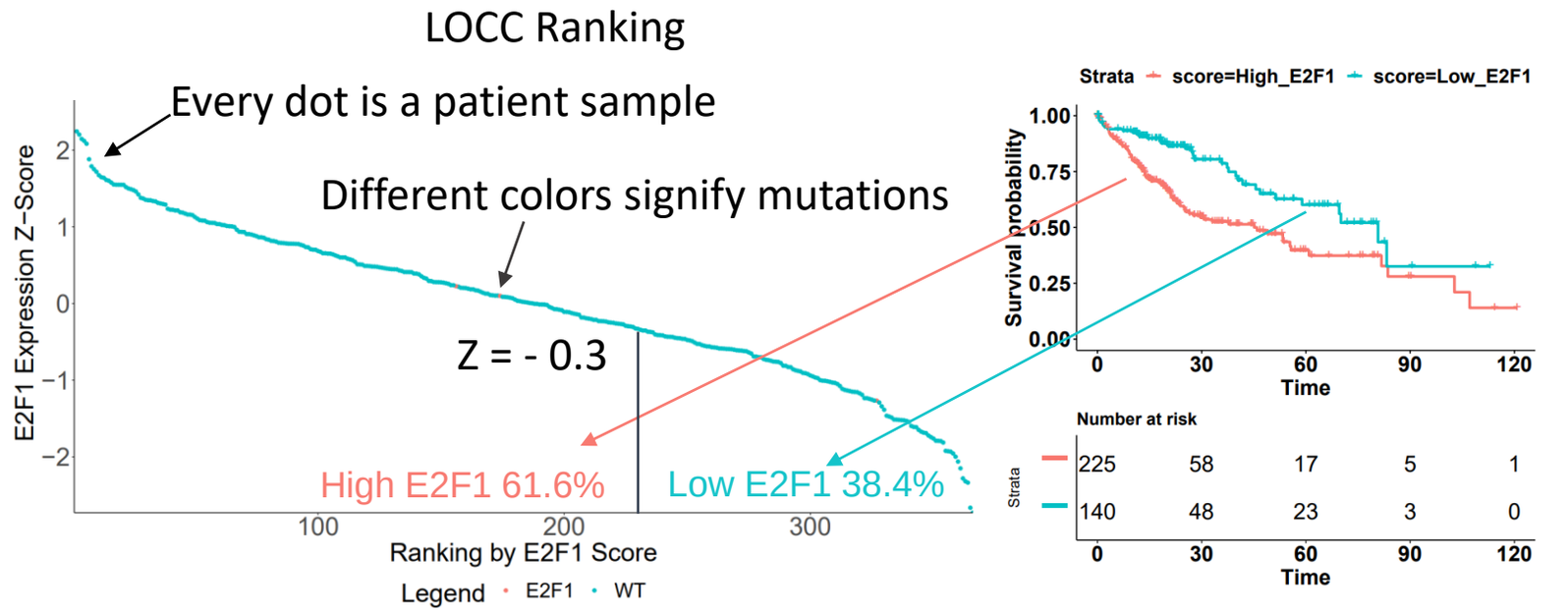
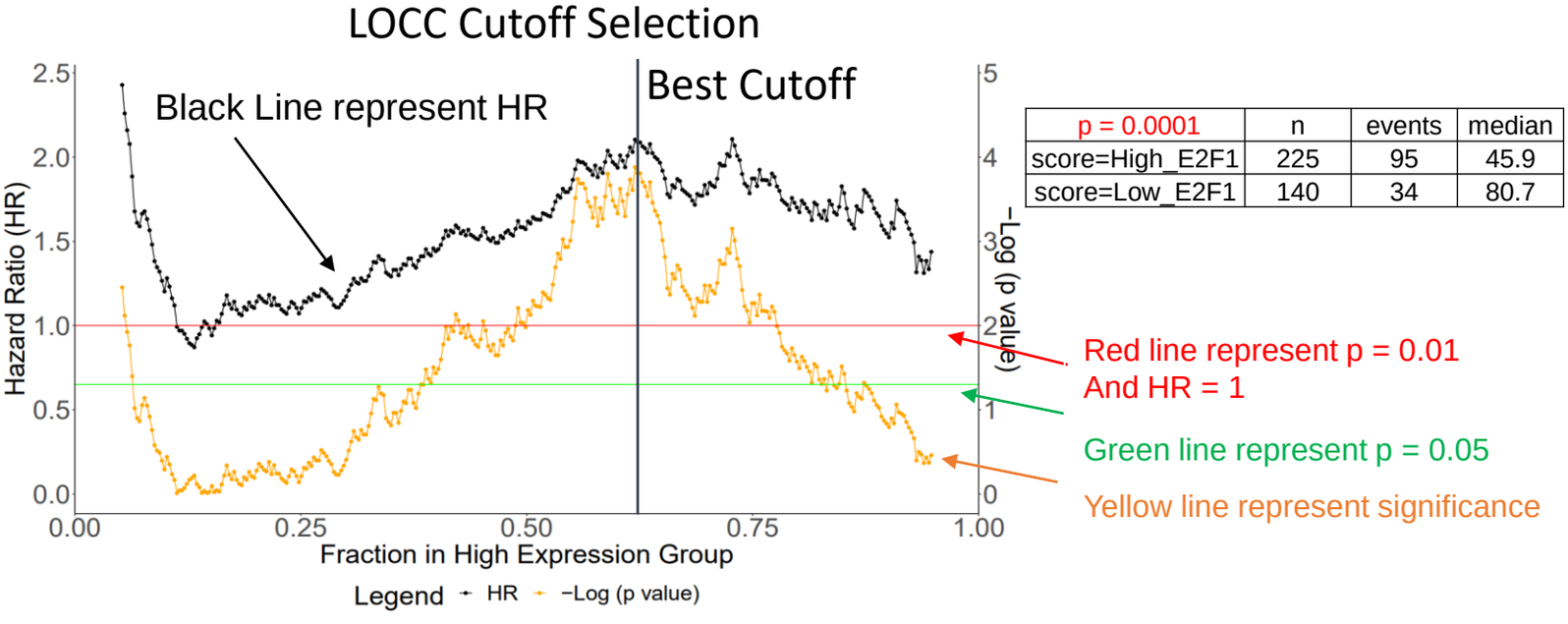




p = 0.0001	n	events	median
score=High_E2F1	225	95	45.9
score=Low_E2F1	140	34	80.7

