibits α-methylacyl coenzyme A racemase (AMACR/P504S), reduces cancer cell proliferation and inhibits in vivo tumor growth in aggressive prostate cancer models. Anticancer Agents Med Chem. 2014; 14:1031–41. https://doi.org/10.2174/1871520614666140327152607. [PubMed]
27. Mäkelä TP, Parvin JD, Kim J, Huber LJ, Sharp PA, Weinberg RA. A kinase-deficient transcription factor TFIIH is functional in basal and activated transcription. Proc Natl Acad Sci USA. 1995; 92:5174–78. https://doi.org/10.1073/pnas.92.11.5174. [PubMed]
28. Fisher RP, Morgan DO. A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. Cell. 1994; 78:713–24. https://doi.org/10.1016/0092-8674(94)90535-5. [PubMed]
29. Andersen G, Busso D, Poterszman A, Hwang JR, Wurtz JM, Ripp R, Thierry JC, Egly JM, Moras D. The structure of cyclin H: common mode of kinase activation

and specific features. EMBO J. 1997; 16:958–67. https://doi.org/10.1093/emboj/16.5.958. [PubMed]

30. Paschen W, Blackstone CD, Huganir RL, Ross CA. Human GluR6 kainate receptor (GRIK2): molecular cloning, expression, polymorphism, and chromosomal assignment. Genomics. 1994; 20:435–40. https://doi.org/10.1006/geno.1994.1198. [PubMed]

31. Wu CS, Lu YJ, Li HP, Hsueh C, Lu CY, Leu YW, Liu HP, Lin KH, Hui-Ming Huang T, Chang YS. Glutamate receptor, ionotropic, kainate 2 silencing by DNA hypermethylation possesses tumor suppressor function in gastric cancer. Int J Cancer. 2010; 126:2542–52. https://doi.org/10.1002/ijc.24958. [PubMed]

32. Towns WL, Begley TJ. Transfer RNA methytransferases and their corresponding modifications in budding yeast and humans: activities, predications, and potential roles in human health. DNA Cell Biol. 2012; 31:434–54. https://doi.org/10.1089/dna.2011.1437. [PubMed]

33. Bourgeois G, Marcoux J, Saliou JM, Cianférani S, Graille M. Activation mode of the eukaryotic m2G10 tRNA methyltransferase Trm11 by its partner protein Trm112. Nucleic Acids Res. 2017; 45:1971–82. https://doi.org/10.1093/nar/gkw1271. [PubMed]

34. Chen ZH, Yu YP, Zuo ZH, Nelson JB, Michalopoulos GK, Monga S, Liu S, Tseng G, Luo JH. Targeting genomic rearrangements in tumor cells through Cas9-mediated insertion of a suicide gene. Nat Biotechnol. 2017; 35:543–50. https://doi.org/10.1038/nbt.3843. [PubMed]

35. He DM, Ren BG, Liu S, Tan LZ, Cieply K, Tseng G, Yu YP, Luo JH. Oncogenic activity of amplified miniature chromosome maintenance 8 in human malignancies. Oncogene. 2017; 36:3629–39. https://doi.org/10.1038/onc.2017.123. [PubMed]

36. Zuo ZH, Yu YP, Ding Y, Liu S, Martin A, Tseng G, Luo JH. Oncogenic Activity of miRNA 650 in Prostate Cancer Is Mediated by Suppression of CSR1 Expression. Am J Pathol. 2015; 185:1991–9. https://doi.org/10.1016/j.ajpath.2015.03.015. [PubMed]

37. Yu YP, Michalopoulos A, Ding Y, Tseng G, Luo JH. High fidelity copy number analysis of formalin-fixed and paraffin-embedded tissues using Affymetrix Cytoscan HD chip. PLoS One. 2014; 9:e92820. https://doi.org/10.1371/journal.pone.0092820. [PubMed]

38. Nalesnik MA, Tseng G, Ding Y, Xiang GS, Zheng ZL, Yu Y, Marsh JW, Michalopoulos GK, Luo JH. Gene deletions and amplifications in human hepatocellular carcinomas: correlation with hepatocyte growth regulation. Am J Pathol. 2012; 180:1495–508. https://doi.org/10.1016/j.ajpath.2011.12.021. [PubMed]

