# Identification of epigenetic readers by DAPPL

**Guang Song**
Johns Hopkins University School of Medicine https://orcid.org/0000-0002-7630-716X
**Guohua Wang**
Harbin Institute of Technology
**Ximei Luo**
Harbin Institute of Technology
**Ying Cheng**
Emory University School of Medicine
**Qifeng Song**
Johns Hopkins University School of Medicine
**Jun Wan**
Johns Hopkins University School of Medicine
**Cedric Moore**
Johns Hopkins University School of Medicine https://orcid.org/0000-0002-3459-392X
**Hongjun Song**
University of Pennsylvania https://orcid.org/0000-0002-8720-5310
**Peng Jin**
Emory University School of Medicine
**Jiang Qian**
Johns Hopkins University School of Medicine https://orcid.org/0000-0002-0476-3596
**Heng Zhu** ( hzhu4@jhmi.edu )
Johns Hopkins University School of Medicine

**Method Article**

**Keywords:** Epigenetic marks, Transcription factor, DNA modification, Methylcytosine, Hydroxymethylcytosine, Formylcytosine, Carboxylcytosine

**Posted Date:** May 26th, 2021

**DOI:** https://doi.org/10.21203/rs.3.pex-1286/v1

# Abstract

Identifying readers of epigenetic marks on DNA is a critical step towards understanding the mechanisms of those modifications in many biological processes. Here, we present an all-to-all approach, dubbed digital affinity profiling via proximity ligation (DAPPL), to simultaneously profile readers for different epigenetic modifications (i.e., 5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine) on CpG dinucleotides using random DNA libraries. Using specific DNA fragments to covalently barcode each human TFs and co-factors, we could connect the identity of a protein to its captured DNA fragments via proximity ligation in a highly multiplexed. Using this approach, we identified numerous readers for different DNA epigenetic modifications

# Procedure

Purification of human transcription factors in 96-well format

1. To prepare the SC-Ura/glucose plates, 1.7 g yeast nitrogen base without amino acids, 5 g ammonium sulfate, 20 g glucose, 2 g amino acid dropout mixes and 20 g agar were mixed and dissolved into 1 L dd$H_2O$. Then the medium was autoclaved at 121°C for 15 minutes.

Pour ~20mL of SC-Ura/glucose agar per 10 cm polystyrene Petri dish.

2. Take out 10 ul of each human ORF expression yeast clone and draw it on the SC-Ura/glucose plates, then put the plates in 30°C incubator for 48 hours.

3. Each of the yeast clones was transferred to 800 µL of SC-Ura/glucose liquid medium (1 L medium contains 1.7 g yeast nitrogen base without amino acids, 5 g ammonium sulfate, 20 g glucose, 2 g amino acid dropout mixes) in deep 96-well plates overnight as primary cultures.

4. 50 µL of the above saturated yeast cultures were inoculated into 8 mL of SC-Ura/raffinose liquid medium (1 L medium contains 1.7 g yeast nitrogen base without amino acids, 5 g ammonium sulfate, 20 g raffinose, 2 g amino acid dropout mixes) in 12 channel plates until the O.D. 600 reached 1.0, following with the induction by adding 2% galactose for 6 hours at 30°C.

5. Yeast cells were harvested by centrifugation at 3,500 x g for 5 min. Each cell pellet was suspended with 1 ml precool water and transferred to the deep 96-well plates. After centrifuging at 3,500 x g for 10 min and discarding the supernatant, the yeast cell pellets were stored at -80°C for further study.

6. Take out the yeast cell pellets in deep-well plates from at -80°C and put them on ice.

7. Add ~250µl zirconia beads (0.5 mm) and 500 µl lysis buffer (50 mM Tris-HCl at pH7.5 with 100 mM NaCl, 1 mM EGTA, 0.01% tritonX-100, 0.1% beta-mercaptoethanol, 1 mM PMSF, and Roche protease inhibitor tablet [Roche]) to each well of the yeast plate.

8. Add a capmat to cover each box, and then lysate the yeast cells using the Harbil paint shaker for 2 min X 8 times with interval of 2 min on ice.

9. The cell lysates were centrifuged at 3,500 g for 10 min at 4°C; the supernatants were transferred to a filter plate with prewashed 50µl glutathione beads (GE Healthcare),

10. Add an additional 500µl of lysis buffer to the yeast pellet plates, and then repeat the step 8, the supernatants were combined with the previous lysates in step 9.

11. Seal the top of the filter plate tightly with adhesive foil and stack horizontally on a platform shaker at 4°C overnight.

12. The beads were washed three times with wash buffer I (50 mM HEPES at pH7.5 with 500 mM NaCl, 1 mM EGTA, 10% glycerol, and 0.1% beta-mercaptoethanol) and three times with wash buffer II (50 mM HEPES at pH7.5 with 100 mM NaCl, 1 mM EGTA, and 10% glycerol). All TFs captured by glutathione beads were stored at -80°C.

Conjugation of anchor oligo to each protein in 96-well format

13. Maleimide Modifier modified (5') Anchor oligo was ordered from Genelink. Please note that the cycloadduct should be treated via a retro-Diels-Alder reaction to active maleimide using toluene and heat.

14. Suspend lyophilized oligo in 1 mL anhydrous Acetonitrile. It will be a suspension as oligo will not dissolve.

15. Evaporate using a speedvac.

16. Re-suspend oligo in 1.5 mL Toluene.

17. Incubate for 4 hrs at 90°C.

18. Cool to room temperature.

19. Evaporate toluene using a speedvac. Now, the oligonucleotide is now ready for conjugation. Please note that the retro-Diels-Alder reaction should be performed immediately prior to conjugation.

