

AGEOptimizer: A Novel Multi-Technique Optimization Tool for Enhanced Model Prediction and Discovery of Cancer Biomarkers

Andrew Shin

Introduction

- Problem Statement:** Existing age-prediction models often suffer from data quality issues, limited generalizability, and biases, hindering their clinical applicability.
- Solution:** This project introduces AGEOptimizer, a novel approach designed to improve the robustness of genomic age prediction through various techniques in the data selection/preprocessing, model selection, parameters,
- Application:** The study also investigates the potential of AGEOptimizer in identifying aging-associated biomarkers in Low-Grade Glioma (LGG) tissues.

Background

- Aging is a key risk factor for chronic diseases, with biological age (BA) reflecting structural and functional aging determinants, differing from chronological age (CA)
- Patients with cancer, cardiovascular diseases, and neurodegenerative disorders often exhibit **significant discrepancies between CA and BA**.
- This divergence suggests that such conditions accelerate the aging process at the cellular and molecular levels, highlighting the need for accurate BA predictions.

Age Prediction Models

Computational tools that return a sample's age based on data patterns

Biomarkers

Measurable indicators of biological processes or conditions

Scientific & Medical Impact:

Advancement of personalized medicine, targeted disease monitoring, disease prognosis, and reduction of related risks

Hypothesis/ Research Goal

The primary goal of this project is to develop an optimized machine-learning model capable of accurately predicting biological age by analyzing tissue-specific biomarkers

(1) **Validate Model Performance:** Use a benchmark model (standard neural network) in order to mark improvements in the model. Note that the effectiveness also depends in predicting biological age across a wide diversity of tissue types

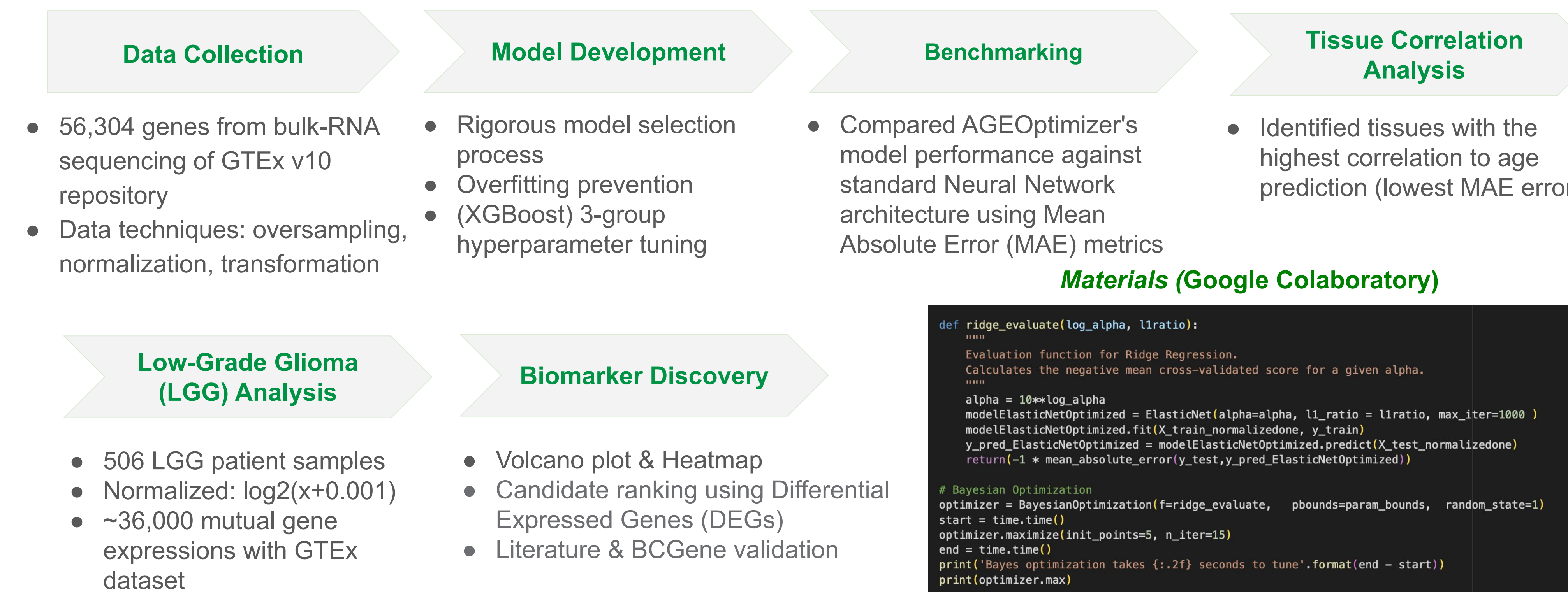
Then, the model enables additional objectives to validate its accuracy and significance