Red line represent p = 0.01
And HR = 1

Green line represent p = 0.05

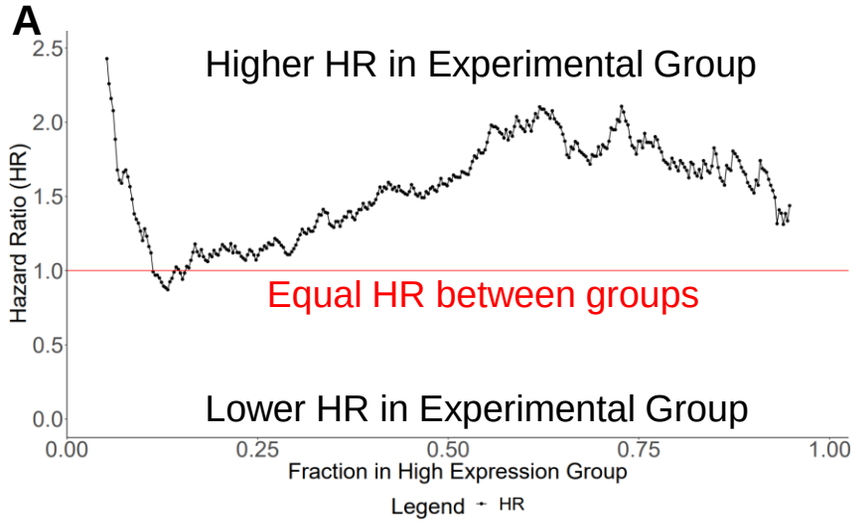
Yellow line represent significance

All tumor samples are ranked by E2F1 expression
Both graphs' samples and x-axis are lined up

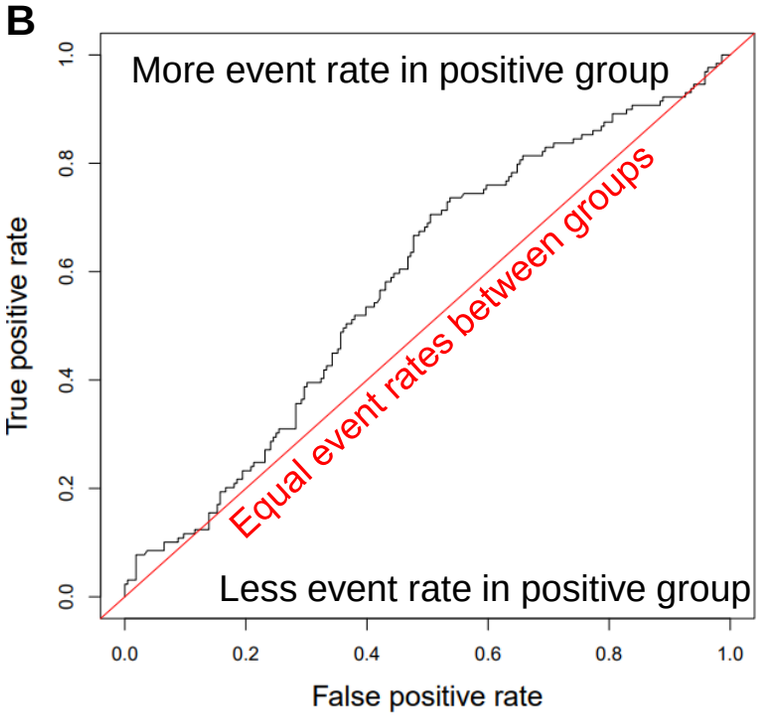
Fig S1.

Supplemental Figure 1. Detailed labeling of LOCC graphs. The LOCC ranking is a graph that plots individual samples values and their respective ranking across all samples. The samples are color coded according to mutation status. The x-axis of ranking directly corresponds to the x-axis of the LOCC cutoff selection graph such that each sample are the same. Therefore, a vertical line from the cutoff selection matches the sample expression and ranking on the LOCC ranking. The cutoff selection has many lines labeled and color coded for ease of visualization. Ideally, the LOCC ranking and cutoff selection are set up so that the hazard ratio is above the red line (HR = 1 as it is easier to visualize HR above 1 than HR below 1. The best cutoff is usually the one with the most significance which is the lowest p-value, visually shown by a peak in the yellow line. However, exceptions can occur if one peak is very close to the edge while comparable peaks exist closer to the middle of the graph. After the cutoff is selected, a corresponding Kaplan-Meier curve is graphed to match the cutoff. The two groups from the Kaplan-Meier curve correspond to the two groups separated by the cutoff on the LOCC ranking. Theoretically, a Kaplan-Meier graph can be plotted for every point on the LOCC graphs which is why there is a large amount of information contained in LOCC.

A



B



C

Test	Death	No Death
Positive	True positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

$$\text{Sensitivity} = \frac{TP}{TP+FN} \qquad \text{Specificity} = \frac{TN}{TN+FP}$$

True Positive Rate (TPR) = Sensitivity

False Positive Rate (FPR) = 1 - Specificity

$$\text{Positive Predictive Value (PPV)} = \frac{TP}{TP+FP}$$

$$\text{Negative Predictive Value (NPV)} = \frac{TN}{TN+FN}$$

$$\text{Hazard ratio} = \frac{\frac{TP}{TP+FP}}{\frac{FN}{FN+TN}} = \frac{PPV}{1-NPV}$$

D

If hazard ratio = 1, then

(1) $\frac{\frac{TP}{TP+FP}}{\frac{FN}{FN+TN}} = 1$

(2) $\frac{TP}{TP+FP} = \frac{FN}{FN+TN}$

(3) TP(FN+TN) = FN(TP+FP)
(4) TP*FN + TP*TN = FN*TP+FN*FP
(5) TP*TN = FN*FP

E

If Sensitivity = 1 – Specificity (TPR=FPR), then

(1) $\frac{TP}{TP+FN} = (1 - \frac{TN}{TN+FP}) = \frac{FP}{TN+FP}$

(2) TP(TN+FP) = FP(TP+FN)
(3) TP*TN + TP*FP = FP*TP + FP*FN
(4) TP*TN = FP*FN

Fig S2.

Supplemental Figure 2. Visual and mathematical comparison of LOCC and ROC curve. (A) The HR line of the LOCC graph for *E2F1* expression in TCGA hepatocellular carcinoma was plotted in black. The red line represent a HR = 1. If the HR is above the red line, there are more a higher risk of death associated with the experimental group whereas if the black line is below the red line, there is a lower risk of death associated with the experimental group. (**B**) A ROC curve was plotted for the *E2F1* expression in TCGA hepatocellular carcinoma. A red line is used to show where the true positive rate (TPR) equals the false positive rate (FPR). This line is also referred to as a random classifier because it cannot differentiate true or false positives. Above the red line is where there is an increased rate of events in the test group while being below the red line is a decreased rate of events in the test group. (**C**) A table shows the groups classified by the predictor and the outcome. Statistical equations and abbreviations are listed. (**D**) Equations are calculated under the assumption HR is 1 to calculate the relationship between groups. (**E**) Equations are calculated under the assumption TPR = FPR to calculate the relationship between groups.