20. Take out the purified TF in 96-well filter plates from at -80°C and put them on ice.

21. The lyophilized oligos were dissolved in PBS buffer to a final concentration of 50 µM and added to each purified protein on glutathione beads arrayed in 96-well filter plates.

22. The conjugation was achieved using a “click” chemistry reaction between the hydro sulphonyl group on cysteine residues of the proteins and a maleimide group tethered to the 5’-end of the anchor oligo at room temperature for 1 hour.

23. The free oligos were removed with three stringent washes (50 mM HEPES at pH=7.5 with 100 mM NaCl, 1 mM EGTA, 10% glycerol, and 0.1% beta-mercaptoethanol) and the anchor oligo-conjugated TFs were stored at -80° C until use.

Assignment of DNA barcodes to TF proteins

24. A collection of 2,000 address oligos (Integrated DNA Technologies) were synthesized, each of which containing a BsaI recognition site and a cutting site (5’-GGTCTCCGACT) at the 5’-end, a 8 nt random sequence as unique molecular identifiers (UMI), a 7-11 nt unique DNA barcode sequence, and a 20-nt consistent sequence, complementary to the anchor oligo.

25. The address oligos were prepared in concentration of 100 μM in $ddH_2O$.

26. Six microliters of each of the above address oligo were used to individually prepare the 50 μl Klenow reaction mixture (1 X NEB Buffer 2 with 0.6 mM dNTP mix, 1 U Klenow [New England Biolabs], and 12 μM address oligos).

27. Add the 50 μl Klenow reaction mixture for each address oligo to the Anchor oligo conjugated TFs on glutathione beads in 96-well filter plates.

28. Klenow polymerase reaction was performed at 37°C for 30 min to synthesize the complementary strands.

29. Free address oligos were removed with three washes (50 mM HEPES at pH=7.5 with 100 mM NaCl, 1 mM EGTA, 10% glycerol, and 0.1% beta-mercaptoethanol).

30. One fifth of the bead slurry for each protein was mixed to prepare the barcoded TF column.

Preparation of epigenetic modified DNA library

31. The template oligos of epigenetic modified DNA libraries were synthesized to encode a 5’ Primer 2 (5’-CACATCCTTCACATTAATCC), an 18-mer sequence with a modified CpG in the middle flanked by two 8-mer random sequences, a short sequence encoding the library identity, and a *Bsa*I recognition site and its

cutting site (Table 1). The both 8-mer random regions of template DNA oligos was synthesized in an A:C:G:T ratio of 30:30:20:20, because this ratio is known to provide a more equal distribution of the four bases. In addition, a three-nucleotide barcode was also added to each sequence to identify the different libraries.

32. The double-stranded DNA libraries were generated by annealing the template oligo with a complementary primer oligo in a 1:1 ratio, followed by a Klenow polymerase reaction according to the manual (M0210, New England Biolabs), in the presence or absence of the modified dCTPs to create symmetric or hemi libraries, respectively. In addition, an unmodified DNA library of the same design was also created as a reference library.

33. Each library was purified with a QIAquick PCR Purification kit (Qiagen) according to the manufacturer's instructions.

34. The four libraries with four different symmetric modifications and the unmodified library were mixed in an equimolar ratio to form the mixture of symmetric modification libraries. A mixture of the hemi-modification libraries was created using the same method in parallel.

Identification of readers for different epigenetic DNA modifications

1. The barcoded TF columns were incubated with 200 nmole of the mixture of symmetric or hemi modification libraries, in a TF binding buffer (10 mM Tris-HCl at pH 7.5 with 50 mM NaCl, 1mM DTT, and 4% glycerol) at room temperature for 30 min.

2. After three stringent washes with binding buffer and 1 X PBS buffer, the protein-DNA complexes on the beads were crosslinked by 5 ml of 0.1% formaldehyde in PBS buffer for 10 min.

3. Centrifugation at 1,000 x g for 5 min was applied to remove the formaldehyde. Then, the beads were incubated with 10 ml of Tris-Glycine buffer (0.1 M glycine, pH 7.5) for 10 min to quench the residual formaldehyde in the system.

4. The beads were washed in 10 ml TBST buffer (0.01% tween-20) three times and equilibrated by 1 X T4 DNA ligase buffer. Then, remove all the buffer in the beads by centrifugation at 1,000 x g for 5 min.

5. Adding a master reaction mixture (227 µL of $ddH_2O$ with 30 µL 10 X T4 DNA ligase reaction buffer, 30 µg bovine serum albumin, 20 µL of 100 U of *Bsa*I [New England Biolabs], and 20 µL of 600 U T4 DNA ligase [Enzymatics]) to the beads and incubated for 1 hour at 37°C to perform the Golden Gate Assembly

reaction, which could connect the identity of a protein to its captured DNA fragments via proximity ligation in a highly multiplexed reaction.

6. Twenty microliter of Proteinase K was added to the beads and incubated for 1 hour at 37°C to digest proteins on beads to release the ligated DAPPL products.

7. Phenol/Chloroform extraction and QIAquick PCR Purification Kit were applied to clean the ligated DAPPL products according to their manuals.

8. The ligated DAPPL products were amplified by PCR using the common primers of the libraries.

9. All the PCR products were separated by 10% TBE gel and the target bands with correct size were cut and subjected to deep-sequencing on an Illumina NextSeq500 sequencer using the commercial kit.

10. The sequencing reads could be readily mapped back to each TF using specific unique barcode for each protein.

11. The binding sequences from a modified library and the unmodified library for each protein were compared and subjected to identification of candidate readers for different epigenetic marks. The binding consensus sequences (i.e., motifs) for each TF protein were computed using HOMER (http://biowhat.ucsd.edu/HOMER).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.jpg