

Artifacts of Sequences Under SELEX Selection Constraints

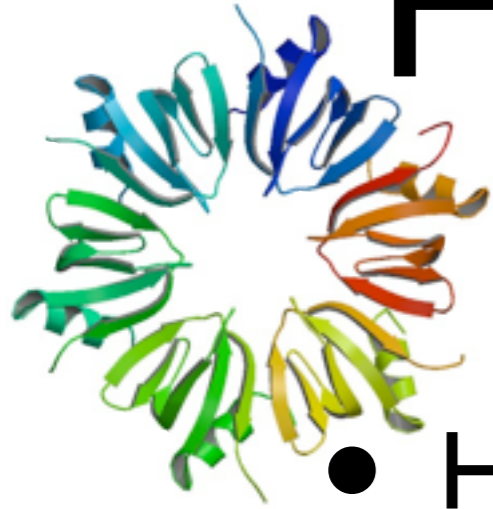
Bob Zimmermann, Tanja Gesell, Doris Chen,
Christina Lorenz and Renée Schroeder
Bled TBI Winterseminar
20 February 2009

Strategies for Detecting Novel ncRNAs

- *in silico*
 - Requires significant homology/structure
- *in culture/vivo* (RNomics)
 - Requires transcription conditions
- *in vitro* (Genomic SELEX)
 - Independent of expression levels
 - May present biases

Evaluation of Genomic SELEX Screen

1. Genomic SELEX for Hfq-binding RNAs
2. *in silico* analysis of sequenced cDNAs
3. Neutral SELEX as control for SELEX
4. 454 sequencing of Neutral and Genomic SELEX cDNAs
5. *in silico* analysis of SELEX pools

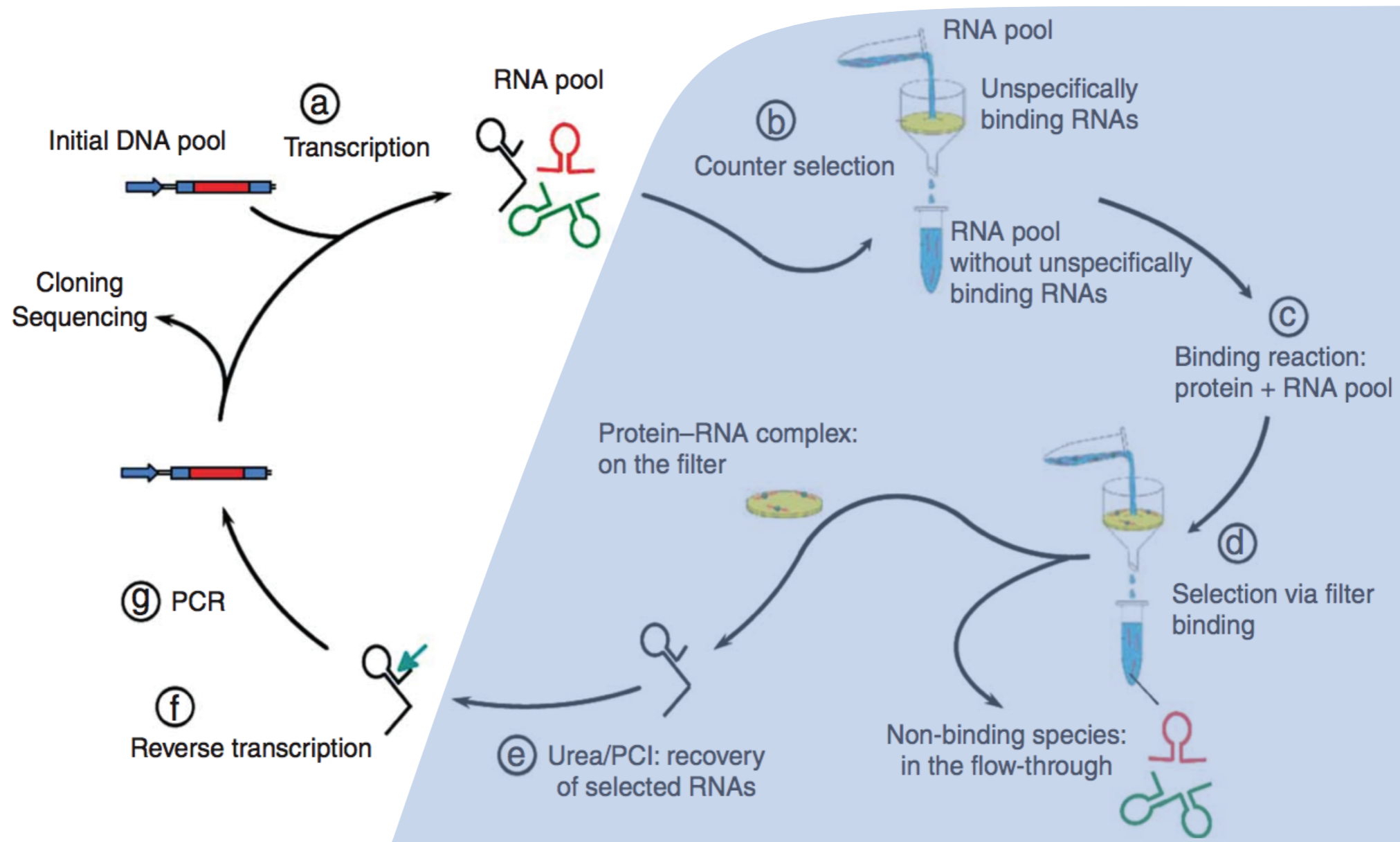


Hfq is a RNA-binding microbial protein

- Highly abundant, pleiotropic regulator involved in ncRNA/mRNA interactions
- OxyS RNA O₂-stress regulation
- Spot42 RNA kinase down-regulation
- Interested in a broader view of Hfq targets
- Develop a method irrespective of expression level: Genomic SELEX w/ *E.coli*

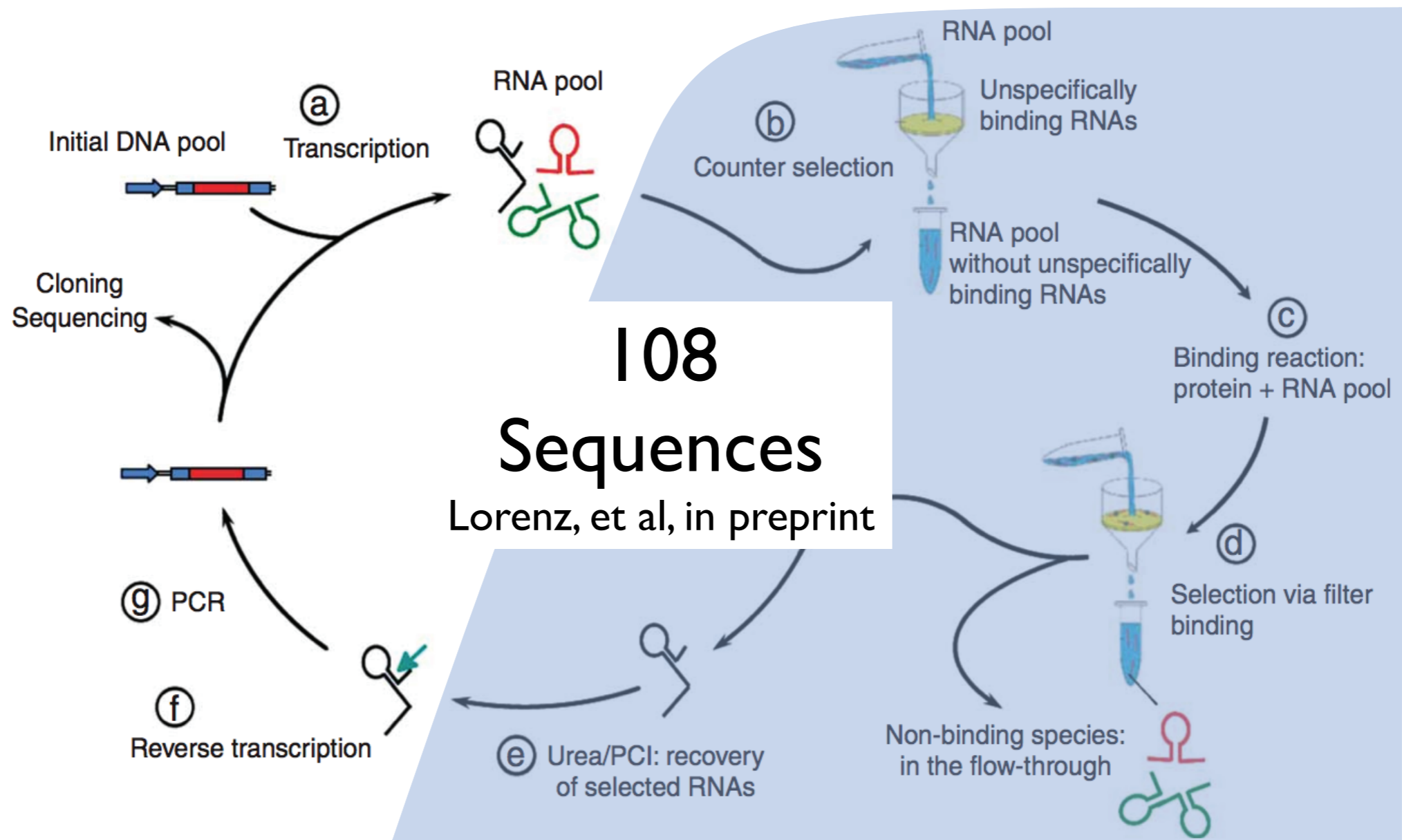
Lorenz, et al, in preprint

I. Genomic SELEX for Hfq-binding RNAs



Lorenz et al, 2006

I. Genomic SELEX for Hfq-binding RNAs



Lorenz et al, 2006

2. *in silico* analysis of sequenced cDNAs

- Measure structural stability: MFE Z-score
- Compares predicted secondary structures of
 - same length
 - same nucleotide content








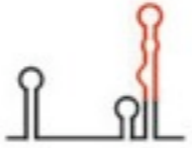
$$Z = \frac{E_x - \mu(\mathbf{E}_{x_{\text{shuff}}})}{\sigma(\mathbf{E}_{x_{\text{shuff}}})}$$

Le, et al 1989

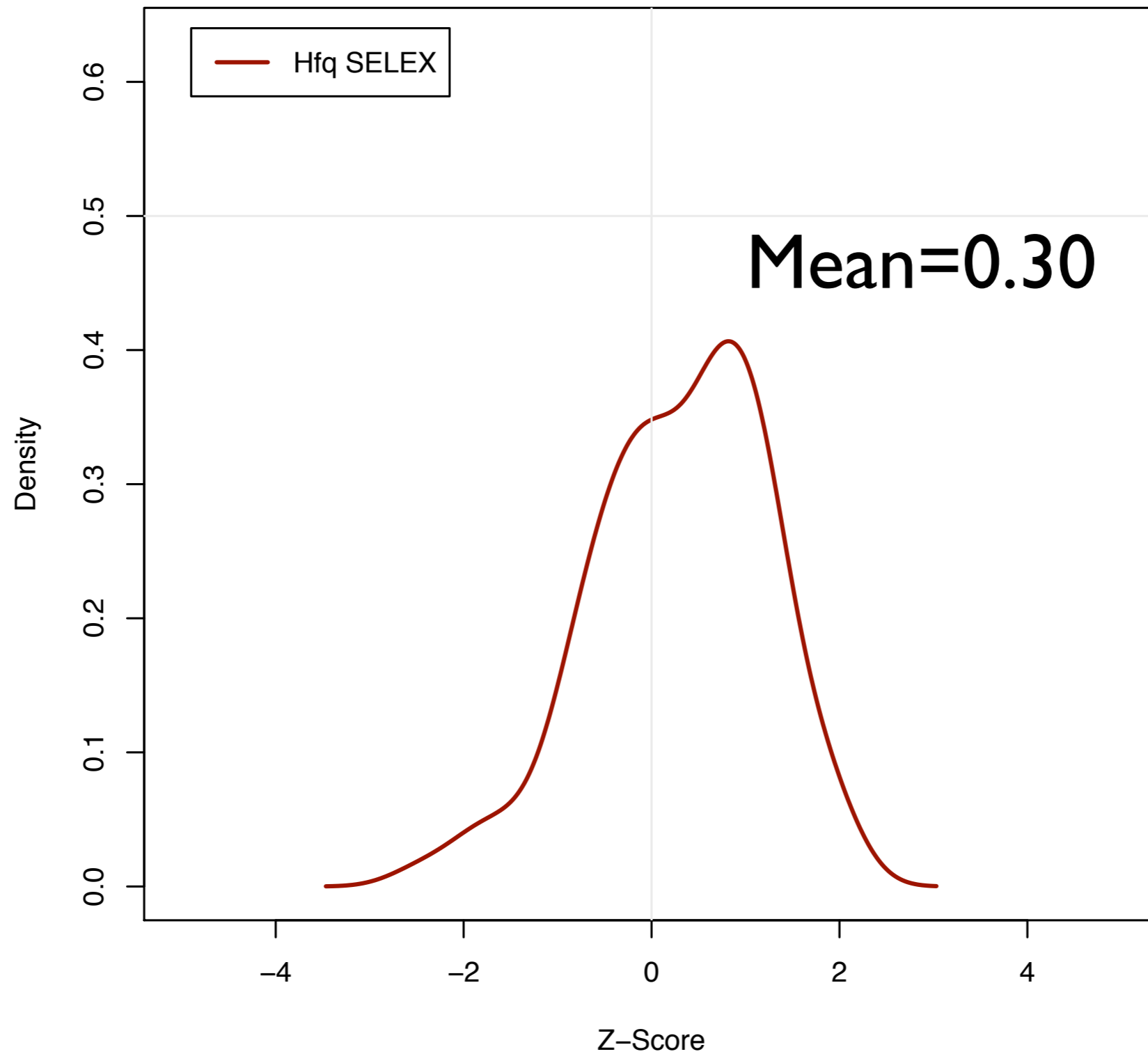
2. *in silico* analysis of sequenced cDNAs

- Measure structural stability: MFE Z-score
- Compares predicted secondary structures of
 - same length
 - same nucleotide content

