

## Supplementary Information (Figures, Tables, and Files)

### Survey of the translation shifts in hepatocellular carcinoma with ribosome profiling

Qin Zou<sup>1-4</sup>, Zhengtao Xiao<sup>1-3,#</sup>, Rongyao Huang<sup>1-3,#</sup>, Xin Wang<sup>1-3</sup>, Xun Wang<sup>5</sup>, Haitao Zhao<sup>6</sup>, and Xuerui Yang<sup>1-3</sup>

<sup>1</sup> MOE Key Laboratory of Bioinformatics, Tsinghua University, Beijing 100084, China

<sup>2</sup> Center for Synthetic & Systems Biology, Tsinghua University, Beijing 100084, China

<sup>3</sup> School of Life Sciences, Tsinghua University, Beijing 100084, China

<sup>4</sup> Joint Graduate Program of Peking-Tsinghua-National Institute of Biological Science, Tsinghua University, Beijing 100084, China.

<sup>5</sup> Department of Hepatobiliary Surgery, The General Hospital of Chinese People's Liberation Army, Beijing 100853, China.

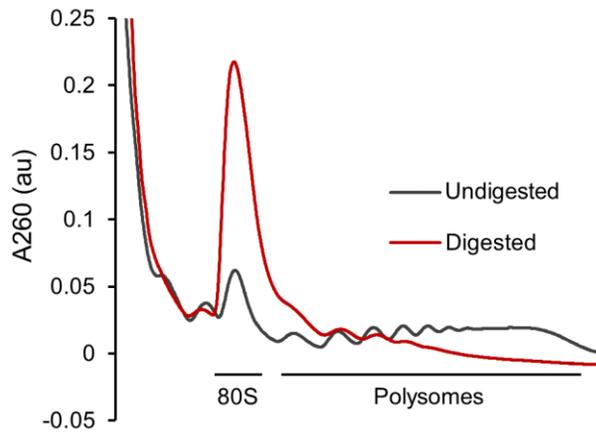
<sup>6</sup> Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100730, China.

# These authors contributed equally to this work.

**Corresponding author:** Xuerui Yang, Medical Science Building D231, School of Life Sciences, Tsinghua University, Beijing 100084, China. Tel: 86-10-62783943. Email: [yangxuerui@tsinghua.edu.cn](mailto:yangxuerui@tsinghua.edu.cn)

