

Comparison of SiMLR and RGCCA: summary of all experiments

Brian B. Avants et al.

2020-10-16

```
## Warning: package 'PAutilities' was built under R version 4.0.2
```

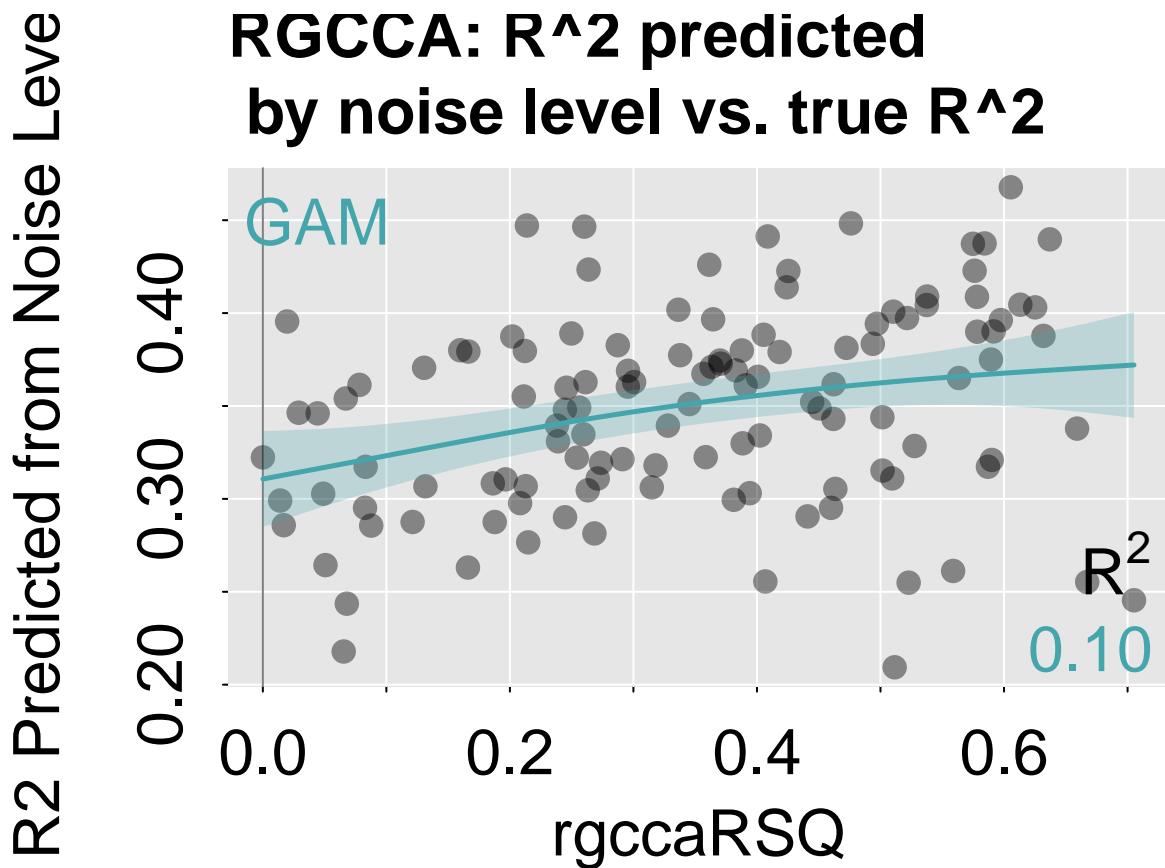
```
## package 'PAutilities' was built under R version 3.5.0
```

Noise and signal in simulation data.

```
mdl=lm(formula = rgccaRSQ ~ corrupt1 + corrupt2 + corrupt3, data = simCCAMIXSVD)
R2_predicted_from_noise_level=predict(mdl)
```

```
rtemis::mplot3.xy( simCCAMIXSVD$rgccaRSQ, R2_predicted_from_noise_level, main='RGCCA: R^2 predicted \n
```

RGCCA: R^2 predicted by noise level vs. true R^2



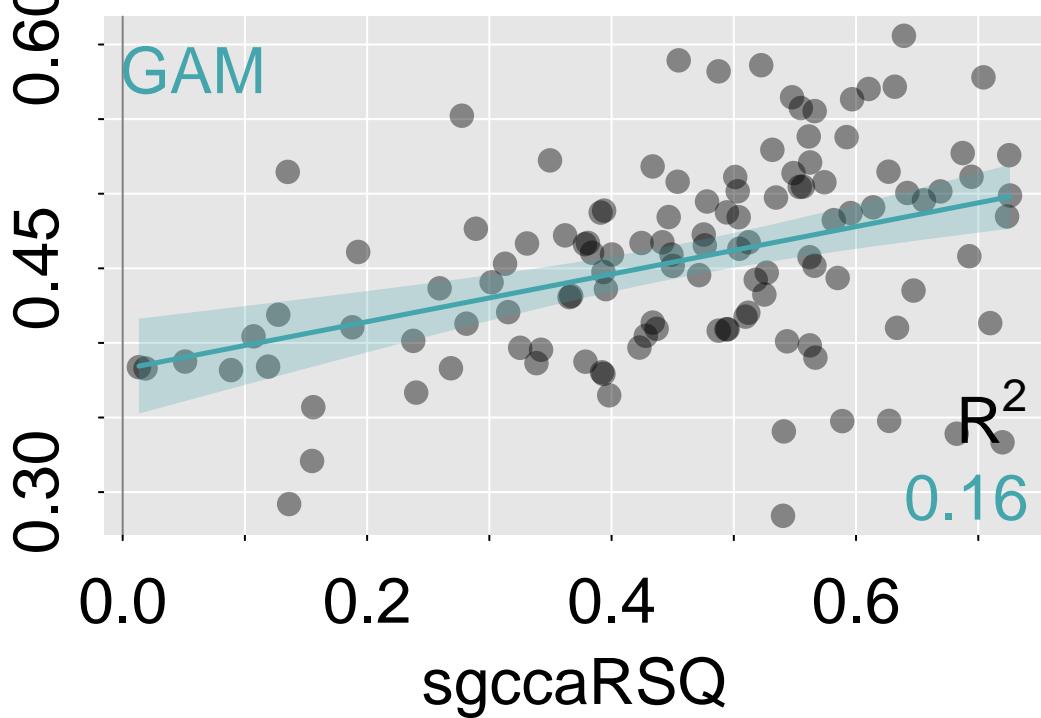
```
mdl=lm(formula = sgccaRSQ ~ corrupt1 + corrupt2 + corrupt3, data = simCCAMIXSVD)
```

```
R2_predicted_from_noise_level=predict(mdl)
```

```
rtemis::mplot3.xy( simCCAMIXSVD$sgccaRSQ, R2_predicted_from_noise_level, main='SGCCA: R^2 predicted \n
```

