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 1

Held-out chromatin-accessibility caller comparison and restriction-site checks in bisulfite sequencing reads.

Online Resource 1

Assay-specific validation of chromatin accessibility and bisulfite reads

Supplement to KilnBio connects assay processing, multiomics integration and biological validation

HepG2 peak callers compared with independent DNase data

I compared MACS2 2.2.7.1 and MACS3 3.1.0 on the same 11,472,997 paired fragments per HepG2 ATAC-seq culture (ENCSR291GJU). Chromosomes 1, 19 and 21 were specified before comparison. In the first replicate (ENCFF990VCP), both callers returned 14,632 unique intervals. In the second (ENCFF624SON), both returned 14,244. Within each replicate the native narrowPeak files were byte-identical.

The independent DNase comparator contained 23,703 replicated positives among 3,152,099 eligible 100 bp bins. I matched genomic coverage within each culture and chromosome. Across callers this yielded 58,385 bins in culture 1 and 56,108 in culture 2. At equal genomic coverage, HMMRATAC gave higher mean agreement with independent DNase data than either MACS caller (F1 0.415 versus 0.338), but lower overlap between ATAC cultures (Jaccard 0.776 versus 0.873). The 0.097 loss exceeded the prespecified 0.020 tolerance. The frozen decision rule therefore retained the existing default in HepG2. Both HMMRATAC fits used MACS3 hmmratac. The first culture reached the fragment-model iteration cap. The held-out cell line tests portability to HepG2.

Native narrowPeak SHA-256, replicate 1 then replicate 2

fd4437304378669fd8d85e80c6b723fe1c9635c33c67cc1f41953ff6c9d3af0d

bcf062c9f32d47f5a777c055fbf230461e17dd93be2a74fba38d21918a8f04ee

Restriction-site enrichment in one RRBS preparation

The YB5 preparation GSM2813724 came from GSE104998 after treatment with 10 μM HH1 for four days. From each of two deposited runs I examined one million untrimmed reads and counted the reads starting with CGG or TGG, the expected MspI-derived prefixes.

Run	Reads	CGG	TGG	Either	Percent
SRR6175605	1,000,000	407,610	182,956	590,566	59.06%
SRR6175606	1,000,000	411,163	184,291	595,454	59.55%

Pooling equal read totals gives 1,186,020 of 2,000,000 reads (59.30%). The two runs sample one biological preparation. The check quantifies restriction-site composition in its deposited reads.

Public sources

ENCODE HepG2 ATAC ENCSR291GJU and DNase ENCSR149XIL. ENCODE Project Consortium et al. Nature (2020). doi:10.1038/s41586-020-2493-4.

Zhang H et al. Cell (2018). doi:10.1016/j.cell.2018.09.051. GEO GSM2813724, runs SRR6175605 and SRR6175606.