Methuselah Foundation Age Prediction

Updates

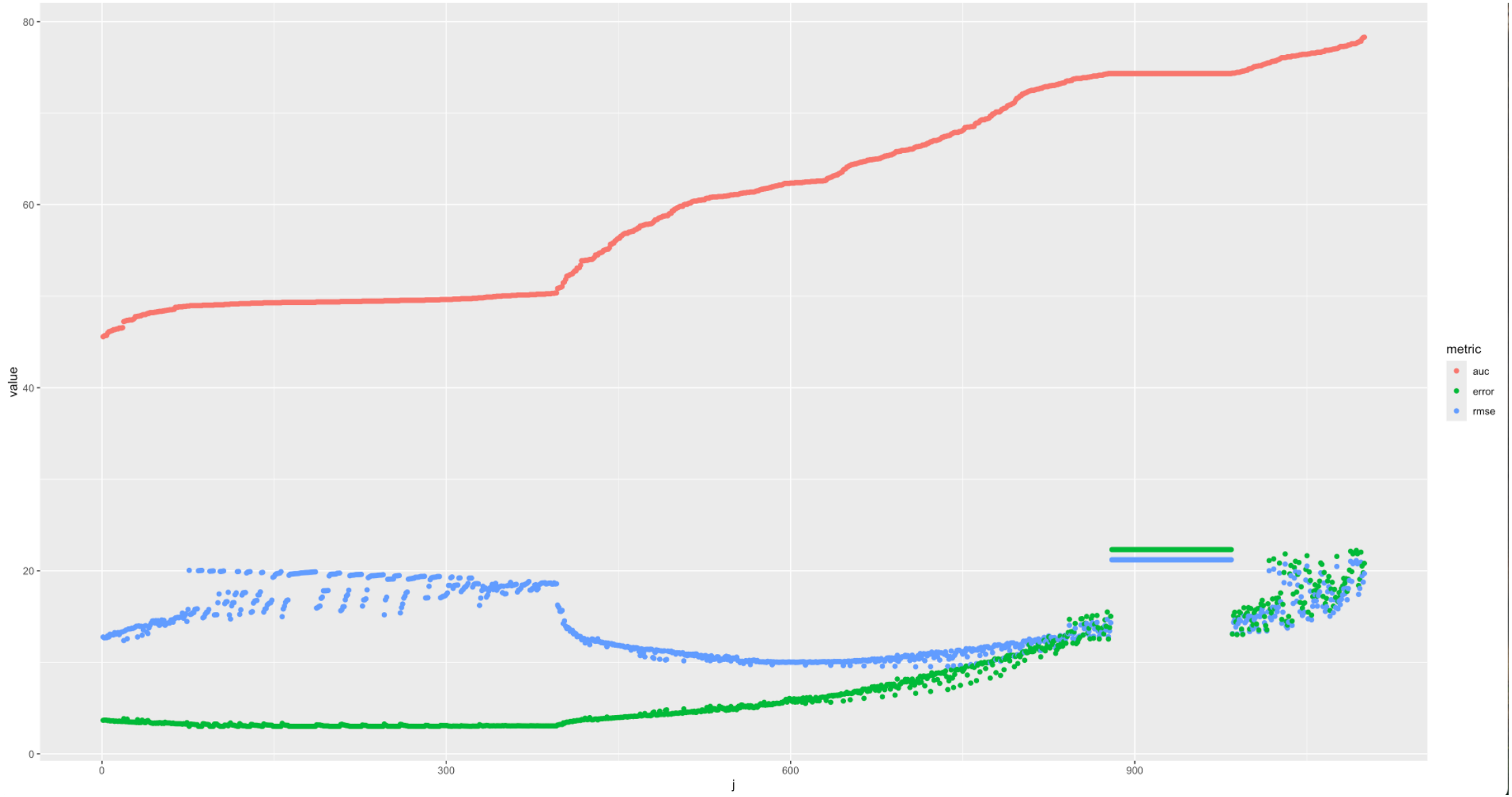
- Tried scaling
  - Did not change prediction performance but feature importance did change

- Added Sex as covariate
  - Did not improve performance, very slight change

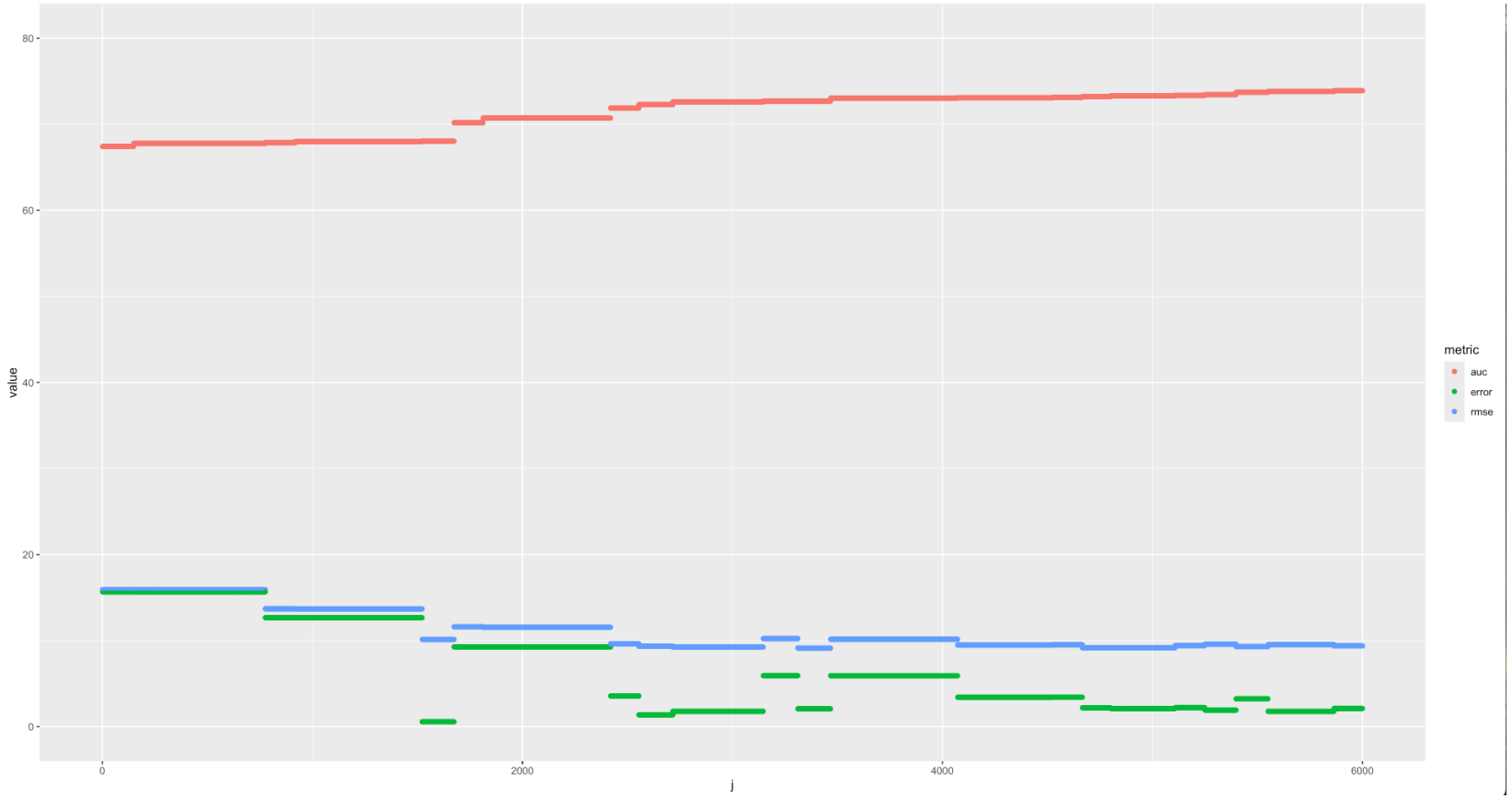
- Univariable Plots of Age vs Protein
  - Tried simple linear regression model of regressing age onto each protein abundance.
  - Used various combination of top proteins in multivariable linear regression to predict age.
    - Performance is moderate