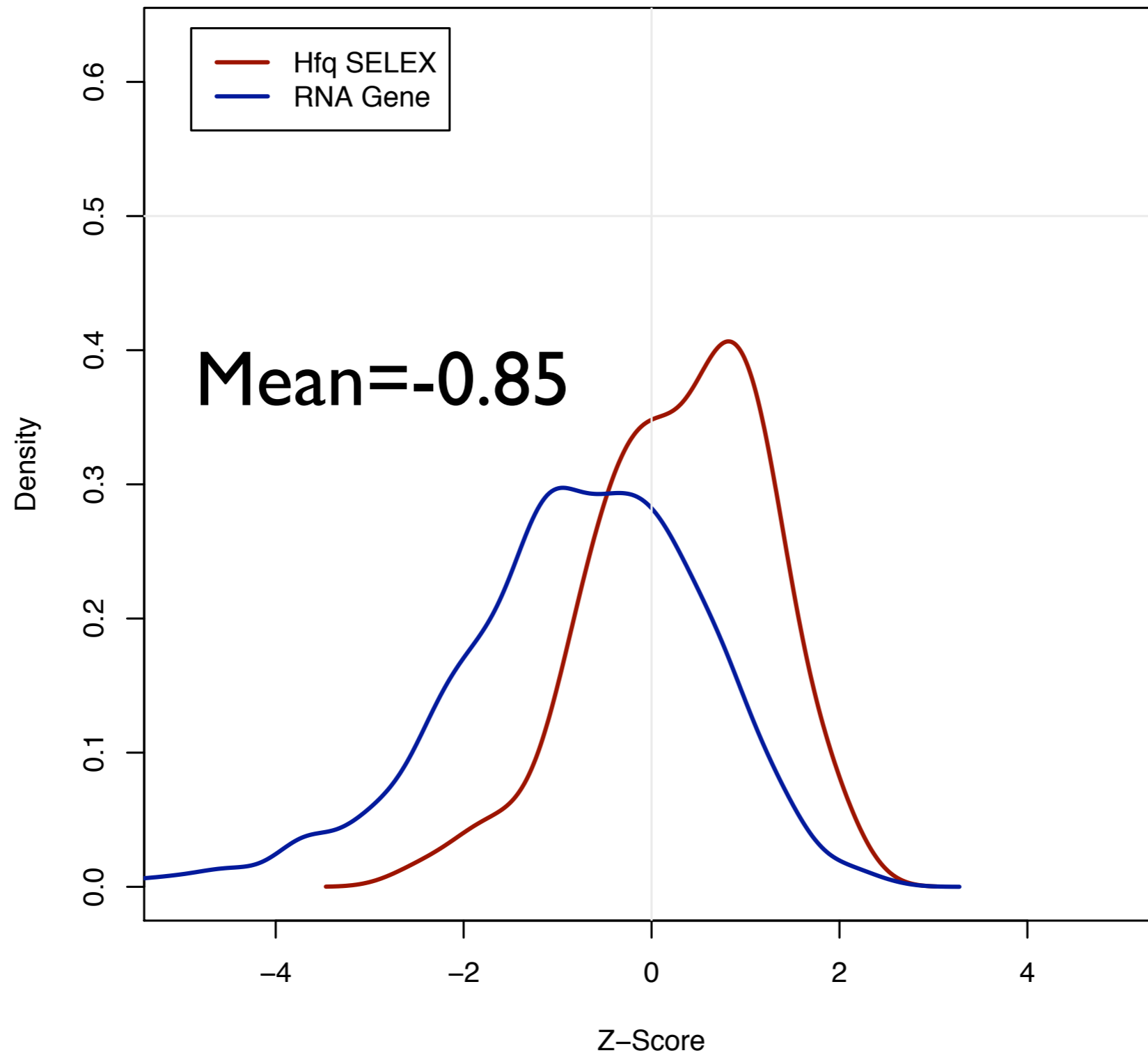
⁷Z-scores of known binders

RNA	Secondary Structure	
6S A		-5.52420
DicF A		-1.60733
DsrA A		-2.09376
MicF A		-1.95915
OxyS A		-3.03948
RhyB A		0.40742
RprA A		-0.75068
Spot42		-0.95073

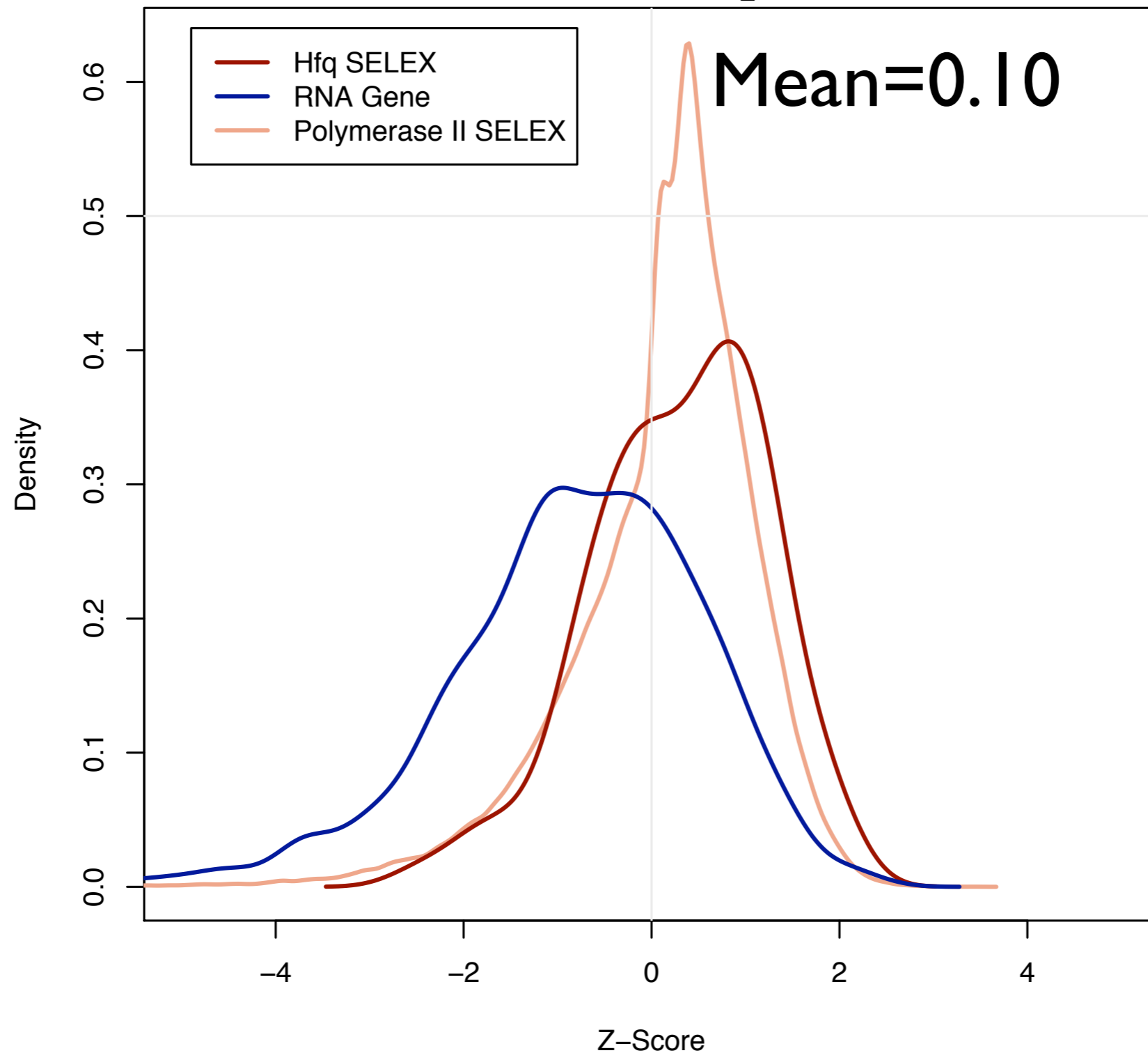
Hfq Sequence Stability



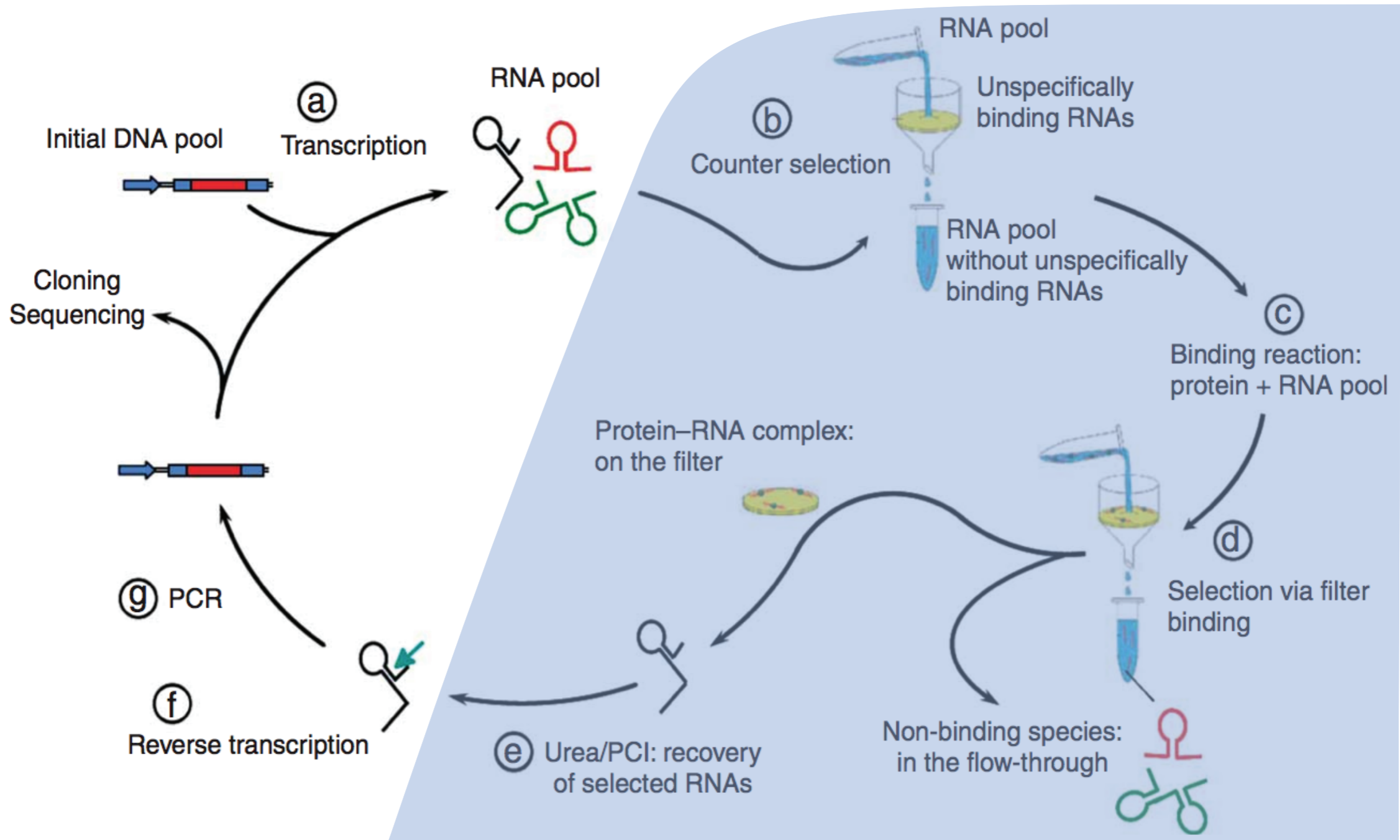
Does Hfq favor binding unstable RNAs?



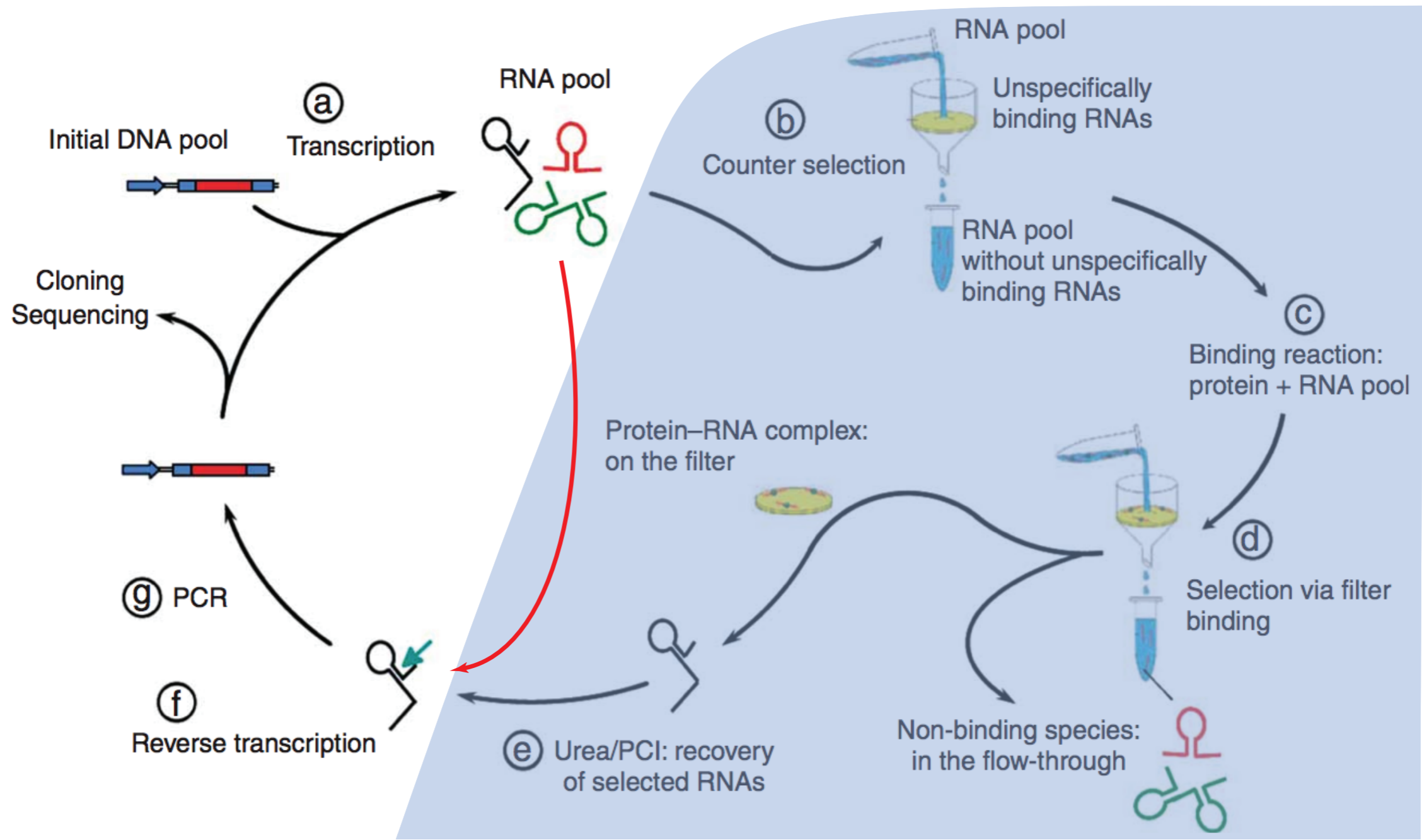
Does SELEX produce unstable sequences?



