

# Towards predictable 5'-UTRs in *Saccharomyces cerevisiae*

Thomas Decoene<sup>a</sup>, Gert Peters<sup>a</sup>, Sofie L. De Maeseneire<sup>b</sup> & Marjan De Mey<sup>a</sup>

<sup>a</sup> Centre for Synthetic Biology, Ghent University, Coupure links 653, 9000 Ghent, Belgium

<sup>b</sup> InBio.be, Centre for Industrial Biotechnology and Biocatalysis, Ghent University, Coupure links 653, 9000 Ghent, Belgium

## Contact

E-mail: Thomas.Decoene@UGent.be  
Tel: +32 9 264 93 83

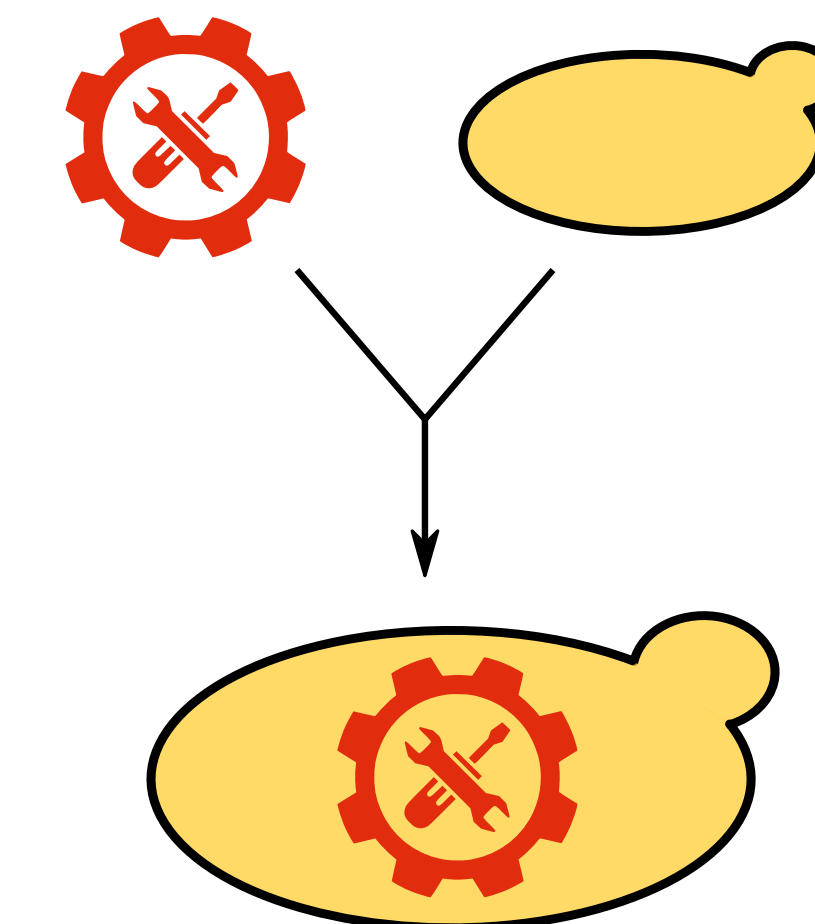
## Introduction

### Need for tools to fine tune pathways in yeast

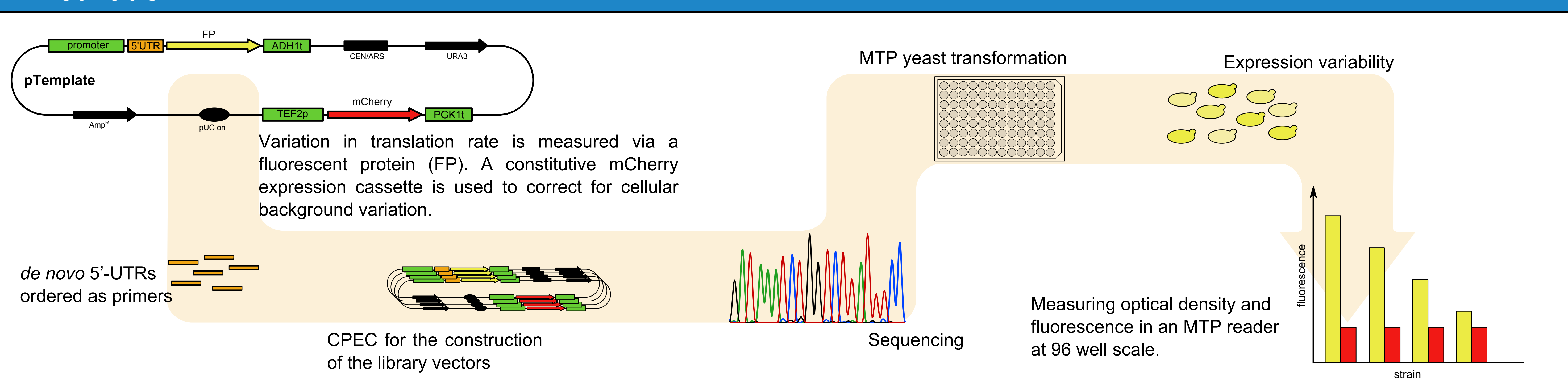
