

Solving Lynch syndrome missing heritability by integrating long-read DNA sequencing and short-read RNA sequencing

Christopher J. Hatch¹, Sami Belhadj¹, John Harting¹, Cassidy Carraway¹, Hoda Mirsafian¹, Stuti Joshi¹, Kenny Chen¹, Esther Hsiao¹, Carolyn Horton¹, Stephen Tanner¹, Bojan Losic¹, Marcy E. Richardson¹, Seth I. Berger¹, Rachid Karam¹

¹Ambry Genetics, Aliso Viejo, CA, USA



BACKGROUND

Lynch syndrome (LS) increases the risk of developing numerous cancers. It is caused by inheritable pathogenic germline variants (PGV) in the DNA mismatch repair (MMR) genes, *MLH1*, *MSH2*, *MSH6*, and *PMS2*, as well as constitutional *MLH1* promoter hypermethylation (CMPH or epimutations). Conventional testing often returns negative results potentially due to inherent limitations in detecting structural variants (SV), regions complicated by pseudogenes, and epimutations. In this study, we sought to explore LS missing heritability by using Long-Read Sequencing (LRS) and leveraging RNA data to identify cases with potentially undetected PGV.

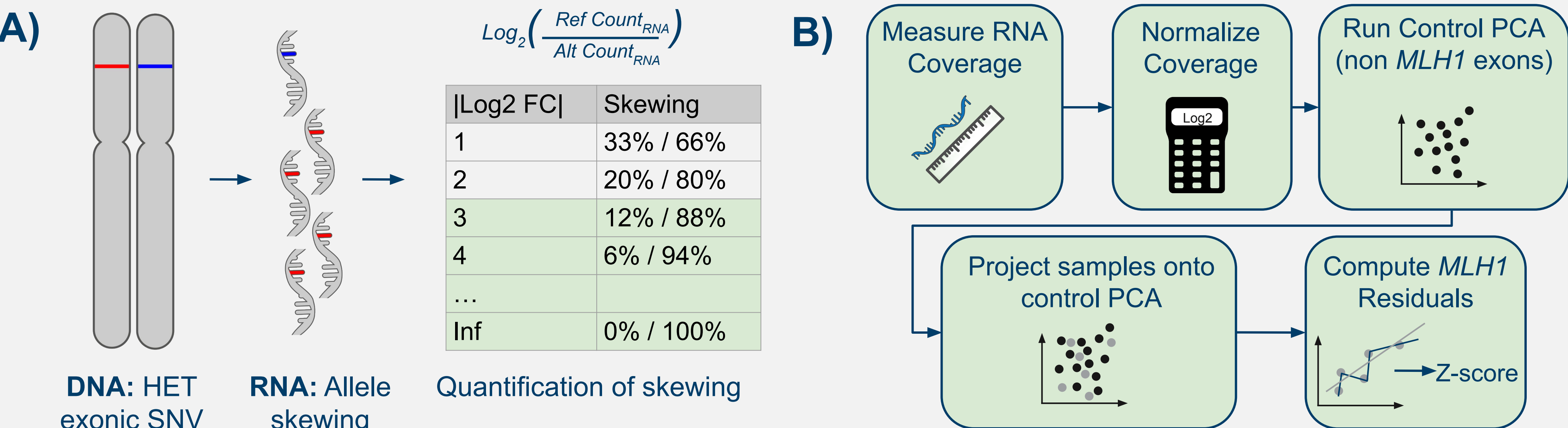
METHODS

• PacBio HiFi sequencing

- T-LRS: Targeted sequencing of the MMR genes using a custom hybrid capture panel by Twist Biosciences, with UTR-UTR coverage. Read lengths averaged 3.5 kb and sequencing depth ~100x.
- LR-WGS: Whole genome sequencing with 15 kb read lengths and 20-30x coverage.

• RNA analysis to prioritize cases for LR-WGS

- A) Allele-specific expression analysis (ASE): identifies RNA skewing in heterozygous exonic variants.
- B) *MLH1* RNA expression analysis: identifies expression outliers based on sequencing coverage.

