

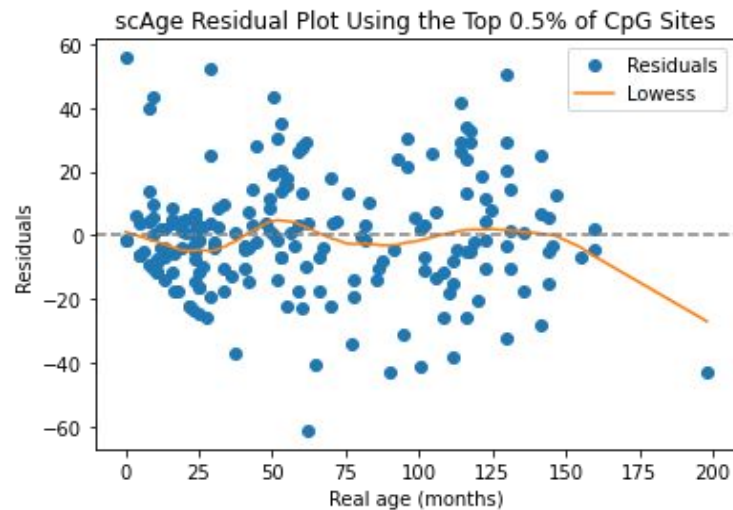
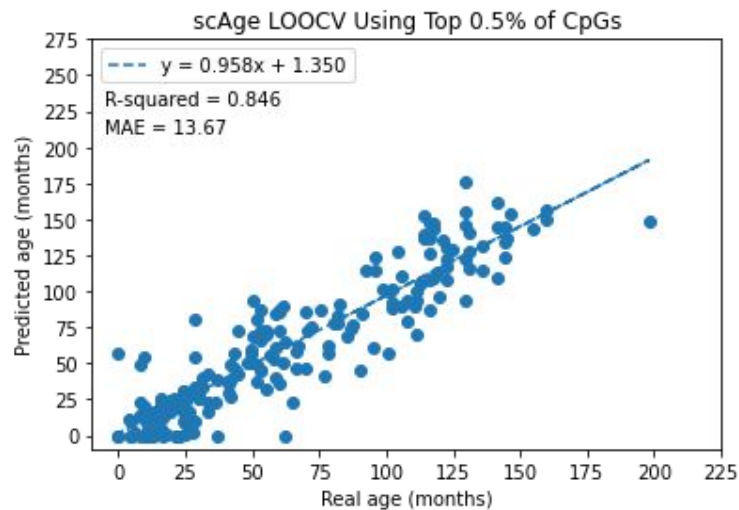
Analysis of Multiple Epigenetic Frameworks using Canine DNA Methylation Data

Jared Paul Guevara, 6/9/2023

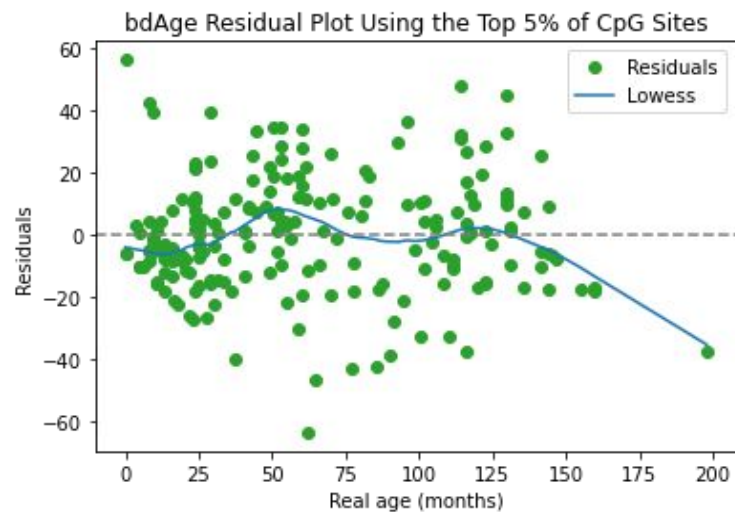
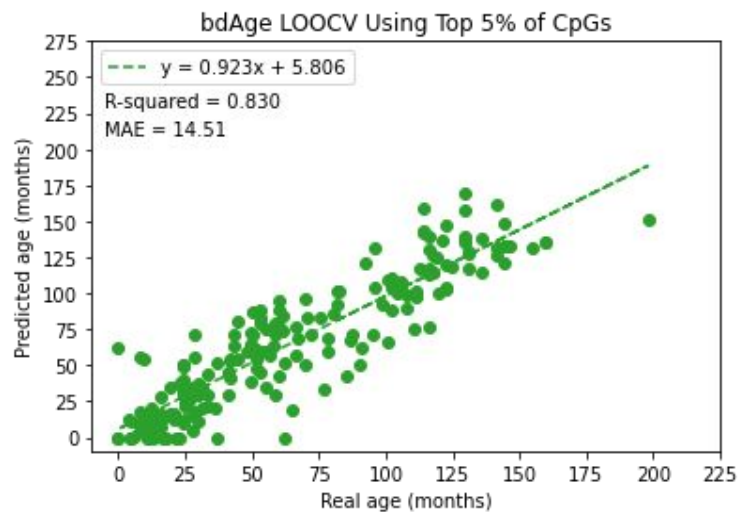
Methods