The ability to precisely fine-tune novel biosynthetic pathways in living cells is a fundamental requirement for the fast development of new microbial cell factories (MCFs). Typically, protein levels are adjusted through engineering of the transcriptional and translational level. Besides efforts in yeast on the transcriptional level like the creation of (synthetic) promoter and terminator libraries, the development of synthetic transcription factors and the modification of transcription factor binding sites, less progress has been made in modifying the translation rate of genes for the predictive regulation of protein levels.

### Varying gene translation rates

We developed a framework in *S. cerevisiae* for the *de novo* design of 5'-untranslated regions (5'-UTRs) with a predictive outcome on translation initiation rate. Therefore, we developed a partial least square (PLS) regression model based on a dataset of 5'-UTR sequences of Dvir *et al.* (2013) [1]. Next, this model was employed for the creation of completely new 5'-UTRs leading to protein abundances (PA) within a specified target range. Finally, *in silico* generated 5'-UTRs with different combinations of promoters, 5'-UTRs and coding sequences were evaluated *in vivo*.



## Methods



## Results

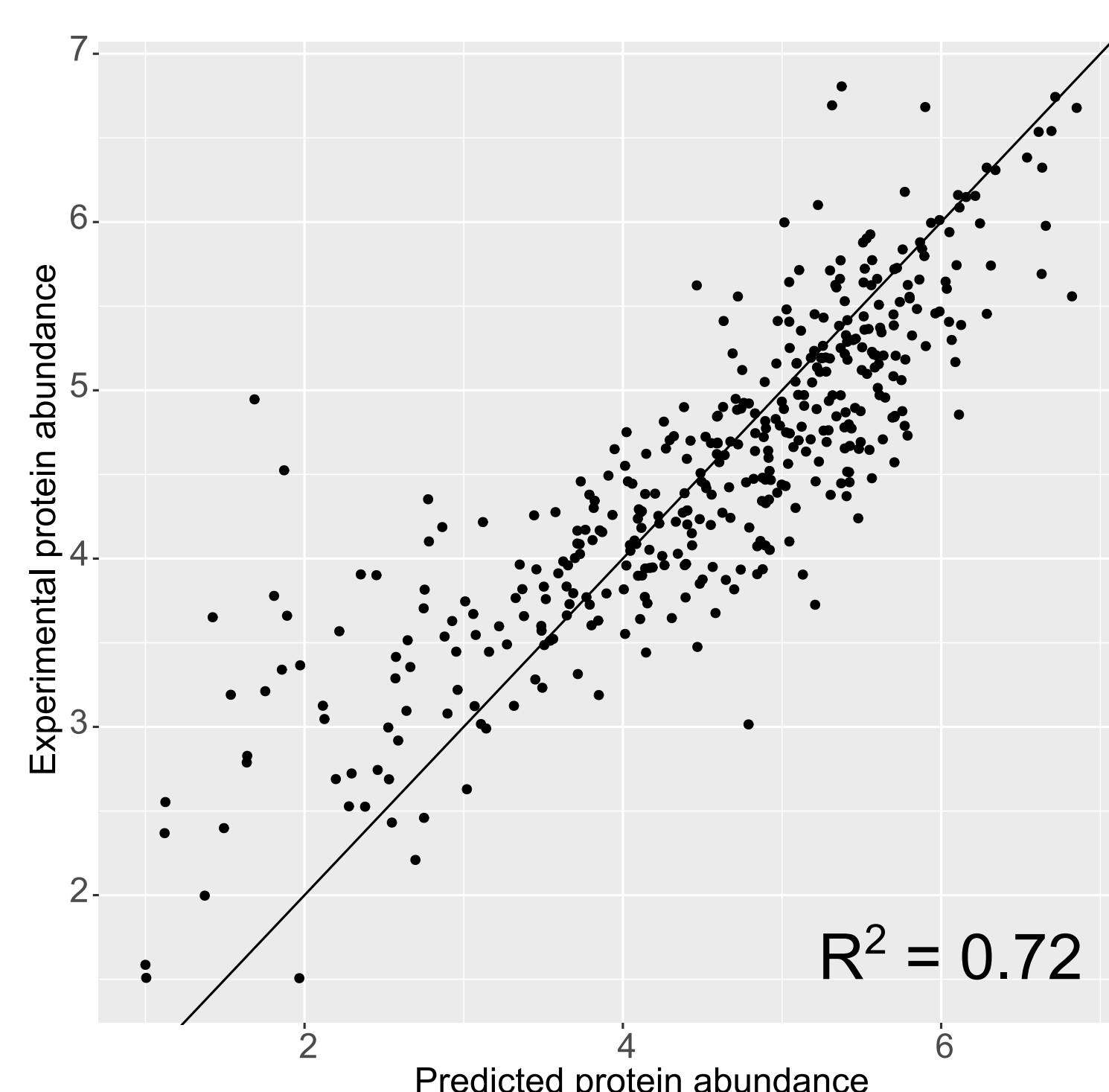
### PLS regression model [2]