R2 Predicted from Noise Level

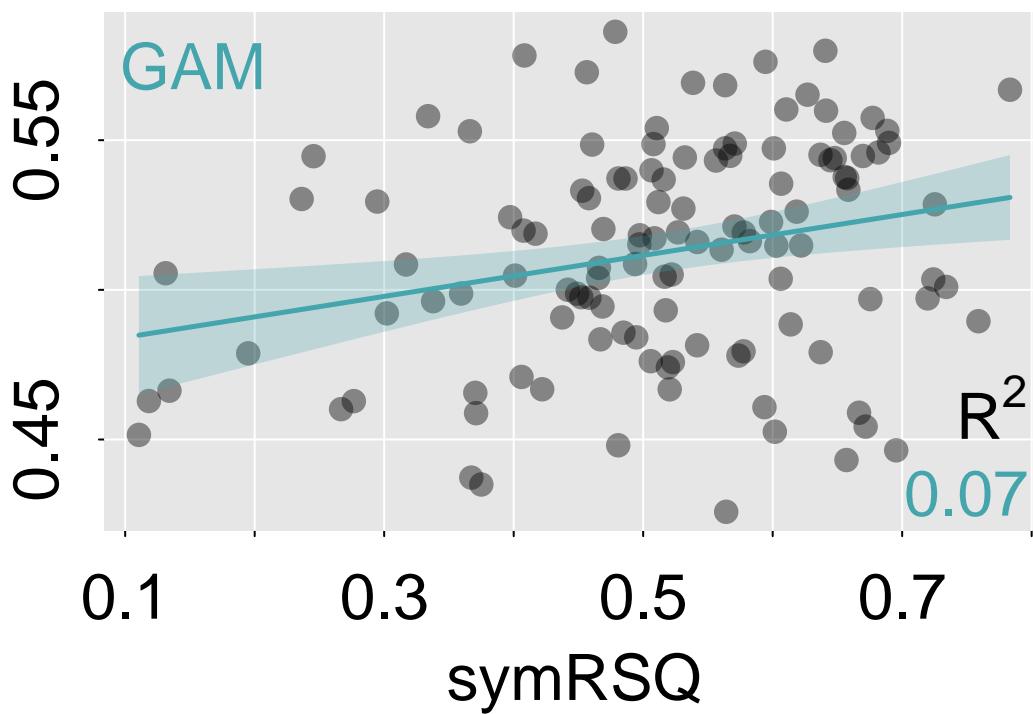
SGCCA: R² predicted by noise level vs. true R²



