

nano-lazar: Validation of read across predictions for nanoparticle toxicities

Christoph Helma¹, Micha Rautenberg¹, Denis Gebele¹

June 8, 2016

¹ in silico toxicology gmbh, Basel, Switzerland

Introduction

Data requirements

Calculation of similarities intersection of physchem descriptors

Experimental data for similar compounds

Use cases

- no nanoparticle information: core+coating properties
- physchem measurements
- proteomics

Objectives

- Evaluate currently available nanoparticle data for read across predictions
- Compare read across predictions based on
 - calculated core and coating properties
 - measured nanoparticle properties
 - nanoparticle protein corona

Methods

Results

Data requirements

Physchem properties

Table 1: Repeated crossvalidation results for models with physchem properties, ** best results of all experiments, * no statistically significant difference to best results ($p > 0.05$)

Algorithm	r^2	RMSE
Weighted average	0.42, 0.46, 0.48	2.02, 1.94, 1.92
Partial least squares	0.53, 0.54, 0.49	1.83, 1.8, 1.9
Random forest	0.53, 0.52, 0.54	1.82, 1.84, 1.79

Protein corona

Table 2: Repeated crossvalidation results for models with protein corona data, ** best results of all experiments, * no statistically significant difference to best results ($p > 0.05$)

Algorithm	r^2	RMSE
Weighted average	0.66, 0.63, 0.63 *	1.58, 1.62, 1.66 *
Partial least squares	0.59, 0.66, 0.63 *	1.74, 1.56, 1.65 *
Random forest	0.66, 0.65, 0.63 *	1.56, 1.59, 1.64 *

Physchem properties and protein corona

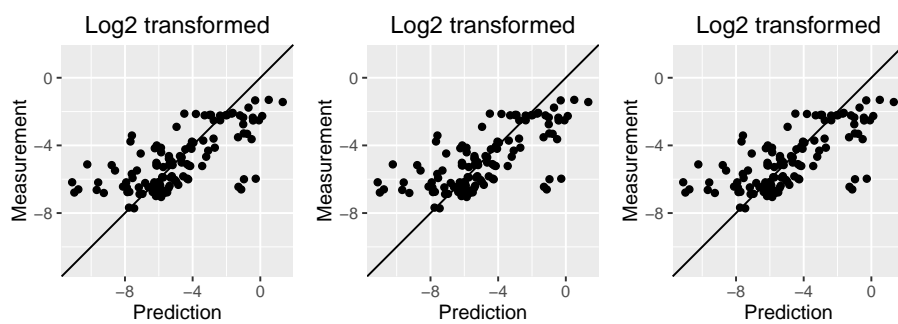


Figure 1: Correlation of log2 transformed net cell association measurements with weighted average predictions using physchem properties.

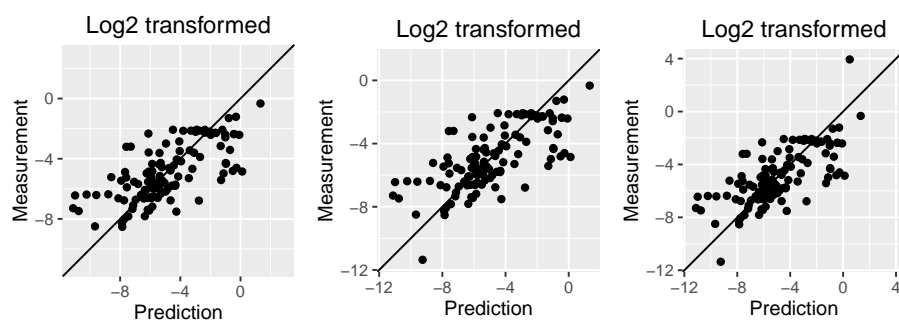


Figure 2: Correlation of log2 transformed net cell association measurements with partial least squares predictions using physchem properties.