## Supplementary Figures

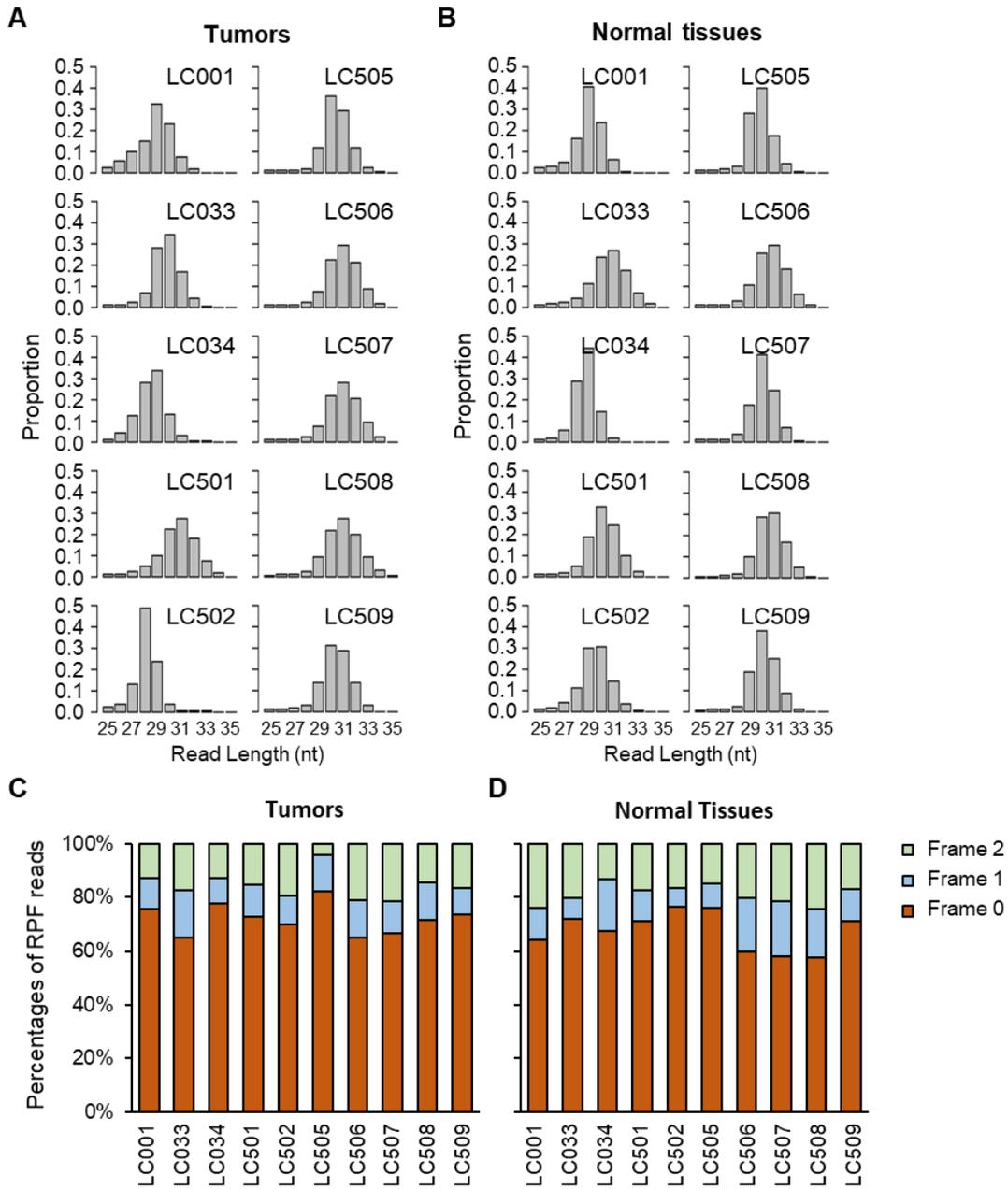
### Supplementary Figure 1



### Supplementary Figure 1. Polysome fractions upon RNase digestion for ribosome profiling.

Fractionation of polysomes upon digestion of RNA samples with RNase I, which was used for extraction of RPFs during ribosome profiling.

