

Group Meeting_Fangjie_2 (2015/03 to 2015/10)

>>>>> Project1: Nucleosome SELEX >>>>>>

The basic packaging unit of chromatin is nucleosome. In spite that the positioning of nucleosome in genome has been well studied^[1], it is difficult to deconvolute such *in vivo* binding preference data to address the underlying contributions from intrinsic sequence affinity, CpG methylation, and the competition bindings of transcription factors (TFs). We hope to dissect separately the contribution from each factor to the overall binding preference of nucleosome, by virtue of SELEX (Systematic Evolution of Ligands by Exponential Enrichment).

1-1 EMSA-SELEX for nucleosome binding signal

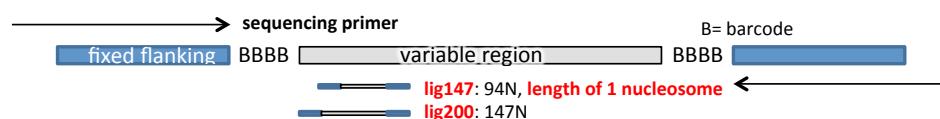
* Workflow

```

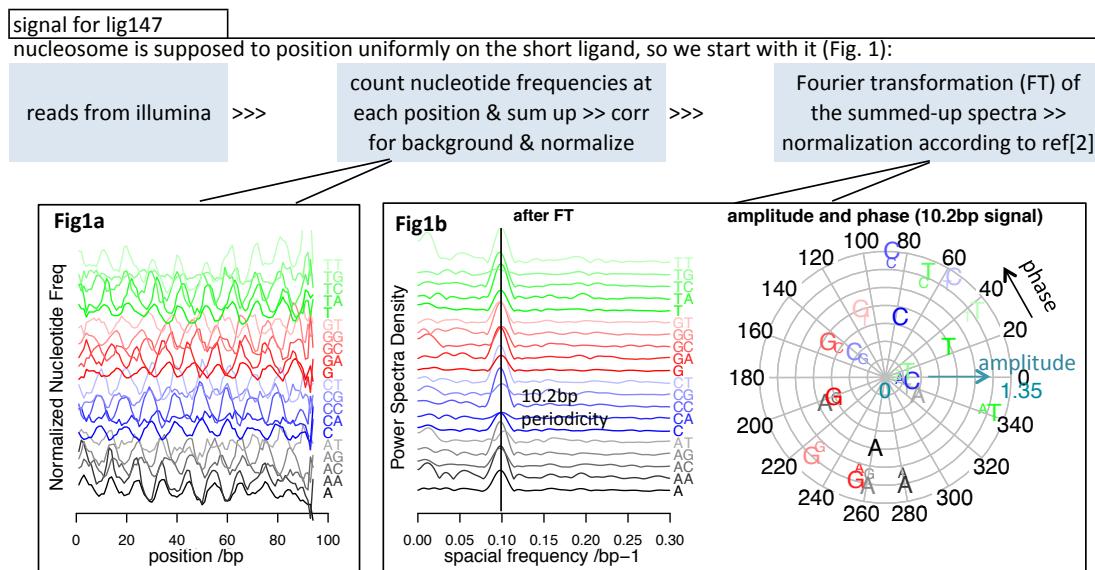
DNA ligands
+      >>    nucleosome   >>    purify for nucleosome   >>    amplify enriched ligands for
Histone octamer          (by EMSA)           sequencing and next cycle
(by Lucas of Patrick lab)

```

* Design of SELEX ligands



* Primary data analysis of the enriched signal

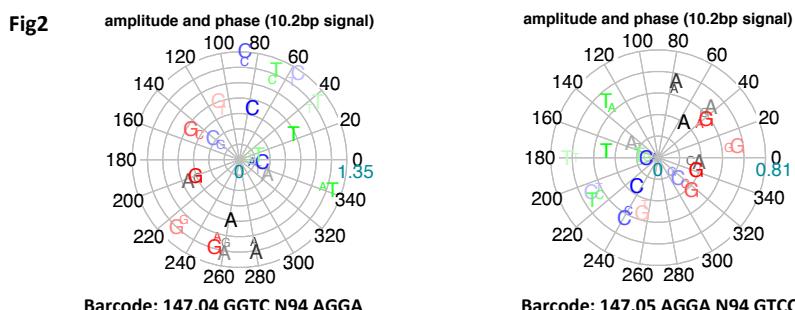


_In addition to the AA=TT and TA signal reported by Widom^[3], CC=GG, TC=GA, and CT=AG also showed relatively large amplitude of the 10.2bp periodicity

_No higher-order harmonics of the fundamental 10.2bp signal are observed, plausibly because the stringency of this SELEX is not enough yet, most sequences within the result library only have short-range periodicity. Can be improved by increasing cycles or using higher conc of TBE buffer during electrophoresis.

