Science Advances
AAAS

Supplementary Materials for

Widespread hypertranscription in aggressive human cancers

Matthew Zatzman *et al.*

Corresponding author: Adam Shlien, adam.shlien@sickkids.ca

Sci. Adv. **8**, eabn0238 (2022)
DOI: 10.1126/sciadv.abn0238

This PDF file includes:

Figs. S1 to S13
Legends for data S1 to S6

Other Supplementary Materials for this manuscript includes the following:

Data S1 to S6

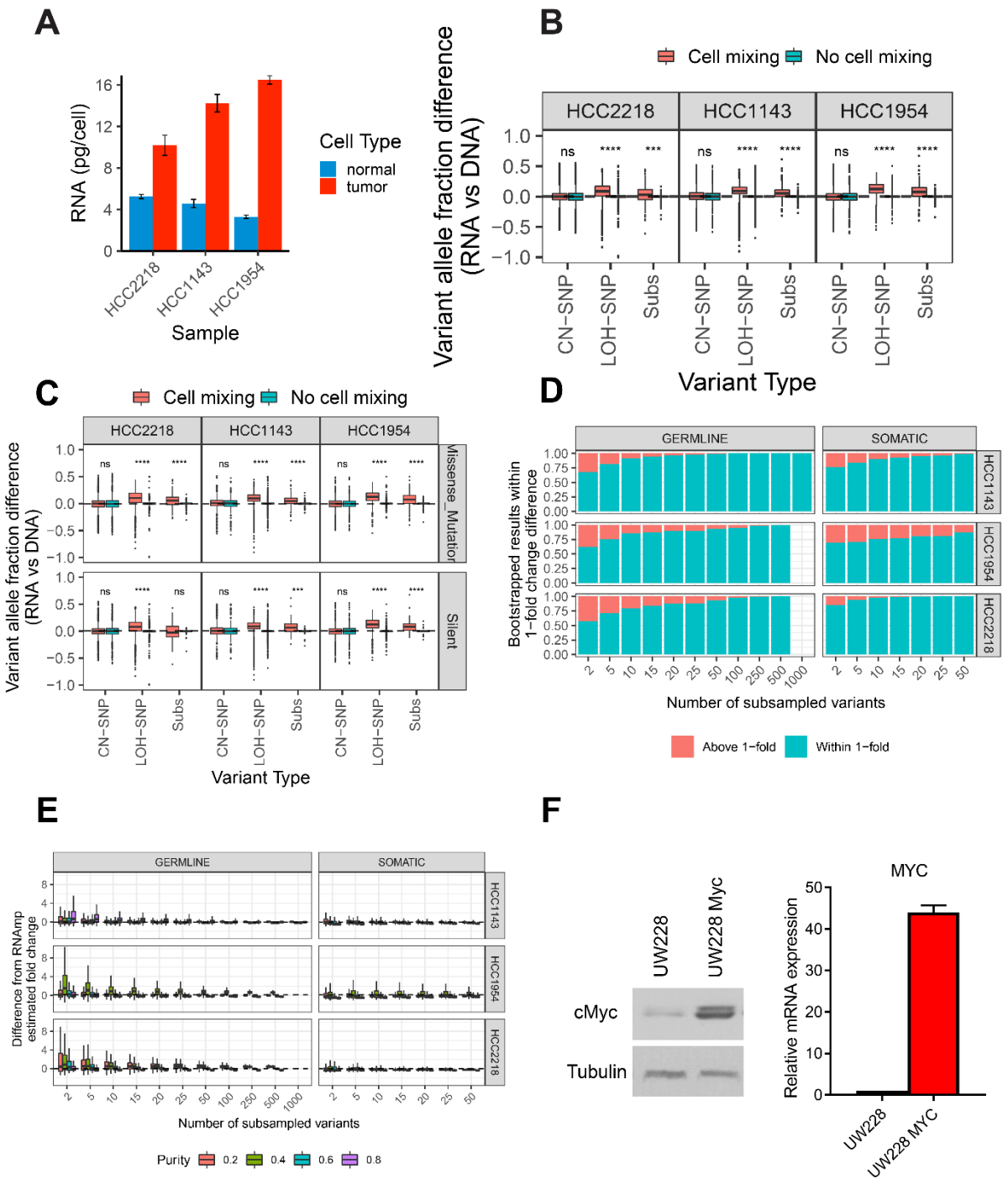


Fig. S1.

RNAmp method validation using cell line mixing and TCGA variants. A) RNA output as measured from cell counting and RNA quantification for three breast cancer cell lines and their matched lymphocyte normal cell line. Tumor cell lines have increased RNA output. **B)** Variant allele fraction difference boxplots of copy-neutral SNP (CN-SNP), LOH-SNP, and somatic substitution variants of each cell line used in either cell mixtures, or purified cell lines. Tumor markers (subs and LOH-SNP) have increased their VAF RNA relative to DNA when mixed with

normal cells, representing their increased RNA contribution (****p < 0.0001, ***p < 0.001, ns = not significant, student's *t*-test). Boxplot center line corresponds to the median, box-limits are upper and lower quartiles, and whiskers represent 1.5 * IQR. **C)** Variant allele fraction difference boxplots of copy-neutral SNP (CN-SNP), LOH-SNP, and somatic substitution variants of each cell line used in either cell mixtures, or purified cell lines split by missense and silent variant types. Both missense and silent tumor markers (subs and LOH-SNP) increase their VAF RNA relative to DNA (****p < 0.0001, ***p < 0.001, ns = not significant, student's *t*-test). Boxplots are defined in Supplemental Figure 1B. **D)** Proportion of bootstrapped results within 1-fold difference of RNAmp's estimates using the full sample variant set. **E)** Boxplots depicted the difference between bootstrapped results and RNAmp's estimates using the full sample variant set**. F)** Left – Western blot of Myc induction in UW228 medulloblastoma cells. Right – qRT-PCR for Myc mRNA expression. Error bars correspond to SD.

