Recurrent Repeat Contractions (rRCs) & Micro-Changing Short Tandem Repeats (mcSTRs): Investigating Novel Polymorphic Variants in 15 Human Cancers

1. REPLICATION ERRORS

- **30.2 million annual** cancer cases expected by **2040**[1] → **stage of cancer at diagnosis** is highly effective in predicting survival (91% 5-year early-stage vs 26% late-stage)[2]
- **Replication errors: two-thirds** of all mutations in human cancers[3], accelerates tumorigenesis[4-9]. **Tandem repeats (TRs):** new and valuable mutation in cancer[10-15]
  - Recurrent repeat expansions (rREs) identified in 7 cancers (Fig. 1)[16-18]

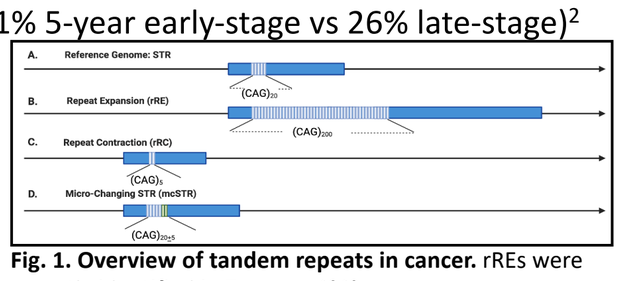

Fig. 1. Overview of tandem repeats in cancer. rREs were recently identified in 7 cancers[16-18]

Micro-changing STRs & Recurrent Repeat Contractions are understudied in cancer

2. DETECTING DISEASE STATE TRs

- Traditional structural variant (SV) callers **disregard rRCs and mcSTRs**[19-20]
- **rRCs: ExpansionHunter Denovo**[21]**/Local Read Depth Normalization (LRDN):** Repeat SV caller + limit range of normalization (±500 bp) → **eliminates false positives**
- **mcSTRs: ExpansionHunter**[22]**/ActiveSTRs Catalog:** Reveals 174,323 TRs (many non-reference TRs; Fig. 2) to study

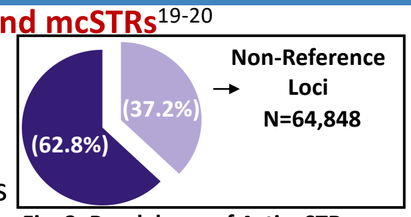

Fig. 2. Breakdown of ActiveSTRs.

3. DNA MUTATIONS IN CIRCULATING CELL-FREE DNA

- Cell-free DNA biomarkers are <150 base pairs[23] → **non-invasive early cancer detection**
- **Cancer sub-type specific mutations** allow for cancer-location specific diagnosis[24-28]

How do rRCs & mcSTRs in cancer compare to pathogenic tandem repeat mutations and is it a novel biomarker for early cancer detection?

METHODOLOGY

1. Identify rRCs/mcSTRs: Pan-Cancer Analysis of Whole Genomes[29-30] (2509 Tumor-Normal Samples; 37 cancer types). rRCs: EH Denovo (EHdn v0.9)[21] + LRDN · **mcSTRs:** EH (EH v5.0)[22]+ ActiveSTRs

2. Compare cancer rRCs'/mcSTRs' genomic role to 1) other known repeats in cancer and 2) repeats other diseases: Microsatellite Instable, Expansions, cis-regulatory elements, JENSEN diseases

3. Identify case-study rRCs/mcSTRs: Experimental validation in isolated cohorts of matching tissue/normal/plasma samples. Evaluate rRCs'/mcSTRs' effectiveness as early diagnostic biomarker

RESULTS/DISCUSSIONS

rRCs: EH Denovo + LOCAL READ DEPTH NORMALIZATION (LRDN)

- **rRCs are significant mutation sources of genetic variation in cancers** (q<0.05)

