

Screw: tools and hardware for building reproducible single-cell epigenomics workflows

Kieran O'Neill^{1,2}, B Decato¹, A Goncarenko¹, A Khandekar¹, B Busby¹, and A Karsan¹

¹Pathology Department, University of British Columbia

²Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, Canada

DNA methylation is a heritable epigenetic mark that shows a strong correlation with transcriptional activity. The gold standard for detecting DNA methylation is whole genome bisulfite sequencing (WGBS). Recently, WGBS has been performed successfully on single cells (SC-WGBS). The resulting data represents a fundamental shift in the capacity to measure and interpret DNA methylation, especially in contexts where subtle cell-to-cell heterogeneity is crucial, such as in stem cells or cancer. However, SC-WGBS comes with unique technical challenges. Furthermore, although some tools have been published, and several existing studies have tended to use similar methods, no standardized pipeline for the analysis of SC-WGBS yet exists.

At the same time, within bioinformatics there has been a drive towards improved reproducibility. Textual descriptions of analyses are deeply inadequate, and often require “forensic bioinformatics” to reproduce. Exact code, accompanied by exact software versions used, is needed to recreate the exact results of a study. Common Workflow Language (CWL) provides a framework for specifying complete workflows, while Docker allows for bundling of the exact software used in an analysis within a container that can be executed anywhere. Together, these have the potential, via repositories such as Dockstore, to enable complete reproducible bioinformatics research.

Here we present Screw (Single Cell Reproducible Epigenomics Workflow). Screw is a collection of standard tools and workflows for analysing SC-WGBS data, implemented in CWL, and with an accompanying Docker image. Screw is intended to provide the parts to build fully-reproducible SC-WGBS analyses.

Tools provided include visualization of pairwise dissimilarity measures, construction of recapitulated-bulk methylomes from single cells of the same lineage, generation of bigWig methylation tracks for downstream visualization, ...

- More detail on tools
- Availability (Dockstore, Github, licenses)