- Averaging performance across 5 folds and choose hyperparameters based on average RMSE performance.
  - Use hyperparameter chosen and the whole Methselah to train; test on other cohorts and Lilly cohort
    - Best model is SVM

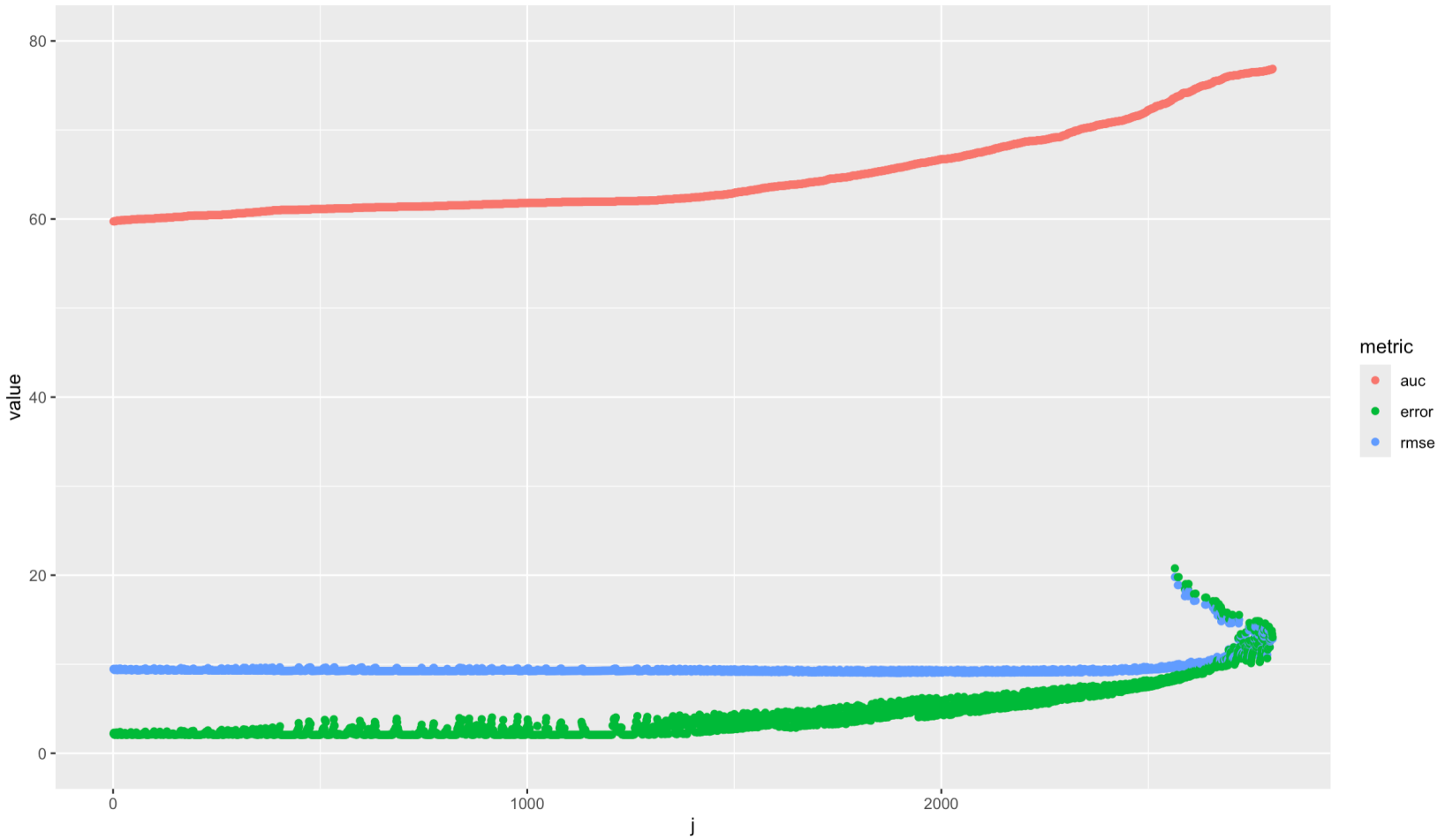
Elastic Net



lightGBM



Svm

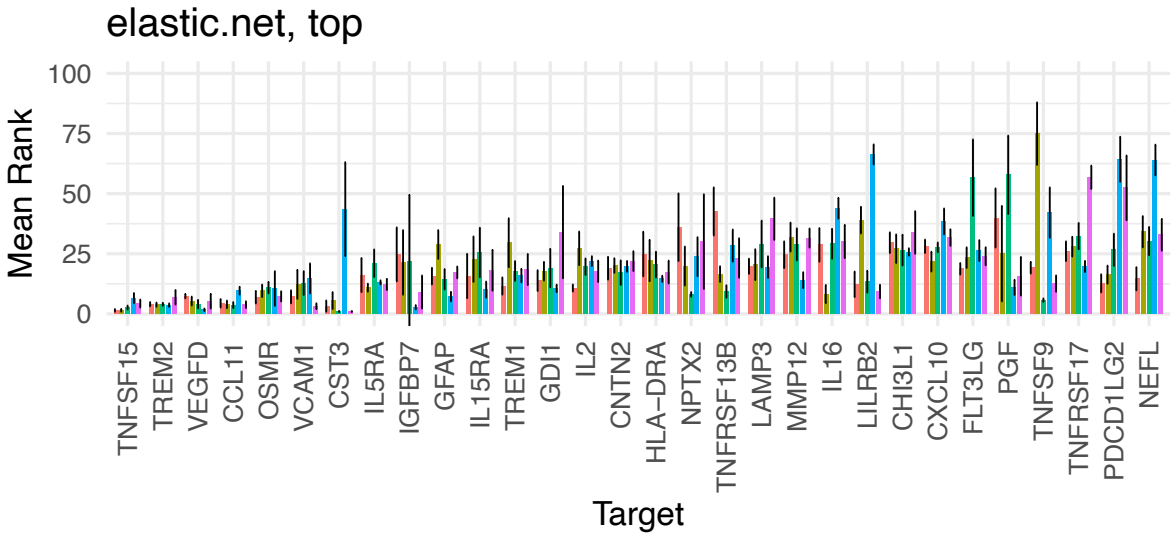


Previous result showed training error (training rmse) to have the same rmse as testing. Turns out a separate validation dataset was used to derive training rmse, leading to svm seemingly underfitting in training set.