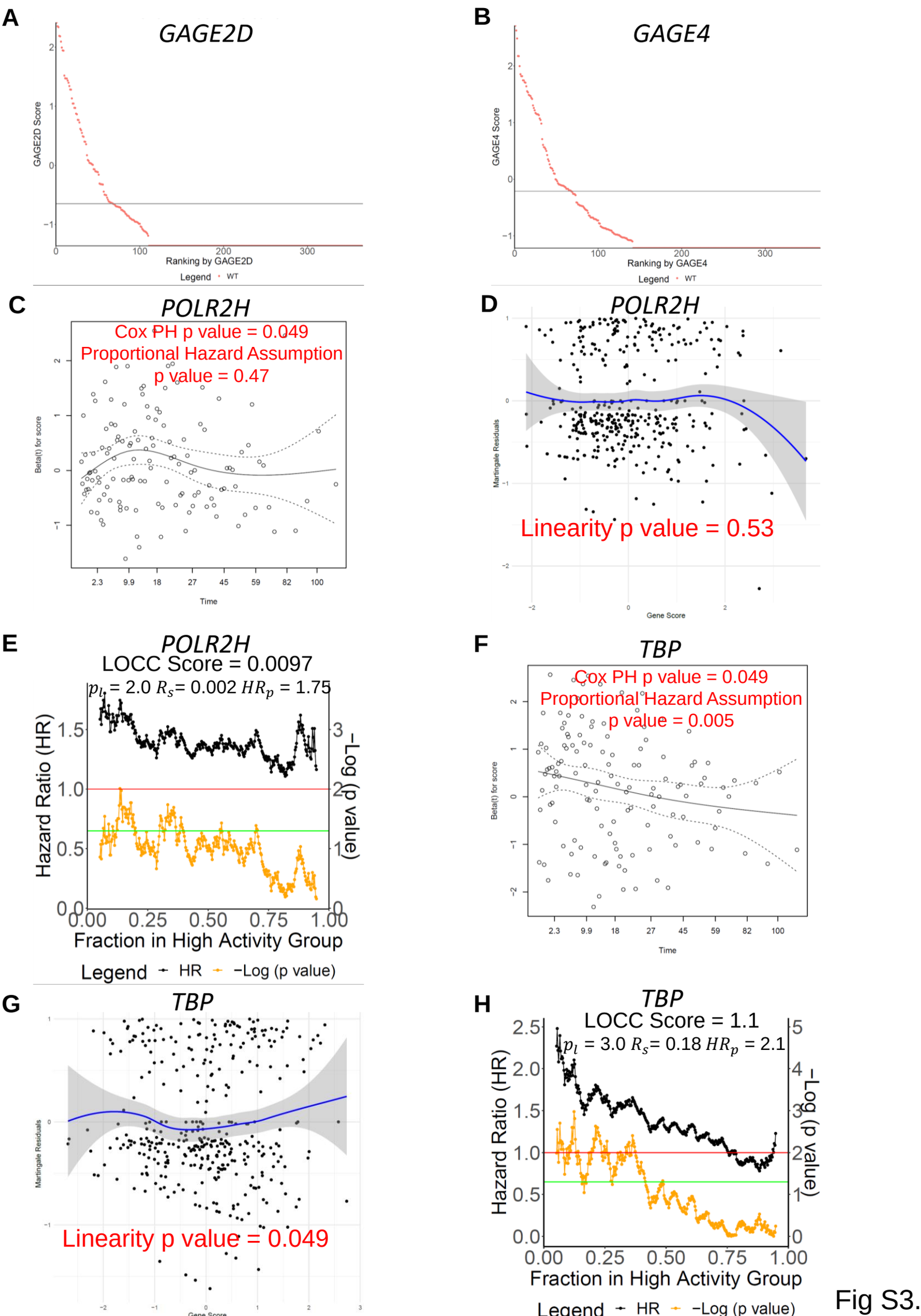
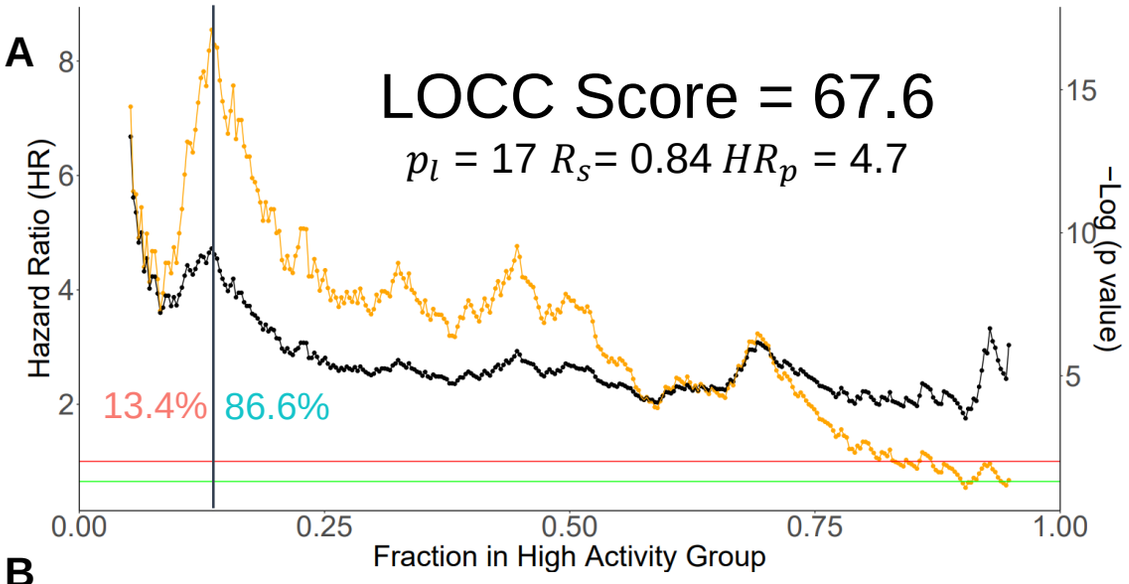
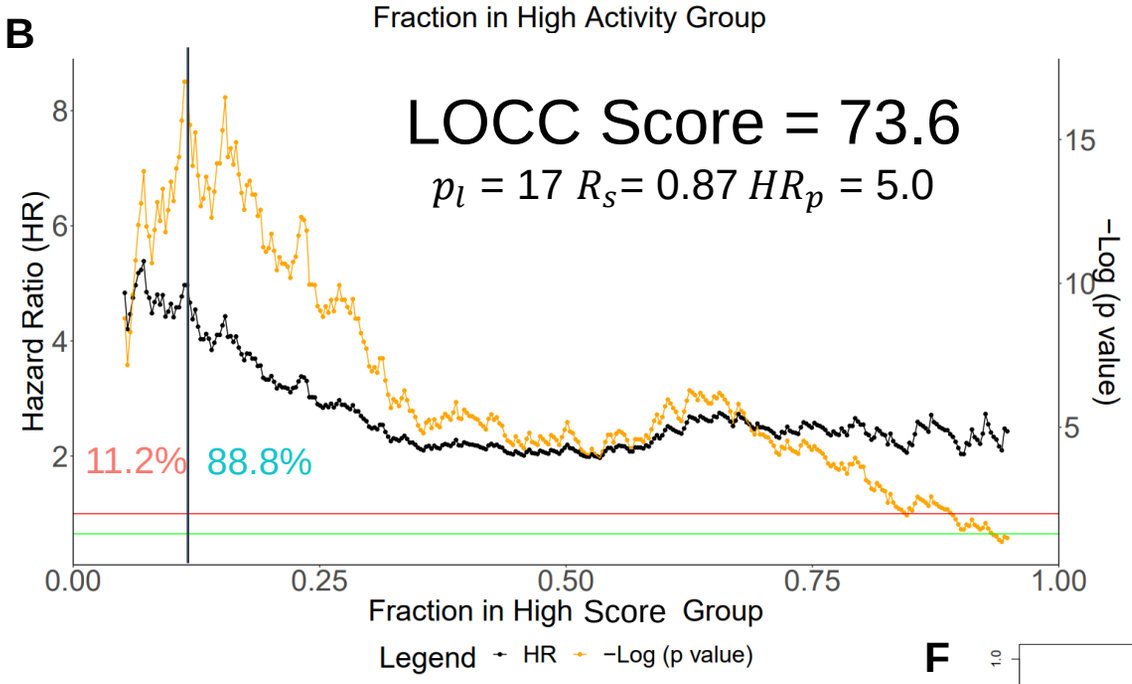
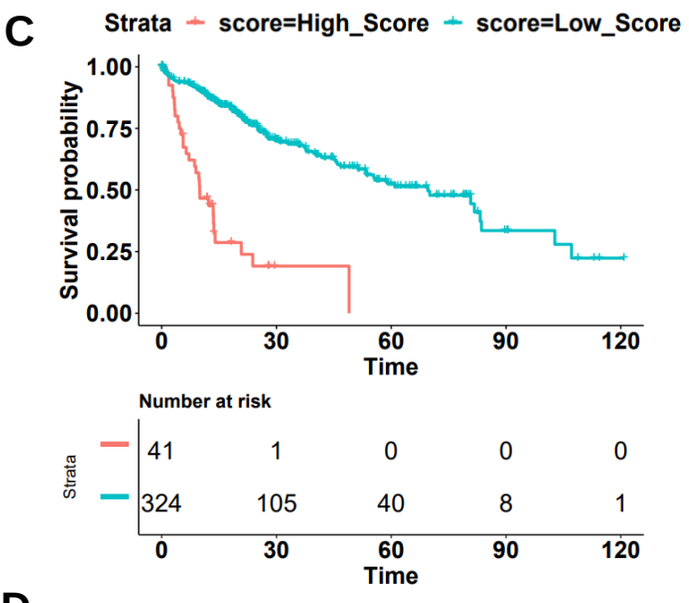