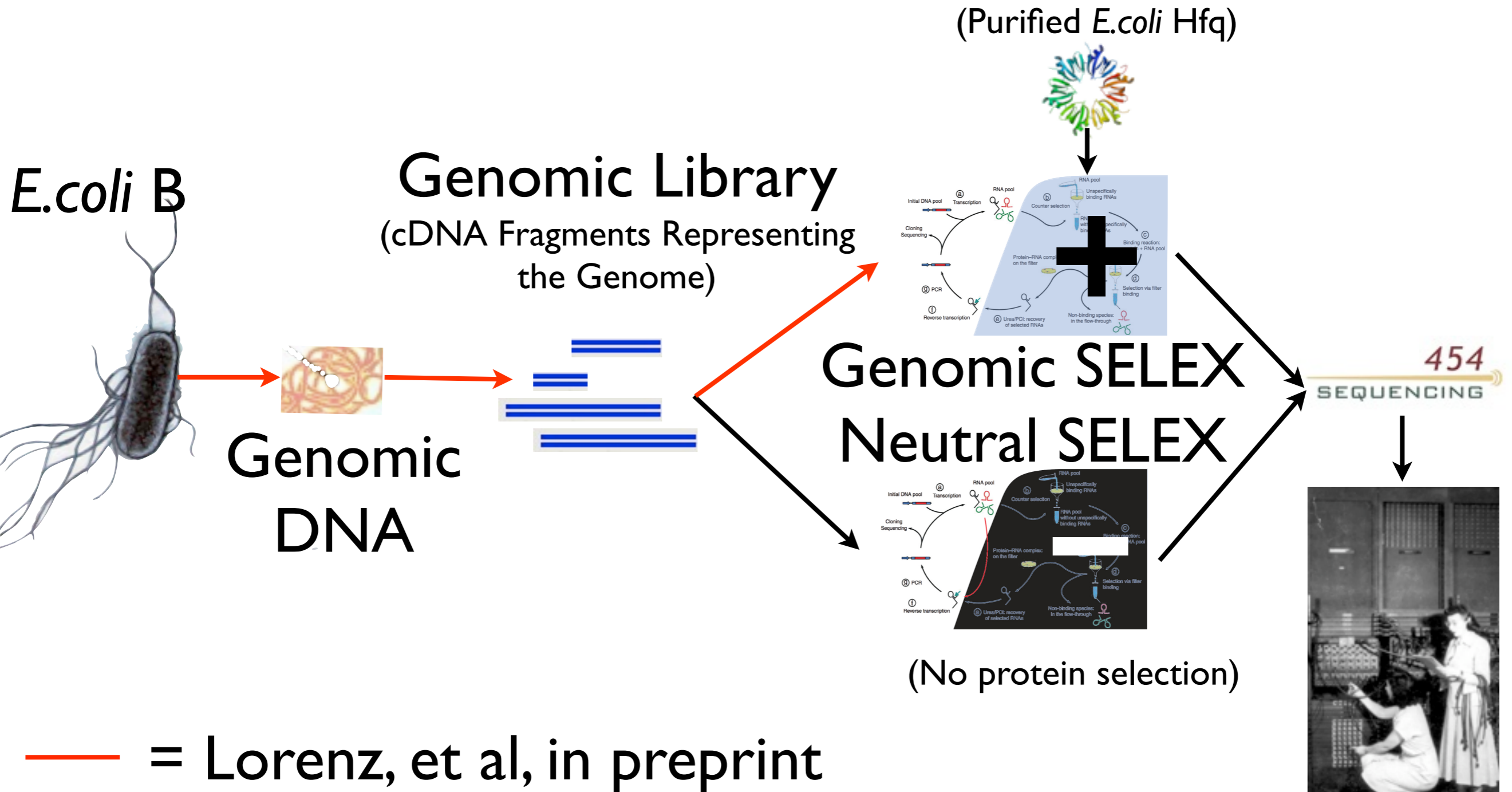
SELEX



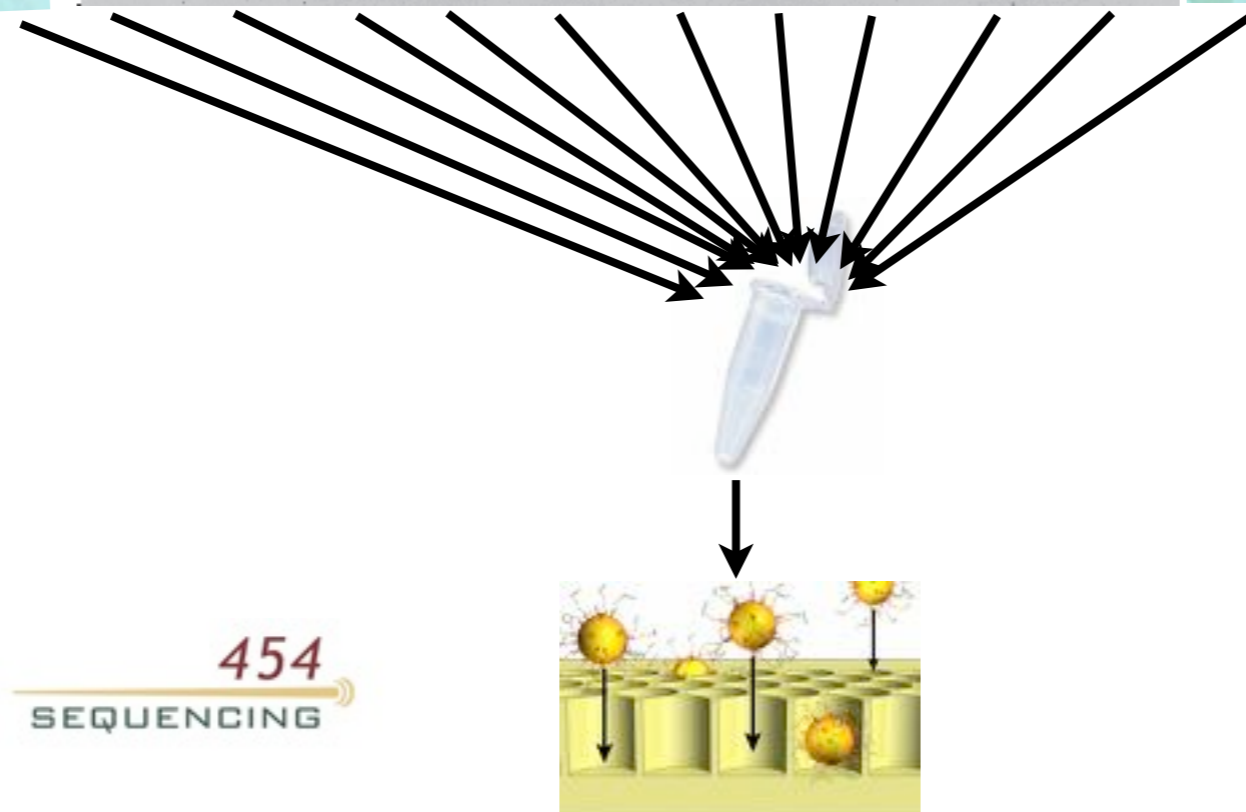
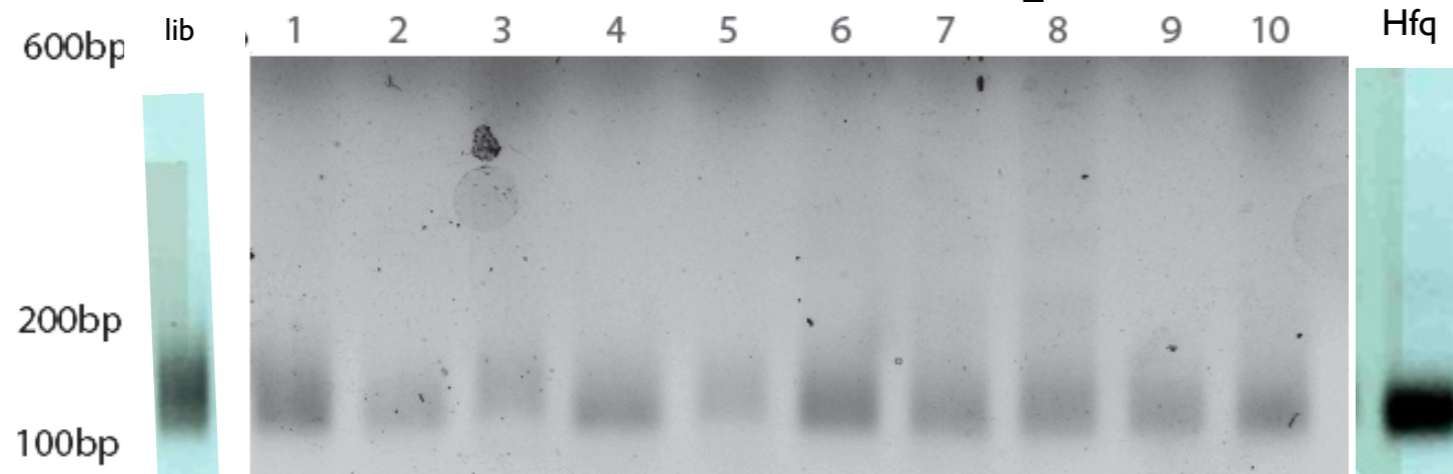
3. Neutral SELEX