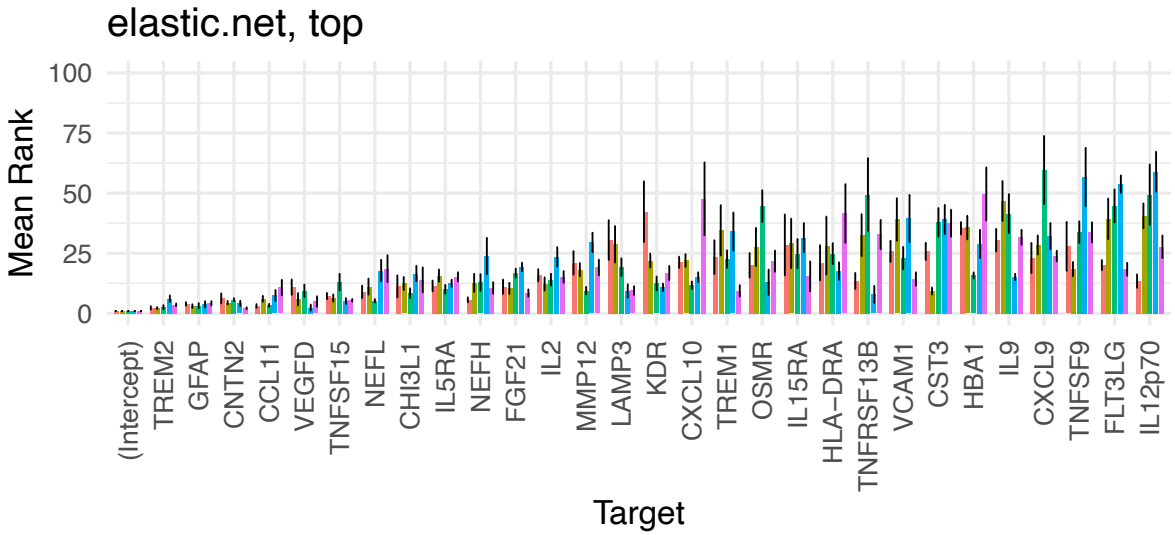
Error was fixed but ultimately the results did not change since testing set rmse was used to decide the selection of hyperparameters

