

# Transfer Learning with AI Improves Estimation of Genomic Variation

Niko Zuppas and Bryan C. Carstens

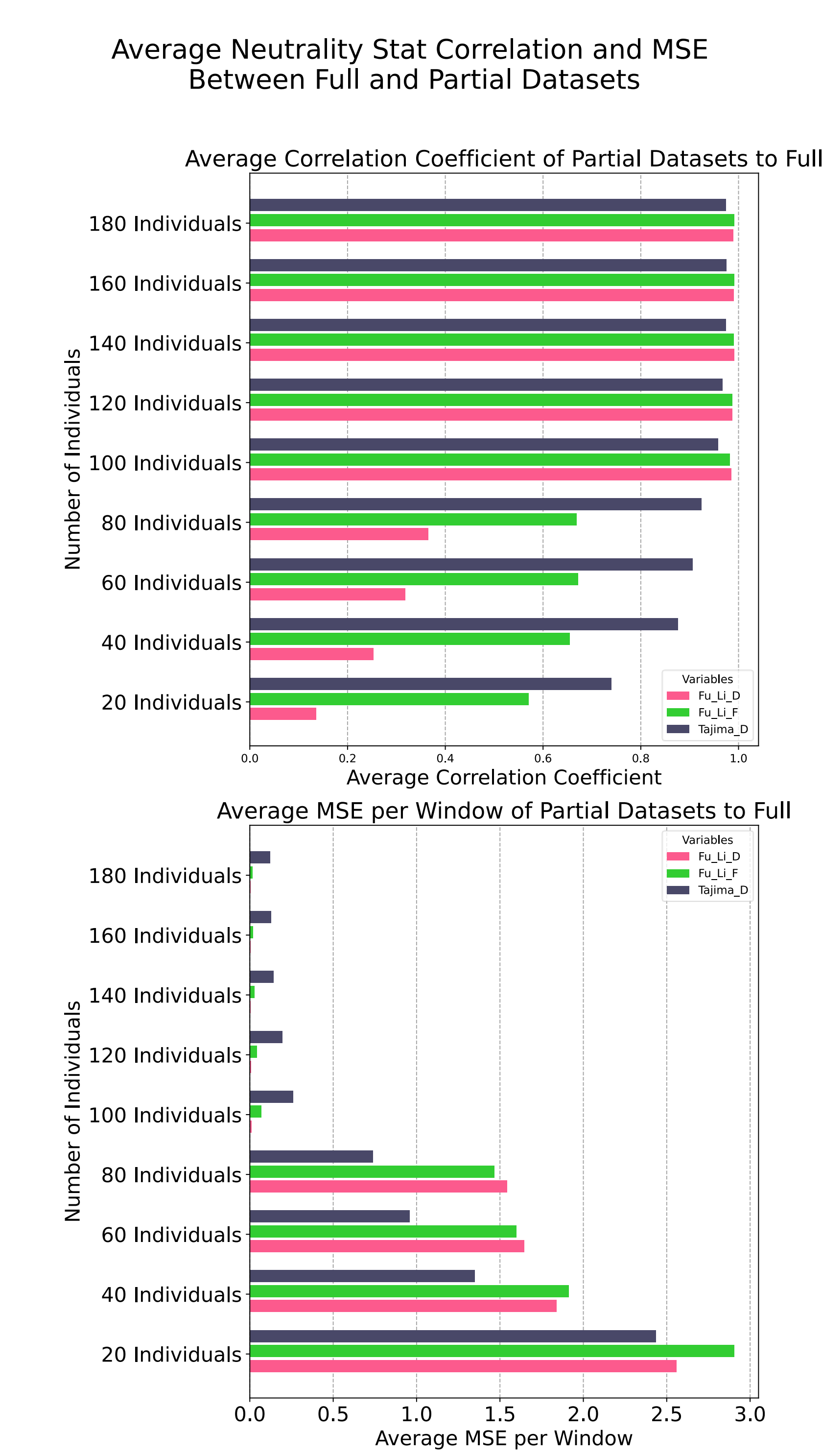
## Background:

Traditional methods of estimating neutrality are sample size dependent

Neutral evolution, parametrized by mutation rate and population size, is a consistent stochastic process among species

Generative models, like GANs<sub>1</sub>, have been used to estimate neutrality from genomic data.

We hypothesize deep learning and transfer learning will improve the estimation of neutrality in small sample sizes.