Schematic of SELEX Experiments



4. 454 Sequencing of SELEX pools

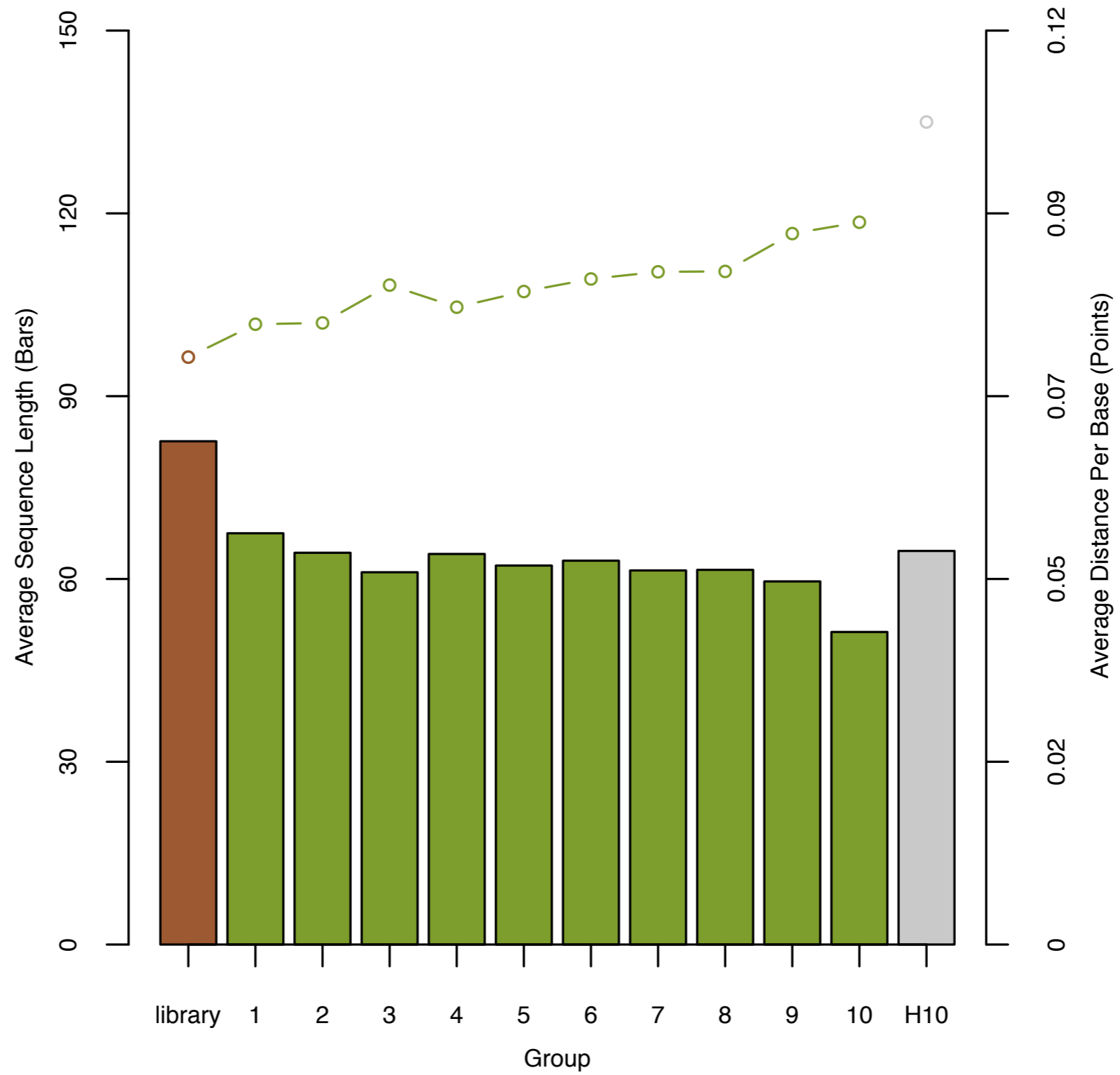


of Sequences

lib	3,276
1-8	68,135
9	5,894
10	69,426
Hfq	9,991

5. *In silico* analysis of SELEX pools

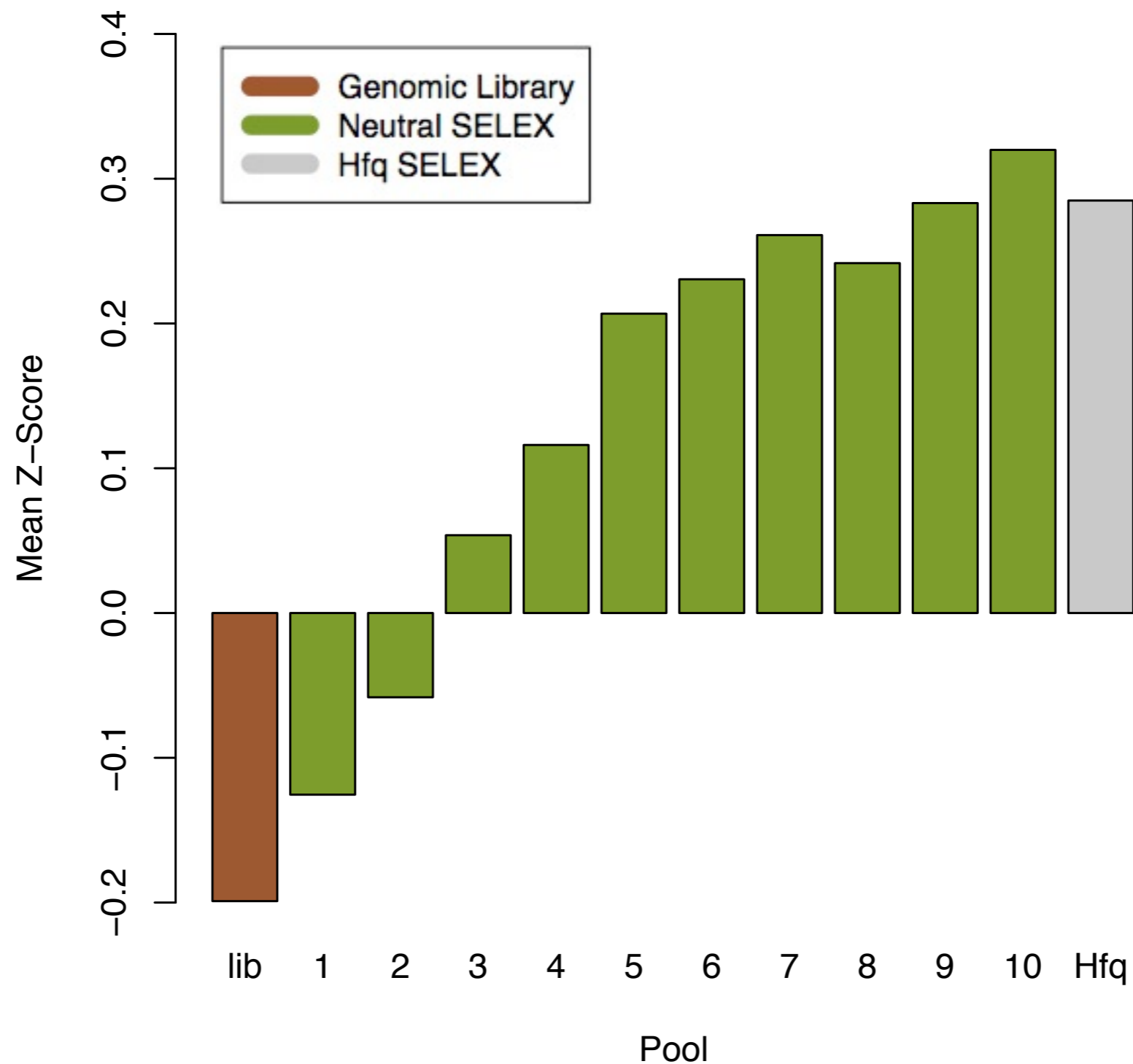
Lengths and Distances of Null SELEX Sequences



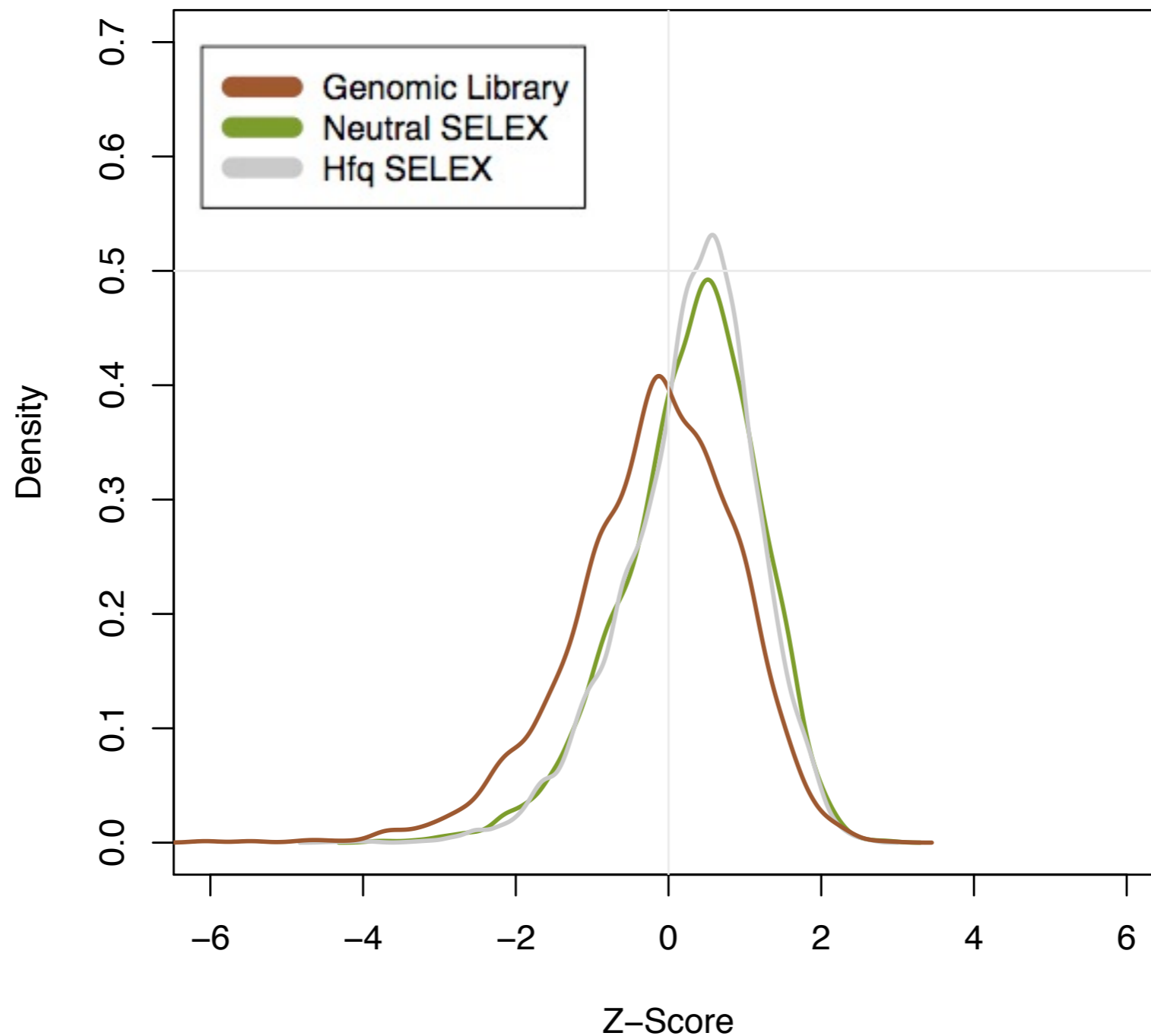
What is the effect of SELEX on the stability of sequences?

- Measure Z-score sequences in each pool
- Take the average of each pool
- Study progression of the pools' averages