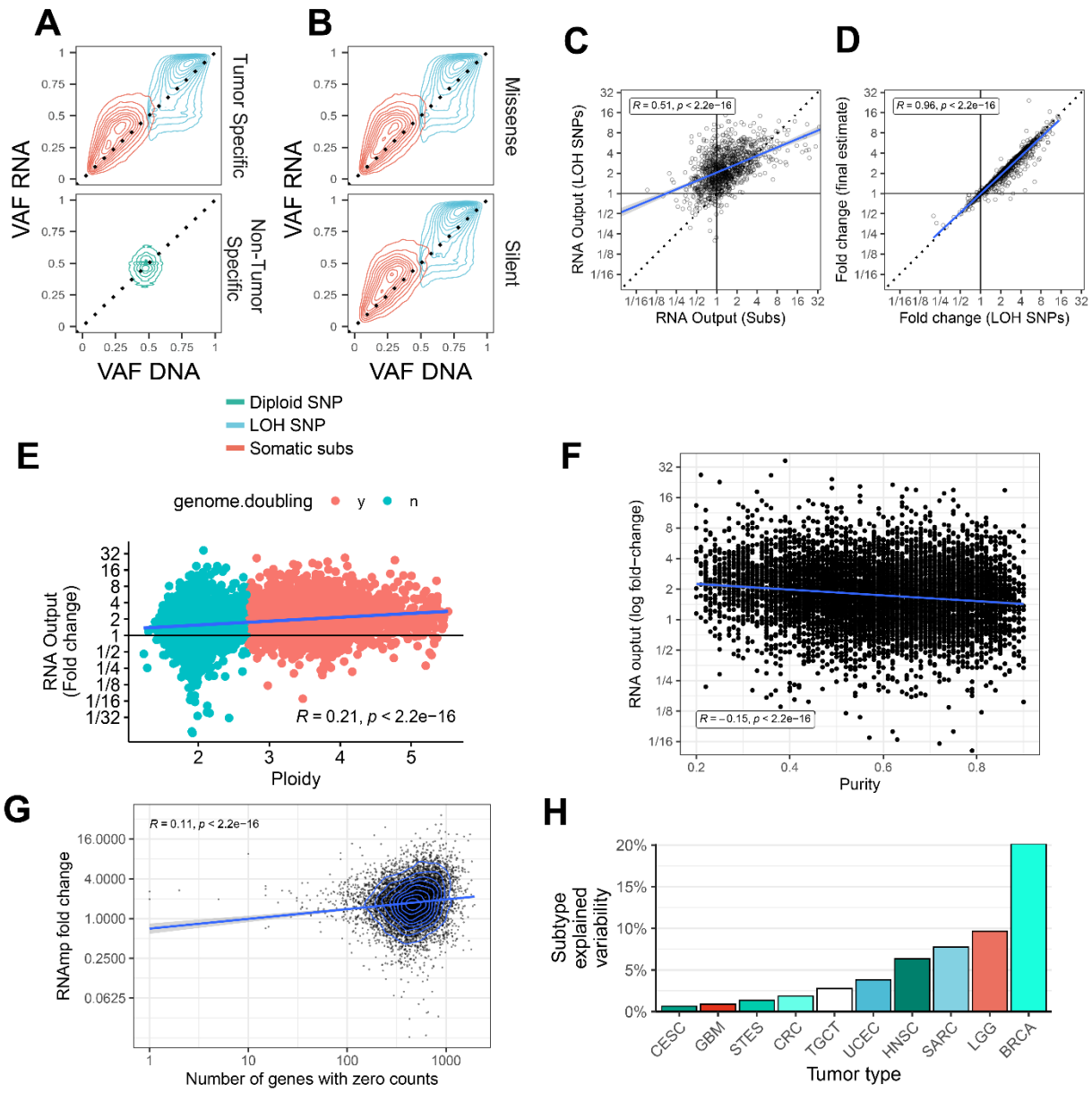


Fig. S2.

RNAmp applied to the TCGA cohort. A) DNA and RNA variant allele fraction distributions for tumor specific (LOH SNPs and Subs) and non-tumor specific variant types (diploid SNPs) in the human pan-cancer dataset. Tumor specific variants have increased VAF RNA relative to DNA. **B)** Missense and silent mutation DNA and RNA variant allele fraction density distributions in the human pan-cancer dataset. Missense and silent mutations have comparable VAF DNA and RNA profiles. **C)** Correlation between RNAmp fold-change results using somatic substitutions (Subs) or LOH-SNPs. **D)** Correlation between RNAmp fold-change results using LOH-SNPs and the final weighted estimate using both LOH-SNPs and substitutions. **E)** Pearson correlation between tumor ploidy and RNA output (R = 0.21, p <2.2e-16). **F)** Pearson correlation between tumor purity and RNA output (R = -0.15, p <2.2e-16). **G)** Pearson

correlation between number of genes with zero counts within a sample and RNA output ($R = 0.11$, $p < 2.2e-16$). **H)** The proportion of variability in RNA output explained within cancers by the differences in tumor subtypes

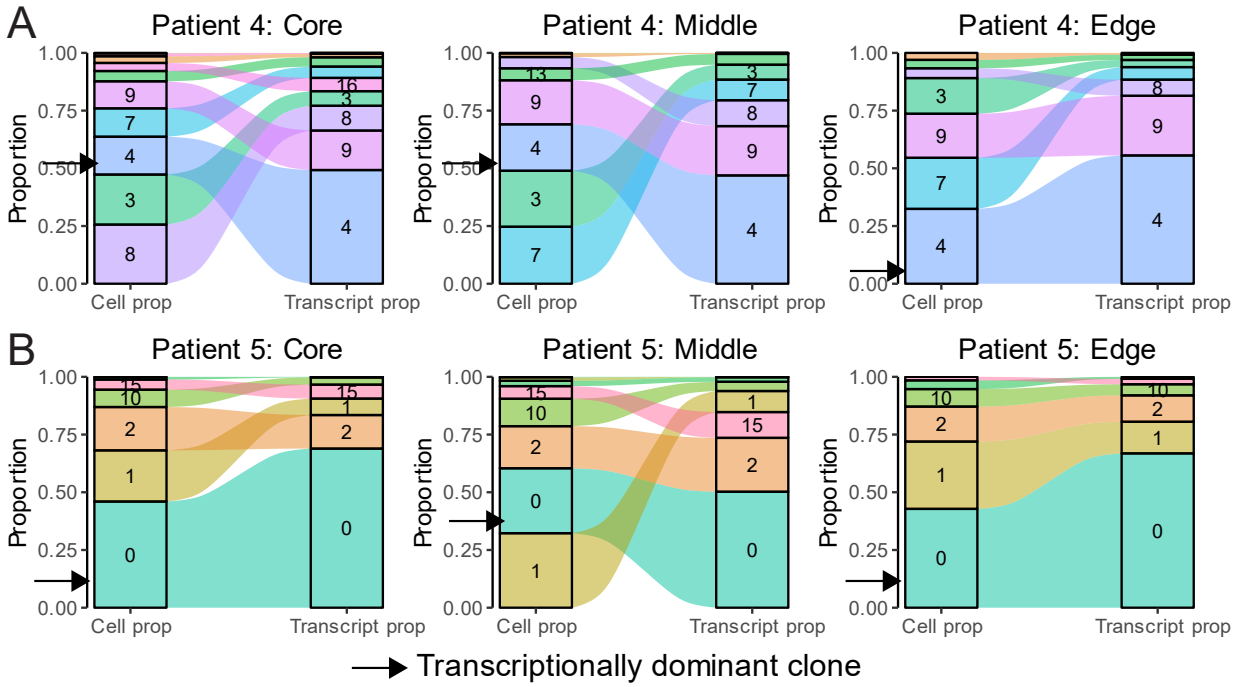


Fig. S3.

Hypertranscription in single tumor cell populations. Flow diagram depicting the proportional cell counts and transcript counts for different tumor subclusters across spatially distinct tumor regions from **(A)** patient 4 and **(B)** patient 5. Arrows indicate transcriptionally dominant clones.


Hypertranscription Association Score
−2 −1 0 1 2
A
PAN
SARC
PRAD
KIRP
PCPG
TGCT
THCA
LIHC
UCEC
LGG
GBM
OV
BLCA
PAAD
HNSC
LUAD
KIRC
STES
BRCA
LUSC
CRC
CESC
SKCM
GLYCOLYSIS
TNFA VIA NFKB
MTORC1
PEROXISOME
IL6 JAK STAT3
APICAL JUNCTION
MITOTIC SPINDLE
IFNa RESPONSE
PI3K AKT MTOR
PROTEIN SECRETION
UV RESPONSE UP
MYC TARGETS V2
BILE ACID METABOLISM
COAGULATION
P53 PATHWAY
HEDGEHOG
TGF BETA
CHOL. HOMEOSTASIS
IFNg RESPONSE
WNT BETA CATENIN
HEME METABOLISM
OXPHOS
MYC TARGETS V1
UPR
ROS PATHWAY
PANCREAS BETA CELLS
G2M CHECKPOINT
ANDROGEN RESPONSE
DNA REPAIR
E2F TARGETS
IL2 STAT5
ANGIOGENESIS
APICAL SURFACE
ER LATE
EMT
HYPOXIA
ER EARLY
KRAS DN
UV RESPONSE DN
SPERMATOGENESIS
XENOBIOTIC METABOLISM
ADIPOGENESIS
NOTCH
INFLAM. RESPONSE
COMPLEMENT
FA METABOLISM
ALLOGRAFT REJECTION
APOPTOSIS
MYOGENESIS
KRAS UP
B
Variability explained (%)
100
75
50
25
0
Variable
Unexplained
Hallmark Pathways
Tumour subtype
Gender
Mutation burden
Age
Ploidy
Tumour Stage
Purity

Fig. S4.

Hallmark expression pathways associated with RNA output. A) Heatmap depicting the association between hallmark expression pathways and RNA output in the pan-cancer cohort, and within individual tumor types. Each box represents the ridge regression coefficient representing the relationship between a pathway's level of relative expression, and global transcriptional output. **B)** The proportion of variability in RNA output explained including hallmark pathway expression data.