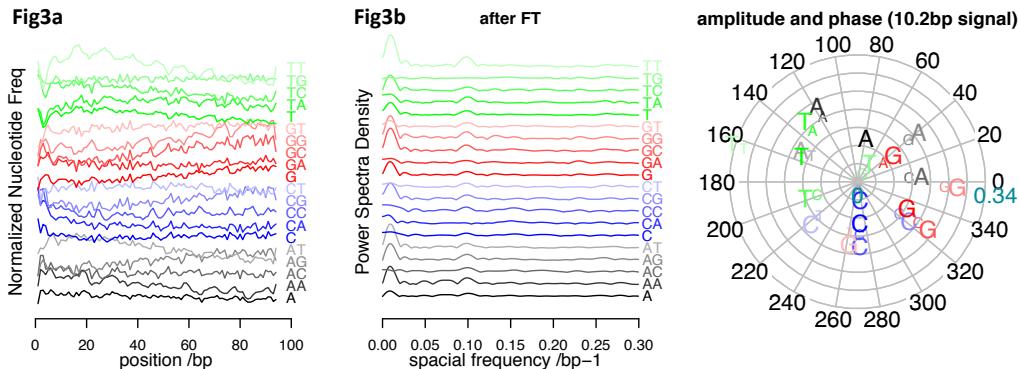
_kmer-space and autoseed failed to identify any significant binding motif, maybe we can try counting evenly-spaced kmer and use the most enriched ones as seeds

Fixed barcodes on SELEX ligands have positioning effect, as suggested by the difference of the absolute phase of the nucleotides between barcodes (Fig. 2). However, the relative phase of all nucleotides remained conserved.

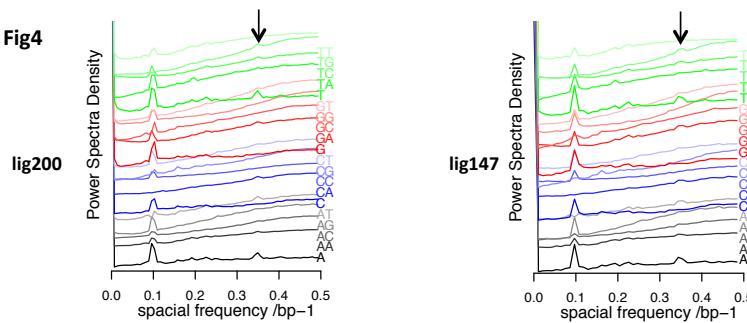


signal for lig200

Nucleosome may assume multiple possible positions on lig200, summed-up spectra only give weak signal (Fig. 3)



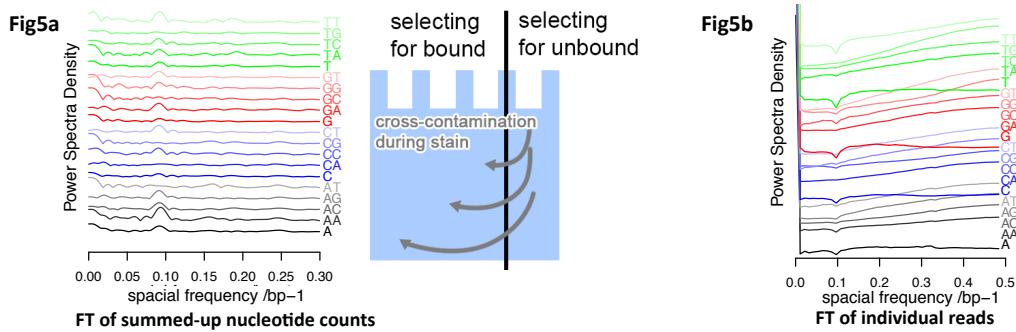
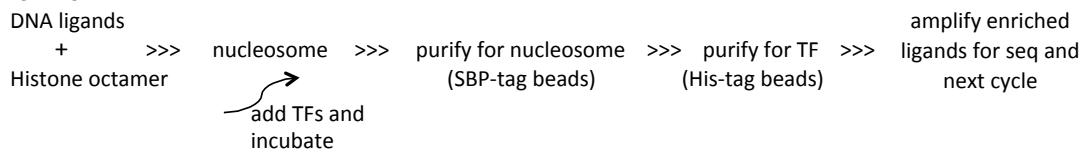
However, if we FT every **individual read** and then sum up the power spectra, the result (Fig.4) suggests that periodicity of the resulted library of lig200 is comparable with that of lig147.