Elastic Net Features Rank

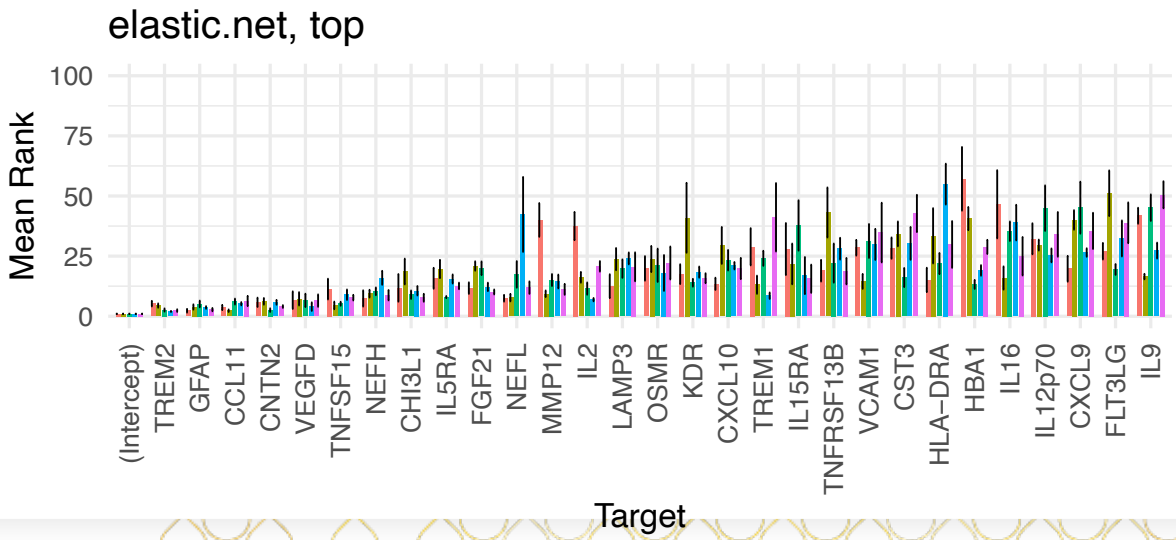
Not scaled + no sex covariate



Scaled



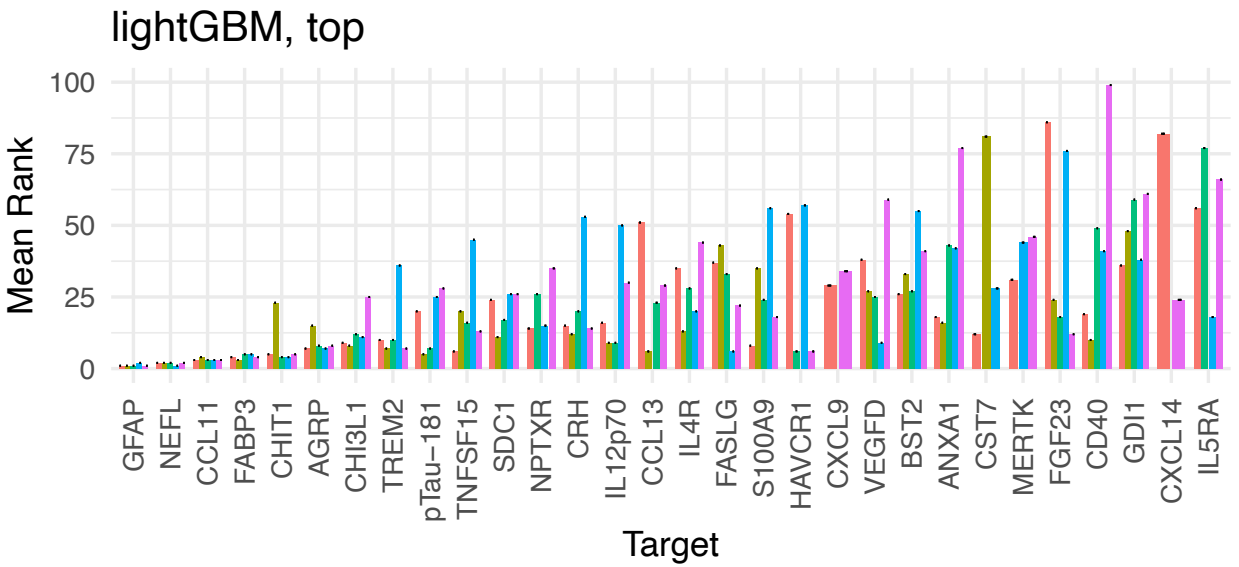
Scaled + Added Sex as covariate



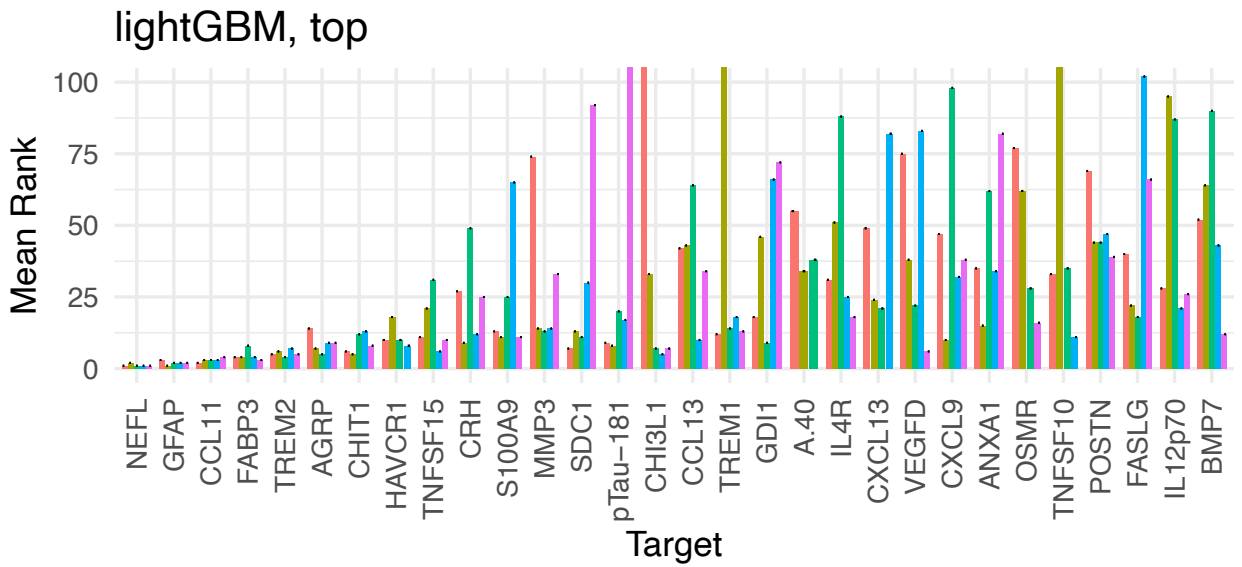
- Top 20 performing models in terms of RMSE derived from different hyperparameters
- averaged rank of features in each of the top 20 models for the 5 cross validations

