KilnBio connects assay processing, multiomics integration and biological validation

Functional & Integrative Genomics
Varun B. Dwaraka and Michael J. Corley
Corresponding author: Varun B. Dwaraka, Independent researcher, Brooklyn, New York, USA
varun.dwaraka@gmail.com

Online Resource 5

Haplotype-specific nanopore methylation, five- and six-letter cytosine imports and targeted sequencing coverage.

Haplotype-specific methylation at SNRPN

I tested whether averaging methylation across chromosome copies obscures a known signal. I used the HG002 human reference specimen and the same imprinted locus previously illustrated in Oxford Nanopore's public five-base data release. This is a reproduction in the original donor.

Reads, genotypes and comparison

I extracted GRCh38 chr15:24,948,000–24,962,000 (1-based inclusive) from the deposited Bonito/Remora 5mC-only aligned reads and companion Clair3 VCF. Of 144 primary mapped reads, 60 carried HP1, 69 carried HP2 and 15 lacked a haplotype assignment. Of these reads, 143 carried a C+m modification tag. I used the deposited haplotags, processed each group with KilnBio, SAMtools 1.24 and modkit 0.6.4, and imported the resulting CpG counts. The modification probability threshold was 0.80.

Pooled and haplotype-specific counts

On the same 229 CpG dyads with at least five valid calls per haplotype, pooled 5mC was 55.4% (9,161/16,526), including unassigned reads. HP1 was 37.0% (2,923/7,901) and HP2 was 74.1% (5,775/7,796). These are call-weighted fractions. The same 229 sites qualified at thresholds of three and ten. All sites and reads describe measurements within one donor.

Of seven PASS biallelic SNP records, five passed genotype depth ≥10 and quality ≥20. Two were heterozygous. At chr15:24,958,408 C>T, 34 HP1 reads supported C and none supported T. Among HP2 reads, 22 supported T and one C (mapping and base quality ≥20). This anchors the haplotype labels to a local allele.

Contribution and source

Starting from deposited alignments, modification calls and haplotags, KilnBio recovered the published haplotype difference and linked the methylation groups to local allele evidence.

Wright C (2022). Phased CpG methylation calling in GM24385 with Remora and Clair3. Source files and original SNRPN example.

Five- and six-letter count import

I used two public NA12878 libraries from the method-development dataset GSE208549. The comparison asks whether KilnBio preserves modification numerators and valid-call denominators when the study's original reports are converted to the supported count schema.

Input and exclusions

I restricted each report to the first 100,000 bases of GRCh38 chromosome 1. I retained positive-depth strand records only when the reported categories summed to the total. I checked CpG context against the reference, converted the legacy columns explicitly and combined complementary strands at the same CpG dyad. I checked imported counts against direct sums of the source categories.

Five-letter data

In GSM6348303, 325 of 2,202 strand records qualified. Eight had unresolved counts and 1,869 had zero coverage. The 192 resulting CpGs had 822 combined modified-cytosine calls among 1,233 valid calls. The five-letter chemistry reports 5mC and 5hmC together.

Six-letter data

In GSM6355491, 43 of 45 strand records qualified and two had unresolved counts. Across 33 CpGs, there were 25 5mC calls and zero 5hmC calls among 51 valid calls. Both channel counts agreed with the study records, preserving the zero observed 5hmC channel in this window.

Contribution and sources

The conversion preserved the modification counts and denominators in the deposited biological reports. Separate synthetic fixtures exercised direct import of contemporary, unmodified commercial duet exports. Together these checks verify the respective study-specific conversion and supported export interfaces.

Füllgrabe J et al. (2023). Simultaneous sequencing of genetic and epigenetic bases in DNA. Nature Biotechnology 41, 1457–1464.

GEO GSE208549. Five-letter GSM6348303 and six-letter GSM6355491.

Target eligibility and depth in capture panels

I tested whether a researcher can distinguish a feature omitted by a capture design from a target with inadequate measured depth. I supplied each kit's covered-target BED, kept probe intervals separate and merged overlapping or adjacent intervals before counting bases.

Human exome sequencing

DRR635838 is a standard exome library from HG002 prepared with Twist Exome 2.0 plus Comprehensive Exome spike-in. I aligned the first 10,000 complete read pairs against the full GRCh38 no-alt reference with Bowtie 2 v2.4.4, then sorted with SAMtools 1.24. Of 20,000 reads, 19,381 mapped. I evaluated only the manufacturer's Exome 2.0.2 core targets. I used minimum mapping and base qualities of 20 and counted overlapping mates once. Depth was calculated before duplicate marking.

Of 36,458,262 unpadded core target bases, 807,103 had positive depth and 35,651,159 had zero depth. None reached 10× in this small prefix. A separate bedtools intersection of genome-wide SAMtools depth gave the same 807,103 covered positions and 817,445 summed base observations. This agreement verifies the target-depth arithmetic on the sampled reads.

Human targeted methylation

GSM8643432 used EM-seq and Twist Human Methylome capture. I compared all 572 deposited bedGraph intervals in GRCh38 chr1:0–100,000 (0-based, half-open) with the manufacturer's target BED, explicitly translating chromosome label 1 to chr1. Of these intervals, 118 were wholly on target and 454 fell outside or crossed a boundary. A separate bedtools intersection gave the same count. The comparison verifies target-interval classification for the deposited methylation percentages.

Sources and implementation

Extended whole-exome sequencing study (2025), Journal of Human Genetics. PRJDB20167, standard exome DRR635838.

GEO GSM8643432. Library preparation, reference and deposited methylation bedGraph.

Twist Exome 2.0 target BED files. Twist Human Methylome covered-target BED.

KilnBio v0.2.0-research.2 for Nanopore, genotype and cytosine imports. KilnBio v0.2.0-research.3 for the capture-panel command, with targets, reference and measured files identified by SHA-256.