**Title of Project**

Developing a frugal transcription factor relative affinity measurement pipeline (TRAMP)

**Primary contact for the team**

Yaomin Cai, Engineering Biology, Earlham Institute, [Yaomin.Cai@earlham.ac.uk](mailto:Yaomin.Cai@earlham.ac.uk)

**Team**

Yaomin Cai ([Yaomin.Cai@earlham.ac.uk](mailto:Yaomin.Cai@earlham.ac.uk), Postdoctoral researcher, Engineering Biology, Earlham Institute)

Yaomin is a synthetic biologist, he will design and set up TRAMP and optimise parameters for TRAMP.

Will Nash ([will.nash@earlham.ac.uk](mailto:will.nash@earlham.ac.uk),Postdoctoral researcher, Bioinformatician/Evolution biology, Earlham Institute)

Will is responsible for designing a bioinformatic pipeline that allows to create new TFBSs and analyse their DNA structure features.

Susana Sauret-Gueto ([ss2359@cam.ac.uk](mailto:ss2359@cam.ac.uk), OpenPlant Research Manager, Prof Haseloff Lab, Plant Sciences Department, University of Cambridge)

Susana and Eftychis have rich experiences on Loop assembly system. They are also compiling a Marchantia DNA toolkit of Loop parts. Susana will design and assembled constructs and screen and characterrise Marchantia transformants.

Eftychios Frangedakis ([ef391@cam.ac.uk](mailto:ef391@cam.ac.uk), Research Associate, Prof Haseloff Lab, Plant Sciences Department, University of Cambridge):

Eftychis is setting up pipelines for throughput transformation and tissue culture of Marchantia. He will design and assembled constructs and transform and screen Marchantia.

**Summary**

Plant synthetic biology applications require the availability of a versatile set of promoters with different strength or specificity. Currently the number of promoters for plant synthetic biology is extremely limited. Our goal is to develop a new method that will facilitate the design of synthetic promoters with varying strength or specificity and will be a useful tool for genome engineering.

It has been shown that both the activity strength and the specificity of a promoter are heavily influenced by the affinity of a transcription factor (TF) to its binding site (TFBS) (Farley et al. 2015; Farley et al. 2016). Techniques to assess the affinity of a TF to its TFBS, such as the protein binding microarray (PBM), are useful tools for identifying novel TFBS and comparing TFs binding affinities (Berger & Bulyk 2006). However, such techniques are still not widely affordable. Microplate-based protein-DNA binding assays have been developed as a cheaper alternative, with the only limiting factor being the cost of the necessary DNA modifications.

In this project, we aim to establish a frugal platform for comparing relative TF-TFBS binding affinities by utilising a non-modified DNA immobilising method. This method is based on the commercial DNA coating solution (Thermo and Abcam) which has been used in the chemiluminescence immunoassay to assess DNA adducts levels (van Gijssel et al. 2002). We plan to proceed in two main stages:

1) Setting up a transcription factor relative affinity measurement pipeline (TRAMP) and optimising its parameters.

2) Demonstration of the usage of TRAMP by applying it to a practical project designing a set of promoters with different activity strengths and function in different species.

**Proposal**

**TRAMP Method in Principle:**

Initially, the coding sequence of TFs of interest will be synthesised or cloned from cDNA with a His6 tag and a HiBit tag (a 11 AA sequence, Promega) which will allow purification and quantification respectively. The concentration of the purified protein will be quantified using the HiBit system to calculate the total input of protein (Pinput).

To measure TF-TFBS binding affinities, double-stranded DNA containing potential TFBSs of corresponding TFs will be coated and immobilised (Dimb) onto the surface of a 96-well microplate with DNA coating solution (Fig. 1 Step I). Purified TFs (Pinput) will be added to the 96-well microplate and incubated in order to bind to the DNA (Fig. 1 Step II). Then, unbound TFs (Pfree) will be removed with a washing buffer leaving only the specific TFs bound (Pbound) to the DNA (Fig. 1 Step III). The bound TFs (Pbound) will be measured using the HiBit system (Fig. 1 Step IV). Finally, a DNA staining dye (picogreen, Thermo) will be added to the plate to assess the unbound DNA (Dfree) (Fig. 1 Step V).