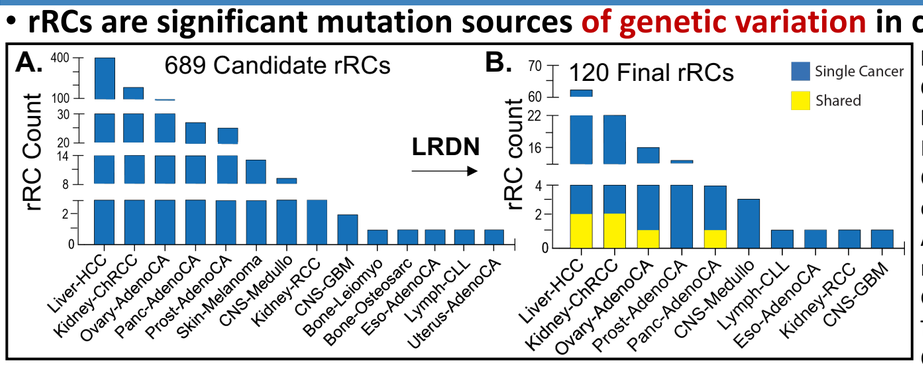

Fig. 3. Comparison of A) Candidate rRCs Before LRDN and B) Final rRCs after LRDN. HCC = Hepatocellular carcinoma. ChRCC = Chromophobe renal cell carcinoma. AdenoCA = Adenocarcinoma. Medullo = medulloblastoma. GBM = Glioblastoma. Wilcoxon Rank Sum Test (p<0.05), False Discovery Rate Correction (q<0.05)

- LRDN eliminated >80% of candidate rRCs as false positives (689 → 120; Fig 3A)
- **EHdn + TROPIC**[31] **+ LRDN** → 120 rRCs across 10 cancer types (Fig. 3B)

rRCs: ELIMINATING NON-rRC REPEAT SIGNALS VIA LRDN

- Repeat count differences in tumor vs normal are rendered **insignificant by LRDN**
- Most false positives in Prost AdenoCA (Fig. 4A) and Kidney ChRCC (Fig. 4B)
- **Local read depth normalization is crucial in finding false positive rRCs**

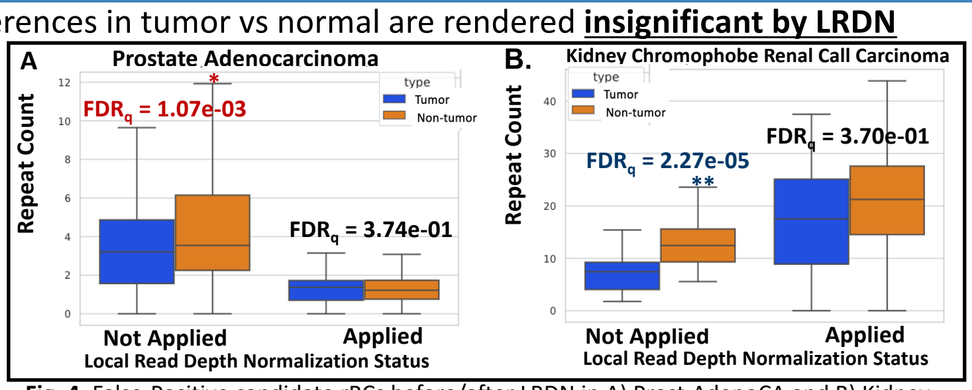

Fig. 4. False-Positive candidate rRCs before/after LRDN in A) Prost-AdenoCA and B) Kidney-ChRCC. Wilcoxon-rank sum test/False Discovery Rate[32-33]. *FDRq= 0.00107 **q=0.0000227

rRCs: VISUALIZATION IN PROSTATE ADENOCARCINOMA

- Electrophoresis visualizations: lower tumor weight vs normal, indicating rRCs (**38%;** Fig. 5)

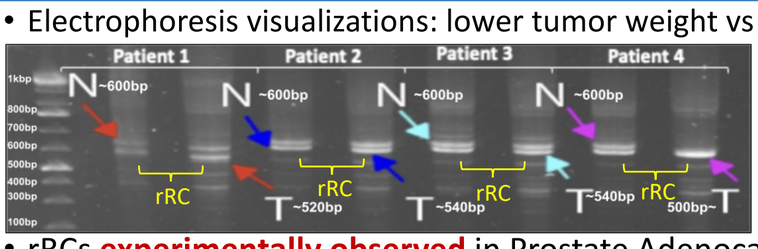

Fig. 5. Sample visualization of rRCs in Prostate Adenocarcinoma. Electrophoresis[34-35] of 4 patients at chr10:22006254-22007372. N: Non-Cancer Sample, T: Tumor Sample. Each pairing shows lower tumor bands, indicating presence of an rRC.