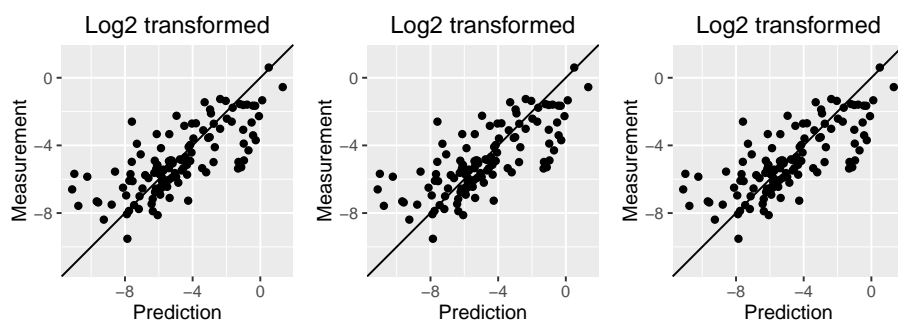


Figure 3: Correlation of log2 transformed net cell association measurements with random forest predictions using physchem properties.

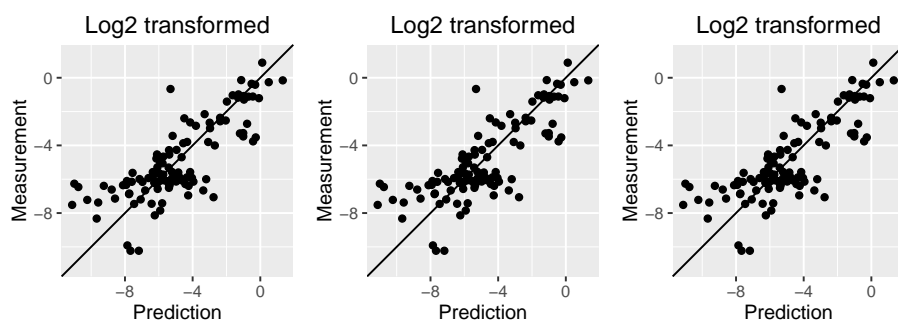


Figure 4: Correlation of log2 transformed net cell association measurements with weighted average predictions using protein corona data.

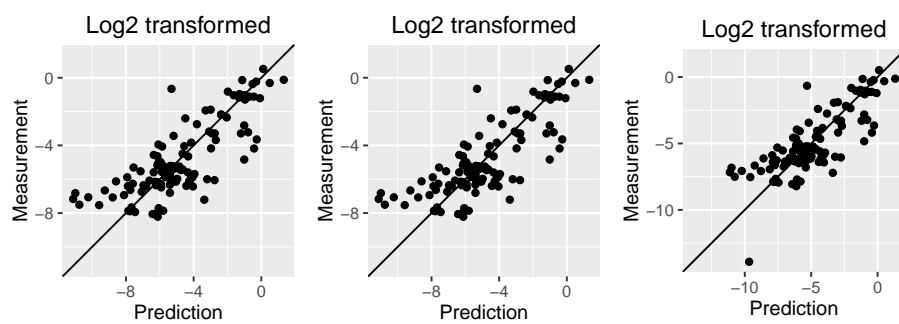


Figure 5: Correlation of log2 transformed net cell association measurements with partial least squares predictions using protein corona data.

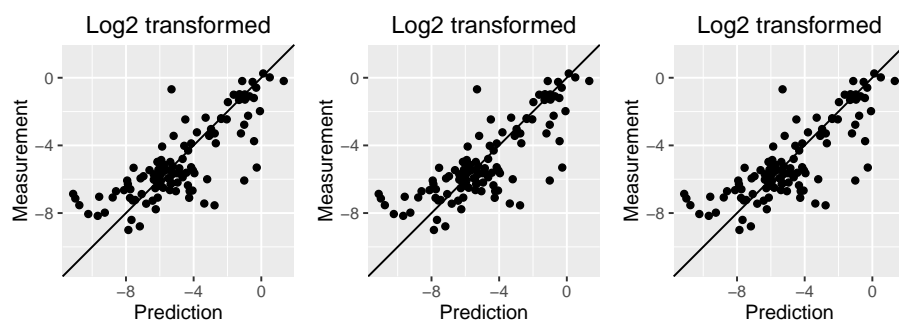


Figure 6: Correlation of log2 transformed net cell association measurements with random forest predictions using protein corona data.