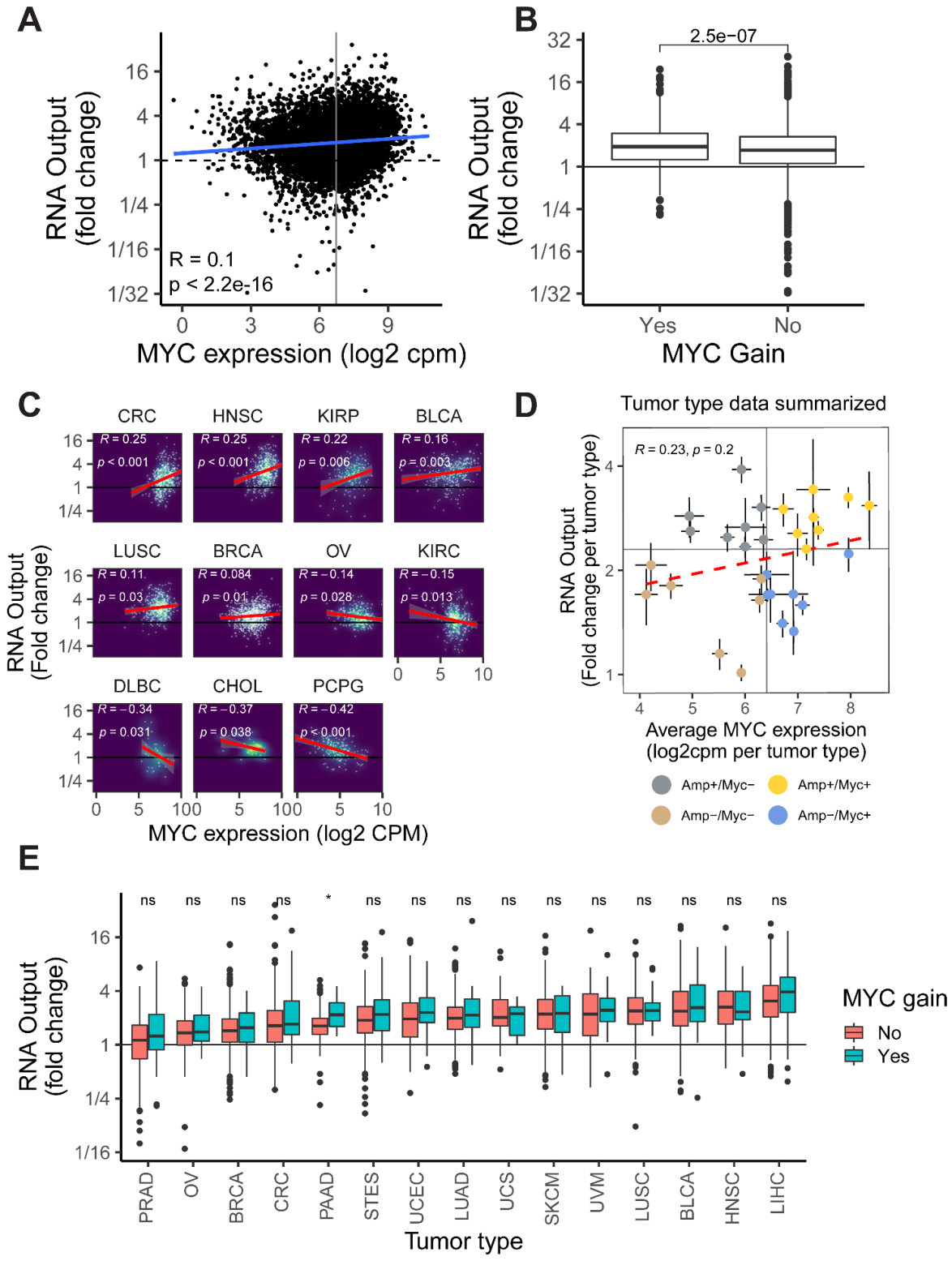
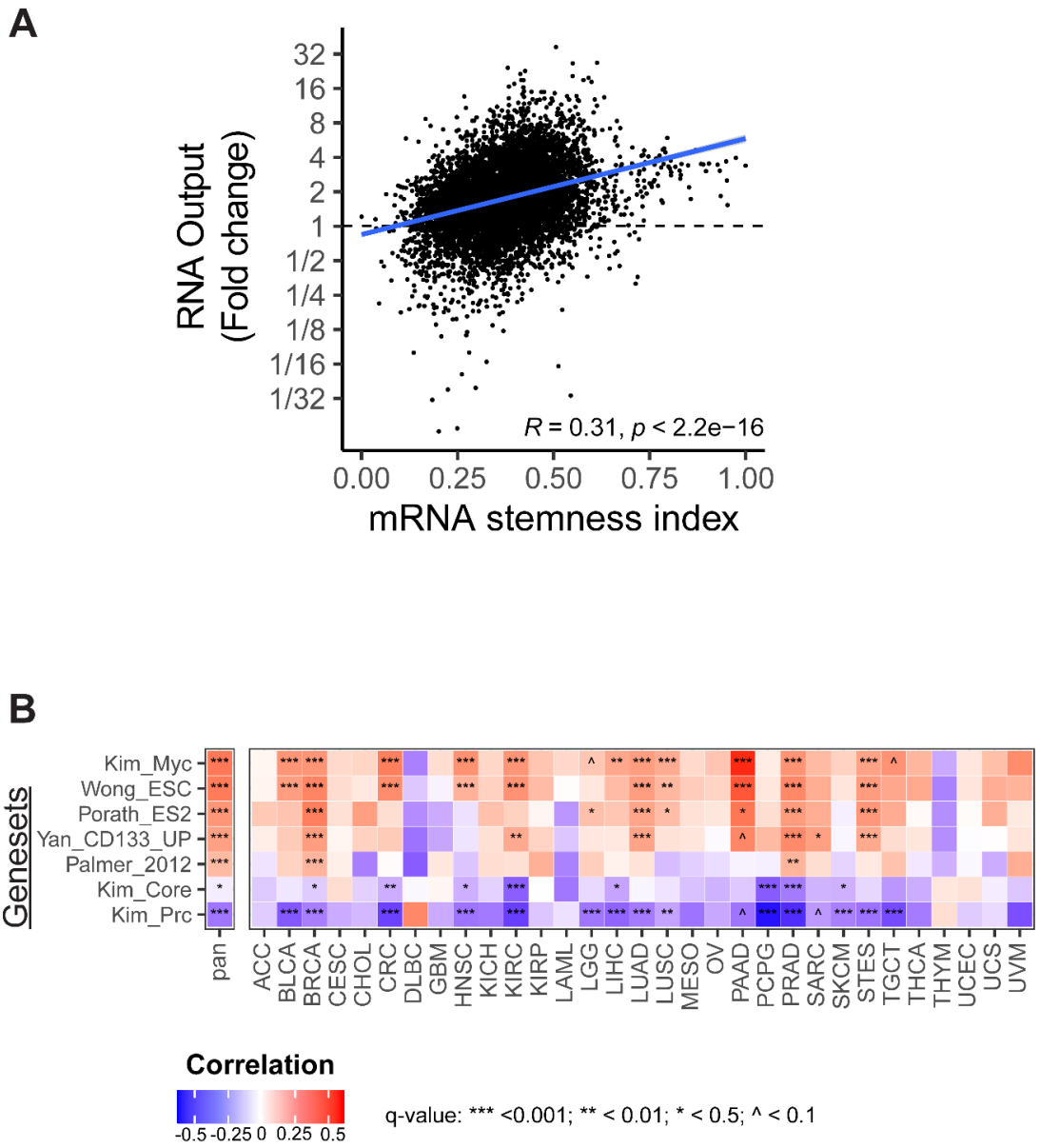
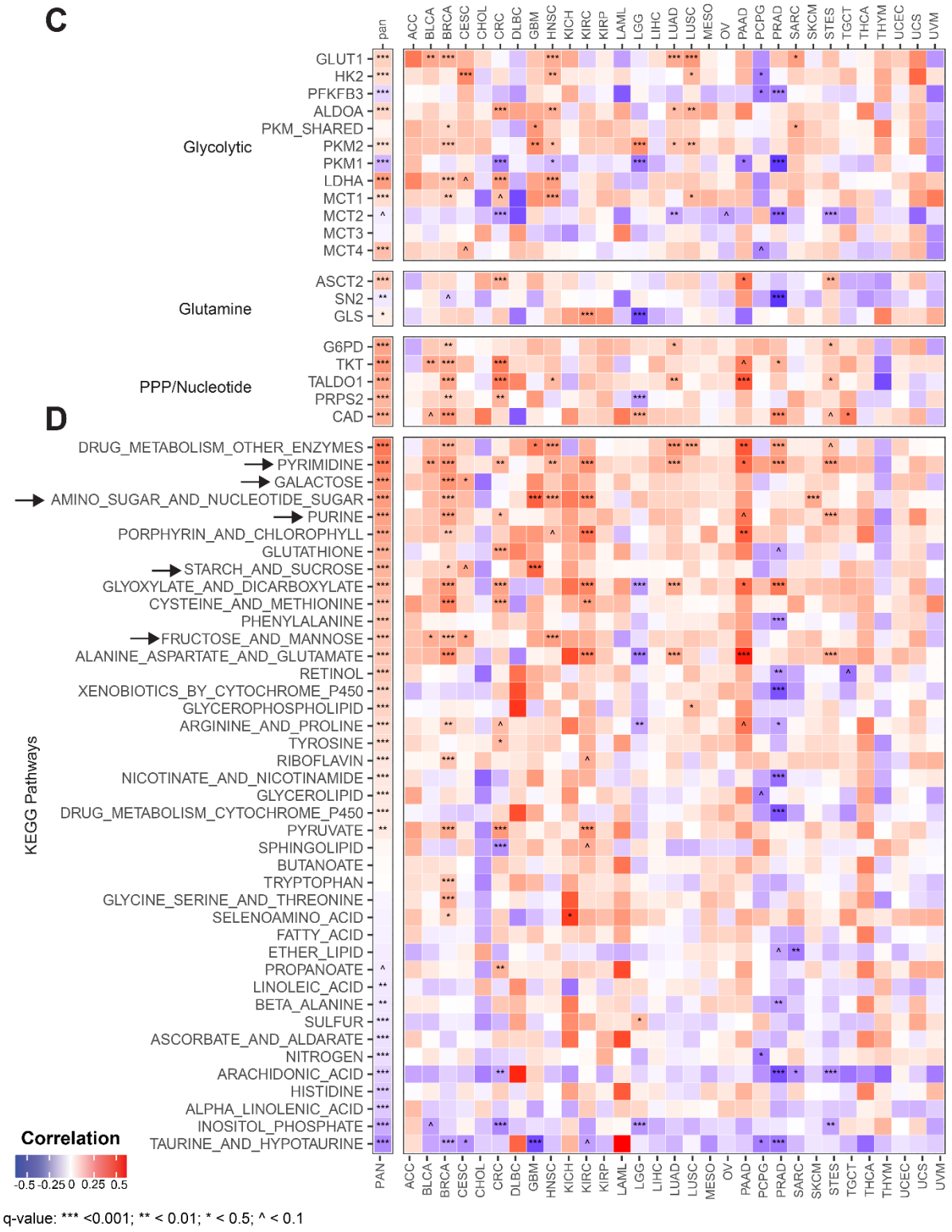

