A130 · A necrosis-associated repeat-element program in metastatic colorectal cancer, and an open problem in separating DNA from RNA

Presenter: Chenyue Lu — Harvard-MIT Health Sciences and Technology · Dana-Farber Cancer Institute

Authors: Chenyue Lu, Cole Nawrocki, Amaya Pankaj, Leon Pappas, Bidish Patel, David Ting, Martin Aryee

Session: Day 2 · Fri Oct 2, 2026 · 2:15–4:15 PM

Pericentromeric satellite HSATII, a repetitive element (RE), is silenced in normal tissue but aberrantly transcribed in epithelial cancers, where it is cancer-specific and co-derepressed with another RE, LINE-1 (Ting 2011). Its RNA is reverse-transcribed into pericentromeric DNA, expanding copy number in colon tumors and tracking worse survival (Bersani 2015). Its location and relation to the microenvironment are unknown. Joint segmentation-free factorization (FICTURE/PUNKST; Si 2024) of seven colorectal cancer Xenium slides (58 patients) places all samples on one comparable basis. Two of 100 factors are dominated by HSATII and LINE-1, as predicted. Instead of viable tumor epithelium, they fill acellular necrotic lumina enclosed by tumor glands, expanding with disease progression from 0.2% of transcripts in primaries to 4.3% and 3.1% in liver and lung metastases. These regions are also enriched for DUX4, the embryonic transcription factor causing facioscapulohumeral muscular dystrophy (FSHD). DUX4 induces HSATII transcription (Shadle 2019), and FSHD patients have higher gastrointestinal cancer incidence (Kurashige 2023). What was detected is ambiguous: a Xenium probe binds a single-stranded target that may be mRNA or genomic DNA. Quality control cannot separate them: genomic-DNA controls decode as confidently as transcripts (median Q-value 40). We tried two decompositions. Subtracting a copy-number-predicted DNA term works for ordinary genes but breaks down at high copy number, the repeats of interest. Regressing on genomic controls fails differently: controls sit at stable-copy-number loci and cannot track a target whose copy number varies. Both assume copy number is fixed; for HSATII it is an outcome of transcription.

A necrosis-associated repeat-element program in metastatic colorectal cancer, and an open problem in separating DNA from RNA

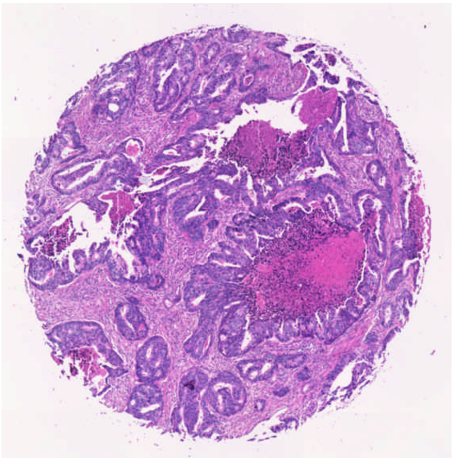
Chenyue Lu[1,4,*], Cole Nawrocki[2], Amaya Pankaj[2], Leon Pappas[2], Bidish Patel[2], David Ting[2], Martin Aryee[3,4]

[1]*Harvard-MIT Health Sciences and Technology, Cambridge, MA, USA* [2]*Massachusetts General Hospital, Boston, MA, USA*
[3]*Broad Institute of MIT and Harvard, Cambridge, MA, USA* [4]*Dana-Farber Cancer Institute, Department of Data Science, Boston, MA, USA*
[*]Presenting author

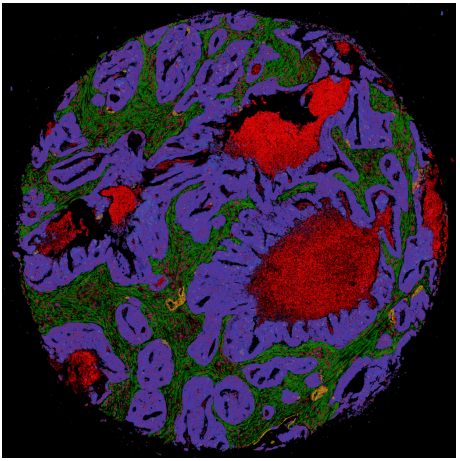
Pericentromeric satellite HSATII, a repetitive element (RE), is silenced in normal tissue but aberrantly transcribed in epithelial cancers, where it is cancer-specific and co-derepressed with another RE, LINE-1 (Ting 2011). Its RNA is reverse-transcribed into pericentromeric DNA, expanding copy number in colon tumors and tracking worse survival (Bersani 2015). Its location and relation to the microenvironment are unknown.

Joint segmentation-free factorization (FICTURE/PUNKST; Si 2024) of seven colorectal cancer Xenium slides (58 patients) places all samples on one comparable basis. Two of 100 factors are dominated by HSATII and LINE-1, as predicted. Instead of viable tumor epithelium, they fill acellular necrotic lumina enclosed by tumor glands, expanding with disease progression from 0.2% of transcripts in primaries to 4.3% and 3.1% in liver and lung metastases. These regions are also enriched for DUX4, the embryonic transcription factor causing facioscapulohumeral muscular dystrophy (FSHD). DUX4 induces HSATII transcription (Shadle 2019), and FSHD patients have higher gastrointestinal cancer incidence (Kurashige 2023).

What was detected is ambiguous: a Xenium probe binds a single-stranded target that may be mRNA or genomic DNA. Quality control cannot separate them: genomic-DNA controls decode as confidently as transcripts (median Q-value 40). We tried two decompositions. Subtracting a copy-number-predicted DNA term works for ordinary genes but breaks down at high copy number, the repeats of interest. Regressing on genomic controls fails differently: controls sit at stable-copy-number loci and cannot track a target whose copy number varies. Both assume copy number is fixed; for HSATII it is an outcome of transcription.



A H&E



B joint factor model

Figure 1: One lung-metastasis sample, two views. (**A**) Cribriform tumor glands enclose lumina filled with eosinophilic necrotic debris. (**B**) The same core under the joint 100-factor model, colored by program: tumor epithelium (indigo), stroma (green), immune (magenta), host lung (gold), and the two repeat-dominated factors (**red**). The repeat program fills the necrotic lumina.