$$Y_j = \beta_0 + \beta_1 X_{j,1} + \dots + \beta_k X_{j,k} + \epsilon_j$$

(k varies from 1 to 13)

Y: protein abundance  
X: 5'-UTR feature

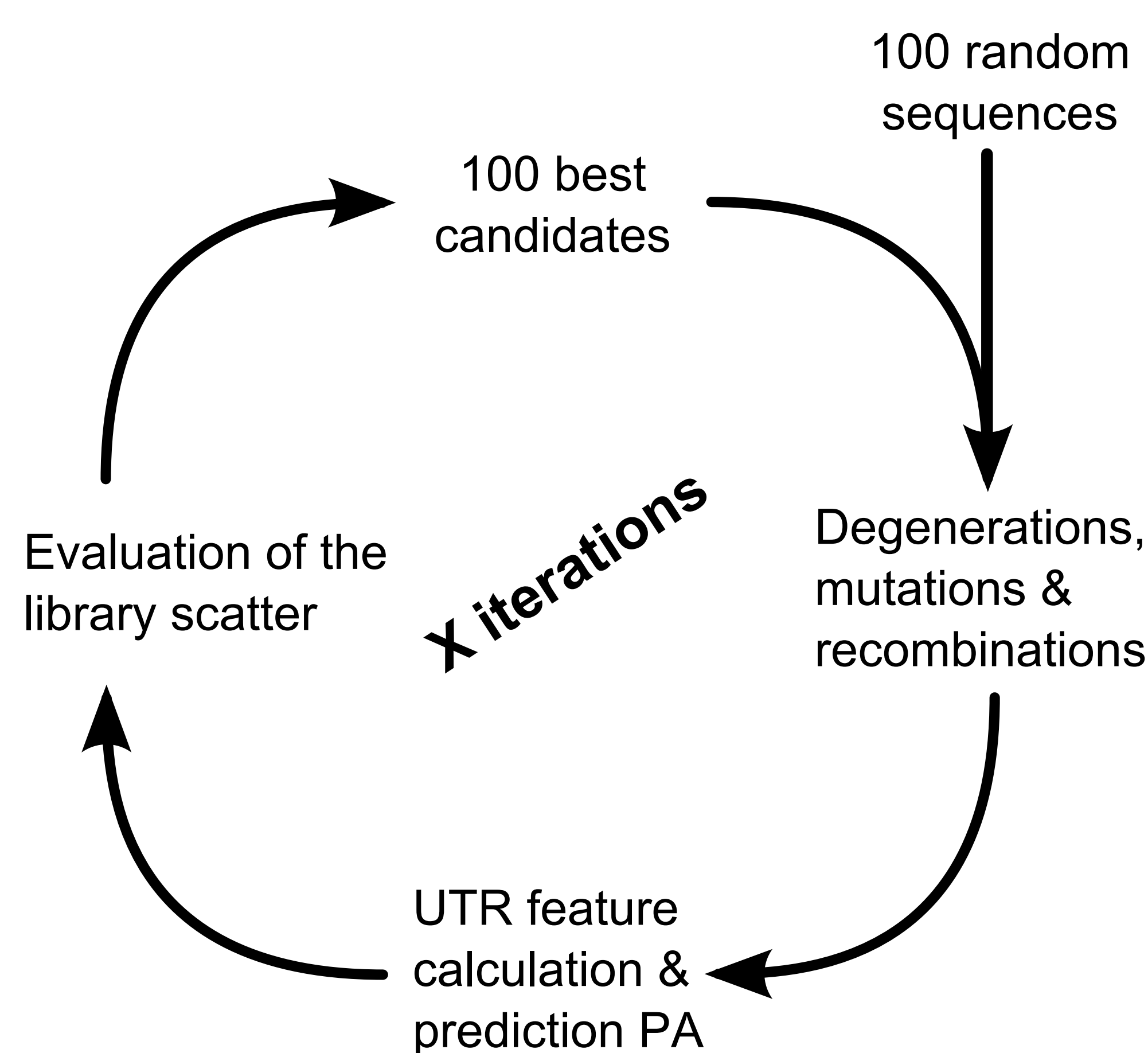
For the model construction, a dataset of 2041 RPL8A 5'-UTR sequences [1] was used. This dataset was split in a training set to build the model and a test set for model validation.



An  $R^2$  of 0.72 was obtained for the test set, indicating that our PLS model has the potential to successfully predict protein abundance.

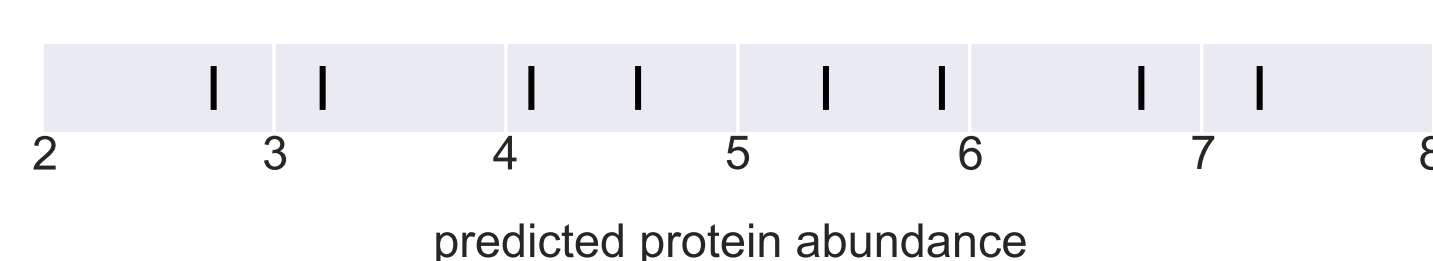
### *in silico* 5'-UTR design

The design of novel 5'-UTR sequences with a predictive outcome of protein expression levels is performed by an iterative process.