LightGBM Features Rank

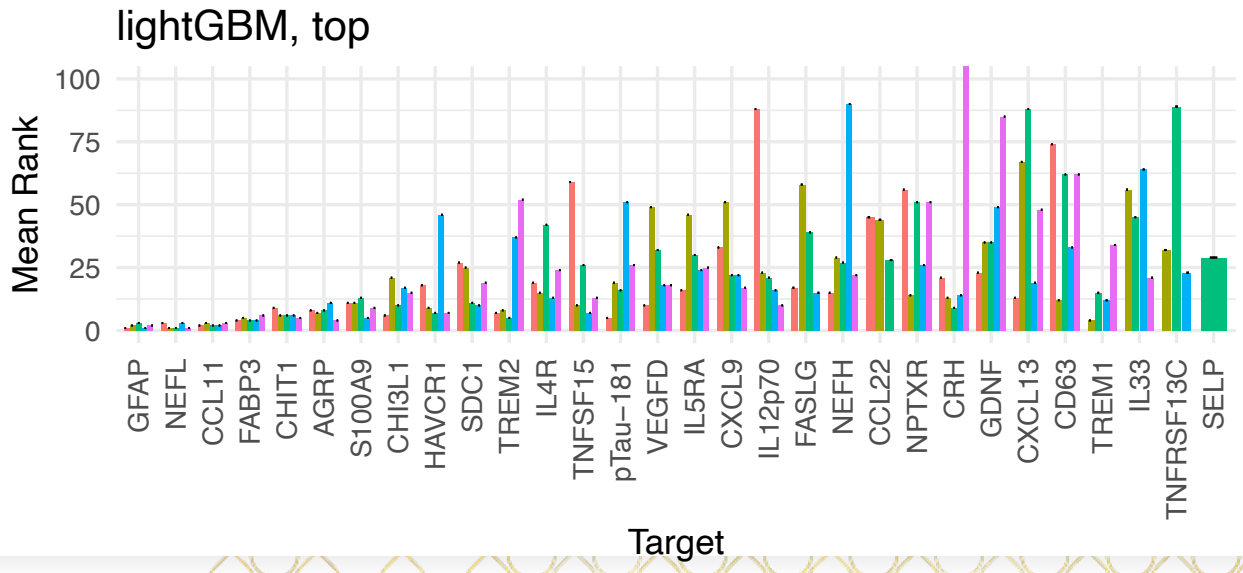
Not scaled + no sex covariate



Scaled

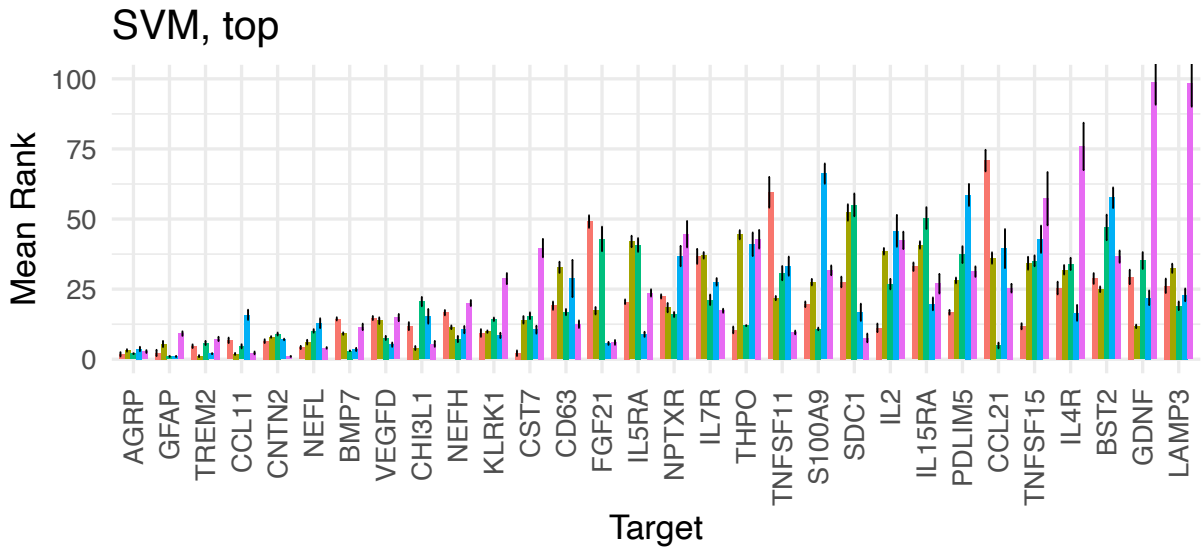


Scaled + Added Sex as covariate

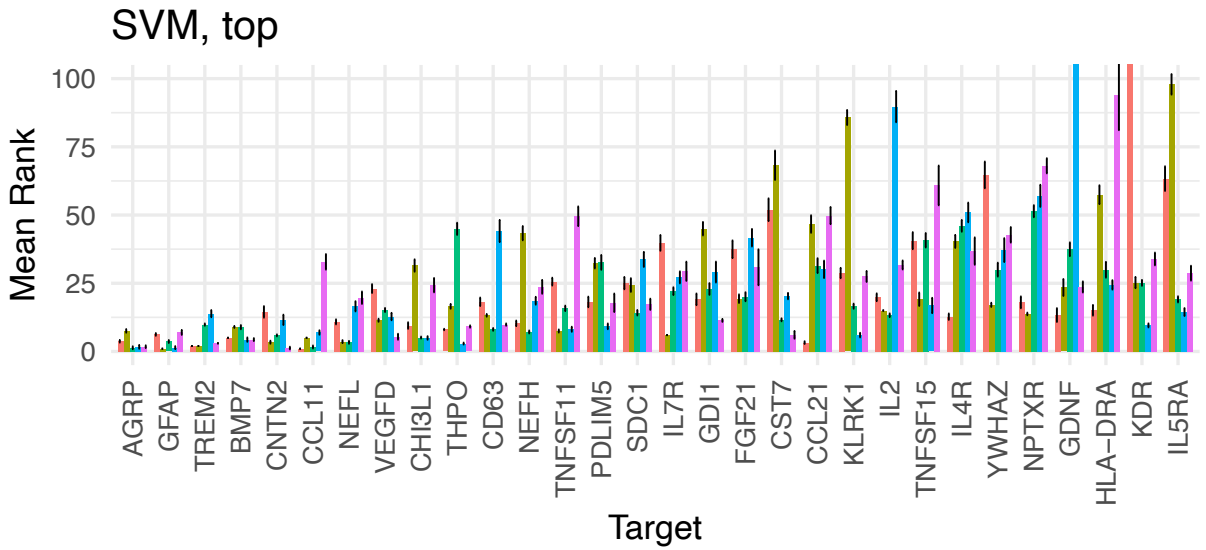


SVM Features Rank

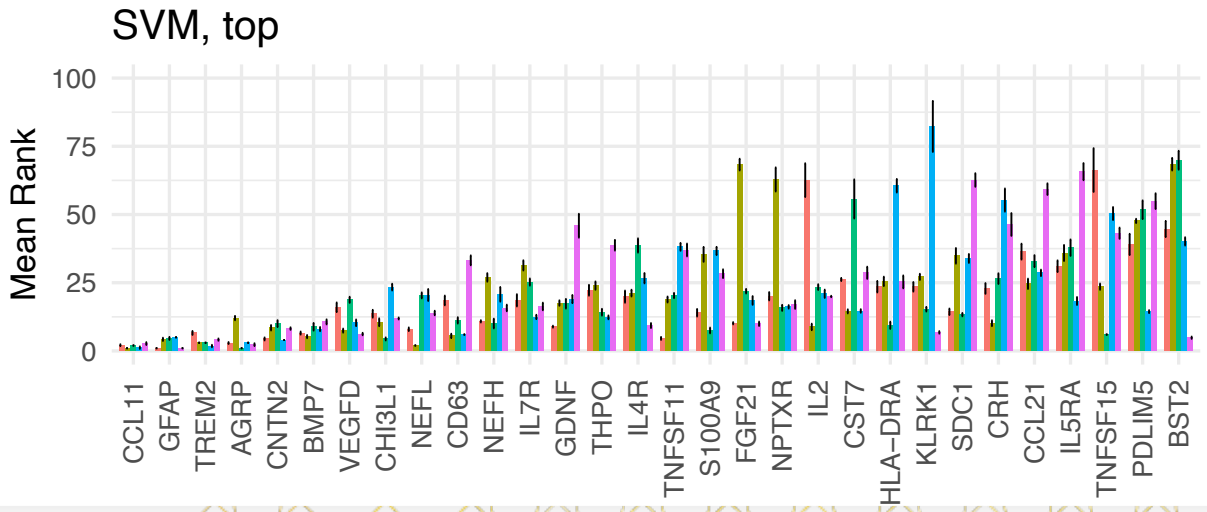
Not scaled + no sex covariate



Scaled



Scaled + Added Sex as covariate

