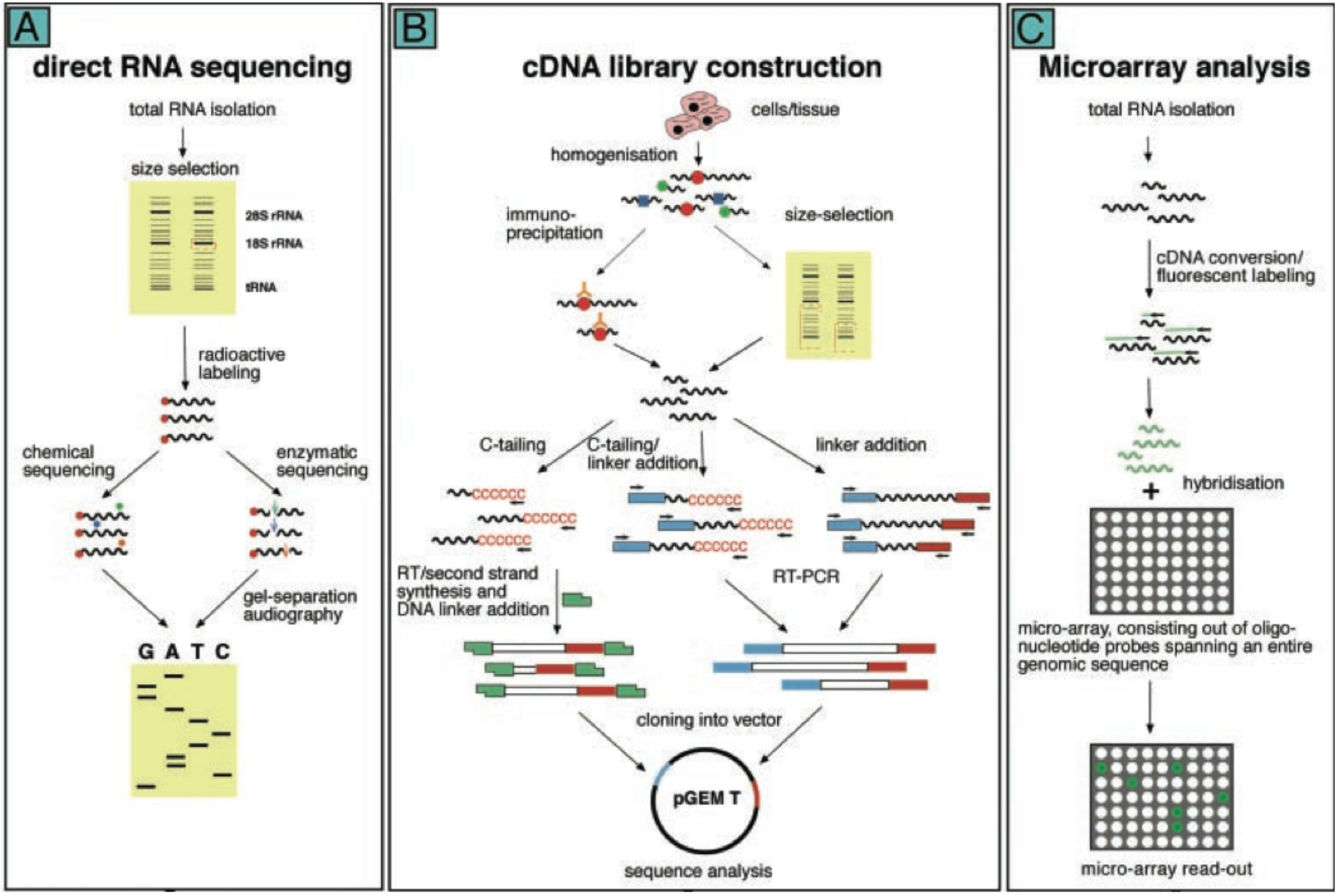
Analysis of High Throughput Data from Genomic SELEX