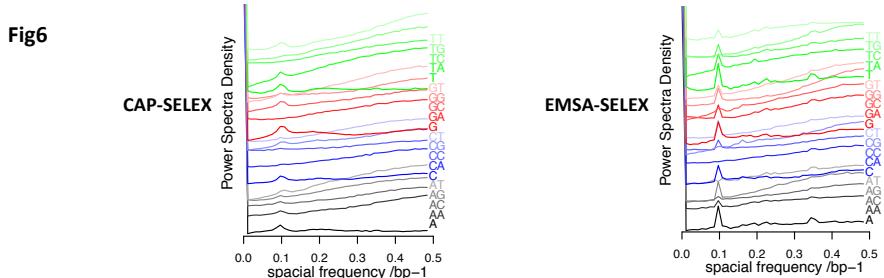
In these FT spectra of individual reads, origin of the peak at 0.34 bp⁻¹ (arrow in Fig. 4) is not yet understood, it is not a higher-order harmonic of the 10.2 bp (0.98 bp⁻¹) periodicity.

unbound fraction

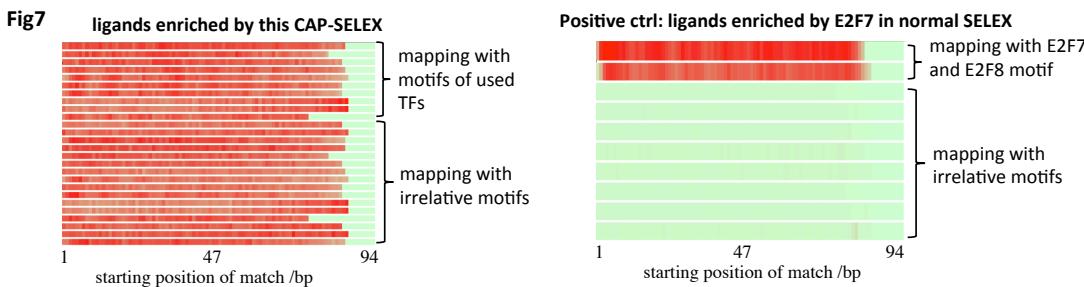
The unbound fraction of lig147 reveals weak signal when FT the summed-up nucleotide counts (Fig. 5a left), this could be due to the cross-contamination of barcodes during staining the EMSA gel (Fig. 5a right) and that the depletion of periodicity does not sum up in the real space. In contrast, the FT spectra of individual reads correctly represented the actual depletion of the 10.2bp periodicity in the unbound fraction (Fig. 5b).

**1-2 CAP (consecutive affinity purification)-SELEX to study the competition of transcription factors and nucleosome***** Work flow***** Results**

Nucleosome signals have been enriched (Fig. 6), however, the peaks in the FT spectra is broader than EMSA SELEX (suggesting a less well-defined periodicity), possibly due to the use of histone octamer with attached SBP-tag. It would be great if we can test histones with a smaller tag, Lucas is still struggling with the expression.

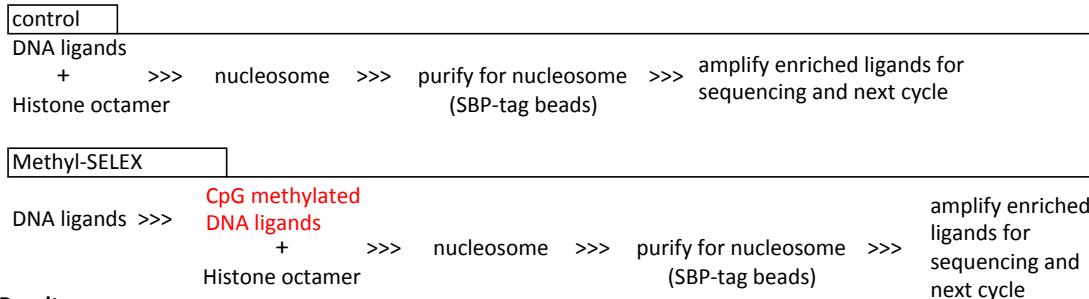


However, the CAP-SELEX failed to enrich any binding signal of TFs (Fig. 7). Yimeng suggested that the large amount of free DNA ligands after nucleosome reconstitution may occupy most TF proteins and result in the poor TF signal. We may improve this by using less DNA and histone octamer in reconstitution, and by increasing the amount of TF proteins.



