

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

Seth I. Berger^{1,2}, Arthur Bandeira de Mello Garcia^{1,2,3}, 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 Vilain^{1,2}

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 which generates both sequence and methylation data in a single assay.

Participants enrolled through the University of California-Irvine Genomics Research to Elucidate the Genetics of Rare Disease (GREGoR) site received Hi-Fi genome sequencing from either blood, buccal, or saliva samples. The PacBio WDLv3 generated genome-wide calls for small variants, structural variants, repeat expansions, and genome-wide methylation calls. Methylation profiles across the dataset were batch normalized and compared to a literature curated database of published methylation signatures for 40 disorders with significant outliers assessed against the genetic variants and clinical histories.

Significant outlier methylation signatures were detected in several cases. Several previously identified VUS, including variants in *CHD4*, *PHIP*, *ARID1A*, and *DNMT3A*, were upgraded to likely pathogenic. Another likely-pathogenic variant in *DNMT3A*, previously not prioritized due to insufficient phenotypic data, was identified due to the methylation signature. An epigenetic signature for fetal valproate embryopathy was detected in a proband with malformations consistent with this disorder and history consistent with *in utero* valproate exposure. Clinically suspected fetal alcohol syndrome and VACTERL association were supported with their respective signatures in additional cases.

Long-read sequencing with integrated methylation calling provides useful methylation signature information for undiagnosed cases through both variant disambiguation and novel findings. Genome sequencing with epigenetic analysis shows promise as a future test consolidation for the diagnostic workup of rare diseases.