Bled Conference
Valentine's Day 2012
Bob Zimmermann

universität wien

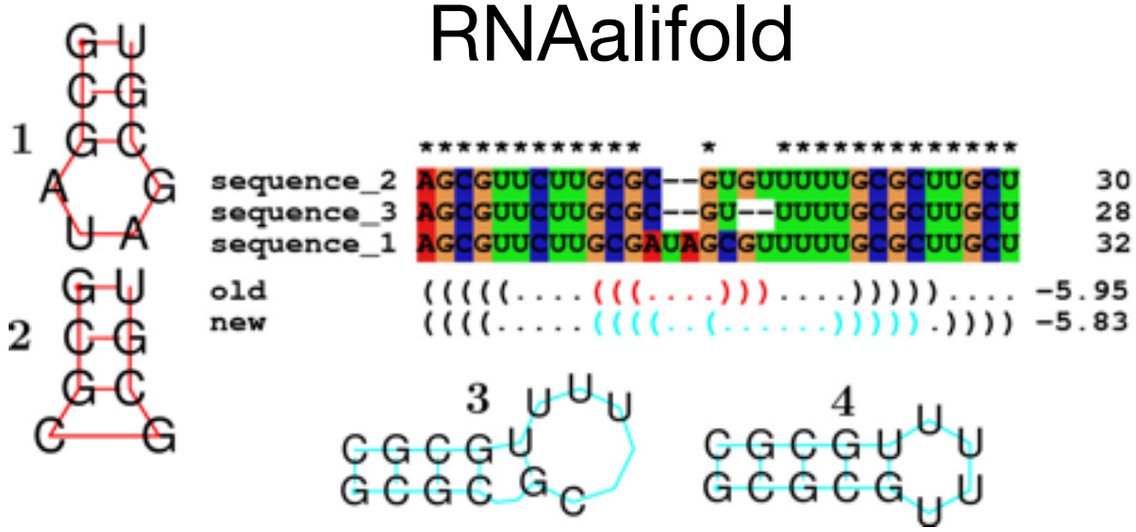
In vivo RNA Discovery



Huttenhofer and Vogel, 2006

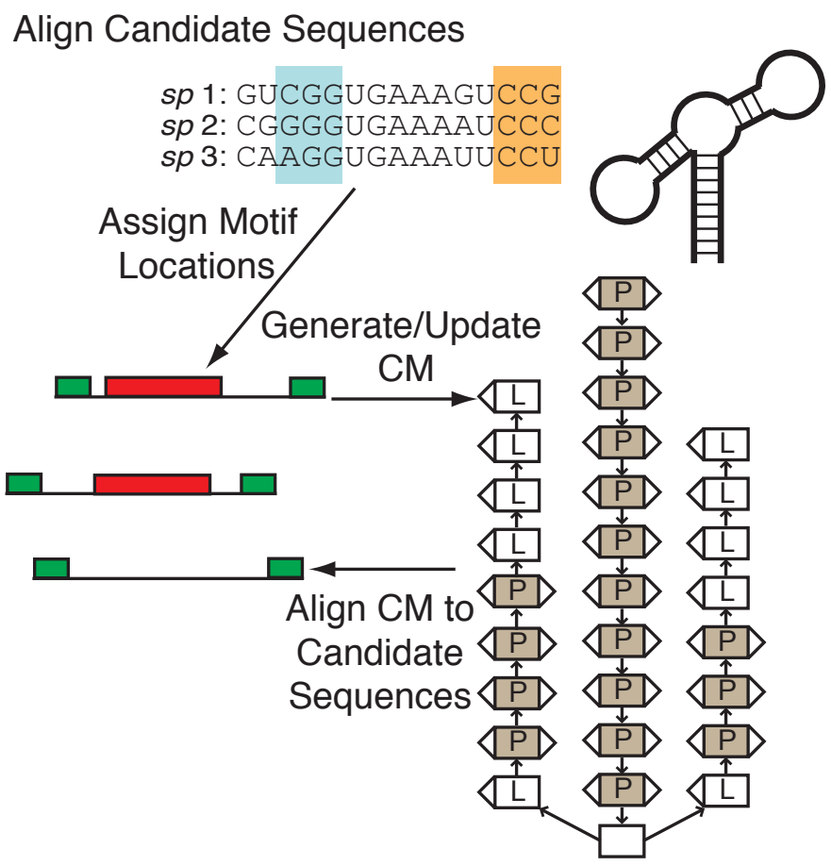
In silico RNA Discovery

RNAalifold

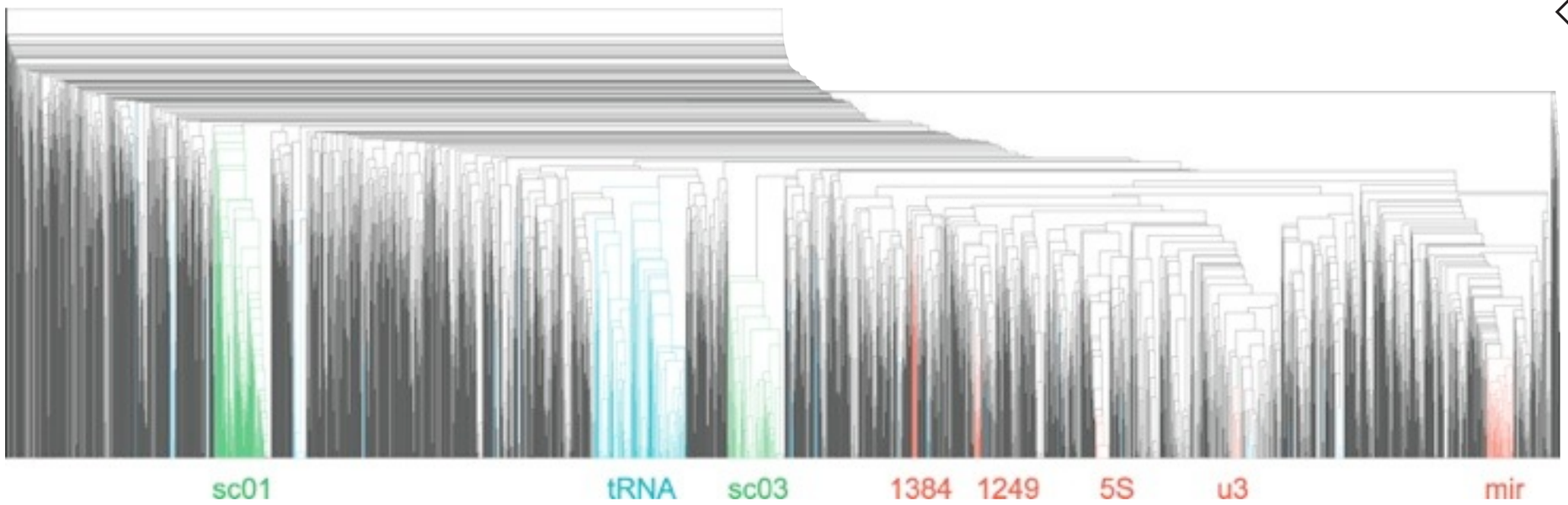


Bernhart et al, 2008

CMfinder



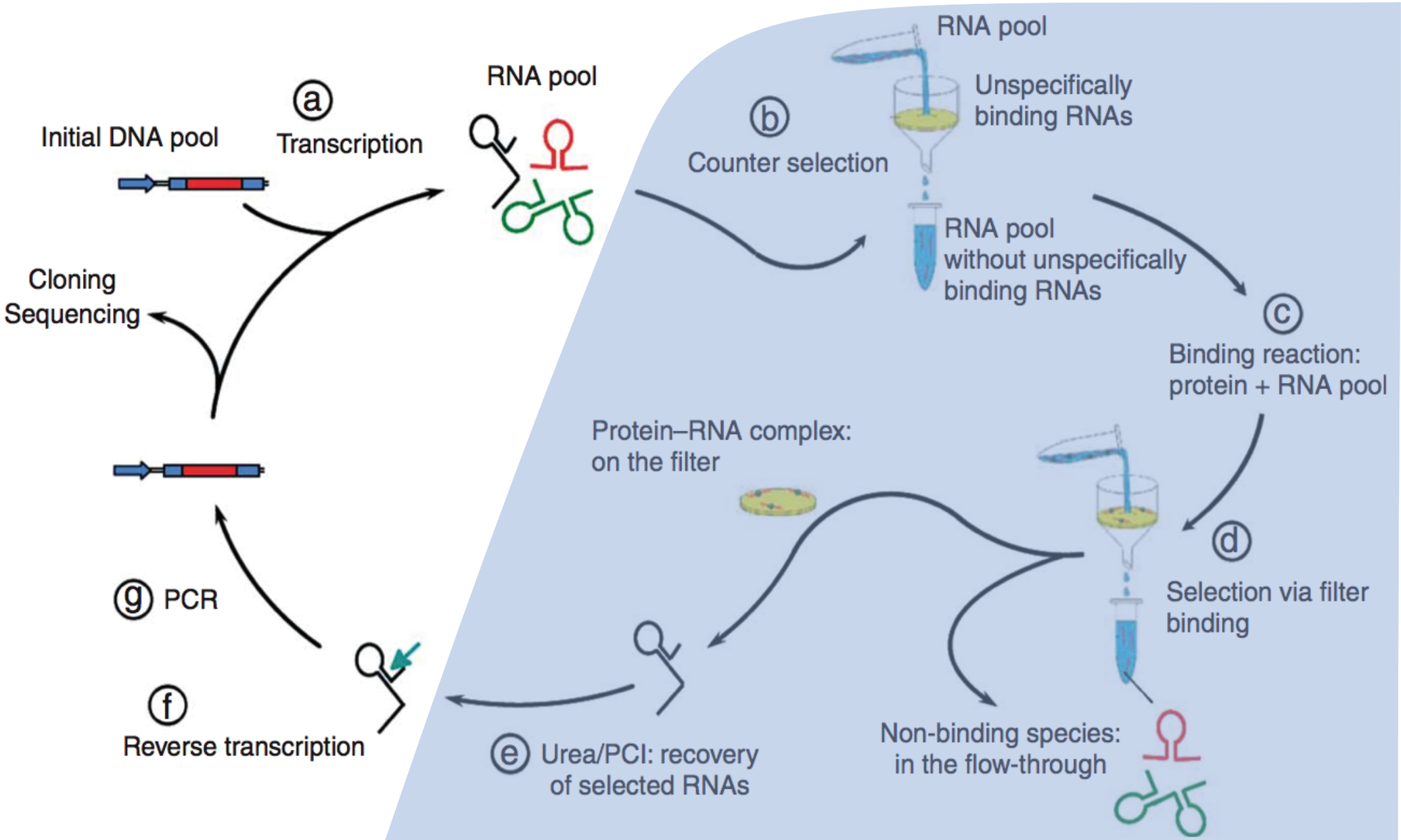
RNAz/mlocarna



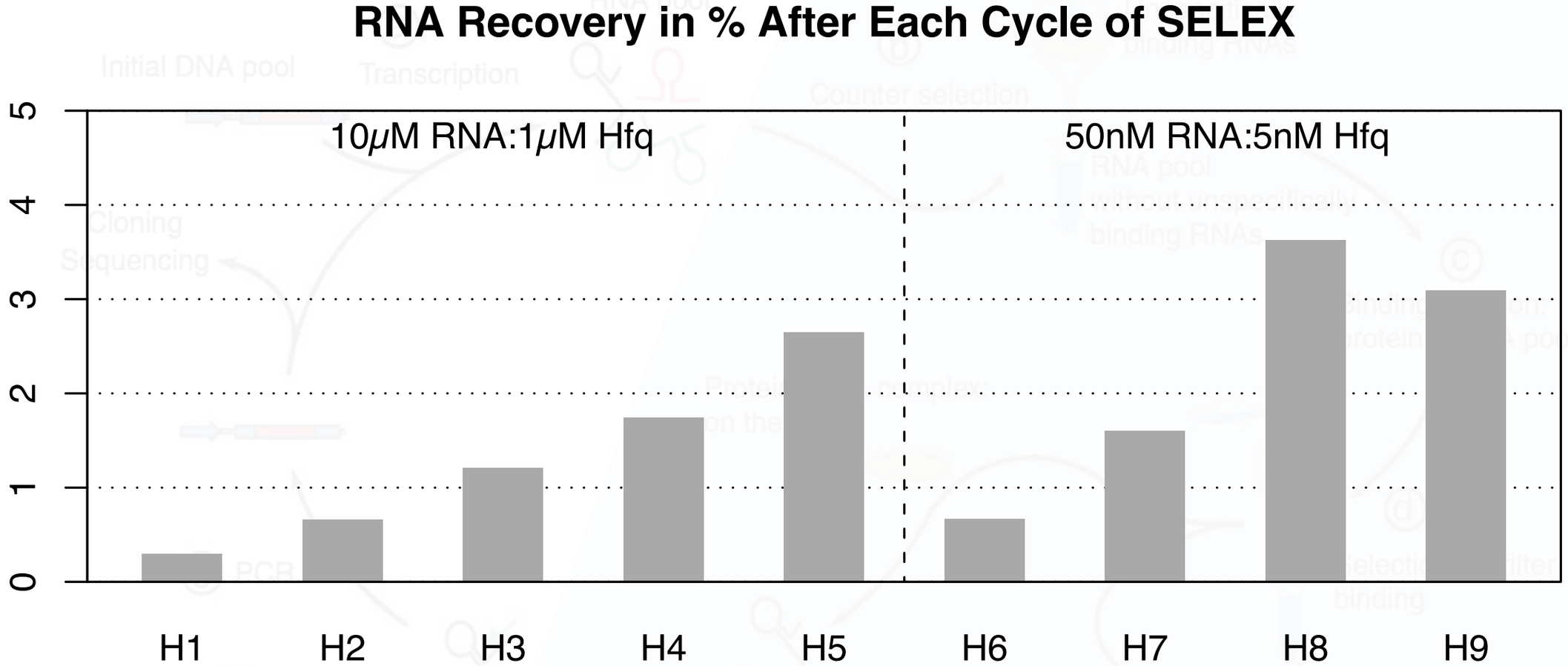
Yao et al, 2006

Will et al, 2007

In vitro RNA Discovery: Genomic SELEX



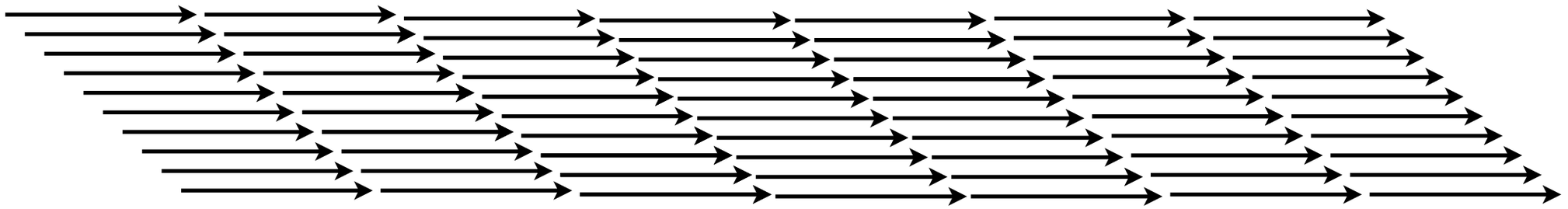
In vitro RNA Discovery: Genomic SELEX



Principle of Selection

Genomic Library

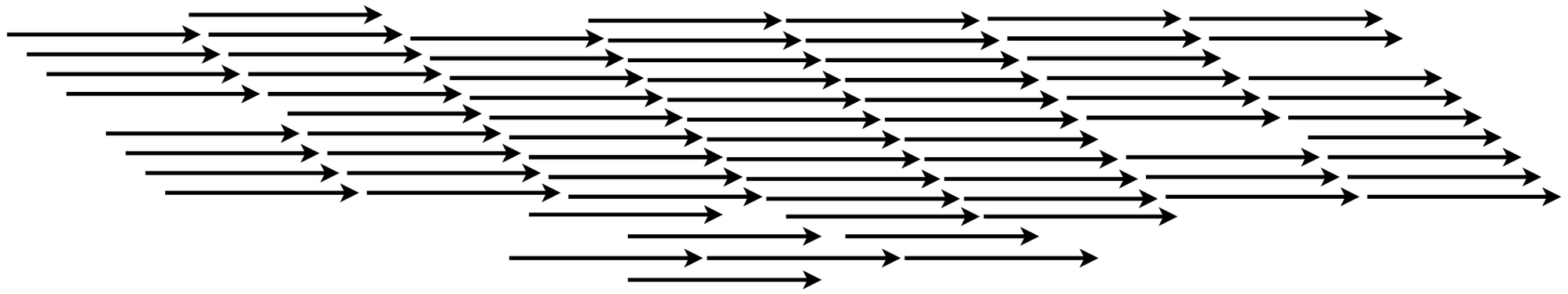
Chromosome



Principle of Selection

Cycle 1

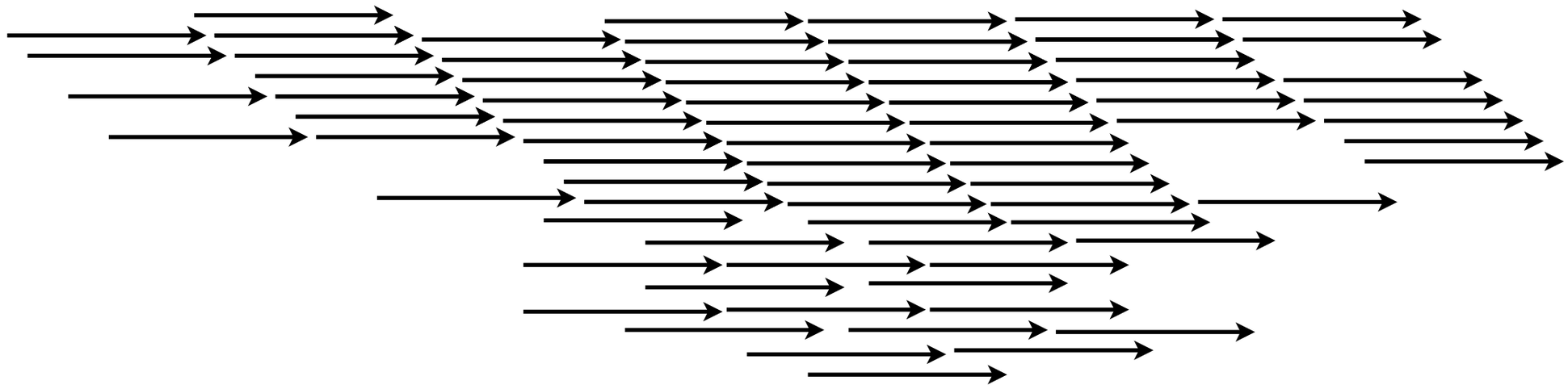
Chromosome



Principle of Selection

Cycle 2

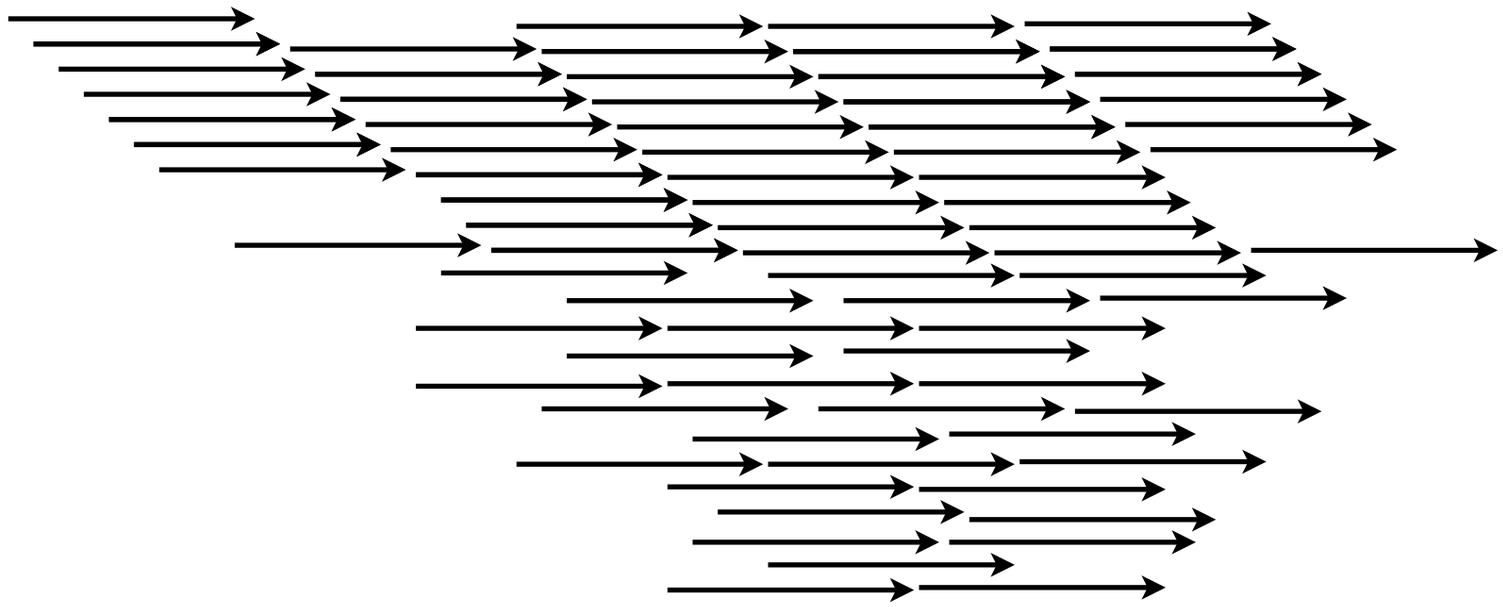
Chromosome



Principle of Selection

Cycle 3

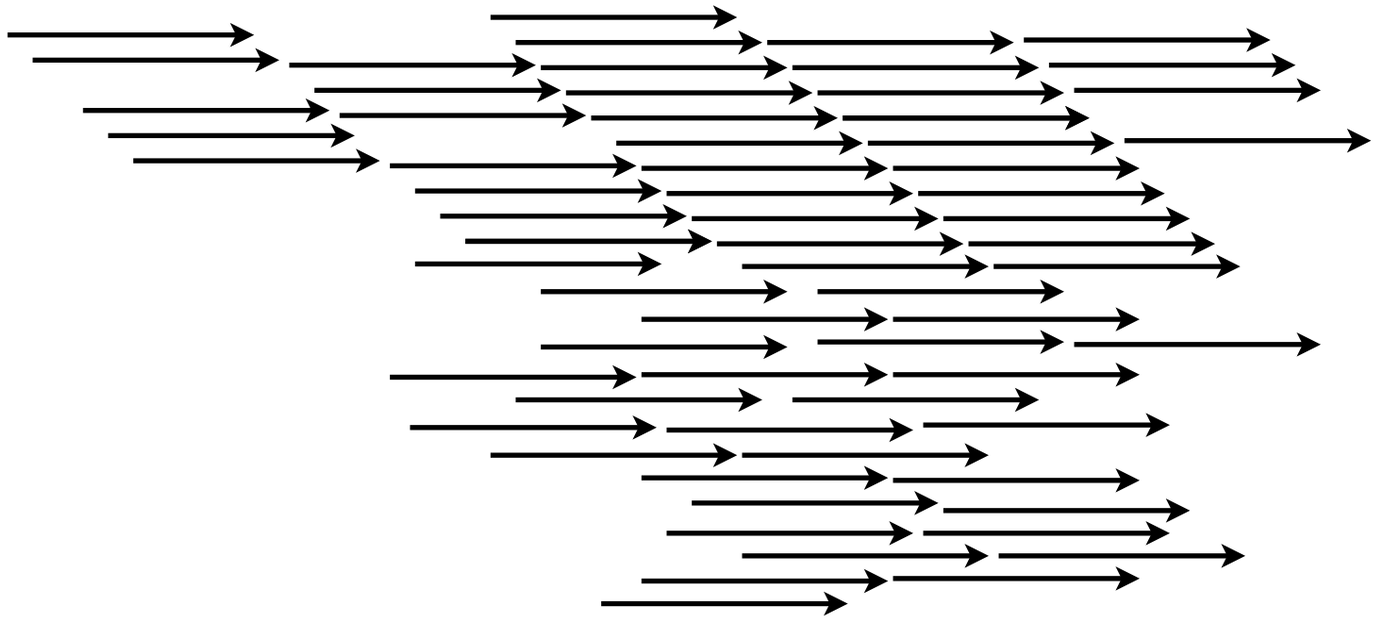
Chromosome



Principle of Selection

Cycle 4

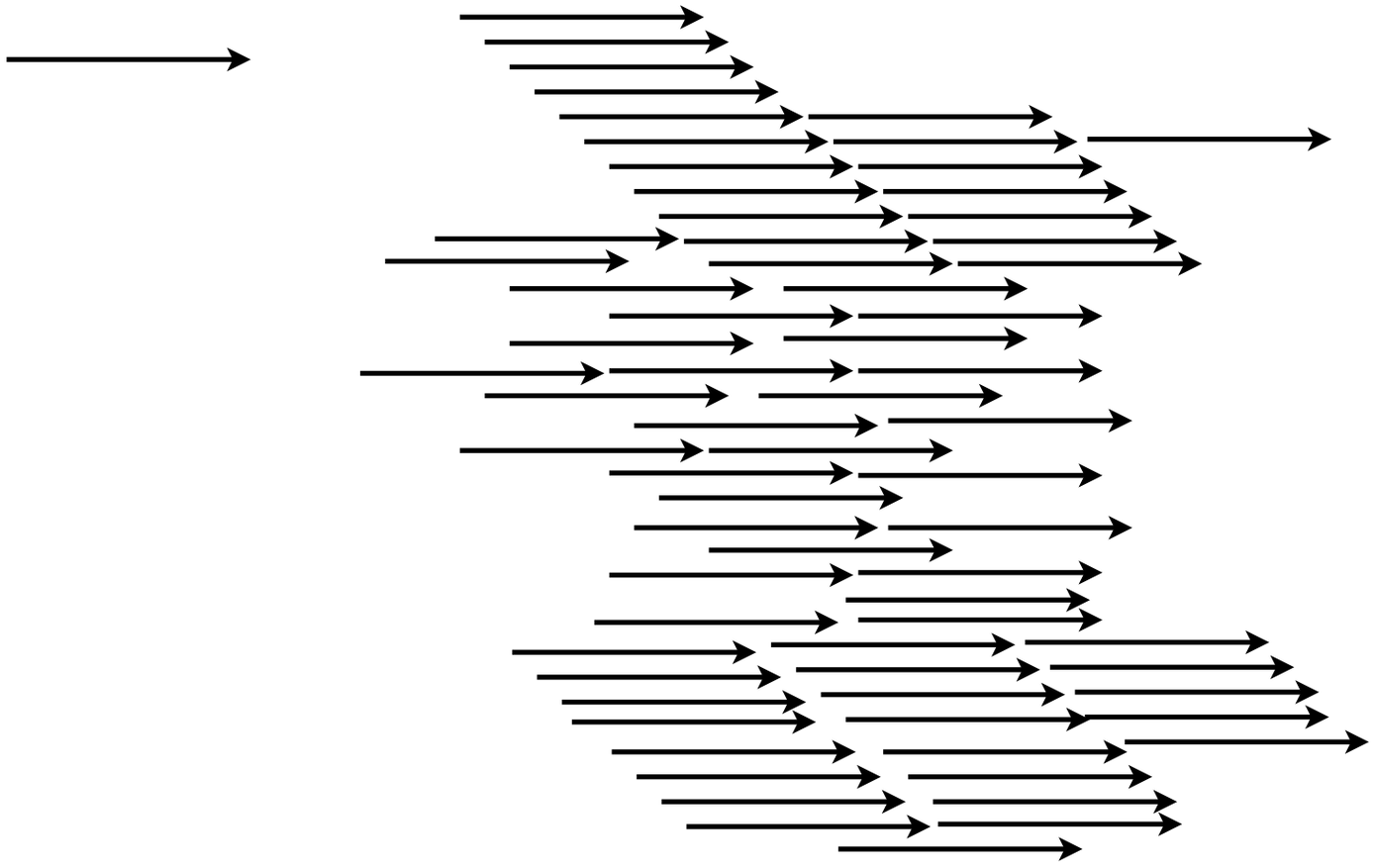
Chromosome



Principle of Selection

Cycle 5

Chromosome

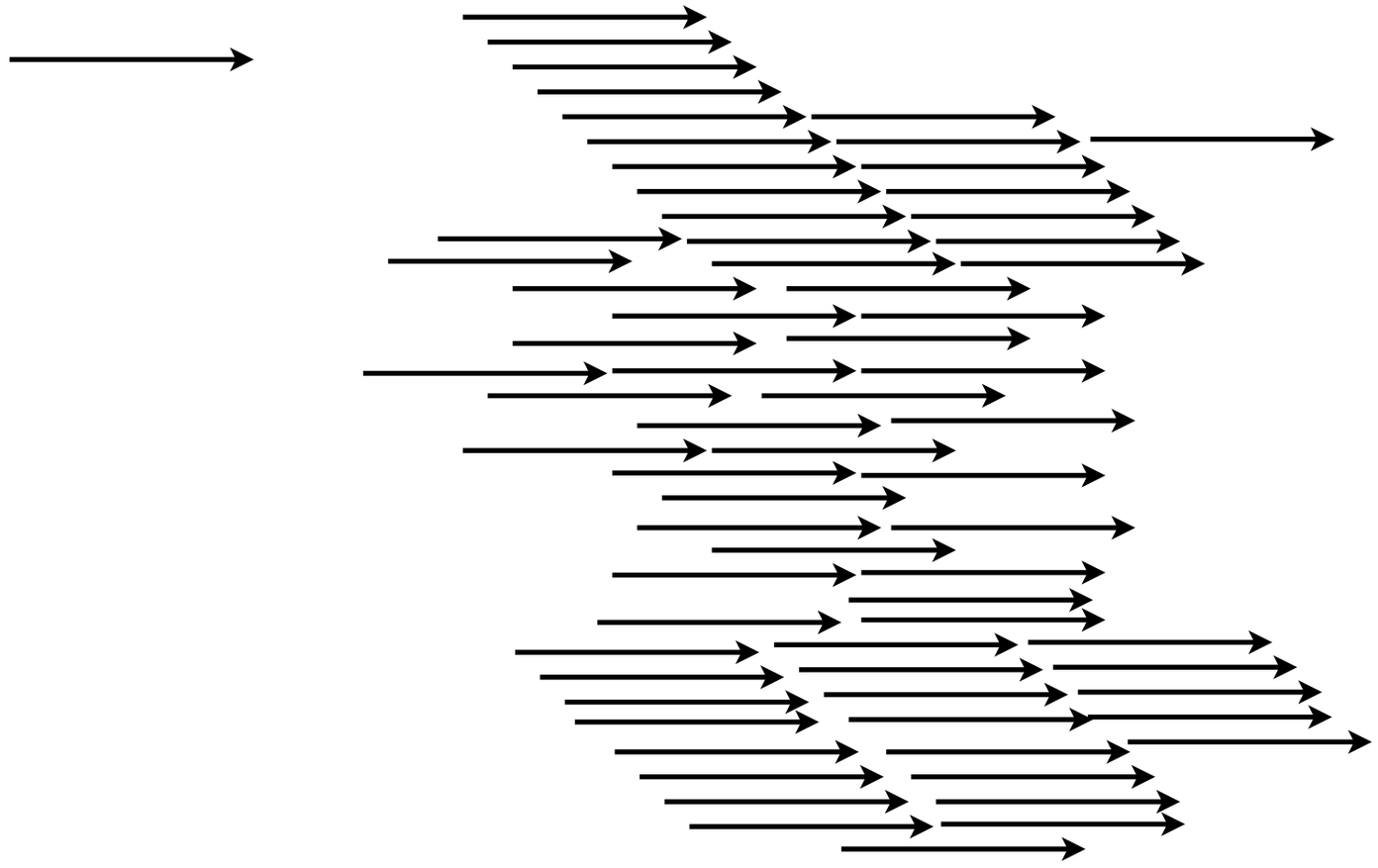


Principle of Selection

Cycle 5

“Genomic Aptamer”

Chromosome



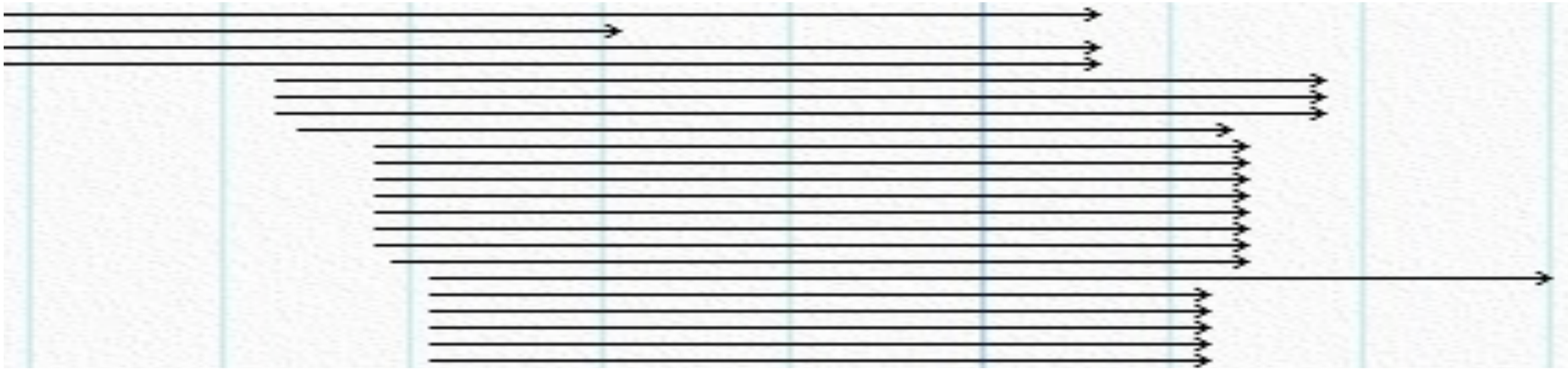
SELEX Experiments

Genome	Protein Target	Scientist(s)
E.coli	None (Neutral SELEX)	Christina Lorenz (library), Bob Zimmermann (Experiment, Analysis), Tanja Gesell (Analysis)
E.coli	Hfq	Christina Lorenz
E.coli	*E.coli* RNA Polymerase	Christina Lorenz (library), Frederike von Pelchrzim
S.cerevisiae	*S.cerevisiae* RNA Polymerase II	Frederike von Pelchrzim
Human	*S.cerevisiae* RNA Polymerase II	Frederike von Pelchrzim

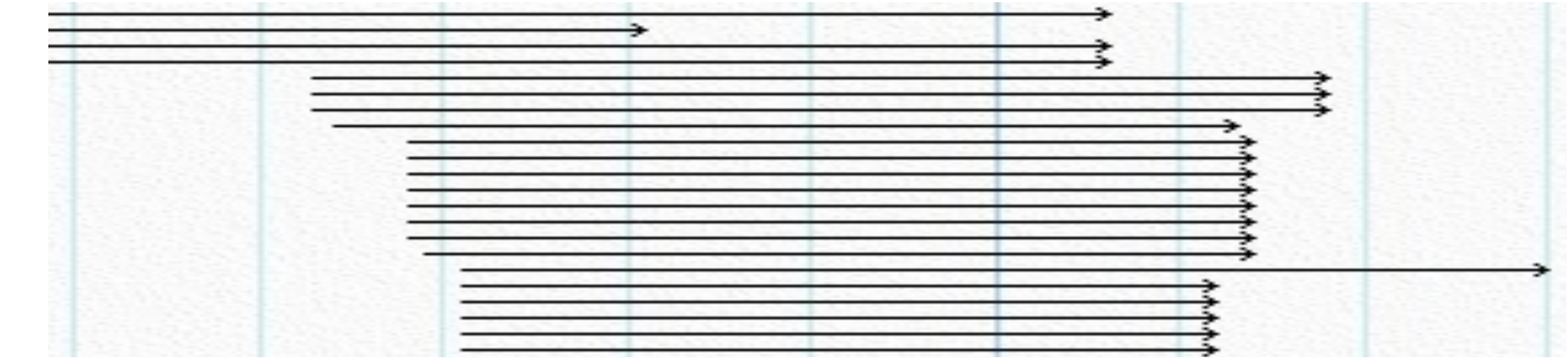
How to interpret the HITS results

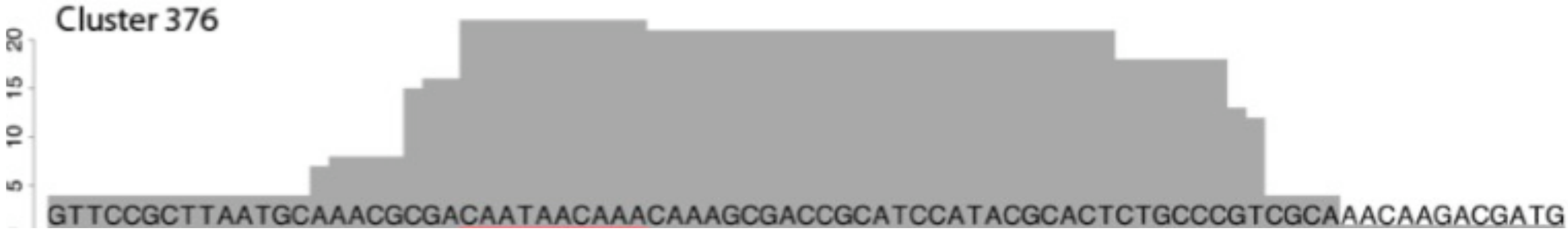
- Location of aptamer domains
- Functional clues from annotations
- Binding motif discovery
- Structural analysis