## Methods:

### Data

WGS 1000 Genomes Project data<sub>2</sub>

Two **foundational populations**, (TSI, IBS), 217 individuals

Two **transfer populations** (CHB, KHV), 20 individuals

### Preprocessing

Filter sites with measurable selective or demographic effects

### Training

Train a Variational Autoencoder<sub>3</sub> (VAE) on remaining neutral sites

### Transfer Learning

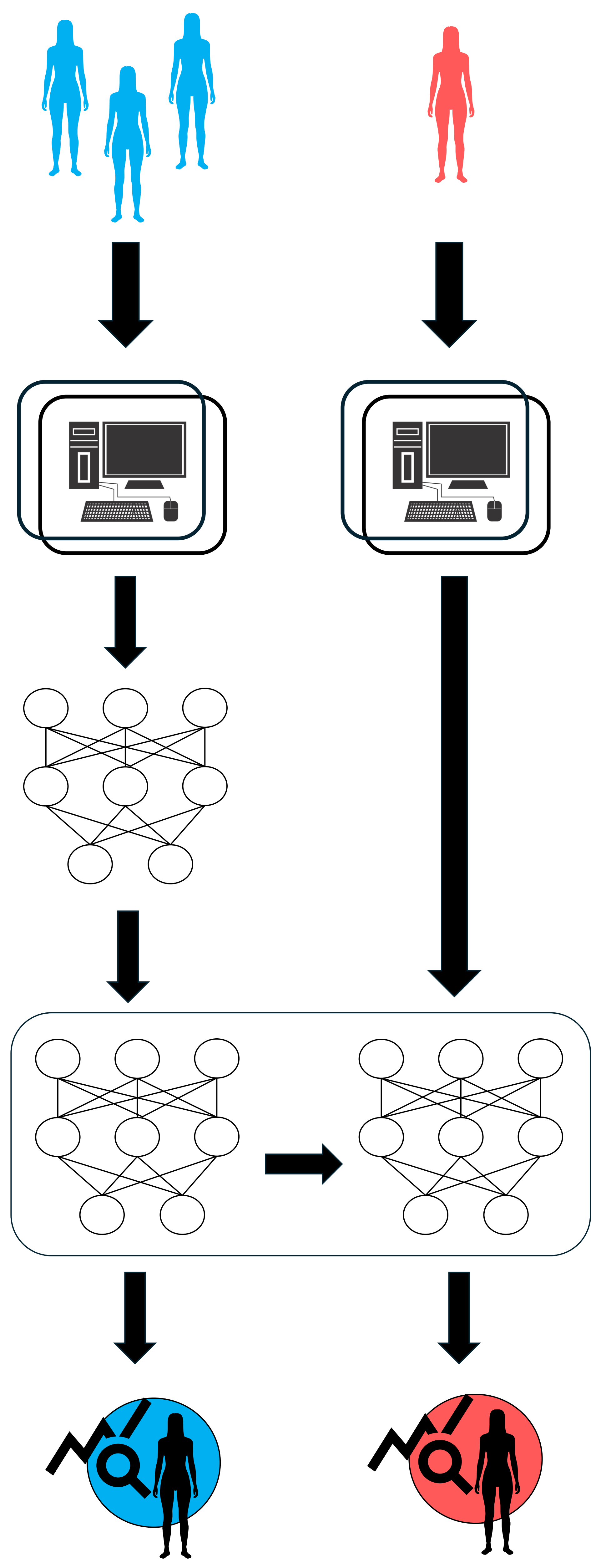
Fine-tune the model on the transfer dataset

Use Lower learning rate, fewer epochs

### Assessment

Analyze reconstruction error scores to assess performance

Use Tajima's D, Fu and Li's D and F as benchmarks for both full and partial datasets