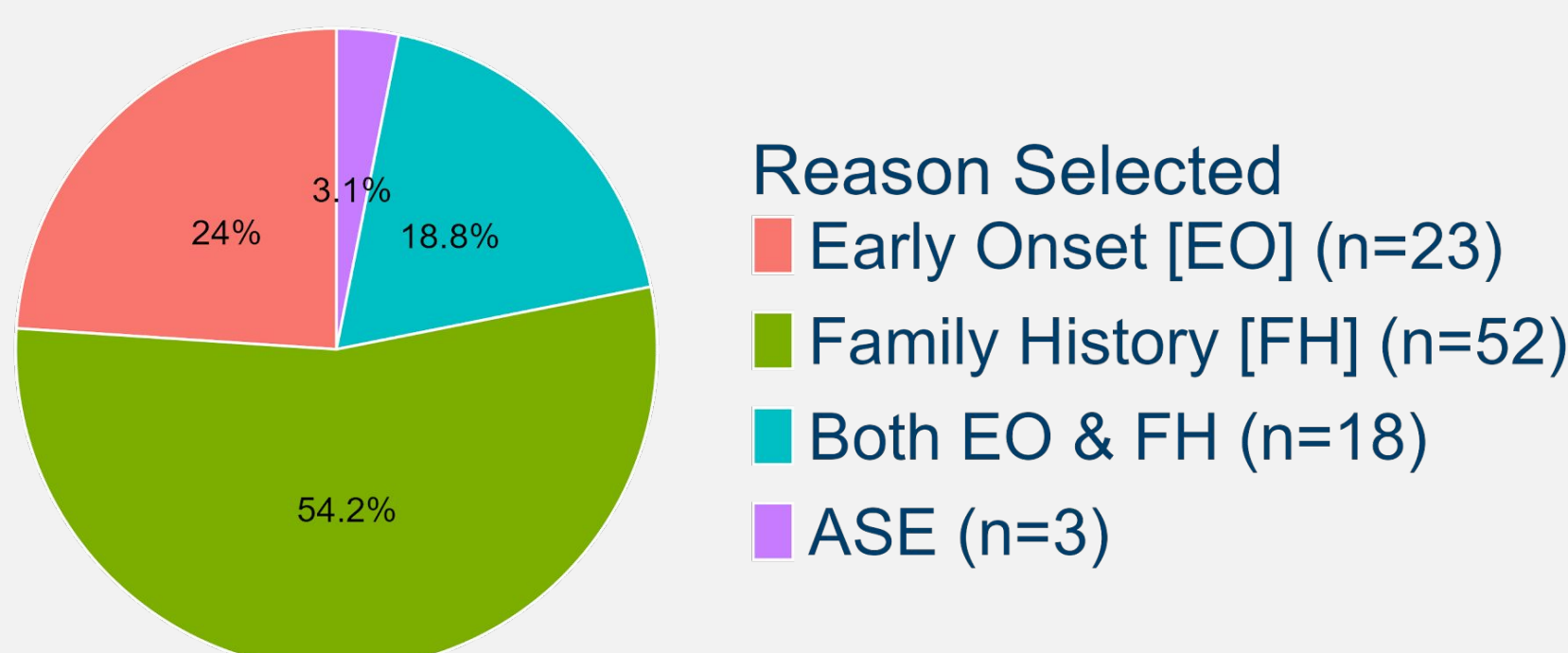


PATIENT COHORTS

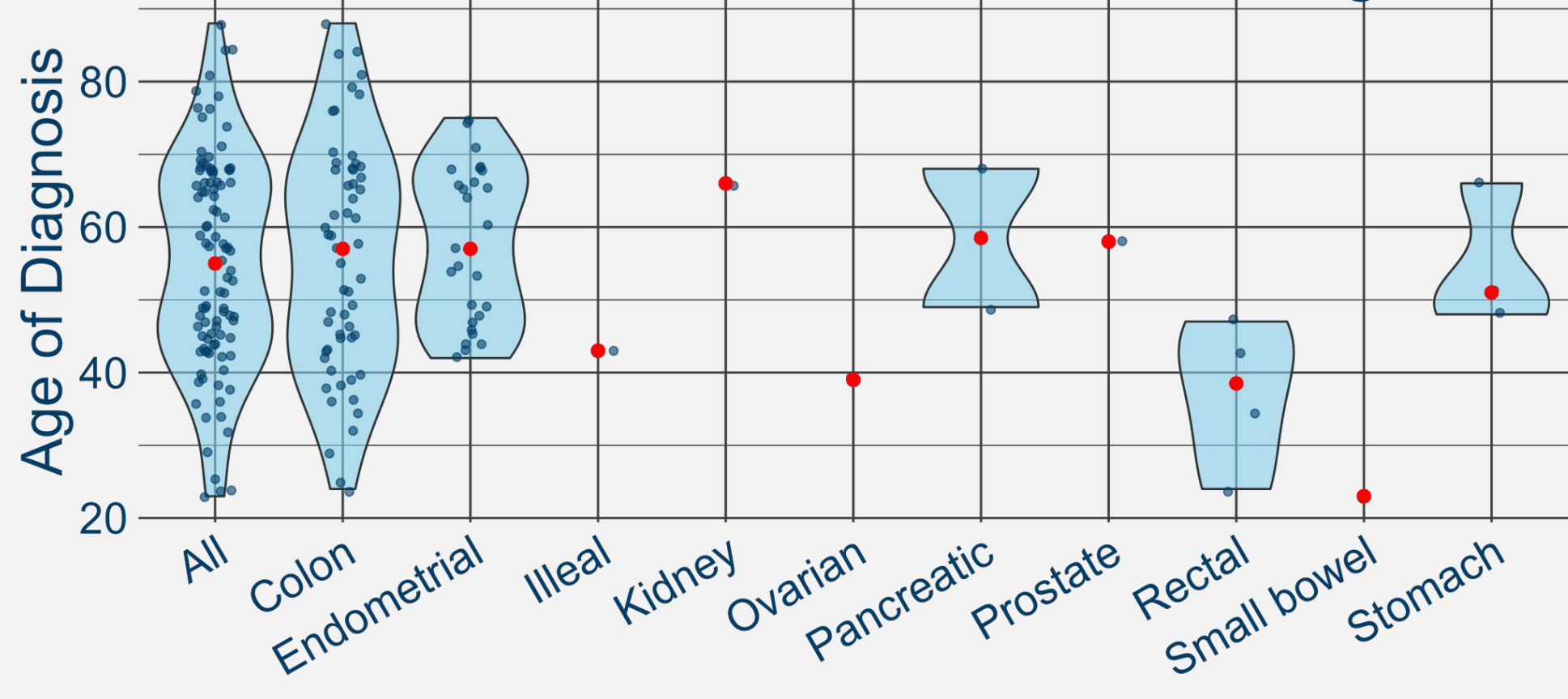
We selected three cohorts to explore LS missing heritability:

Cohort	N	Inclusion Criteria	ASE analysis	<i>MLH1</i> expression outlier analysis	T-LRS	LR-WGS
1	96	MMR-deficient, manually curated personal and family LS cancer history.	✓	✓	96/96	3/96
2	250	MMR-deficient, unselected for personal and family history.	✓	-	3/250	3/250
3	9,420	Non-curated. Referred for LS testing.	✓	✓	-	16/9,420

Cohort 1: Inclusion Criteria



Cohort 1: Personal Cancer Diagnoses



COHORT 1: T-LRS does not identify causative variants

Table 1. T-LRS detected variants in MMR genes. 96 patients underwent T-LRS. **Left)** Small variants prioritized by *in-silico* predictors were either known benign variants (n=4) or had no supporting RNA evidence (n=1). **Right)** Structural variants had intronic breakpoints without intersecting coding exons. RNA data was leveraged to discard an impact on RNA splicing.