Enrichment Sequences

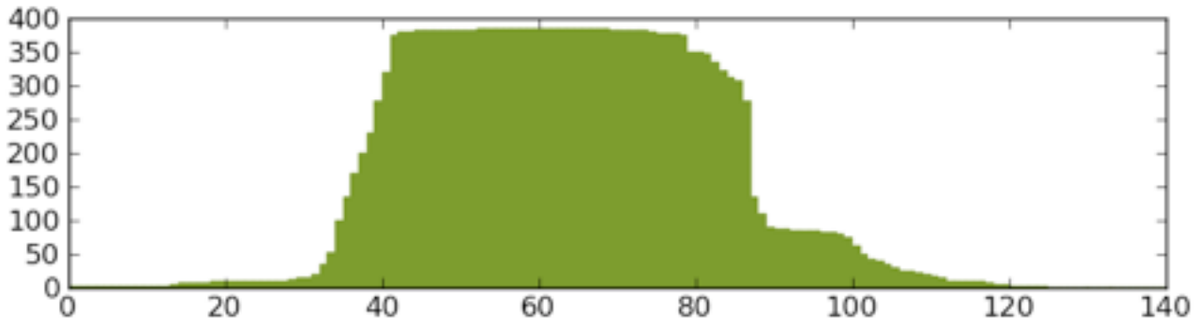


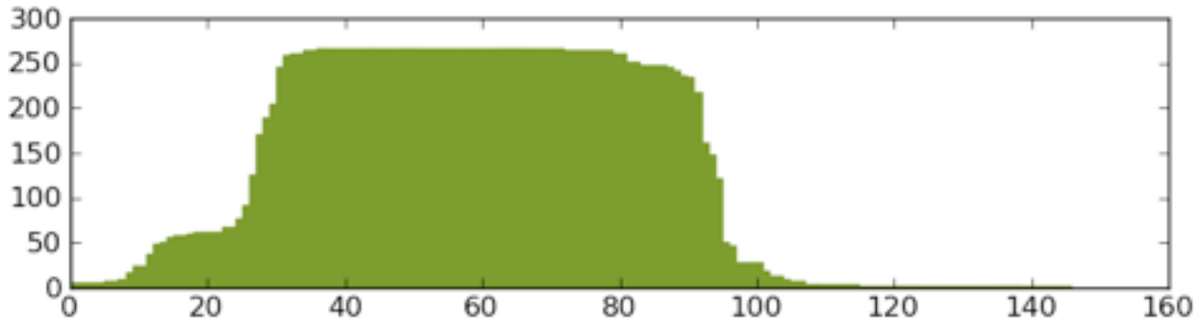
Enrichment Sequences

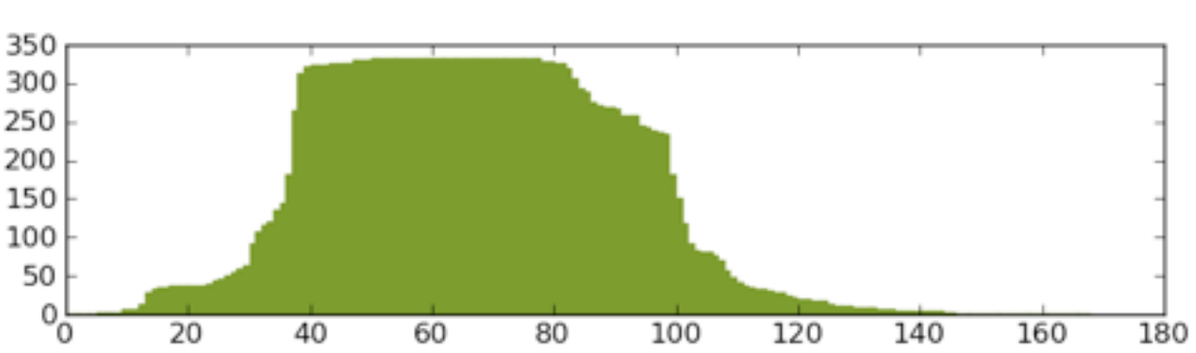


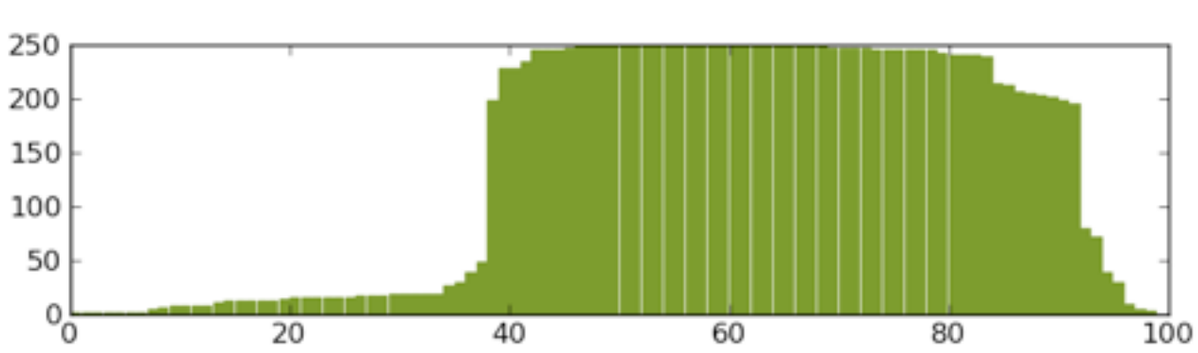


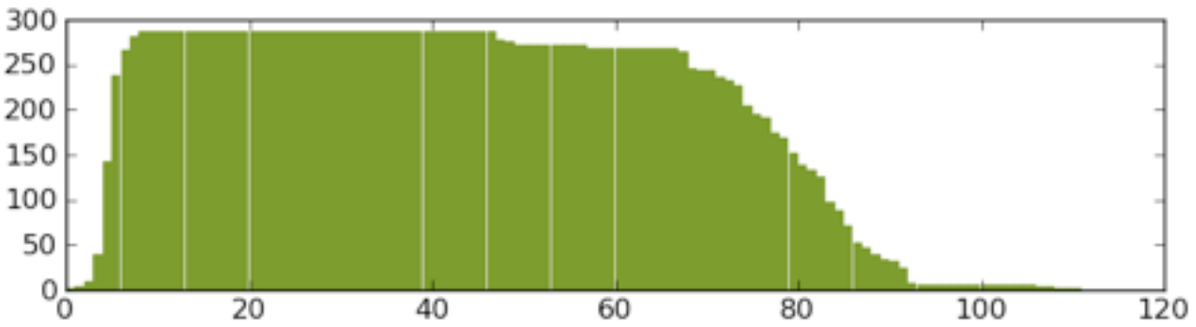
Finding Genomic Aptamer Boundaries

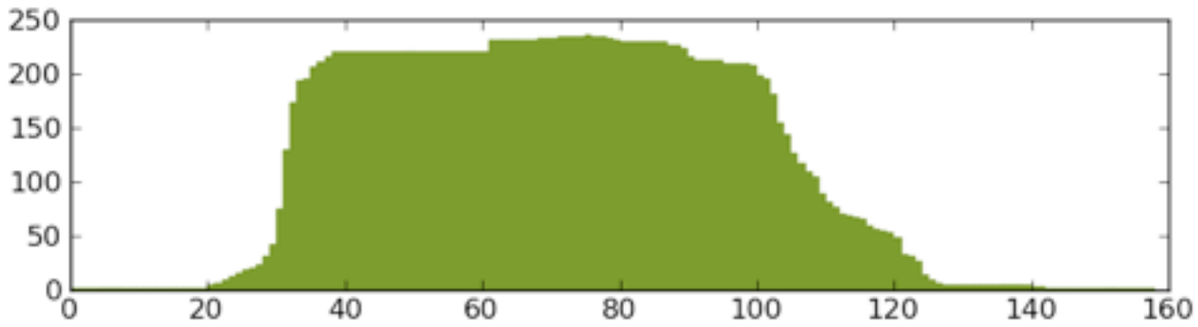


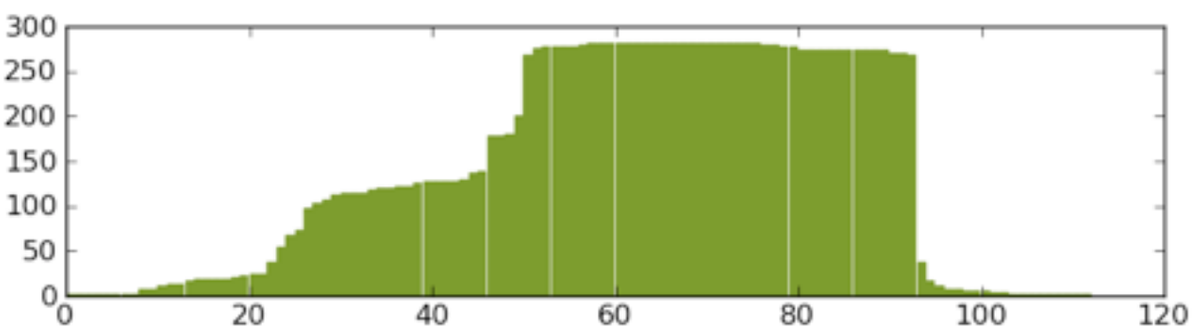


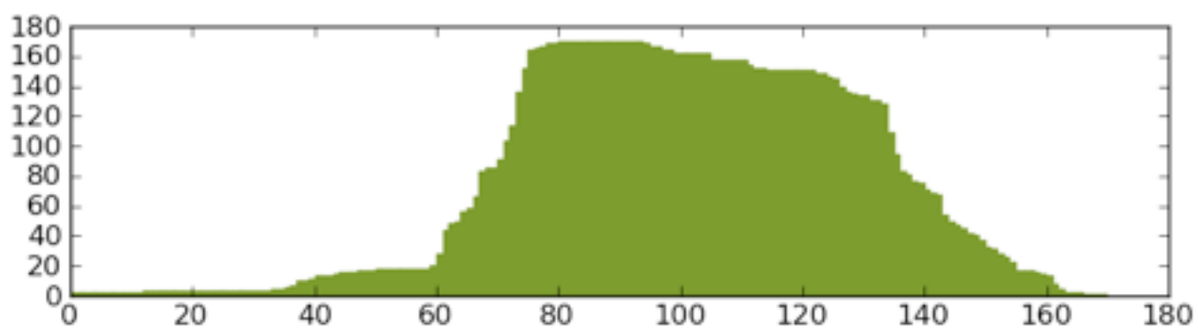


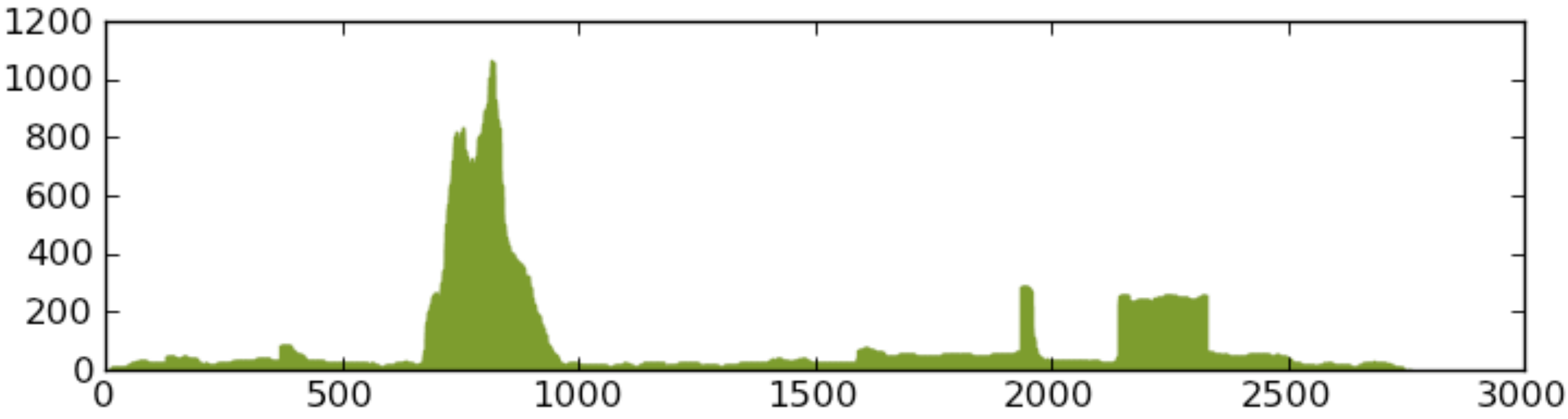


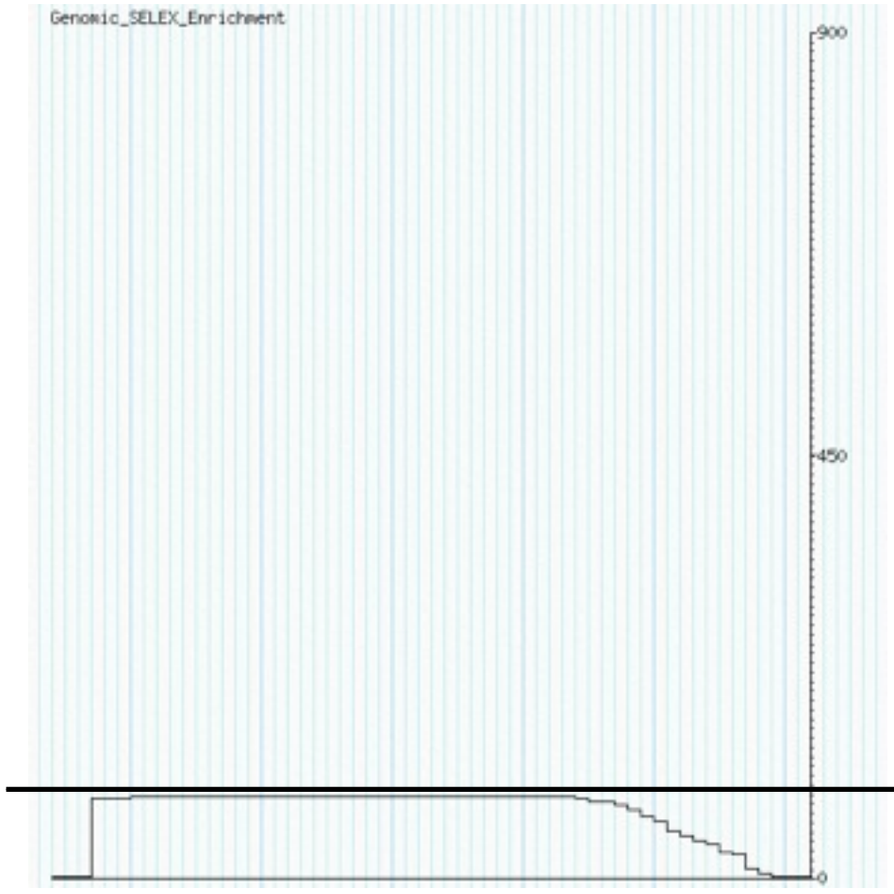


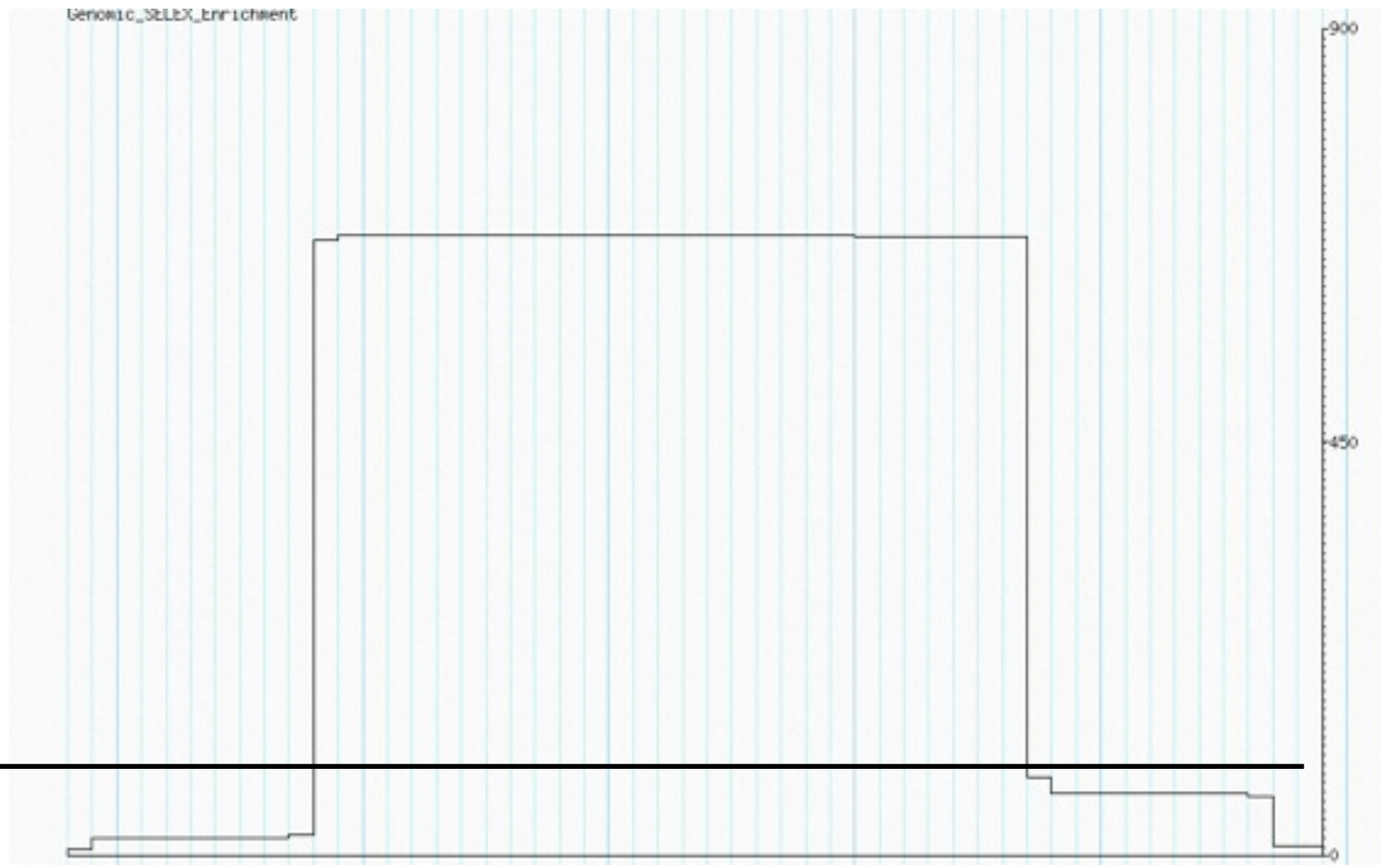


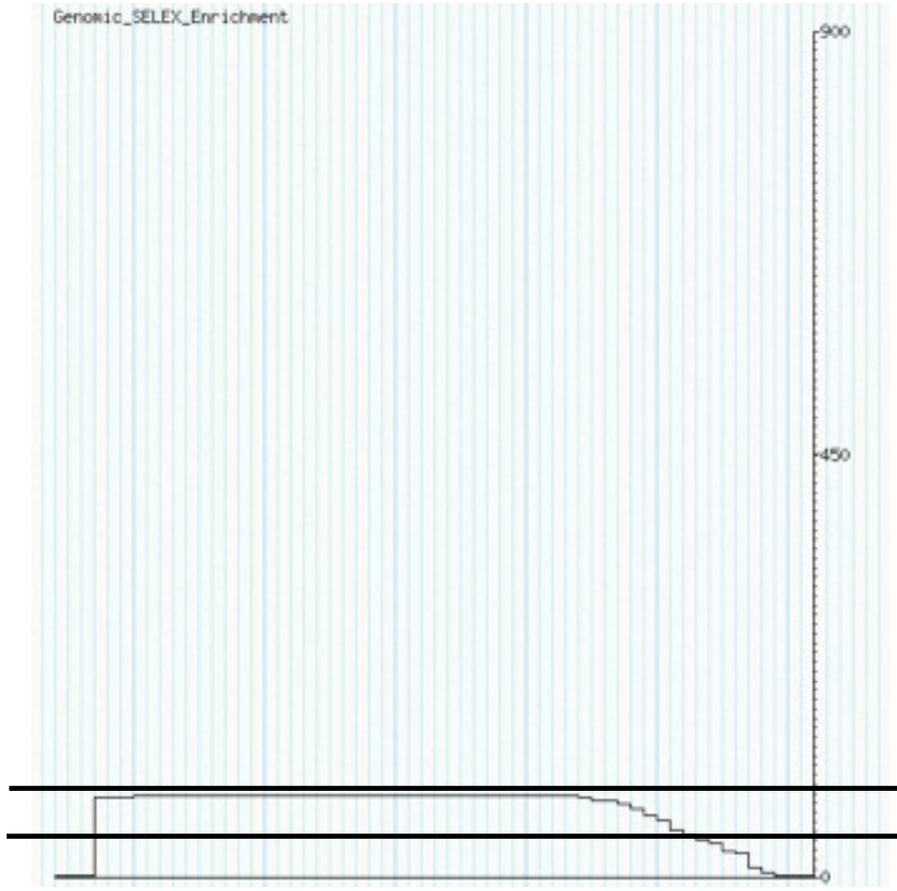


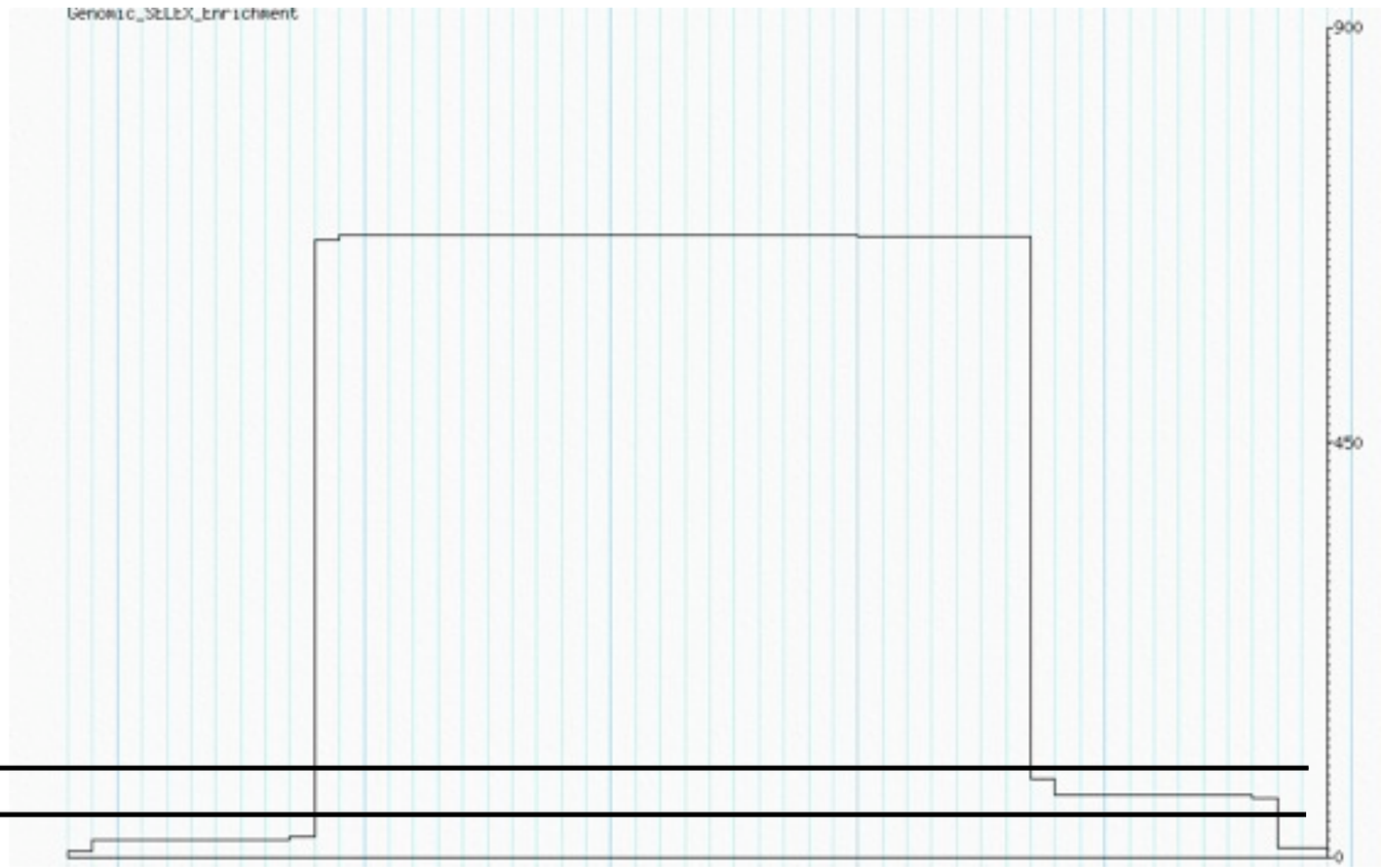


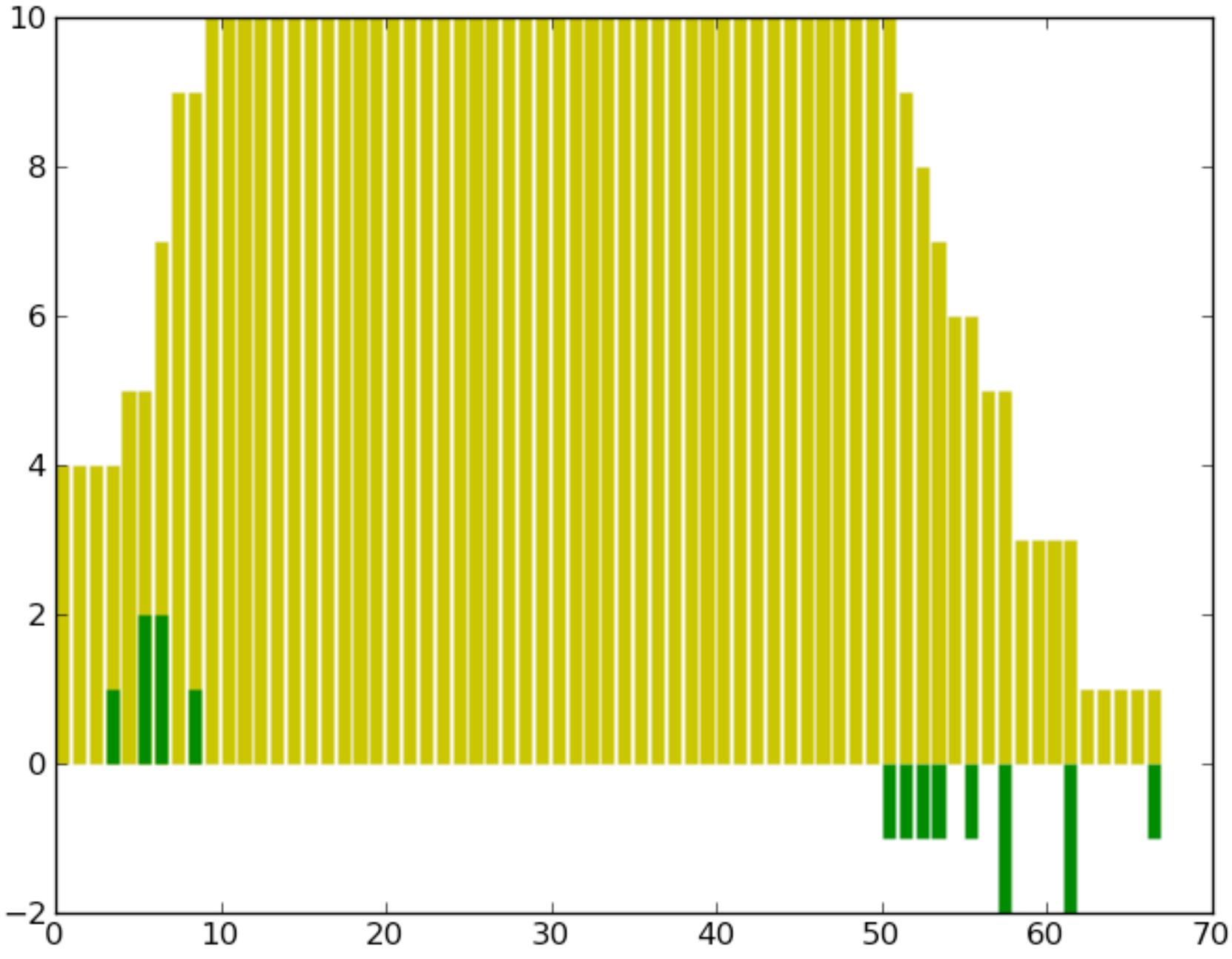

1200
1000
800
600
400
200
0
0
500
1000
1500
2000
2500
3000

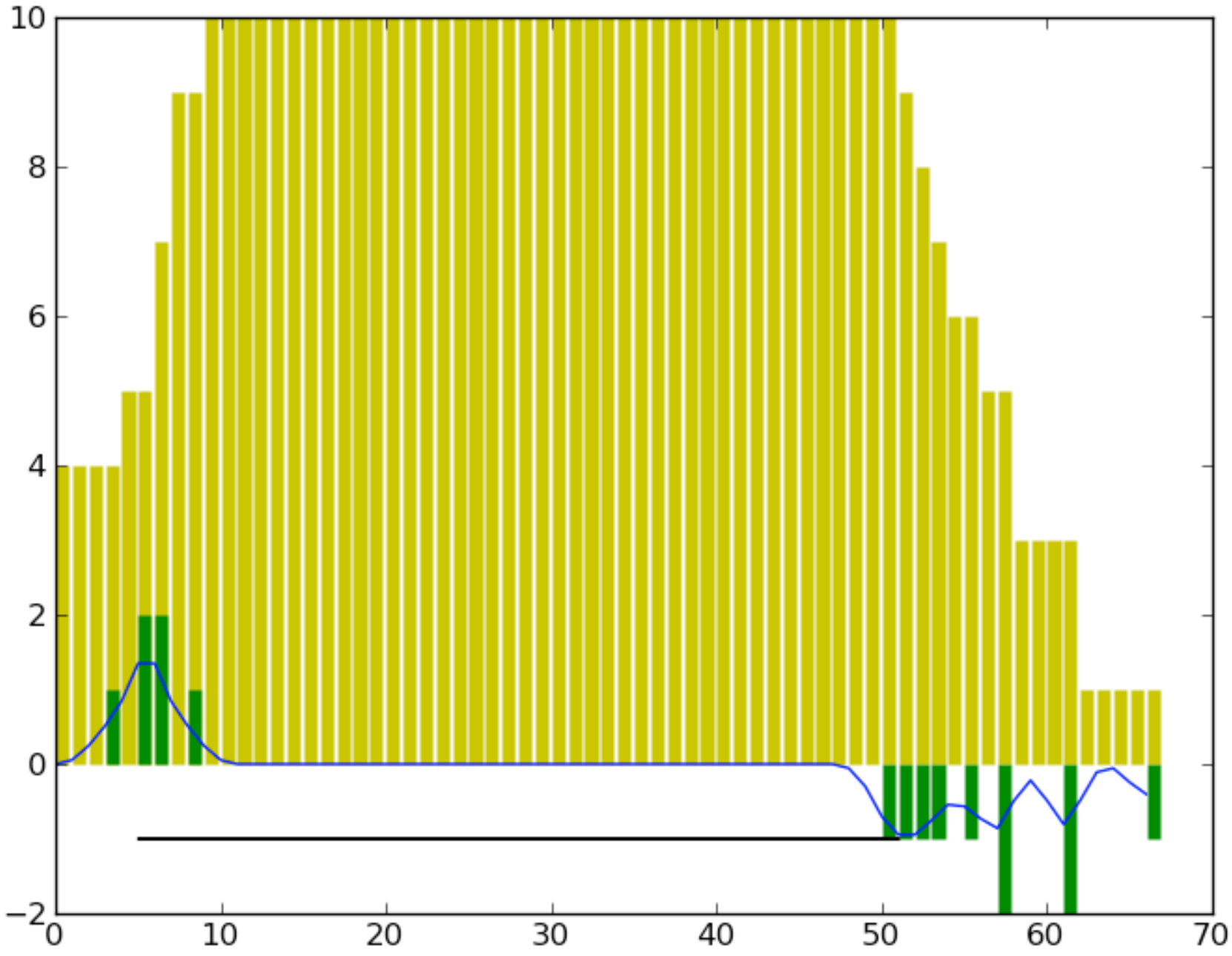

Genomic_SELEX_Enrichment
900
450


Genomic_SELEX_Enrichment
900
450

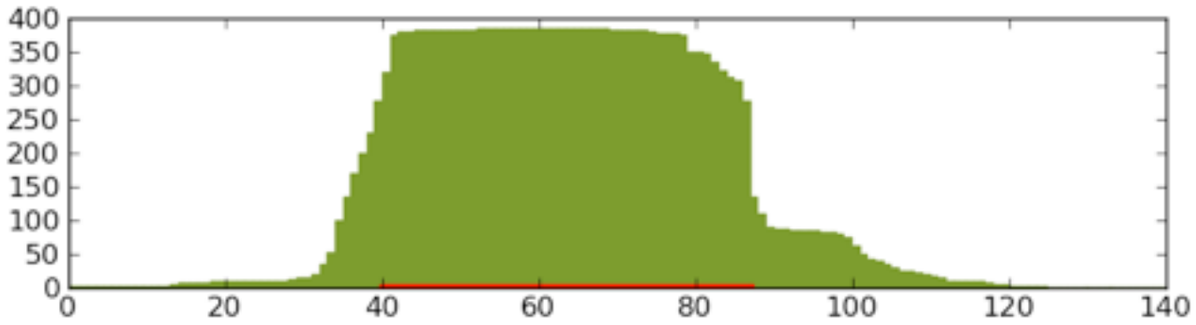

Genomic_SELEX_Enrichment
900
450

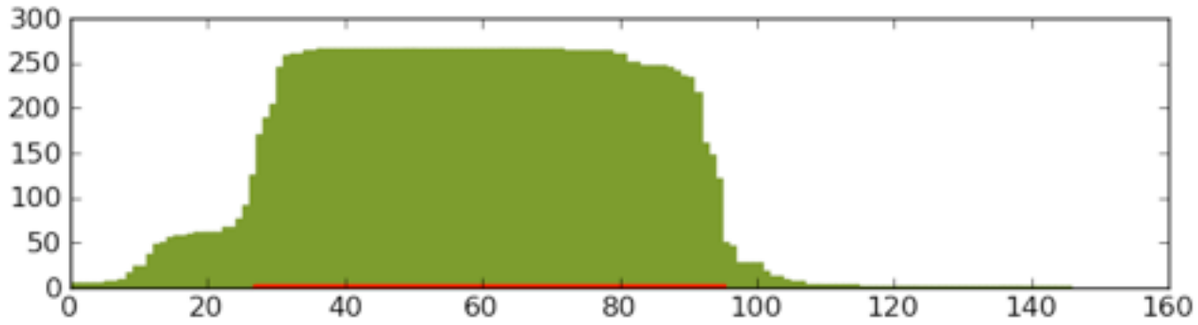

Genomic_SELEX_Enrichment
900
450

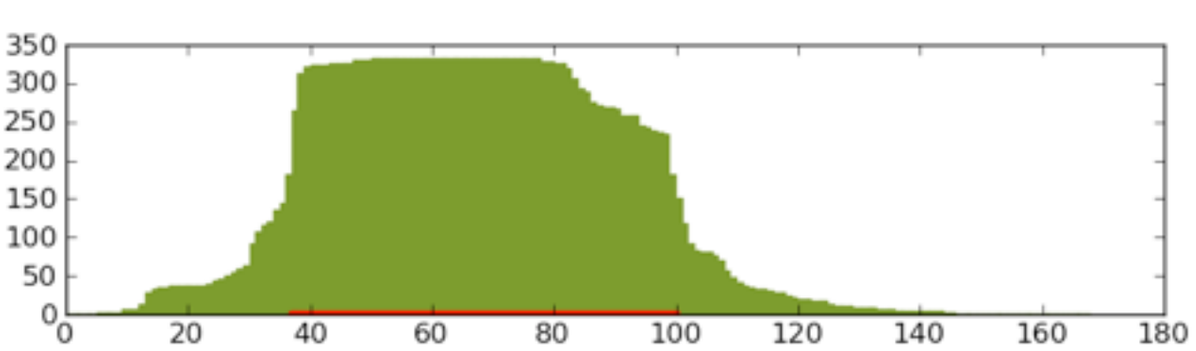

10
8
6
4
2
0
−2
0
10
20
30
40
50
60
70

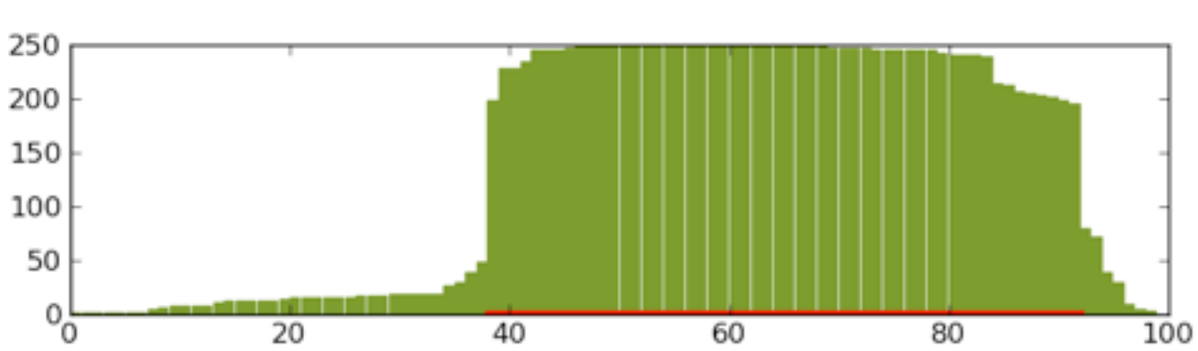

10
8
6
4
2
0
−2
0
10
20
30
40
50
60
70

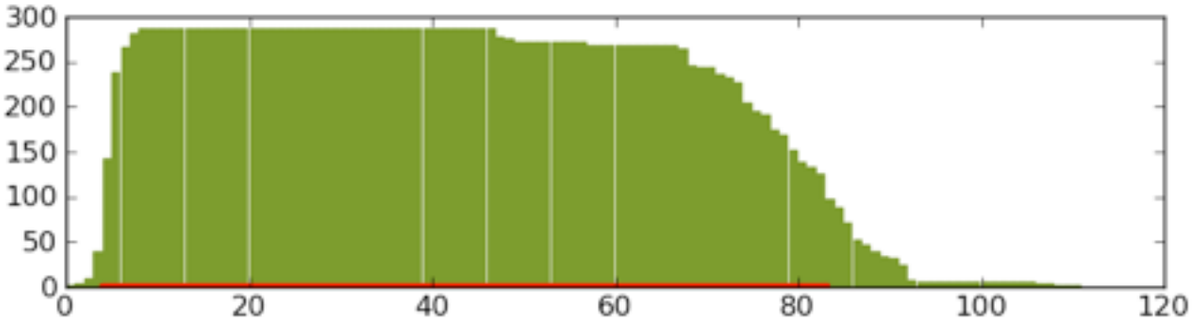
“Hills”