SELEX decreases stability of sequences



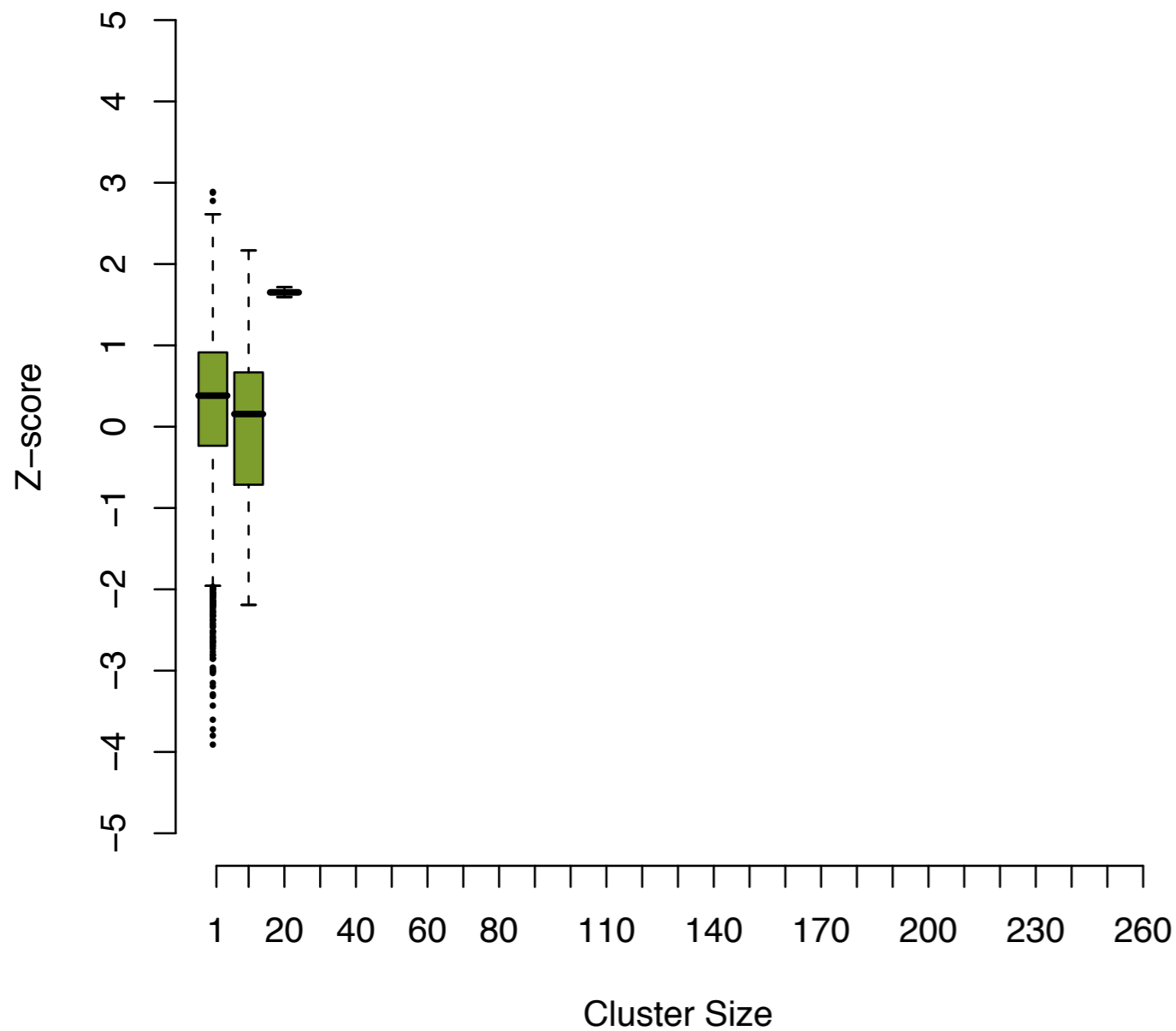
Hfq sequences are unstable from SELEX



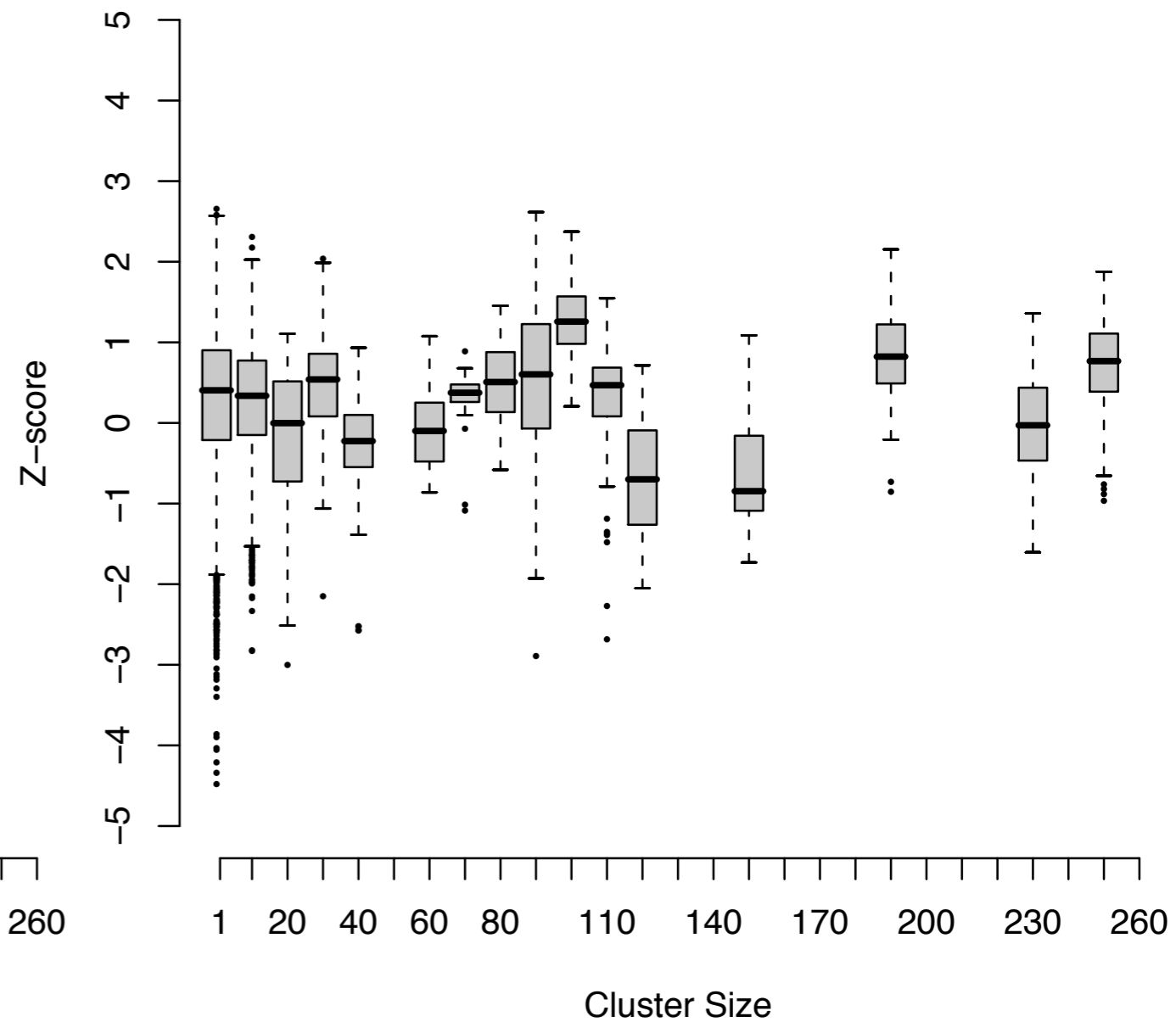
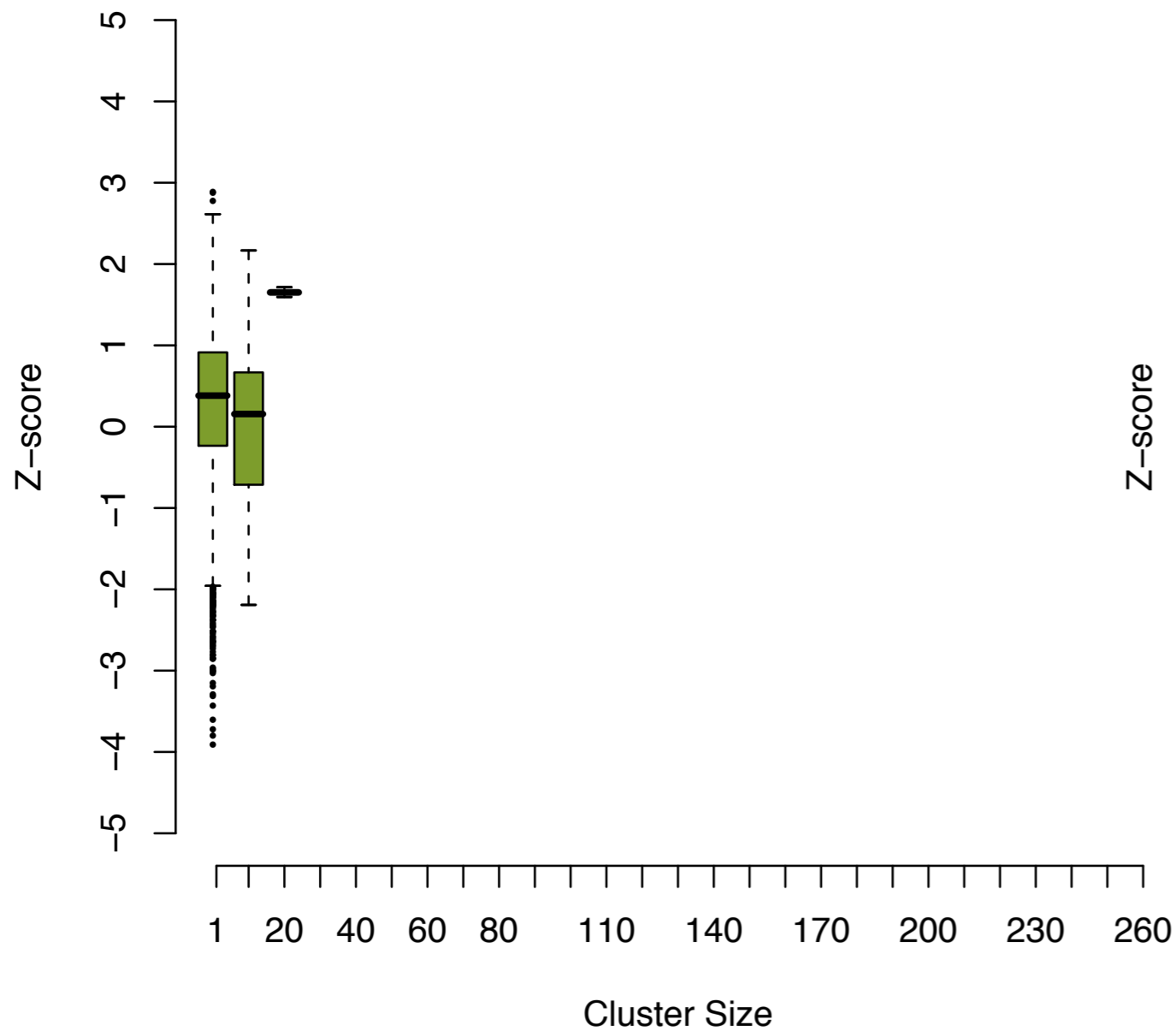
Must enriched sequences be unstable?

- Cluster sequences through pairwise alignments
- If the sequences are 85% identical, group them in a cluster
- Size of the cluster is a measure of the level of enrichment
- Plot the enrichment against Z -score

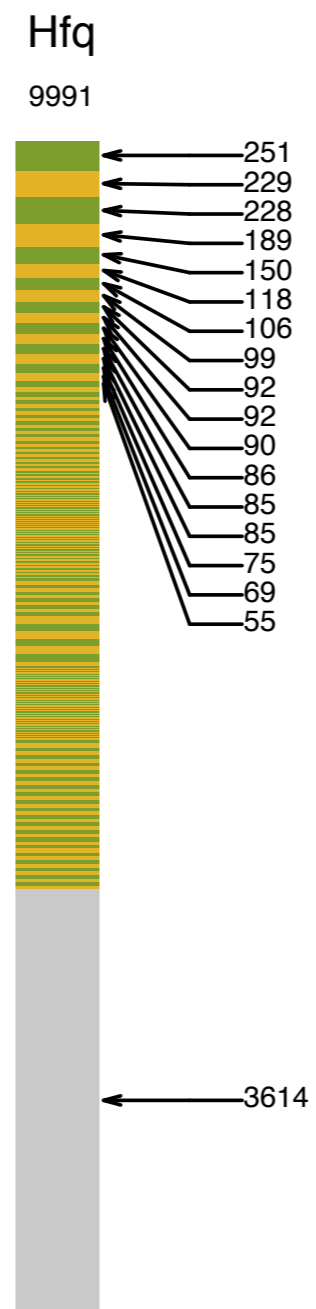
Enrichment does not imply instability



Enrichment does not imply instability

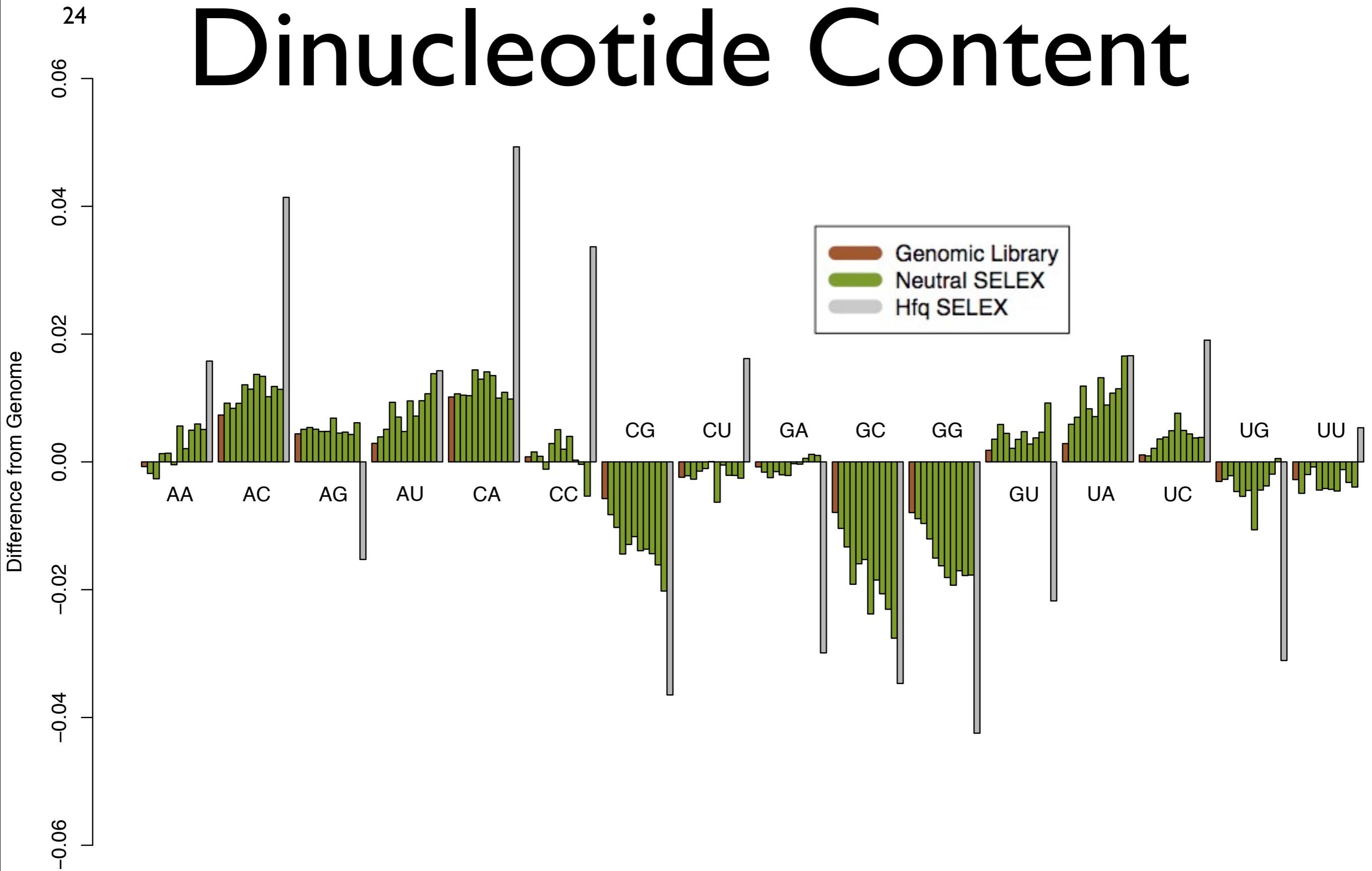


Enrichment is a Strong Signal

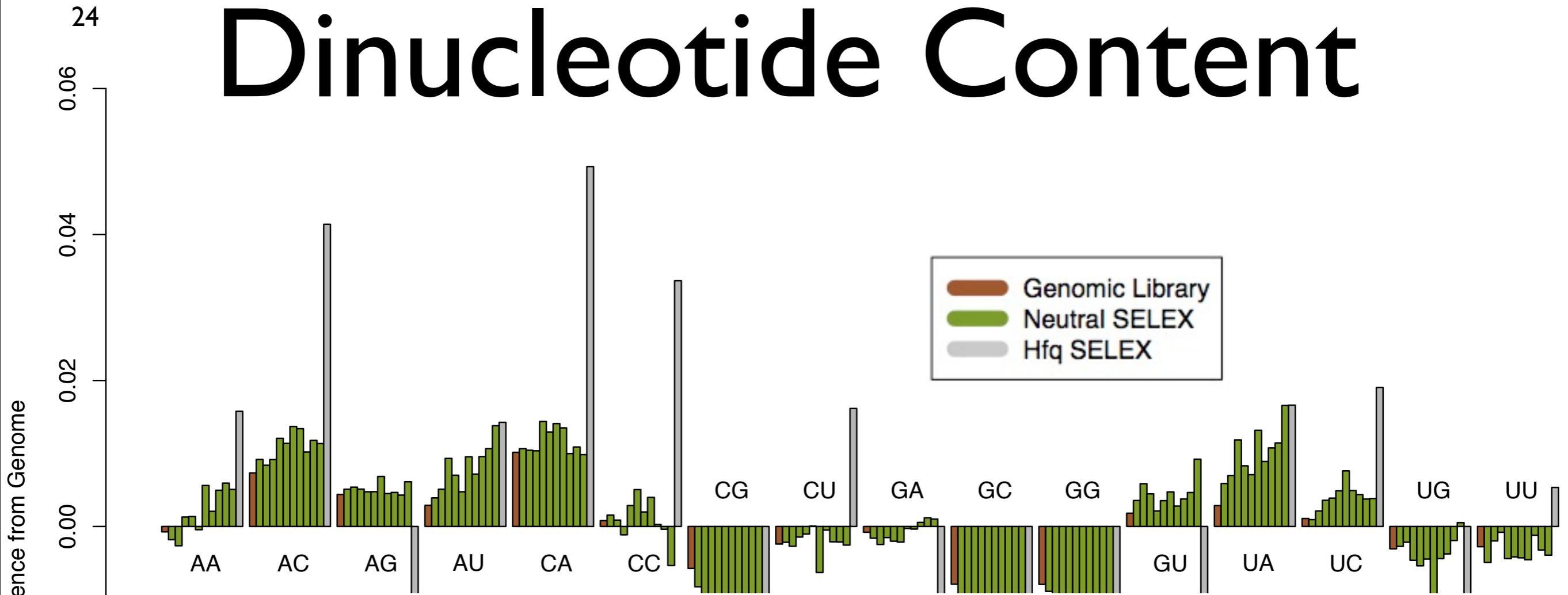


Enrichment is a Strong Signal

Dinucleotide Content



Dinucleotide Content



Winning Sequences: (of 9991)

1. CCCACG-TTAATCCCAATAAGAACAGAATCCAGATTATCCGTTTCAACATCACTCCTTCACTG (251x)
2. ACCA--ATAACTACAACATCAGCCCCAGTTCTTCTGCCAACTCATTGACTT (229x)
3. CCATAACA-ATAATCCATAGCTTCCCAGCCCACAAGGTACGAAACAACCCACGAATATCCAGTTCATTTTCAGCGT (228x)
4. CTACACCATGAAAATAAAGAACATTCTACTCACCCTTTGCACCTCACTCCTGCTTACCAACGTTTC (189x)
5. ACTTCATCACTAACAACAATAATACGGGAGACATTGGTTTCTT (150x)
6. AACAA-CAATAAA-GCGGCACACCCTACCCAAAACGCTGCTCGCATCTCTTCTTCCTCTGATCTTCAAGC (118x)
7. AATTCCACCATTATATCCATCTCCTCCCCAAACACTTCCATTATTATTTAGCGTAAAGTTTTACCGTA (106x)
8. CAAAACCTACAAGACCATAGAACCGCCAGCCTTTTCGCCTTTTTTTTCACCACTCATTGCCACATTCCTTGTGTATA (99x)
9. GTCTTACGAACGGTATCCGCAGCCACTTCATCACTAACAACAATAATACGGGAGACATTGGTTTCTT (92x)
9. CGCCAGACCAATAATAATAGTGACGGAACCAAGCACACCCGCCATCCCGGCGATCACTTTACCCACTT (92x)
10. CATAAGAACAATAACAAGACAATAAATTGCCATTACAGCCACAATAGCCAGCGTGC (86x)

Conclusions

- Genomic SELEX presents artifacts
 - Unstable sequences are favored
 - Certain dinucleotides are enriched
- Genomic SELEX: an *in vitro* alternative
 - Can screen genome for binding motifs independent of expression level
 - Sequence enrichment is a strong signal
 - It is possible to enrich stable RNAs

Future Plans

- Can Neutral SELEX aid in the bioinformatic analyses of SELEX pools?
- How does SELEX compare to Co-IP
 - Polymerase II: Genomic SELEX and Co-IP
- What bioinformatic analyses could lead up to the discovery of these RNAs *ab initio*?

