

Concurrent Long-Read Genome Sequencing with Methylation Pattern Analysis Enhances Diagnostic Yields in the UCI-GREGoR Rare Disease Cohort

Seth I. Berger¹, Arthur Bandeira de Mello Garcia^{1,2}, Greta Pitsava^{1,2}, Krista Bluske¹, Rebekah Barrick², Nida Karra¹, Sarah L. Stenton¹, Sami Belhadj¹, Sara E. Kalla¹, Stuti Joshi¹, Kenny Chen¹, John Harting¹, Ivan De Dios², Miguel Almalvez², Arthur Ko², Charles Hadley King², Jonathan LoTempio², UCI-GREGoR Consortium, Emmanuèle C. Délot², Changrui Xiao², Rachid Karam¹, Eric Vilain²

¹Ambry Genetics, Aliso Viejo CA; ²University of California-Irvine, Irvine CA



Ambry Genetics[®]
A TEMPUS COMPANY



BACKGROUND

Genome-wide reproducible patterns of hyper- and hypo-methylation characteristic of specific disorders are described in the literature for a range of monogenic syndromes. These epigenetic signatures also exist for syndromes caused by fetal teratogenic exposures, such as alcohol and valproate, and disorders without known genetic variants, such as the recurrent constellation of embryonic malformations (e.g. VACTERL). Array-based testing for these signatures is used when clinical suspicion is high or to disambiguate variants of uncertain significance (VUS), while testing guidelines for malformations and neurodevelopmental disorders focus on sequencing assays. We studied concurrent evaluation of disease methylation patterns in an undiagnosed disease cohort with long-read whole genome sequencing generating sequence and methylation data in a single assay.

METHODS

- Participants enrolled through the University of California-Irvine Genomics Research to Elucidate the Genetics of Rare Disease (GREGoR) or selected from internal controls from Ambry Genetics
- Hi-Fi genome sequencing on PacBio Revio from either blood, buccal, or saliva samples.
- PacBio WDLv3 aligned to hg38 and generated genome-wide calls for small variants, structural variants, repeat expansions, and genome-wide methylation calls
- Methylation data aggregated, normalized, and scored against literature curated pattern database