- rRCs **experimentally observed** in Prostate Adenocarcinoma.

rRCs: PRDM16 INTRONIC rRC in HCC CASE STUDY

- **PRDM16 intronic rRC may play functional roles in transcript usage and survival rate**

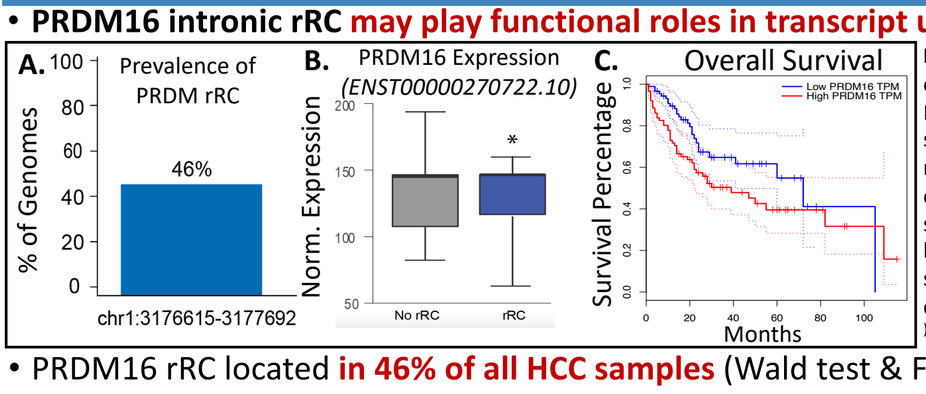

Fig. 6. PRDM16 rRC qualities. A) % of rRC expressed HCC Patients. B) PRDM16 isoform expression in HCC samples as a function of PRDM16 rRC detection[36-37]. Wald test/FDR correction[30]. C) Kaplan-Meier plots stratified by PRDM16 expression levels in HCC (lowest quartile by Z score; high quartile, RNA-seq quantitation; log-rank p=2.2 x10^{-7}).[38-39] *p = 0.0048

- PRDM16 rRC located **in 46% of all HCC samples** (Wald test & FDR; Fig. 6A; p=0.0048)
- Associated with a **significant decrease in transcript isoform usage** (Fig. 6B)
- High PRDM16 expression correlates to **poor HCC survival** (Fig. 6C; p=2.2x10^{-7})

rRCs/mcSTRs: GENOMIC QUALITIES TO OTHER REPEATS

- Cancer rRCs/mcSTRs are **genomically similar** to other repeats[40-46], yet **remain distinct**