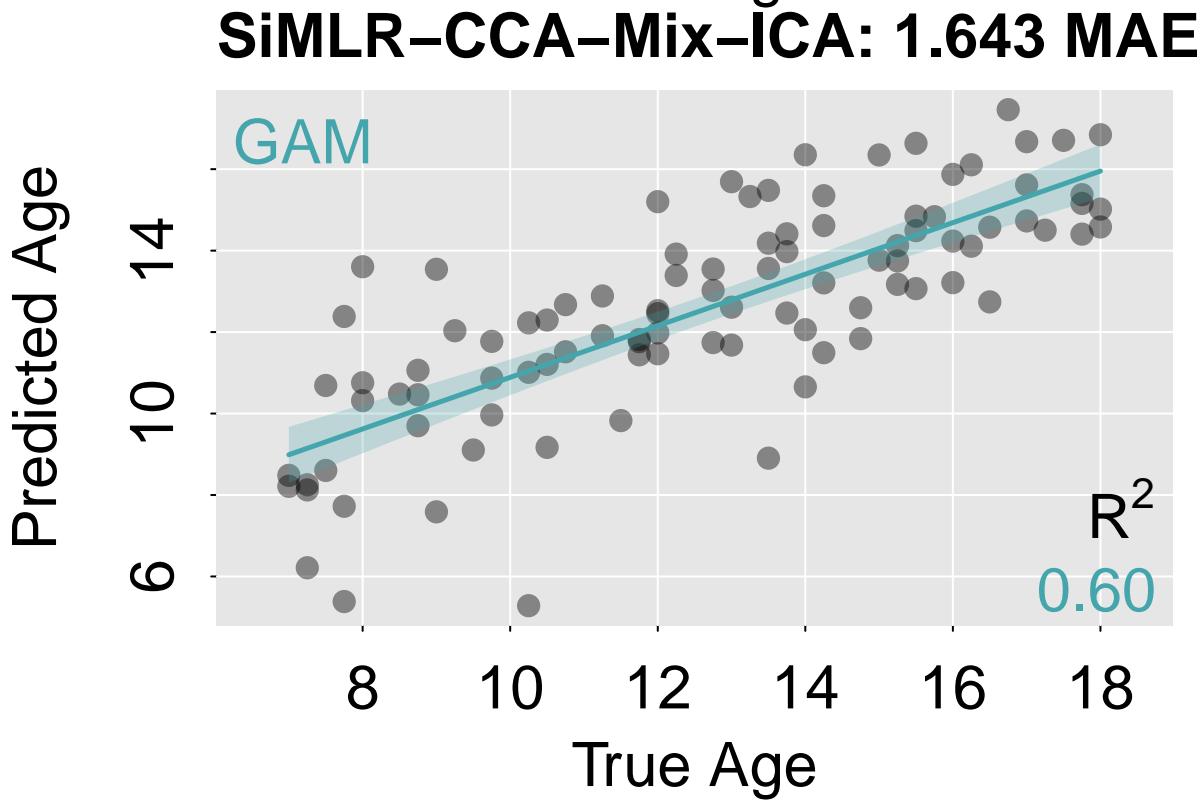
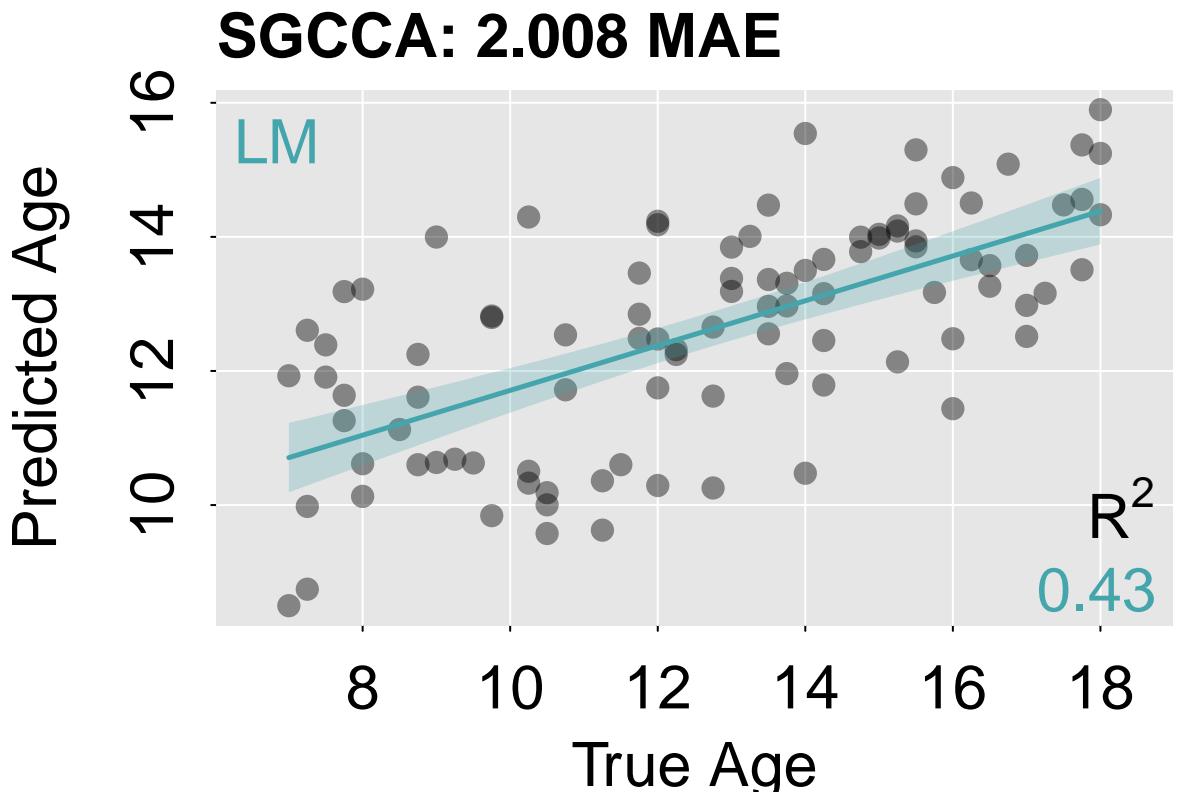
```
mdl=lm(formula = symRSQ ~ corrupt1 + corrupt2 + corrupt3, data = simCCAMIXSVD)
R2_predicted_from_noise_level=predict(mdl)
rtemis::mplot3.xy( simCCAMIXSVD$symRSQ, R2_predicted_from_noise_level, main='SiMLR-CCA-SVD: R^2 predicted by noise level vs. true R^2')
```

