

Evaluating Perturbation Effects Predictions in Perturb-seq: All you Need is the Right Metric

Sesili Pirtskhalava

*The thesis has been submitted to the Faculty on Natural Sciences and Medicine of Ilia State
University following the requirements for obtaining the degree of Master of science*

Program: Applied Genetics

Research Supervisor: Vincenzo Lagani, Ph.D.

Ilia State University

Tbilisi 2026

Statement

To the Dean of Faculty of Natural Sciences and Medicine LEPL Ilia State University

Ms. Elene Zhuravliova

Faculty of Natural Sciences and Medicine

Program: Applied Genetics

Master's student: Sesili Pirtskhalava

As the author of the submitted thesis, I declare that the thesis is my original work and does not contain materials previously published, accepted, or defended by other authors, which are not mentioned or cited according to the citation rules.

Sesili Pirtskhalava

15.06.2026

Abstract

Large-scale genetic perturbation experiments, particularly Perturb-seq, have enabled systematic investigation of gene function by linking CRISPR-based perturbations to single-cell transcriptomic responses. The growing availability of perturbation datasets has led to the development of computational models that aim to predict cellular responses to previously unseen perturbations, reducing the need for exhaustive experimental screening. However, recent studies have questioned whether complex models consistently outperform simple baseline approaches and whether commonly used evaluation metrics accurately capture perturbation-specific biological effects.

In this study we addressed the conflicting conclusions regarding the performance of perturb-seq prediction models by replicating and evaluating GEARS, a graph-based model for perturbation-response prediction, using Norman, Adamson, and Frangieh datasets. GEARS predictions were compared against simple average-based baselines and assessed using both standard performance metrics, including Pearson correlation and RMSE, as well as perturbation-centered metrics inspired by the Systema framework, including Pearson Δ and centroid accuracy.

GEARS occasionally showed modest evidence of capturing perturbation-specific structure through centroid accuracy, particularly in the Adamson dataset. However, Pearson Δ values were generally low and did not indicate strong recovery of perturbation-specific transcriptional effects. These findings highlight the importance of benchmarking perturbation-prediction models against simple baselines and demonstrate that the interpretation of model performance depends strongly on the evaluation metric used.

Keywords: Perturb-seq; GEARS; genetic perturbations; gene expression prediction; perturbation-response modeling.

Acknowledgment

I would like to express my sincere gratitude to my supervisor, Professor Vincenzo Lagani, Ph.D., for his continuous guidance, support, and expertise throughout the development of this thesis.

His insightful advice and scientific knowledge played a crucial role in shaping this research. His assistance with data analysis and interpretation was invaluable to the successful completion of this work. The knowledge and skills I acquired through his bioinformatics course further strengthened my understanding of the field and greatly contributed to this study.

I am also deeply appreciative of his encouragement, and willingness to provide constructive feedback at every stage of the research process. His mentorship and dedication have been a constant source of motivation, and this thesis would not have been possible without his support.