This method will allow to:

1. Qualitatively compare TF-TFBS binding preferences expressed as Pbound/Dimb.

2. Quantifitatively compare TF-TFBS binding affinities by calculating the binding constant Kb= Pbound/ Dfree \* Pfree (Pfree = Pinput - Pbound).



Figure 1. The working procedure of TRAMP. Step I, immobilise DNA to wells. Step II, incubating TFs with DNA. Step III, remove unbound TFs. Step IV, measure bound TFs. Step V, measure unbound DNA.

**Stage I. Set up TRAMP**

A pair of known TF/TFBS, *Arabidopsis thaliana* TGA1 (AT5G65210) / NVTGACGHNN (IPUAC DNA codes) will be used to validate TRAMP. The *TGA1* gene as well as genes used in the following stages will be synthesised or cloned from cDNA. The TGA1 protein will have a His6 tag attached at the N-terminal and a HiBit tag at the C-terminal. The recombined TGA1 will then be expressed by an *E. coli* expression system and purified with HisPur™ Cobalt Resin (Thermo). The TGA1 TFBS double-stranded DNA will be made with annealing single-stranded oligonucleotides. In order to achieve the best performance, several parameters need to be optimised.

**Risks and solutions**

1. **Plate choice**. The coating solution is compatible with both polystyrene and polypropylene surface. Both types of plates will be tested for coating efficiency.
2. **The coating efficiency**. The DNA molarity and length are two main factors affecting the coating capacity. To optimise the combination of DNA moles and length, an array with DNA of 0.1, 0.2,0.5,1, 5, 10, 20, 40pmol and 15, 20, 25, 30, 35, 40, 45, 50, 55, 60 bp will be tested. Each oligo will have the TGA1 TFBS in the centre flanked by random sequences. Random sequences will be scanned to ensure the absence of accidently introduced TFBS like motifs.
3. **Coating variation**. Variation may occur during coating from batch to batch. To minimise the variation, a saturated amount DNA will be used to coat the microplate.
4. **Quantify Dfree**. Since TF may not be able to bind the entire DNA sequence motif, there is still a short part of DNA that can be bound by picogreen. So, the shortest possible motifs, according to the results in comparing coating efficiency, will be used to maximise the differences between TF bound and unbound DNA.

**Stage II. Demonstration of usage of TRAMP**

**Design bioinformatics pipeline**

Design of a pipeline for the generation of TFBSs sets with a broad range of affinities to its corresponding TFs. The proposed in silico analysis and design of new motifs will be conducted in the following stages:

* 1. Generation of all possible sequences which match the position weight matrix (PWM) of testing TFs motifs.
  2. Removal of sequences which recruit TFs different from the target by checking against the PlantTFDB PWM database (Jin et al. 2017).
  3. For each sequence, prediction of 5 biophysical metrics determining DNA shape features (DNA 3D structure features) using DNAshapeR (Zhou et al. 2013).
  4. Analysis the diversity of the feature patterns of all motifs, and combinations of motifs.
  5. Curation of a representative test set which spans the broadest range of feature patterns calculated in the previous stage, but remains within the frugal framework of the proposal (~2 x 96 test sequences).

This process will be followed by use of the TRAMP assay developed in Stage I to allow comparison of the binding affinities of the TFBS set.

**Risks and solutions**

1. **Unknown effects of nucleotide variation on DNA shape**. The effects of nucleotide variation on the DNA shape parameters we propose to measure has not been tested in this context. Several dummy motifs will be used to estimate the effect of altering nucleotides to the integrities of the motifs in silico.
2. **Alteration of binding sites may attract competitive binding.** It will be important to consider the effect on alteration of motif sequences on the structure of the promoter sequence as a whole. Changing the nucleotides in a motif may break its affinity with one TF and create a novel motif in its place, or in combination to the nucleotide context of the flanking sequences. The pipeline will be designed to take this into account and all sequences will be scanned against all known PWMs from the PTFBD prior to consideration as an experimental target.
3. **Motif proximity/complexity/diversity will vary in other systems.** Although we may be able to test a range of *A. thaliana* regulatory sequences, it will be essential to make our analytical approach generalisable, and as such able to apply to any promoter. As an example, we envisage scenarios where motifs in close proximity have influence on the biophysical properties of their neighbours. We will address this by designing scripts to utilise the potentially large amounts of regulatory sequence variants it will be possible to generate to maximise the variation in binding affinity tested using TRAMP. This will be accompanied by reporting of the levels of variation generated. It will also be important to design thresholds in the scripts which will filter low quality motifs, and also report promoters which do not generate a wide variation in DNA shape parameters.