**Supplementary Figure 2**

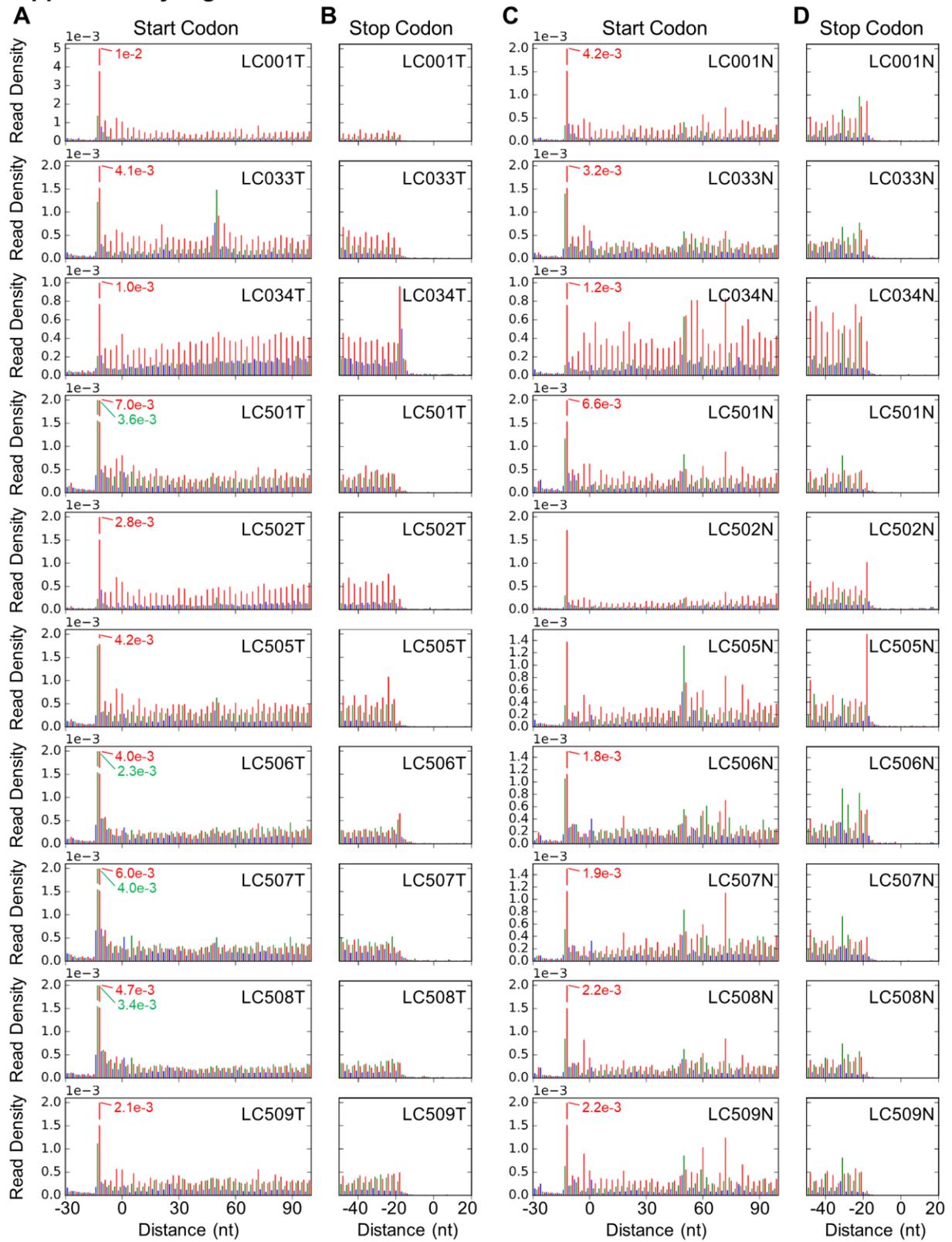


**Supplementary Figure 2. Length distributions and frame allocations of the RPF reads.**

**(A, B)** Length distributions of the RPF reads in the ribosome profiling data of the 10 tumors (A) and their adjacent normal tissue samples (B). **(C, D)** Percentages of the RPF reads allocated, according to their P-site positions, to the 3 open reading frames of the previously annotated CDS

regions. The algorithm RiboCode was used to determine the P-site positions of the RPF reads with different lengths and in different samples.