A
RNA Output (fold change)
16
4
1
1/4
1/16
1/32
R = 0.1
p < 2.2e-16
0
3
6
9
MYC expression (log2 cpm)
B
2.5e−07
32
16
4
1
1/4
1/16
1/32
RNA Output (fold change)
Yes
No
MYC Gain
C
CRC
HNSC
KIRP
BLCA
R = 0.25
p < 0.001
R = 0.25
p < 0.001
R = 0.22
p = 0.006
R = 0.16
p = 0.003
LUSC
BRCA
OV
KIRC
R = 0.11
p = 0.03
R = 0.084
p = 0.01
R = −0.14
p = 0.028
R = −0.15
p = 0.013
DLBC
CHOL
PCPG
R = −0.34
p = 0.031
R = −0.37
p = 0.038
R = −0.42
p < 0.001
RNA Output (Fold change)
16
4
1
1/4
0
5
10
100
MYC expression (log2 CPM)
D
Tumor type data summarized
R = 0.23, p = 0.2
RNA Output (Fold change per tumor type)
4
2
1
4
5
6
7
8
Average MYC expression (log2cpm per tumor type)
Amp+/Myc−
Amp+/Myc+
Amp−/Myc−
Amp−/Myc+
E
ns
*
RNA Output (fold change)
16
4
1
1/4
1/16
PRAD
OV
BRCA
CRC
PAAD
STES
UCEC
LUAD
UCS
SKCM
UVM
LUSC
BLCA
HNSC
LIHC
Tumor type
MYC gain
No
Yes

Fig. S5.

Relationship between MYC expression or copy status and RNA output. A) Pearson correlation between MYC gene expression and RNA output (R = 0.1, p<2.2e-16). **B)** Boxplot depicting RNA output levels of tumors with and without MYC copy gains (p = 2.5e-07, student's two-sided *t*-test). Boxplots are defined in Supplemental Figure 1B. **C)** Pearson correlation between MYC expression and RNA output in individual tumor types. Only tumors with a statistically significant (p <0.05) positive or negative correlation are shown. **D)** Pearson correlation between tumor type average MYC expression and RNA output. Vertical and horizontal lines indicate median MYC expression and RNA output respectively. Each point represents a single tumor type, with 95% confidence intervals for RNA output and MYC expression measure shown as lines extended from each point (R = 0.23, p = 0.2). Boxplots are defined in Supplemental Figure 1B. **E)** Boxplots depicting RNA output of samples with or without MYC copy gains in individual tumor types (*p < 0.05, ns = not significant, student's *t*-test).


A
RNA Output
(Fold change)
32
16
8
4
2
1
1/2
1/4
1/8
1/16
1/32
R = 0.31, p < 2.2e−16
0.00
0.25
0.50
0.75
1.00
mRNA stemness index
B
Stemness
Genesets
Kim_Myc
Wong_ESC
Porath_ES2
Yan_CD133_UP
Palmer_2012
Kim_Core
Kim_Prc
pan
ACC
BLCA
BRCA
CESC
CHOL
CRC
DLBC
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
PCPG
PRAD
SARC
SKCM
STES
TGCT
THCA
THYM
UCEC
UCS
UVM
Correlation
-0.5 -0.25 0 0.25 0.5
q-value: *** <0.001; ** < 0.01; * < 0.5; ^ < 0.1



q-value: *** <0.001; ** < 0.01; * < 0.5; ^ < 0.1

Fig. S6.

Metabolic and stemness pathways associated with RNA output. A) Pearson correlation between mRNA stemness index scores and RNA output (R=0.31, p <2.2e-16). **B)** Heatmap depicting the Pearson correlation values for stemness gene-sets and RNA output. **C)** Heatmap depicting the Pearson correlation between metabolic genes and RNA output. **D)** Heatmap depicting the Pearson correlation between KEGG metabolic pathways and RNA output. For figures 1B-D, significant correlations between RNA output and the respective gene or pathways are indicated (after FDR correction).