Acknowledgements

Collaborators:

Tanja Gesell, Doris Chen,
Christina Lorenz

Schroeder Lab:

Oliver, Jen, Kasia, Boris,
Sabine, Ivana, Martina,
Marina, Andreas, Elli, Silke,
Katharina, Peter



Renée Schroeder



MFPL **Max F. Perutz
Laboratories**



Extra Help:
Christina Waldsich

Committee:
Eric Westhof
Ivo Hofacker

Polishing Up

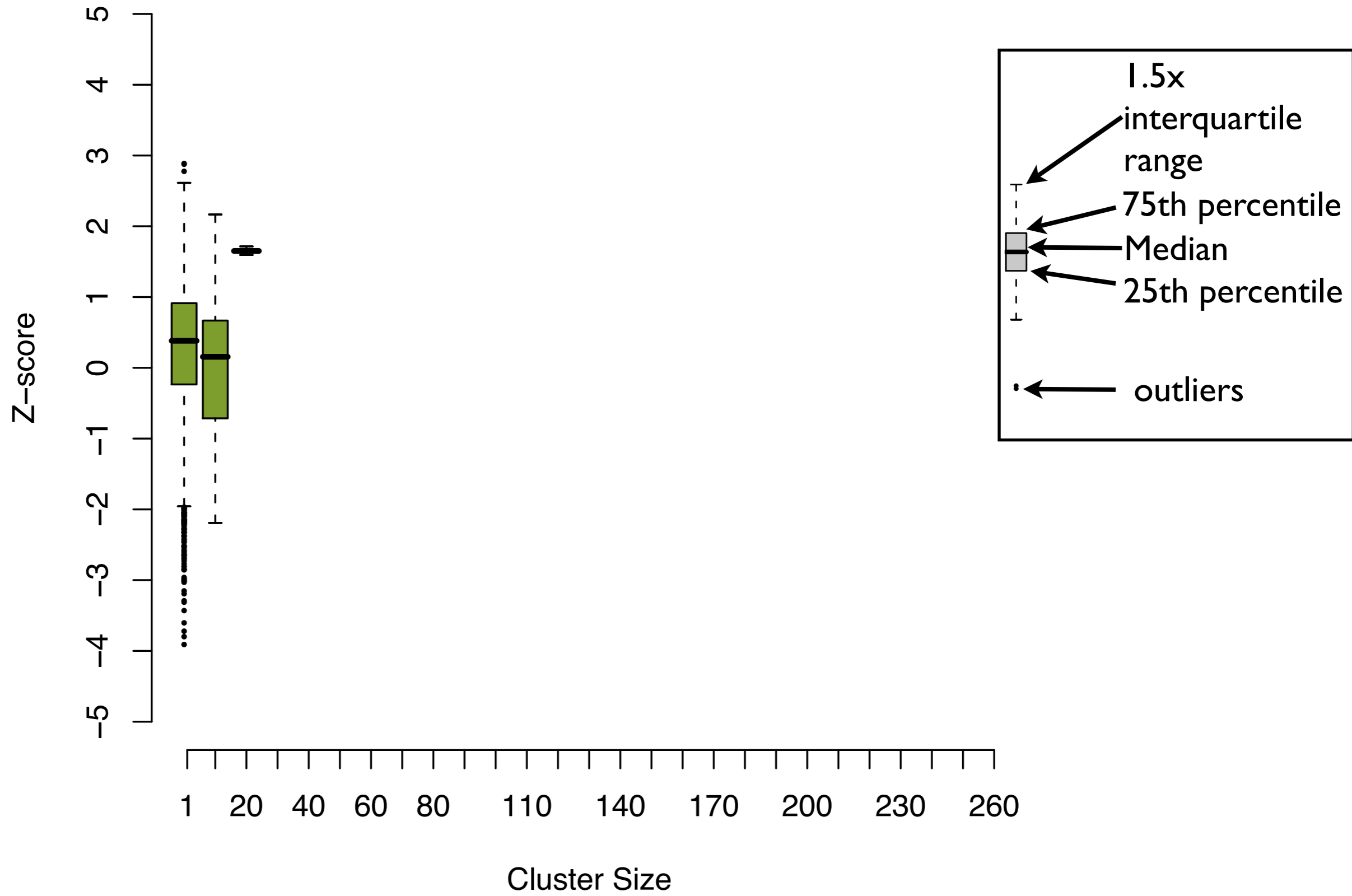
- To prove the worthiness of Genomic SELEX, we need the positive Hfq results
- We believe (are almost sure) the reverse transcriptase is at fault, but some biochemical grounding for the biases needs to be established. Could it be better?
- Previous work....

How is the nucleotide content affected by Hfq and Neutral selection?

- Count occurrence of each nucleotide of each pool
- Normalize to frequencies ($f_A + f_C + f_G + f_T = 1$)
- Compare SELEX pools' frequencies to genomic frequency

How stable are the Hfq-selected RNAs?

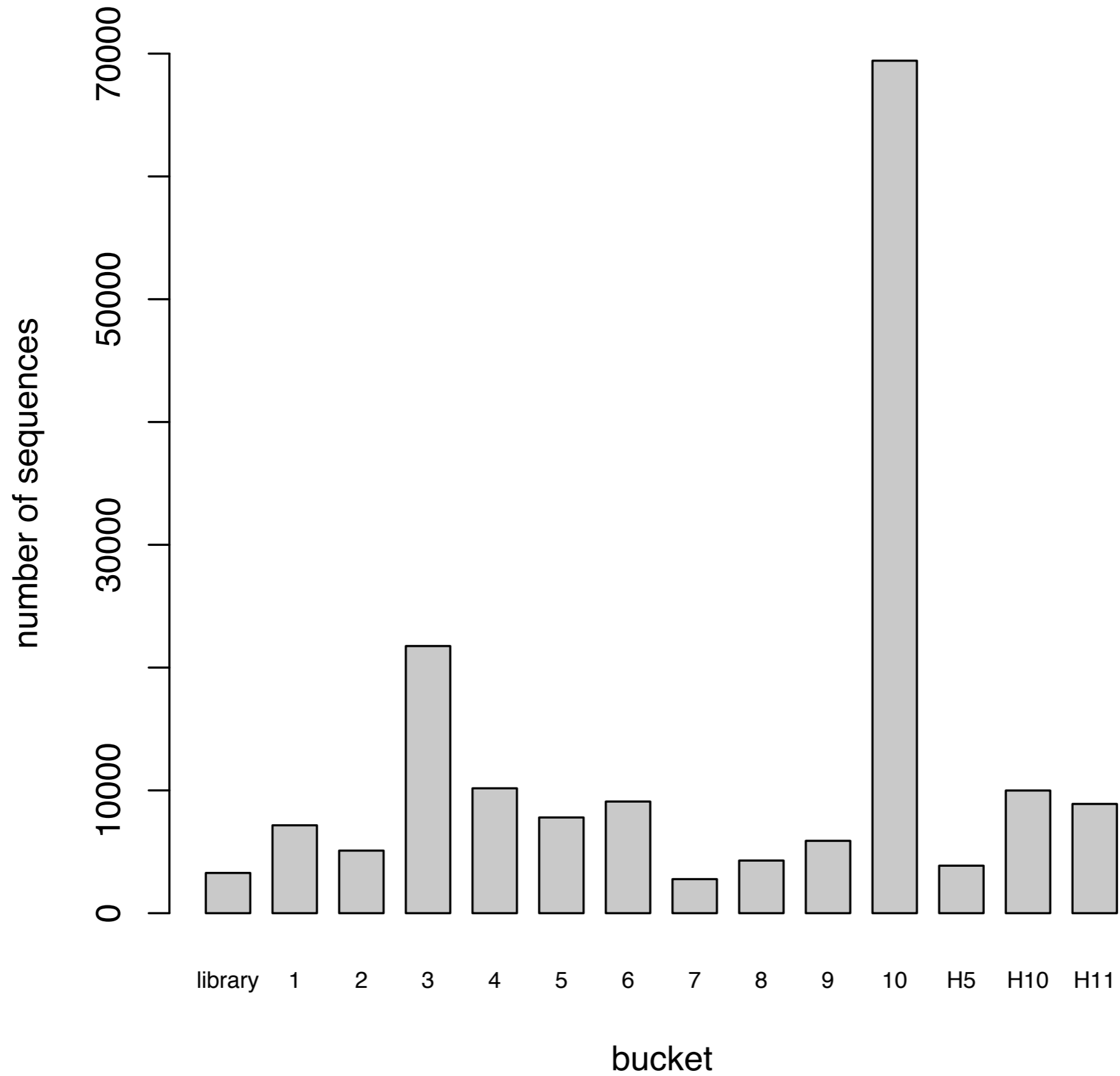
- Measure Z-score of each sequence in pool
- Look at the resulting distribution

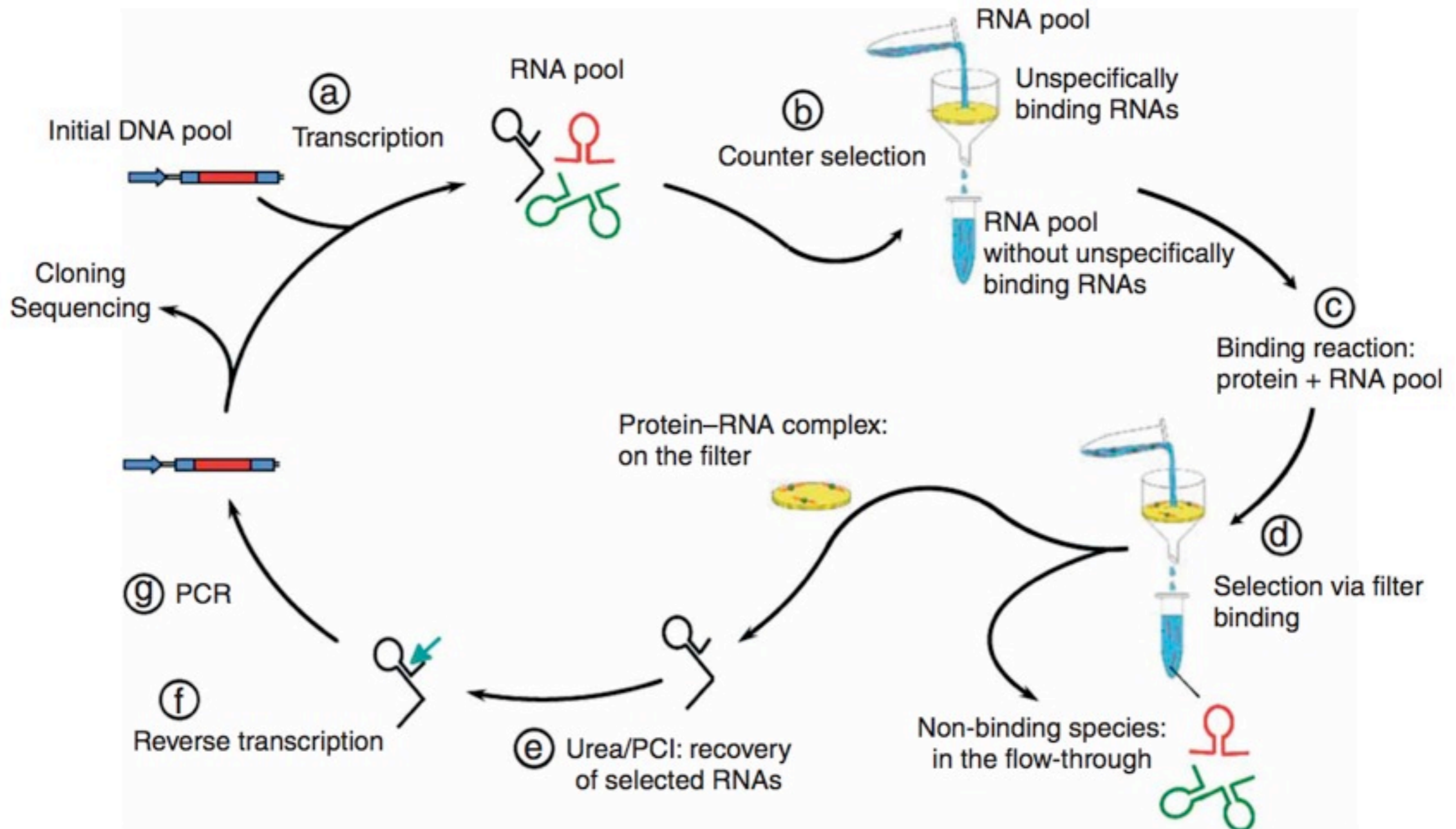


Future Directions

- Look into the enriched Hfq sequences
 - What are the motifs?
 - In what ways do they differ from the background?
 - Can Neutral SELEX aid in the bioinformatic analyses of SELEX pools?
 - Why are certain dinucleotides enriched?