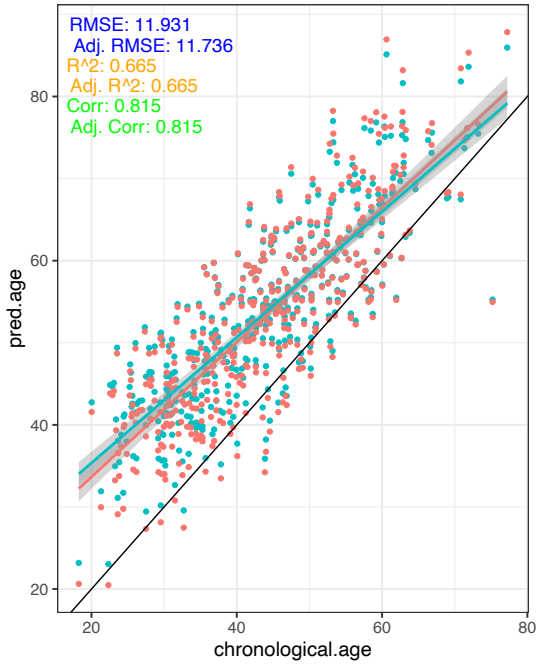


Basel Cohort

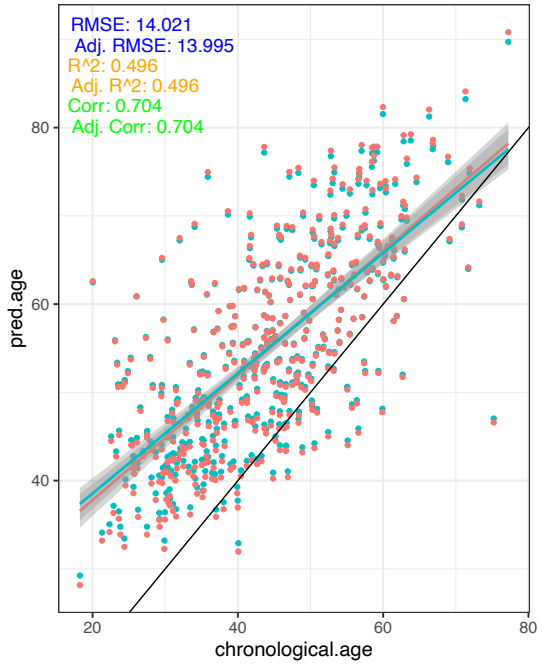
Scaled + Added Sex as covariate

chose hyperparameters based on average RMSE performance

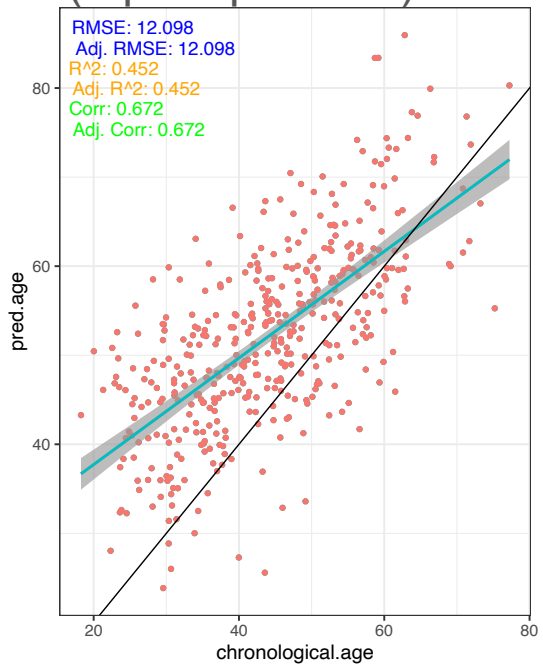
Elastic Net



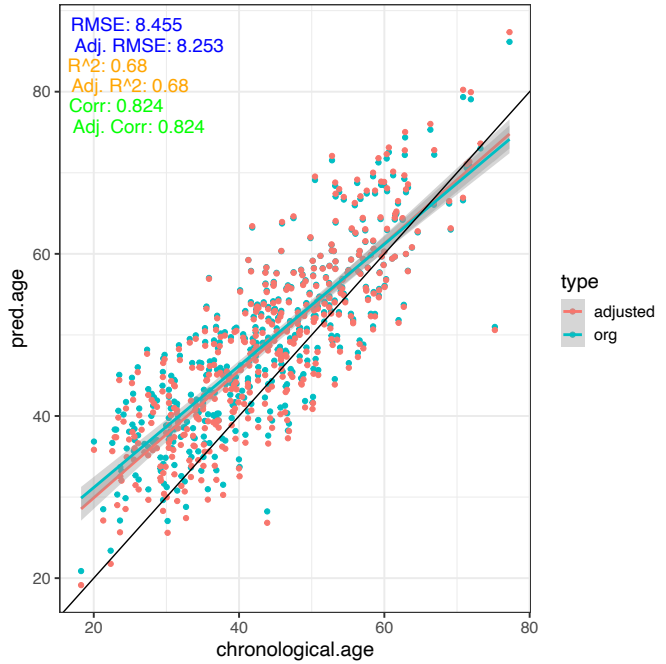
LightGBM



Linear Model (top 20 proteins)