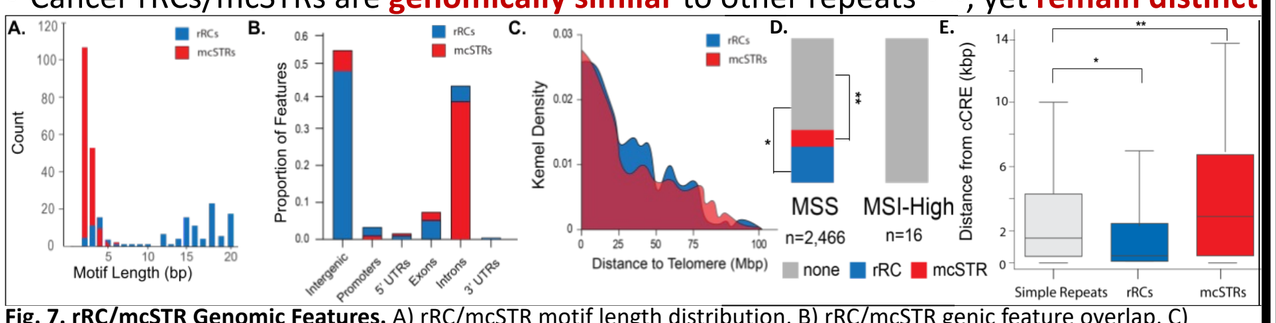

Fig. 7. rRC/mcSTR Genomic Features. A) rRC/mcSTR motif length distribution, B) rRC/mcSTR genic feature overlap, C) Telomeric distance rRC/mcSTRs, D) rRC/mcSTR distribution across MSS & MSI-high cancers. Chi-square test with Yates' correction[47-48] *p=0.04, **p=0.03 E) rRCs/mcSTRs cCRE distance vs Simple Repeats Catalog[49-51].*p=0.072 **p=0.002

- **Bimodal repeat distribution** (Fig. 7A); various genic features (Fig. 7B) favors telomeres (Fig. 7C)
- **No microsatellite instability enrichment** in samples ±mcSTR/rRC (Fig. 7D)[52-53]; rRCs are closer to cCREs than expected; mcSTRs located closer than expected (Fig. 7E)
- mcSTRs and rRCs may play pathogenic roles in **cancer cell death & gene expression**[54-56]

mcSTRs: EXPANSIONHUNTER (EH) + ACTIVESTRs CATALOG

- **mcSTRs are significant mutation sources of genetic variation in cancers** (q<0.05)

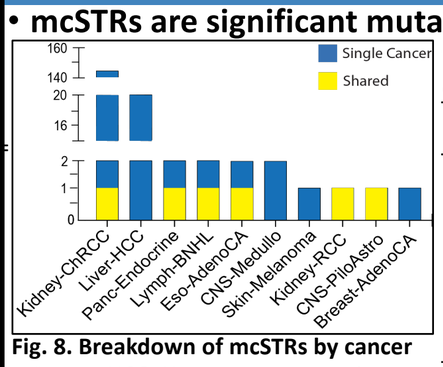

Fig. 8. Breakdown of mcSTRs by cancer subtype. HCC=Hepatocellular carcinoma.

Table 1. Overlap of mcSTRs to other repeat types in cancer. mcSTRs had little overlap to known repeats in cancer. rRCs and mcSTRs represent distinct sets

Repeat Type	No. of Repeats	Overlap to mcSTRs
Recurrent Repeat Contractions	120	**Zero**
Recurrent Repeat Expansions	160[51]	**Zero**
Microsatellites	200,000+[52-54]	**One (partial)**
Extrachromosomal Circular DNA	1270[55]	**149 (proximal/in)**

- **182 final mcSTRs** identified across **10 cancer types; 176/182 subtype specific (Fig. 8), allowing for cancer tissue-of-origin detection by mcSTR**
- **mcSTRs are distinct from other repeat types (Table 1);** may form from **alternative** cancer pathways/mechanisms; **precedent to be found in eccDNA**

mcSTRs: LIVER HEPATOCELLULAR CARCINOMA (HCC) CASE STUDY

- **Full contractions observed** in eight mcSTRs (Table 2), distinct **biomarkers in HCC**

Table 2. HCC mcSTRs contract in tumors. mcSTRs named proprietarily

Locus	Gene	Repeat	Tumor Full Contraction	Tumor Any Contraction	Normal Any Contraction
mcSTR1	MCG1 (intron)	(CGC)*	52%	60%	2%
mcSTR2	MCG2 (intron)	(GCG)*	60%	88%	29%
mcSTR3	MCG3 (intron)	(CGG)*	51%	56%	14%
mcSTR4	MCG4 (intron)	(CGCGCC)*	56%	65%	39%
mcSTR5	MCG6 (intron)	(GGC)*	41%	43%	8%
mcSTR6	intergenic	(GCG)*	53%	82%	33%
mcSTR7	intergenic	(GCC)*	26%	77%	59%
mcSTR8	MCG8 (exon)	(GCN)*	9%	31%	3%

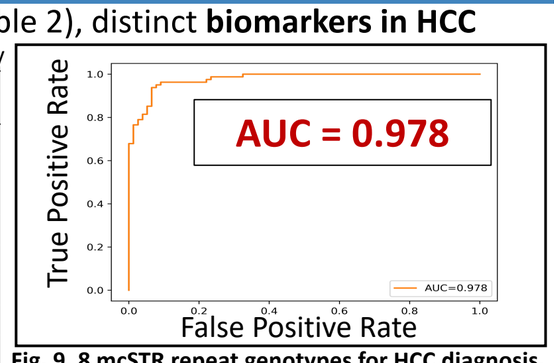

Fig. 9. 8 mcSTR repeat genotypes for HCC diagnosis.