39. Han YC, Yu YP, Nelson J, Wu C, Wang H, Michalopoulos GK, Luo JH. Interaction of integrin-linked kinase and miniature chromosome maintenance 7-mediating integrin {alpha}7 induced cell growth suppression. Cancer Res. 2010; 70:4375–84. https://doi.org/10.1158/0008-5472.CAN-09-4403. [PubMed]

40. Yu YP, Yu G, Tseng G, Cieply K, Nelson J, Defrances M, Zarnegar R, Michalopoulos G, Luo JH. Glutathione peroxidase 3, deleted or methylated in prostate cancer, suppresses prostate cancer growth and metastasis. Cancer Res. 2007; 67:8043–50. https://doi.org/10.1158/0008-5472.CAN-07-0648. [PubMed]

41. Ren B, Yu YP, Tseng GC, Wu C, Chen K, Rao UN, Nelson J, Michalopoulos GK, Luo JH. Analysis of integrin alpha7 mutations in prostate cancer, liver cancer, glioblastoma multiforme, and leiomyosarcoma. J Natl Cancer Inst. 2007; 99:868–80. https://doi.org/10.1093/jnci/djk199. [PubMed]

42. Yu G, Tseng GC, Yu YP, Gavel T, Nelson J, Wells A, Michalopoulos G, Kokkinakis D, Luo JH. CSR1 suppresses tumor growth and metastasis of prostate cancer. Am J Pathol. 2006; 168:597–607. https://doi.org/10.2353/ajpath.2006.050620. [PubMed]

43. Luo JH, Ren B, Keryanov S, Tseng GC, Rao UN, Monga SP, Strom S, Demetris AJ, Nalesnik M, Yu YP, Ranganathan S, Michalopoulos GK. Transcriptomic and genomic analysis of human hepatocellular carcinomas and hepatoblastomas. Hepatology. 2006; 44:1012–24. https://doi.org/10.1002/hep.21328. [PubMed]

44. Yu YP, Landsittel D, Jing L, Nelson J, Ren B, Liu L, McDonald C, Thomas R, Dhir R, Finkelstein S, Michalopoulos G, Becich M, Luo JH. Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. J Clin Oncol. 2004; 22:2790–99. https://doi.org/10.1200/JCO.2004.05.158. [PubMed]

45. Ren B, Yu YP, Jing L, Liu L, Michalopoulos GK, Luo JH, Rao UN. Gene expression analysis of human soft tissue leiomyosarcomas. Hum Pathol. 2003; 34:549–58. https://doi.org/10.1016/S0046-8177(03)00014-5. [PubMed]

46. Luo JH, Yu YP, Cieply K, Lin F, Deflavia P, Dhir R, Finkelstein S, Michalopoulos G, Becich M. Gene expression analysis of prostate cancers. Mol Carcinog. 2002; 33:25–35. https://doi.org/10.1002/mc.10018. [PubMed]

47. Yu YP, Lin F, Dhir R, Krill D, Becich MJ, Luo JH. Linear amplification of gene-specific cDNA ends to isolate full-length of a cDNA. Anal Biochem. 2001; 292:297–301. https://doi.org/10.1006/abio.2001.5066. [PubMed]

48. Yu YP, Lin F, Bisceglia M, Krill D, Dhir R, Becich M, Luo JH. Identification of a novel gene with increasing rate of suppression in high grade prostate cancers. Am J Pathol. 2001; 158:19–24. https://doi.org/10.1016/S0002-9440(10)63939-9. [PubMed]

49. Lin F, Yu YP, Woods J, Cieply K, Gooding B, Finkelstein P, Dhir R, Krill D, Becich MJ, Michalopoulos G, Finkelstein S, Luo JH. Myopodin, a synaptopodin homologue, is frequently deleted in invasive prostate cancers. Am J Pathol. 2001; 159:1603–12. https://doi.org/10.1016/S0002-9440(10)63006-4. [PubMed]