A
GAGE2D
GAGE2D Score
Ranking by GAGE2D
Legend WT
B
GAGE4
GAGE4 Score
Ranking by GAGE4
Legend WT
C
POLR2H
Cox PH p value = 0.049
Proportional Hazard Assumption
p value = 0.47
Beta(t) for score
Time
D
POLR2H
Martingale Residuals
Linearity p value = 0.53
Gene Score
E
POLR2H
LOCC Score = 0.0097
p_l = 2.0 R_s= 0.002 HR_p = 1.75
Hazard Ratio (HR)
−Log (p value)
Fraction in High Activity Group
Legend HR −Log (p value)
F
TBP
Cox PH p value = 0.049
Proportional Hazard Assumption
p value = 0.005
Beta(t) for score
Time
G
TBP
Martingale Residuals
Linearity p value = 0.049
Gene Score
H
TBP
LOCC Score = 1.1
p_l = 3.0 R_s= 0.18 HR_p = 2.1
Hazard Ratio (HR)
−Log (p value)
Fraction in High Activity Group
Legend HR −Log (p value)

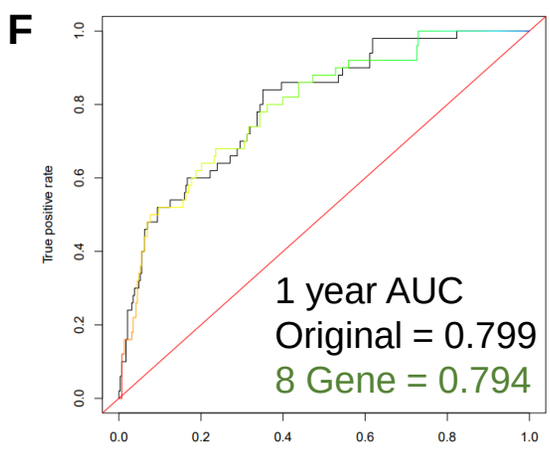
Fig S3.

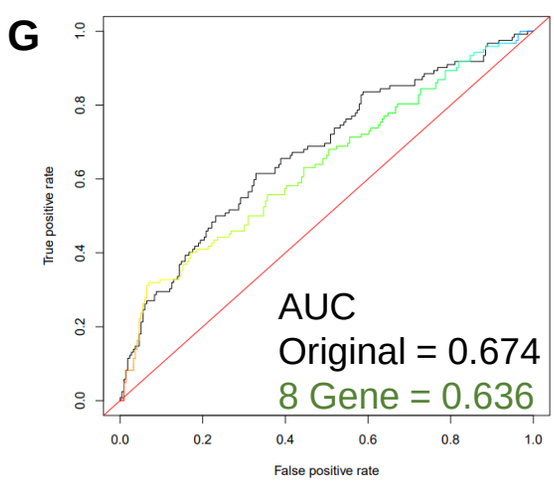
Supplemental Figure 3: Issues with Cox PH modeling in LIHC analysis. (A) An activity graph of GAGE2D show no expression for the majority of tumor samples. **(B)** An activity graph of GAGE4 show no expression for the majority of tumor samples. **(C)** A proportional hazard test using Schoenfeld residuals plot demonstrates *POLR2H* does not violate the proportional hazard assumption. **(D)** A linearity test using Martingale residual plot demonstrates *POLR2H* does not violate the linearity assumption. **(E)** LOCC analysis of *POLR2H* is shown. **(F)** A proportional hazard test using Schoenfeld residuals plot demonstrates *TBP* does violate the proportional hazard assumption. **(G)** A linearity test using Martingale residual plot demonstrates *TBP* does violate the linearity assumption. **(H)** LOCC analysis of *TBP* is shown.