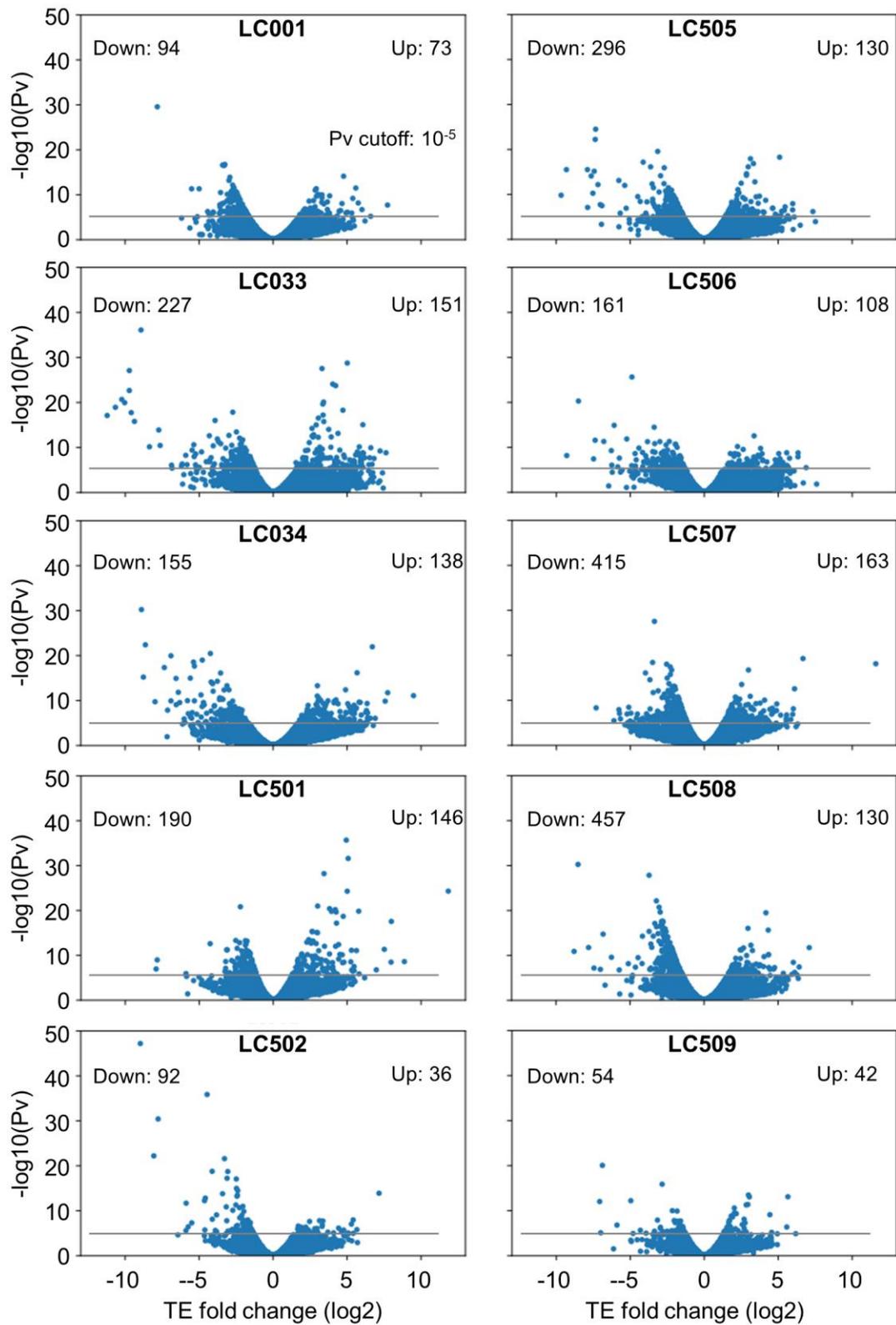
### Supplementary Figure 3



**Supplementary Figure 3. Metagene analysis of the RPF reads mapped to the protein coding genes.**

**(A-D)** The RPF reads from the 10 tumors (A, B) and their adjacent normal tissue samples (C, D) were mapped to the previously known CDS regions of the protein coding genes. Each bar plot shows the normalized density of the RPF reads whose 5' ends were allocated to each position around the start (A, C) or stop (B, D) codons of the protein coding genes.

