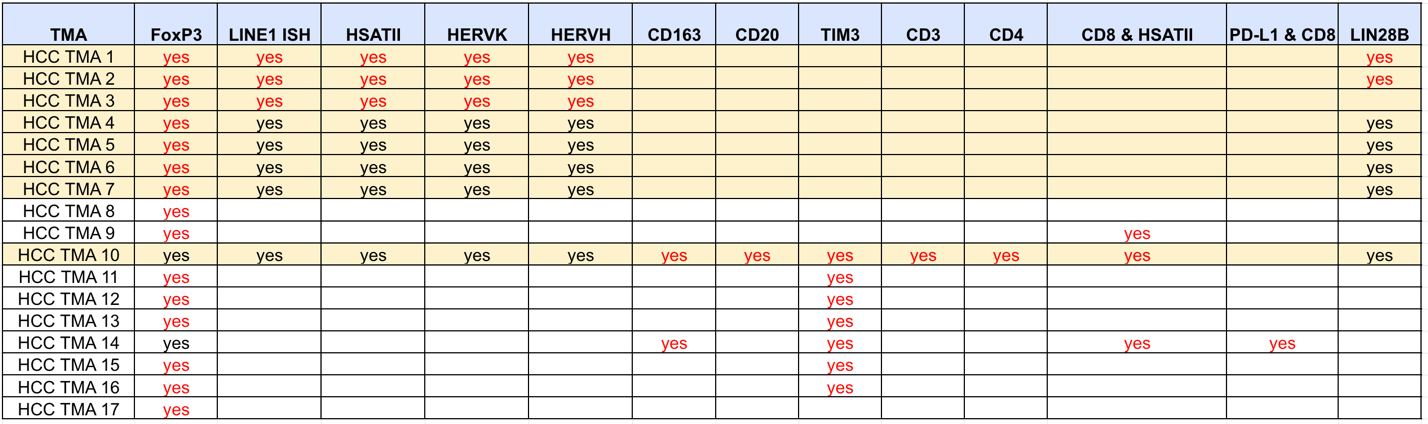
**HCC TMA Project Outline**

Step 1: ISH analysis

* Use all ISH data to make survival plots for LINE1 high and low based on ISH
* Correlate LINE1 and FOXP3, if possible
* Red = not quantified, black = quantified, yes = stained and imaged



* Specifically, use the data that Linda sent for TMAs 4-7 and TMA 10
  + Clean data
  + Plot (LINE1 area) / (tissue area) 🡪 bar plot
  + For patients with multiple cores, sum the LINE1 area and sum the tissue area
  + Decide on high vs. low groups
  + Kaplan-Meyer plot for all patients

Step 2: CosMx analysis on patient level

* Check concordance between ISH groups and tx-omics groups
* DE analysis: try to show that high LINE1 tracks with some EMT/neuro-endocrine transition genes, other repeats, and ICB genes that tumor would express
* Abundance analysis:
  + pool all counts of cells for each patient
  + high patients vs low patients for immune cell type numbers only (may need to remove biased FOVs), offset by total number of cells from the patient
  + number of TLSs (should be quick to count) across two groups
* Co-localization analysis:
  + Do this for all cell types, but only show/test the immune cell types (we want to reduce unnecessary hypothesis testing)
  + Try suggestions Martin made

Step 3: CosMx analysis on the cell level

* Random effect in all models will still be the patient ID
* We want to see if directly characterizing cells as LINE1 positive/negative will lead to more detailed information about the TME
* Do the DE analysis and co-localization, as above

Overall Idea

* LINE1 can be a biomarker for poor survival
* LINE1 modulates tumor state
* LINE1 can contribute to immune evasion by tumor, which can be seen both on the patient level and on the cell-cell level

Joseph’s outline:

***Dysregulated HCC viral-like repeat RNA expression correlates with poor outcomes and an immune-suppressive microenvironment***

Nawrocki, Coley, …., Nieman, Aryee, Franses, Ting.