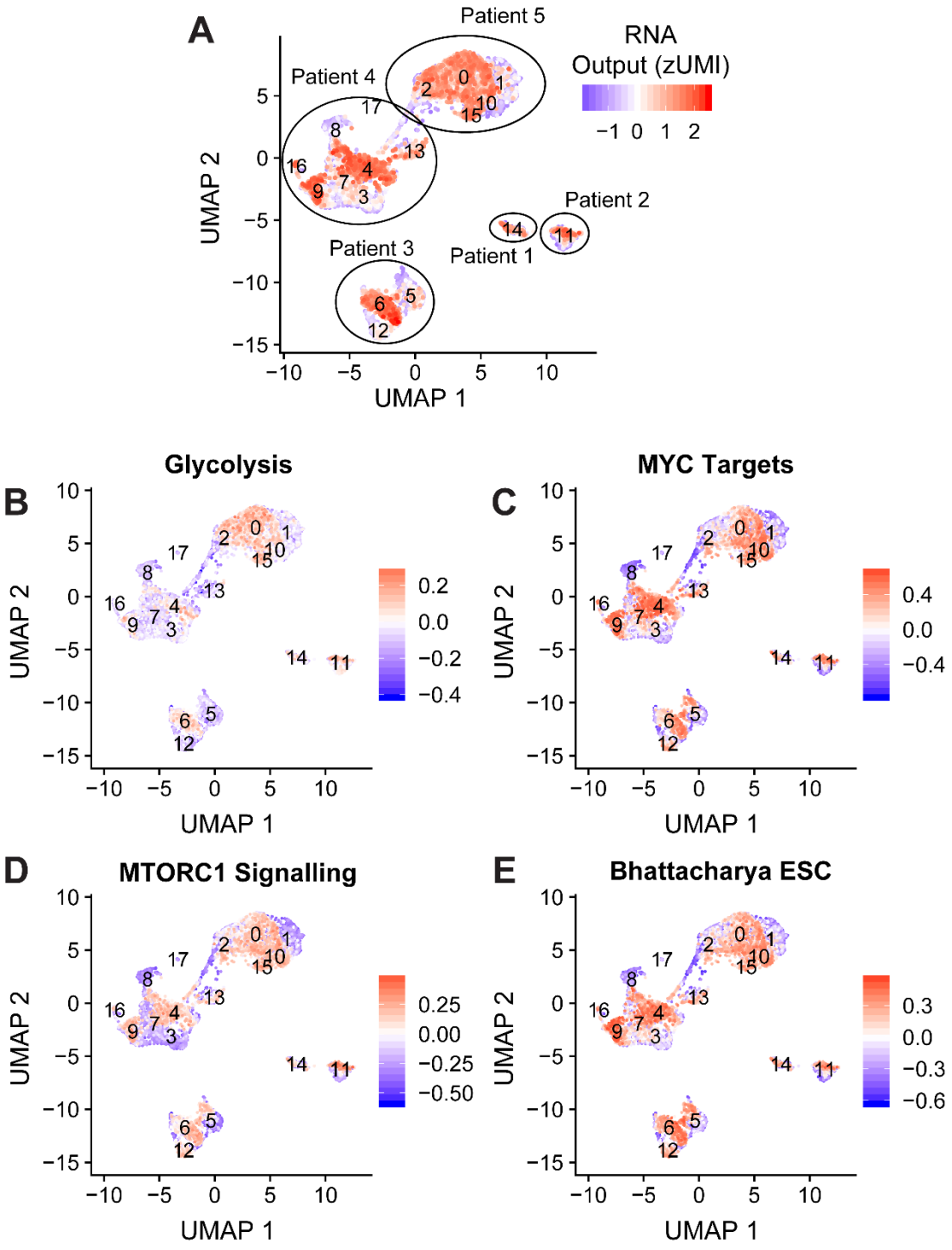
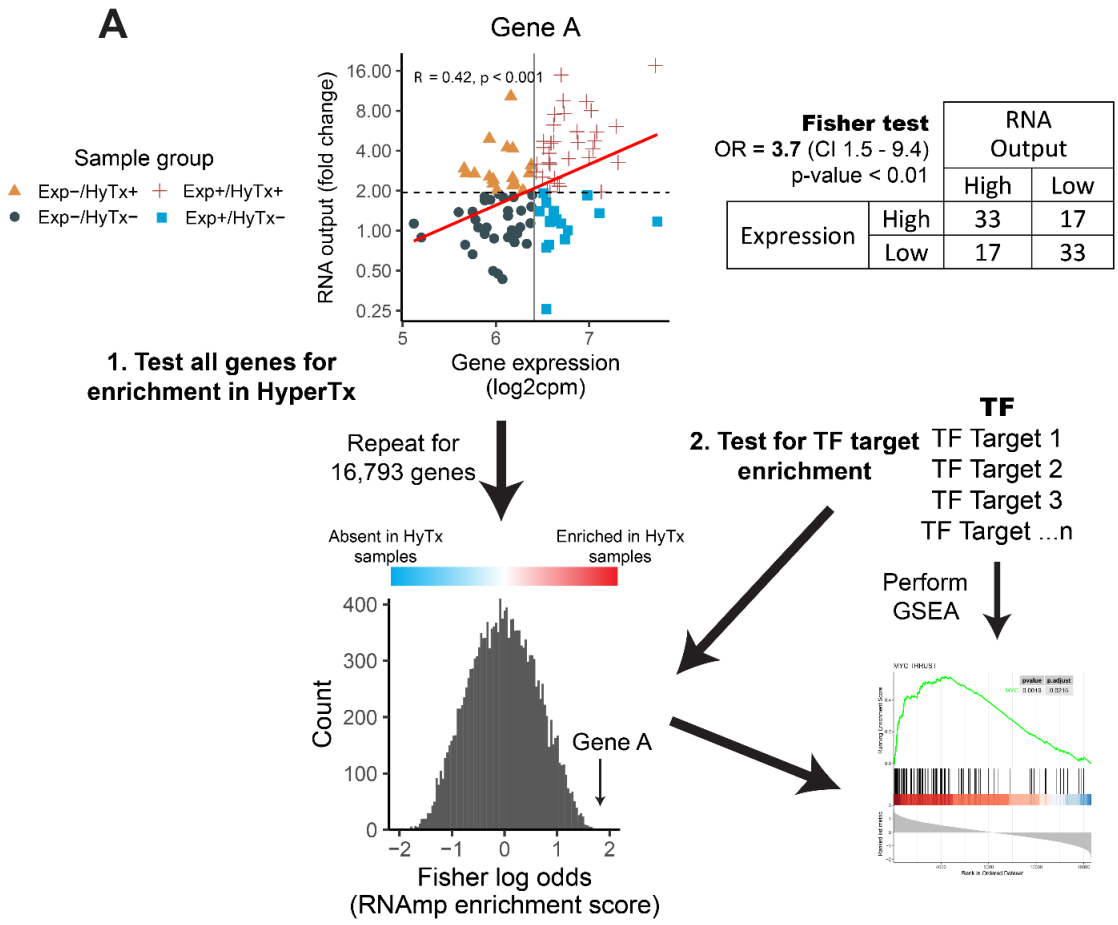


Fig. S7.

Gene expression pathways associated with RNA output in single cells. A) RNA output of single-cells overlaid onto the UMAP expression clusters. **B-E)** Expression levels for glycolysis, MYC targets, MTORC1 and ESC signaling pathways overlaid onto the tumor cell UMAP**.**



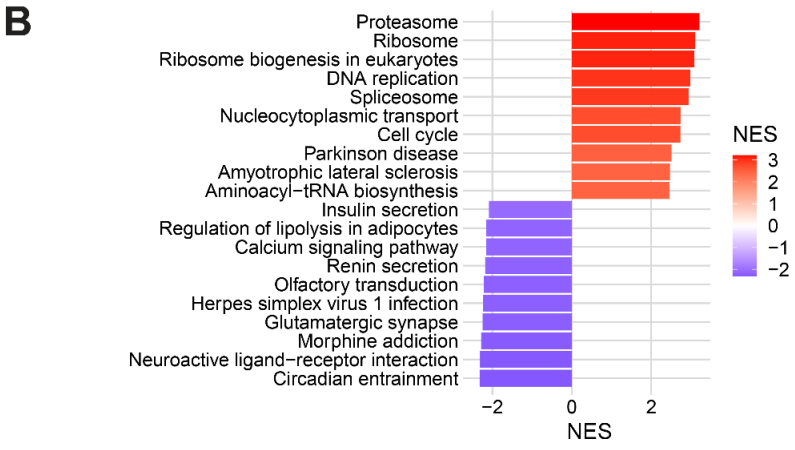


Fig. S8.

Method for determining specific transcription factor regulators of hypertranscription. A) To find genes regulating RNA output, first all genes are scored by Pearson correlation and fisher test to assess each genes' relationship with RNA output. The resulting distribution of odds ratio

(OR) is used to test enrichment of transcription factors (TFs) and their targets using gene-set enrichment analysis (GSEA). Putative drivers are filtered for those where both the transcription factor targets and the transcription factor itself are either highly enriched or depleted in the hypertranscriptional state. **B)** Barplot depicting the normalized enrichment score (NES) for the top-10 and bottom-10 most enriched KEGG pathways in the pan-cancer Fisher's log odds distribution for hypertranscriptional genes (in **A**).