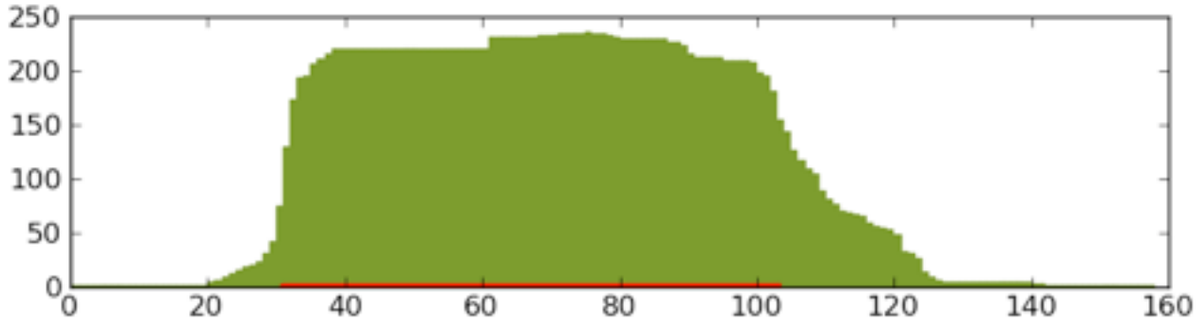

0 20 40 60 80 100 120 140
0 50 100 150 200 250 300 350 400

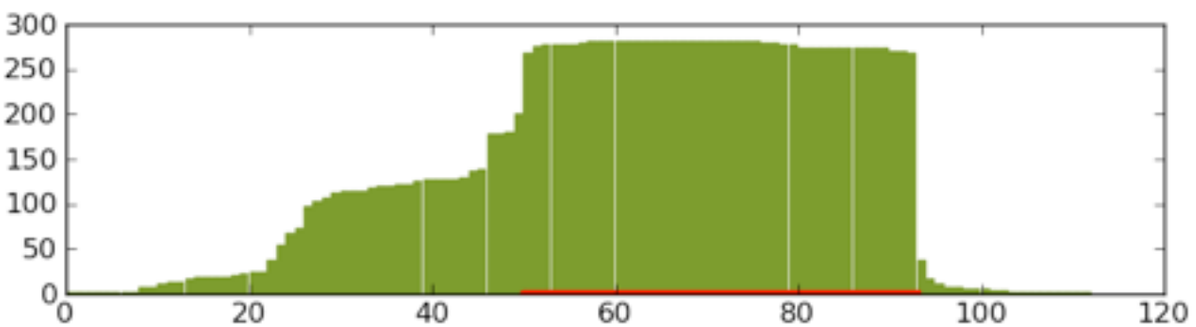

0 20 40 60 80 100 120 140 160
0 50 100 150 200 250 300

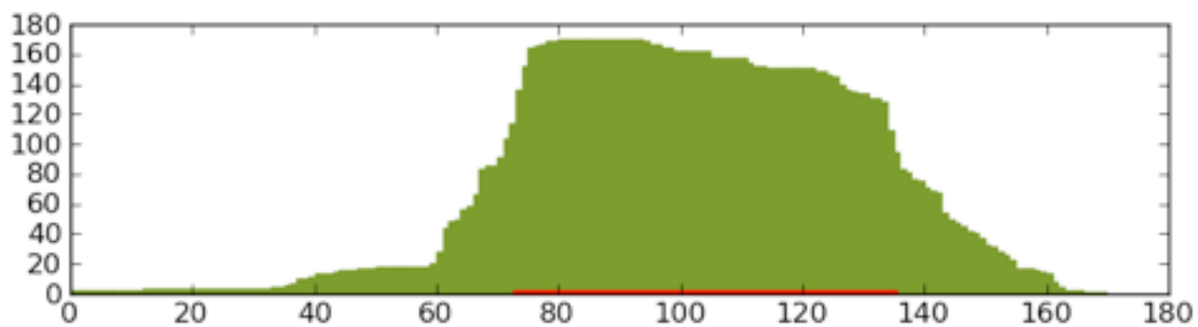

0 20 40 60 80 100 120 140 160 180
0 50 100 150 200 250 300 350

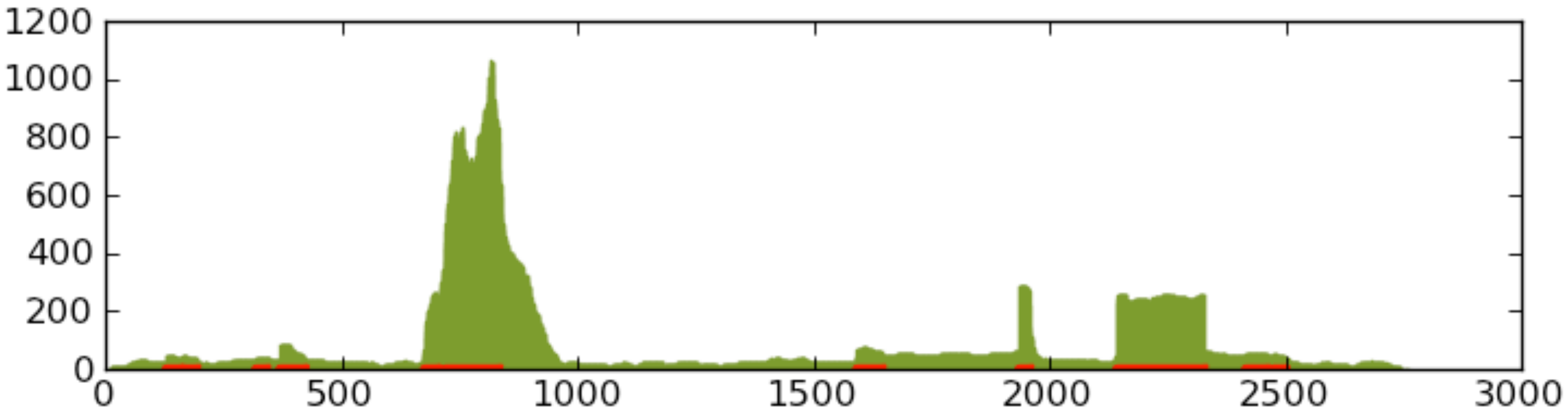

0 20 40 60 80 100
0 50 100 150 200 250


0 20 40 60 80 100 120
0 50 100 150 200 250 300

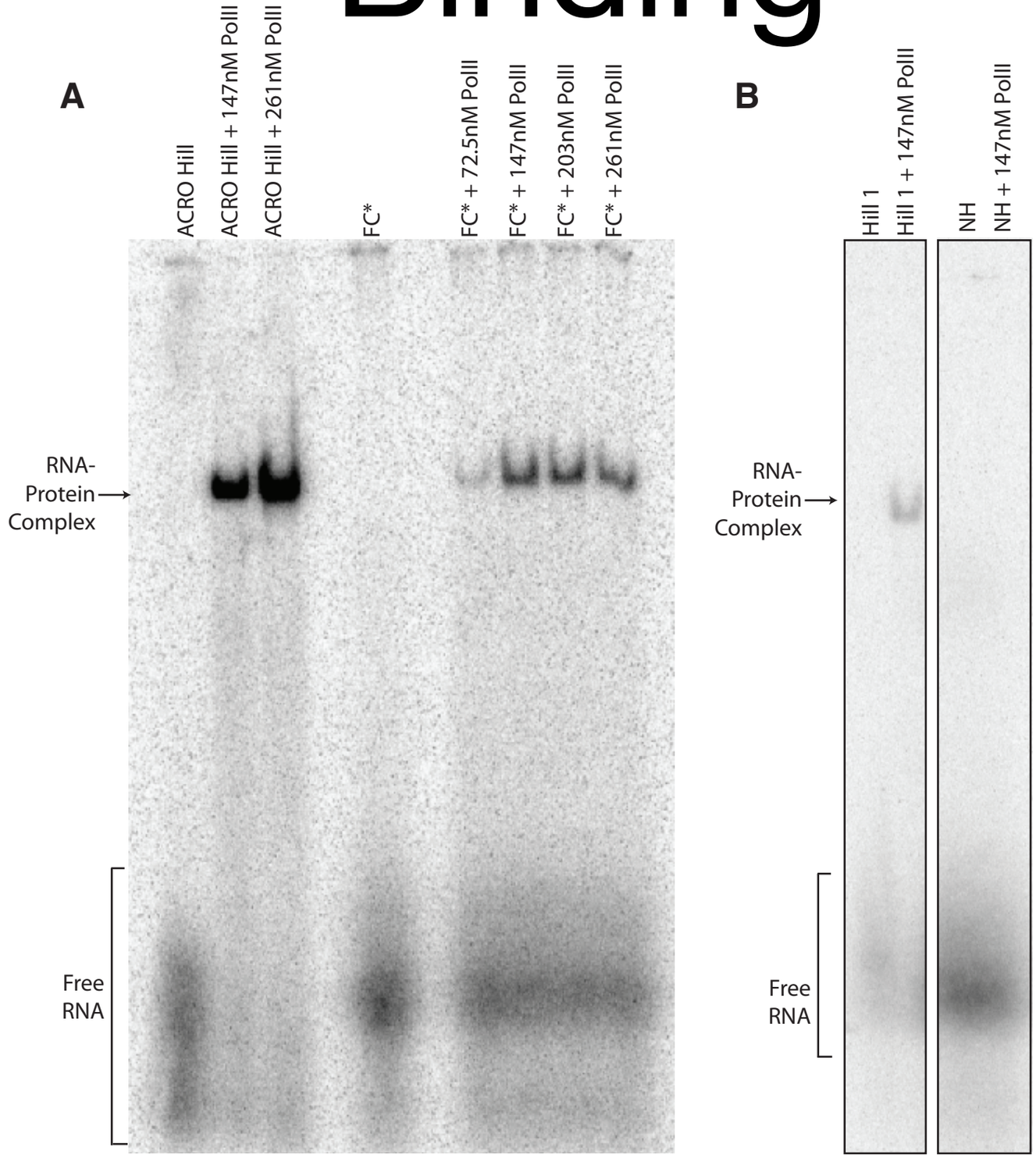

0 20 40 60 80 100 120 140 160
0 50 100 150 200 250


0 20 40 60 80 100 120
0 50 100 150 200 250 300


0 20 40 60 80 100 120 140 160 180
0 20 40 60 80 100 120 140 160 180


1200
1000
800
600
400
200
0
0
500
1000
1500
2000
2500
3000

Confirmation of Hill Binding



Other Uses of Hills

- Enrichment-aware motif detection
- Discovery of *in cis* binding domains
- Discovery of binding domains to separate proteins in complex

Future Perspectives: Computational

- Refinement of hill algorithm
  - Potential discovery of convoluted hills
  - Refined look at hill-dense regions
- Motif discovery for domains of RNA polymerase-binding RNAs
- Enrichment-aware structure prediction

Future Perspectives: Biochemical

- Activity assays on Polymerase binding genomic aptamers
  - Do our RNAs confer inhibition of Polymerase activity
- Structure determination through co-crystallization (collab. with Patrick Cramer)

Acknowledgements

Hfq/Neutral:
Christina Lorenz
Tanja Gesell
Ivana Bilusic
Ursula Schoeberl
Meghan Lybecker

PBE:
Frederike von Pelchrzim
Jennifer L Boots
Katarzyna Matylla-Kulinska
Marek Zywicki
Doris Chen
Adam Weiss

The Rest:
Oliver Mayer
Martina Dötsch
Boris Fürtig
Johanna Bisich

EYBE:
Frederike von Pelchrzim
Nadia Tukhtubaeva
Ece Ergir

Committee: Eric Westhof, Ivo Hofacker

Acknowledgements

Acknowledgements

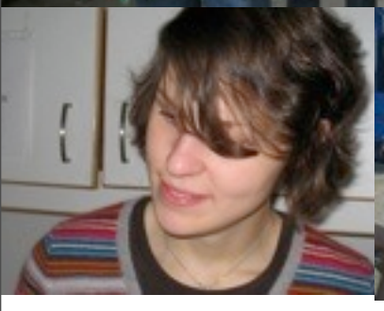
Renée!

