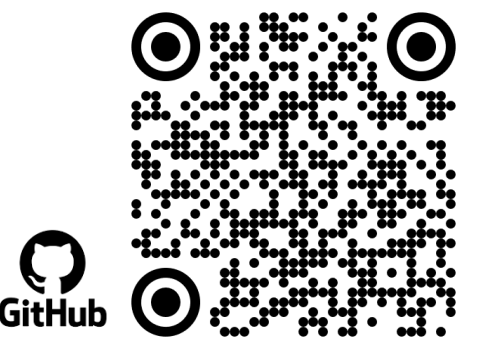


Nextflow based pipeline for association studies

Praveen Krishna Chitneedi, Jose Espinosa-Carrasco, Daniel Fischer, Christa Kuehn

¹Research Institute for Farm Animal Biology (FBN), Germany, ²Centre for Genomic Regulation (CRG), Spain, ³Natural Resources Institute Finland (Luke), Finland, ⁴Friedrich-Loeffler-Institut (FLI), Germany.

Contact: chitneedi@fbn-dummerstorf.de



- Genome wide expression association studies are often performed with huge amount of genomic and phenotype data. Handling such large data and optimizing a bioinformatic pipeline for such studies is always challenging.
- To address this issues, we developed **eQTL-Detect** pipeline with modular approach using Nextflow DSL2, which provide flexibility and ease of use.
- This modular design facilitates the confidential distribution of the computational workload among partners to pre-process the raw input data and combine the intermediate results to generate eQTL statistics.

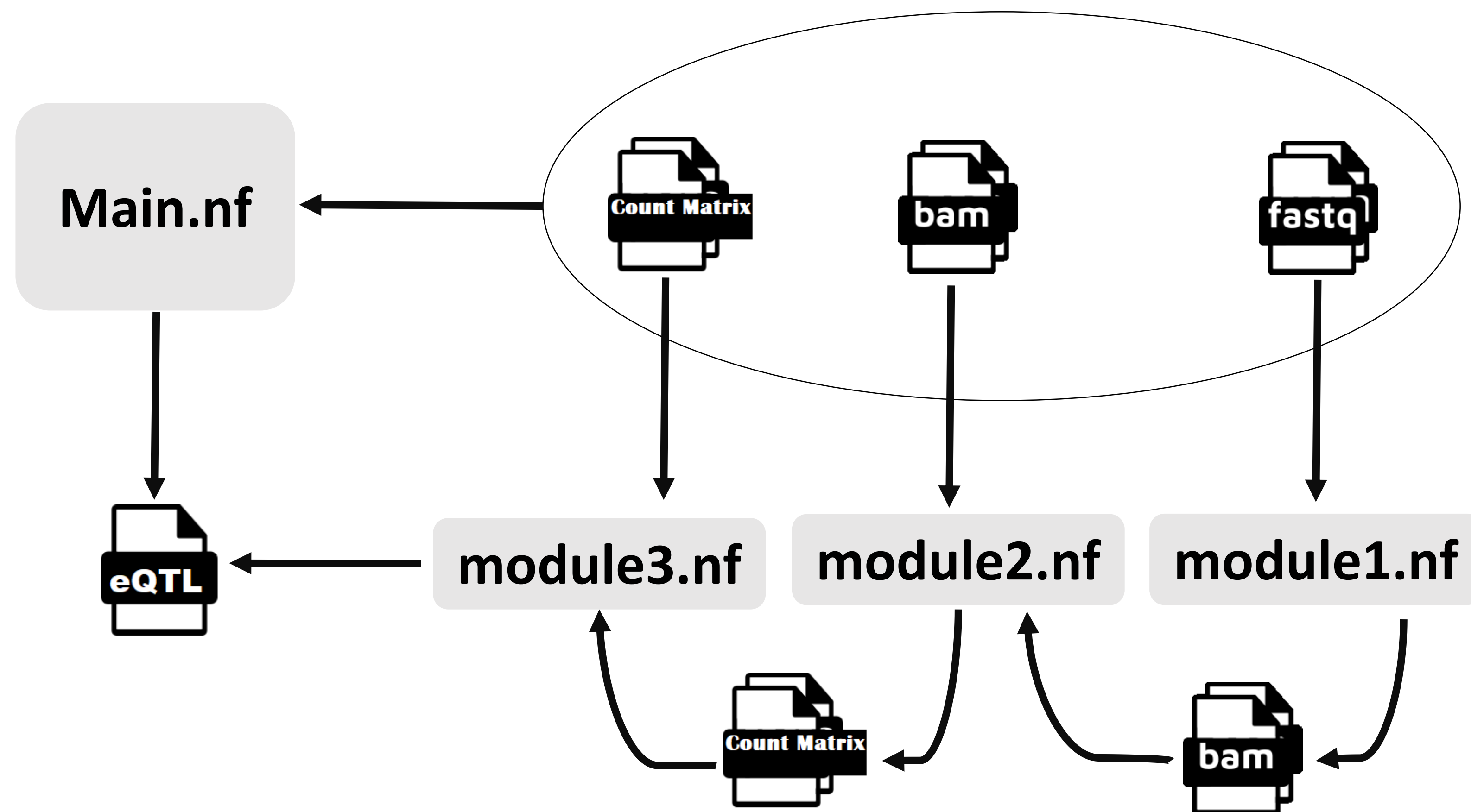


Figure 1. The flow of input expression data in different formats (fastq, bam or count matrices) after execution from different modules and from standalone script main.nf. Modules can take the expression data generated from previous modules or generated externally. The main.nf accepts any of these three formats as input and runs the entire analyses until eQTL detection.