- Models were trained on 204 samples
 - Each sample containing 6,706 features
- Leave One Out Cause Validation implemented to train each sample against the remaining 203 samples
- Tested following frameworks:
 - scAge, bdAge, nbAge, LASSO regression, mle bdAge, mle nbAge
- Tested increasing coverages of CpG sites for each framework: top 0.01%, 0.1%, 0.5%, 1%, 5%, 10%, 25%, 50%, 100%
 - Of the results shown, only the coverage with the highest R^2 value was chosen and subsequently had their residuals plotted
- Tested increasing values of tau for the mleAge frameworks: 0.1, 0.5, 1.0

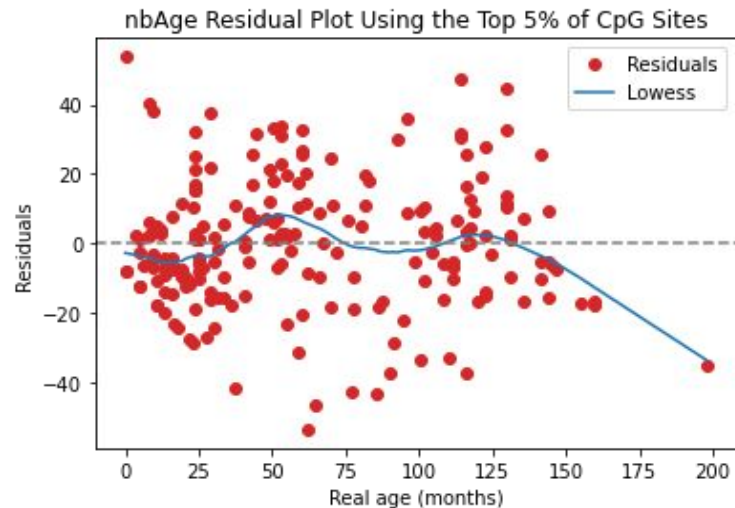
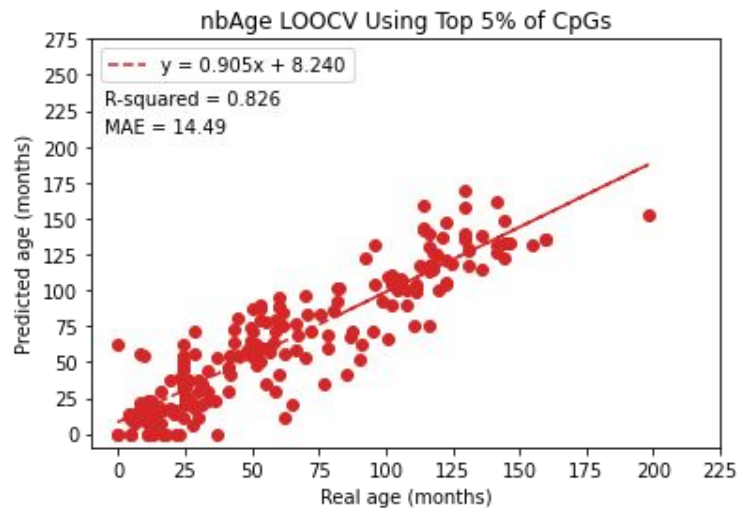
Results: scAge