R2 Predicted from Noise Level

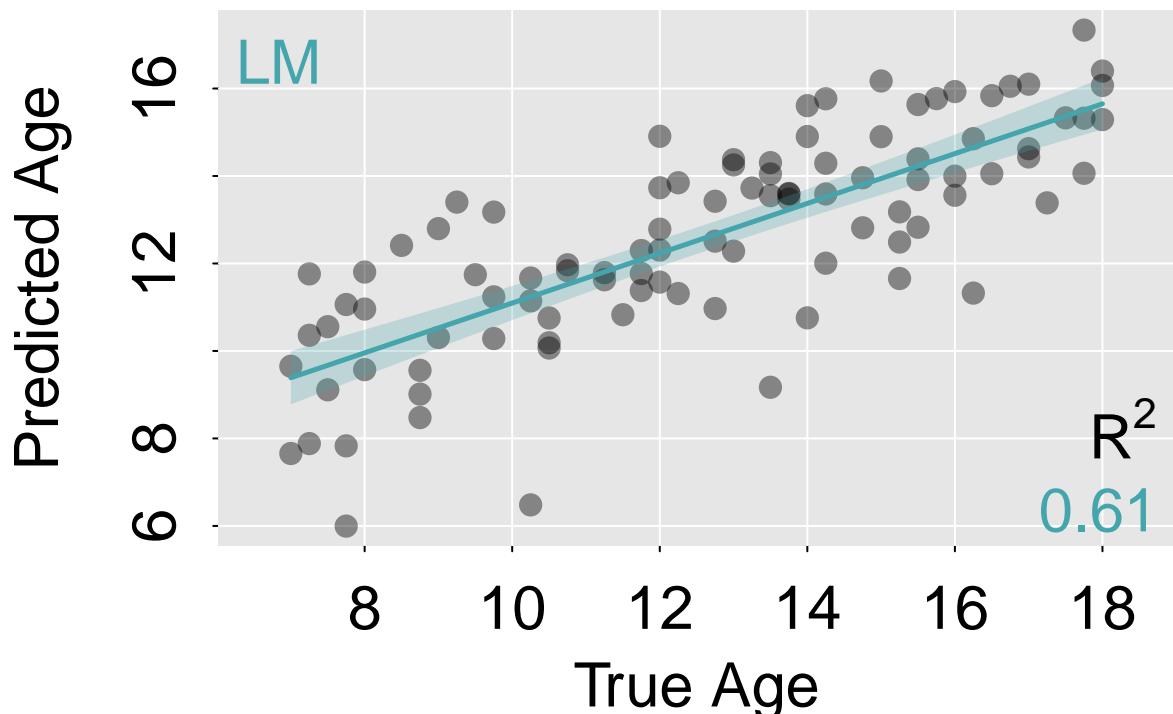
SiMLR–CCA–SVD: R² predicted by noise level vs. true R²