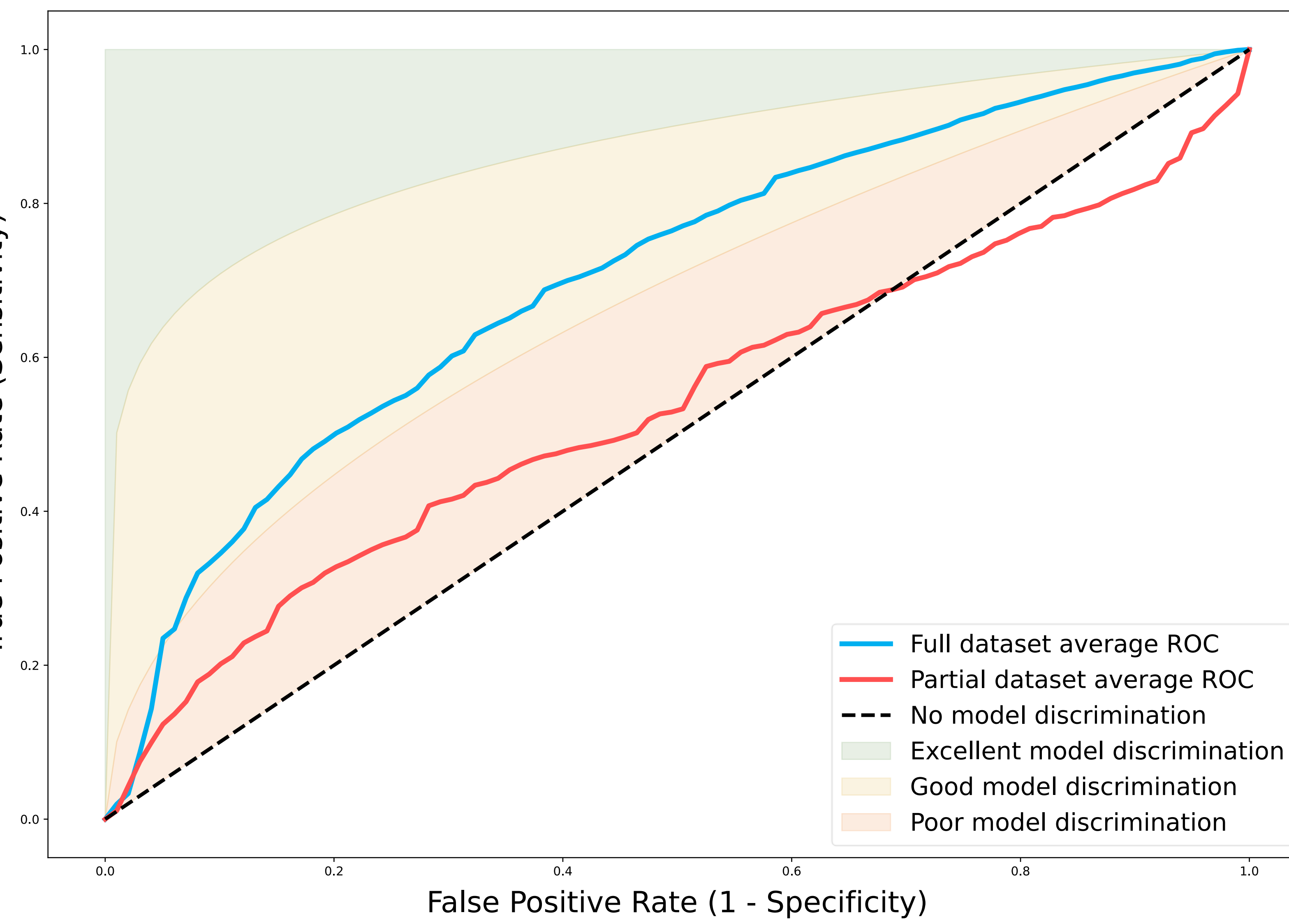
## Results/Discussion:

Transfer learning neutrality estimates were more consistent with larger sample sizes (202 individuals) than smaller ones (20 individuals).

Population genetic inference in low-sampled populations was improved with transfer learning, given sufficient training data.

Given the broad importance of estimating in population genetics analyses, transfer learning might be able to help improve inferences from the kind of data biologists are able to collect.

Transfer learning ROC Curves on Full/Partial Datasets Using Summary Stats



## Citations:

1. Riley et al. 2024, *Genetics*
2. 1000 Genomes Project Consortium 2015, *Nature*
3. Higgins et al. 2017, *ICLR*

## Acknowledgements:

We would like to thank the Ohio Supercomputer Center for essential computational resources. We also acknowledge Lisle Gibbs, Laura Kubatko and members of the Carstens Lab for feedback and support.

Niko Zuppas  
zuppas.3@osu.edu

THE OHIO STATE UNIVERSITY

Ohio Supercomputer Center

DEPARTMENT OF EVOLUTION, ECOLOGY, AND ORGANISMAL BIOLOGY