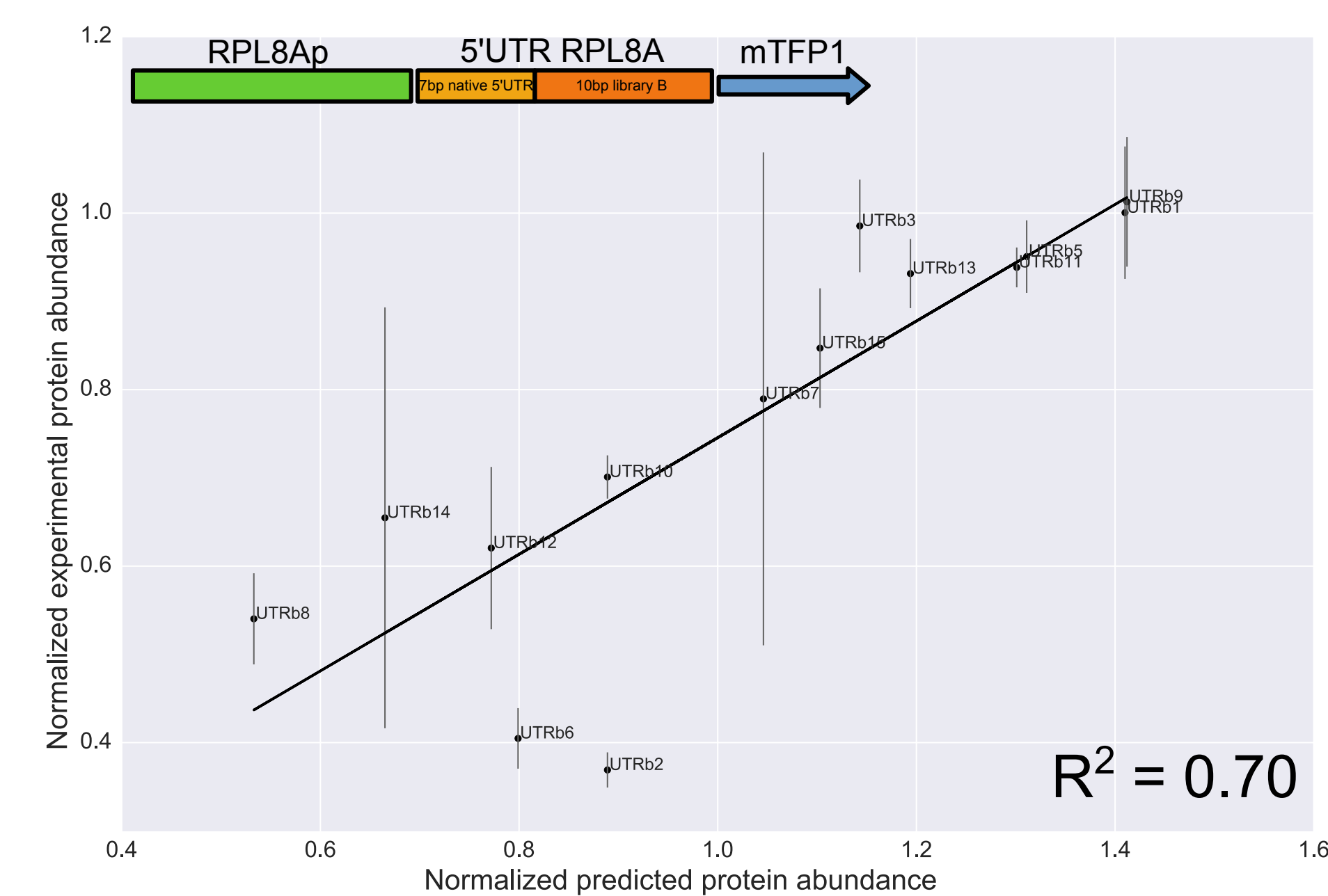
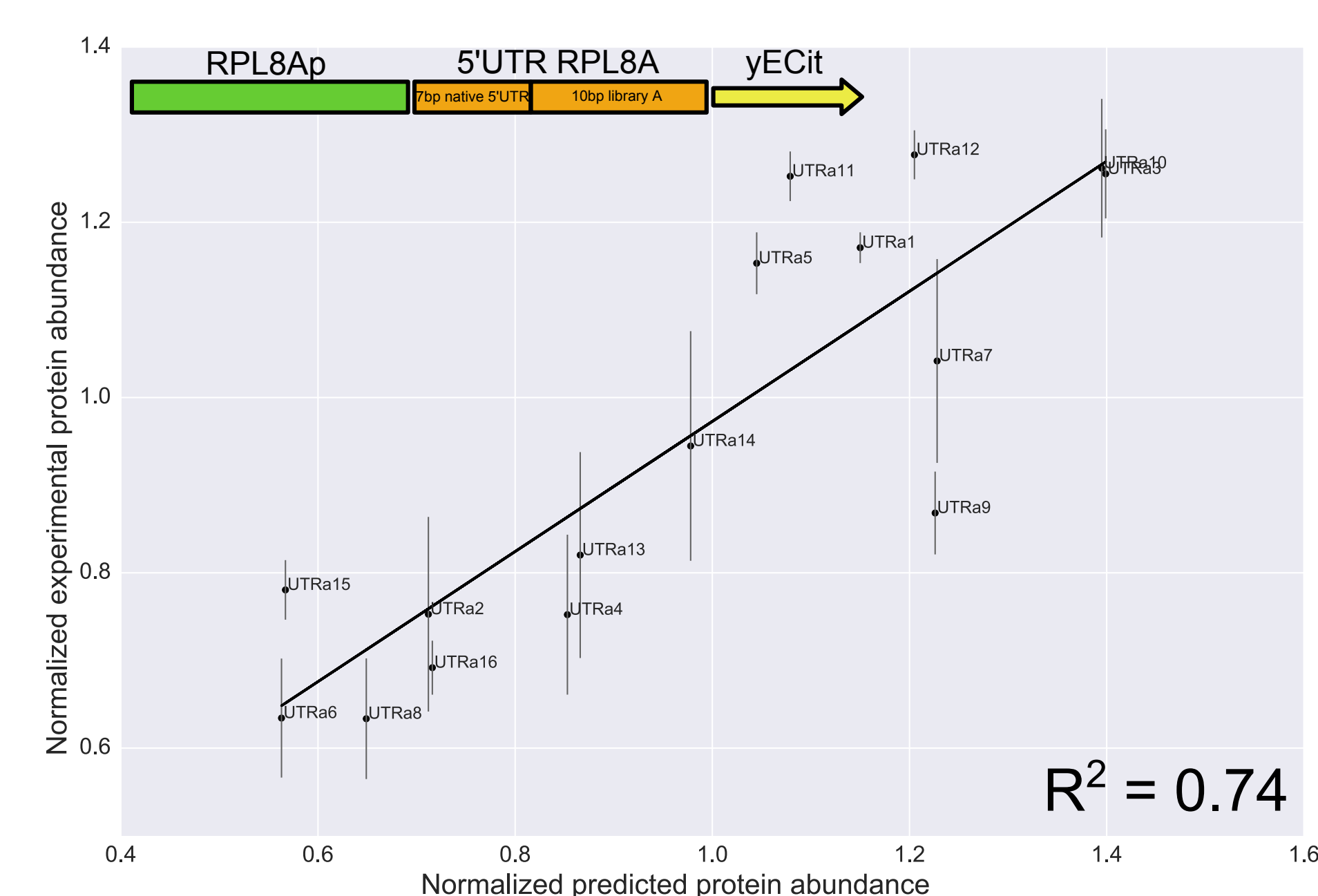


As output, 2 libraries of eight 5'-UTRs covering the wanted protein abundance range is given.

Example of the scatter of an *in silico* generated library.



### *in vivo* validation of new 5'-UTRs



[Error bars represent the standard deviation (n = 4)]

In both cases, an  $R^2$  of 0.70 or higher was obtained, indicating the predictive power of our 'yUTR calculator'. In future work, the generality of our model will be tested with other promoters and UTRs.

## Acknowledgements

The authors wish to thank Flanders Innovation & Entrepreneurship (Agentschap INNOVEREN & Ondernemen) for financial support in the framework of the Ph.D grant of T. Decoene.

## References

- [1] Dvir, S., Velten, L., Sharon, E., Zeevi, D., Carey, L. B., Weinberger, A., & Segal, E. (2013). *PNAS*, 110 (30), e2792-801.
- [2] Dayal, B. S., & MacGregor, J. F. (1997). *Journal of Chemometrics*, 11 (1), 73-85.