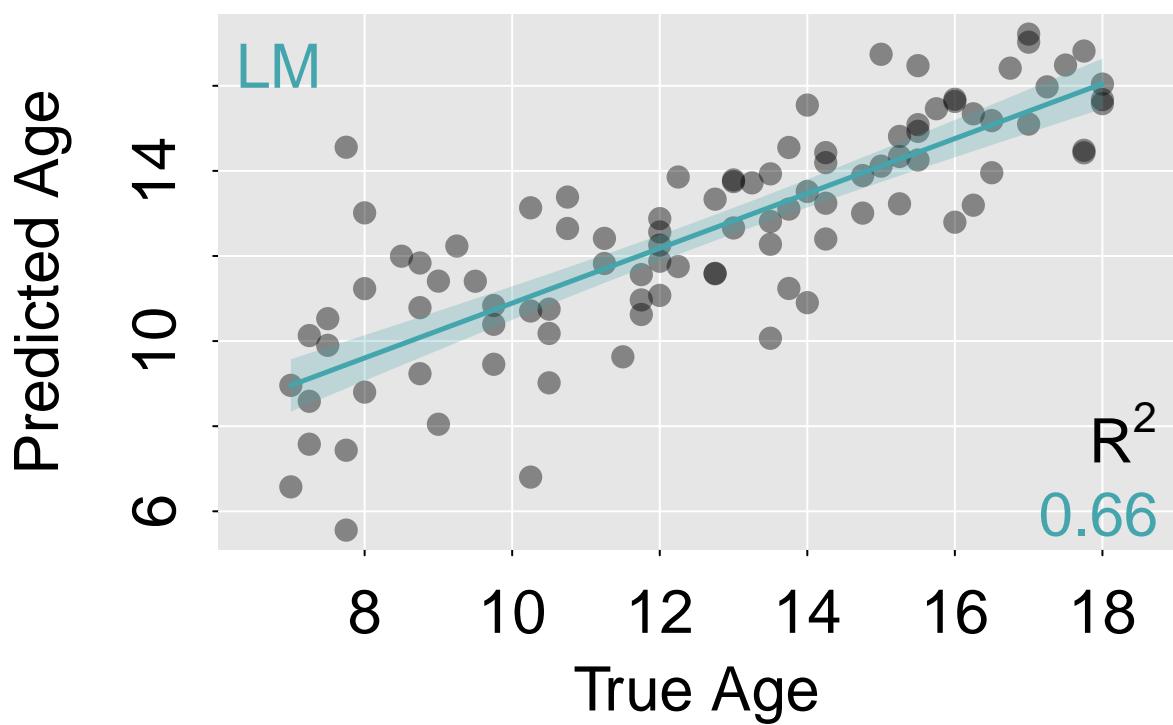
brain age figures



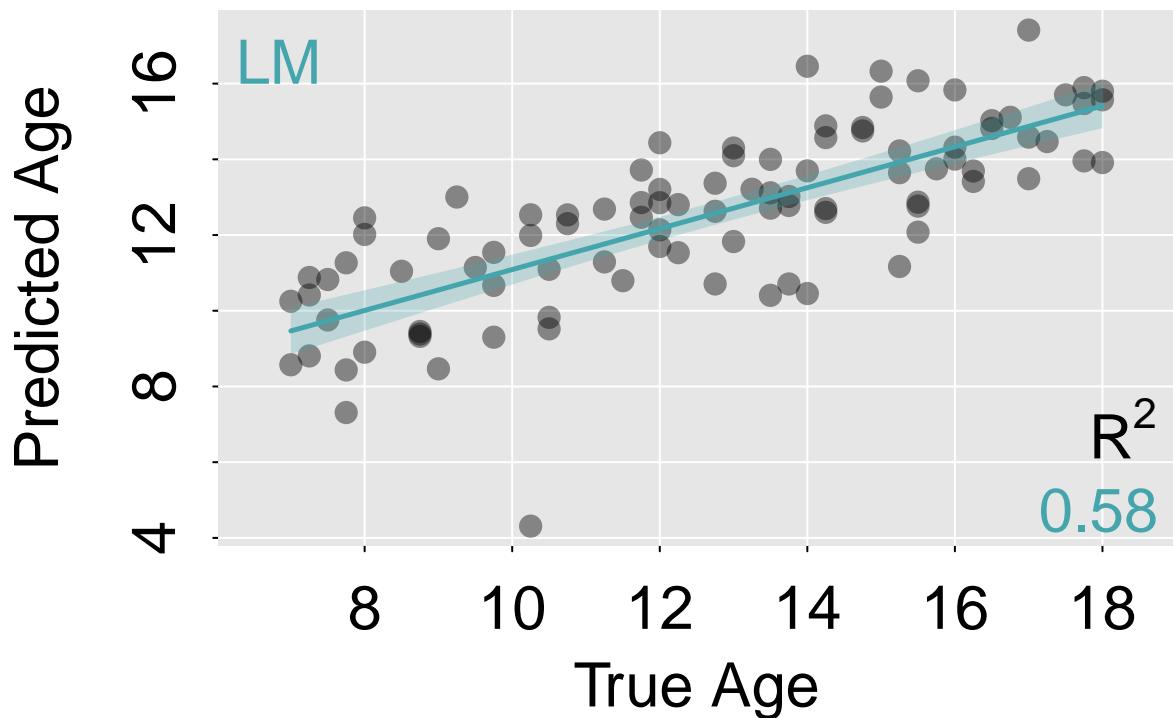
SiMLR–Reg–Mix–ICA: 1.575 MAE



SiMLR–CCA–Mix–SVD: 1.444 MAE



SiMLR–Reg–Mix–SVD: 1.701 MAE



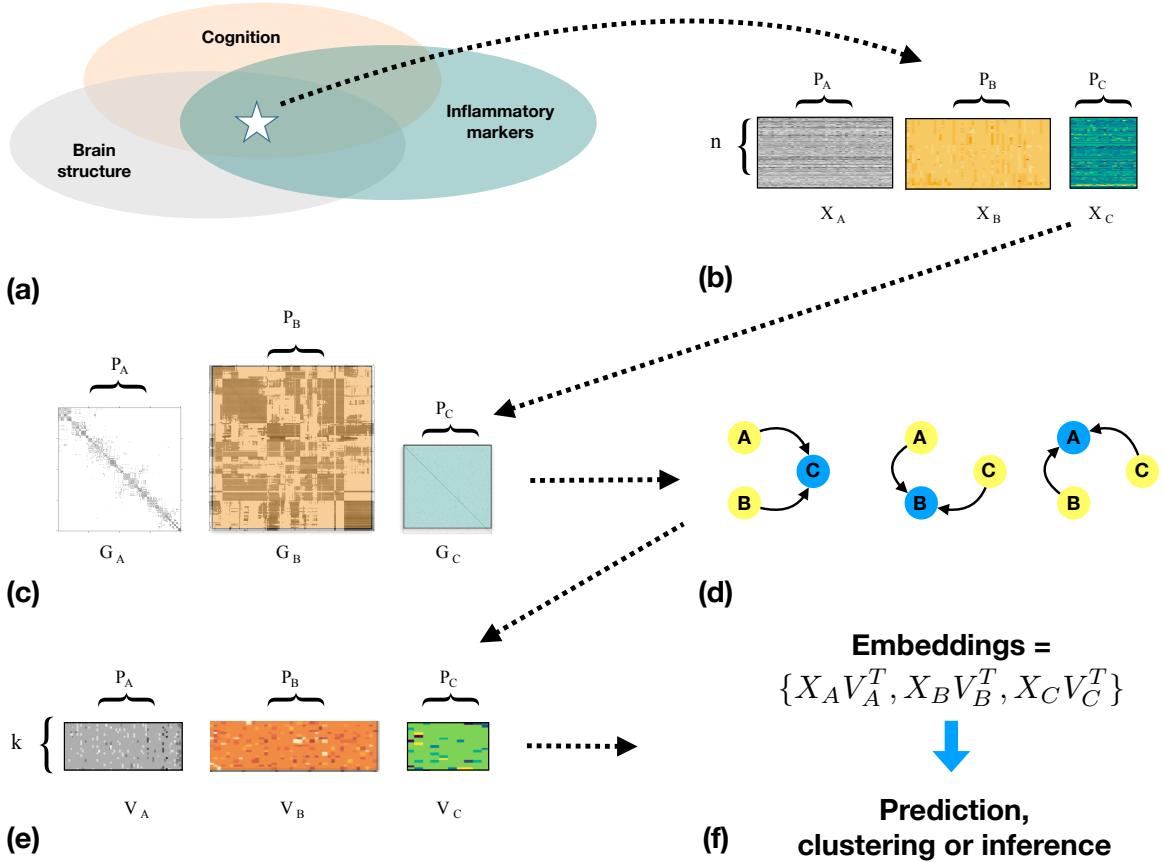


Figure 1: Example SiMLR study overview.