D

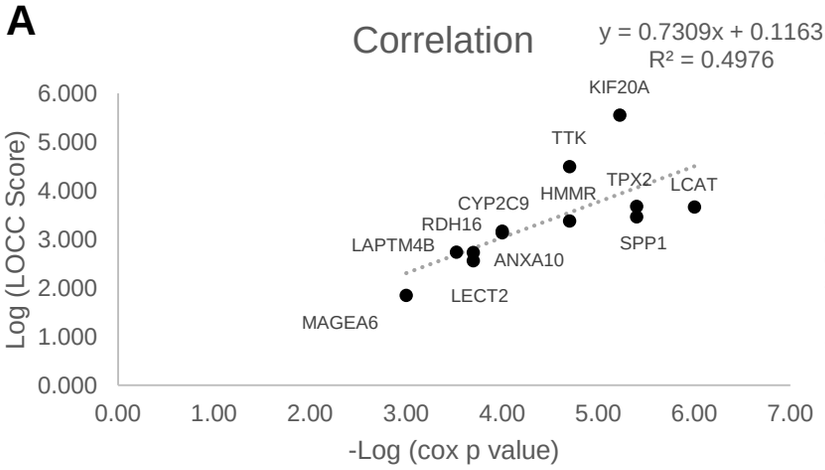
Cox Regression	β Coefficient	Exp (β)	Concordance	Log rank test (p value)
Original Risk	0.893	2.44	0.699	1.0E-11
8-Gene Risk	1.04	2.82	0.696	2.0E-10

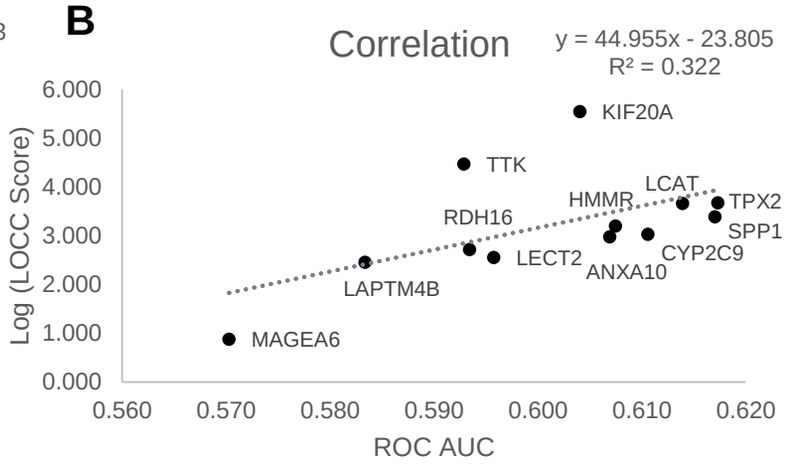
E

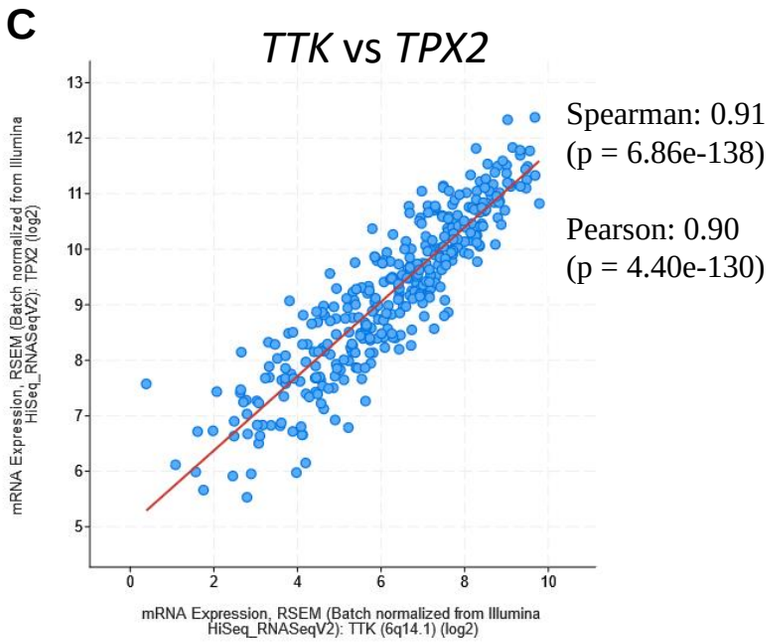
LOCC Analysis	-log (p value)	Percentage (<0.01)	Highest HR	Lowest HR	LOCC Score	AIC
Original Risk	17.093	0.837	4.723	1.942	67.6	1283.8
8-Gene Risk	17.008	0.871	4.971	1.972	73.6	1282.2

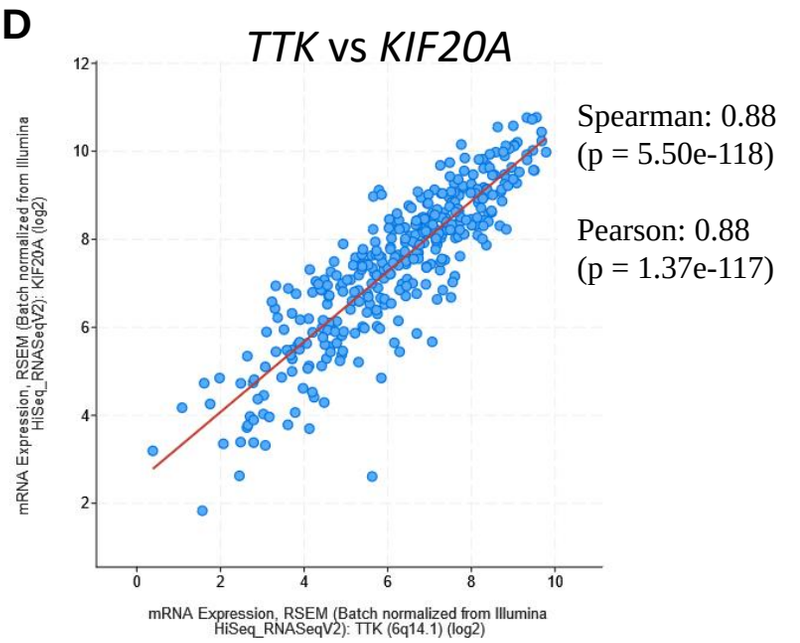
Fig S4.