- 8 mcSTR genotypes effectively **distinguish** between **tumor vs normal samples** (Fig. 9)

mcSTRs EXPERIMENTALLY VALIDATED IN HCC TUMOR/PLASMA

- Sequencing of 20 HCC tumor/normal tissues displayed **contractions in 55% of tumor samples**

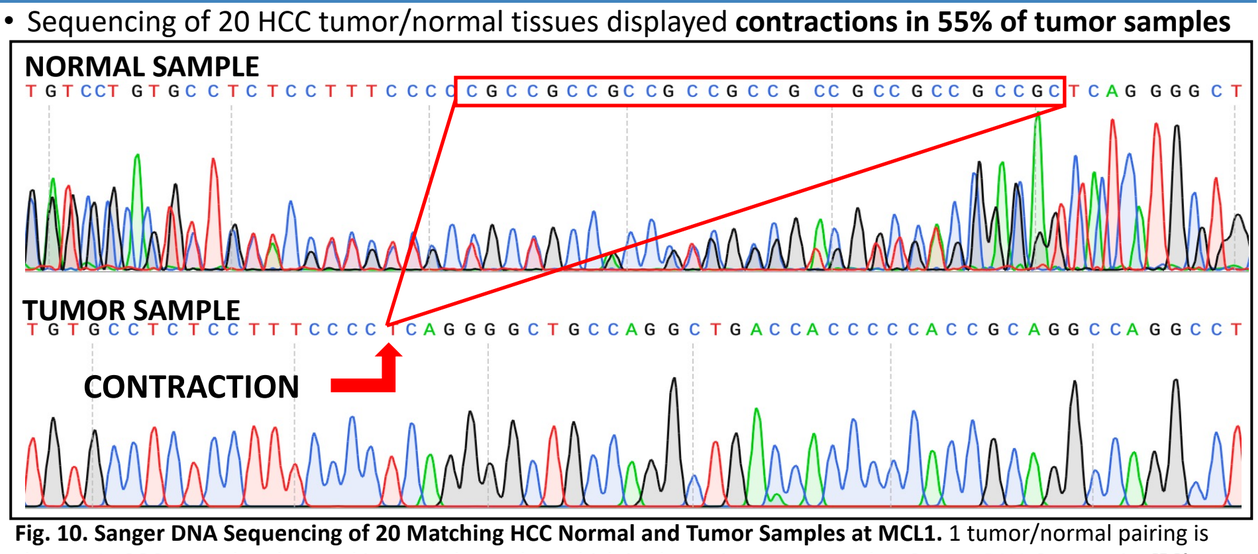

Fig. 10. Sanger DNA Sequencing of 20 Matching HCC Normal and Tumor Samples at MCL1. 1 tumor/normal pairing is shown. A $(CGC)_{10}$ can be observed in normal samples, which is absent in tumor samples. Sanger DNA Sequencing[57-58]

- **$(CGC)_{10}$ mcSTR present in normal tissue; contracted in tumor tissue** (Fig. 10)
- **mcSTR1 contraction experimentally observed in hepatocellular carcinoma**
- Two primer sets: Control primers (always-bind) and Contraction-only primers[59-60]