Acknowledgements

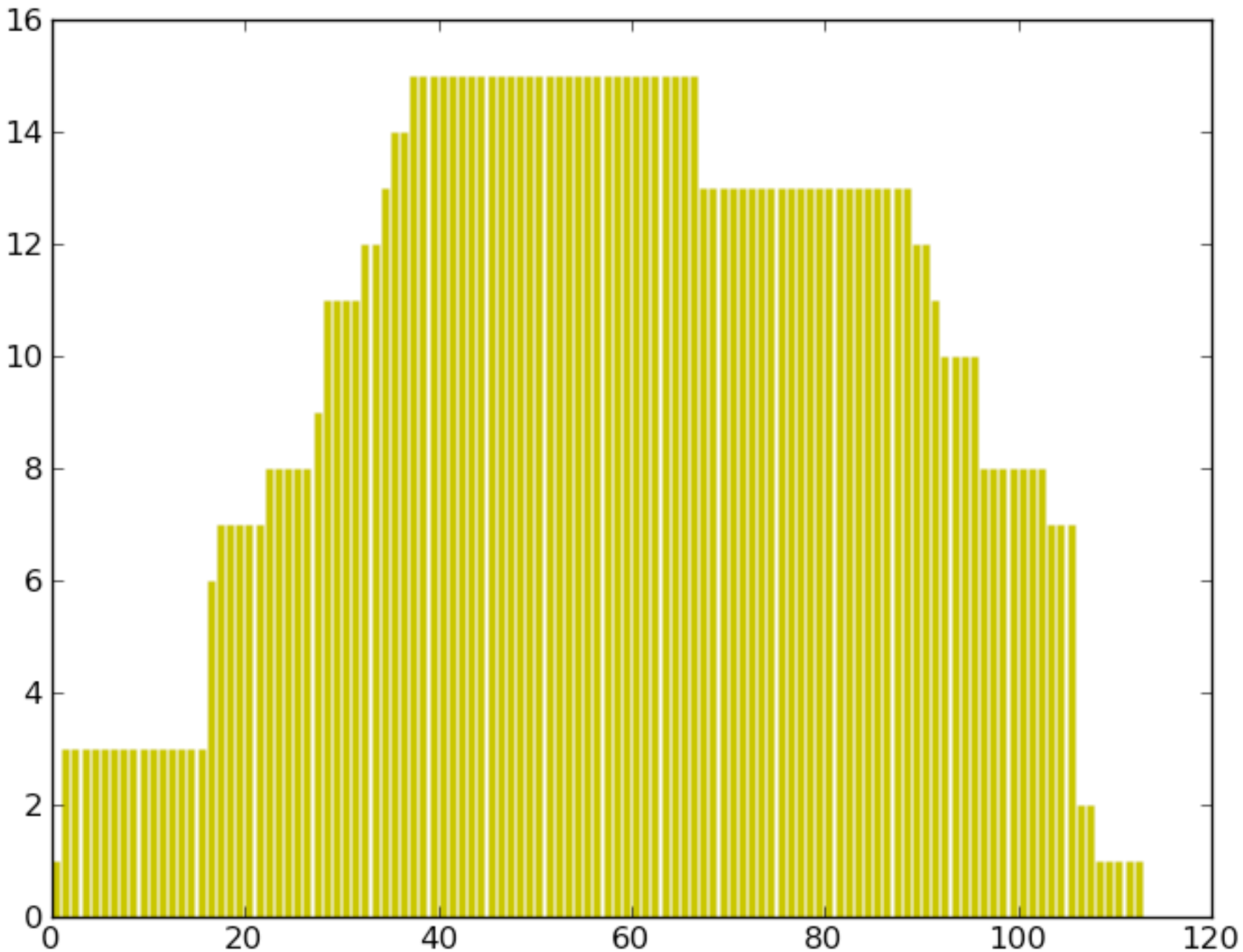




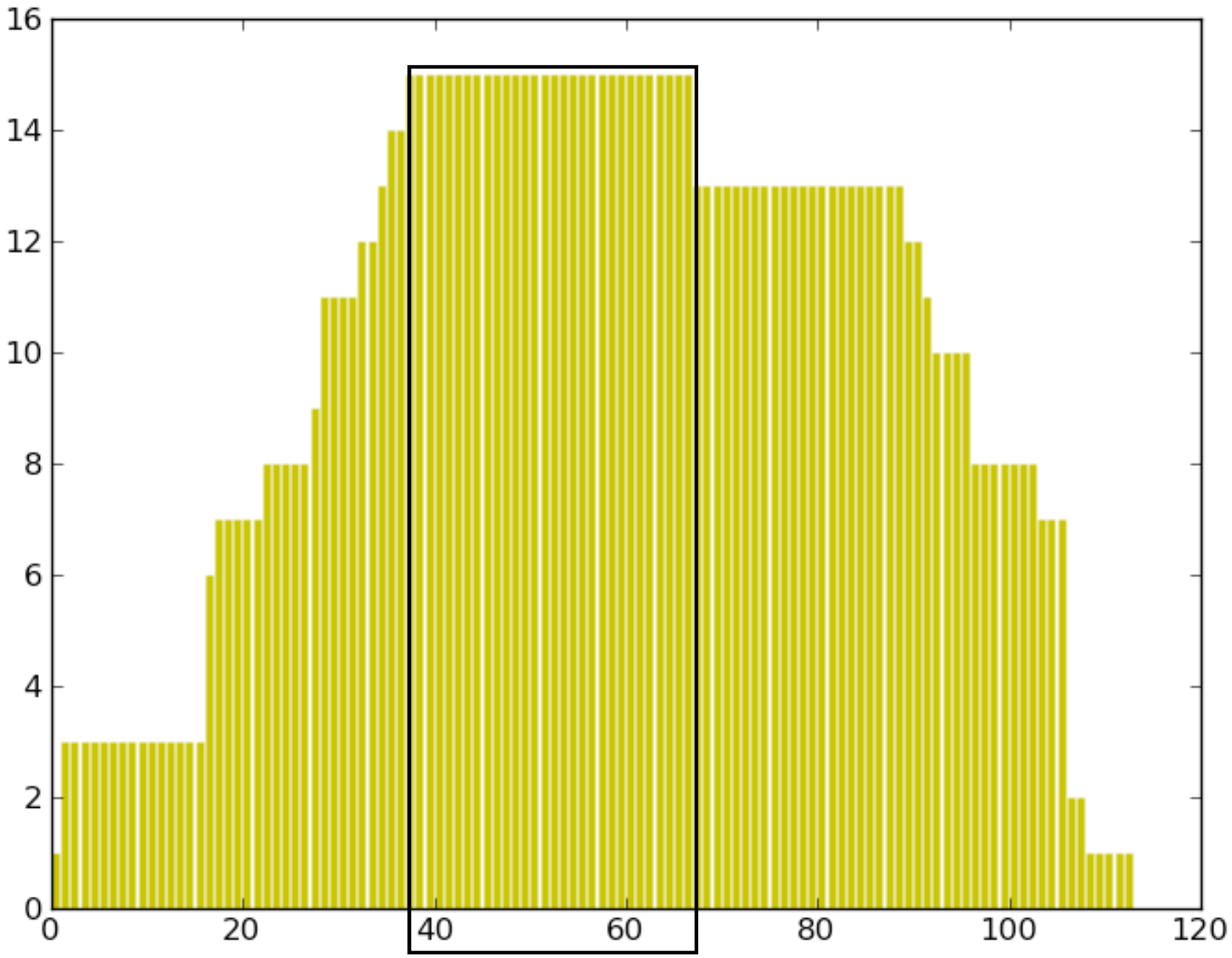


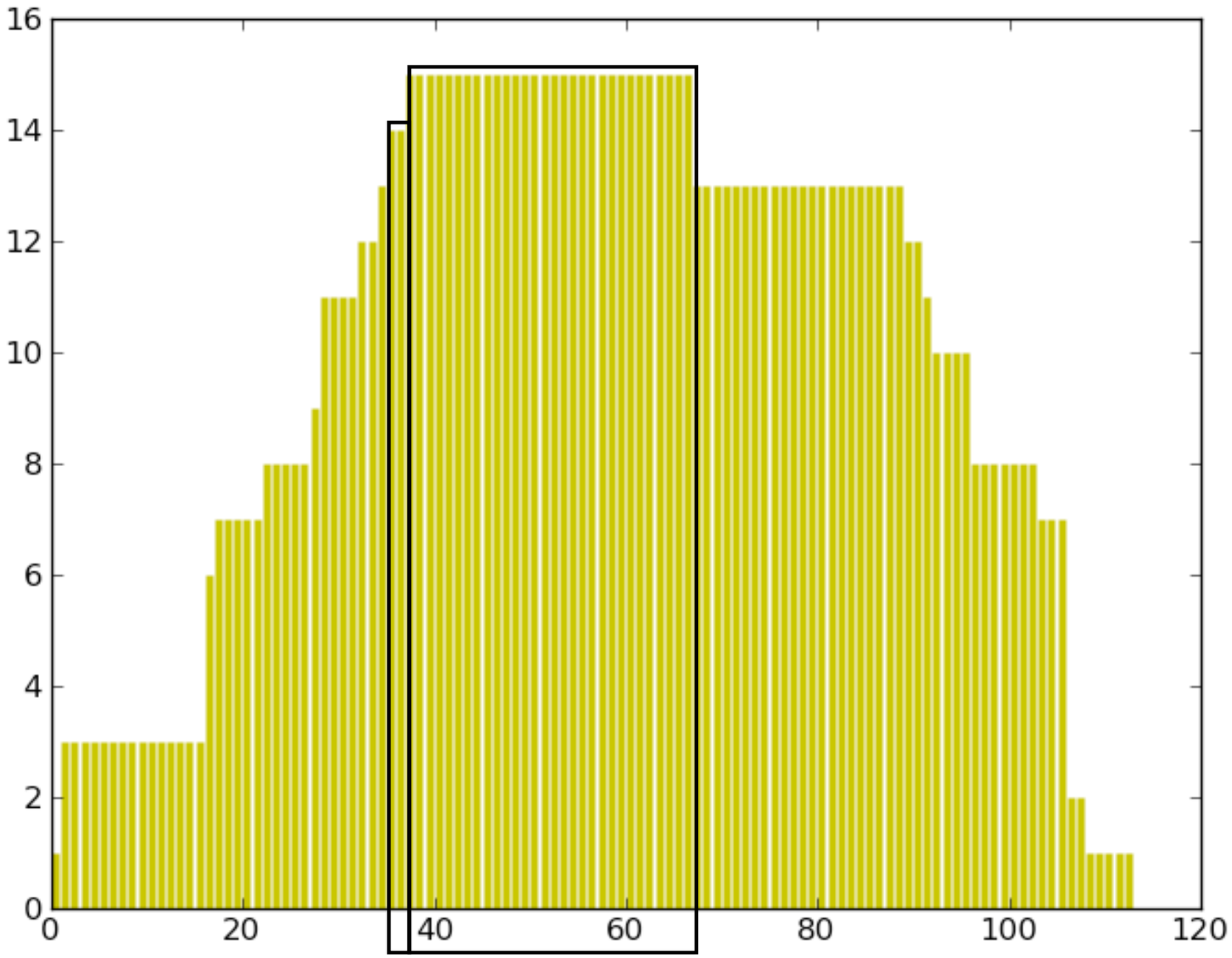




16
14
12
10
8
6
4
2
0
0
20
40
60
80
100
120

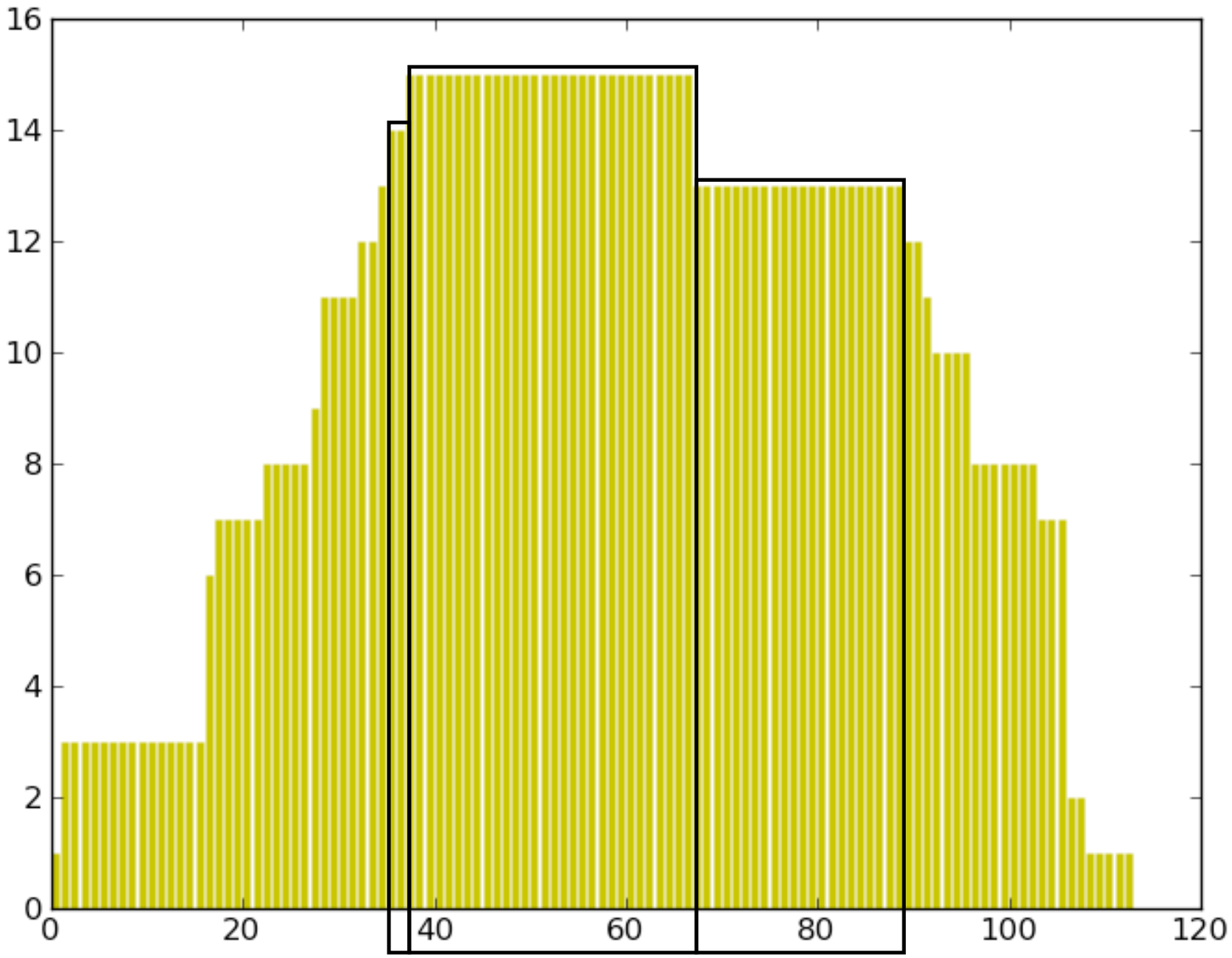
Mass = 39%

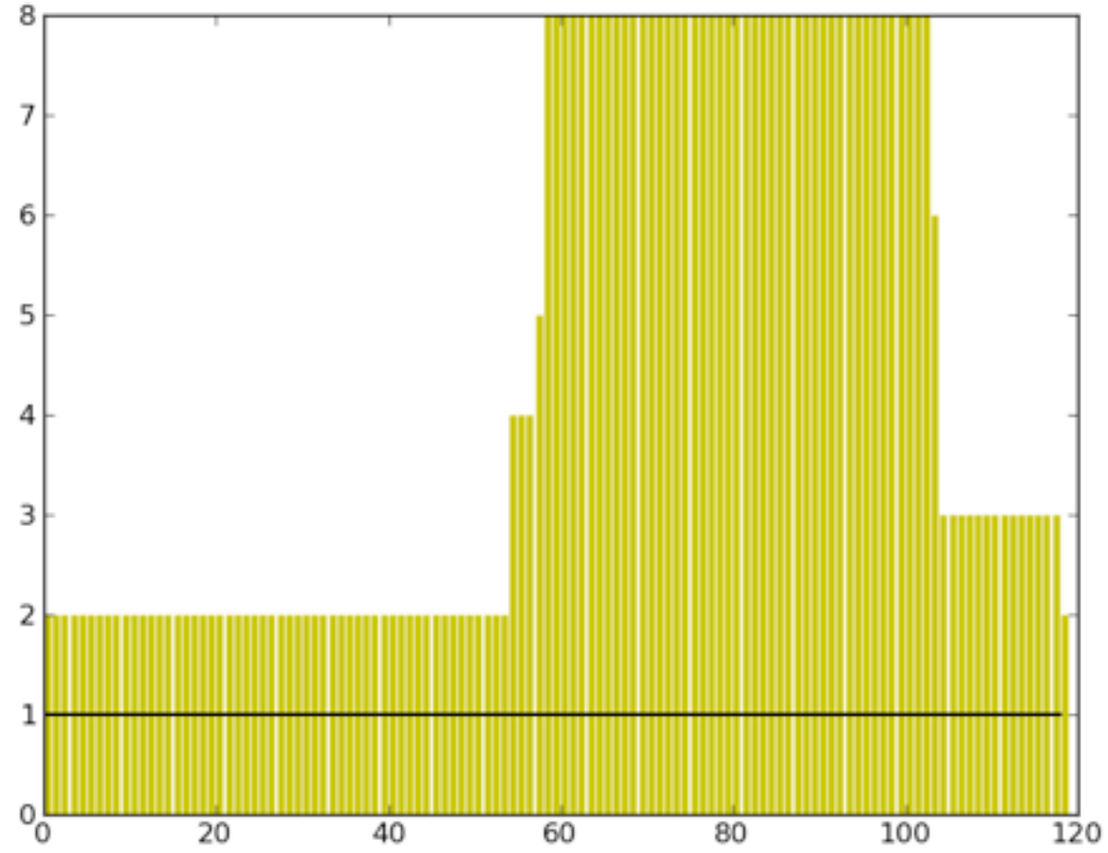


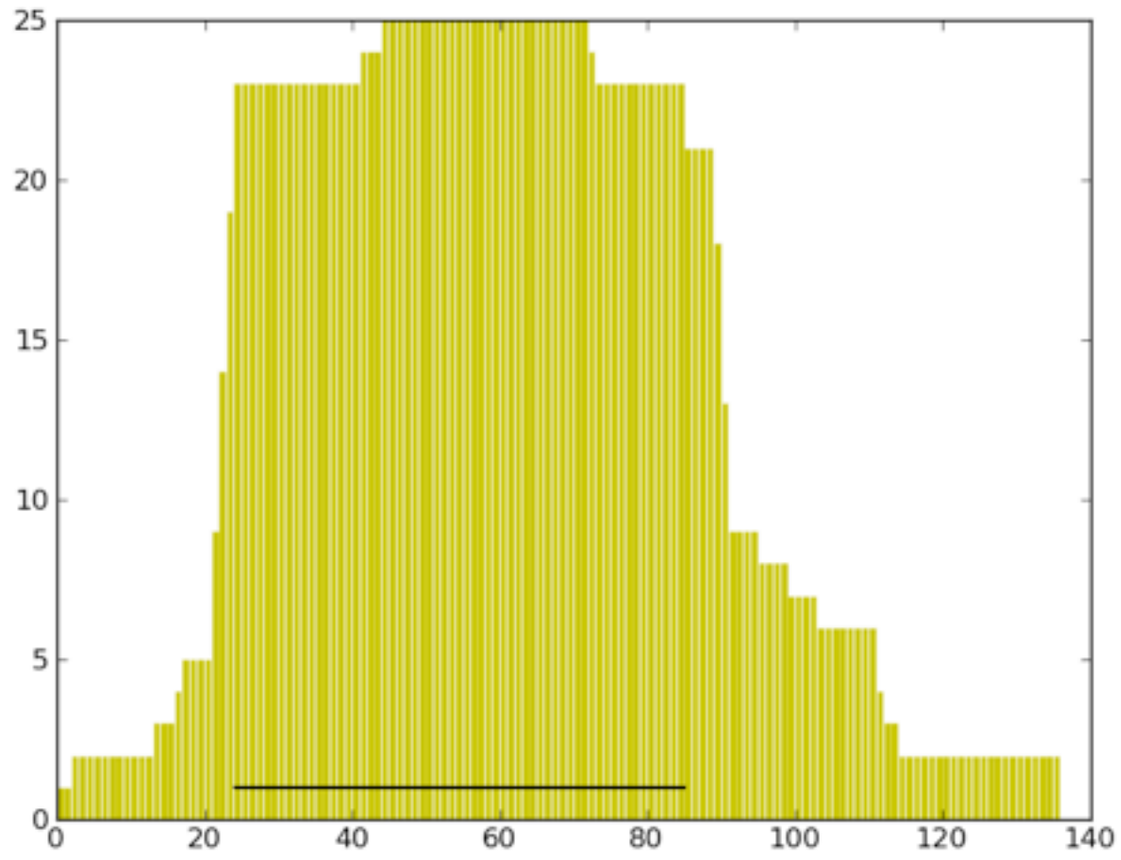
Mass = 42%

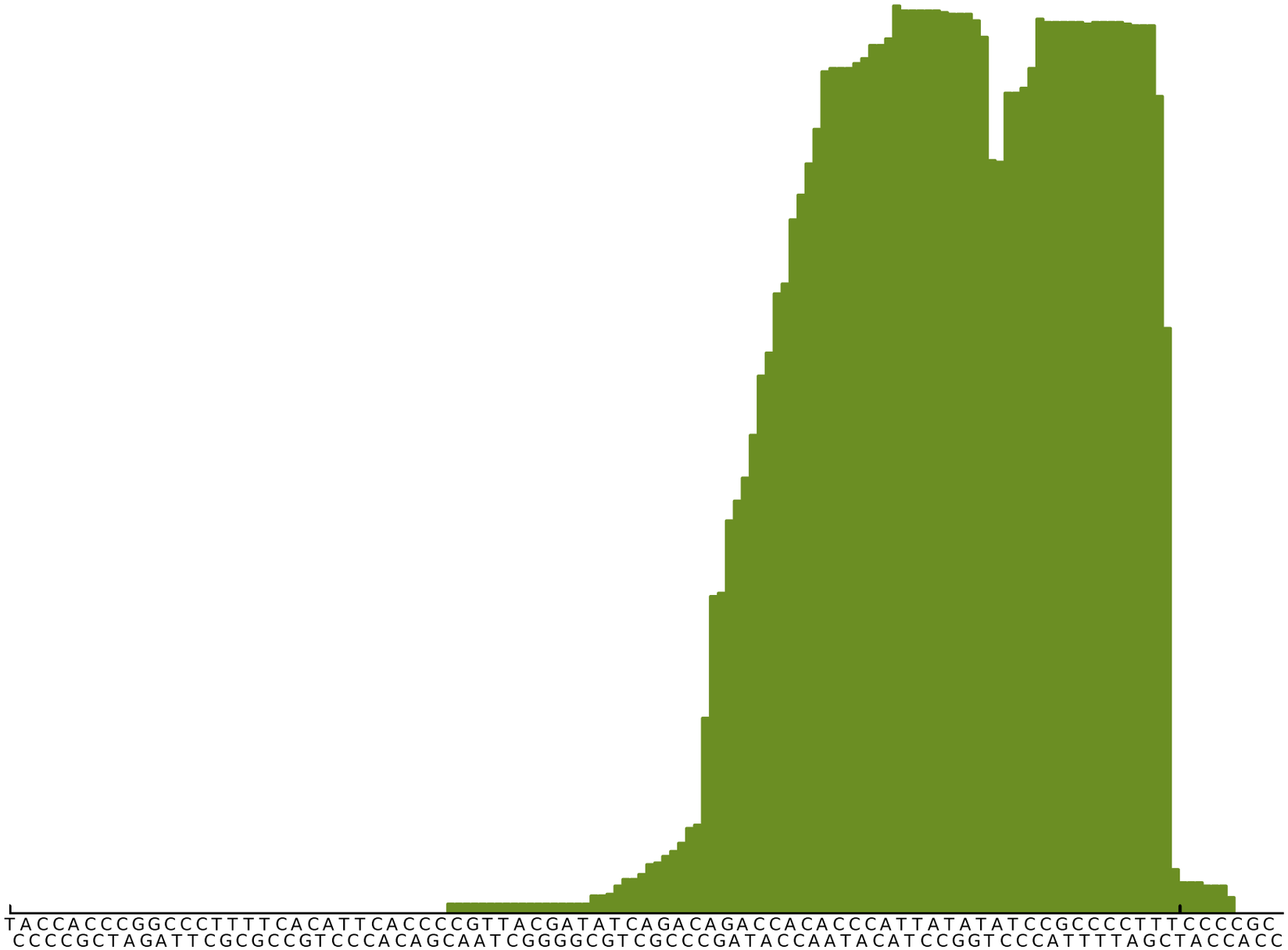


Mass = 67%




0
1
2
3
4
5
6
7
8
0
20
40
60
80
100
120


0
5
10
15
20
25
0
20
40
60
80
100
120
140


TACCACCCGGCCCTTTTCACATTCACCCCGTTACGATATCAGACAGACCACACCCATTATATATCCGCCCCTTTCCCCGC
CCCCGCTAGATTCGCGCCGTCCCACAGCAATCGGGGCGTCGCCCGATACCAATACATCCGGTCCCATTTTAGCTACCACC

libybe

3365450

284421

283687

247992

84949

71786

63990

63023

59226

58237

54277

43893

37833

37650

36527

35332

27179

26565

25001

24359

23416

22687

22163

1132936

ybe

2263135

79423

74805

53338

31313

29338

26521

25002

21934

21668

1067326

preebe

1698006

39132

33935

32761

26898

25703

21485

21006

1005832

ebe

1545422

40437

35427

32627

31289

30815

23844

22609

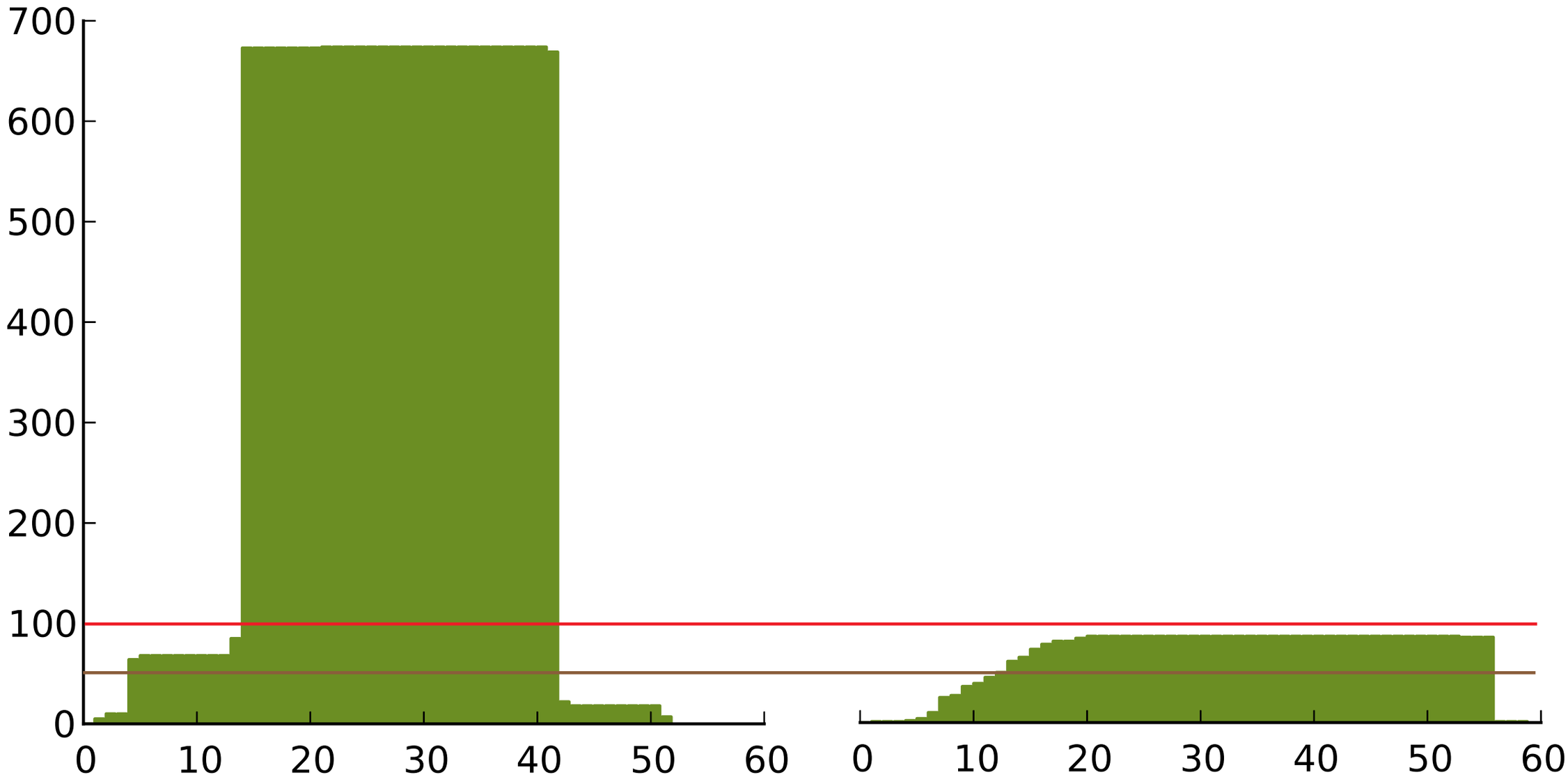
20841

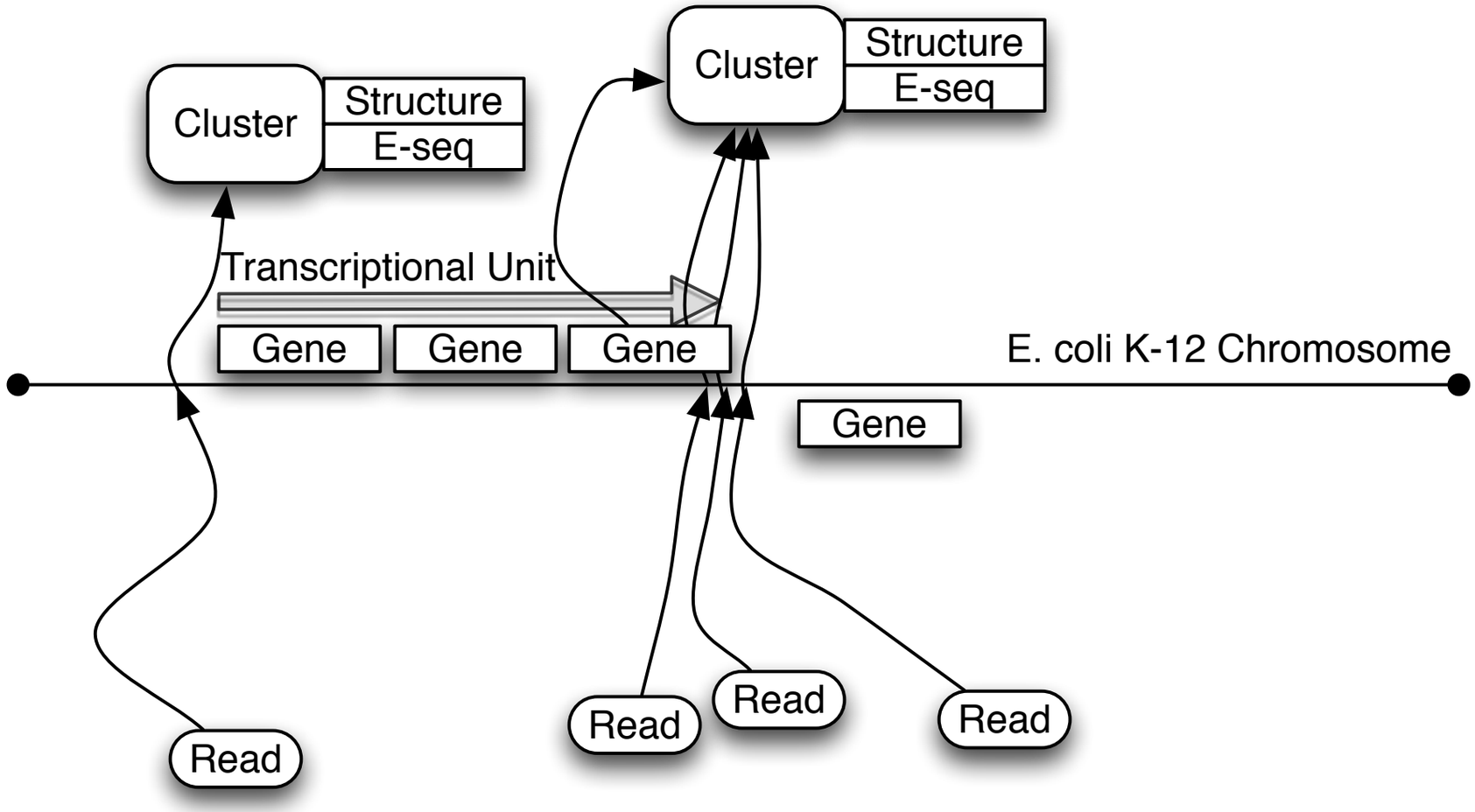
781974

libebe

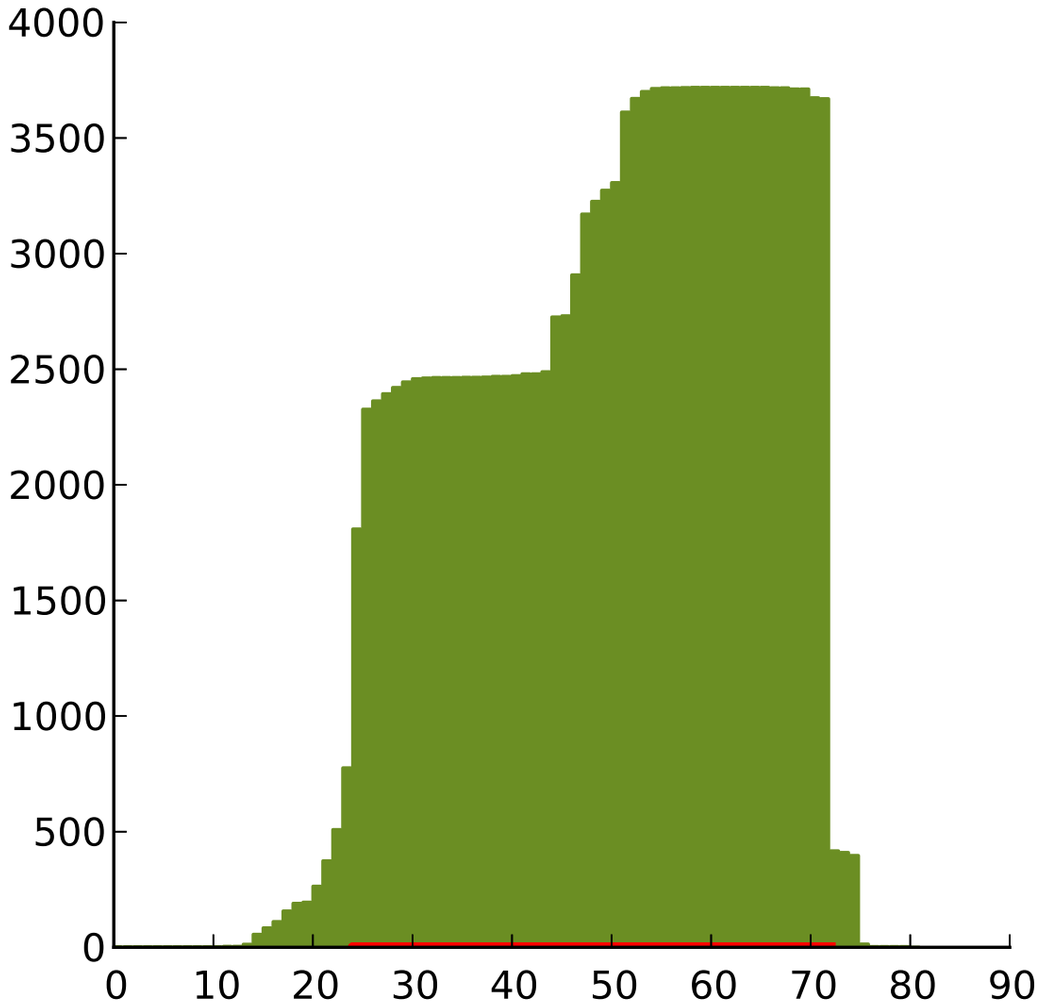
398654

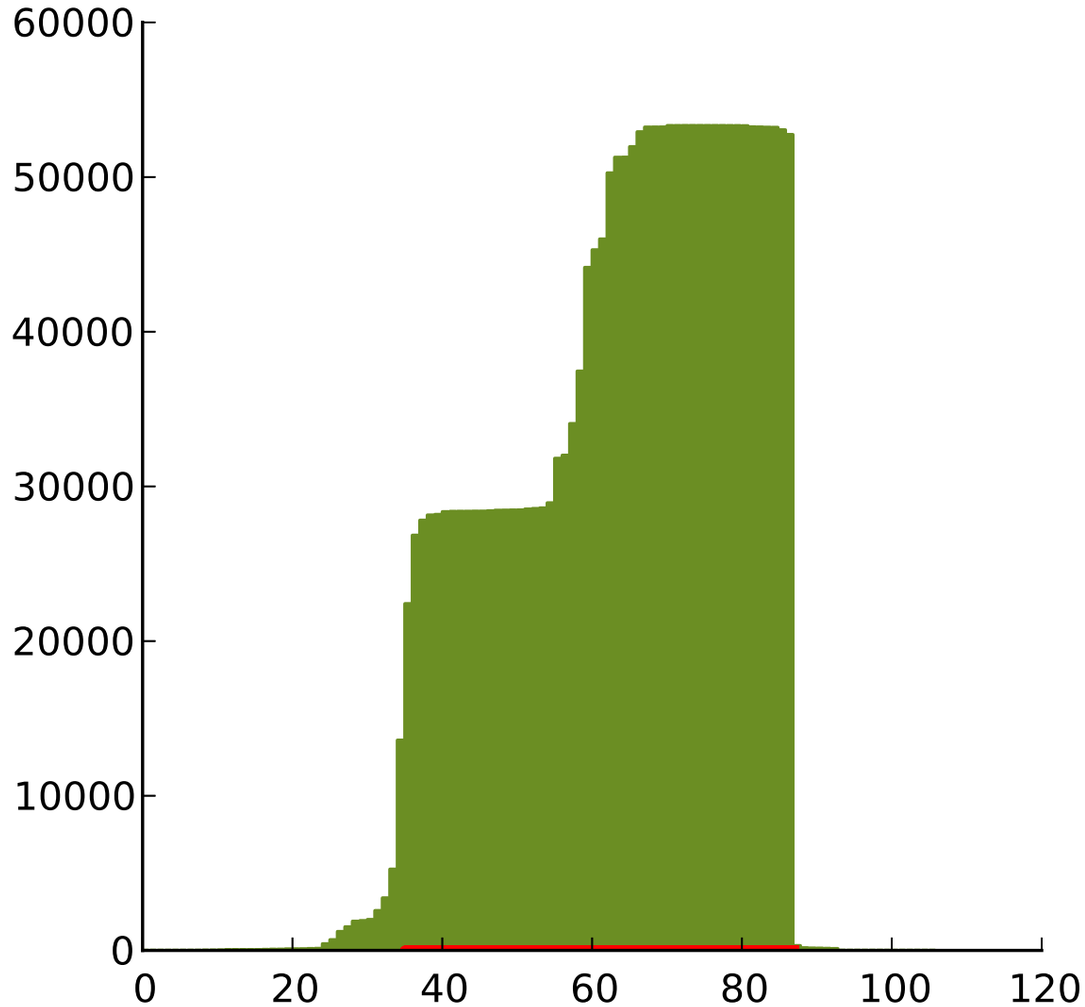
395863


700
600
500
400
300
200
100
0
0
10
20
30
40
50
60
0
10
20
30
40
50
60

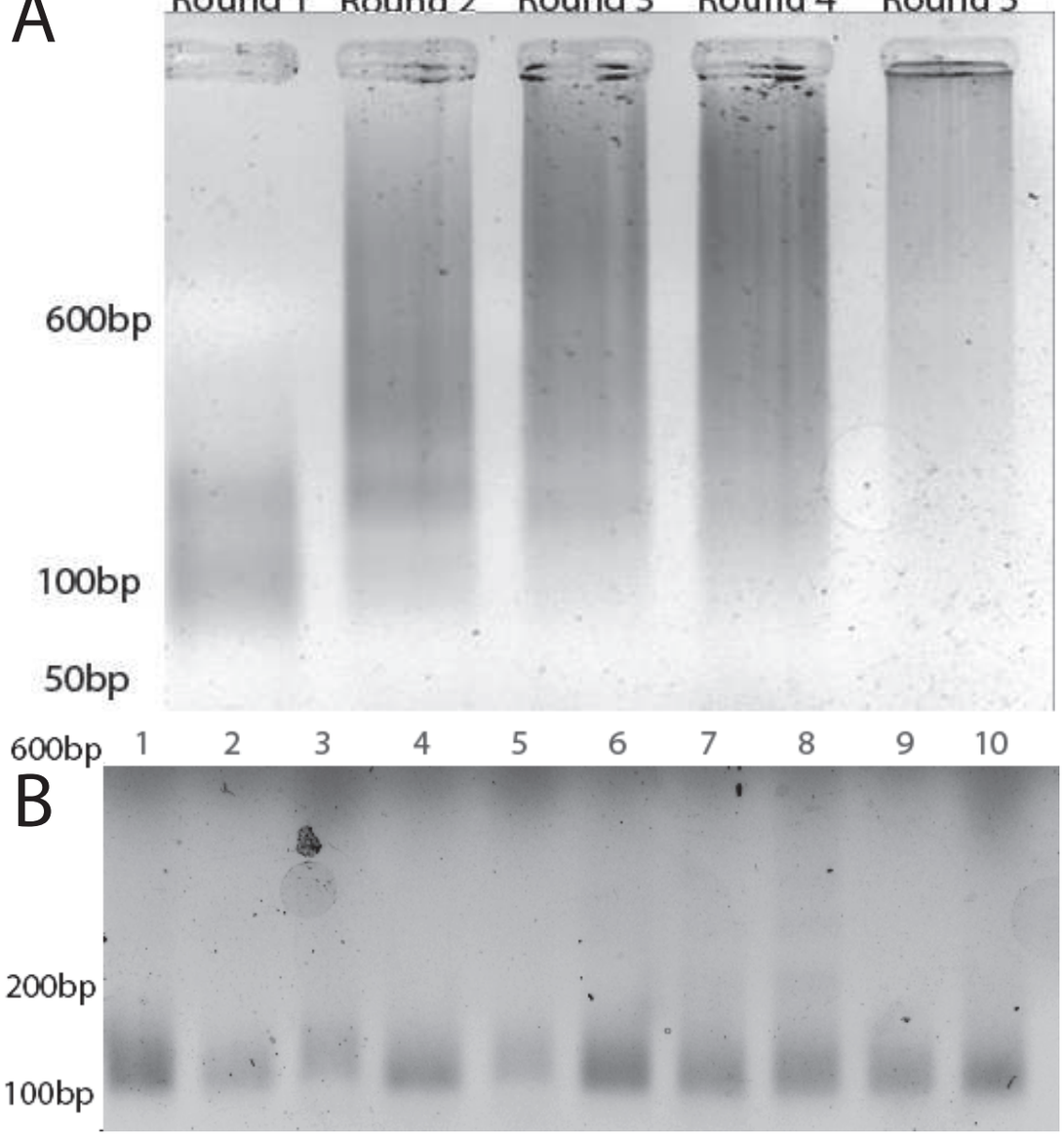

Cluster
Structure
E-seq
Cluster
Structure
E-seq
Transcriptional Unit
Gene
Gene
Gene
E. coli K-12 Chromosome
Gene
Read
Read
Read
Read