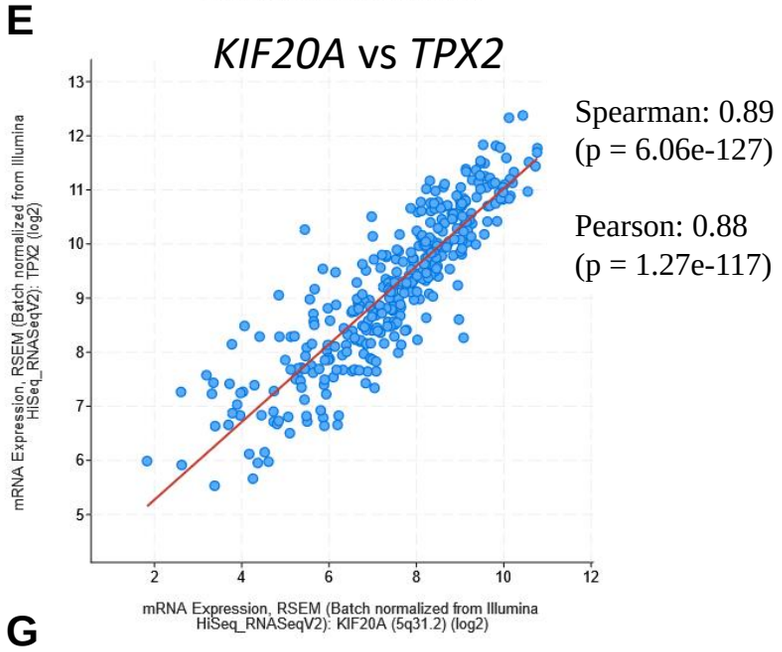
Supplemental Figure 4: Analysis and Comparison of 12-gene original and 8 gene modified RISK score (A) LOCC graph of the original risk gene set is shown. The most significant cutoff is marked with a vertical line. **(B)** LOCC graph of the modified 8-gene set is shown. The most significant cutoff is marked with a vertical line. **(C)** The Kaplan-Meier shows difference in survival using the most significant cutoff for the modified 8-gene set. **(D)** A table shows the Cox regression analysis results for original risk and 8-gene risk gene sets. **(E)** A table shows LOCC analysis between the original risk and 8-gene risk gene set. Additionally, the Akaike information criterion (AIC) is shown between the two models.**(F)** A comparison of 1-year ROC Area Under the Curve (AUC) for original risk and 8-gene risk is shown. The black is the original risk while the colored is the modified 8-gene set. **(G)** A comparison of ROC Area Under the Curve (AUC) for original risk and 8-gene risk is shown. The black is the original risk while the colored is the modified 8-gene set.

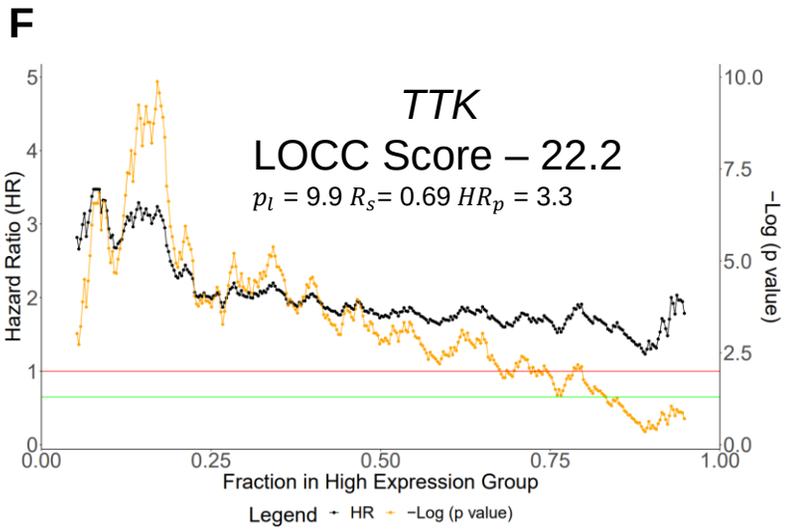

A
Correlation
y = 0.7309x + 0.1163
R² = 0.4976
KIF20A
TTK
TPX2
LCAT
HMMR
CYP2C9
RDH16
LAPTM4B
ANXA10
SPP1
LECT2
MAGEA6
Log (LOCC Score)
-Log (cox p value)


B
Correlation
y = 44.955x - 23.805
R² = 0.322
KIF20A
TTK
LCAT
HMMR
TPX2
RDH16
SPP1
LECT2
CYP2C9
ANXA10
LAPTM4B
MAGEA6
Log (LOCC Score)
ROC AUC


C
TTK vs TPX2
Spearman: 0.91
(p = 6.86e-138)
Pearson: 0.90
(p = 4.40e-130)
mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2): TPX2 (log2)
mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2): TTK (6q14.1) (log2)


D
TTK vs KIF20A
Spearman: 0.88
(p = 5.50e-118)
Pearson: 0.88
(p = 1.37e-117)
mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2): KIF20A (log2)
mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2): TTK (6q14.1) (log2)


E
KIF20A vs TPX2
Spearman: 0.89
(p = 6.06e-127)
Pearson: 0.88
(p = 1.27e-117)
mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2): TPX2 (log2)
mRNA Expression, RSEM (Batch normalized from Illumina HiSeq_RNASeqV2): KIF20A (5q31.2) (log2)


F
TTK
LOCC Score – 22.2
p_l = 9.9 R_s= 0.69 HR_p = 3.3
Hazard Ratio (HR)
−Log (p value)
Fraction in High Expression Group
Legend HR −Log (p value)