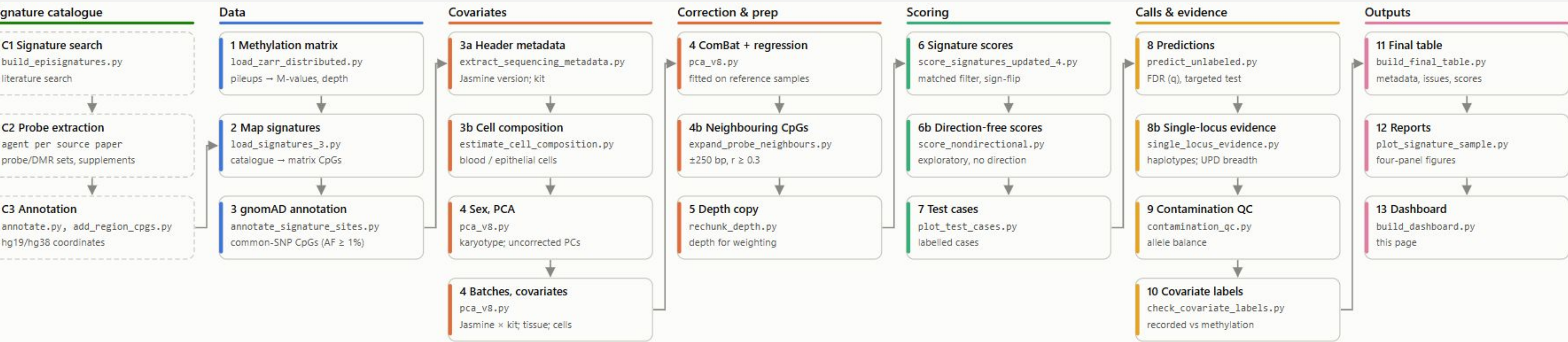
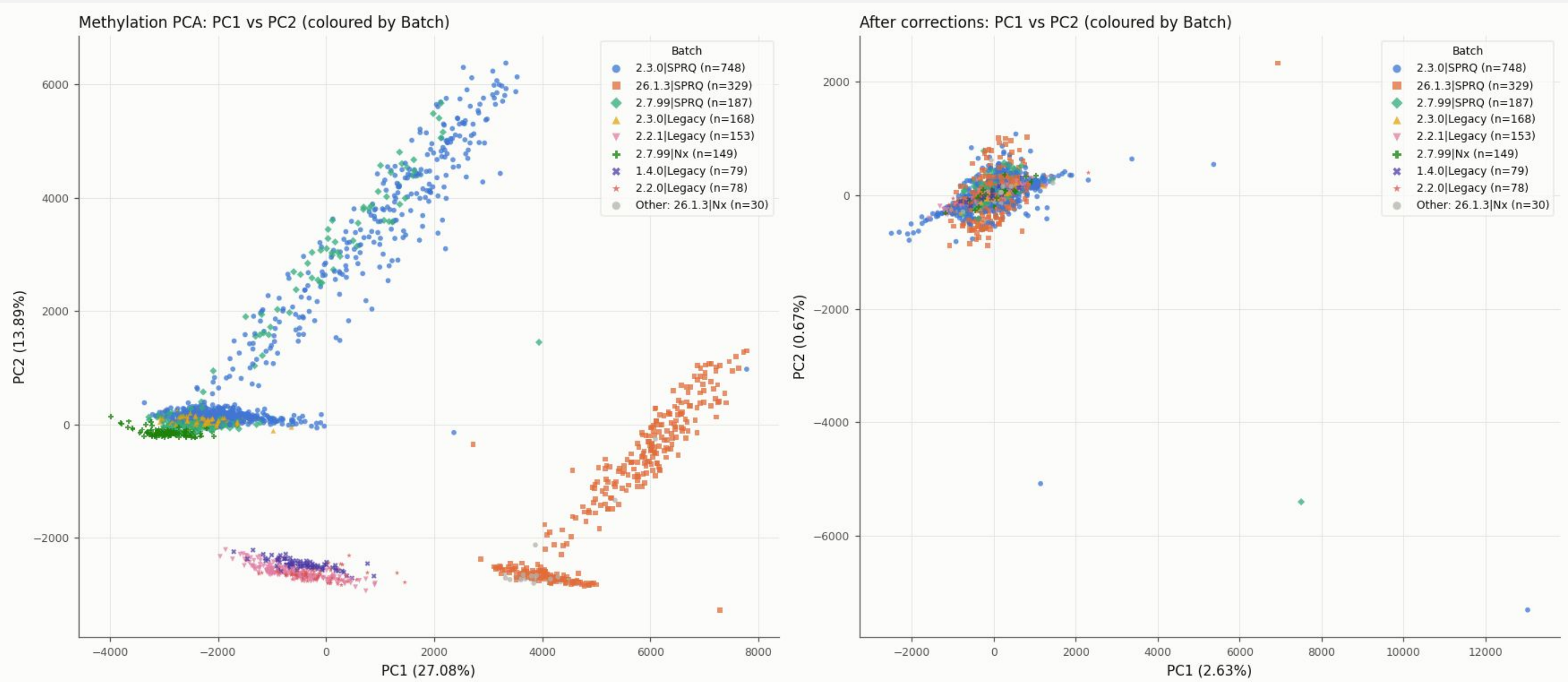


Figure 1: Epigenetic pattern scoring pipeline using literature agent-curated dataset. Literature derived signatures included **91** Mendelian disorders, **28** imprinting/locus-specific disorder, and **3** teratogen exposures. These were scored across **1,921** HiFi-WGS samples.

Batch Correction

Figure 2: Batch

normalization of global methylation calls corrects for Jasmine Methylation Caller and sequencing chemistry version. Additional imputed continuous tissue composition, methylation age, and sequencing metric normalizations were applied from derived covariates.