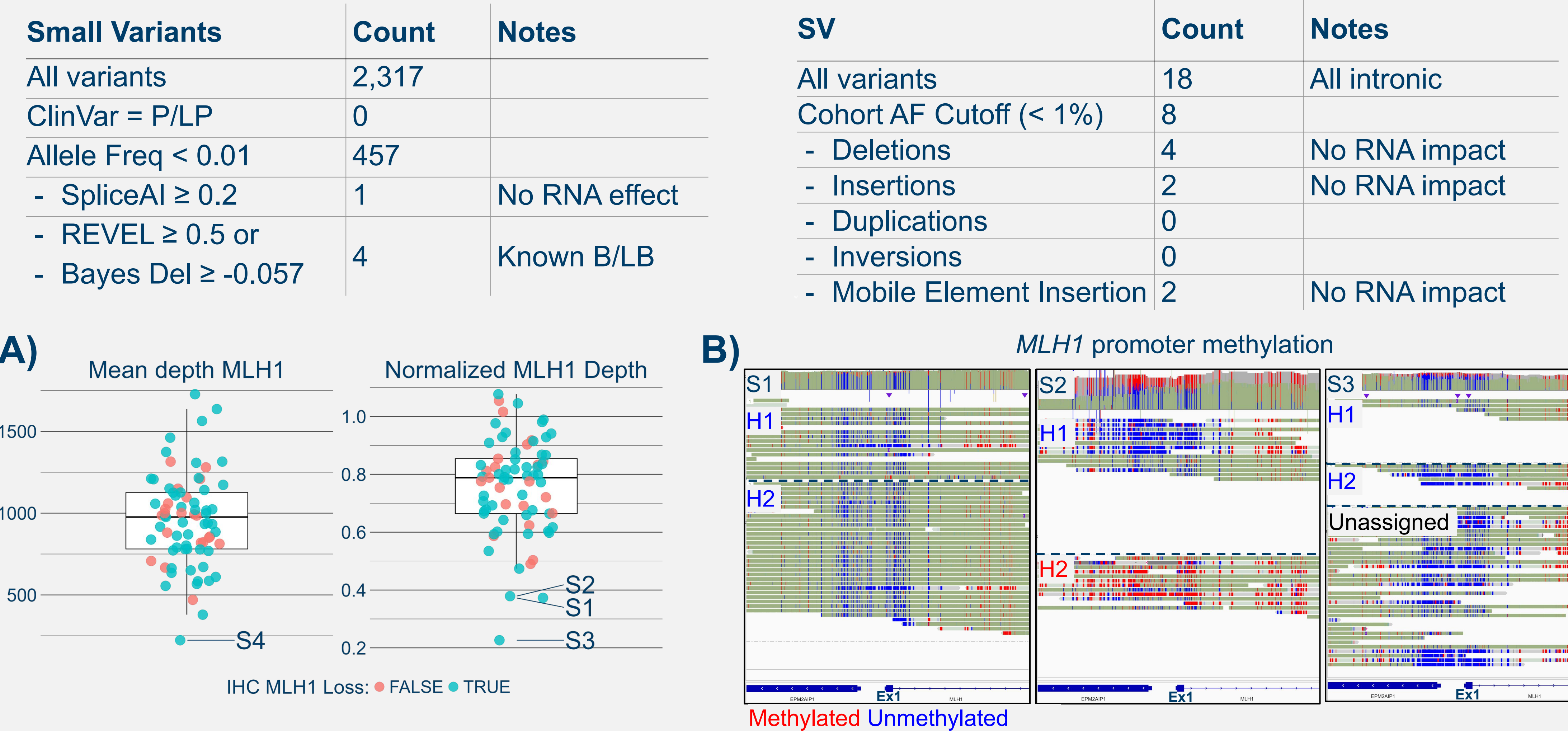


Figure 1. Cohort 1 has 1 *MLH1* epimutant detected by RNA expression. **A)** Normalized depth (*MLH1* depth vs Panel depth) detects 3 potential epimutants (S1-S3). **B)** LR-WGS of outliers shows S2 has constitutional *MLH1* promoter methylation. For S1-S3, no additional causative variants were found.

COHORT 3: RNA Expression detects *MLH1* epimutants

Category	#	Metric cutoff	Candidates	Selected for LR-WGS	Methylated	SV	SNV	Suspected
Expression only	5,468	Strong: Z-score ≤ -3	14 (0.15%)	8	1	0	0	2
ASE	3,952	Strong: abs(Log2FC) ≥ 2	7 (0.07%)	2	0	0	0	0
Expression + ASE	3,952	Strong: Z-score ≤ -3 & abs(Log2FC) ≥ 1.4	4 (0.04%)	4	1	2	0	0
		Moderate: Z-score ≤ -1.4 & abs(Log2FC) ≥ 1.8	2 (0.02%)	2	0	0	0	0