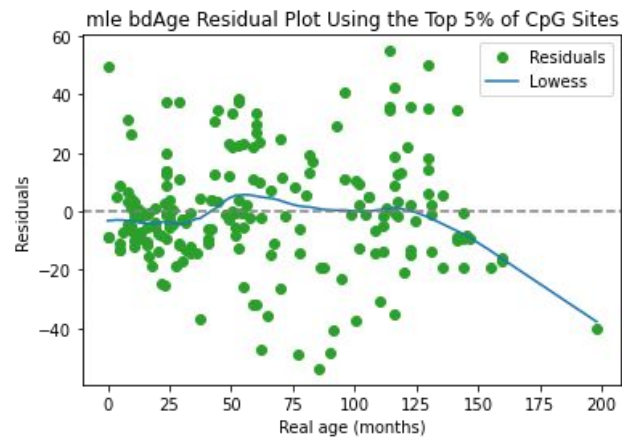
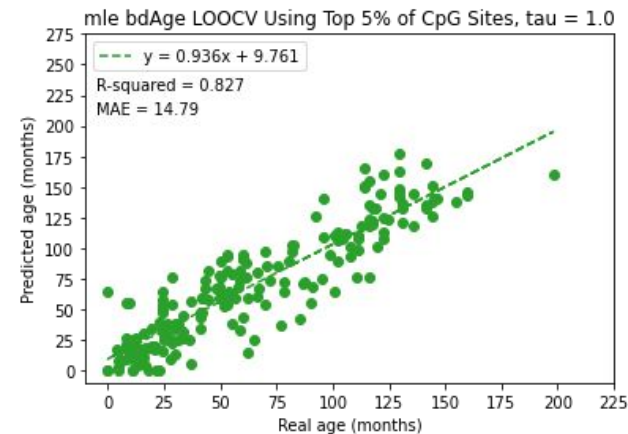
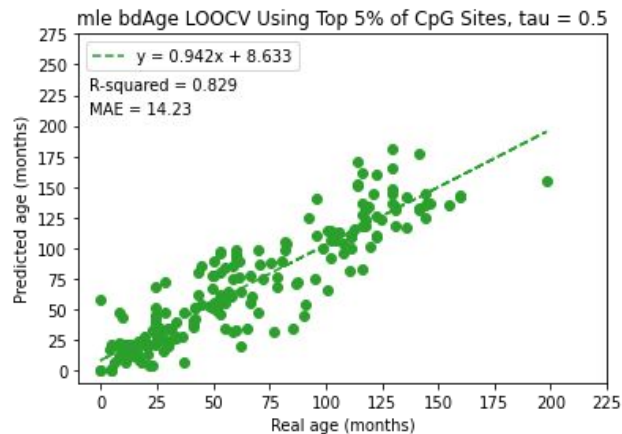
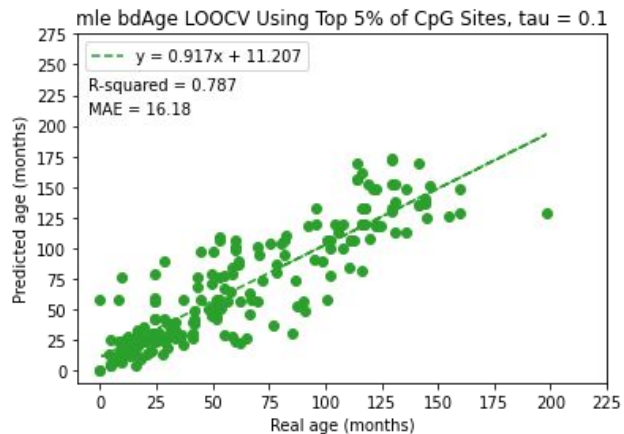
Results: bdAge



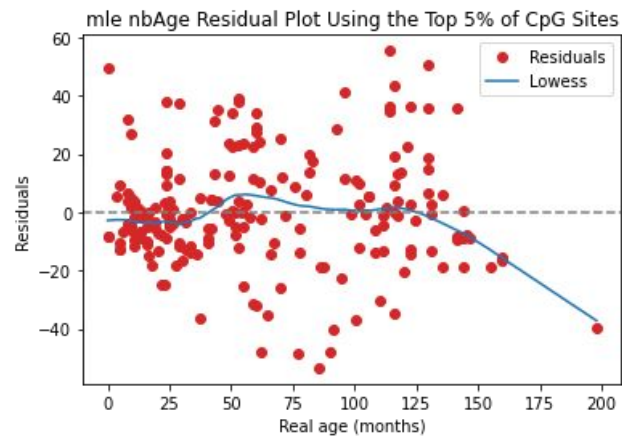
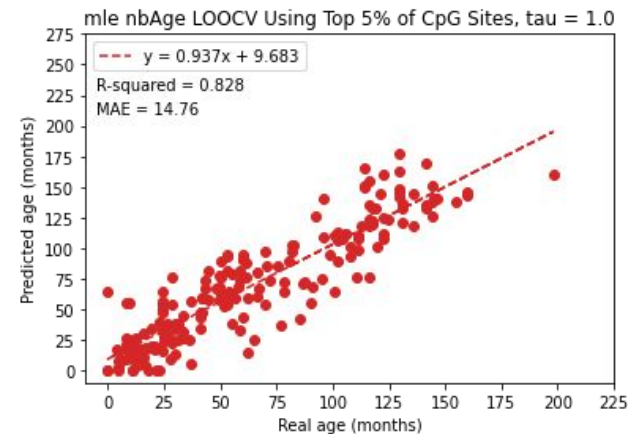
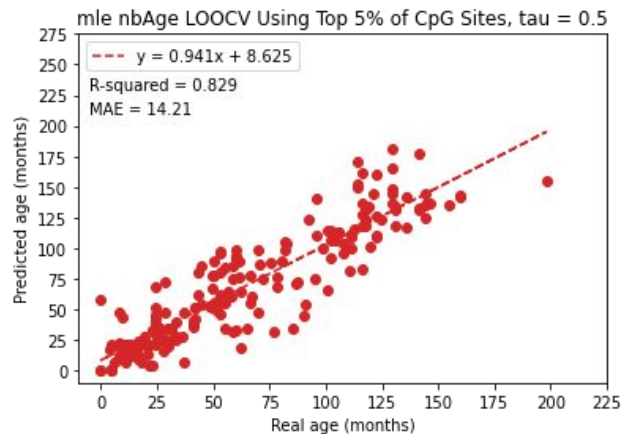
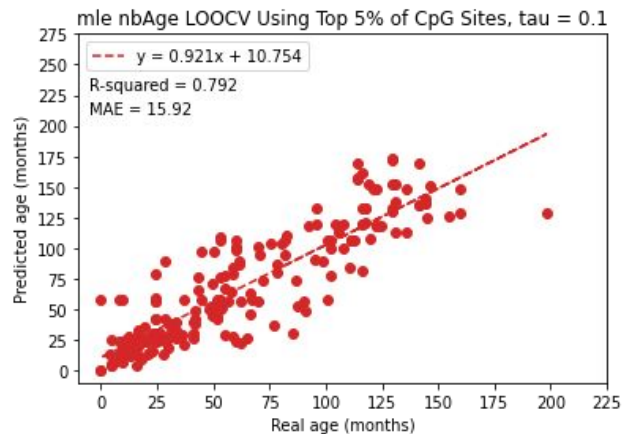
Results: nbAge



