Sequencing Coverage and Quality Statistics

The Editors of IJC request that (epi)genomic next generation sequencing (NGS) data should be uploaded to an appropriate (restricted access) public data repository for public release upon publication (e.g., GEO, EGA, dbGAP). In addition, the authors must perform a quality control assessment of the data and provide a detailed summary of the sequencing coverage and quality statistics.
Library preparation, sequencing technology information (e.g., platform, read length, paired-end/single read, etc.) as well as preprocessing, quality control and filtering of the raw NGS data should be described in detail in the (Supplementary) Materials and Methods. The sequencing coverage and quality statistics of each sample should be summarized in a Supplementary Table, to which should be clearly referred in the main text. The minimum information that should be included in this Supplementary Table is for different methods listed below.

Bulk sequencing methods

Whole exome sequencing (WES) /whole genome sequencing (WGS)

Sample ID	Total number of sequenced reads	Total number of uniquely mapped non-duplicate reads[a]	Total number of covered bases[b]	Median coverage (and range) per base[b]	Percentage of targeted bases with coverage ≥10[b,c,d]

[a]Specify in table description or legend which reference genome was used (e.g., GRCh38).
[b]After removing unmapped, non-uniquely mapped and duplicate reads.
[c]Define "targeted bases" in table description or legend (e.g., whole genome, whole exome).
[d]A higher minimum coverage threshold is permitted.

Reduced representation bisulfite sequencing (RRBS) / whole genome bisulfite sequencing (WGBS)

Sample ID	Total number of sequenced reads	Total number of uniquely mapped non-duplicate reads[a]	Total number of covered CpGs[b]	Median coverage (and range) per CpG[b]	Total number of CpGs with coverage ≥5[b,c]

[a]Specify in table description or legend which reference genome was used (e.g., GRCh38).
[b]After removing unmapped, non-uniquely mapped and duplicate reads.
[c]A higher minimum coverage threshold is permitted.

RNA sequencing (RNA-seq)

Sample ID	Total number of sequenced reads	Total number of uniquely mapped reads[a]	RNA integrity number (RIN)	Ratio of all reads aligned to rRNA regions to total uniquely mapped reads (rRNA rate)	Ratio of exon-mapped reads to total uniquely mapped reads (Expression Profile Efficiency)	Total number of detected transcripts with reads ≥1[b]

[a]Specify in table description or legend which reference genome was used (e.g., GRCh38).
[b]A higher minimum coverage threshold is permitted.

ChIP sequencing (ChIP-seq)[a] and similar enrichment-based genomics NGS data (e.g., ATAC-seq)

Sample ID	Total number of sequenced reads	Total number of uniquely mapped non-duplicate reads[b]	Strand cross-correlation coefficient[c,d]	Total number of peaks[d]	Fraction of reads in peaks (FRiP)[d]	Irreproducibility discovery rate (IDR)[d]

[a]Alternatively, for ChIP-seq authors can also use the QC report of the NGS-QC generator (www.ngs-qc.org).
[b]Specify in table description or legend which reference genome was used (e.g., GRCh38).
[c]Please define whether the normalized strand cross-correlation coefficient (NSC) or the relative strand cross-correlation coefficient (RSC) is listed. See for details https://genome.ucsc.edu/ENCODE/qualityMetrics.html and https://www.encodeproject.org/data-standards/terms/.
[d]After removing unmapped, non-uniquely mapped and duplicate reads.

High-throughput chromatin conformation (Hi-C)

Sample ID	Total number of sequenced reads	Total number of uniquely mapped di-tags[a]	Total number of uniquely mapped valid di-tags	Total number of uniquely mapped valid non-duplicated di-tags	Percentage of intra-chromosomal (*cis*) di-tags[b]	Percentage of inter-chromosomal (*trans*) di-tags[b]	Hi-C reproducibility score[b,c]

[a]Specify in table description or legend which reference genome was used (e.g., GRCh38).
[b]Calculated based on the uniquely mapped valid non-duplicated di-tags.
[c]Please report one of the following Hi-C reproducibility scores: HiC-Spector, GenomeDISCO, QuASAR-Rep, or HiCRep. See for details Yardımcı et al., (Genome Biology, 2019)(PMID:30890172).

Capture high-throughput chromatin conformation (CHi-C) (e.g., promoter CHi-C)

Sample ID	Total number of sequenced reads	Total number of uniquely mapped di-tags[a]	Total number of uniquely mapped valid di-tags	Total number of uniquely mapped valid non-duplicated di-tags	Total number of uniquely mapped valid non-duplicated captured di-tags	Capture efficiency (%)	Percentage of intra-chromosomal (*cis*) di-tags[b]	Percentage of inter-chromosomal (*trans*) di-tags[b]	Hi-C reproducibility scor[b,c]

[a]Specify in table description or legend which reference genome was used (e.g., GRCh38).
[b]Calculated based on the uniquely mapped valid non-duplicated di-tags.
[c]Please report one of the following Hi-C reproducibility scores: HiC-Spector, GenomeDISCO, QuASAR-Rep, or HiCRep. See for details Yardımcı et al., (Genome Biology, 2019)(PMID:30890172).

Single-cell sequencing methods

Single-cell DNA sequencing (scDNA-seq)

Sample ID	Total number of sequenced reads per sample	Total number of called cells[a,b]	Median number (and range) of uniquely mapped reads per called cell[b]	Median percentage (and range) of bases with coverage ≥1 per called cell[b]	Median coverage (and range) per targeted base per called cell[b,c]

[a]Specify in table description or legend which reference genome was used (e.g., GRCh38).
[b]After quality control and filtering for e.g., "noisy" cells.
[c]Define "targeted bases" in table description or legend (e.g., whole genome, whole exome).

Single-cell bisulfite sequencing (scBS-seq) / single-cell reduced-representation bisulfite sequencing (scRRBS)

Sample ID	Total number of sequenced reads per sample	Total number of called cells[a,b]	Median number (and range) of uniquely mapped reads per called cell[b]	Median number (and range) of covered CpGs per called cell[b]	Median coverage (and range) of covered CpGs per called cell[b]	Median percentage (and range) of targeted CpGs with coverage ≥1 per called cell[b,c]

[a]Specify in table description or legend which reference genome was used (e.g., GRCh38).
[b]After quality control and filtering for e.g., cells with low conversion rates, insufficient coverage.
[c]Define "targeted CpGs" in table description or legend (e.g., whole genome).

Single-cell RNA sequencing (scRNA-seq)

Sample ID	Total number of sequenced reads per sample	Total number of uniquely mapped reads per sample[a]	Total number of called cells[b]	Median number (and range) of uniquely mapped reads per called cell[b]	Median rRNA rate (and range) per called cell[b]	Median number (and range) of detected genes per called cell[b]

[a]Specify in table description or legend which reference genome was used (e.g., GRCh38).
[b]After quality control and filtering for e.g., possible cell doublets, potential apoptotic cells.
rRNA rate = Ratio of all reads aligned to rRNA regions to total uniquely mapped reads.

Single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq)

Sample ID	Total number of sequenced reads per sample	Median number (and range) of uniquely mapped reads per called cell[a,b]	Total number of called cells[b]	Median fraction (and range) of reads in regions of interest per called cell[b,c]

[a]Specify in table description or legend which reference genome was used (e.g., GRCh38).
[b]After quality control and filtering for e.g., low-quality barcodes.
[c]Define "regions of interest" in table description or legend (e.g., ATAC peaks from pseudo bulk, pre-defined open regions, etc.).