MLH1 RNA Expression Results

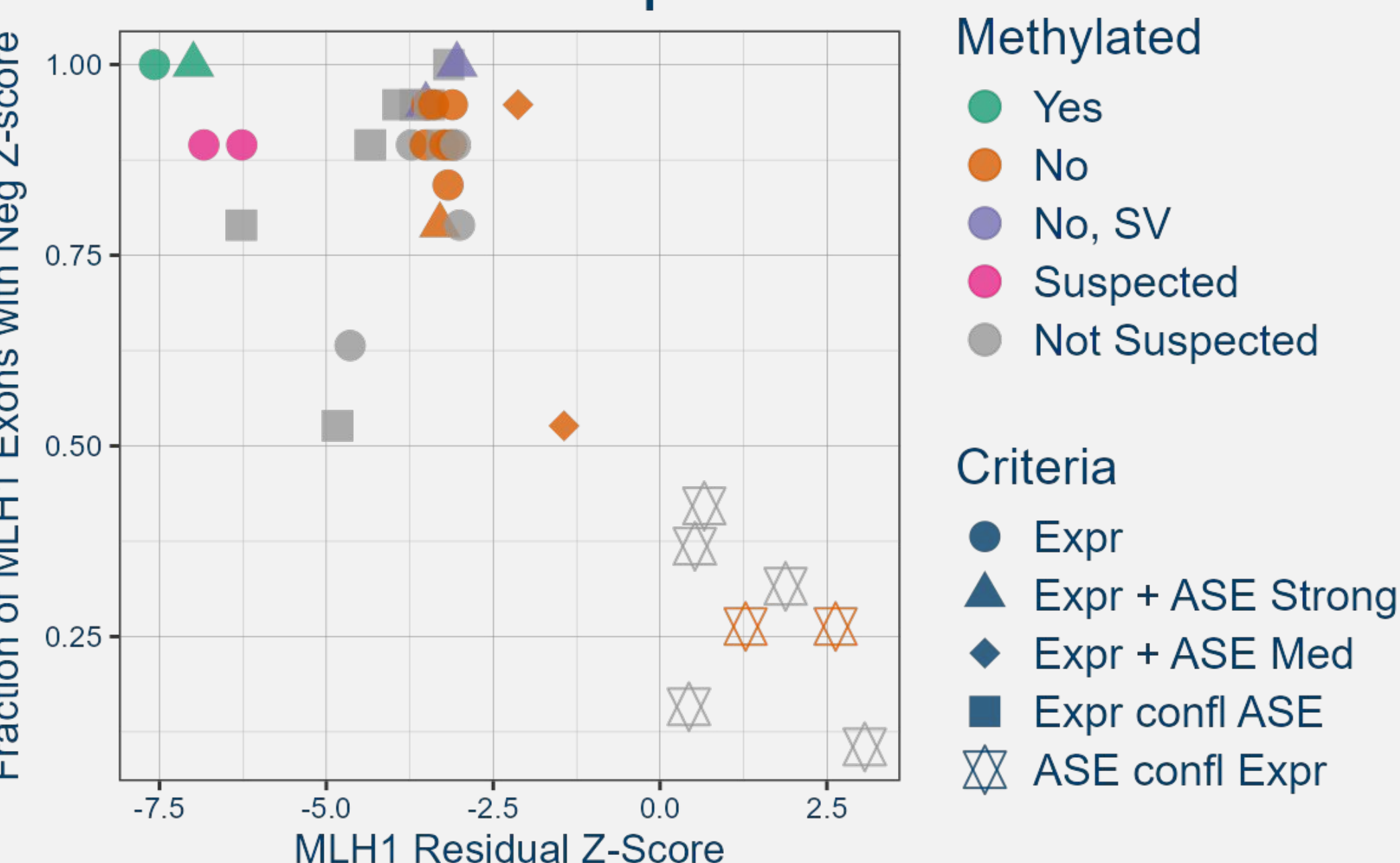


Figure 3. *MLH1* RNA Expression Outlier analysis and ASE detect *MLH1* epimutants in a non-curated cohort of 9,420 patients. Confirmed epimutants (teal) cluster close to suspected epimutants (pink) that were not sequenced due to sample availability. Combined, an estimated solve rate is 0.04%. 2 samples have likely benign structural variants (purple).

Expr: Expression Outlier, confl: conflicting, Med: Medium

COHORT 2: ASE analysis identifies *MLH1* epimutants

Table 2. Allele Specific Expression identifies 3 patients with monoallelic *MLH1* expression. phASER¹ detected monoallelic expression at 1.2% (3/250). All 3 cases were diagnosed with early onset CRC, *MLH1*/*PMS2* loss in their tumors, but no strong family history.

Case ID	Cancer Hx	IHC	Family Cancer History (LS)	Gene	# Ref	# Alt	# Total	Log ₂ _aFC	Variant name
AMB259	CRC (40s)	MLH1/PMS2	2 SDR, Prostate Ca. (60s)	<i>MLH1</i>	0	299	299	Inf	c.655A>G
					111	0	111	Inf	c.*35_*37delCTT
AMB790	CRC (20s)	MLH1/PMS2	None	<i>MLH1</i>	95	1	96	6.57	c.*35_*37delCTT
AMB918	CRC (30s)	MLH1/PMS2	1 TDR, CRC (70s)	<i>MLH1</i>	36	4	40	3.17	c.*35_*37delCTT, c.*162_*165delGATT

CRC: colorectal cancer, IHC: immunohistochemistry, SDR: second-degree relative, TDR: third-degree relative

