

PipeRat: A high-throughput python package to perform and visualize large-scale genetic association analysis

Thiago Missfeldt Sanches¹ and Apurva S. Chitre², Oksana Poleskaya¹, Benjamin B. Johnson², Riyan Cheng, Daniel Munro^{1,3}, Khai-Minh Nguyen¹, Denghui Chen², Montana Kay Lara², Faith Okamoto (to add,gavi,sakina,elaine,faith), Abraham A. Palmer¹

¹ Bioinformatics and System Biology Program, University of California San Diego, La Jolla, CA, USA

² Department of Psychiatry, University of California San Diego, La Jolla, CA, USA

We present **PipeRat**, a versatile End-2-End Python package designed to facilitate large-scale genetic analyses, specifically tailored for genome-wide association studies (GWAS). The primary goal of the **PipeRat** is to integrate a multitude of genetic analysis tools to perform GWAS and subsequent ad-hoc analysis and provide an in-depth report of the findings with publication quality figures.

PipeRat can provide the following analysis: **1)** filtering SNPs for quality **2)** regressing out covariates, with the default of linear regression but also allowing GAMs for time-series traits and mixed-models for traits that originate from distinct sources **3)** Estimate SNP heritability and genetic correlation between traits **4)** Perform GWAS analysis **5)** Estimate P-value thresholds for the genotypes provided **6)** Identify QTLs of interest, performing conditional analysis and Bayesian posterior probabilities of inclusion (PPI) **7)** Generate Q-Qplots and volcano plots **8)** Perform PheWAS, eQTL, sQTL analysis for all the QTLs identified and generate a regional association plot **9)** Unify all the results into a single report.

PipeRat can be downloaded on GitHub <https://github.com/sanchestm/GWAS-pipeline>. All necessary installations are managed via conda for interoperability and it runs efficiently both in local computers and High-Performance Computing Clusters. We provide both a command-line interface and a Python class for experienced users. In sum, **PipeRat** represents a step forward in lowering the complexity of using genetic analysis.