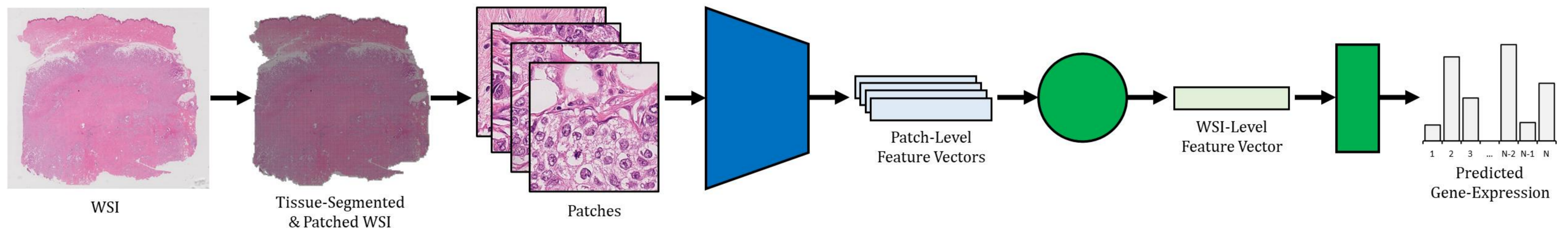


# Evaluating Deep Regression Models for WSI-Based Gene-Expression Prediction

Fredrik K. Gustafsson<sup>1</sup>, Mattias Rantalainen<sup>1,2</sup>

<sup>1</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

<sup>2</sup>MedTechLabs, BioClinicum, Karolinska University Hospital, Solna, Sweden



## Introduction

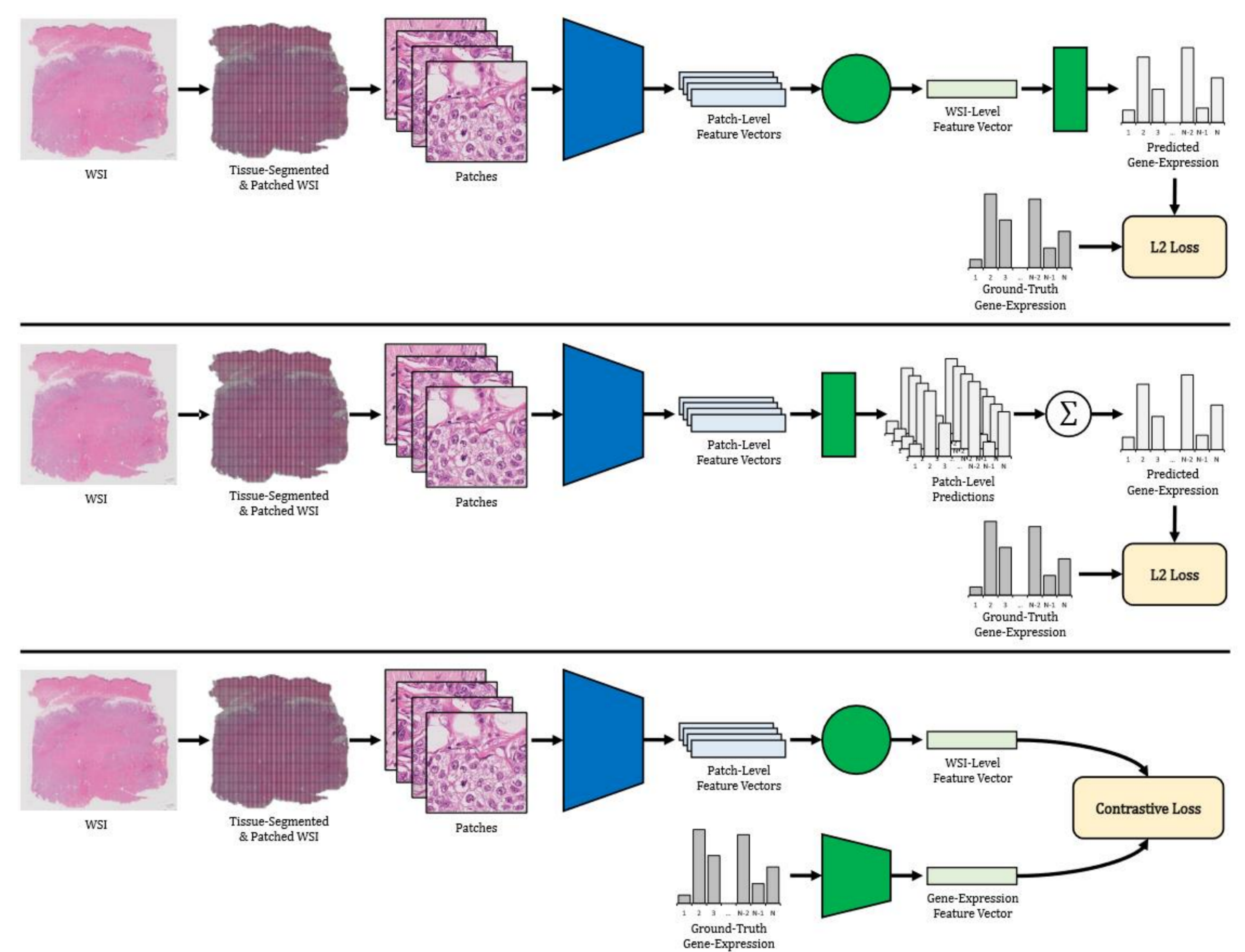
Prediction of mRNA gene-expression profiles directly from routine WSIs using deep learning models could potentially offer cost-effective and widely accessible molecular phenotyping. While such WSI-based gene-expression prediction models recently have emerged, the high-dimensional nature of the corresponding regression problem offers numerous design choices which remain to be analyzed in detail.

This study provides recommendations on how deep regression models should be trained for WSI-based gene-expression prediction.

## Main Takeaways

1. Training regression models on top of UNI features gives accurate WSI-based models for gene-expression prediction (TCGA-BRCA: 4927 genes with Pearson corr. above 0.4, mean Pearson of 0.56 for PAM50 genes).
2. Despite conceptual differences, *Direct - ABMIL* and *Contrastive* achieve very similar performance and should both be considered go-to regression models.
3. Training a single model to regress all 20530 genes is a computationally efficient and very strong baseline, this should be the starting point given any new dataset.
4. Training one model for each individual gene incurs an extremely high computational cost yet achieves comparatively low accuracy.

## Evaluated Models



Top: *Direct - ABMIL*. Middle: *Direct - Patch-Level*. Bottom: *Contrastive*.

## Results