1-3 Methyl-SELEX to study the effect of methylation on nucleosome sequence preference

* Work flow



* Result

Signals have been enriched for both ctrl and methyl ligands but were extremely weak. I made a mistake by reducing too much the soaking time during wash, because I assumed that nucleosome is unlikely to dissociate under physiological salt concentrations and it is just the buffer change that matters during wash, but it seems the dissociation step is actually playing a crucial role.

Common issues with all nucleosome SELEX

_Signals are much weaker than normal TF SELEX. Although the binding of nucleosome to DNA is intrinsically less specific than TFs, the wash step (or the run time and buffer in EMSA) also needs further optimization.

_Primer concatenation appeared in the later rounds of SELEX. Should be able to solve by reducing the amount of histone octamer in reconstitution, which in turn eliminates the necessity of concentrating PCR products by heat evaporation. At worst we can rely on nested primer pairs.

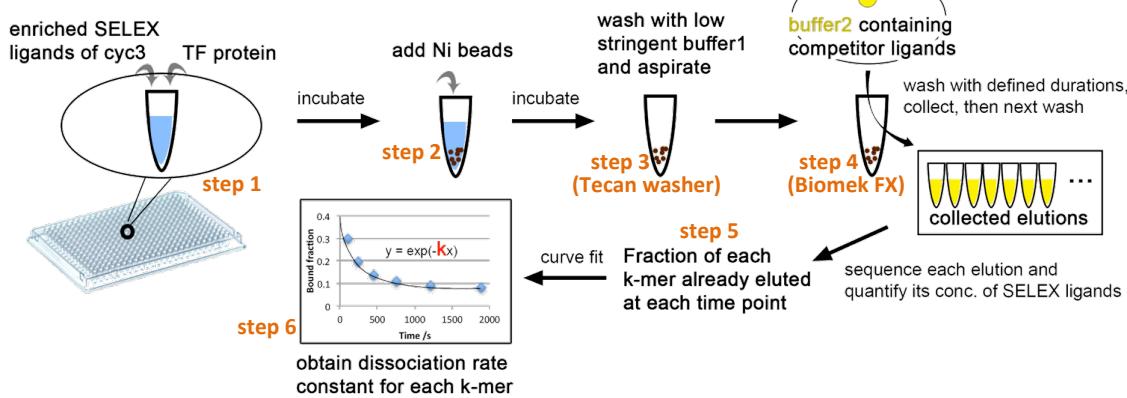
_Not yet understood why the periodic signal of nucleosome is stronger when using higher histone octamer/DNA ratio. Could be some stoichiometric effect.

>>>>> Project2: high throughput measurement of TF binding kinetics and thermodynamics

Currently, most binding models of TFs are probabilistic models representing the relative likelihood of sequences to be bound under a certain experiment condition. These models cannot predict the absolute amount of binding, and also cannot quantitatively predict the binding probabilities once TF or site concentration changes. To further improve the prediction power of TF binding models, deriving models by energetic approaches would be preferable.

2-1 HT measurements of dissociation kinetics

To obtain a less complicated starting library, 3 cycles of monomer SELEX were run in 4 replicates using 384 plate with help from Yimeng, ca. 80% of the wells enriched signals.

*** Workflow**

nth wash (step4)	1	2	3	4	5	6	Final heat elution
duration of wash	1min 30s (1.5 ¹ min)	2min 15s (1.5 ² min)	3min 23s (1.5 ³ min)	5min 4s (1.5 ⁴ min)	7min 36s (1.5 ⁵ min)	11min 24s (1.5 ⁶ min)	80°C for 15min

The run has finished but encountered some issue with library preparation. In order to decrease rebinding as complete as possible, we directly used the PCR product of competition ligand as wash **buffer2** without dilution (so the wash buffer contains lots of competition ligand and their **primers**). This significantly suppressed the amplification of target SELEX ligands when preparing sequencing library from the elution, and caused inaccuracy of measurement of lib concentration.