Key features

- It offers flexibility to run on a local server and also on a computer cluster.
- All the tools used in the pipeline are containerized and it offers flexibility to choose either Docker or Singularity or Podman container technology to install all the required bioinformatic tools.
- In terms of runtime and memory management our pipeline performed similar or better than other eQTL pipelines, more details can be found in the published scientific paper.

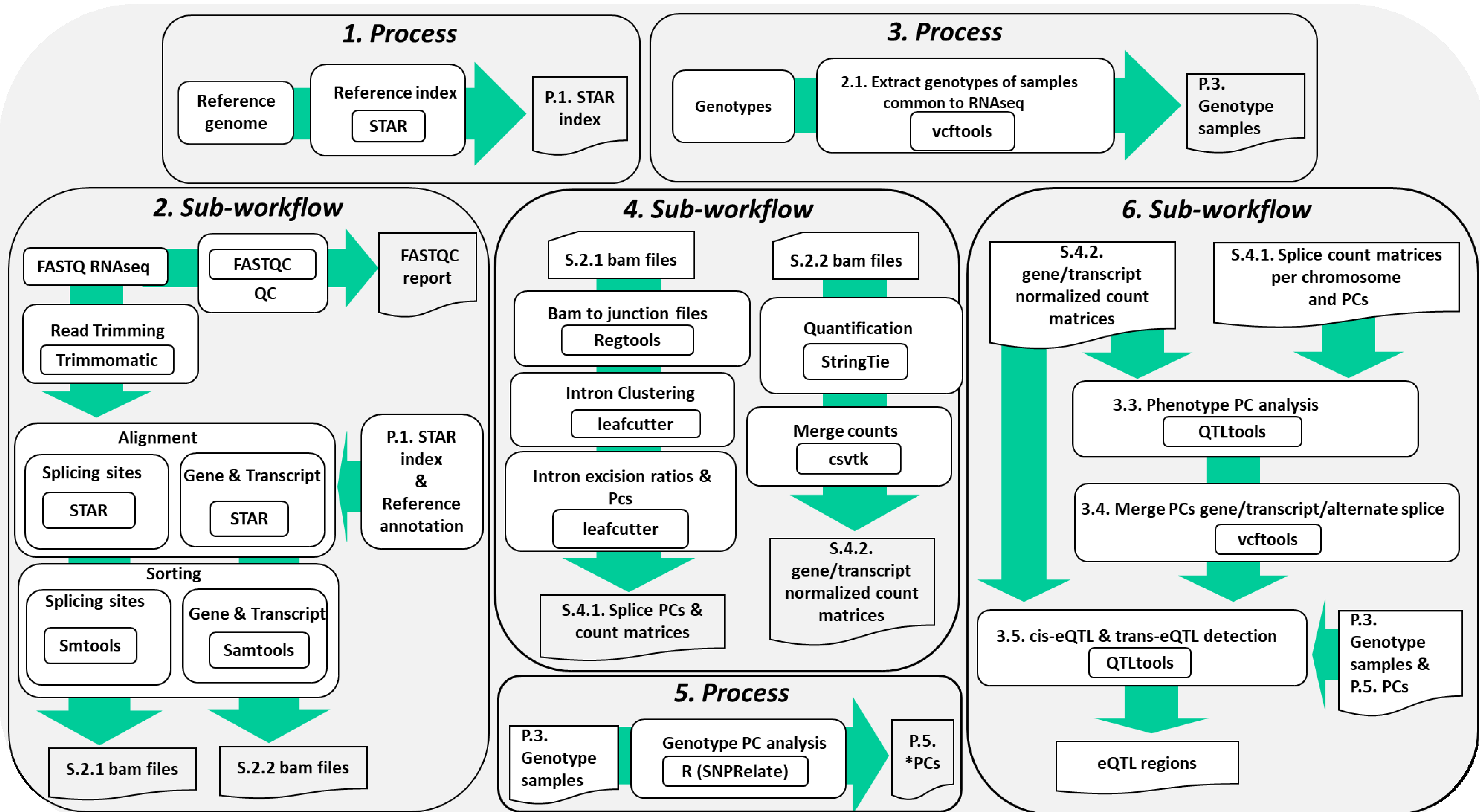
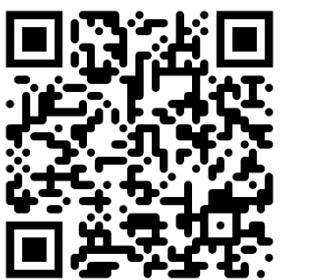


Figure 2. eQTL-Detect bioinformatics pipeline: Each block represents an independent Nextflow process or sub-workflow.