VUS disambiguation and variant prioritization



Figure 3: Representative example signature. Signature distribution (above) for *DNMT3A* associated Tatton-Brown-Rahman syndrome.

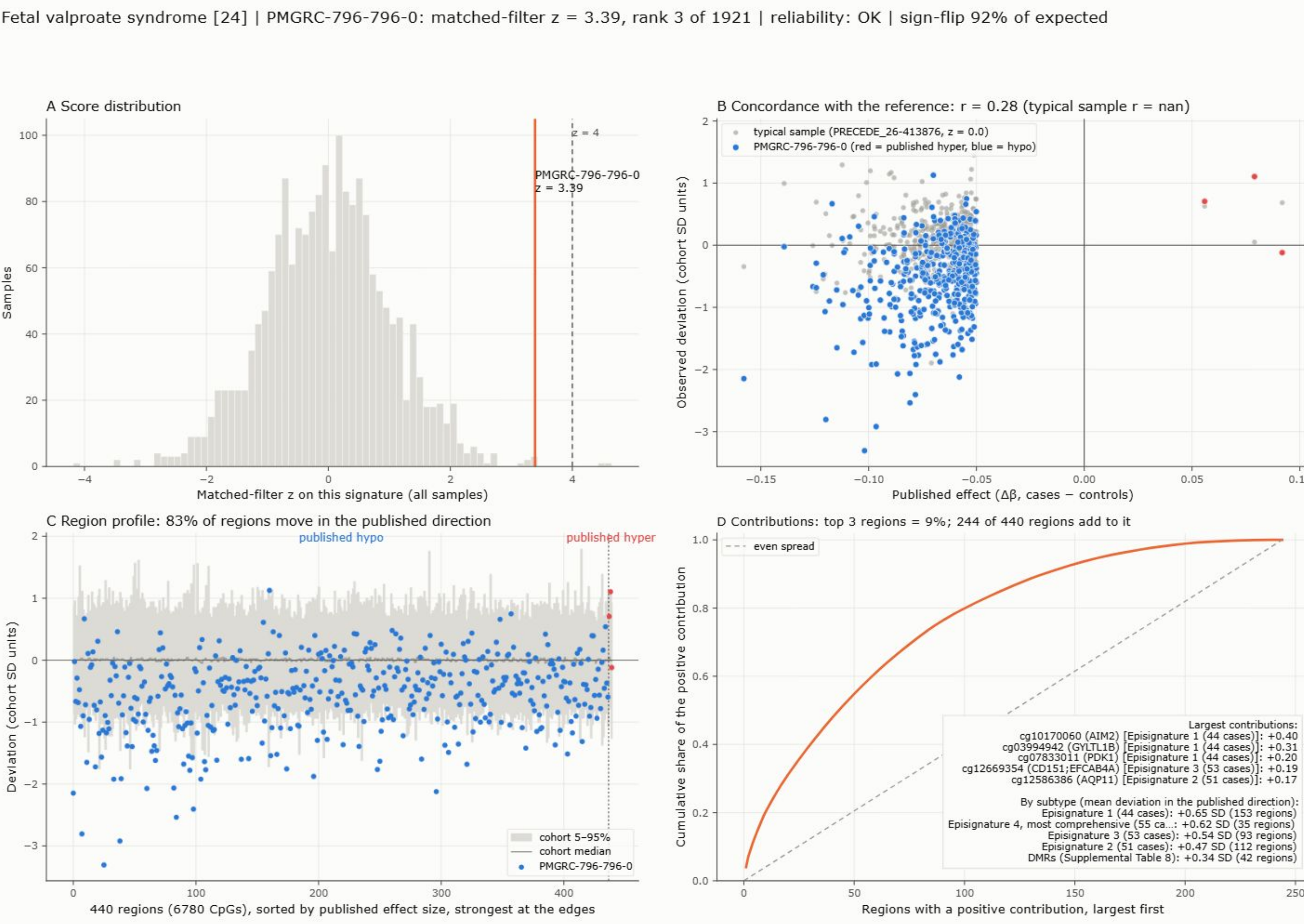
- 3 cases with known pathogenic variants
- 3 cases with VUS
- All ranked among top 7 samples
- One VUS not originally prioritized due to limited phenotypic match and subsequently prioritized based on the signature finding.
- Labeled *DNMT3A* test cases clustered together (right), had targeted signature $p < 0.0001$ providing evidence for VUS reclassification.