Number of Full Length Sequences Found in Each Bucket



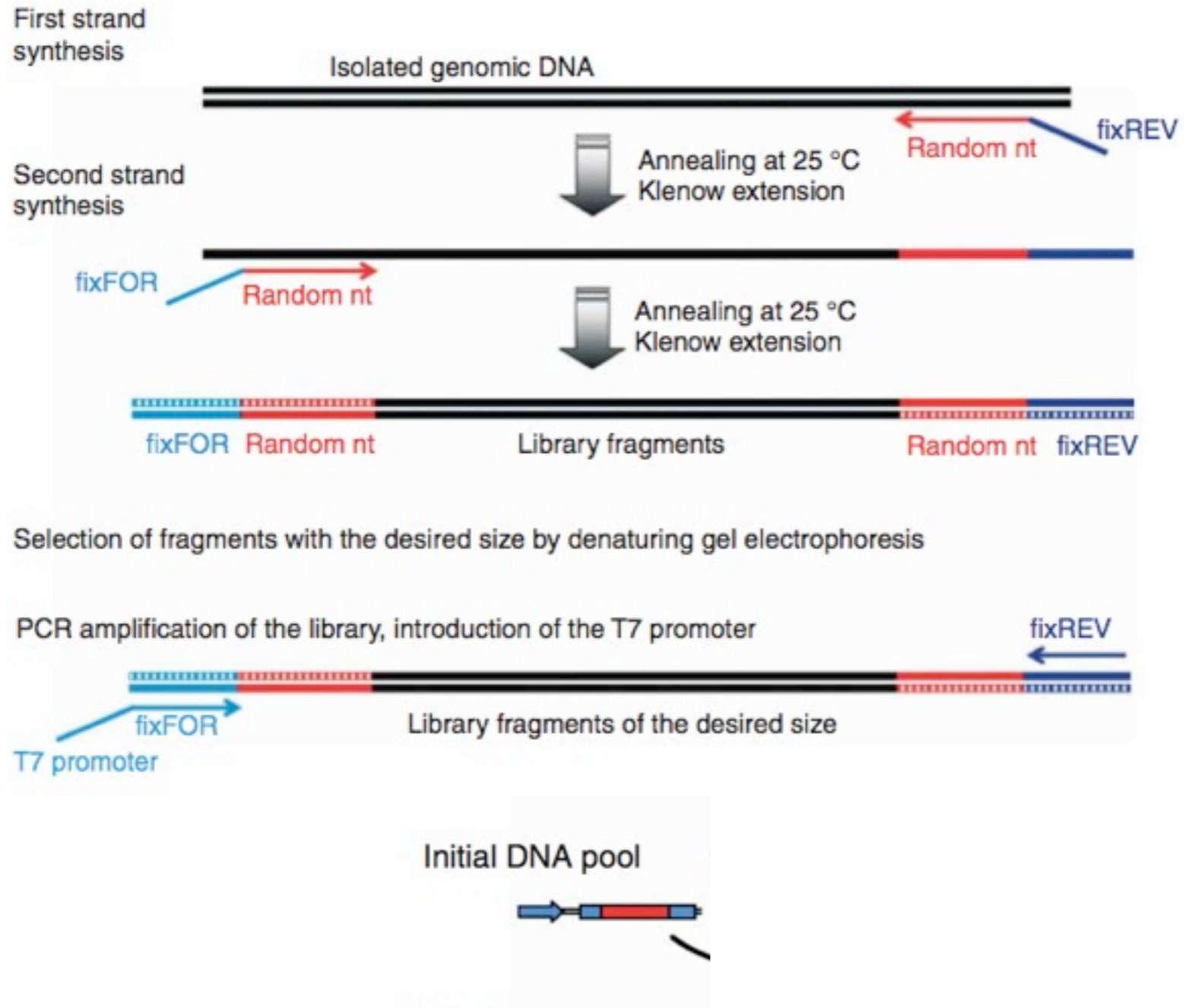


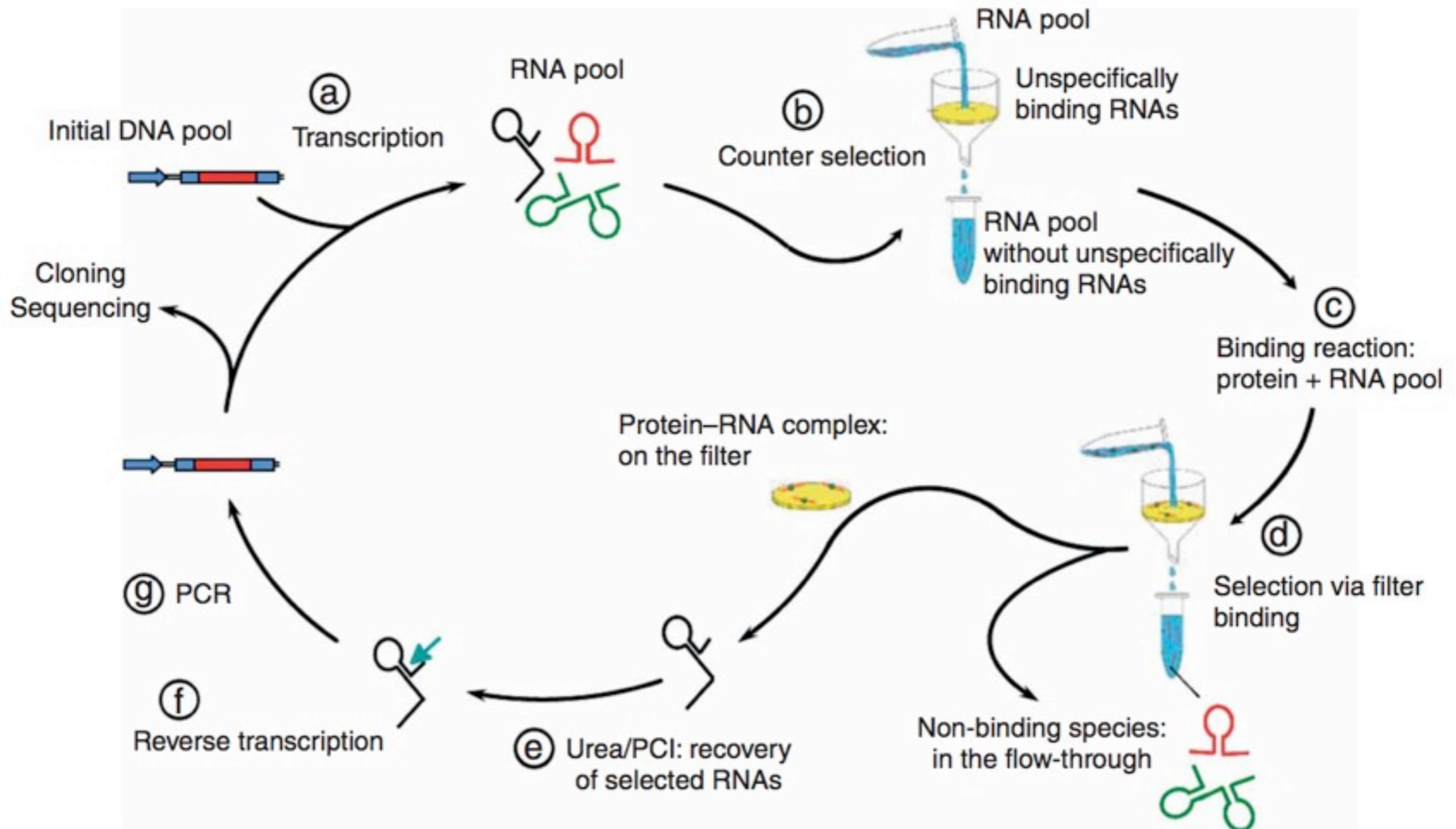
Reverse transcription

(e) Urea/PCI: recovery of selected RNAs

Non-binding species: in the flow-through





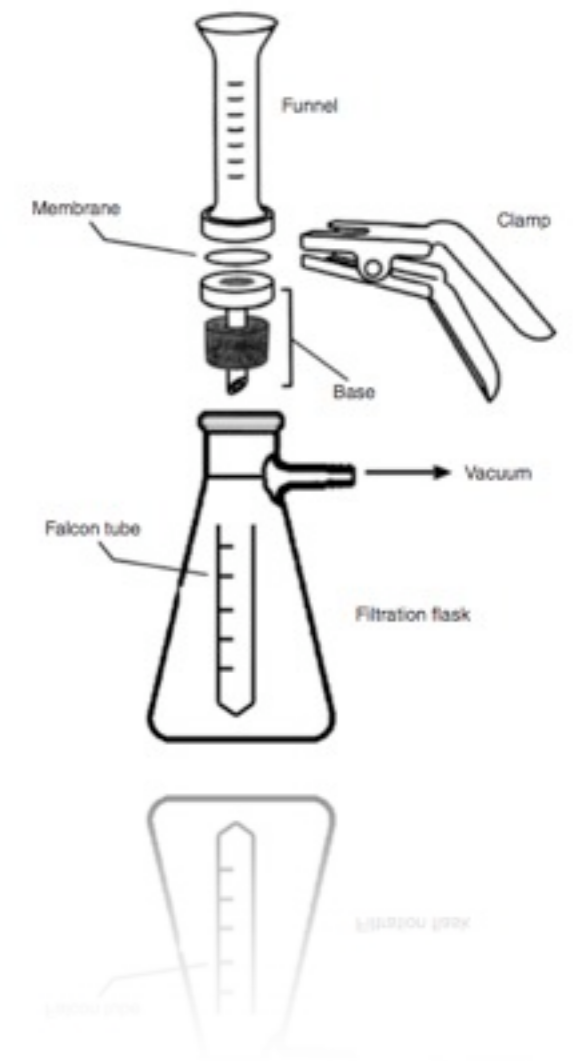
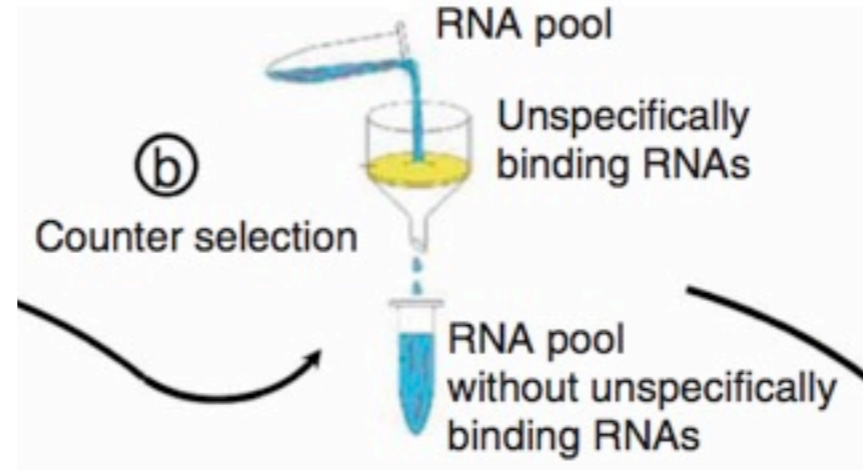