For the present run we can try to use less elution as template in PCR. In the future, we hope to synthesize double-stranded competitor ligands by annealing two complement oligos together, because amplifying large amount of competitor ligands by PCR was quite a hassle and will not help to save money (consumes a lot of polymerase).

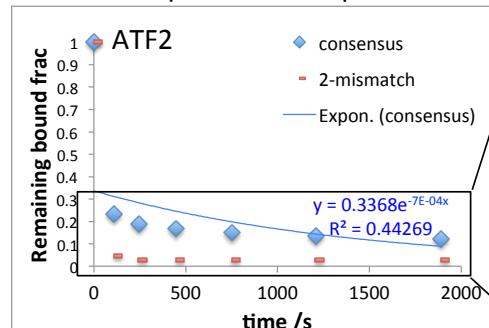
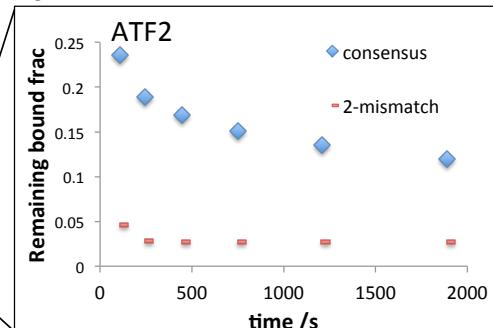
We also included controls in the 4th-quad of the 384 plate designed as follows, each design was tested with 12 proteins and the same 12 proteins are used for all the designs.

control design	ligand	protein	competitor
cyc3 PCR product	+	100% conc	
cyc3 PCR product	+	50% conc	
cyc3 PCR product	+	10% conc	
cyc3 PCR product	+	-	
cyc3 PCR product	+	non-specific	
synthetic consensus lig.	+	100% conc	
synthetic 2-mismatch lig.	+	100% conc	
cyc3 PCR product	-	100% conc	

*** Data analysis**

We first considered to use the basic rate equation of dissociation $f_b = e^{-kt}$ (cf. supplement S-1 for details). To fit this model the concentrations of SELEX ligands in each elution is required. These concentrations can be measured by qPCR. The amplification efficiency of SELEX ligands was measured as 1.95–2.06 so basically no correction is needed. But as mentioned above, the elutions contained ca. 35ng/ μ L competition ligand. The elutions should be diluted 50x before used as template in qPCR, therefore a bit caution should be taken to prevent DNA contamination. The quantification by qPCR has been tested and seems doable.

For controls using synthetic consensus and 2-mismatch ligands, we were already able to monitor dissociation from the qPCR data (supplement S-2). One example is shown in Fig. 8.

Fig8a measured points with mono-exponential fit**Fig8b**

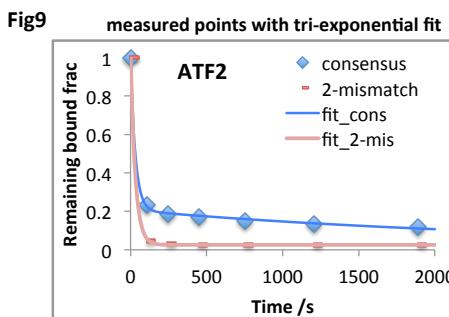
The dissociation data in Fig. 8 cannot be satisfactorily described by a mono-exponential model (Fig. 8a), instead we assumed that a tri-exponential decay model should be applied, including the following off-rate components:

1. **Weakly bound nonspecific ($k_{nsp,w}$)**. The fast dissociation at the beginning (Fig.8a) is obviously suggesting a weakly-bound fraction. Due to its large amount, this fraction is less likely to be carried by the protein. Plausibly it represents the nonspecific binding by beads or plates.
2. **Specific (k_{sp})**. The later points during the dissociation of consensus (Fig. 8) seem to illustrate a reasonable dissociation curve for a consensus ligand of a TF.
3. **Strongly bound nonspecific ($k_{nsp,s}$)**. The later points for the dissociation of the 2-mismatch ligand (Fig. 8) suggest the existence of a strongly bound fraction. Dissociation rate of this fraction is so low that it is less likely to be accounted for by a specific binding, which should usually be reversible. However, further examination is required to verify its non-specificity. Ratio of this fraction differs between wells of different proteins (supplement S-2) therefore carrier of this fraction seems to be the protein.