Table 3: Repeated crossvalidation results for models with physchem properties and protein corona data, ** best results of all experiments, * no statistically significant difference to best results ($p > 0.05$)

Algorithm	r^2	RMSE
Weighted average	0.73, 0.66, 0.66 *	1.41, 1.57, 1.58 *
Partial least squares	0.67, 0.64, 0.69 *	1.53, 1.63, 1.5 *
Random forest	0.69, 0.69, 0.7 **	1.51, 1.5, 1.46 **

Discussion

Liu paper:

descriptor selection not included in cv!! prediction accuracy != r^2 uses bootstrap and strange r^2 which includes training set performance

all papers: no silver particles

Conclusion

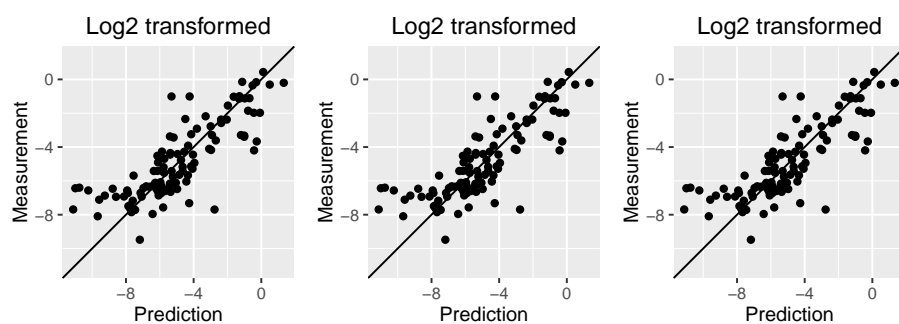


Figure 7: Correlation of log2 transformed net cell association measurements with weighted average predictions using physchem properties and protein corona data.

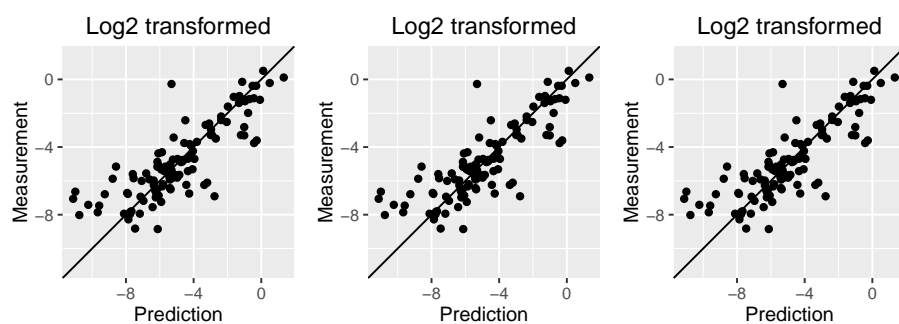


Figure 8: Correlation of log2 transformed net cell association measurements with partial least squares predictions using physchem properties and protein corona data.

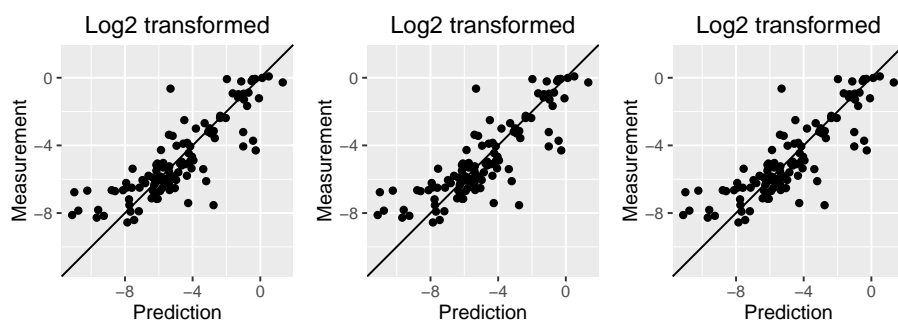


Figure 9: Correlation of log2 transformed net cell association measurements with random forest predictions using physchem properties and protein corona data.