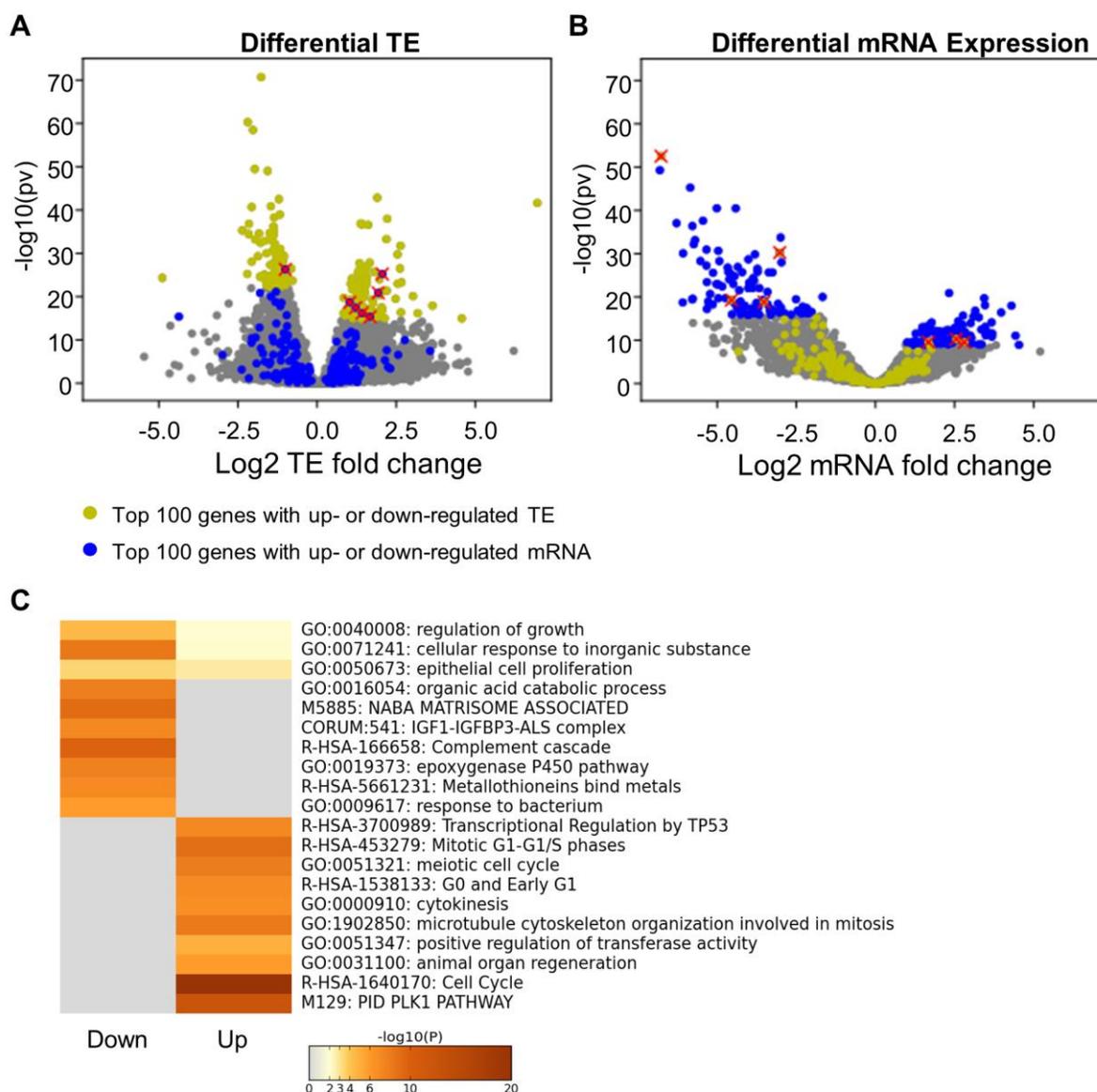
### Supplementary Figure 4



**Supplementary Figure 4. Volcano plots showing gene differential translations in 10 paired comparisons between HCC tumors and their adjacent normal samples.**

Gene differential translation analysis was performed for each pair of tumor and adjacent normal samples. The results from the 10 HCC patients are summarized as 10 volcano plots. Log<sub>2</sub> of the TE fold change is shown on the horizontal axis, and  $-\log_{10}$  of the P-value is shown on the vertical axis. The numbers of genes showed significantly up- or down-regulated translation efficiencies in tumor are provided on each plot.