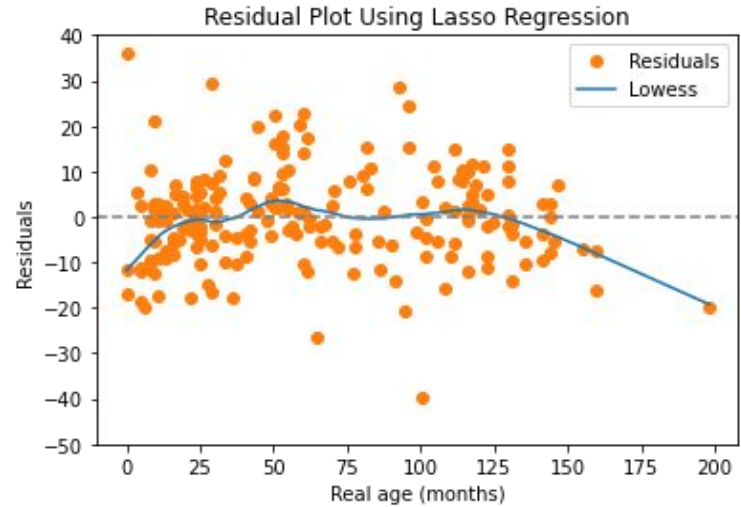
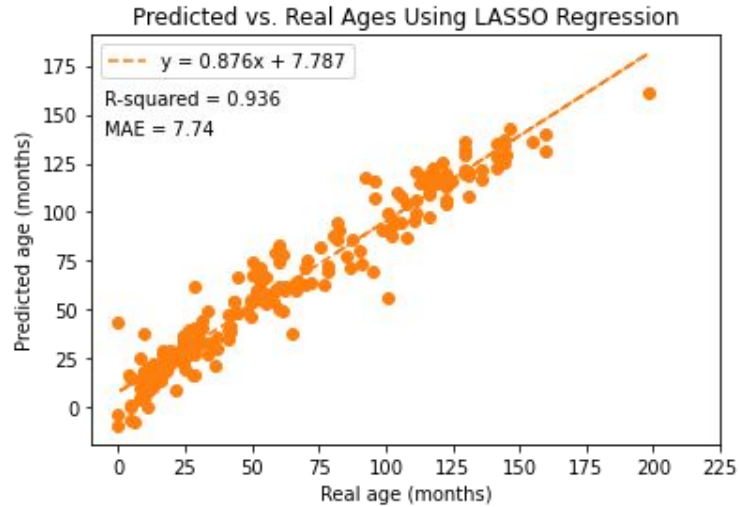
Results: mle bdAge



Results: mle nbAge



Results: LASSO



Alpha = 0.19

Conclusion

- Of all frameworks, LASSO regression fits the data set best, at 100% coverage, followed by scAge, at 0.5% coverage
- Can better optimize script to find best parameters (coverage %, tau)