

BunDLE-seq (Binding to Designed Library, Extracting and Sequencing) -

A quantitative investigation of various determinants of TF binding; going beyond the characterization of core site

Einat Zalckvar*^{1,2}, Michal Levo*^{1,2} and Eran Segal^{1,2}

* These authors contributed equally to this work

¹ *Department of Computer Science and Applied Mathematics*, ² *Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel*

Introduction

Deciphering the binding determinants of transcription factors (TFs) is fundamental to understanding the mechanisms underlying the formation of robust and timely gene expression patterns. Great advances in the characterization of TF binding sites were made in recent years, with the development of platforms for high-throughput and accurate *in vitro* binding measurements of TFs to thousands of short sequences. However, the complementary development of protocols for genome-wide TF binding measurements (*e.g.*, ChIP-chip and ChIP-seq) revealed that although some binding events are well accounted for by the underlying presence of an *in vitro*-deduced binding site, many gaps still remain. These instances demonstrate the need for a more comprehensive understanding of the various factors influencing TF binding to regulatory sequences, going beyond the characterization of core binding site.

Several recent studies that aim to address this gap employed *in vitro* based methods (*e.g.* DIP-seq, PB-seq, (gc-)PBMs, EMSA-seq, SELEX-seq, MITOMI, HiTS-FLIP), and identified various mechanisms that affect TF binding, including chromatin accessibility, cofactors binding that influences specificity, TF dimer interactions and the effect of sequences flanking the core TF binding site (TFBS) that can be mediated by DNA shape properties.

Here, we present a novel experimental approach, termed BunDLE-seq that allows the quantitative investigation of several determinants of TF binding within a single experiment. Specifically, the assay provides quantitative TF binding measurements to large-scale libraries of fully-designed sequences. Our assay is unique in its ability to study thousands of long and systematically designed sequences, and its capacity to isolate different states of TF binding formed on these sequences. Notably, the assay can be performed with both TFs and histone proteins. Furthermore, the assay is particularly appealing as a complementary method for the rapidly emerging protocols for high-throughput reporter assays, as the same library of sequence variants can serve as input to both *in vitro* binding and *in vivo* expression measurements, with the first thereby offering mechanistic insights on the latter.

Reagents

BunDLE-seq reaction buffer:

0.15 M NaCl

0.5 mM PMSF (Sigma)

1 mM BZA (Sigma)

0.5X TE (100X TE: 1M Tris (pH 8.0), 0.1 M EDTA (pH 8.0))

0.16 µg/µl Poly-L-glutamic acid (PGA) (Sigma)

TBE buffer:

Prepare a 5X stock solution in 1 L of H₂O:

54 g of Tris base

27.5 g of boric acid

20 mL of 0.5 M EDTA (pH 8.0)

Method

This protocol includes detailed information from the protein binding step up to the library preparation for high throughput sequencing. More on a specific application of BunDLE-seq can be found in Levo and Zalckvar et al. 2015.

Since DNA library synthesis and high throughput sequencing is constantly developing, we avoided overloading the protocol with specific information regarding these steps. Further information about DNA library description and preparation as well as mapping of the sequencing reads can be found in Levo and Zalckvar et al. 2015, Blecher-Gonen et al. 2013 Nature Protocols¹ and Sharon et al. 2012 Nature Biotechnology².

* Important –

Before starting make sure that your protein can bind DNA. If possible, conduct a simple EMSA experiment with a DNA sequence containing a known binding site, and a DNA without a binding site. Use this experiment to also establish the protein concentration with which you will carry out the experiment; try to minimize non-specific binding on the one hand, while still producing sufficient binding for low-to-medium affinity sites (i.e. accounting for the amount of DNA you will need to extract from the gel in order to proceed to high-throughput sequencing).

Consider using a protein without a tag.

Day 1

1. Incubate the BunDLE-seq reaction buffer in low binding tubes (Sorenson) for 2hr at room temperature while rotating.

* To enable loading on gel, the final volume of the reaction (including proteins and DNA) is 30 μ L.

When histone octamers are used, incubate them with the BunDLE-seq reaction buffer at this stage.

2. Cool the tubes for 30 min at 4°C.
3. To minimize non-specific binding, add 0.067 μ g/ μ L BSA (Sigma).
4. Add the purified protein (e.g. a transcription factor).
5. Add 200ng of the library DNA.
6. Incubate the proteins and DNA for 1 hr at 4°C.

* When Gal4 was used the tubes were not cooled, and the protein and DNA were incubated for 30 min at room temperature.

7. Add 6 μ L 15% Ficoll (Sigma) to each tube.

* Do not use loading dye. It might mask some of the bands.

* Do not use glycerol. The DNA will not run properly in gel.

8. Run the samples in 7.5% acrylamide gel in cold 0.25XTBE buffer.
 - * Use cold TBE buffer.
 - * Do not use higher concentrations of the TBE buffer.
 - * Load the samples while the gel is running. This minimizes the time of incubation of the samples in the wells, and thus reduces detachment of protein-DNA in the presence of the high salt-containing running buffer.
9. Stain the gel for 30 min with GelStar (Lonza), at room temperature.
10. Cut the bands under UVIblue blue light transilluminator (UVItec).
 - * Avoid using a UV table.
 - * Avoid cutting different bands with the same scalpel.
11. Elute the DNA from the gel slices using electroelution Midi GeBAflex tubes (Gene Bio-Application).
 - * To minimize GC bias avoid using a chaotropic salt-containing gel extraction kits.
12. Precipitate the DNA with 1 volume isopropanol, 1/10 volume 3M NaOAc (pH 5.2), and 1 μ g/ μ l glycogen (Fermentas) over night at -20°C.
 - * When ~150 bp DNA is used it is highly recommended to use glycogen as mention above, and use an over night precipitation.

Day 2

13. Re-suspend the DNA in 10 μ L 1XTE buffer.
14. Measure the DNA concentration.
 - * Avoid using NanoDrop. We found that measuring the concentration of 150bp DNA using this spectrophotometer is inaccurate.

We used the Qubit dsDNA HS Assay Kit (Molecular Probes, Life, Technologies).
15. Dilute the DNA to 0.1 ng/ μ L concentration in H₂O.

16. Amplify the DNA by 8 cycles of PCR using 3' primer that is common to all bands, and 5' primer with a unique upstream 5 bp barcode sequence to each band.

* It is recommended to perform a trail PCR on one sample to verify that the amount produced after 8 amplification cycles is sufficient.

* Avoid using more PCR cycles to minimize any PCR biases.

* If Illumina sequencer is used, it is recommended to add 5 random nucleotides to the 5' end of the common 3' primer. This will increase the complexity on the sequencing flow cell.

17. Clean the DNA using MinElute PCR purification tubes.

* Elution in 11 μ L H₂O per tube.

18. Measure the DNA concentration.

* Avoid using NanoDrop. We found that measuring the concentration of 150bp DNA using this spectrophotometer is inaccurate.

We used the Qubit dsDNA HS Assay Kit (Molecular Probes, Life, Technologies).

* It is recommended to run a sample from each tube on 2% long agarose gel in which one lane contains DNA from the original (shorter) library.

Day 3

19. Prepare a library containing all samples for high throughput sequencing.

* If Illumina sequencer is used it is recommended to use a control lane (for example with phage lambda DNA), or mix the samples with another sample with less homogenous ends.

* Avoid using many PCR cycles to minimize any PCR biases.

References

1. Blecher-Gonen, R. *et al.* High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat Protoc* **8**, 539-54 (2013).
2. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat Biotechnol* **30**, 521-30 (2012).