## Supplementary Figure 5

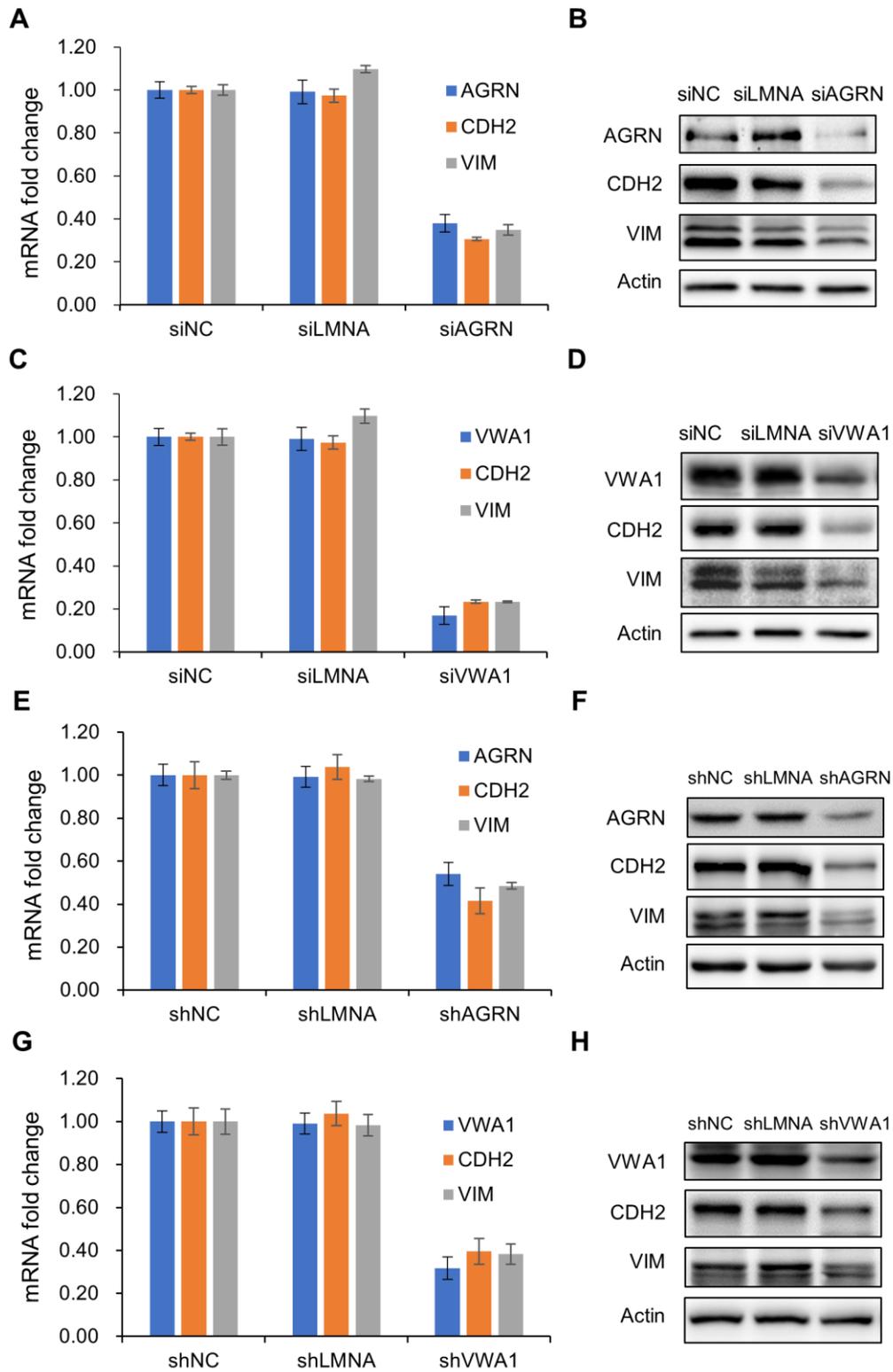


### Supplementary Figure 5. Results of consensus differential translation and differential mRNA expression analyses.

**(A)** A volcano plot showing the consensus differential translation of the protein coding genes across the 10 HCC patients. As described in the section of methods, this was based on the results of 10 differential translation analyses for the 10 pairs of HCC tumors and normal samples. **(B)** A volcano plot showing the result of mRNA differential expression analysis for the 10 pairs of tumor

and normal tissue samples. For each gene,  $\log_2$  of the TE fold change (A) or mRNA expression fold change (B) is shown on the horizontal axis, and  $-\log_{10}$  of the P-value is shown on the vertical axis. In both plots, the top 100 genes with the most significant translational up- or down-regulations are marked in green (200 in total), and the top 100 genes with significantly up- or down-regulated mRNA expressions are in blue (200 in total). The overlapping genes with both differential TEs and differential mRNA expressions are highlighted in red. **(C)** The biological processes enriched in the genes with up- or down-regulated mRNA expressions in the 10 HCC tumors compared to their adjacent normal tissues. For each term, saturation of the color indicates the statistical significance ( $-\log_{10}(P_v)$ ) of the enrichment.