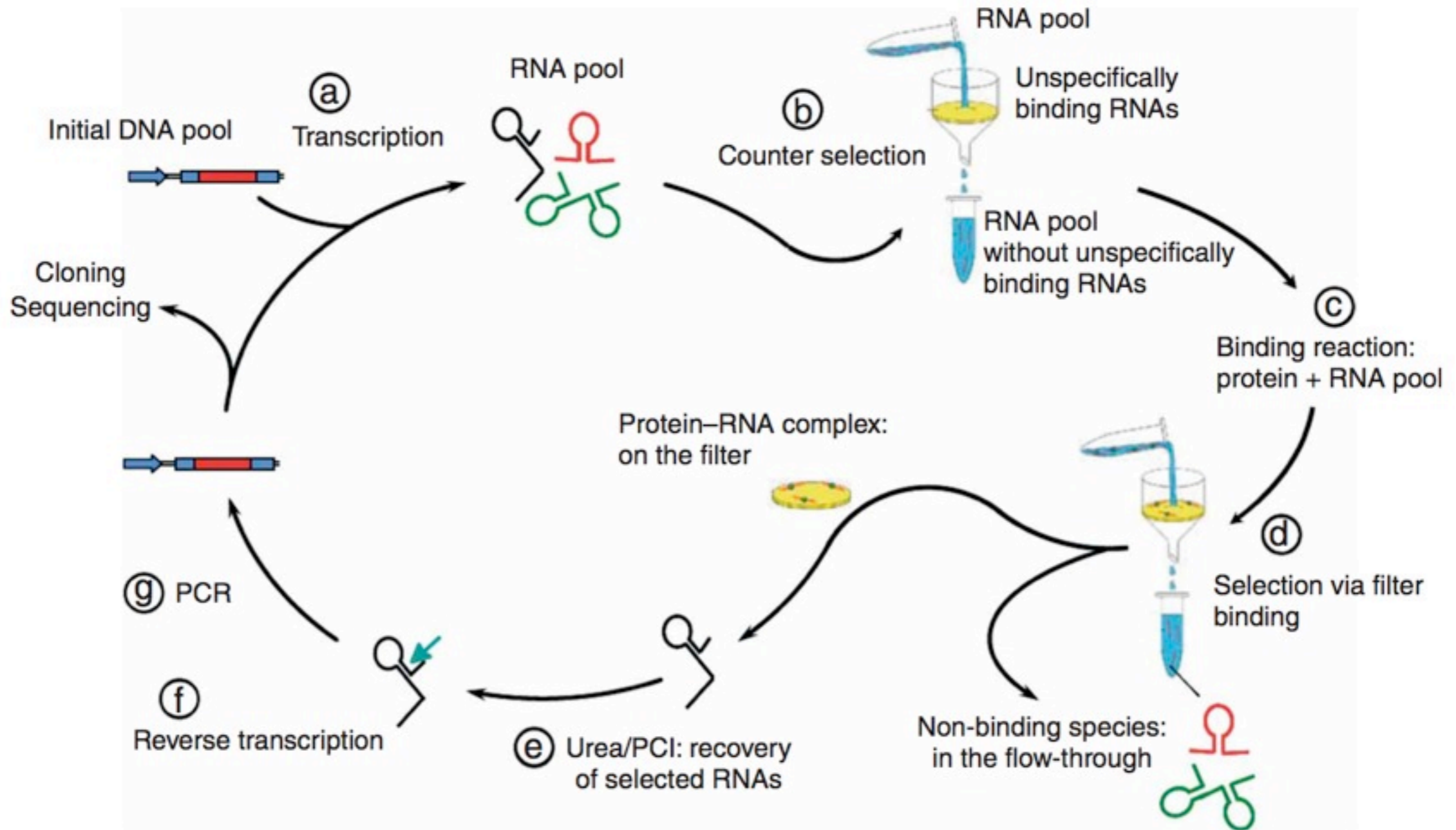
Reverse transcription

(e) Urea/PCI: recovery of selected RNAs

Non-binding species: in the flow-through





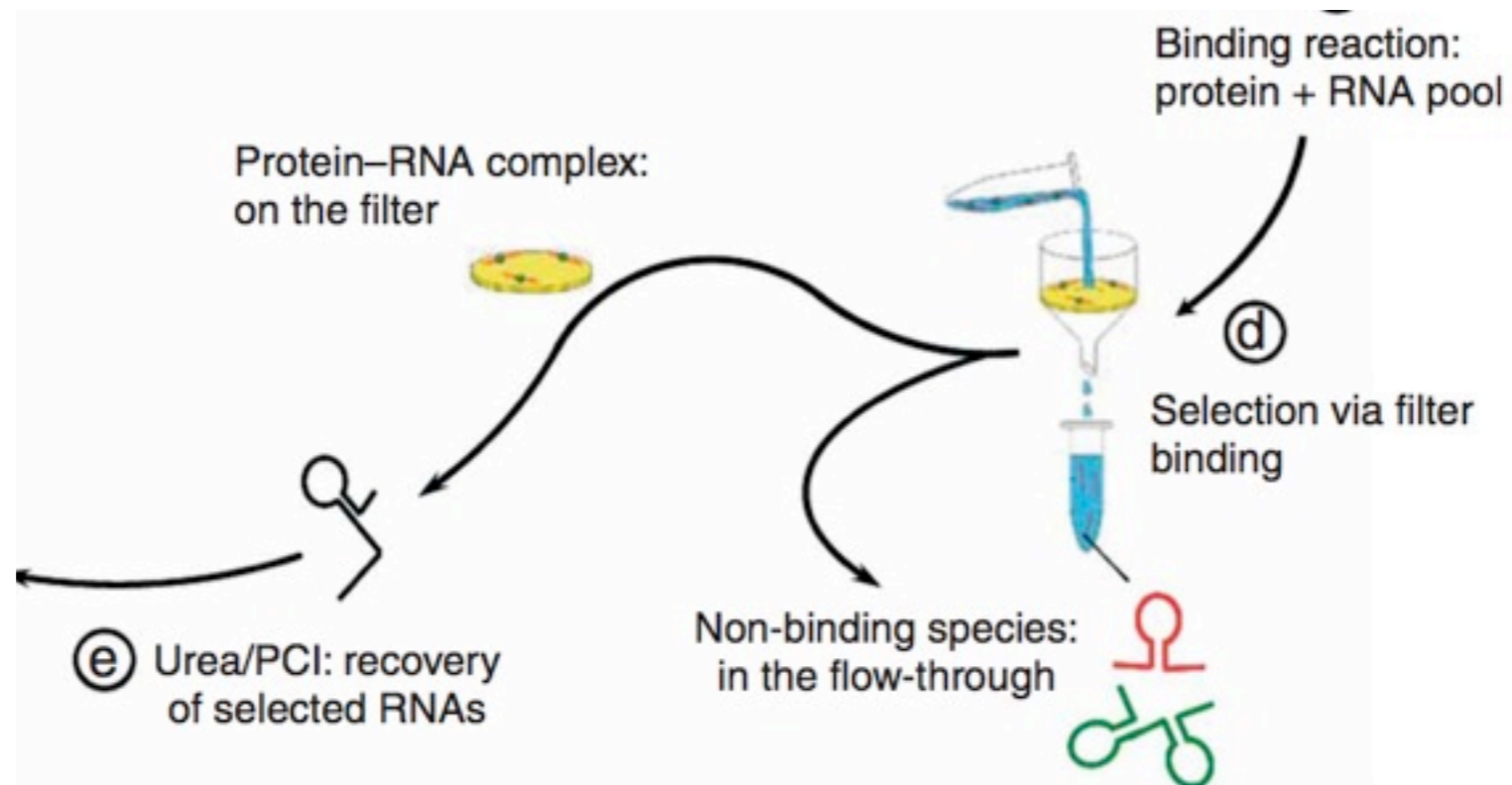


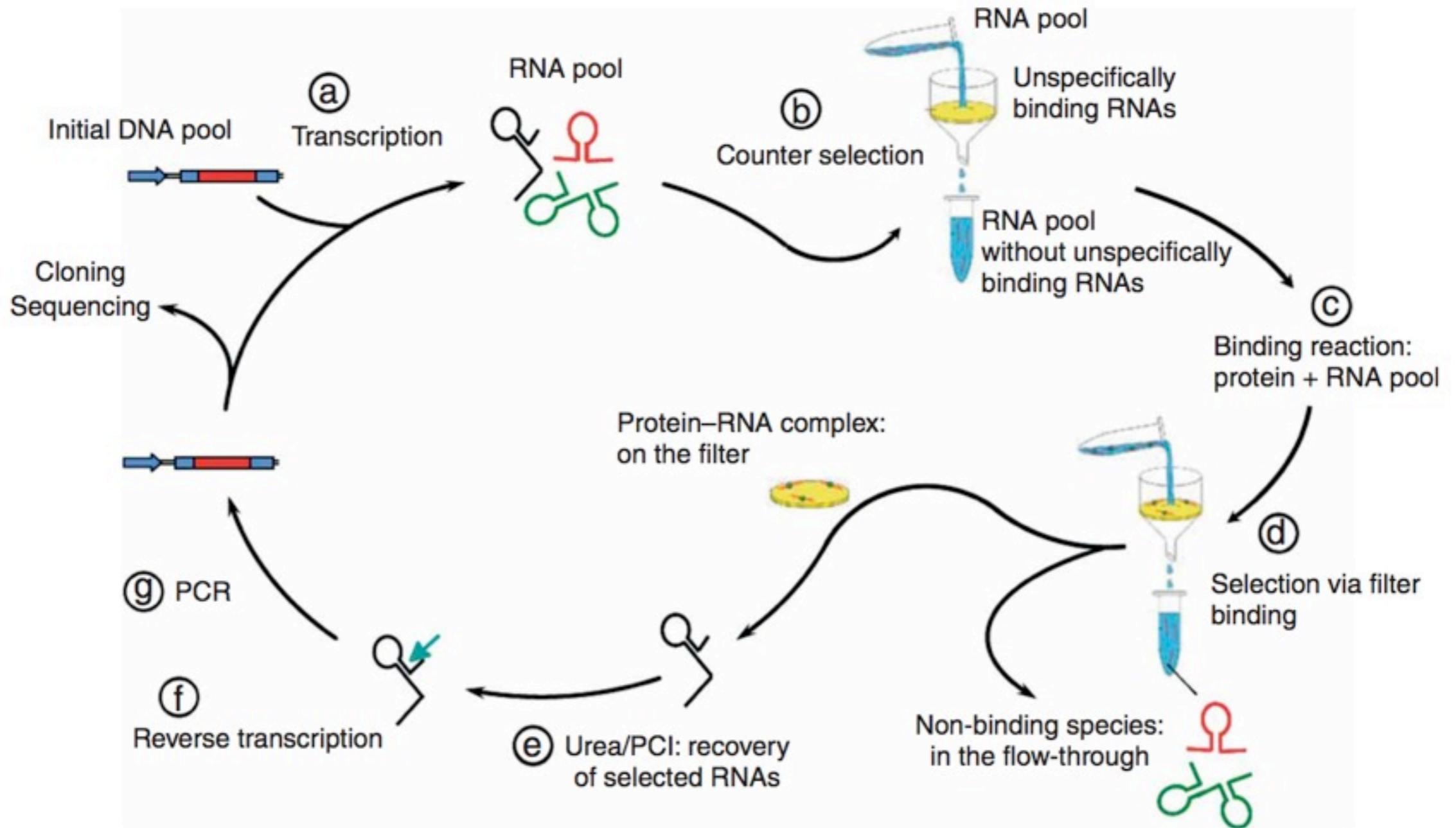
Reverse transcription

(g) Urea/PCI: recovery of selected RNAs

Non-binding species: in the flow-through





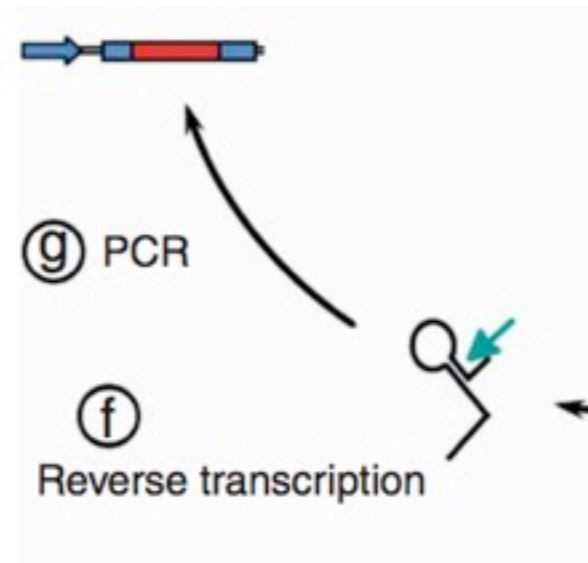


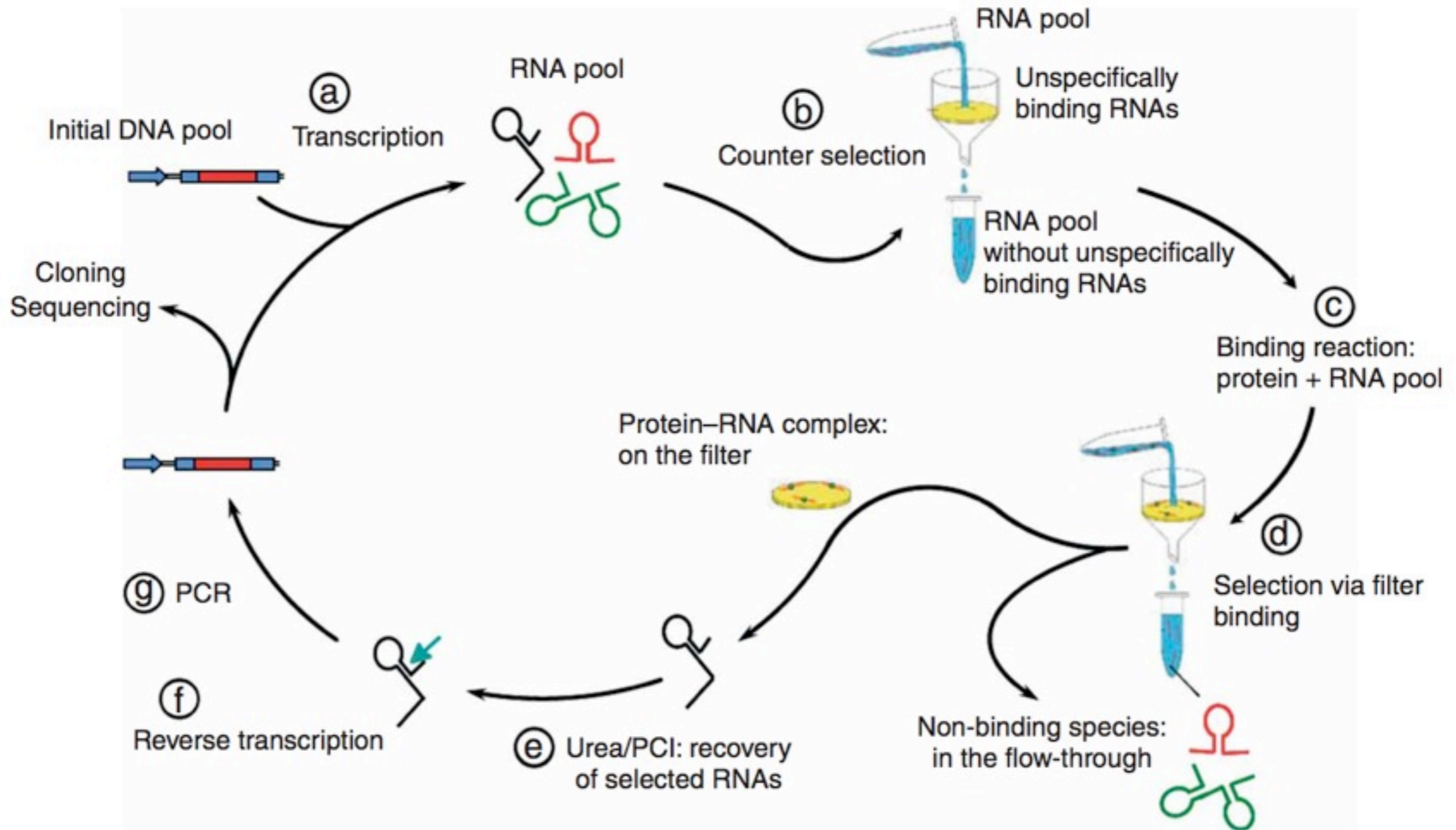
Reverse transcription

(g) Urea/PCI: recovery of selected RNAs

Non-binding species: in the flow-through







Reverse transcription

(g) Urea/PCI: recovery of selected RNAs

Non-binding species: in the flow-through