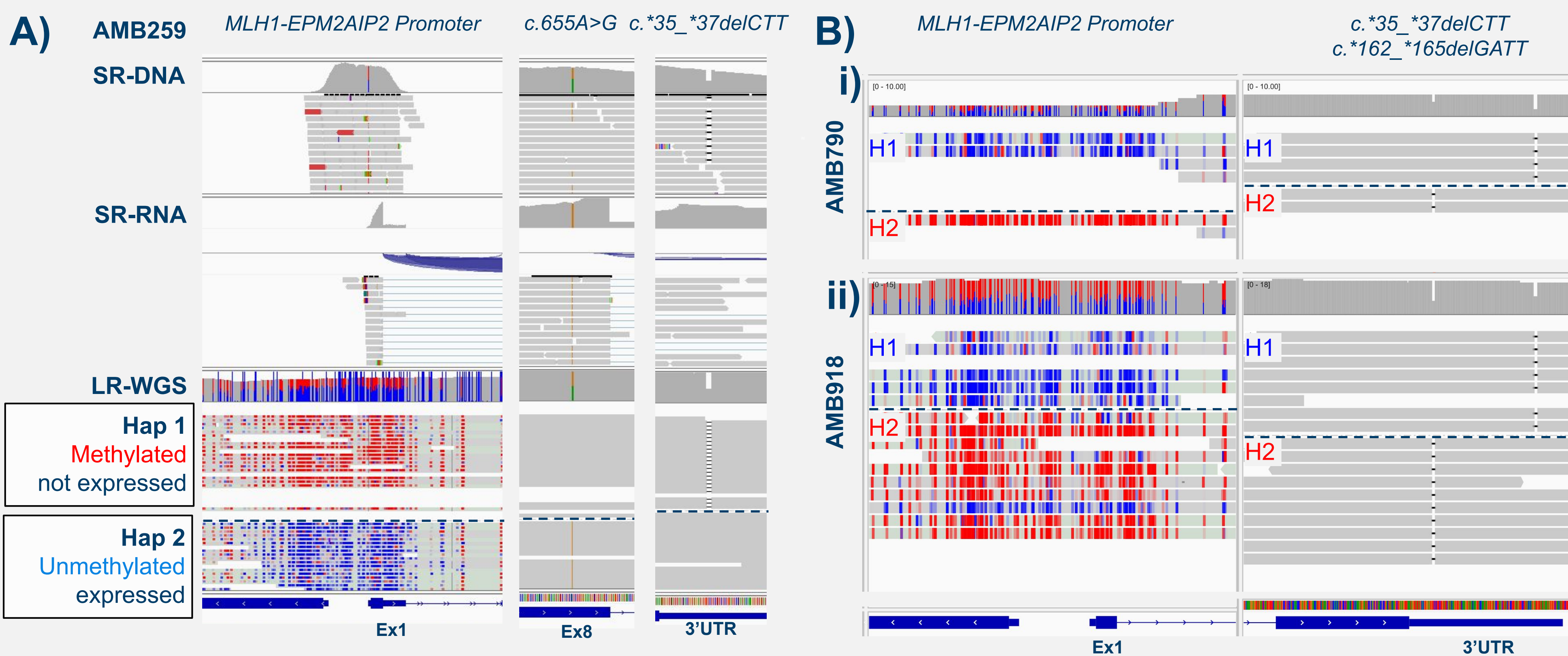


Figure 2. Cohort 2 has 3 *MLH1* epimutants prioritized by ASE analysis. **A)** LR-WGS of AMB259 uncovers haplotype specific methylation that silences the allele not expressed in RNA. **B)** LR-WGS in i) AMB790 and ii) AMB918 uncovers haplotype specific methylation.

1. SE Castel, et al. *Nat Comms* (2016). doi: doi: 10.1038/ncomms12817

CONCLUSIONS

- Targeted long read sequencing did not improve diagnostic yield in our cohort of patients suspected to have Lynch syndrome; however, a larger cohort may be required to determine the full diagnostic utility of LRS.
- Short-read RNA data played a vital role, either by facilitating the interpretation of LRS findings or by helping in the prioritization of cases for LR-WGS.
- Constitutional *MLH1* promoter hypermethylation may be an underappreciated driver of Lynch syndrome:
 - In a cohort restricted to patients presenting with MMR-deficiency, ASE analysis detected *MLH1* epimutations in 1.2% of cases.
 - In an unselected cohort of patients referred for LS testing, RNA expression outlier analysis detected *MLH1* epimutation in up to 0.04% of cases, many of which could not undergo ASE analysis due to the lack of informative exonic variants.