SVM

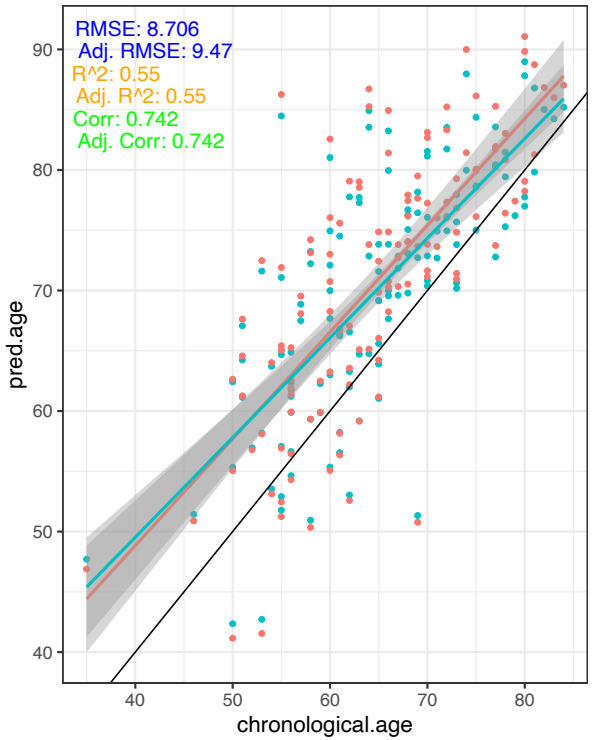


LMU Cohort

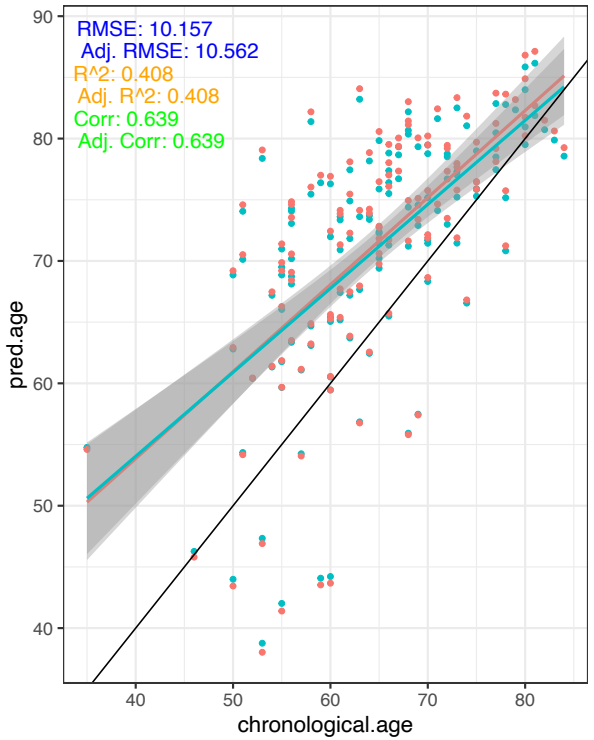
Scaled + Added Sex as covariate

chose hyperparameters based on average RMSE performance

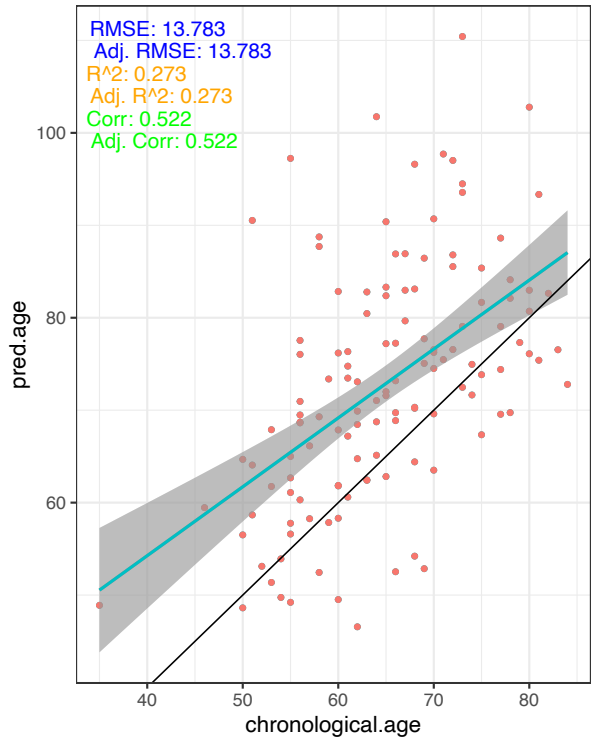
Elastic Net



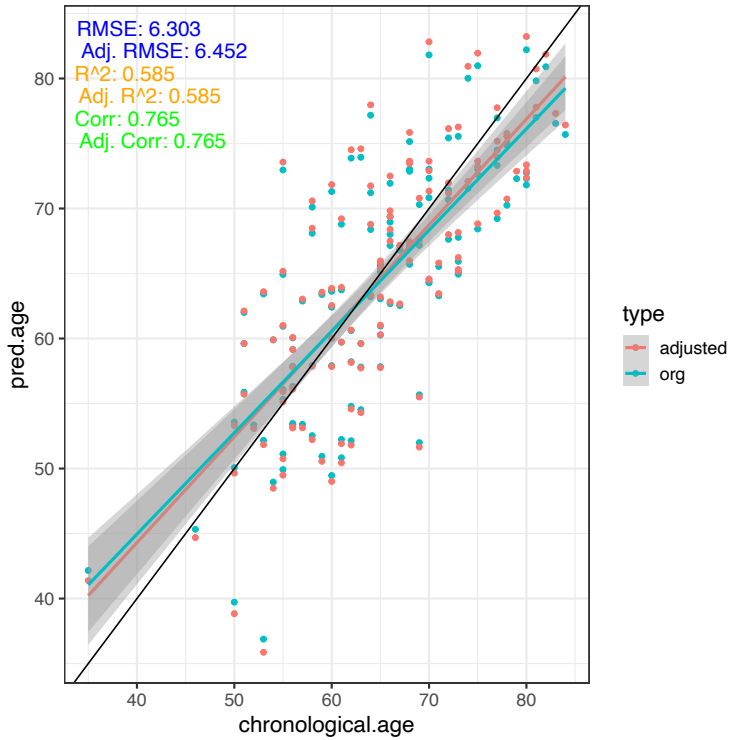
LightGBM



Linear Model (top 20 proteins)



SVM



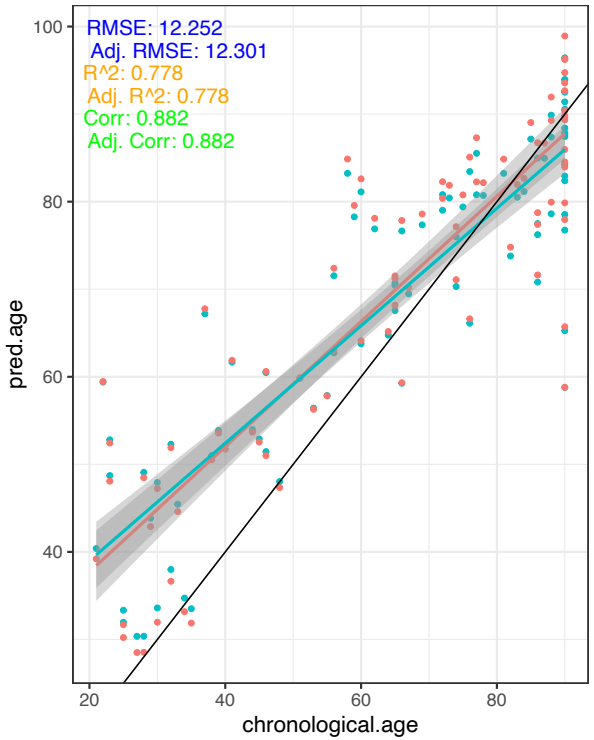
Methuselah Cohort Age Prediction

Scaled + Added Sex as covariate

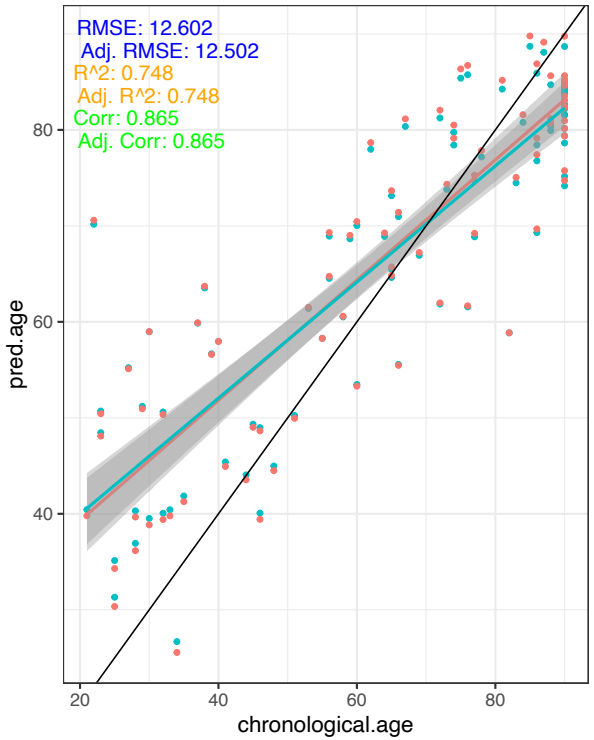
chose hyperparameters based on average RMSE performance

