

REPRODUCIBLE PIPELINE for VISUALIZING METABOLIC PATHWAYS from SINGLE-CELL TRANSCRIPTOMICS: Application to Schizophrenia

Alba Coto Escofet

Master Thesis in Bioinformatics



June 2024

Supervisors: Postdoc, Romain de Oliveira. PhD candidate, Isabel Evi Naumann.

Professor Konstantin Khodosevich

Biotech Research & Innovation Centre (BRIC)

Khodosevich Group

UNIVERSITY OF
COPENHAGEN



ABSTRACT:

This thesis presents a reproducible and modular computational pipeline. The pipeline integrates single-cell RNA sequencing (scRNA-seq) data with genome-scale metabolic modeling using the Recon3D model and Escher visualization platform.

The pipeline created in this project enables researchers to translate transcriptomic differences into functional interpretations of metabolic activity across biological conditions. As proof of concept, the pipeline is validated using a curated bacterial model (*E. coli*) and then applied to human brain single-cell data from schizophrenia (SZ) and control samples.

The brain single-cell data are raw and must therefore undergo demultiplexing and clustering with Seurat. Subsequently, differential gene expression is mapped to metabolic reactions in Recon3D, and condition-specific models are generated using Flux Balance Analysis (FBA).

The nucleotide interconversion pathway is prioritized due to its transcriptional dysregulation and relevance to neuronal signaling. Simulations revealed elevated flux through enzymes such as ADK3, PDE1, and PDE4 in schizophrenia, indicating disease-associated changes in energy metabolism and cyclic nucleotide signaling.

Importantly, the pipeline also identified reactions where gene expression and metabolic activity diverged, offering insight into the multifactorial nature of metabolic dysregulation.

This integrative approach aims to demonstrate the potential of combining scRNA-seq data with metabolic modeling to explore disease mechanisms and offer a generalizable framework for studying metabolic alterations in complex disorders.

TABLE OF CONTENTS:

1.	INTRODUCTION	5
2.	METHODOLOGY	9
2.1.	DATASET AND DATA STRUCTURE	9
2.1.1.	<i>Gene Expression and Metadata Structure</i>	10
2.2.	PREPROCESSING, DEMULTIPLEXING AND CLUSTERING WITH SEURAT	12
2.2.1.	<i>Setup</i>	12
2.2.2.	<i>HTO Barcoding and Demultiplexing</i>	13
2.2.3.	<i>Downstream analysis</i>	14
2.3.	DIFFERENTIAL EXPRESSION CALCULATION	17
2.4.	RECON3D AND ITS ROLE IN METABOLIC MODELING	17
2.4.1.	LOADING AND PREPARING METABOLIC AND TRANSCRIPTOMIC DATA	19
2.4.2.	<i>Gene Identifier Mapping for Recon3D compatibility</i>	19
2.5.	CONDITION-SPECIFIC MODEL CONSTRUCTION AND SIMULATION	20
2.5.1.	<i>Generation of Condition-Specific Metabolic Models</i>	20
2.5.2.	<i>Expression Thresholding for Model Constraints</i>	21
2.5.3.	<i>Simulating Metabolic Activity with Flux Balance Analysis (FBA)</i>	21
2.6.	ESCHER FOR VISUALIZATION OF METABOLIC PATHWAYS	23
2.6.1.	<i>Overview of Escher</i>	23
2.6.2.	<i>Integration with COBRA and COBRApy</i>	24
2.6.3.	<i>Escher's file structure: Models and Maps</i>	24
3.	RESULTS	27
3.1.	METABOLIC PATHWAY VISUALIZATION IN A BACTERIAL MODEL: <i>E. COLI</i>	27
3.1.1.	<i>Reaction Flux Visualization</i>	28
3.1.2.	<i>Gene Expression Overlay</i>	29
3.2.	KEY OUTPUT FILES FROM THE SEURAT PIPELINE	32
3.3.	SEURAT ANALYSIS OF SINGLE CELL TRANSCRIPTOMIC DATA	33
3.4.	GENE MAPPING TO THE METABOLIC MODEL	37
3.5.	PATHWAY PRIORITIZATION FOR METABOLIC ANALYSIS IN SCHIZOPHRENIA	38
3.6.	METABOLIC PATHWAY MODELING AND VISUALIZATION IN SCHIZOPHRENIA	41
3.6.1.	<i>Structural Visualization of the Nucleotide Interconversion Pathway</i>	41
3.6.2.	<i>Flux Simulation of the Nucleotide Interconversion Pathway</i>	42
3.6.3.	<i>Condition-Specific Flux Visualization Using Subnetworks</i>	47
4.	DISCUSSION	50
4.1.	VALIDATION OF THE PIPELINE USING A BACTERIAL MODEL	50
4.2.	SEURAT ANALYSIS	51
4.3.	GENE MAPPING AND INTEGRATION EFFICIENCY	52
4.4.	PATHWAY PRIORITIZATION AND JUSTIFICATION	52
4.5.	CONDITION SPECIFIC METABOLIC MODELING OF THE NUCLEOTIDE INTERCONVERSION PATHWAY	52
4.5.1.	<i>Condition-Specific Flux Simulation Using FBA</i>	53
4.5.2.	<i>Integrating Gene Expression with Flux Predictions in Escher</i>	54
4.5.3.	<i>Comparative Subnetwork Visualization of SZ and CON Conditions</i>	55
4.6.	LIMITATIONS	57
4.7.	FUTURE DIRECTIONS FOR THE PIPELINE EXPANSION	58
5.	CONCLUSIONS	59
6.	REFERENCES	60