2
bits
1
0
5′
1
2
3
4
5
6
7
3′
weblogo.berkeley.edu


4000
3500
3000
2500
2000
1500
1000
500
0
0
10
20
30
40
50
60
70
80
90

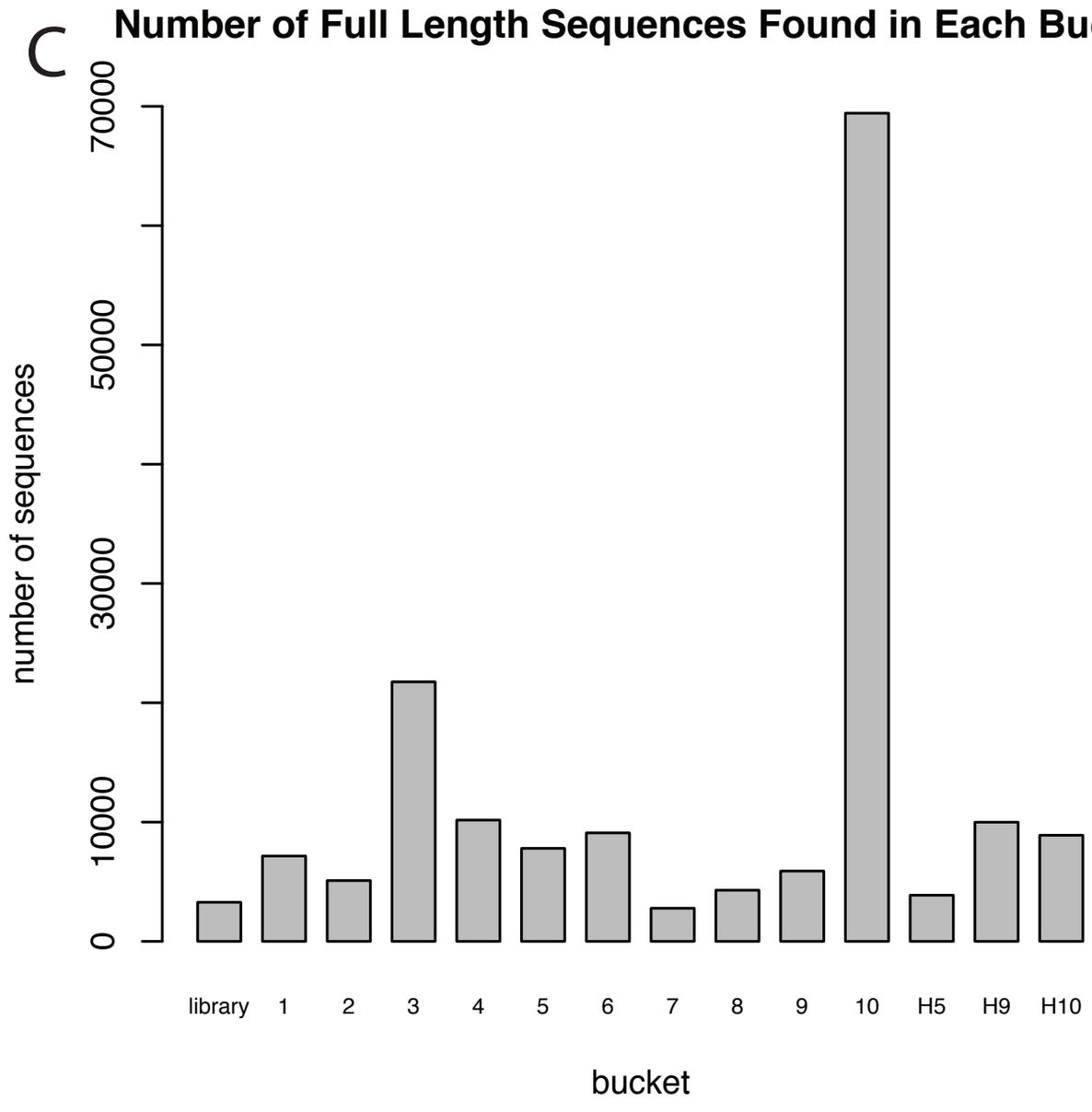

60000
50000
40000
30000
20000
10000
0
0
20
40
60
80
100
120

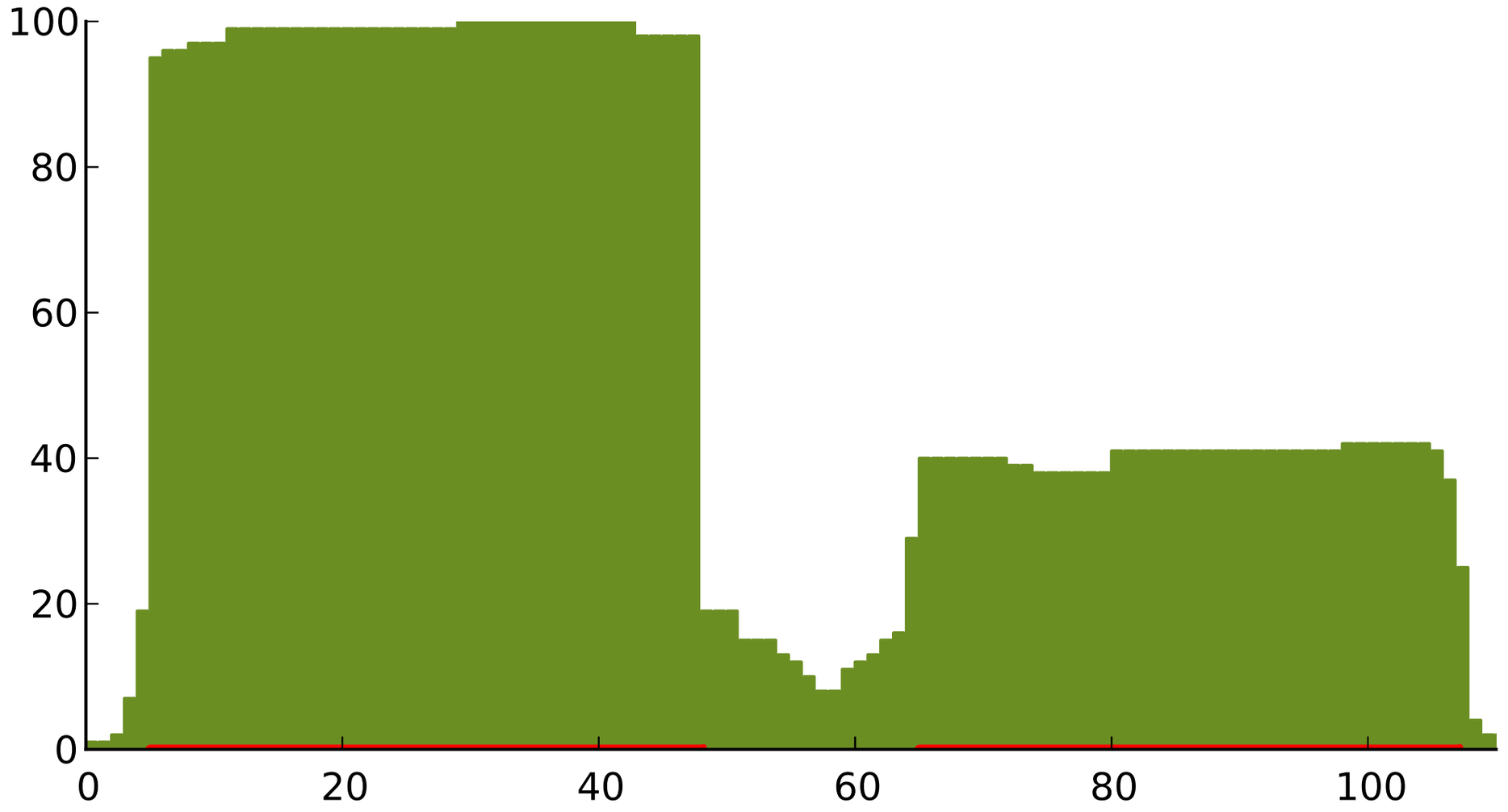
A


Round 1
Round 2
Round 3
Round 4
Round 5
600bp
100bp
50bp

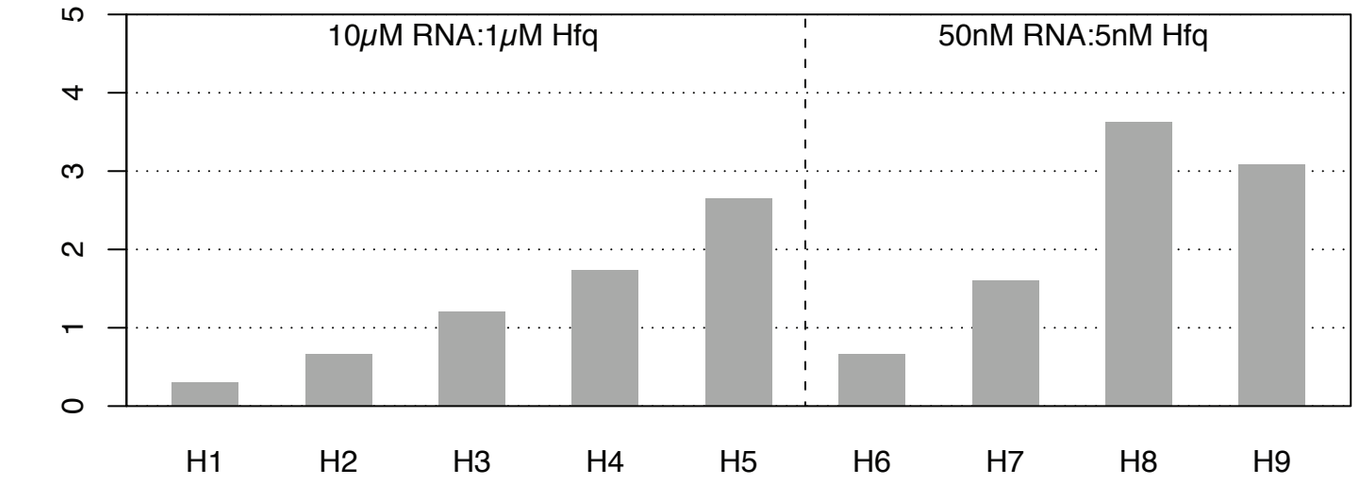
600bp
1
2
3
4
5
6
7
8
9
10
B
200bp
100bp

C

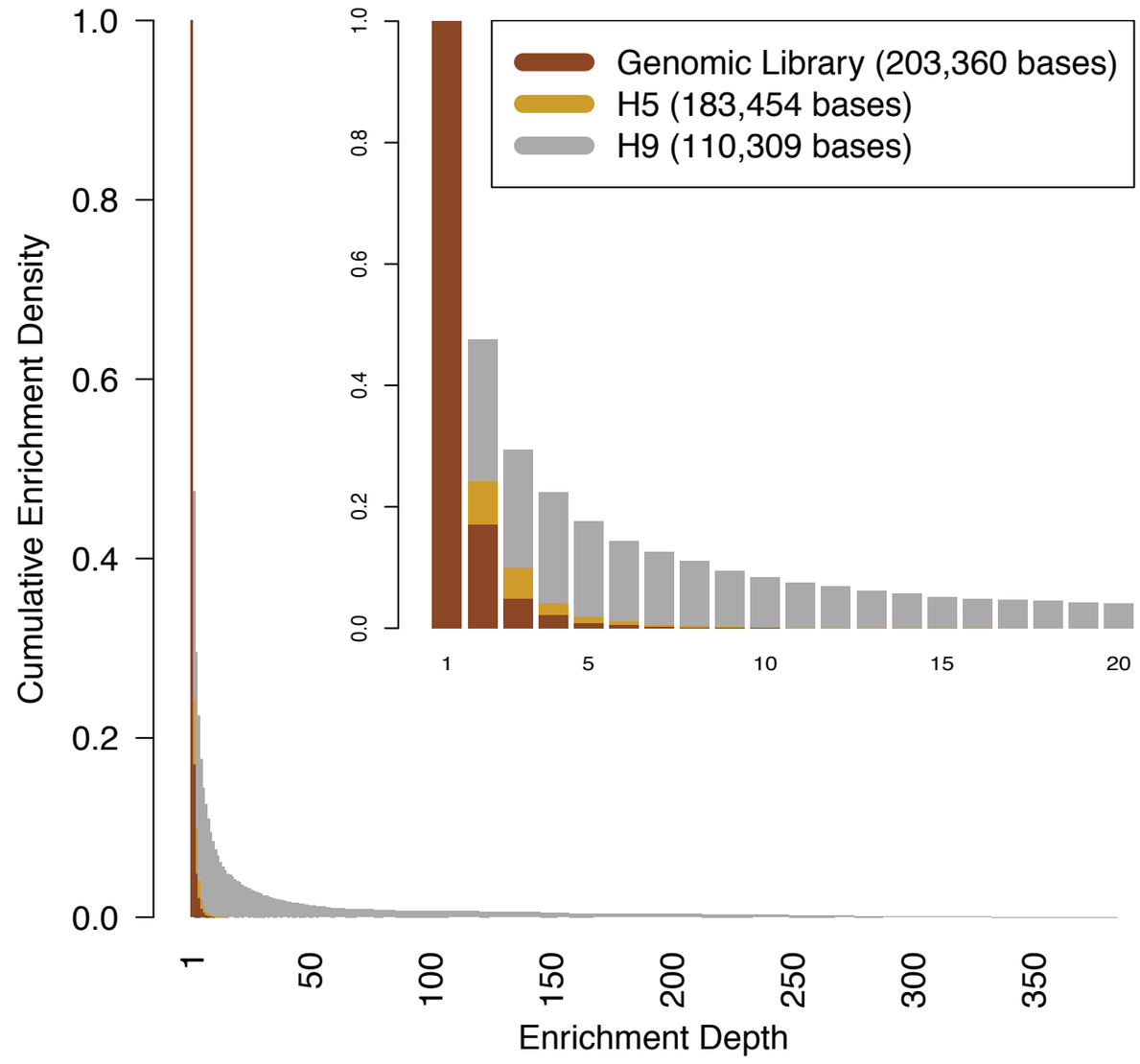

Number of Full Length Sequences Found in Each Bucket
number of sequences
0
10000
30000
50000
70000
library
1
2
3
4
5
6
7
8
9
10
H5
H9
H10
bucket

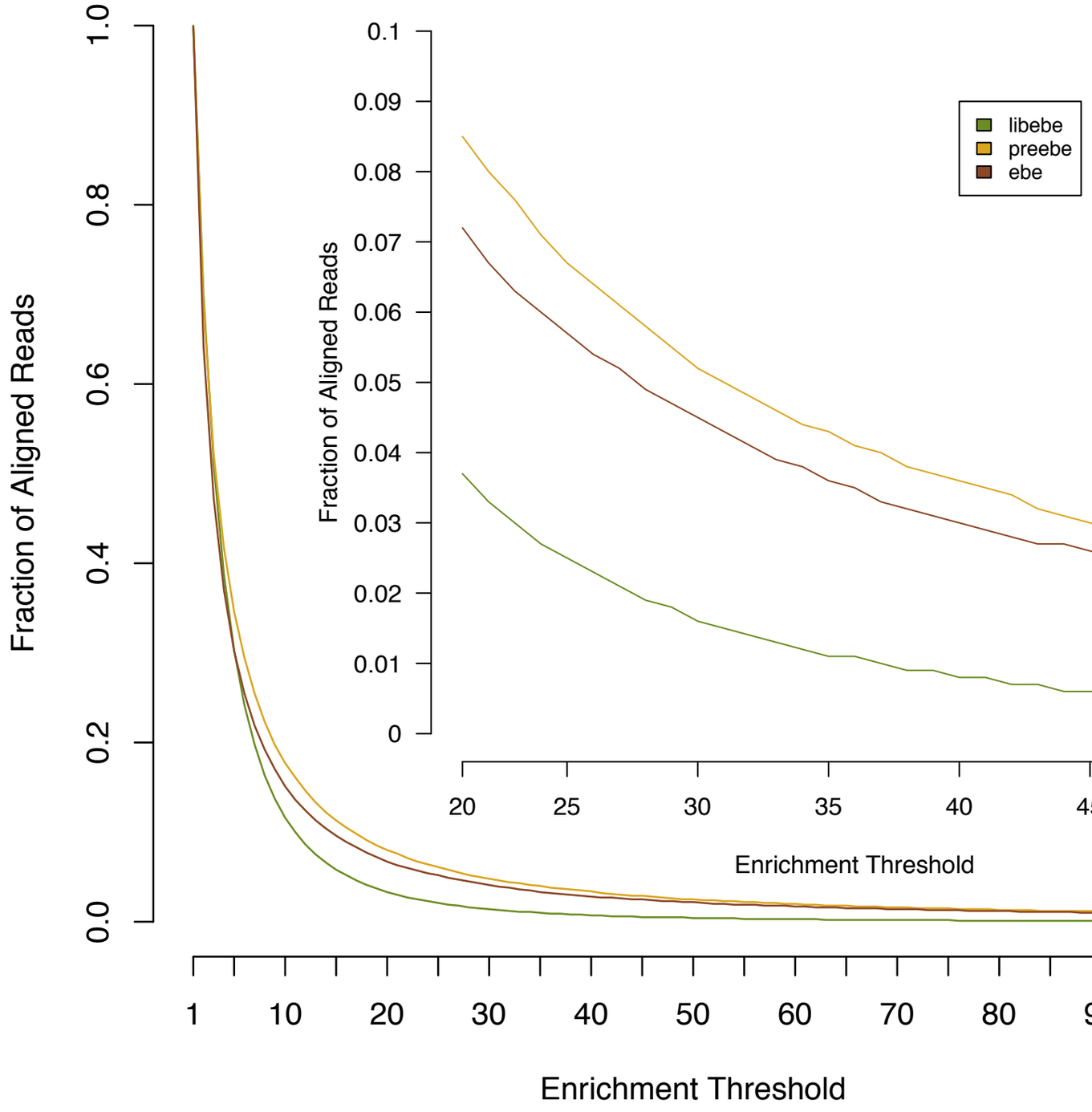

100
80
60
40
20
0
0
20
40
60
80
100

A

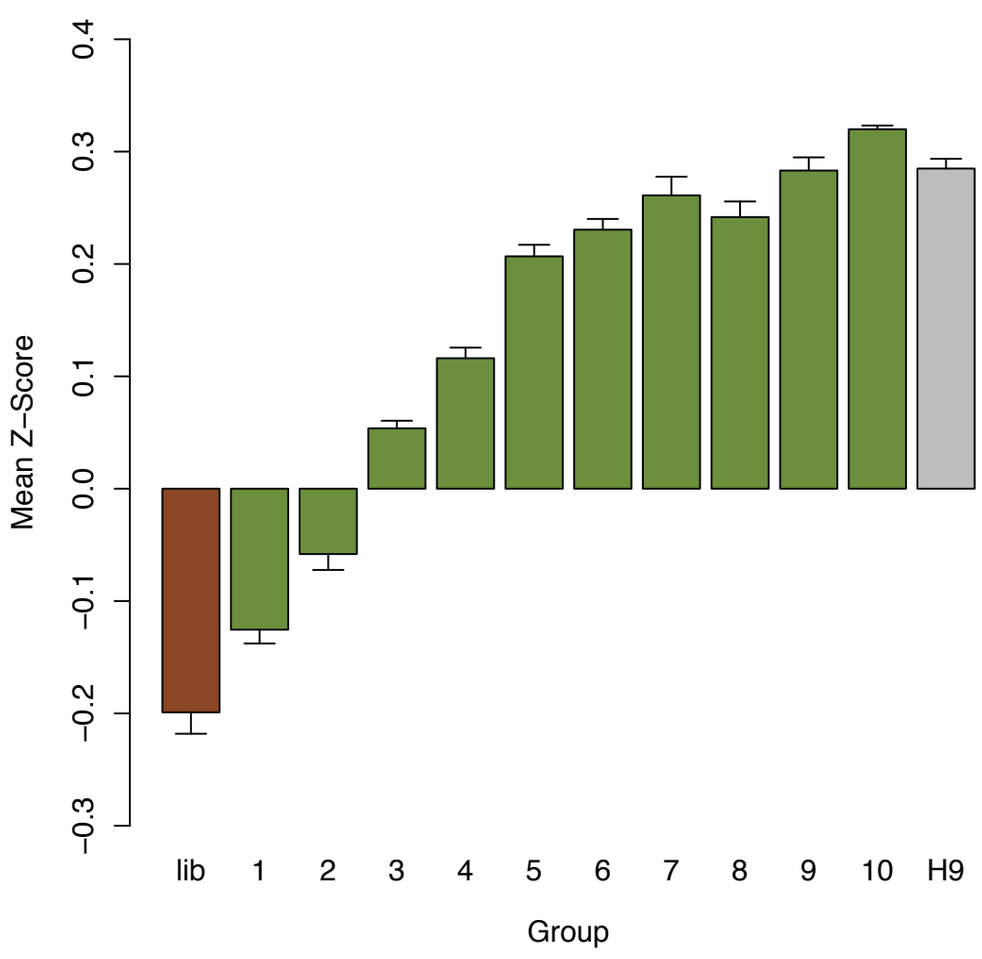

RNA Recovery in % After Each Cycle of SELEX
10μM RNA:1μM Hfq
50nM RNA:5nM Hfq
0
1
2
3
4
5
H1
H2
H3
H4
H5
H6
H7
H8
H9

B

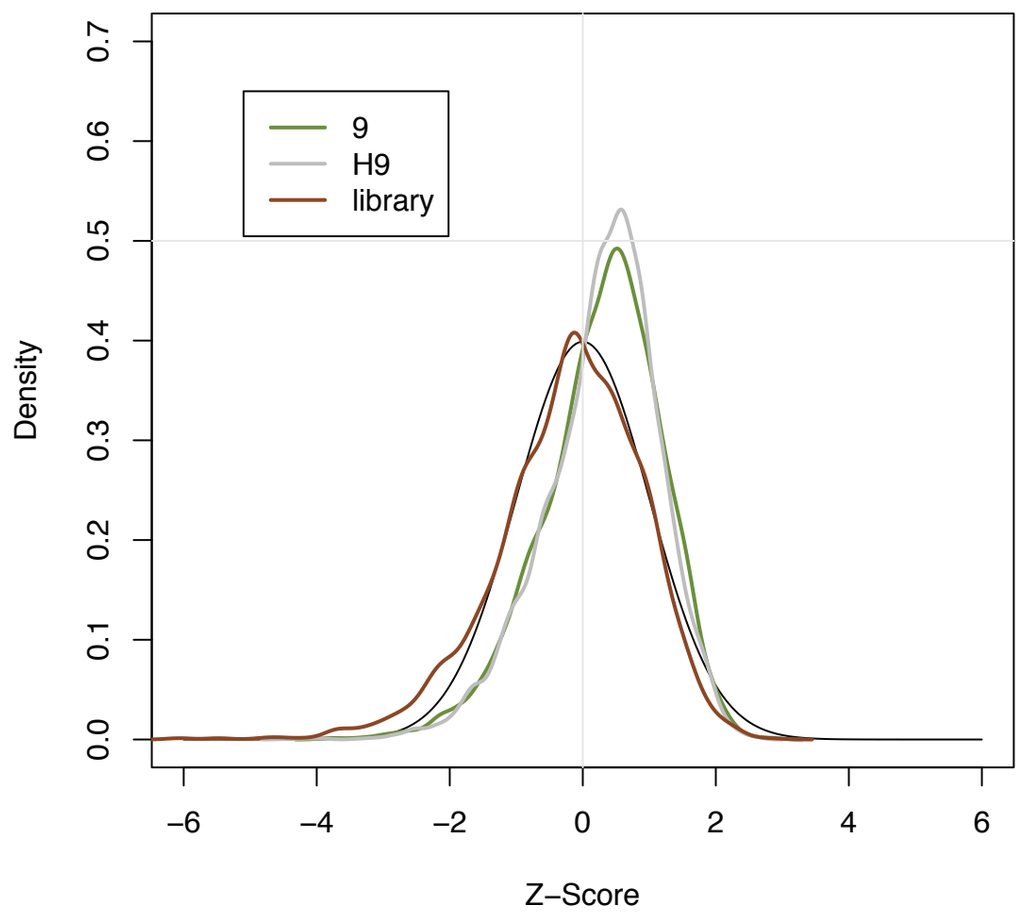

Cumulative Base–Resolution Enrichment as a Fraction of the Pool Size
Genomic Library (203,360 bases)
H5 (183,454 bases)
H9 (110,309 bases)
Cumulative Enrichment Density
Enrichment Depth
0.0
0.2
0.4
0.6
0.8
1.0
1
50
100
150
200
250
300
350
1
5
10
15
20


Fraction of Aligned Reads
0.0
0.2
0.4
0.6
0.8
1.0
1
10
20
30
40
50
60
70
80
90
Enrichment Threshold
Fraction of Aligned Reads
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
20
25
30
35
40
45
50
Enrichment Threshold
libebe
preebe
ebe

A


Mean Z−scores
Mean Z−Score
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
lib
1
2
3
4
5
6
7
8
9
10
H9
Group

B


Z−score Distributions in Neutral and Genomic SELEX
9
H9
library
Density
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−6
−4
−2
0
2
4
6
Z−Score