(2) Identification of optimal tissues for age prediction

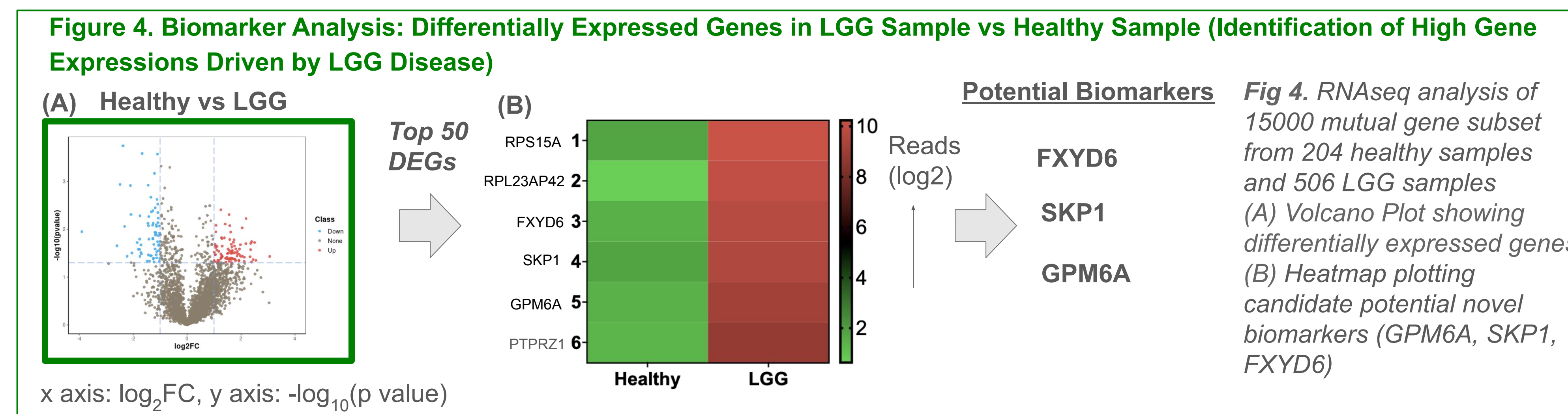
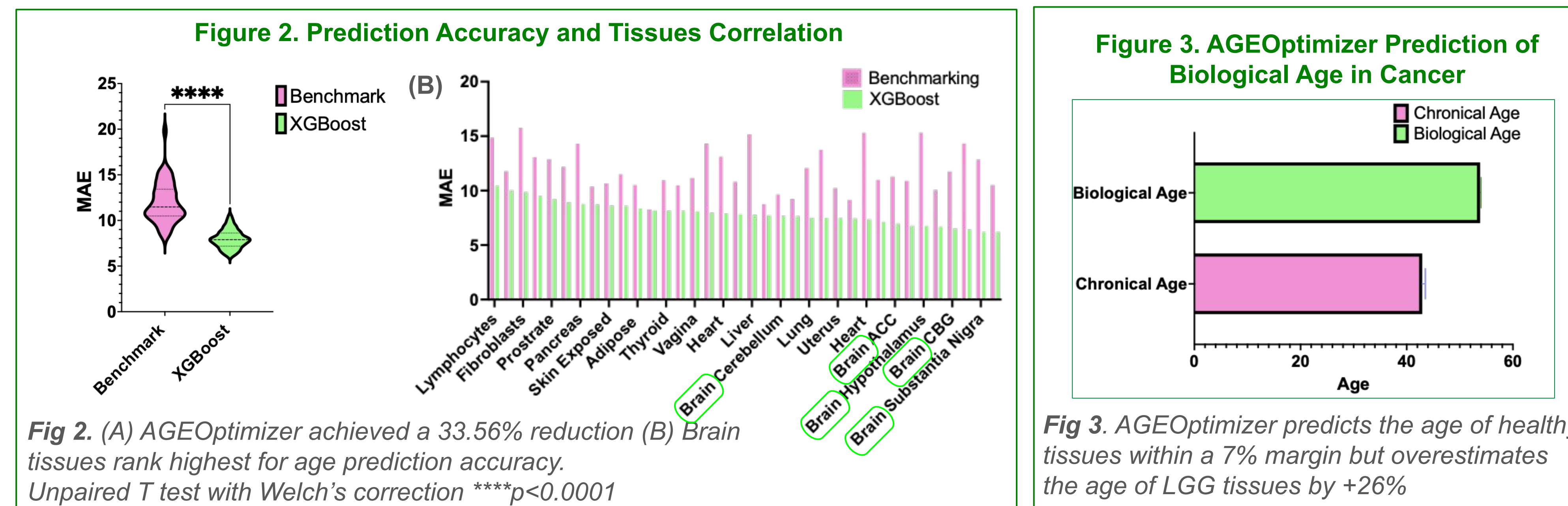
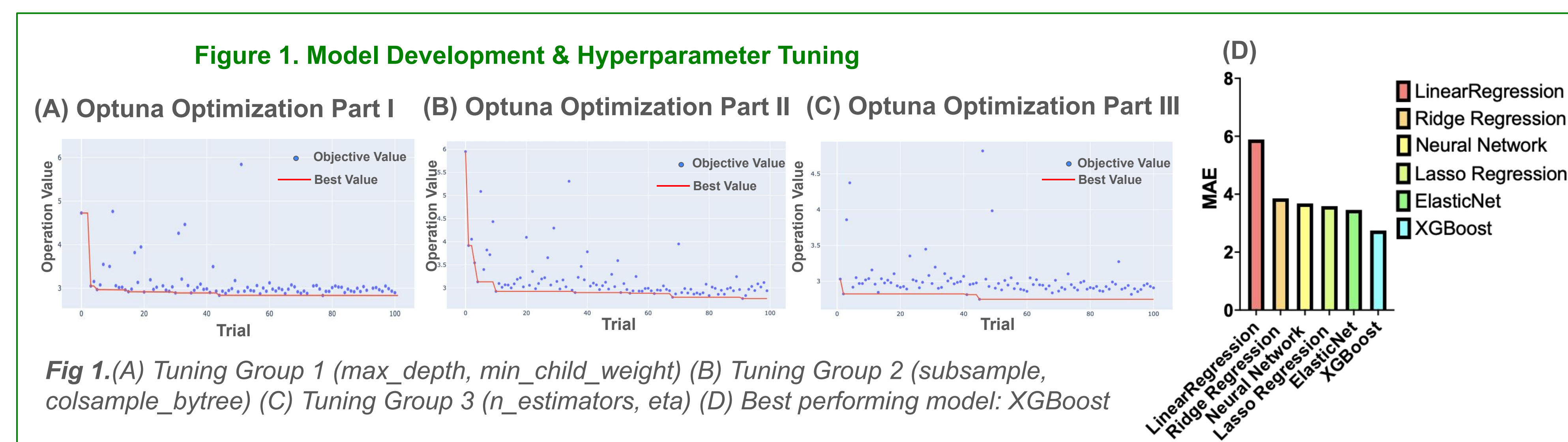
(3) Evaluation of if a disease is 'age-correlated'