**Test new promoters**

The target promoter used in the Stage II is the CaMV35S promoter (pro35S). Pro35S has different versions with diverse performances in different organisms. The cause of these variations is still elusive. In the Haseloff lab, two versions of pro35S have been tested in Marchantia, concluding that pro35S\_BL used in previous studies (Boisnard-Lorig 2001; Ishizaki et al. 2016) are stronger than the long version of pro35S from the MoClo kit (pro35SLong). Interestingly, the same TFs seem to recognise the same TFBSs between pro35SLong and pro35S\_BL (Fig. 2). However, sequences of those TFBSs contain several mismatches (Fig. 2). We reason that mutations in TFBSs is one important factor that causes different performances of pro35SLong and pro35S\_BL in Marchantia. By using the bioinformatics pipeline which will be developed, we will design new TFBSs with different DNAshape features. The DNAshape features have been shown to influence the TFBS binding affinity (Zhou et al. 2013; Käppel et al. 2018). New TFBSs’ affinities will be tested using TRAMP. The best TFBSs candidates, will be used to synthesise new pro35S versions and test their activities in Marchantia.



Figure 2. Comparison of pro35SLong and pro35S\_BL sequences. Diagrams of pro35SLong and pro35S\_BL are aligned. Blue box with grids are TFBSs identified by FIMO and Arabidopsis Cistrome database(O’Malley et al. 2016). The sequence alignment of TFBSs between pro35SLong and pro35s\_BL are shown under the diagram. Mismatches are highlighted with a pink background. TFs are listed above corresponding TFBSs. Green frames are Arabidopsis TFs, orange frames are the orthologous in Marchantia.

More specifically:

1. Synthesising new promoters. In order to synthesise as many promoters as possible, motif sequences and non-changing sequences will be synthesised separately. The full promoter sequences will be assembled with new motifs and non-changing sequences by the Loop type IIS assembly (Pollak et al. 2018 under review). The principle is illustrated in Figure 3.



Figure 3. promoter assembly. Promoters which are to be tested can be divided into changeable (motifs) and non-changeable regions. Non-changeable sequences only need to be synthesised once. Motifs can be synthesised as primers with appropriate type II restriction enzyme cloning site (e.g. BsaI). The whole promoter will be assembled into a standard L0 acceptor and then assembled with a reporter and other necessary parts into a transcriptional unit through the Loop system.

1. Testing new promoters in Marchantia. New promoters will be used to drive HygR or a fluorescence protein marker in Marchantia. It is possible to assess the activity of the promoters in planta in about 3 weeks.

**Timeline**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Month | 1 | 2 | 3 | 4 | 5 | 6 |
| Set up TRAMP |  |  |  |  |  |  |
| Design new motifs |  |  |  |  |  |  |
| Measure affinity with TRAMP |  |  |  |  |  |  |
| Synthesis new promoter and test in Marchantia |  |  |  |  |  |  |

**Benefits and outcomes**

Describe how your project fits the remit of OpenPlant and the judging criteria, including details of any new interdisciplinary interactions between Cambridge and Norwich.

1. A frugal method to compare TF/TFBS binding affinity.
2. Binding affinities of a set important TF/TFBS which recognise 35s promoters.
3. Identify sources of variations of pro35S activity in different species.
4. Lead to publications.
5. This proposal has already promoted interdisciplinary working between the Earlham Institute and The University of Cambridge, having begun a collaboration between the Synthetic Biology, Evolutionary Genomics and Plant Synthetic Biology groups.
6. An open protocol of TRAMP and an open-source bioinformatics pipeline to design TFBSs.

**Sponsor for the research and cost centre**

Dr. Nicola Patron

Earlham Institute

[Nicola.Patron@earlham.ac.uk](mailto:Nicola.Patron@earlham.ac.uk)

I confirm that I have the full support of the sponsor listed above and that they can be added to the OpenPlant Fund mailing list to receive project updates (to which they can unsubscribe at any time).