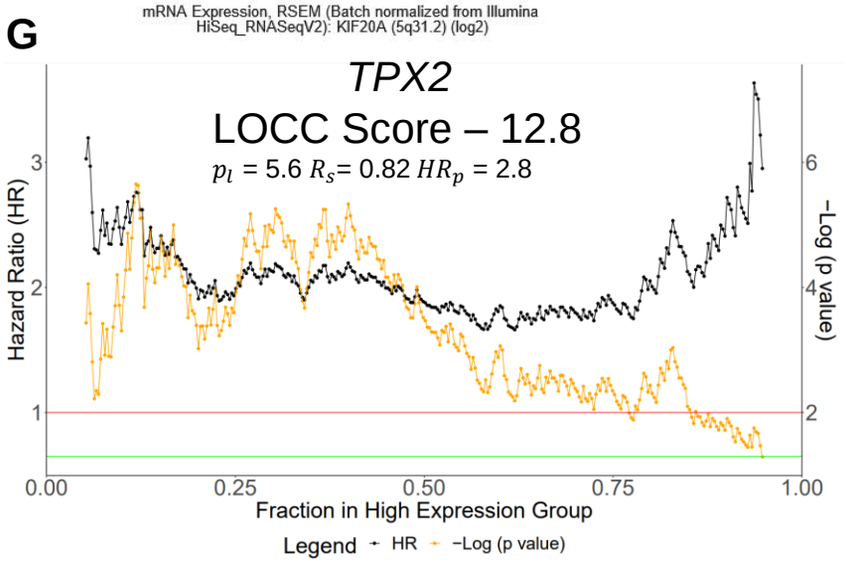

G
TPX2
LOCC Score – 12.8
p_l = 5.6 R_s= 0.82 HR_p = 2.8
Hazard Ratio (HR)
−Log (p value)
Fraction in High Expression Group
Legend HR −Log (p value)

Fig S5.

Supplemental Figure 5. Analysis of Gene Predictors from the 12-gene RISK score

(A) A scatterplot is plotted between the Log_2(LOCC Score) and -Log_{10} (cox p-value). There is a strong correlation (r = 0.705) between the two markers. **(B)** A scatterplot is plotted between the Log_2(LOCC Score) and ROC AUC. There is a modest correlation (r = 0.576) between the two markers. **C-E.** Scatterplots of gene expression are plotted between the following genes: (**C**) *TTK* vs *TPX2*, (**D**) *TTK* vs *KIF20A*, (**E**) *KIF20A* vs *TPX2*. (**F**) The LOCC graph is shown for *TTK.* (**G**) The LOCC graph is shown for *TPX2.*

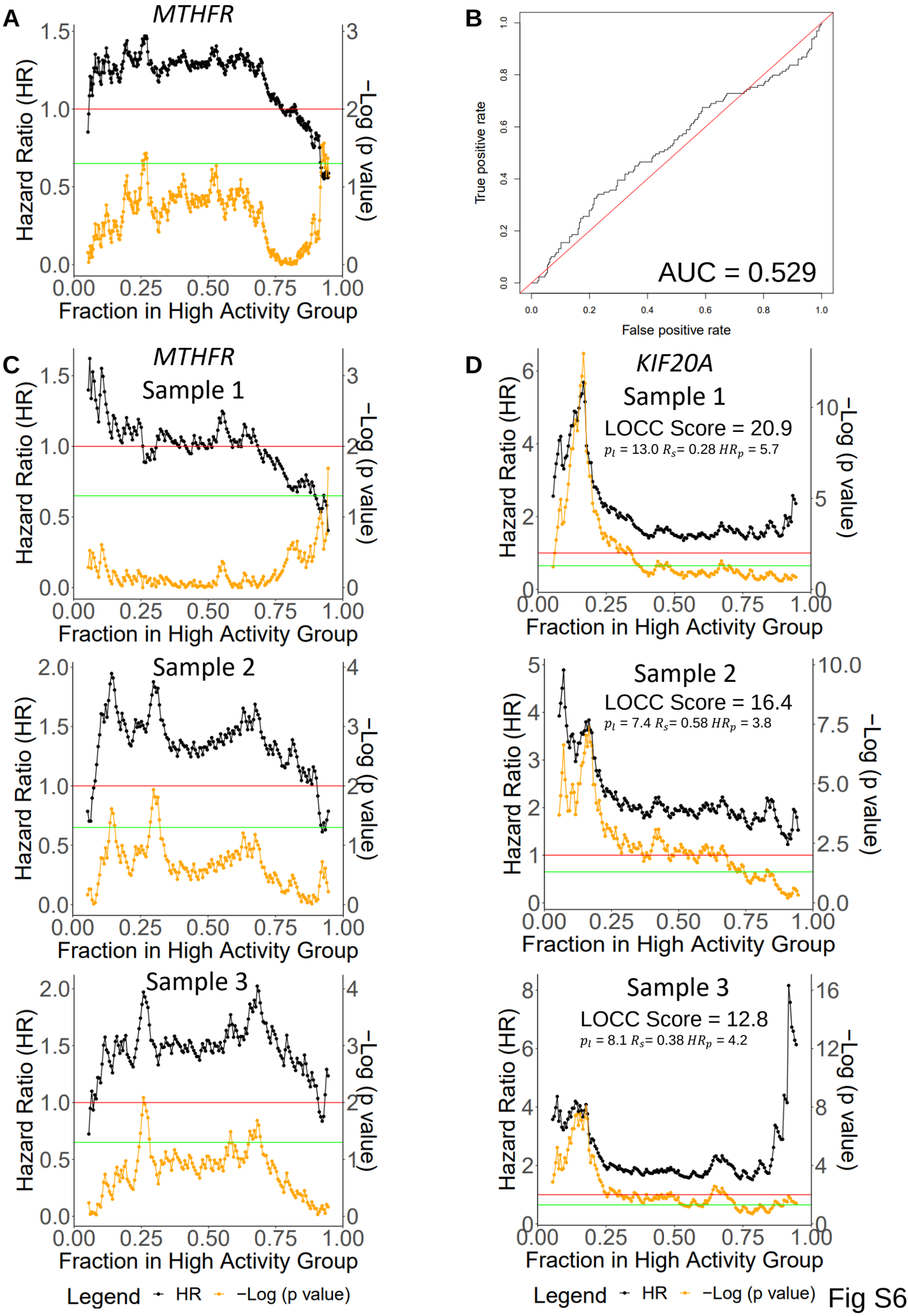
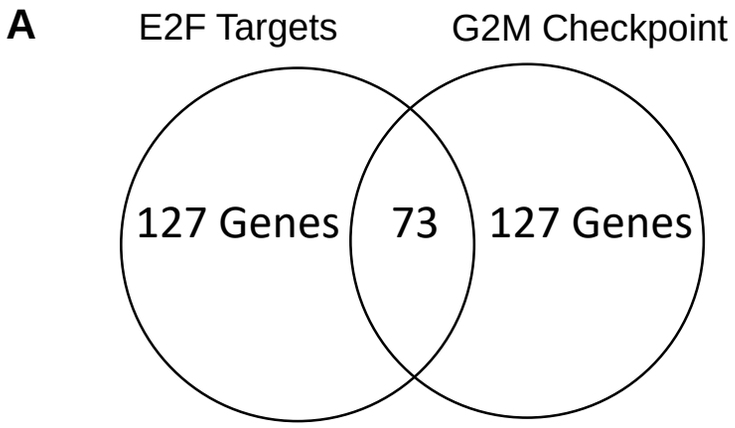
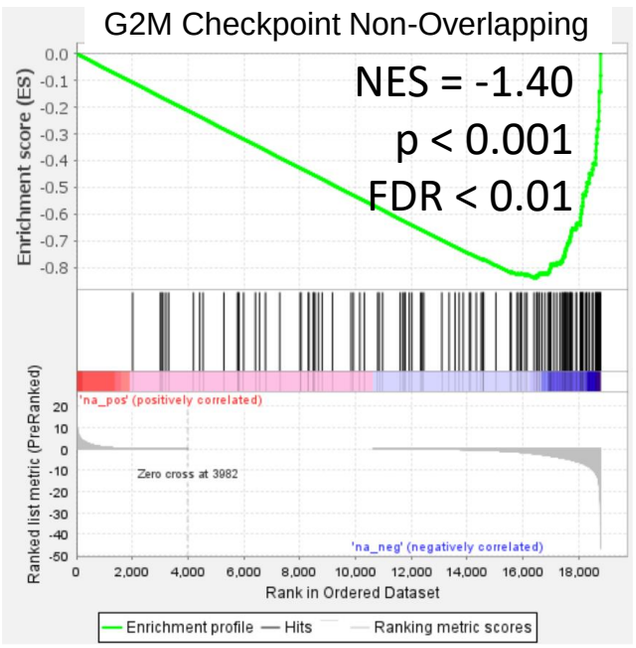
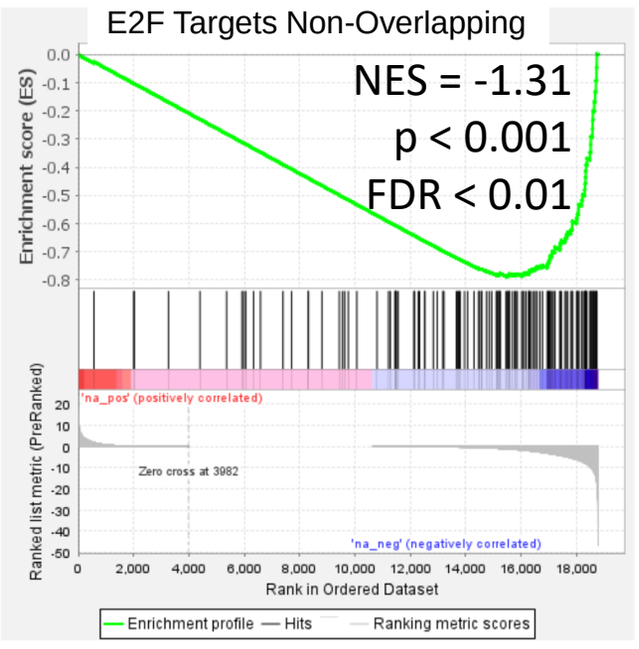

