biomodal software release notes duet pipeline 1.6.0 and CLI 2.1.0 August 2026

Contents

duet pipeline 1.6.0 and CLI 2.1.0

We have issued the following feature improvements and bug fixes in the current pipeline 1.6.0 release (covering changes from 1.5.0 and CLI 2.0.0):

Essential pipeline parameters

New pipeline modes for the duet +modC mosaic and duet 6-base mosaic assays

- The pipeline `mode` parameter now accepts `mosaic_5bp` and `mosaic_6bp` modes for processing data from the new **duet +modC mosaic** and **duet 6-base mosaic** assays, respectively.
- Continue to use the existing `5bp` mode for the **duet +modC** assay and `6bp` mode for the **duet evoC** assay.

Summary reports and metrics

Genome mismatch rate metric replaces lambda genetic accuracy metrics for the duet +modC mosaic and duet 6-base mosaic assays

- For the `mosaic_5bp` and `mosaic_6bp` modes used for processing data from the new **duet +modC mosaic** and **duet 6-base mosaic** assays, genetic accuracy metrics are not calculated from the lambda control. Instead, a `genome mismatch rate` metric measures the mismatch rate on `chr20` relative to a db-SNP masked reference.
- This metric is only calculated if the genome used is `GRCh38Decoy`. If the pipeline is run in targeted mode, but there are no reads on `chr20`, then no metric will be calculated.
- The chromosome used for calculating this metric can be overridden by overriding the value in the `genomes.GRCh38Decoy.genetic_accuracy_chrom` parameter.
- See also: 'Introduce GRCh38 dbSNP reference file'.

Epigenetic accuracy metrics support methylated pUC19 and unmethylated lambda for the duet +modC mosaic and duet 6-base mosaic assays

- For the `mosaic_5bp` and `mosaic_6bp` modes used for processing data from the new **duet +modC mosaic** and **duet 6-base mosaic** assays, epigenetic accuracy metrics are calculated relative to methylated pUC19 and unmethylated lambda.
- Sensitivity for 5hmC is calculated relative to a new truth set for the short 5hmC oligo that features in the **mosaic** assays.

Excel, HTML, and MultiQC summary reports for the duet +modC mosaic and duet 6-base mosaic assays

- For the `mosaic_5bp` and `mosaic_6bp` modes used for processing data from the new **duet +modC mosaic** and **duet 6-base mosaic** assays, the summary reports have been adapted to display metrics that are applicable to those assays.

Resolution algorithm

Q-table support for NovaSeqX 1.4

- New Q-tables have been introduced for compatibility with the NovaSeqX 1.4 software update from Illumina.

Q-tables updated for NextSeq2000 and Element AVITI24

- Q-tables have been updated to improve their accuracy for the allocation of resolved quality scores for the NextSeq2000 and the Element AVITI24 platforms.

Trimming and masking strategy for the duet +modC mosaic and duet 6-base mosaic assays

- For the `mosaic_5bp` and `mosaic_6bp` modes used for processing data from the new **duet +modC mosaic** and **duet 6-base mosaic** assays, a different trimming strategy is used.
- The new strategy accommodates assay-specific hairpin sequences, including removal of a poly-T sequence associated with single-stranded hairpin ligation.
- The XL tag which indicates the original fragment length, is updated after trimming.
- Furthermore, because there is no end repair step in the **mosaic** assays, no masking of C's in read tails is performed when the `mosaic_5bp` and `mosaic_6bp` modes are used.

Automatic use of q-models during quality score resolution with NovaSeqX data in `mosaic_6bp` mode

- For the `mosaic_6bp` mode used for processing data from the new **duet 6-base mosaic** assay, if the sequencer is detected as a NovaSeqX, then new q-models are used to resolve Phred scores instead of q-tables.
- The new q-models take account of additional information such as read cycle, proximity to fragment end, and proximity to a resolution error when assigning quality scores to resolved bases.
- This automatic application of q-models instead of q-tables can be deactivated by setting the `prelude.phred` parameter to `qtable`.

Methylation quantification

Automatic Q20 filtering during hmC quantification with NovaSeqX data in `mosaic_6bp` mode

- For the `mosaic_6bp` mode used for processing data from the new **duet 6-base mosaic** assay, if the sequencer is detected as a NovaSeqX, then 5hmC calls with a resolved quality score < Q20 are filtered out during methylation quantification and during generation of the 5hmC m-bias plots for MultiQC.
- This automatic filtering of 5hmC calls with resolved quality scores < Q20 can be deactivated by setting the `epiquant.min_hmc_quality` parameter to `0`.

Support for alternative pipeline modes and use cases

BAM entry point

- The duet pipeline can now be started directly from previously generated duet BAM files, enabling re-analysis of modified datasets (e.g. downsampled BAMs), and generation of updated results (e.g. Zarr stores) without reprocessing from FASTQ files.
- If both FASTQ and BAM files are present for different samples, the pipeline automatically processes FASTQs from the start and later incorporates BAM files at the appropriate stage. BAM files can also be supplied from a separate location using the `input_bam_path` parameter.

Reference files, the reference file pipeline, and support for alternative reference genomes

Introduce support for a `hg19Decoy` reference

- A `hg19` reference including decoy contigs has been introduced. This can be activated by setting the `ref_genome` parameter to `hg19Decoy`.
- This build uses the UCSC naming convention for contigs, includes unplaced and unlocalised contigs, excludes haplotype contigs, and includes decoy contigs.

Introduce a PDX combined GRCh38Decoy and GRCm38.p6 reference genome

- A new reference genome has been introduced to support PDX (patient-derived xenograft) experiments.
- The combined `PDX_GRCh38Decoy_GRCm38p6` reference genome features both human and mouse contigs and has been configured such that the mouse contigs are excluded from the list of primary contigs.
- Reads that align to the mouse contigs rather than the human contigs will get filtered out as decoy reads.

Introduce GRCh38 dbSNP reference file

- To enable the calculation of the new genome mismatch rate metric referred to above, a new GRCh38 dbSNP reference file has been introduced to the reference bundle for use with the `GRCh38Decoy` reference genome, and consequently the reference pipeline version has been updated.

Targeted mode

Target panel parameter validation improved

- Validation checks on target panel parameters have been improved so that if the `targeted` mode is activated without either selecting a target panel, or providing target reference files, then the pipeline aborts early with an error message.

Resource allocation and multi-platform support

Resource allocation for mouse reference files

- Elevated resources applied when using mouse reference genomes will now take effect when the reference genome name features `GRCm`, rather than only when the reference genome name begins with `GRCm`.