Figure 2. Simulated latent signal embedded within modality-specific matrices

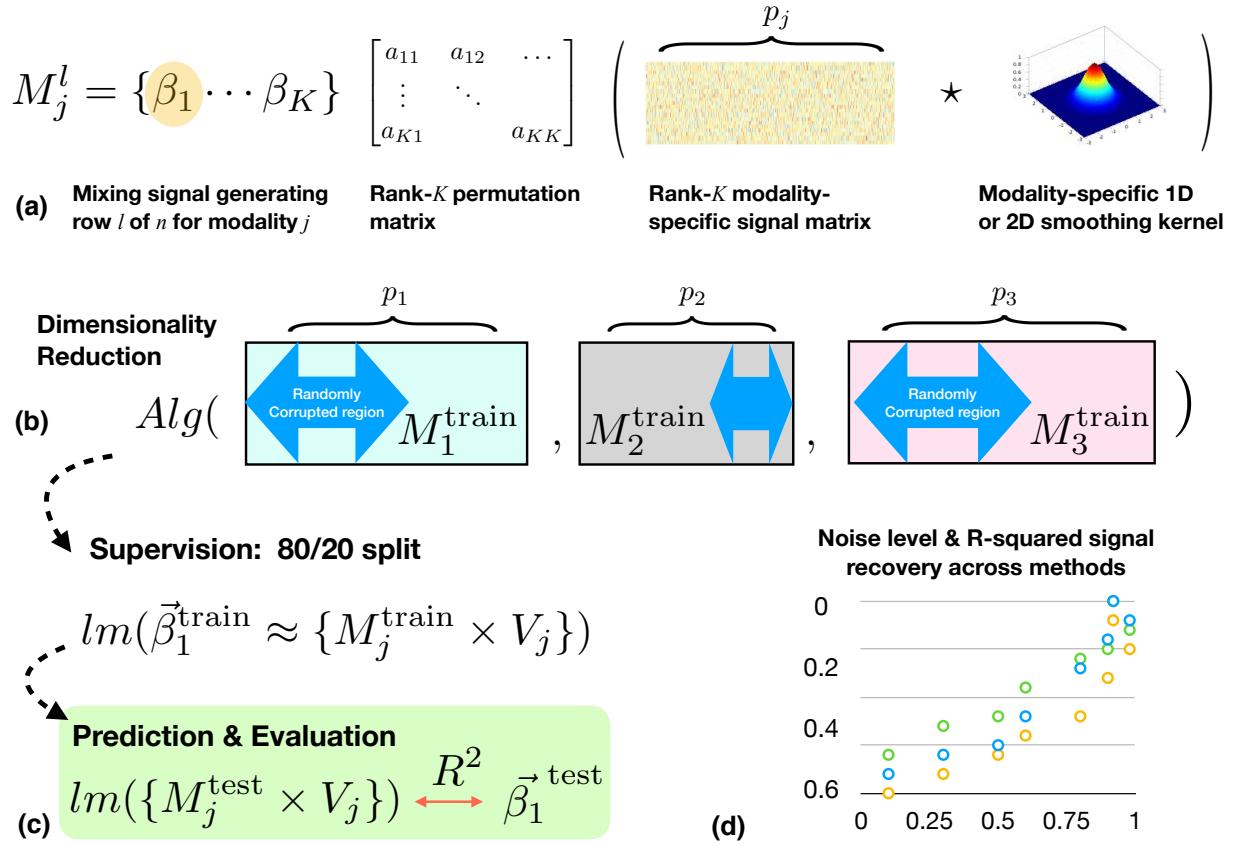


Figure 2: SiMLR simulation study design.

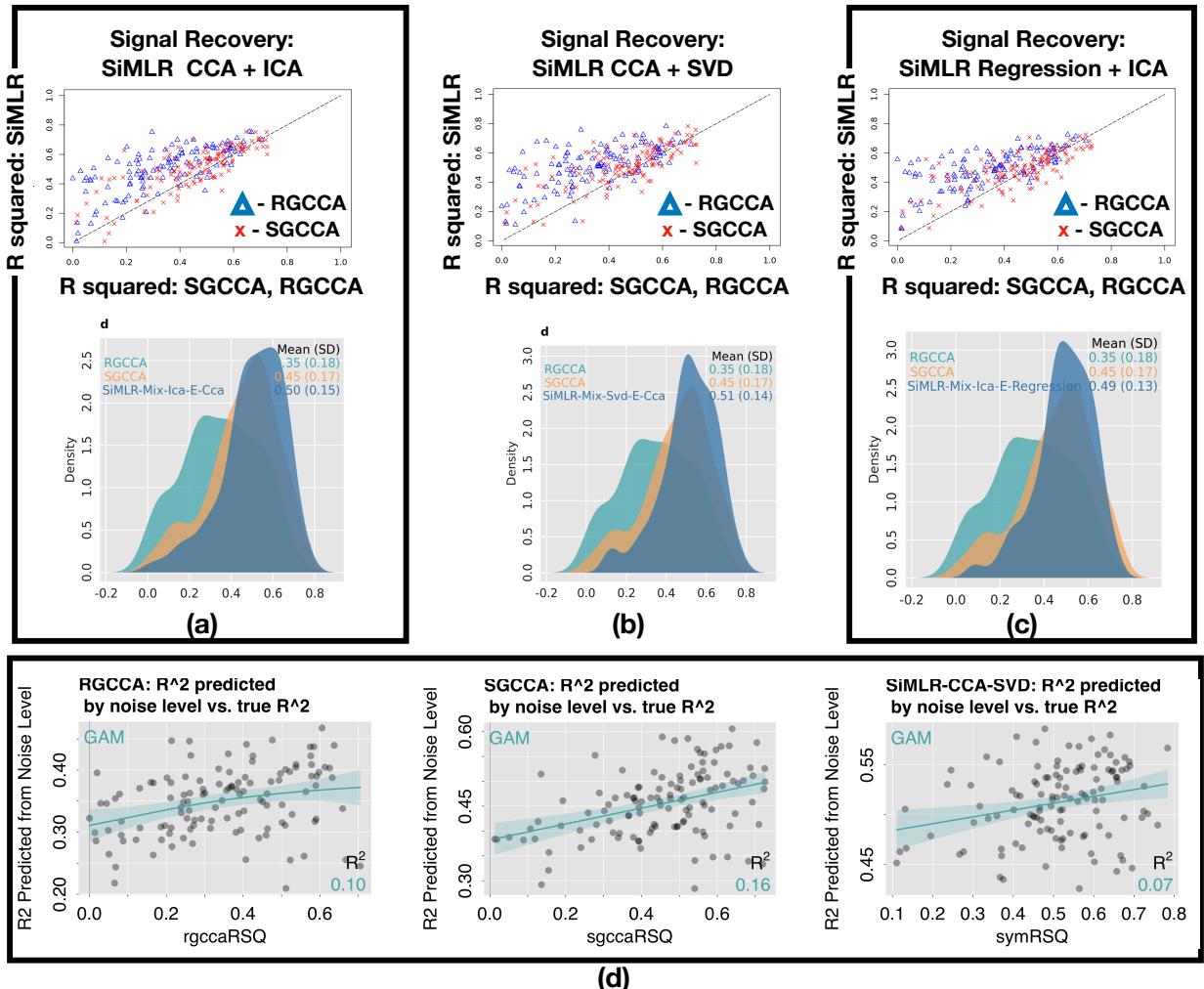


Figure 3: SiMLR simulation study results: sensitivity to noise and ability to recover signal.

```
## Warning: package 'huxtable' was built under R version 4.0.2
```