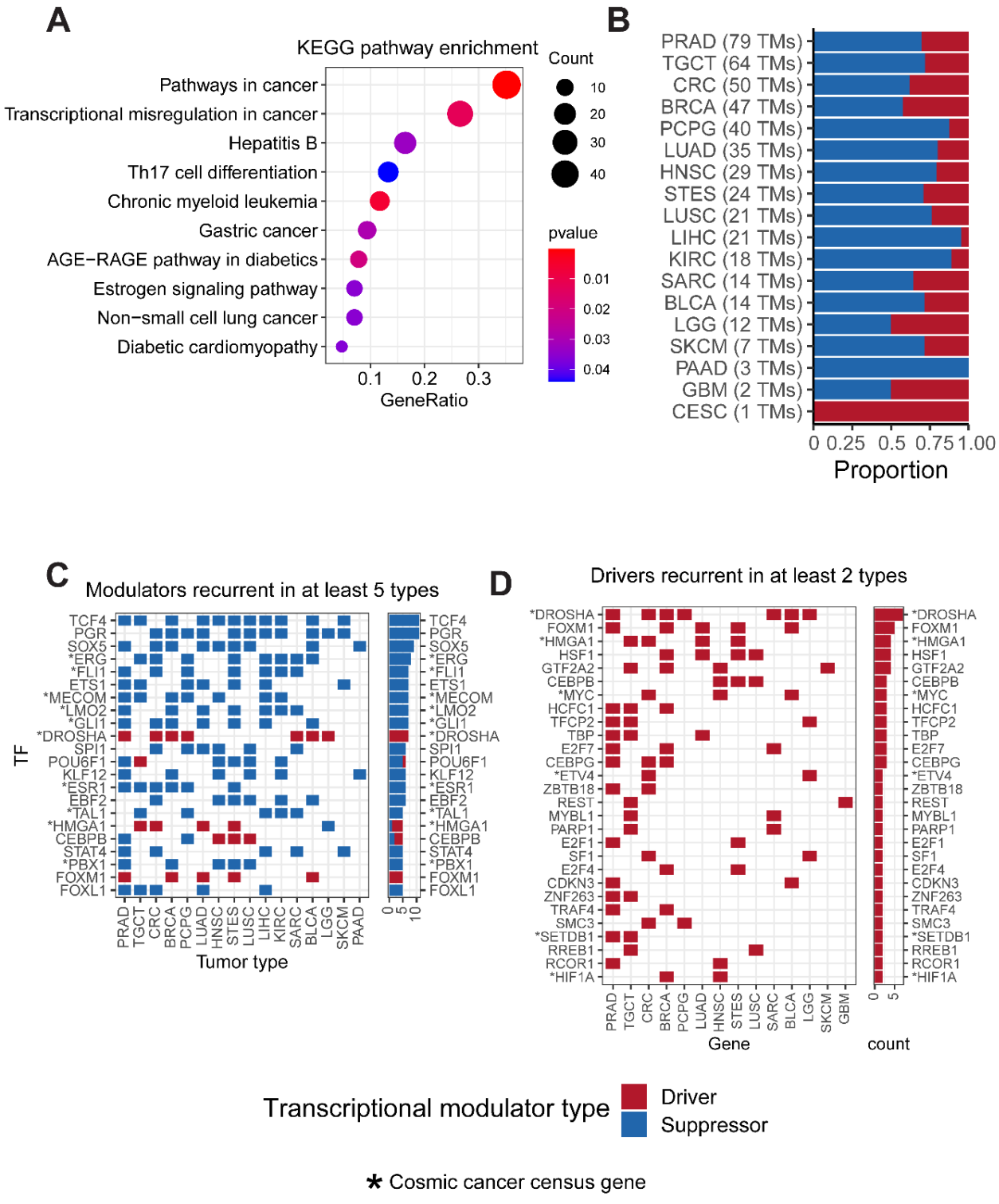


Fig. S9.

Regulators of RNA output in cancer. A) KEGG pathway enrichment for transcription factors found to regulate RNA output. **B)** Proportion of transcriptional modulators in each tumor type colored by their correlation to RNA output. **C)** Plot depicting genes found to regulate RNA

output in at least 5 tumor types, colored by TM correlation with RNA output. **D)** Positive TMs found in at least 2 tumor types. Cosmic cancer census genes are indicated by gene names with an asterisk (*).

