A Population-Scale Targeted Methylation Sequencing Platform for Human Disease and Exposure Biology

Michelle Cipicchio[1], Can Kockan[1], Sophie Low[1], Jeff Cao[2], Andrew Ratanatharathorn[2], Fabiana Duarte[2], Mark Fleharty[1], Chris Kachulis[1], Sam Pollock[1], Katie Larkin[1], Sean Hofherr[1], Niall Lennon[1], Carrie Cibulskis[1], Jason Buenrostro[2,3]

1 Broad Clinical Labs, Burlington, Massachusetts, USA
2 Broad Institute, Epigenomics Program, Cambridge, Massachusetts, USA
3 Harvard University, Department of Stem Cell and Regenerative Biology, Cambridge, Massachusetts, USA

Introduction

Goal

DNA methylation is a primary mechanism for cellular response to the environment. Our project aims to identify the epigenetic signatures associated with trauma and adversity, to uncover the biological pathways at the root of chronic diseases associated with trauma.

Due to the complexity of the underlying biology, a massive and diverse cohort of samples is required to identify and decode the epigenetic molecular markers of adversity.

Challenge

Epigenome-wide association studies (EWAS) are an essential tool to identify these biomarkers across a range of disease types. Despite this, population scale epigenomics is currently understudied due to challenges in the standard assays for DNA methylation.

Resolving Challenges in DNAm Assay Landscape

Challenge	Solution
Cost	Cost effective short read next generation sequencing
Sensitivity / Coverage	Enables standard targeted capture workflows with panels for high coverage of regions of interest
DNA Integrity	Only methylated cytosines are converted in a gentle process that does not damage or fragment DNA
Scalability	Massive scale enabled with automation friendly workflow and standard NGS

Solutions

Our workflow overcomes the barriers in standard DNA methylation assays by leveraging new methods, using short read sequencing after gentle conversion of methylated cytosines.

Multimodal: this method preserves the sequence diversity of the genomic DNA, which improves sequence alignment, enables DNA variant calling, and is compatible with targeted capture using panels for regions of interest.

We have optimized a workflow that leverages high throughput scale to reduce costs and streamline lab processing, laying the groundwork to enable population EWAS projects. The methods implemented here enable high throughput EWAS due to efficiency of workflow and cost that have never before been available to facilitate epigenetic research.

Materials & Methods

This workflow addresses each of the challenges in the methylation assay landscape. The Watchmaker TAPS+ chemistry employs a two step gentle conversion after library preparation, enabling detection of both 5mC and 5hmC bases in the library.

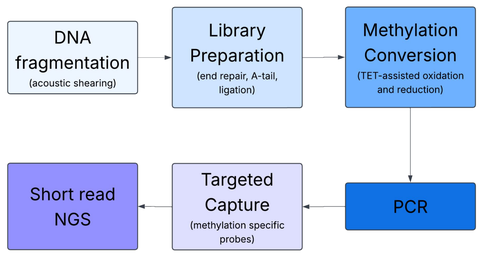


Figure 1. The wet lab workflow is modular and automatable enabling high throughput scale in batches of 96 samples.

Sample Cohort

We leverage the Nurses' Health Study II, a prospective cohort pairing validated adversity exposure data, decades of health outcomes, and available biospecimens.

- Responses for childhood and adolescent physical/sexual abuse for 68,505 women
- Linked to downstream cardiovascular, metabolic, and mortality outcomes
- 100 ng of DNA from whole blood used per sample

Targeted Capture

The capture panel is 8.24 Mbp covering 406,000 CpGs, targeted at ~1000x depth for regions with known + novel EWAS-associated CpGs.

The design uses methylation-aware probes from Twist Biosciences to minimize off target capture.

Selection Efficiency: % On and Near Target Reads

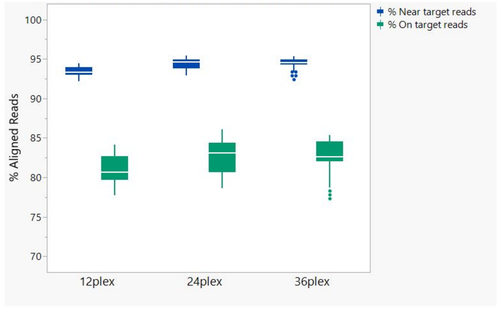


Figure 2. Percent of reads on or near target for different plexing conditions, highlighting the efficiency of the assay and the high scale enabled.