Table 1: Summary of RGCCA-SGCCA-SiMLR comparison results. RGCCA = regularized generalized canonical correlation analysis; SGCCA = sparse generalized canonical correlation analysis; Sim = similarity-driven multivariate linear reconstruction (SiMLR); Reg = regression; CCA = absolute canonical covariance; ICA = ICA mixing method; SVD = SVD mixing method.

Best results are highlighted in cadet blue; worst in antiquewhite. SiMLR with the absolute canonical covariance similarity measurement and SVD (SimCCASVD) as a mixing method performs best overall. SiMLR with the regression energy and ICA mixing method (SimRegICA) outperforms SGCCA most consistently across sparseness levels, provides closely competitive performance overall, and is highlighted in pink. The PING examples are exploratory analyses described in the supplementary information as we cannot directly share the data. The "n comp" description in the PING table refers to the number of significant components related to either anxiety or depression.

study	RGCCA	SGCCA	SimCCAICA	SimCCASVD	SimRegICA	SimRegSVD	metric
Signal-Sens.	0.35+/-0.18	0.45+/-0.17	0.5+/-0.15	0.51+/-0.14	0.49+/-0.13	0.49+/-0.14	R-squared
Noise-Sens.	0.09	0.16	0.09	0.06	0.07	0.1	R-squared
Mixomics	N/A	0.62 +/ 0.01	0.64 +/ 0.03	0.65 +/ 0.03	0.65 +/ 0.04	0.61 +/ 0.03	Concordance
brainAge	N/A	2+/-1.5	1.6+/-1.2	1.4+/-1.2	1.6+/-1.3	1.7+/-1.2	MAE
PING-Anx	N/A	1 comp.	N/A	3 comp.	5 comp.	N/A	Inferential
PING-Dep	N/A	1 comp.	N/A	1 comp.	5 comp.	N/A	Inferential

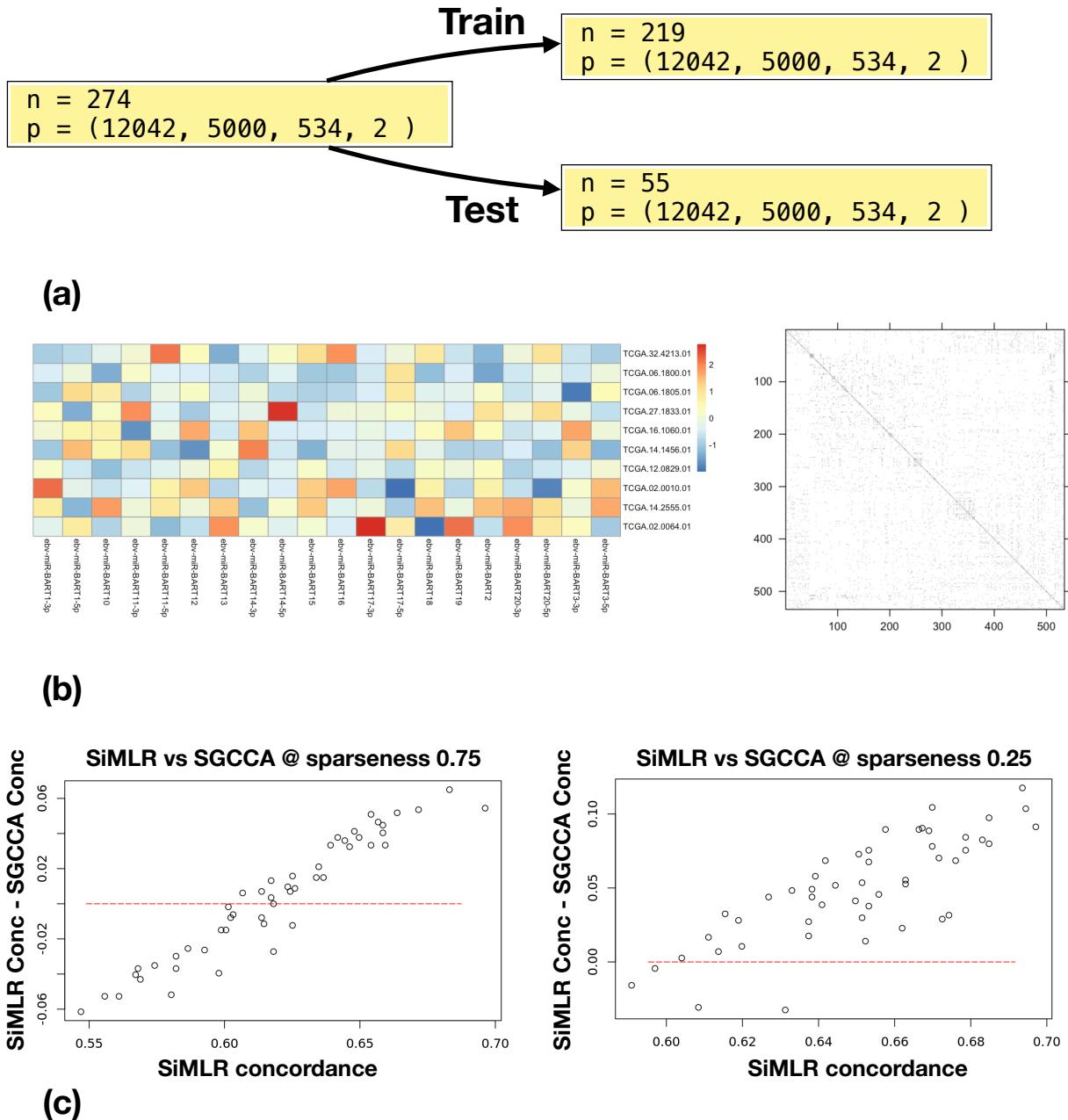


Figure 4: Mixomics study: supervised survival prediction comparison to RGCCA.