Trained on BRIDGED alpha training set, test on Argo test set

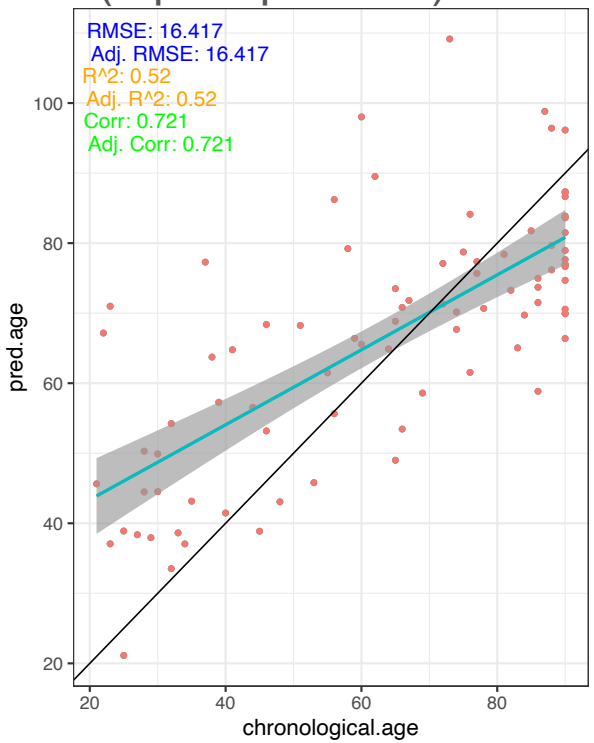
Elastic Net



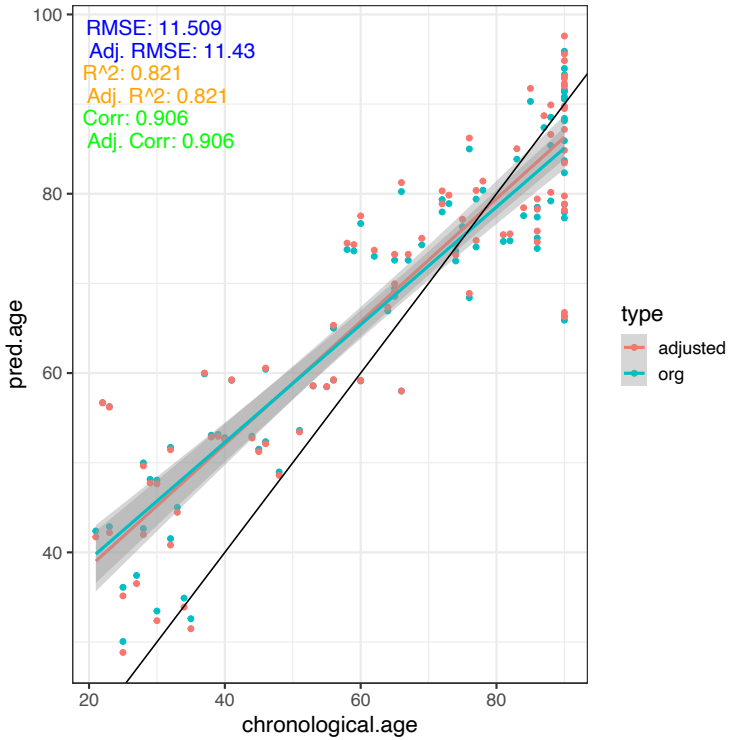
LightGBM



Linear Model (top 20 proteins)



SVM



Basel Cohort

scaled

model	test.fold	rmse	adj.rmse	r2	adj.r2	corr	adj.corr
lm	best.param	12.1331474	12.1331474	0.44983032	0.44983032	0.67069391	0.67069391
lightGBM	best.param	14.0214095	13.9947698	0.49619008	0.49619008	0.70440761	0.70440761
SVM	best.param	8.05334075	7.88975272	0.67837549	0.67837549	0.82363553	0.82363553
elastic.net	best.param	11.946302	11.7517621	0.66487105	0.66487105	0.81539626	0.81539626

Scaled + Added Sex as covariate

model	test.fold	rmse	adj.rmse	r2	adj.r2	corr	adj.corr
lm	best.param	12.0979148	12.0979148	0.45161666	0.45161666	0.6720243	0.6720243
lightGBM	best.param	14.0214095	13.9947698	0.49619008	0.49619008	0.70440761	0.70440761
SVM	best.param	8.45523046	8.25310565	0.67979238	0.67979238	0.82449523	0.82449523
elastic.net	best.param	11.9313664	11.736092	0.66479162	0.66479162	0.81534754	0.81534754

LMU Cohort

scaled

model	test.fold	rmse	adj.rmse	r2	adj.r2	corr	adj.corr
lm	best.param	13.6598544	13.6598544	0.26869217	0.26869217	0.51835526	0.51835526
lightGBM	best.param	10.1568589	10.5623166	0.40809818	0.40809818	0.63882562	0.63882562
SVM	best.param	6.34395534	6.47552791	0.58589344	0.58589344	0.76543677	0.76543677
elastic.net	best.param	8.70547006	9.46952566	0.55041253	0.55041253	0.74189792	0.74189792

Scaled + Added Sex as covariate

model	test.fold	rmse	adj.rmse	r2	adj.r2	corr	adj.corr
lm	best.param	13.7828032	13.7828032	0.27273133	0.27273133	0.52223686	0.52223686
lightGBM	best.param	10.1568589	10.5623166	0.40809818	0.40809818	0.63882562	0.63882562
SVM	best.param	6.30339142	6.45232233	0.58507177	0.58507177	0.76489984	0.76489984
elastic.net	best.param	8.70553059	9.46984801	0.55010028	0.55010028	0.74168746	0.74168746

Methuselah Cohort Age Prediction

Trained on BRIDGED alpha training set, test on Argo test set

scaled

model	test.fold	rmse	adj.rmse	r2	adj.r2	corr	adj.corr
lm	best.param	16.3093544	16.3093544	0.52298636	0.52298636	0.72317796	0.72317796
lightGBM	best.param	12.5497429	12.4609626	0.7519101	0.7519101	0.8671275	0.8671275
SVM	best.param	11.5937957	11.4985981	0.81547008	0.81547008	0.90303382	0.90303382
elastic.net	best.param	12.255431	12.3049224	0.77744566	0.77744566	0.88172879	0.88172879

Scaled + Added Sex as covariate

model	test.fold	rmse	adj.rmse	r2	adj.r2	corr	adj.corr
lm	best.param	16.4170091	16.4170091	0.5204918	0.5204918	0.72145117	0.72145117
lightGBM	best.param	12.6016935	12.5017993	0.7480554	0.7480554	0.86490196	0.86490196
SVM	best.param	11.5089525	11.429593	0.82073132	0.82073132	0.90594223	0.90594223
elastic.net	best.param	12.2515505	12.3008163	0.7776771	0.7776771	0.88186002	0.88186002