Results

Spike-in Positive and Negative Controls for Methylation Conversion

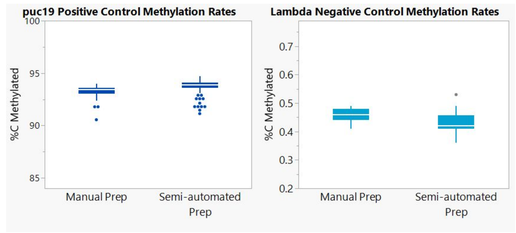


Figure 3. Conversion efficiency is monitored using spiked-in controls: a fully CpG-methylated pUC19 plasmid as a positive control, and unmethylated lambda DNA as a negative control. Positive controls showed 90–94% methylation call rates while negative controls showed 0.3–0.5% false-positive methylation calls, consistent with high specificity. Conversion rates did not differ significantly between manual and automated library preparation. N = 147 healthy donor controls and NHS2 samples.

M-Bias (Methylation Bias Plots)

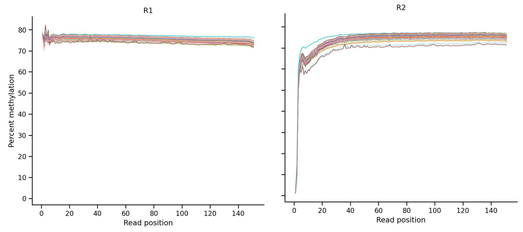


Figure 4. M-bias by read position for 95 samples, split by R1/R2. Genome-wide CpG methylation percentage plotted by position along the read for each sample. The drop at the start of R2 reflects nucleotides introduced during end-repair and adapter ligation.

Low-Pass WGS: SNP Calling and PRS Correlation

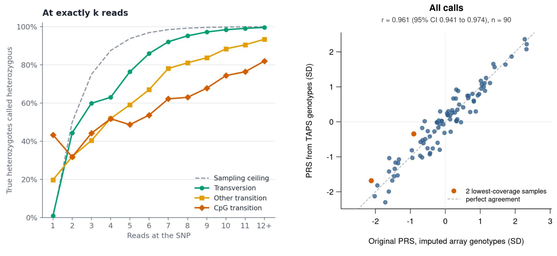


Figure 5. Left: Depth-dependent sensitivity for heterozygous variant calling, by mutation class. Right: Even at the measured depth, PTSD polygenic risk score from TAPS genotypes correlates strongly with the original score from imputed array genotypes. N = 95 NHS2 samples

Results

Sensitivity and Reproducibility

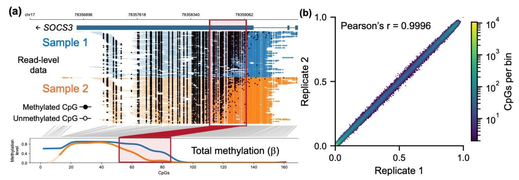


Figure 6. Our platform achieves high coverage and excellent assay-to-assay reproducibility. Preliminary Data show sensitive detection of interindividual DNAm differences (a), while technical replicates have high DNAm correlation (b).

Known BMI biomarkers across multiplexing levels

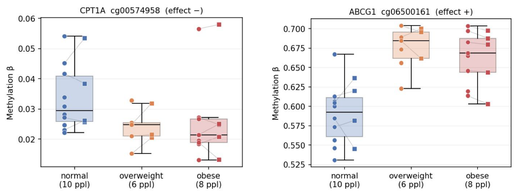


Figure 7. Verification of established obesity-associated methylation markers across multiplexing levels. Methylation rates at two well-characterized CpGs in NHS2 pilot samples stratified by BMI category. For each sample, paired points show the same library captured at 36-plex (circle) and 24-plex (square), demonstrating consistent methylation calls across multiplexing levels.

Conclusion

Population-scale EWAS has been limited by costly, damaging, or low-throughput methylation assays. Our platform, built on Watchmaker TAPS+ chemistry and methylation-aware capture, overcomes these barriers: automation-compatible, highly reproducible, and scalable.

In pilot NHS2 samples, it recovers established obesity-methylation signals and supports genotype calling accurate enough to match conventional array-based PRS — demonstrating true dual-omic utility from one low-input assay. This lays the foundation for population-scale studies, including the Biology of Adversity Project, to decode trauma's epigenetic signatures.

Contact

Broad Clinical Labs | BCLservices@broadinstitute.org
Michelle Cipicchio | michellec@broadinstitute.org