Diagnosis of fetal teratogen exposure

Figure 4: Signature distribution for fetal valproate syndrome.

- A) Undiagnosed sample scored highly ($Z=3.39$) on fetal valproate syndrome signature consistent with phenotype of congenital malformations including radial anomalies.
- B) Individual probes used for scoring signature concordant with published beta values
- C) Consistent hyper- and hypo-methylation at probe sites.
- D) Signature is not predominated by subset of sites.

Re-contact of participant confirmed previously undisclosed prenatal valproate exposure.



Decreased Sensitivity on Buccal/Saliva

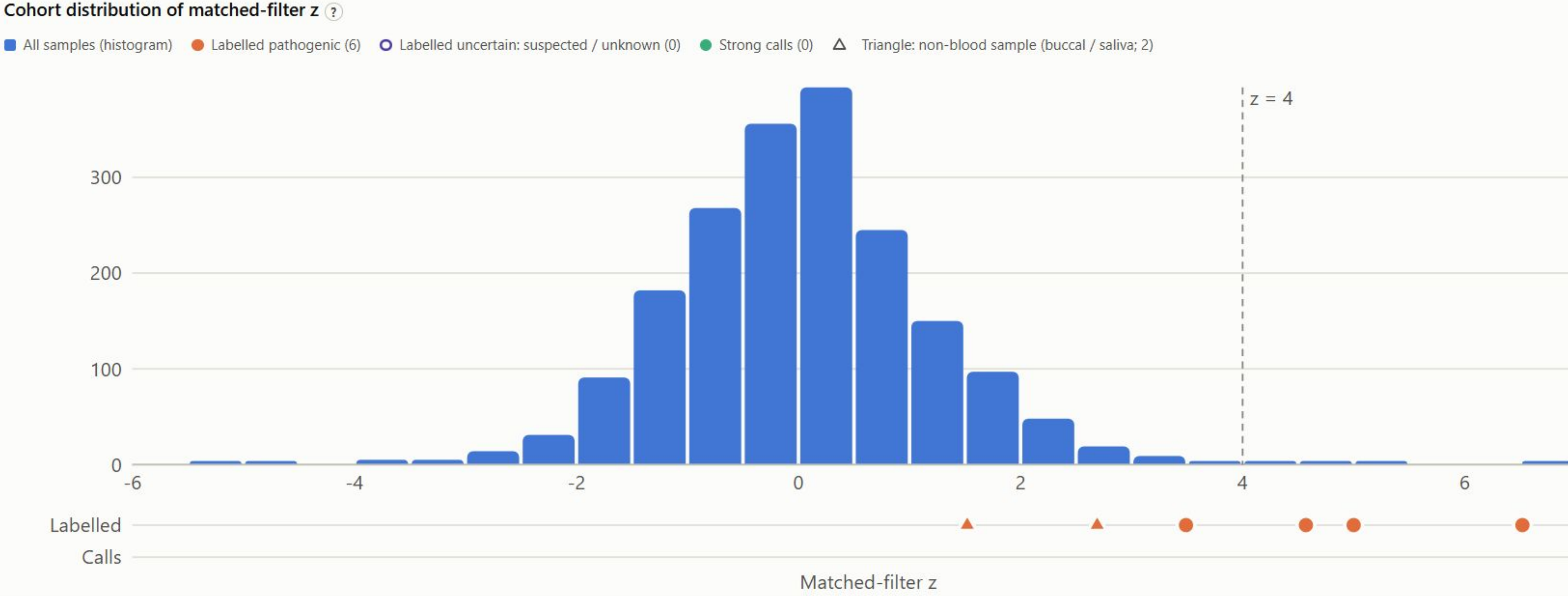
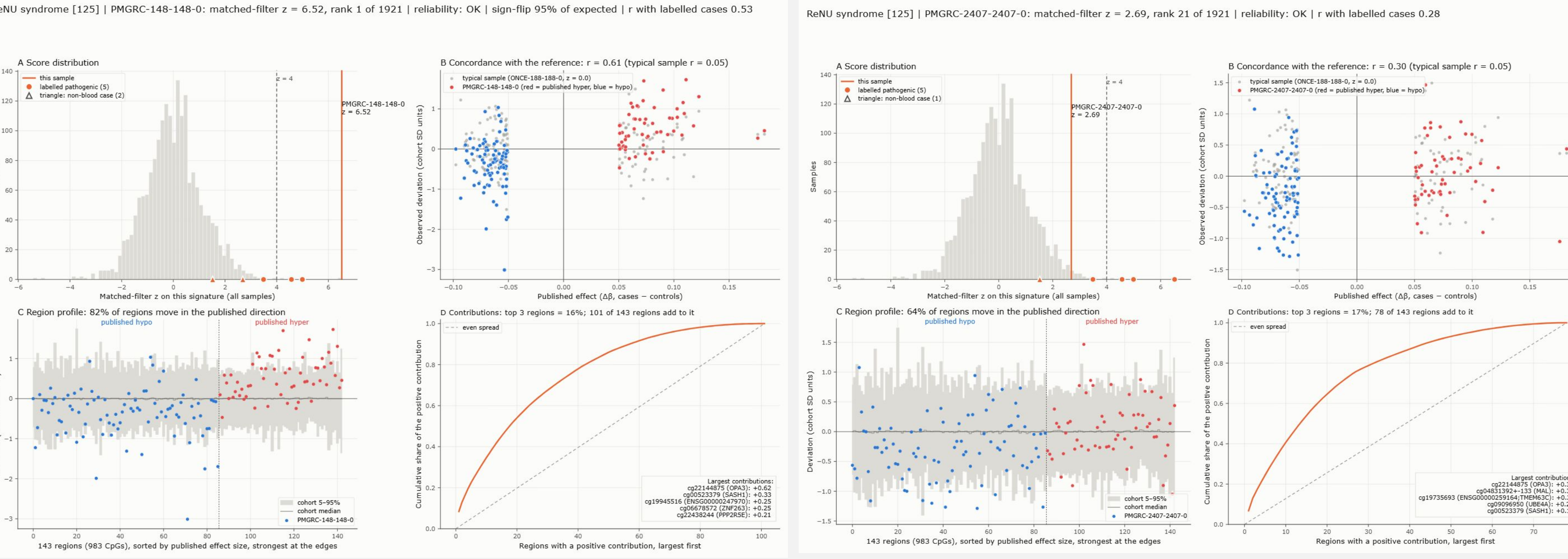


Figure 5. Signature distribution (above) for ReNU syndrome including 6 known positive samples (4 from blood, and 2 buccal samples). Buccal samples show decreased sensitivity. Representative reports from a strong positive blood (below left) and buccal sample (below right).



CONCLUSIONS

- Long-read methylation profiles can be scored using epigenetic patterns published from array data with concordant signatures for 37 of 47 testable known positive samples.
- Methylation signatures from long-read genomes can resolve undiagnosed cases.
 - Reflex epigenetic profiling provided evidence for VUS upgrades with 13/39 samples with VUS concordant with a pathogenic signature.
 - Epigenetic profiles can diagnose disorders without genetic etiologies (e.g. teratogens) including fetal valproate syndrome and fetal alcohol syndrome.
 - Integrating epigenetic profiling with sequencing can assist variant prioritization.
- Blood-derived signatures are detected in buccal/saliva with decreased sensitivity.
- Long-read genome sequencing with epigenetic analysis may provide further test consolidation for the diagnostic workup of rare diseases.