

Summary

Description: Validation checks whether the output files of a pipeline match the *expectation*.

Options:

- *validation.compareOn*
- *validation.disableValidation*
- *validation.expectationFile*
- *validation.reportOn*
- *validation.sizeWithinPercent*
- *validation.stopPipeline*

The validation utility creates a table for the output of each module where it reports the file name, size and md5. These tables are saved in the validation folder; the *validation* folder generated by a pipeline can be used as the expectations when re-running the same pipeline.

If there are no expectations, these values are reported in the *validation* folder.

If there are expectations, these values are reported *and* compared against the expected values; the result of the comparison is reported as either PASS or FAIL for each file.

If `validation.stopPipeline=Y`, the validation utility will halt the pipeline if any outputs FAIL to meet expectations, otherwise the result is reported and the pipeline moves forward.

Soft Validation

Many components of a pipeline have the potential for tiny variation: maybe a date is stored in the output, or a reported confidence level is based on a random sampling. With these tiny variations, the file is practically the same, but it will FAIL md5 validation. The file might also be a few bytes bigger or smaller, so it will also FAIL size validation. "Soft validation" is the practice of allowing some wiggle room. If the config file gives `validation.sizeWithinPercent=1`, then an output file will PASS size validation as long as it is within 1.0% of the expected file size. By default, this value is 0, and a file must be exactly the expected size to pass size validation.

Expectations

Give the file path to the expectation file using `validation.expectationFile=/path/to/saved/validation`.

This path can either point to a tab-delimited table giving the expectations for a single module, or it can point to a folder, in which case BioLockJ assumes that a file within this folder has a