Rai
Rai
RAI
RAI


DARČEK ZDARMA

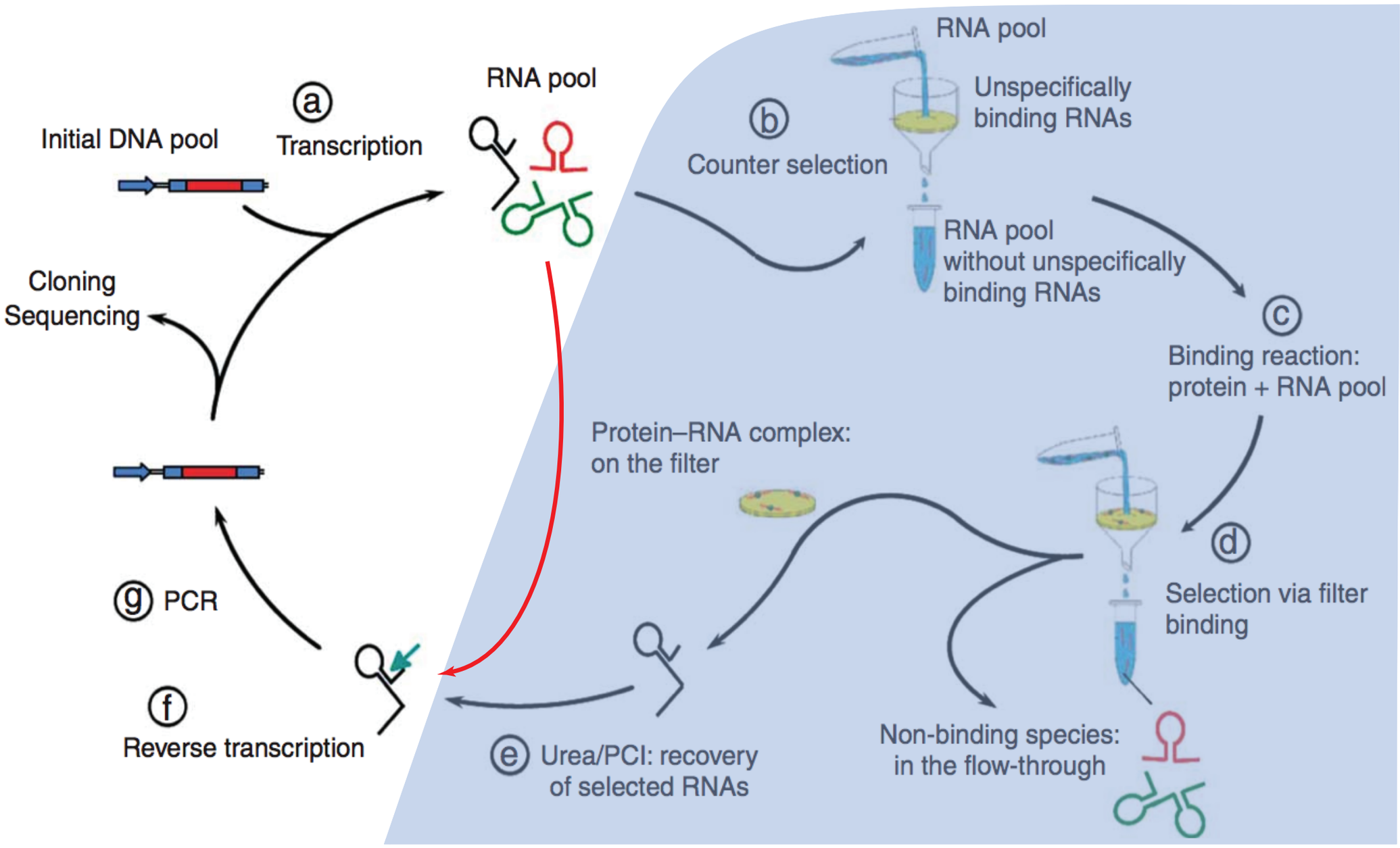

Moser
RAW3301
RAW3301
RAW3301
SONY
SIGMA
EOS


RAW 3301
RAW 3301
RAW 3301


PRI NÁKUPE NAD 3.500,- Sk
DARČEK ZDARMA
Hello Kitty

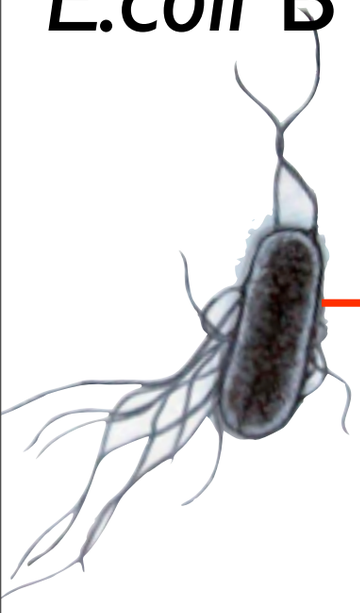

3
UV400
CE

Neutral SELEX



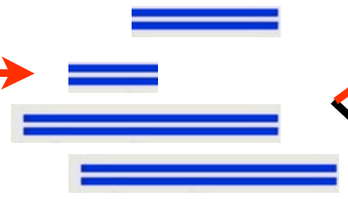
Experimental Setup

E.coli B

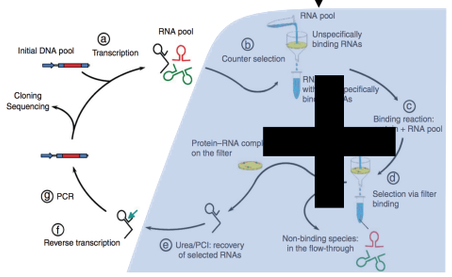


Genomic DNA

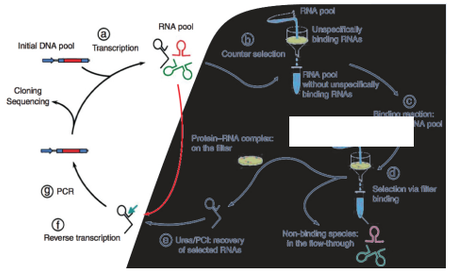
Genomic Library
(cDNA Fragments Representing the Genome)



Genomic SELEX
(Purified *E.coli* Hfq)

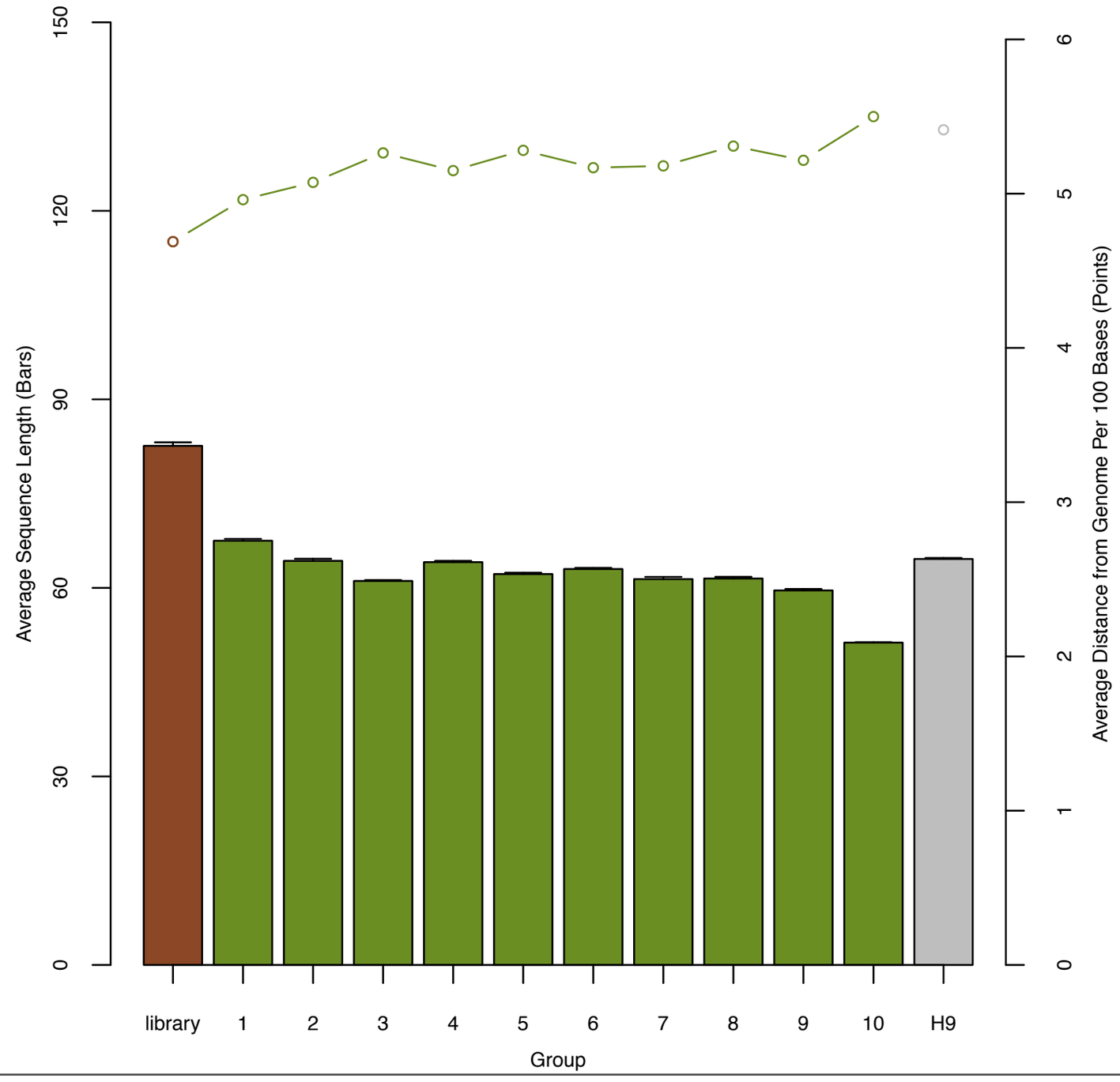


Neutral SELEX
(No protein selection)

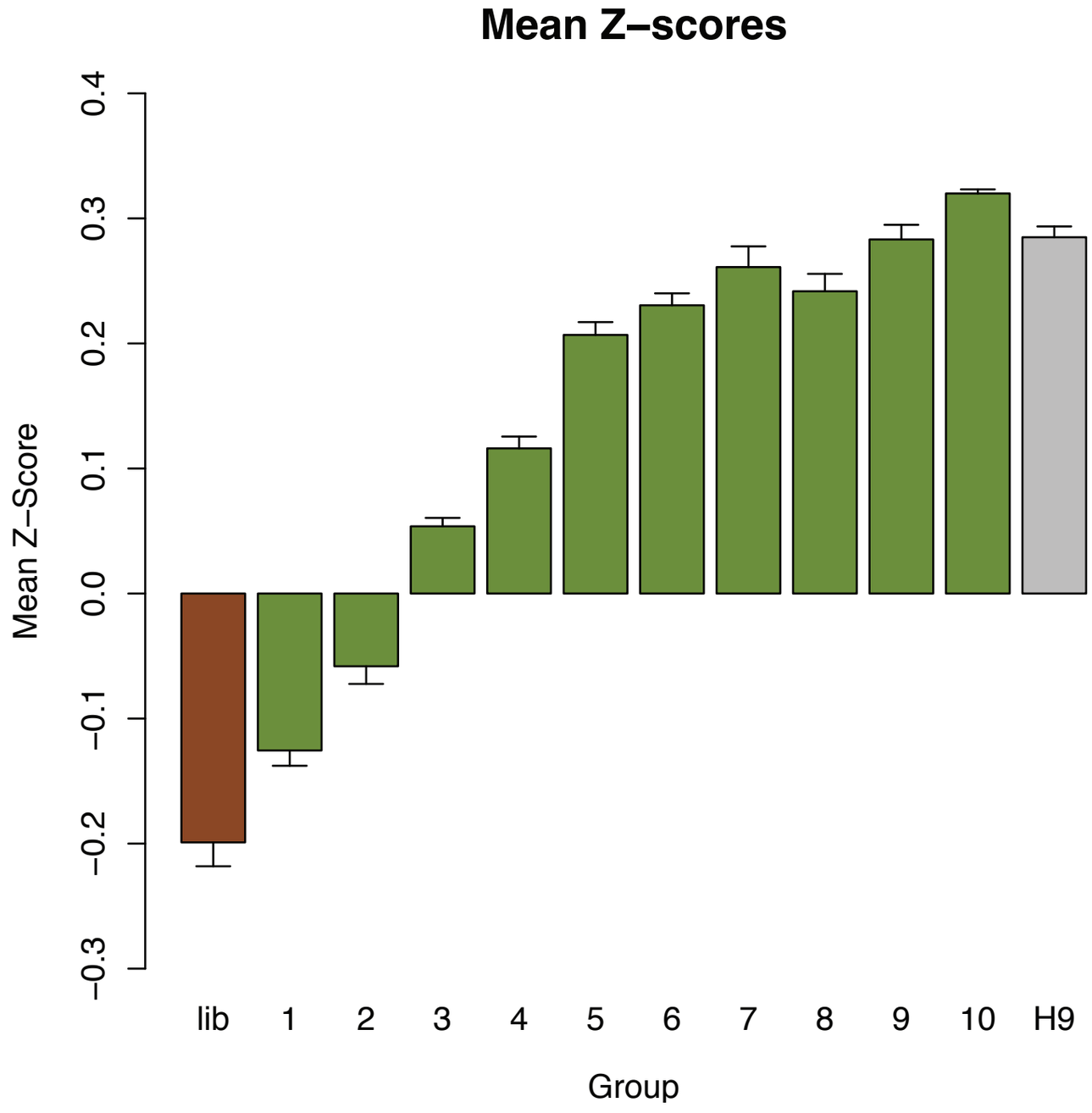


454 SEQUENCING

Lengths and Sequence Not Strongly Affected

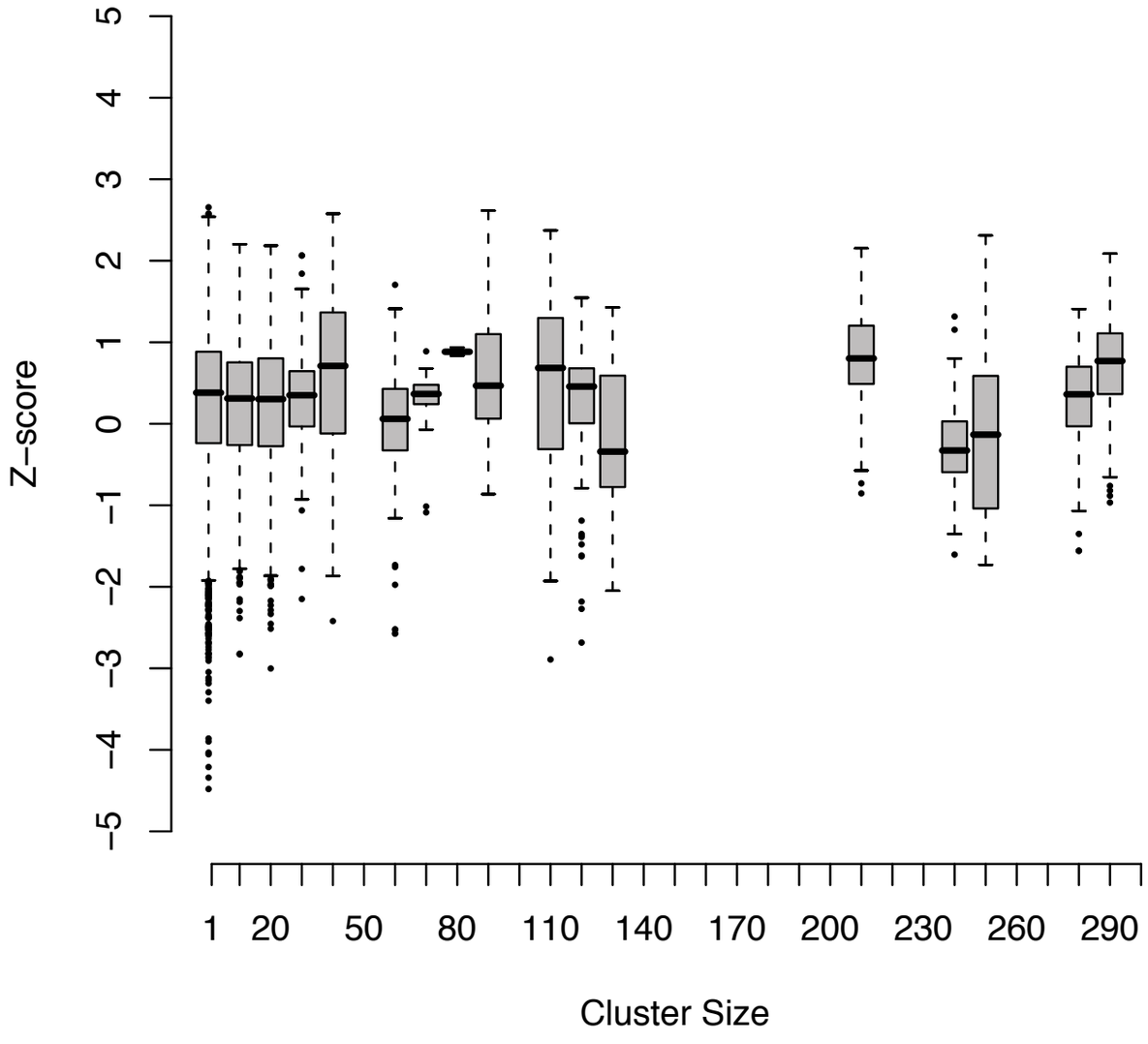


Structural Stability Decreases

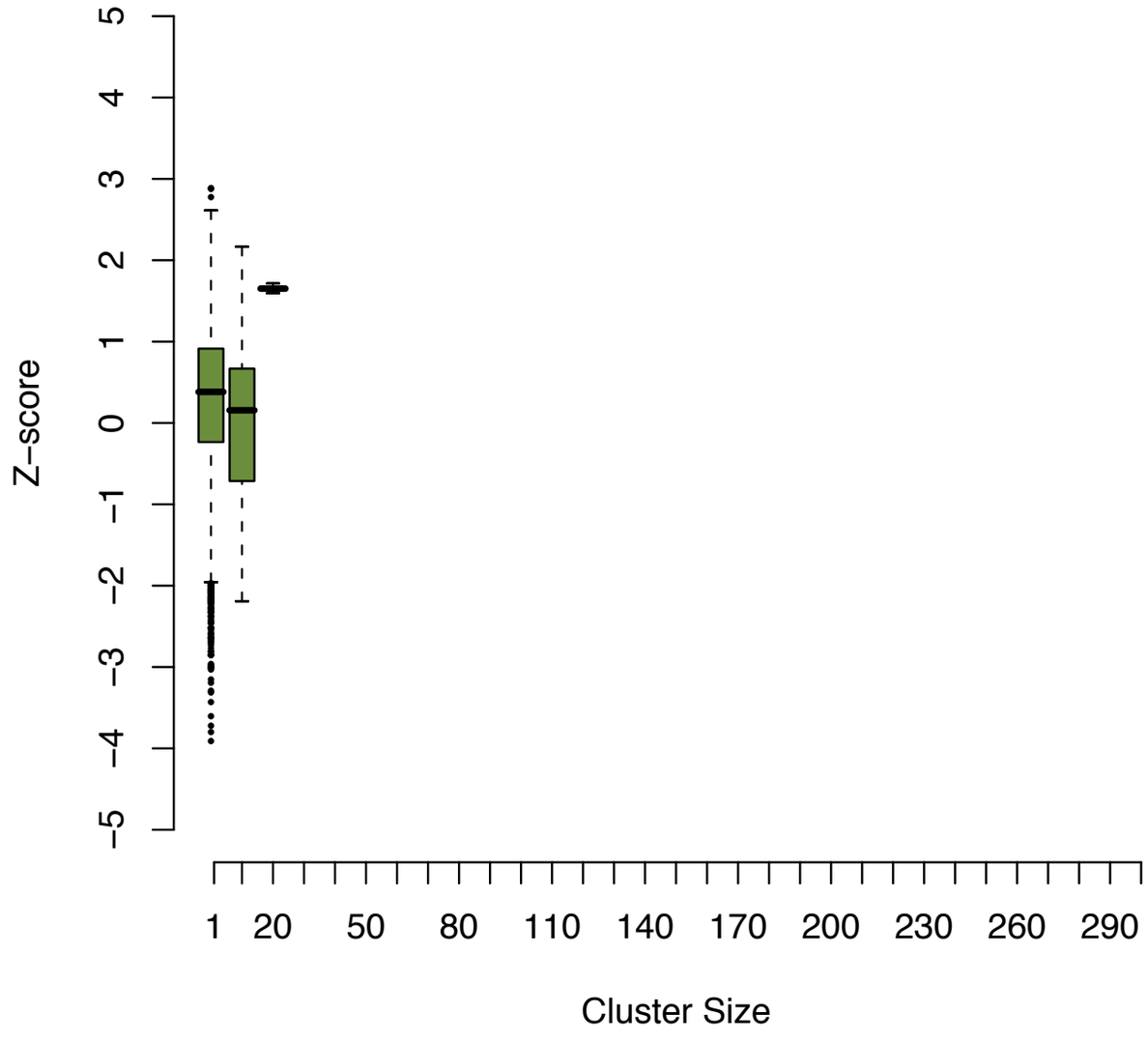


Enrichment Not Structurally Dependent

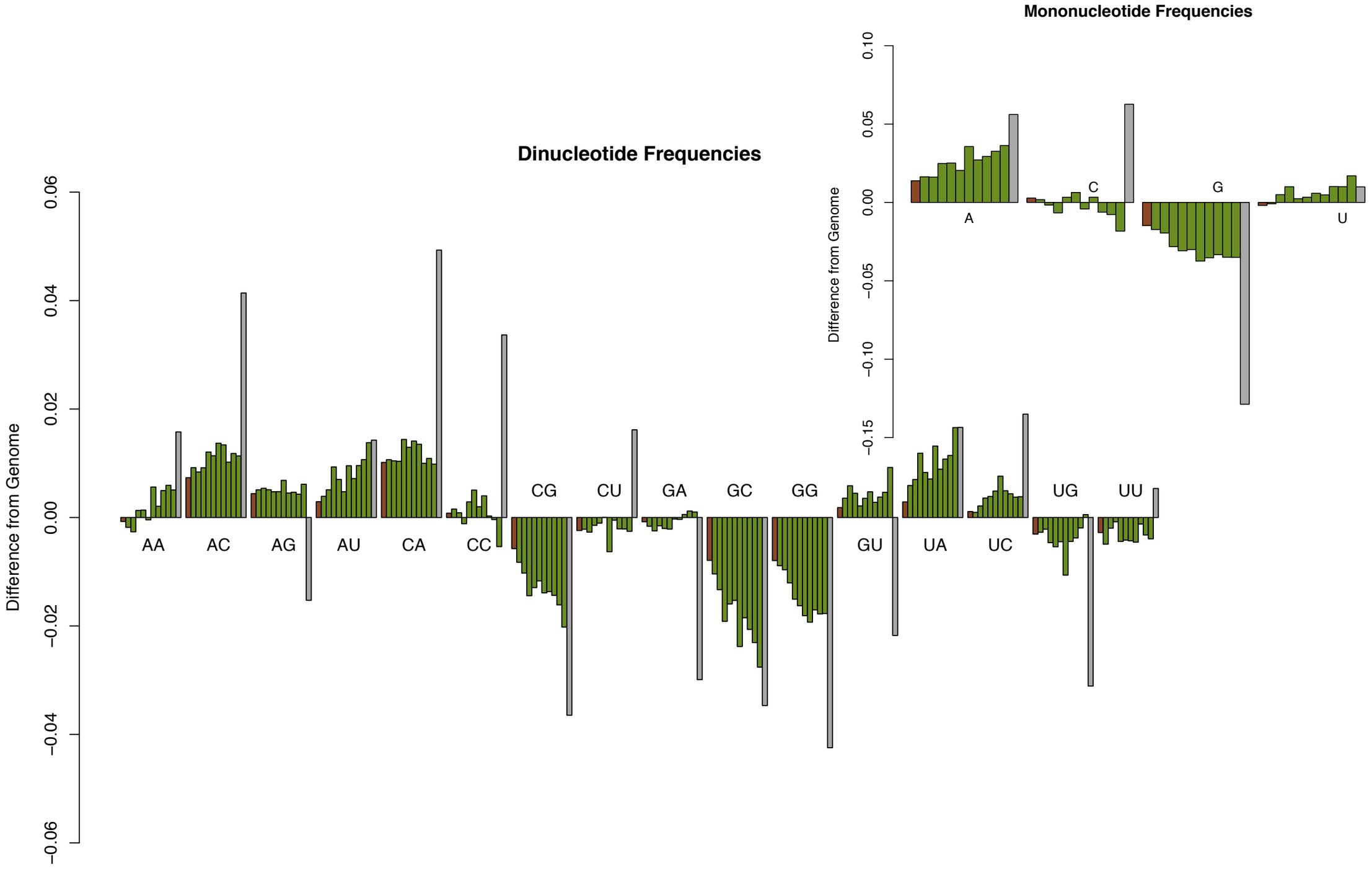
A



B



Changes in Nucleotide Content

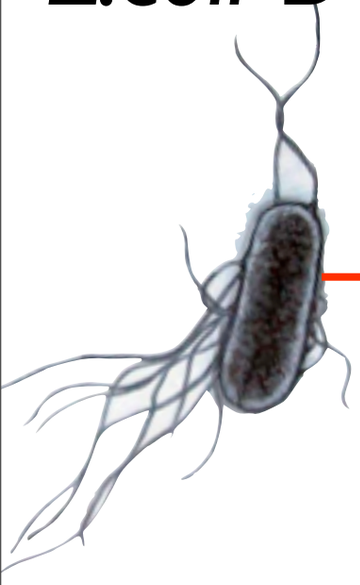


Conclusions

- Length and distance from the genome not very strongly affected
- The average stability of the structure is affected, however is not requisite for enrichment
- Nucleotide content changes are significant

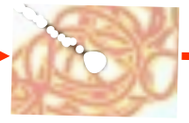
Experimental Setup

E.coli B

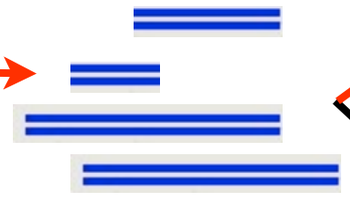


Genomic Library

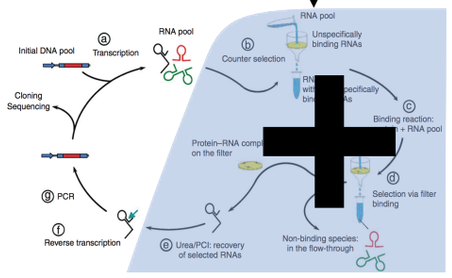
(cDNA Fragments Representing the Genome)



Genomic DNA

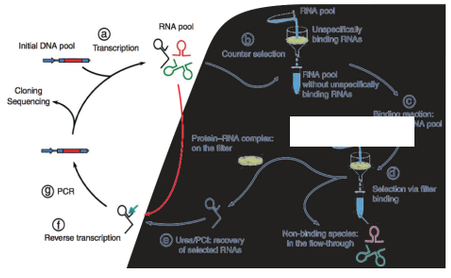


(Purified *E.coli* Hfq)

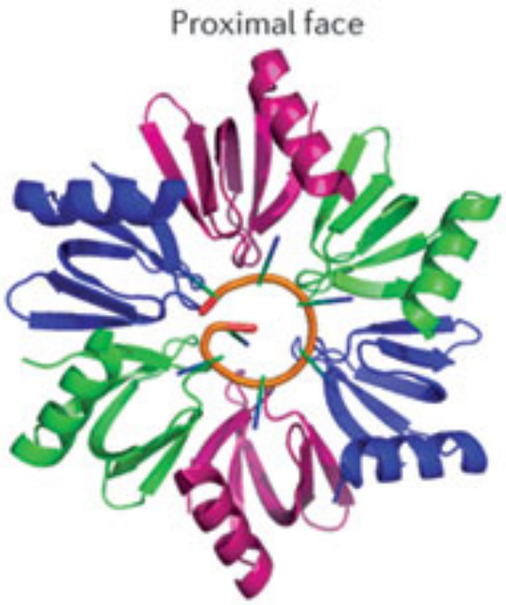


Genomic SELEX

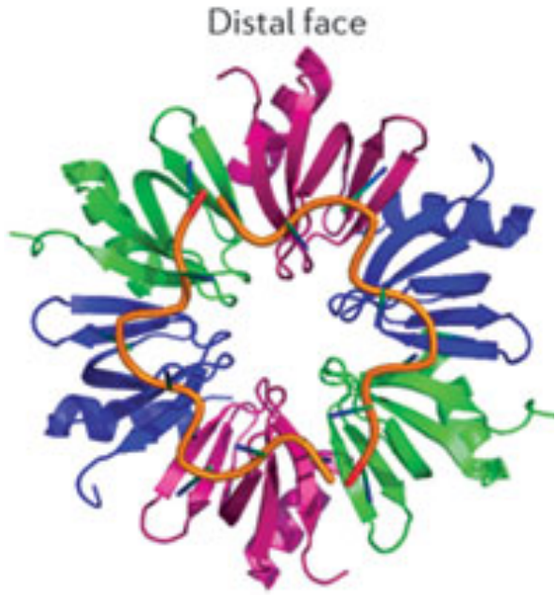
Neutral SELEX



(No protein selection)



Hfq is...