**Budget**

|  |  |
| --- | --- |
| **Material** | **Cost (£)** |
| **DNA** **immobilising** |  |
| Polystyrene plate x30: Vision PlateTM 96 Well, Non-Sterile (4ti-0224, 4titude) | 156.5 |
| Polypropylene plate x80: PlateOne microplate, flat bottom (S1837-9600, Starlab) | 148.9 |
| DNA coating solution 100mL (17250, Thermo) | 86.18 |
| DNA coating solution 30mL (ab156917, Abcam) | 109 |
| TGA1 TFBSs | 150 |
| Picogreen (P7581, Thermo) | 351.73 |
| **Protein expression and purification** |  |
| Cloning TF genes | 342 |
| TGX stain free gel, pkg of 10 (4568123, Bio-Rad) | 107 |
| HisPur cobalt resin 10mL (89964, Thermo) | 79.71 |
| **Detection** |  |
| \*pro35s TFBSs variants | 1843 |
| Nano-Glo® HiBiT Lytic Detection System, 10mL (N3030, Promega) x3 | 342 |
| **Synthesis of promoter** |  |
| Synthesis of non-changeable sequences | 180 |
| Assemble promoters (Enzymes, plasmid prep kit etc.) | 100 |
| **Total** | 3996.02 |

\*the cost of pro35s TFBSs variants varies dependent on the optimum length of TFBS for TRAMP. This table shows the maximum cost.

**Reference**

Berger, M.F. & Bulyk, M.L., 2006. Protein Binding Microarrays (PBMs) for Rapid, High-Throughput Characterization of the Sequence Specificities of DNA Binding Proteins. In *Gene Mapping, Discovery, and Expression*. New Jersey: Humana Press, pp. 245–260. Available at: http://link.springer.com/10.1385/1-59745-097-9:245 [Accessed June 13, 2018].

Boisnard-Lorig, C., 2001. Dynamic Analyses of the Expression of the HISTONE::YFP Fusion Protein in Arabidopsis Show That Syncytial Endosperm Is Divided in Mitotic Domains. *the Plant Cell Online*, 13(3), pp.495–509. Available at: http://www.plantcell.org/cgi/doi/10.1105/tpc.13.3.495.

Farley, E.K. et al., 2015. Suboptimization of developmental enhancers. *Science (New York, N.Y.)*, 350(6258), pp.325–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26472909 [Accessed May 18, 2018].

Farley, E.K. et al., 2016. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proceedings of the National Academy of Sciences*, 113(23), pp.6508–6513. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27155014 [Accessed May 18, 2018].

van Gijssel, H.E. et al., 2002. Semiquantitation of polycyclic aromatic hydrocarbon-DNA adducts in human esophagus by immunohistochemistry and the automated cellular imaging system. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 11(12), pp.1622–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12496053 [Accessed July 6, 2018].

Ishizaki, K. et al., 2016. Molecular Genetic Tools and Techniques for *Marchantia polymorpha* Research. *Plant and Cell Physiology*, 57(2), pp.262–270. Available at: https://academic.oup.com/pcp/article-lookup/doi/10.1093/pcp/pcv097 [Accessed July 6, 2018].

Jin, J. et al., 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, 45(D1), pp.D1040–D1045. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw982 [Accessed July 4, 2018].

Käppel, S. et al., 2018. The floral homeotic protein SEPALLATA3 recognizes target DNA sequences by shape readout involving a conserved arginine residue in the MADS&#x2010;domain. *The Plant Journal*. Available at: http://doi.wiley.com/10.1111/tpj.13954 [Accessed June 5, 2018].

O’Malley, R.C. et al., 2016. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape (Cell (2016) 165(5) (1280???1292)). *Cell*, 166(6), p.1598. Available at: http://dx.doi.org/10.1016/j.cell.2016.04.038.

Pollak, B. et al., 2018. Loop Assembly: a simple and open system for recursive fabrication of DNA circuits. *bioRxiv*, p.247593. Available at: https://www.biorxiv.org/content/early/2018/01/18/247593 [Accessed July 6, 2018].

Zhou, T. et al., 2013. DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic acids research*, 41(Web Server issue), pp.56–62.