A
MTHFR
Hazard Ratio (HR)
−Log (p value)
Fraction in High Activity Group
B
True positive rate
False positive rate
AUC = 0.529
C
MTHFR
Sample 1
Sample 2
Sample 3
D
KIF20A
Sample 1
LOCC Score = 20.9
p_l = 13.0 R_s= 0.28 HR_p = 5.7
Sample 2
LOCC Score = 16.4
p_l = 7.4 R_s= 0.58 HR_p = 3.8
Sample 3
LOCC Score = 12.8
p_l = 8.1 R_s= 0.38 HR_p = 4.2
Legend HR −Log (p value)

Fig S6.

Supplemental Figure 6. Examples of Sampling by LOCC Visualization (A) A LOCC graph of *MTHFR* is shown. **(B)** A ROC curve of MTHFR is shown with the AUC calculated to be 0.529. **(C)** LOCC graphs of *MTHFR* in various samples of the full dataset are shown. **(D)** LOCC graphs of *KIF20A* in various samples of the full dataset are shown.


A
E2F Targets
G2M Checkpoint
127 Genes
73
127 Genes


B
G2M Checkpoint Non-Overlapping
NES = -1.40
p < 0.001
FDR < 0.01
Enrichment score (ES)
'na_pos' (positively correlated)
Zero cross at 3982
'na_neg' (negatively correlated)
Ranked list metric (PreRanked)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores


C
E2F Targets Non-Overlapping
NES = -1.31
p < 0.001
FDR < 0.01
Enrichment score (ES)
'na_pos' (positively correlated)
Zero cross at 3982
'na_neg' (negatively correlated)
Ranked list metric (PreRanked)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

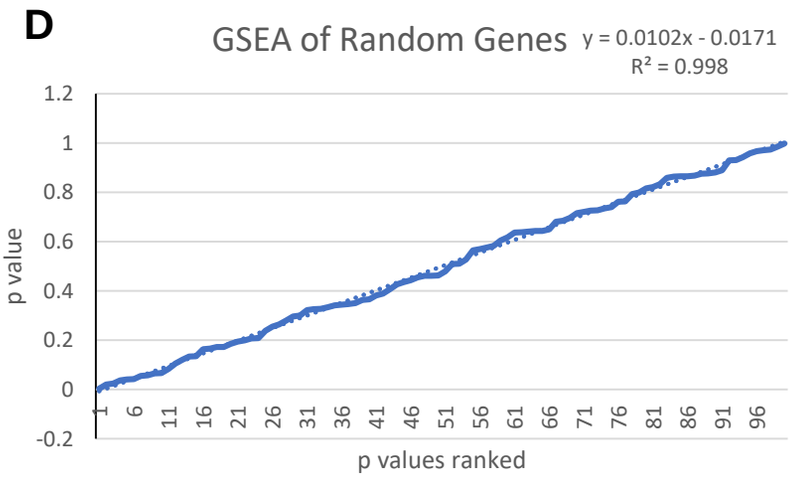

D
GSEA of Random Genes
y = 0.0102x - 0.0171
R² = 0.998
p value
p values ranked

Fig S7.

Supplemental Figure 7. GSEA without overlapping genes (A) A Venn diagram showing the number of overlapping and distinct genes in Hallmark E2F target gene and G2M checkpoint gene sets. **(B)** The enrichment plot of the hallmark G2M checkpoint with removal of overlapping genes with E2F target genes is shown using LOCC scores. **(C)** The enrichment plot of the hallmark E2F target genes with removal of overlapping genes with G2M checkpoint is shown using LOCC scores. **(D)** Gene sets of 200 randomly selected genes were evaluated in GSEA preranked using LOCC scores. A graph of the p values is plotted against the ranking of the p values with the line of best fit shown.