Background: Dysregulated repeat RNA expression has been shown to contribute significantly to the pathogenesis of many solid tumors, but their roles in HCC haven’t yet been carefully explored. We leveraged conventional and novel tissue profiling technologies to examine how repeat RNAs are expressed in HCC.

Methods: RNA-ISH, clinical dataset, CosMx 1000-plex dataset

Table 1: Clinical characteristics of the patient data corresponding to the entire cohort (and a subset illustrating which subset of patients were used for the CosMx *(similar to Table 1 in the most recent manuscript version)*

Fig 1: HCC tumor tissue expresses diverse repeat RNA species, with the highest expressing species - LINE-1 – correlating with poor outcome and with coordinate expression of other repRNA species.

1. Representative RNA-ISH stains of LINE-1, HSATII, HERV-K, and HERV-H *(similar to current Fig 1B-E)*
2. Quantification of each of the above 4 repeat RNA species across all samples (*similar to current Fig. 1F*) using the same vertical axis and units (? To show that there are more LINE-1 dots per square micron of tissue compared with the other species)
3. Correlation of the expression of LINE-1 RNA-ISH expression with: HSATII RNA, HERV-K RNA, and HERV-H RNA; probably easiest to just show 3 dot plots and regression lines

*\*\* New analysis of existing data, hopefully easy. \*\**

1. Kaplan-Meier curves of LINE-1 high vs low expression with overall survival (using the data that Linda just generated; *similar to current Fig 2D*)

Supplemental Fig. 1: Data pre-processing and QC metrics for HCC-1000 plex TMA CosMx dataset (# of cores, # cells, # transcripts/cell, etc.; *can incorporate similar data/schematics as in slides 2-3*)

Fig 2: Single-cell spatial analysis can provide high-plex spatial analysis of transcriptomic-level expression states.

1. Schematic (Biorender) of 1000-plex data analysis workflow
2. Illustration of FOV placement on TMA low-res maps
3. UMAP with cell type annotations
4. Bubble heatmap of cell types identified in UMAP using marker genes
5. Sample tissue images (2-3 of the “best”; *similar to slide 4*)

Supplemental Fig. 2: Proportions of each cell type across each FOV (or core?) in the dataset.

Fig 3: Spatial analyses confirm coordinate expression of LINE-1 with multiple repeat RNAs and with de-differentiated cancer cell phenotype.

1. Volcano plot of differentially expressed genes between LINE1-high vs LINE1-low (whichever threshold is best; *similar to either slide 5 or slide 8*)
2. Representative images from b of LINE-1 expression with HERV-K and HSATII from CosMx dataset
3. GSEA showing less hepatocyte-like gene set expression overlap (*similar to left plot in slide 9*)
4. Representative images from c of LINE-1 expression and specific hepatocyte marker genes in representative LINE1 high and LINE1-low cells

Fig 4: Spatial analyses demonstrate an immune-suppressive phenotype correlating with LINE-1 expression within HCC tumor tissue.

1. Colocalization / niche analyses of CD8+ exhausted cells (or Tregs or some other subset) with LINE-1 (what Cole was working on most recently; *like slide 20*)

\*\* I don’t know if that reference that Leo sent to you helps or not at this stage, but you could give it a try

1. Representative images from a
2. Some of Avril’s IHC data (like Fig 3-4 from most recent manuscript) showing correlation of LINE1 with Treg marker (FoxP3)?

\*\* Or multiplex IF analyses…not sure if DTT/LN wanted to do this for this specific manuscript/project or not but I seem to recall this being a thought in a recent e-mail

1. Is there a way to identify the highest risk group even within LINE1-high sets? E.g. does the highest Treg density within the LINE1-high group correlate with the worst outcomes (or is there another similar subgroup)? ◊ if so, then this would be a nice Kaplan-Meier curve to bookend the figure

Supplemental Fig. 3: IHC/IF image data, representative images, quantification