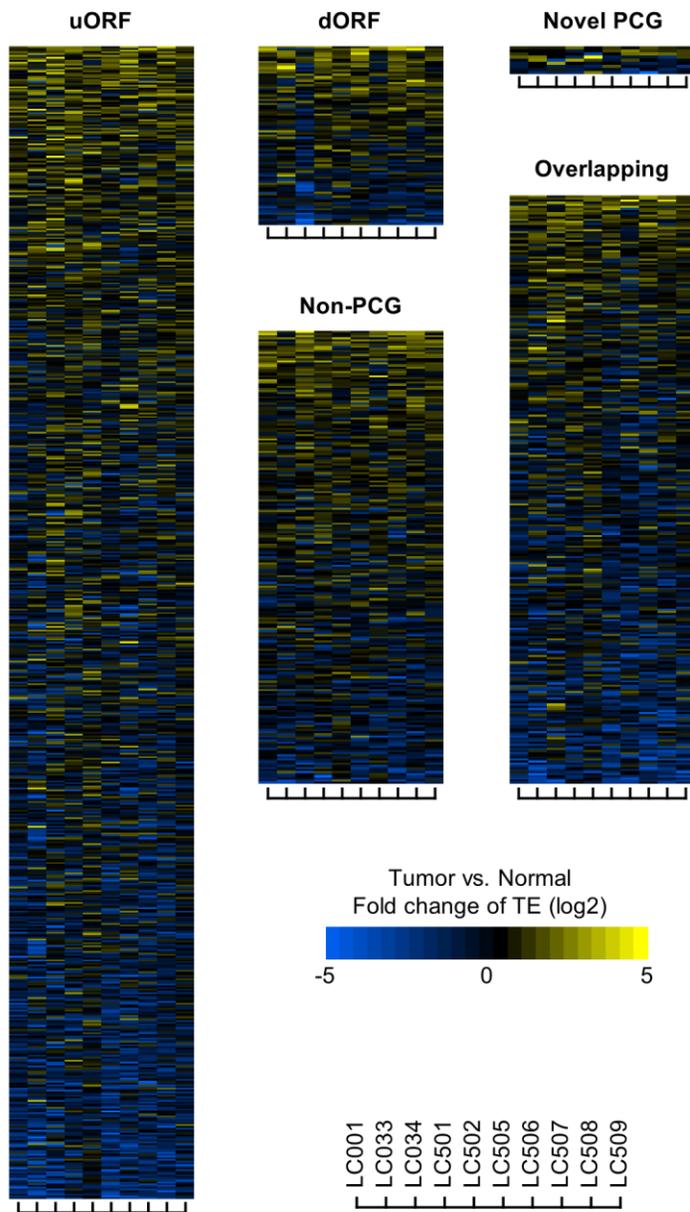
## Supplementary Figure 6



**Supplementary Figure 6. mRNA and protein expressions upon silencing of AGRN and VWA1.**

**(A-D)** mRNA (A, C) and protein (B, D) expression fold changes upon siRNA-mediated knock-down of AGRN (A, B) or VWA1 (C, D) in Huh7 cells. **(E-H)** mRNA (E, G) and protein (F, H) expression fold changes upon lentivirus-mediated shRNA knock-down of AGRN (E, F) or VWA1 (G, H) in Huh7 cells.

## Supplementary Figure 7



**Supplementary Figure 7. TE fold changes of the ORFs in the 10 HCC tumors compared to the adjacent normal tissues.**

**(A-E)** For the 5 categories of non-canonical ORFs, which were annotated in the translatomies of the HCC tumors and normal tissues, the fold changes of their TE in each tumor compared to the adjacent normal tissue are summarized as heat maps. Within each category, the ORFs were

sorted from top to down by their average TE fold changes in the 10 pairs of tumor and adjacent normal samples.

## Supplementary Tables

**Supplementary Table 1. Summary of the ribosome profiling data from HCC patients.**

Sample	Tissue type	Mapped reads		No. of Genes	
		RPF	RNA	RPF	RNA
LC001	Tumor	4257175	4853555	12957	15136
	Adjacent normal	5613864	4437075	13168	16105
LC033	Tumor	8929475	5602870	13492	14781
	Adjacent normal	4772817	3744776	11971	15791
LC034	Tumor	3403678	3249872	14420	15311
	Adjacent normal	12436189	5042714	13659	15920
LC501	Tumor	4985689	4723425	13209	14610
	Adjacent normal	5136939	4097488	13871	15528
LC502	Tumor	10209583	7748878	14068	15097
	Adjacent normal	9275703	4861635	13783	15871
LC505	Tumor	4823874	4948992	13559	15078
	Adjacent normal	5715826	3571356	12317	14753
LC506	Tumor	4299663	4809507	13664	16033
	Adjacent normal	4020289	4550724	11354	15658
LC507	Tumor	3982890	5154486	12700	15600
	Adjacent normal	4753715	4775113	13887	16157
LC508	Tumor	6912199	6130614	13663	15514
	Adjacent normal	4876965	4275438	12890	15722
LC509	Tumor	5309565	5700362	13423	18344
	Adjacent normal	5148473	3296396	12924	15096

## **Supplementary Files**

**Supplementary File 1. Read counts of RPF and RNA in 10 pairs of tumor and adjacent normal tissue samples from HCC patients.**

**Supplementary File 2. Metagene analysis of the RPF reads with different lengths aligned on the protein-coding genes.**

**Supplementary File 3. Differential TE of the genes in each tumor compared to the adjacent normal tissue.**

**Supplementary File 4. Consensus differential TE of the genes based on differential TE analyses in 10 pairs of tumor and adjacent normal samples.**

**Supplementary File 5. Different categories of ORFs in the translomes of tumors and normal tissues.**

**Supplementary File 6. TE changes of the different types of ORFs in the tumors compared to normal tissues.**

**Supplementary File 7. Associations between TE abnormalities of uORFs or dORFs with their corresponding main CDS regions of the hosting genes.**