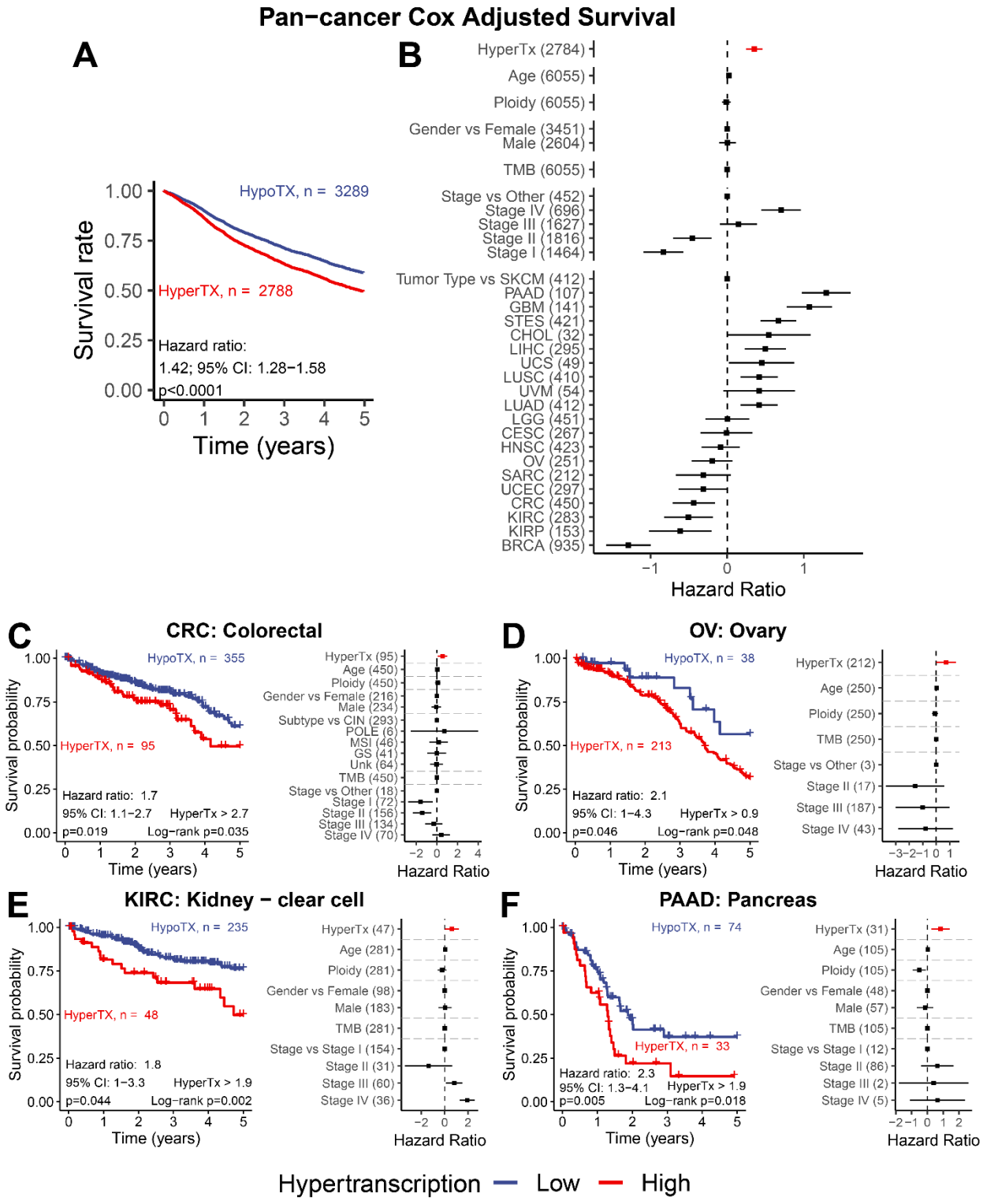

Pan−cancer Cox Adjusted Survival
A
HypoTX, n = 3289
HyperTX, n = 2788
Hazard ratio:
1.42; 95% CI: 1.28−1.58
p<0.0001
Survival rate
Time (years)
B
HyperTx (2784)
Age (6055)
Ploidy (6055)
Gender vs Female (3451)
Male (2604)
TMB (6055)
Stage vs Other (452)
Stage IV (696)
Stage III (1627)
Stage II (1816)
Stage I (1464)
Tumor Type vs SKCM (412)
PAAD (107)
GBM (141)
STES (421)
CHOL (32)
LIHC (295)
UCS (49)
LUSC (410)
UVM (54)
LUAD (412)
LGG (451)
CESC (267)
HNSC (423)
OV (251)
SARC (212)
UCEC (297)
CRC (450)
KIRC (283)
KIRP (153)
BRCA (935)
Hazard Ratio
C
CRC: Colorectal
HypoTX, n = 355
HyperTX, n = 95
Hazard ratio: 1.7
95% CI: 1.1−2.7
HyperTx > 2.7
p=0.019
Log−rank p=0.035
Survival probability
Time (years)
HyperTx (95)
Age (450)
Ploidy (450)
Gender vs Female (216)
Male (234)
Subtype vs CIN (293)
POLE (6)
MSI (46)
GS (41)
Unk (64)
TMB (450)
Stage vs Other (18)
Stage I (72)
Stage II (156)
Stage III (134)
Stage IV (70)
Hazard Ratio
D
OV: Ovary
HypoTX, n = 38
HyperTX, n = 213
Hazard ratio: 2.1
95% CI: 1−4.3
HyperTx > 0.9
p=0.046
Log−rank p=0.048
Survival probability
Time (years)
HyperTx (212)
Age (250)
Ploidy (250)
TMB (250)
Stage vs Other (3)
Stage II (17)
Stage III (187)
Stage IV (43)
Hazard Ratio
E
KIRC: Kidney − clear cell
HypoTX, n = 235
HyperTX, n = 48
Hazard ratio: 1.8
95% CI: 1−3.3
HyperTx > 1.9
p=0.044
Log−rank p=0.002
Survival probability
Time (years)
HyperTx (47)
Age (281)
Ploidy (281)
Gender vs Female (98)
Male (183)
TMB (281)
Stage vs Stage I (154)
Stage II (31)
Stage III (60)
Stage IV (36)
Hazard Ratio
F
PAAD: Pancreas
HypoTX, n = 74
HyperTX, n = 33
Hazard ratio: 2.3
95% CI: 1.3−4.1
HyperTx > 1.9
p=0.005
Log−rank p=0.018
Survival probability
Time (years)
HyperTx (31)
Age (105)
Ploidy (105)
Gender vs Female (48)
Male (57)
TMB (105)
Stage vs Stage I (12)
Stage II (86)
Stage III (2)
Stage IV (5)
Hazard Ratio
Hypertranscription
Low
High

Fig. S10.

Prognostically significant hypertranscriptional subgroups in cancer. A) Cox adjusted survival curves for hyper- and hypotranscriptional groups in the pan-cancer cohort. **B)** Forest plot showing hazard ratios for the pan-cancer cox regression model. **C-F)** Kaplan-Meier survival plots (left) and Cox regression model hazard ratios (right) for hyper- and hypotranscriptional subgroups in **C**) colorectal carcinoma, **D**) ovarian cancer, **E**) kidney renal clear cell carcinoma and **F**) pancreatic adenocarcinoma. *Error bars on all hazard ratio coefficients represent the 95% CI.*

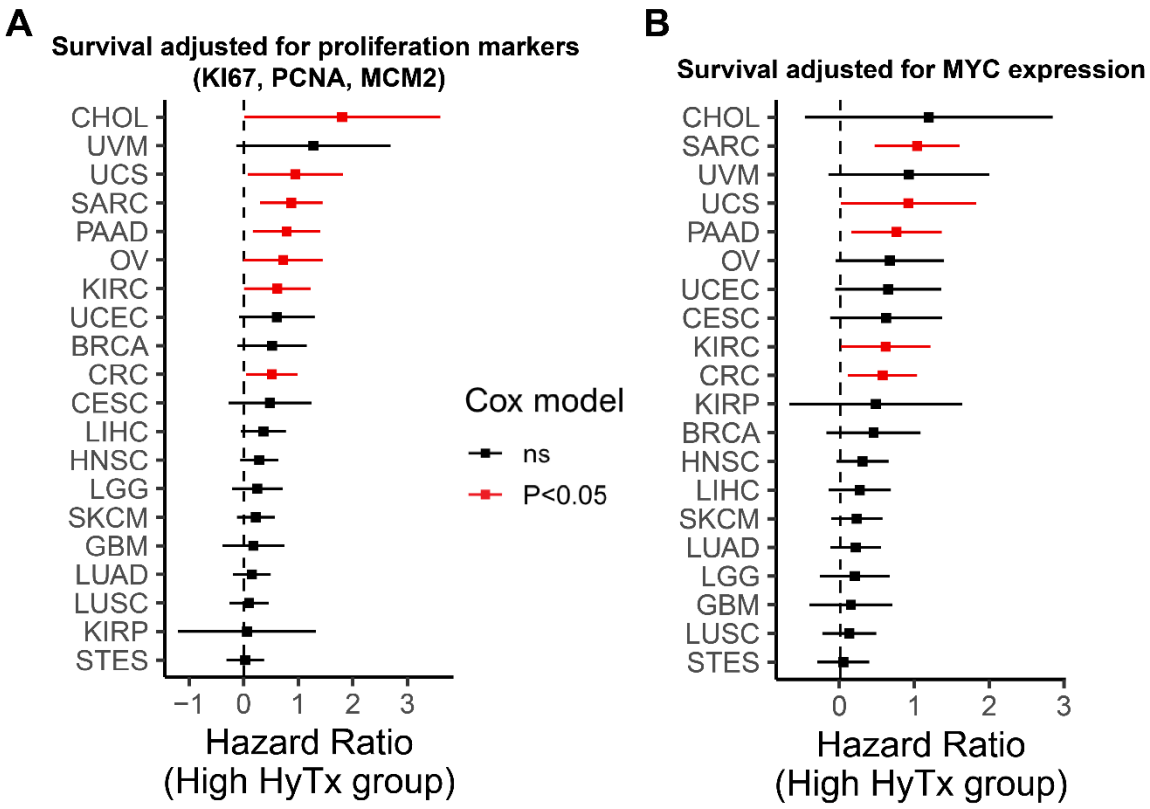


Fig. S11.

Analysis of RNA output and survival in cancer types accounting for proliferation associated markers. A) Hazard ratios for hypertranscriptional subgroups across cancer types while accounting for expression of proliferation associated markers KI67, PCNA, and MCM2 or **B)** MYC. E*rror bars on all hazard ratio coefficients represent the 95% CI.*