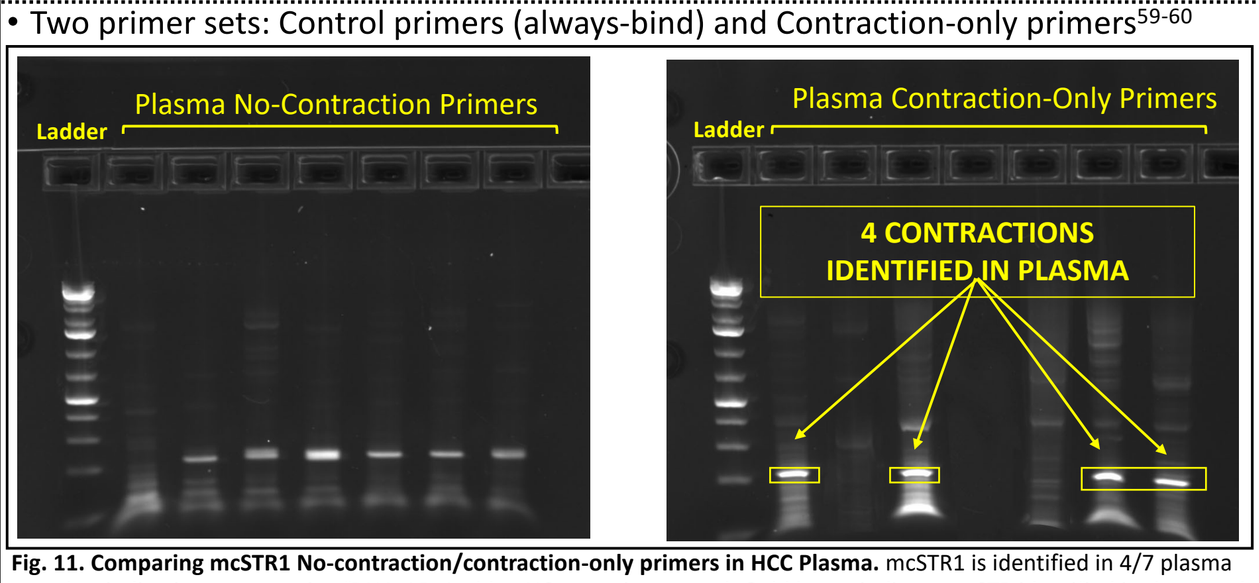

Fig. 11. Comparing mcSTR1 No-contraction/contraction-only primers in HCC Plasma. mcSTR1 is identified in 4/7 plasma samples, indicating presence in cfDNA. LD=Ladder. NC=negative control. Gold boxes indicate mcSTR identified in the sample.

- **0/20 HCC normal, 11/20 HCC tumor, 4/7 plasma HCC samples exhibit mcSTRs** (Fig. 11)
- **mcSTR1 experimentally observed in HCC cfDNA, indicating early cancer biomarker**

CONCLUSIONS

I. **Local Read Depth Normalization:** new algorithm that removes false-positive rRCs
II. **ActiveSTRs Catalog:** effective **benchmark dataset** for evaluating pathogenic mcSTRs
III. **rRCs/mcSTRs: NOVEL mutations** spanning 15 cancers; **functionally similar to other repeats**
IV. **Liver-HCC mcSTRs:** effective diagnostic panel; **experimentally identified in tumor/plasma**
V. **PRDM16 rRC:** related to **HCC expression/survival**, clinically applicable
VI. **Global mcSTR Burden:** associated to DNA repair gene pathways and cancer expression

FUTURE INVESTIGATIONS/APPLICATIONS

- cfDNA rRCs/mcSTRs are promising **biomarkers in early, localized cancer detection**(Fig. 12)[61-62]
- rRC/mcSTR-targeted therapies may allow for **novel cancer medicines**[63-65]

**All images/graphics/figures created by finalist*

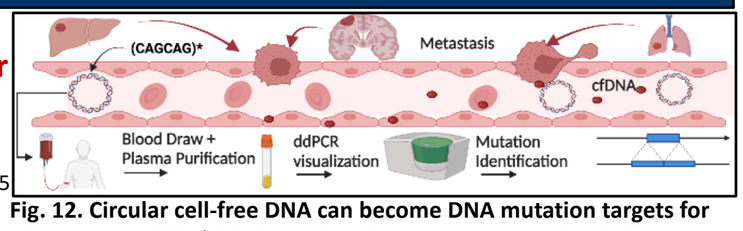

Fig. 12. Circular cell-free DNA can become DNA mutation targets for diagnosis. mcSTRs/rRCs identified via sequence detection in DNA