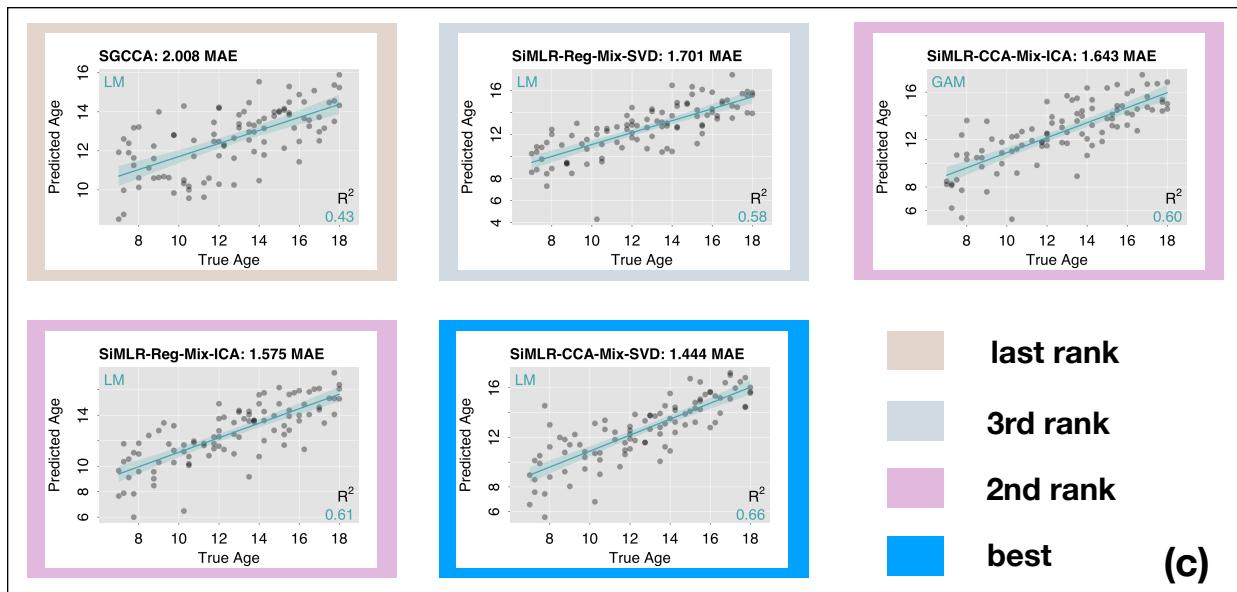
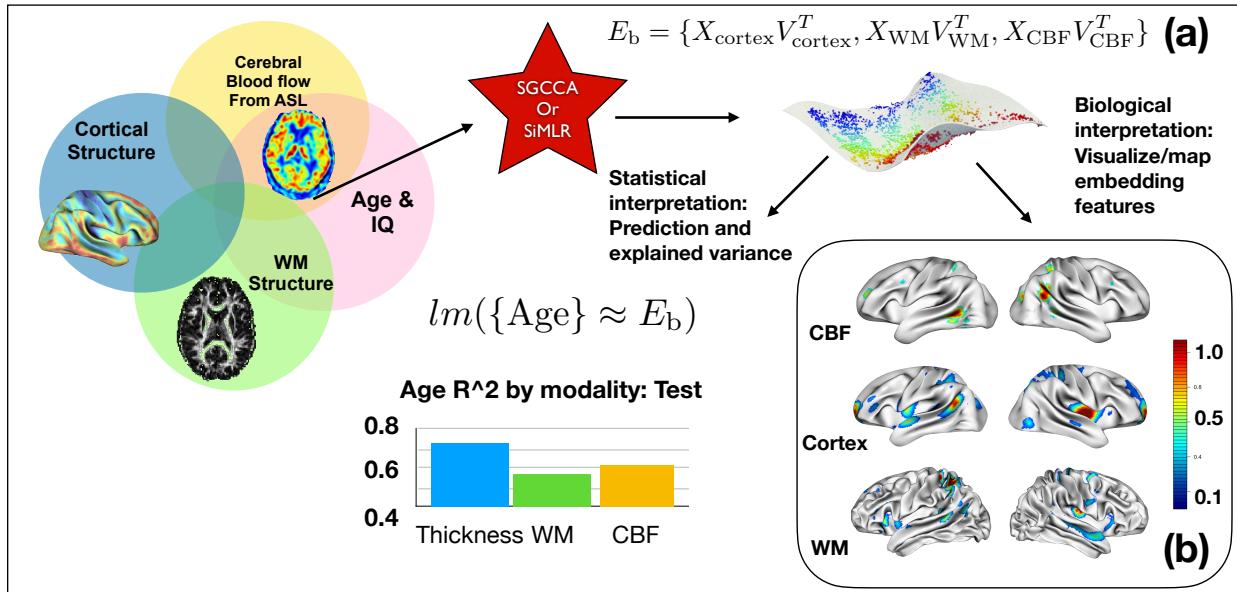


Figure 5: PTBP fully supervised brain age prediction: comparison to RGCCA.