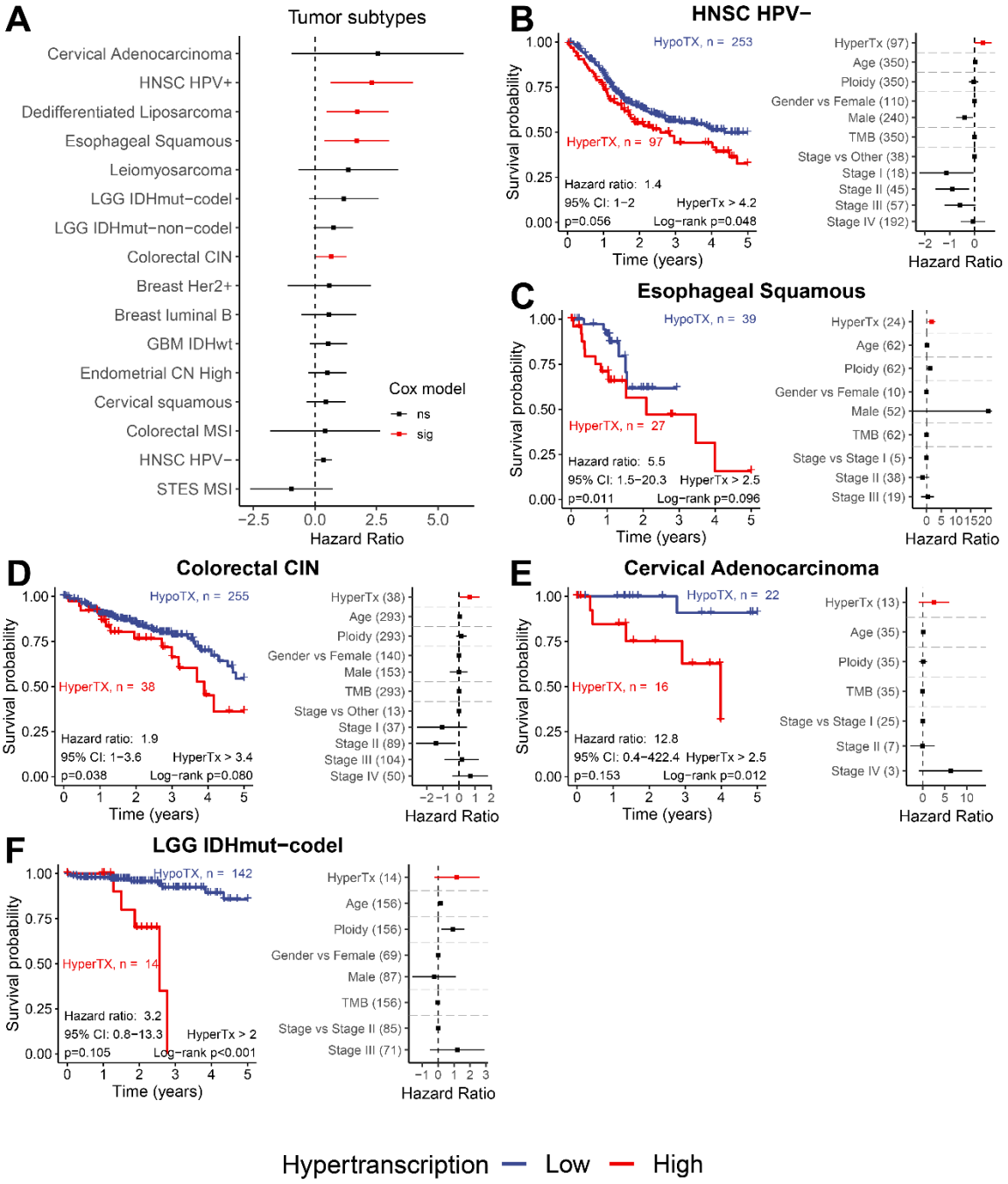


Fig. S12.

Analysis of RNA output and survival in additional cancer subtypes. A) Hazard ratios for hypertranscriptional subgroups across cancer subtypes. **B-F)** Kaplan-Meier survival plots (left) and Cox regression model hazard ratios (right) for hyper- and hypotranscriptional subgroups in

B) HPV- head and neck squamous cell carcinoma, **C**) esophageal squamous cell carcinoma, **D**) chromosomal instable colorectal carcinoma, **E**) cervical adenocarcinoma, **F**) IDHmutant-codel low grade glioma. *Error bars on all hazard ratio coefficients represent the 95% CI.*

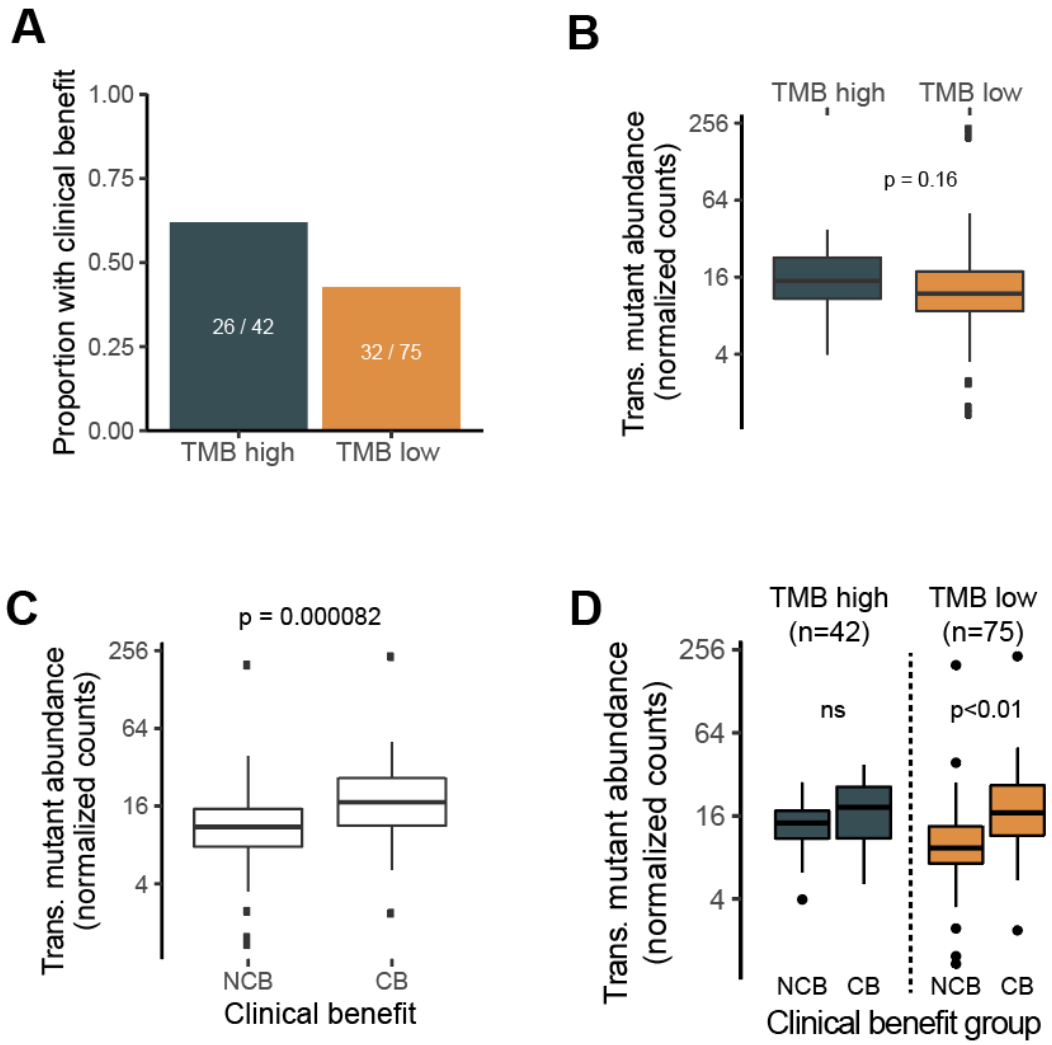


Fig. S13.

Transcriptional mutation abundance and ICI response. A) Proportion of patients with clinical benefit from ICI in either high or low TMB groups. TMB high is defined as greater than 10 mutations per megabase. **B)** Average transcriptional mutation abundance of TMB high and TMB low ICI patients (student's *t*-test, p=0.16). **C)** Average transcriptional mutation abundance of ICI patients with and without clinical benefit to ICI. Patients with clinical benefit have significantly increase average mutation abundance (student's *t*-test, p = 0.000082). **D)** Average transcriptional mutation abundance of ICI patients with and without clinical benefit to ICI split by TMB high and low groups. TMB low patients have significantly increase expression of their mutations. CB = clinical benefit. NCB = no clinical benefit. Boxplots are defined in Supplemental Figure 1B.

Data S1. (separate file)
TCGA sample numbers analyzed by RNAmp.

Data S2. (separate file)
Cohort hypertranscription measures.

Data S3. (separate file)
Proportion of variability explained in hypertranscription by various features

Data S4. (separate file)
Ranking of hallmark pathways associated with hypertranscription

Data S5. (separate file)
Putative regulators of RNA output across cancers.

Data S6. (separate file)
TCGA and GTEx analysis sample and gene numbers