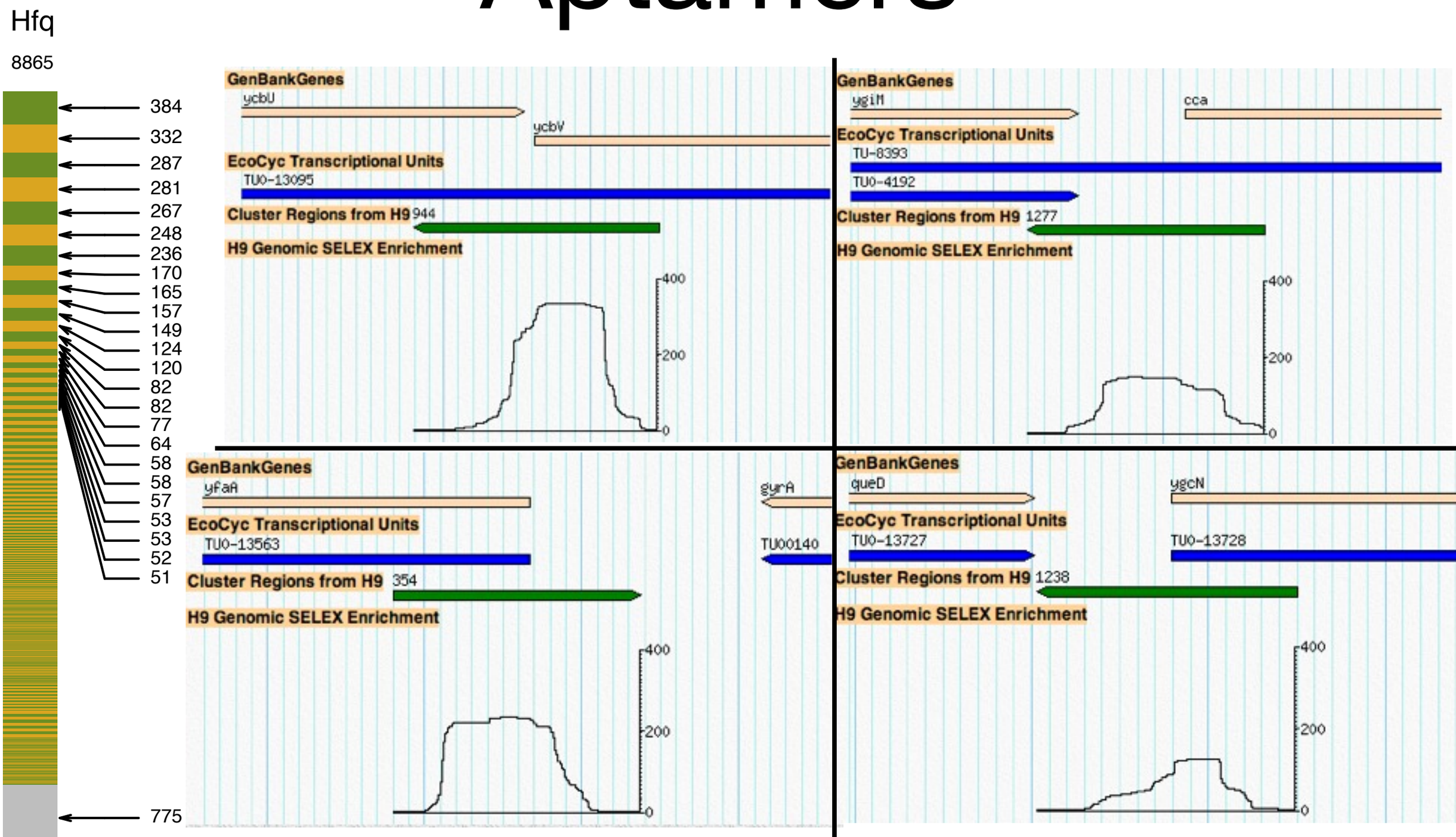
- Bacteria's prominent RNA-binding protein
- Present in half of the sequenced bacteria
- Direct sRNA binder facilitating mRNA binding
- Translational repressor and activator
- Modulator of mRNA decay through Poly-A stimulation
- Loss of Hfq deregulates >70 abundant proteins
- Required for virulence
- Enhancer of stress response
- Regulator of σ^E and σ^S transcription factors
- Regulator of outer membrane protein expression
- Has a role in tRNA biogenesis
- Helps survival in low-gravity conditions

Vogel, et al 2011

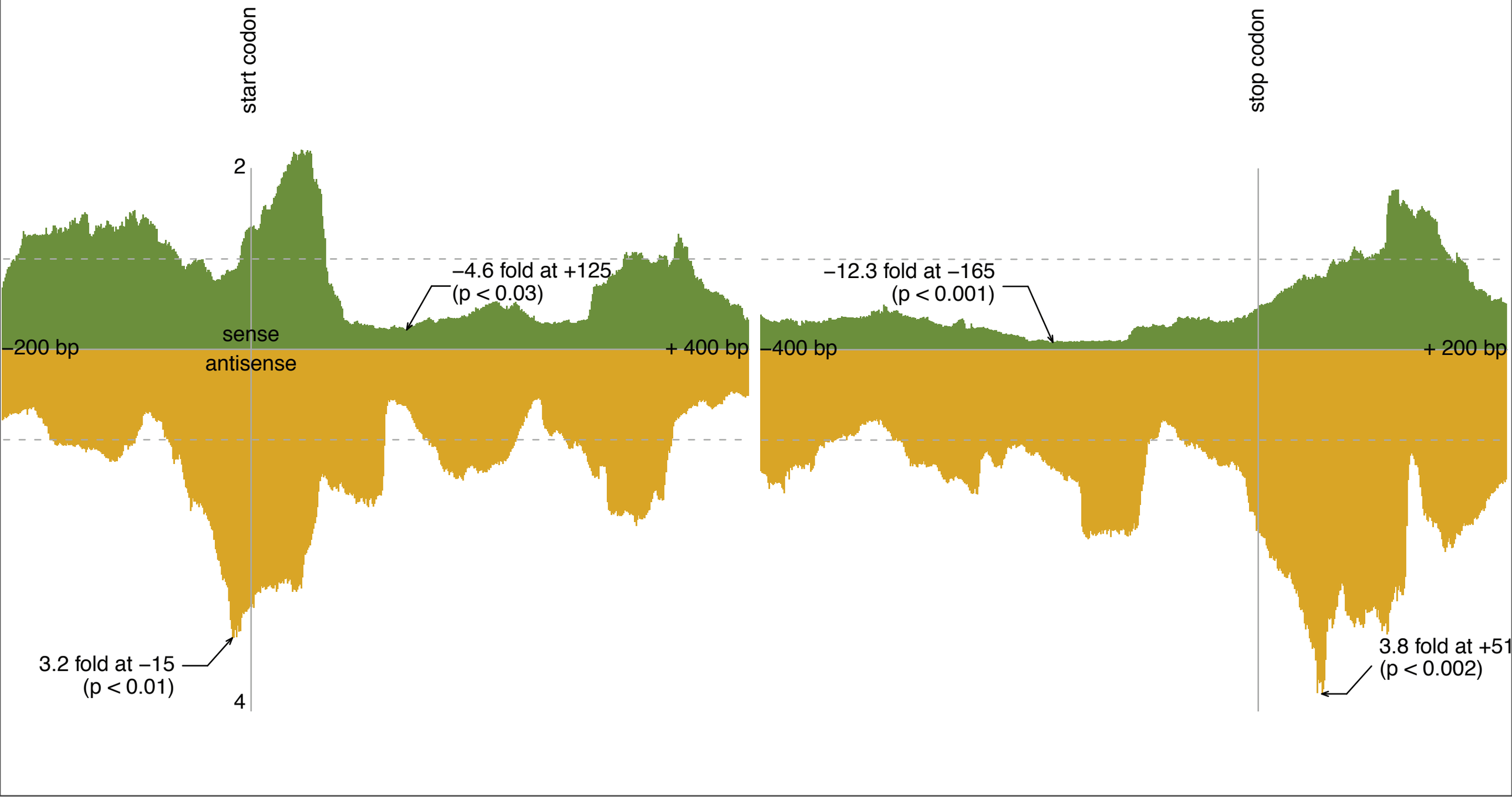
Locations of Hfq Genomic Aptamers

Feature Type	Genomic Distribution	Hfq SELEX Distribution	
		Sense	Antisense
CDS	90.8%	13.9% (-3.0)	68.7% (+1.5)
Intergenic	4.2%	2.7% (+1.3)	3.2% (+1.5)
misc ncRNA	1.0%	0.1% (*)	0.4% (-1.3)
Unattributed	4.0%	9.9% (+2.5)	

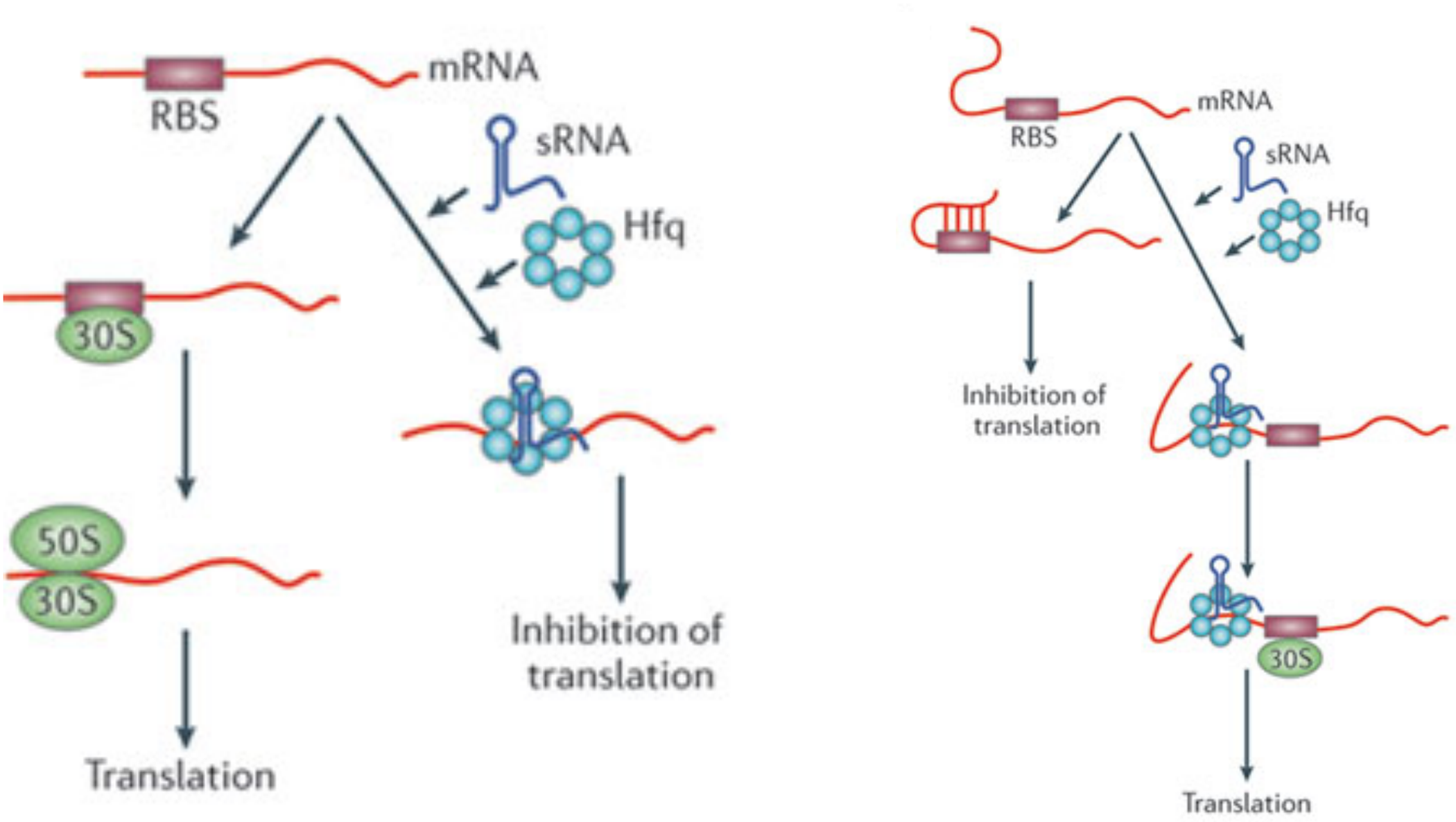
Some Top Genomic Aptamers



Enrichment Near Translational Signals

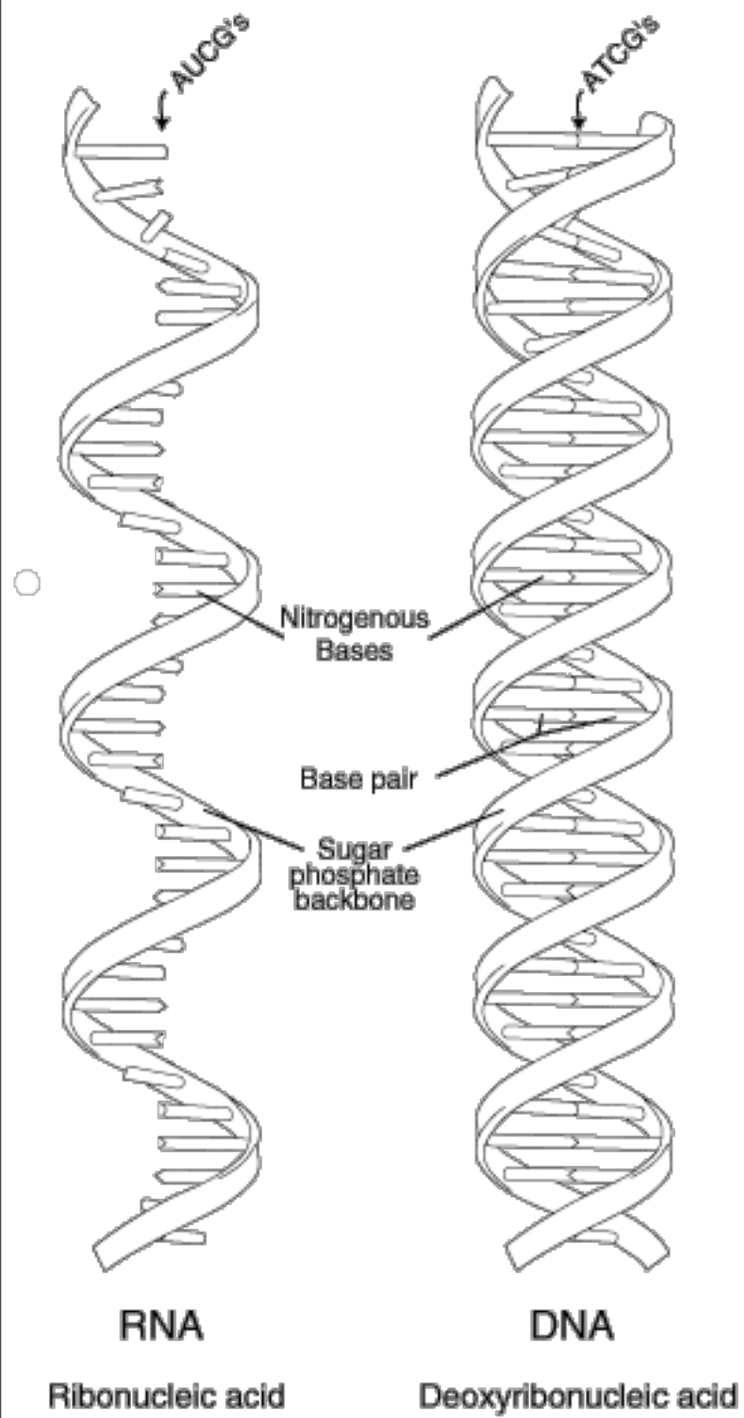


Models

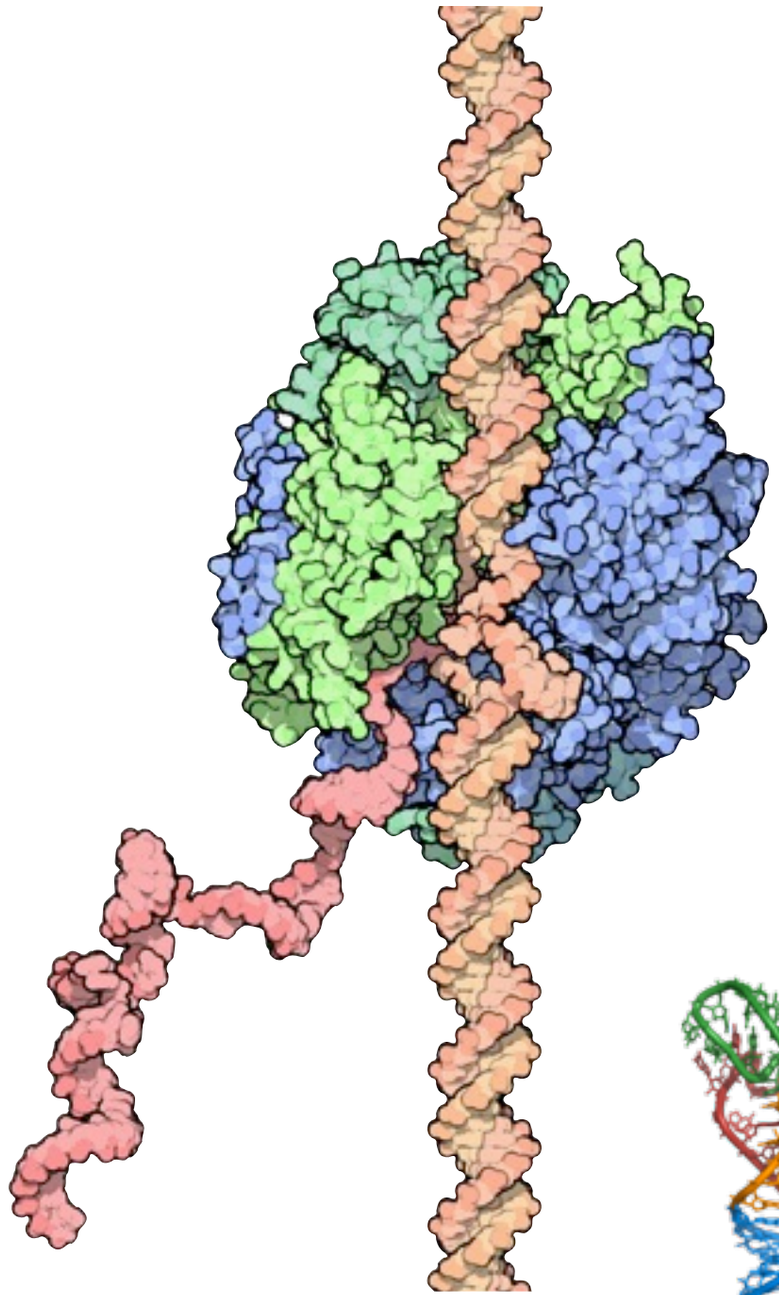


Vogel, et al 2011

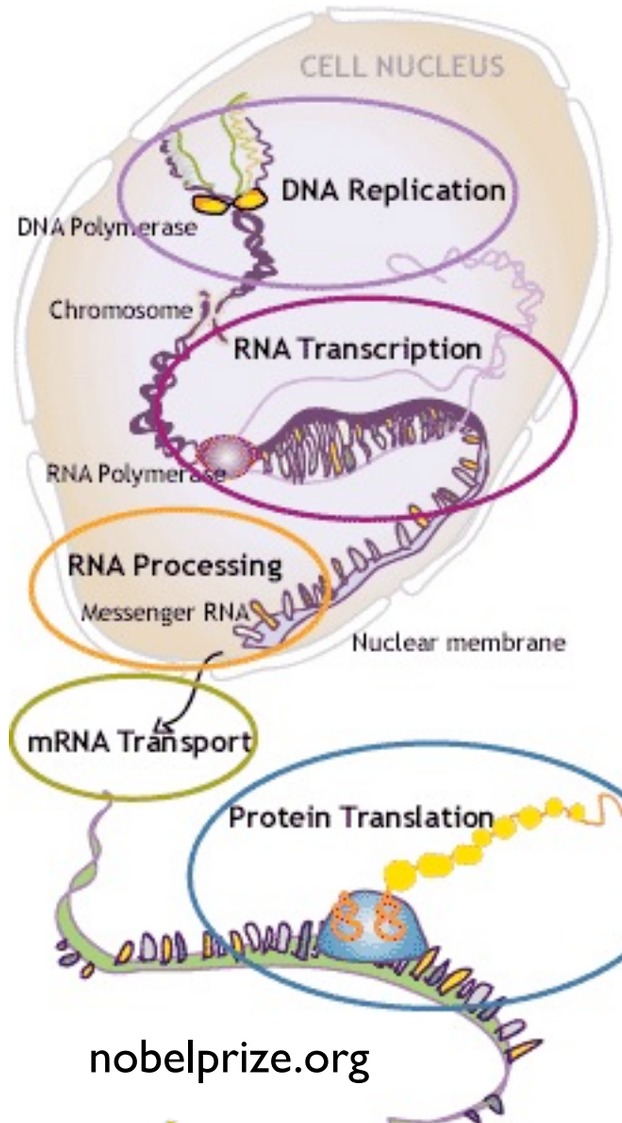
RNA



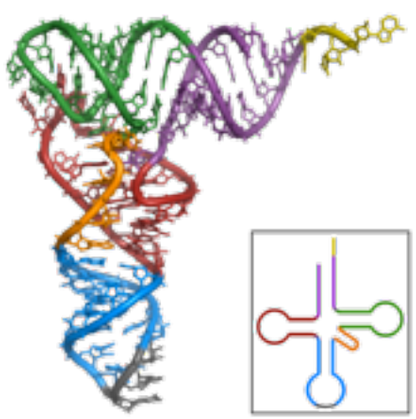
Darryl Leja, NHGRI



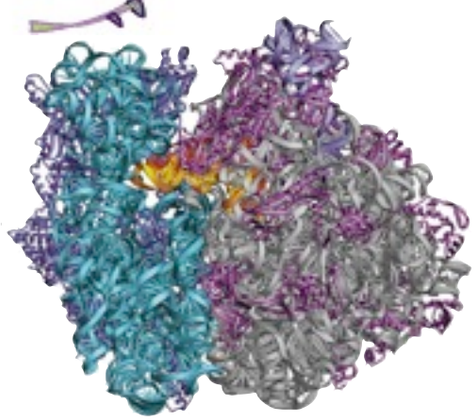
Goodsell, 2003



nobelprize.org



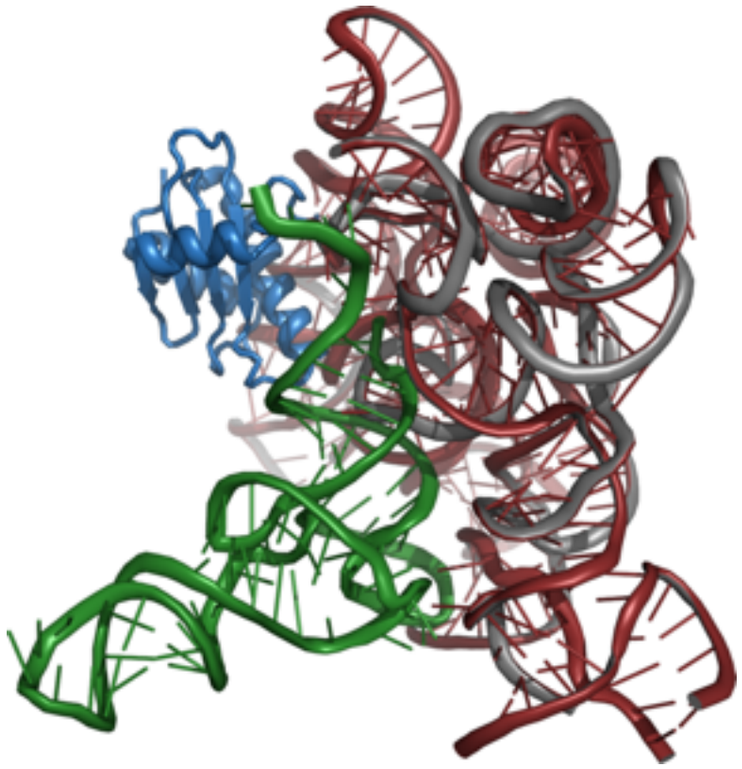
Yikrazuul 2010



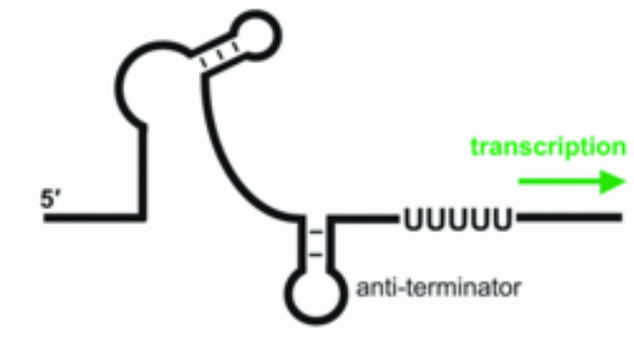
Barton, et al 2007

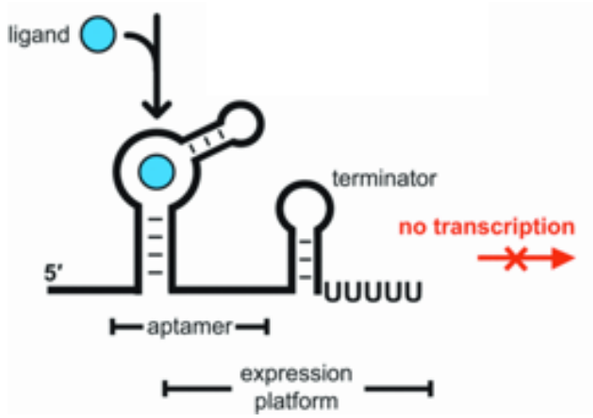
Non-coding RNAs

RNase P

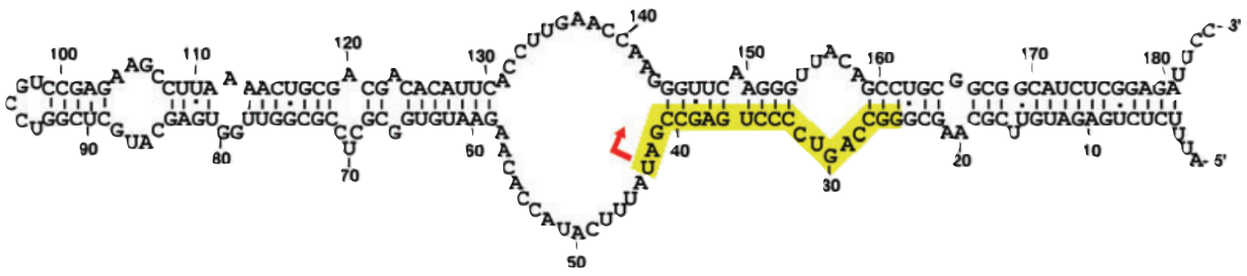


Riboswitches





6S RNA



- ~1% of 3.2B bases of human DNA code for protein

Kazantsev et al 2005

Kim and Breaker 2008

Wassarman 2007