Then the dissociation model looks like:

$$f_b = F_{nsp,w}^0 e^{-k_{nsp,w}t} + F_{sp}^0 e^{-k_{sp}t} + F_{nsp,s}^0 e^{-k_{nsp,s}t}$$

Where f_b is the fraction of a k-mer remaining bound at time t . F^0 represents the fraction of each component of the total bound at $t=0$. Jointly fitting both the consensus and 2-mismatch data in figure 8 by this model (assuming some reasonable constraints) gives the results in Fig.9. It seems two mismatch from the consensus significantly increased the dissociation rate.



	Fitted parameters		
	k_{off}	Initial frac	
consensus	$k_{nsp,w}$	0.031578	$F_{nsp,w}^0$
	k_{sp}	0.000385	F_{sp}^0
	$k_{nsp,s}$	0	$F_{nsp,s}^0$
2-mismatch	$k_{nsp,w}$	0.031578	$F_{nsp,w}^0$
	k_{sp}	0.031578	F_{sp}^0
	$k_{nsp,s}$	0	$F_{nsp,s}^0$

* yellow cells indicate the adjusted parameters

* Challenges for this project

The major challenge stems from the need of fitting for multiple parameters from limited data points (each data point accords to 1 illumina lane therefore hard to measure a lot). However, if the dissociation rate of nonspecific components are the same for all TFs, or at least the same for all k-mers against the same TF, we can get a relatively accurate fit of k{sp} by jointly fitting for all k-mers.

_Another challenge is that I am a bit concerned the inability of this methodology to measure fast kinetics may end up only giving the dissociation kinetics for k-mers which are very closely related to the consensus. As suggested in supplement S-2, most 2-mismatch ligands showed too fast dissociation for an accurate measurement. Probably we can adjust the salt concentration or temperature of the wash buffer to slow down the dissociation. I guess the vigorous pipetting during wash by Biomek is also accelerating the dissociation and maybe better to have a shaker accessory for Biomek.

_Rebinding of the dissociated ligand may be a issue but will not affect the result qualitatively

* Some Trivial points to improve for better accuracy of measurement

_After the slight wash of step 3, liquid drops were attached to the wall of some wells, making the volume of the 1st wash in step 4 inaccurate. >> blot against paper towel

_Timing accuracy of biomek has limit, may affect the first few washes.

_Carryover of one elution to the next. (To improve the timing accuracy, no tip changing was performed between washes) >> can add a specific ligand and using qPCR to monitor carryover rate

In summary, this project is doable but may need some effort to optimize. If possible I hope to get some help or discussion with professionals about the data analysis after the results come out.

References

- [1] aK. Brogaard, L. Q. Xi, J. P. Wang, J. Widom, *Nature* **2012**, 486, 496-501;
- bT. K. Kelly, Y. P. Liu, F. D. Lay, G. N. Liang, B. P. Berman, P. A. Jones, *Genome Res* 2012, 22, 2497-2506.
- [2] C. K. Collings, A. G. Fernandez, C. G. Pitschka, T. B. Hawkins, J. N. Anderson, *Plos One* **2010**, 5.
- [3] P. T. Lowary, J. Widom, *J Mol Biol* **1998**, 276, 19-42.

Supplemental Materials

S-1

We start with the simplest dissociation model. Eq. (1) can be used to fit for dissociation rate constant of every single k-mer.

$$f_b = e^{-kt} \quad (1)$$

k : dissociation rate constant of the k-mer

t : time of wash

f_b : fraction of the k-mer remaining bound after time t , we only measured the elutions but f_b can be derived from Eq. (2).

$$f_b = 1 - \frac{\text{eluted amount at time } t}{\text{total bound at } t = 0} = 1 - \frac{\sum_{n=1}^{n(t)} c_n f_n}{\sum_n c_n f_n} \quad (2)$$

c_n : concentration of SELEX ligand in n th elution, measured by qPCR

f_n : fraction of total reads containing a certain k-mer, measured by sequencing

Measurement of c_n for each elution by qPCR is necessary because we cannot fit for the ratio of $k/k_{\text{consensus}}$, even though we have the counts of both consensus and the target k-mer for every elution.

S-2

Measured points of dissociation curve for consensus and 2-mismatch sequences of 12 TFs.