(4) Confirming existing and finding new disease biomarkers

Methodology



Results



Results

- Prediction Accuracy:** AGEOptimizer achieved a **33.56%** reduction in overall MAE and a **46.17%** improvement in MAE for the best-correlating tissues compared to benchmark models. (Fig. 2a)
- Tissue Correlation:** The **Brain Amygdala**, **Brain Substantia Nigra**, and **Brain Cortex** were identified as the most correlated tissues for accurate age prediction (Fig. 2b)
- LGG Tissue Analysis:** AGEOptimizer predicted the age of healthy tissues within a 7% margin but overestimated the age of LGG tissues by **+26%**, indicating that LGG tissues exhibit transcriptomic signatures akin to older tissues (Fig. 3)
- Biomarker Discovery:** Differentially expressed gene analysis confirmed known LGG biomarkers PTPRZ1 and MT1X from our dataset. Additionally, potential new biomarkers, **GPM6A**, **SKP1**, and **FXYD6** were identified for potential future therapeutic targeting (Fig. 4)

Conclusions and Future Study

Conclusions:

- AGEOptimizer significantly enhances decreases the error of genomic age prediction and generalizes across multiple data types
- Additionally, it can help identify age-related biomarkers in LGG tissues through its ability to detect age-correlated diseases.
- Helping identify both established and novel biomarkers, AGEOptimizer reduces reliance on large disease cohorts for accurate biomarker discovery, potentially shaping future biomedical treatment strategies.

Future Study:

- Find new techniques / models to improve AGEOptimizer
- Validation of identified biomarker candidates and their biological pathways via pathway enrichment analysis
 - This analysis will explore molecular mechanisms of brain cancer, potentially allowing the found biomarkers to become novel therapeutic targets.

Reference

- Wang J, Gao Y..., Zhou Y. Proc Natl Acad Sci U S A. 2024 Jan 16;121(3);PMID: 38190540
- Wang F, Yang J, Lin H, Li Q, Ye Z, Lu Q, Chen L, Tu Z, Tian G. Front Genet. 2020 Sep 24;11:1025. PMID: 33101366
- Niccoli T, Partridge L. Curr Biol. 2012 Sep 11;22(17):R741-52. doi: 10.1016/j.cub.2012.07.024. PMID: 22975005.
- Ibrahimi E., D'Elia D, Berland M, Marcos-Zambrano L.J. Front Microbiol. 2023 Oct 5;14:1250909. doi: 10.3389/fmicb.2023.1250909. PMID: 37869650; PMCID: PMC10588656.
- Xiaolin Ni a..., Huiping Yuan Computer Methods and Programs in Biomedicine. Volume 240, October 2023, 107686.
- <https://towardsdatascience.com/xgboost-concept-explained-in-layman-terms-1d2bcdeaf12f>
- <https://www.gtportal.org/home/>
- <https://www.cancer.gov/ccg/research/genome-sequencing/tcga>