



**University of
Zurich**^{UZH}

Zurich Open Repository and
Archive
University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2017

A thousand empirical adaptive landscapes and their navigability

Aguilar-Rodríguez, José; Payne, Joshua L.; Wägnér, Andreas

DOI: <https://doi.org/10.1088/1367-0332/19/1/015701>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.1088/1367-0332/19/1/015701>

Journal Article

Accepted Version

Originally published at:

Aguilar-Rodríguez, José; Payne, Joshua L.; Wägnér, Andreas (2017). A thousand empirical adaptive landscapes and their navigability. *Nature Ecology and Evolution*, 1(2):005

DOI: <https://doi.org/10.1088/1367-0332/19/1/015701>

A thousand empirical adaptive landscapes and their navigability

Abstract

Results

Adaptive landscapes of transcription factor binding affinity

E

E

!

dominant

"

!

1

additive

shuffled

Landscape navigability: The number of peaks

Landscape navigability: Epistasis

" !

in vivo

Navigability influences the *in vivo* abundance of binding sites

in vivo

Mus musculus

in vivo

in vivo

P

#

P-

Gene expression reflects landscape topography



P

Global peak breadth affects the diversity of binding sites

S. cerevisiae

P - #

P -

Discussion

in vivo

Drosophila melanogaster

in vitro *in vivo*

Materials and Methods

***In vitro* data**

!

!

***In vivo* data**

M. musculus

P #

et al.

P

in vivo

P

%

in vitro

in vivo

Genotype networks

E

Quantitative measures of landscape navigability

δ

δ

δ

E_i i

E_j j E_i E_j δ

E_i i

E_j j E_i δ E_j

δ

E &

δ

δ

number of peaks

i

j

i *j*

i *j*

peak

accessibility

epistasis

$$\varepsilon = E_{AB} + E_{ab} - E_{Ab} - E_{aB} \tag{3}$$

$$\begin{array}{ccc} & E_{AB} & \\ E_{Ab} & & E_{ab} \\ & E_{aB} & \\ & |\varepsilon| & \delta \end{array}$$

Magnitude epistasis

$$|\Delta E_{ab \rightarrow Ab} + \Delta E_{aB \rightarrow AB}| = |\Delta E_{ab \rightarrow Ab}| + |\Delta E_{aB \rightarrow AB}| \tag{4}$$

$$\begin{array}{lll} \Delta E & \text{mutational effect} & ab \rightarrow \\ Ab & \text{Simple sign epistasis} & \end{array}$$

$$|\Delta E_{ab \rightarrow Ab} + \Delta E_{aB \rightarrow AB}| < |\Delta E_{ab \rightarrow Ab}| + |\Delta E_{aB \rightarrow AB}| \tag{5}$$

$$\begin{array}{ll} \text{Reciprocal sign epistasis} & [3] \end{array}$$

$$|\Delta E_{ab \rightarrow aB} + \Delta E_{Ab \rightarrow AB}| < |\Delta E_{ab \rightarrow aB}| + |\Delta E_{Ab \rightarrow AB}| \tag{6}$$

$$\begin{array}{ccc} & \delta & \\ \delta & & |\varepsilon| \geq \delta \end{array}$$

Null models

additive

$$\# \, L$$

$$L$$

$$f_{i,b}$$

$$b$$

$$i.$$

$$L$$

$$S$$

$$L$$

$$S_{\text{pwm}} = \sum_{i=1}^l f_{i,b} \, I(i), \tag{7}$$

$$l$$

$$I(i)$$

$$i$$

$$I(i) = 2 + \sum_b f_{i,b} \log_2 f_{i,b} \tag{8}$$

δ

δ

shuffled

δ

References

Proceedings of the Sixth International Congress on Genetics **1**,

J. Stat. Mech. Theory Exp. **2013**,

Nat. Rev. Genet. **15**,

J.

Theor. Biol. **128**,

et al.

J. R. Soc. Interface **7**,

Proc. Natl. Acad. Sci. U. S. A. **110**,

Nat. Rev. Genet. **8**,

Proc. Natl. Acad. Sci. U. S. A. **104 Suppl**,

Nature **457**,

PLoS Genet. **6**,

et al.

Nat. Biotechnol. **30**,

J.

BMC Evol. Biol. **4**,

Science **315**,

Proc. Natl. Acad. Sci. U. S.

A. **105**,

PLoS Comput. Biol. **10**,

et al.

Proc. Natl. Acad. Sci. U. S. A. **107**,

Proc. Natl. Acad.

Sci. U. S. A. **110**,

et al.

Nat. Biotechnol. **32**,

Nucleic Acids Res. **37**,

et al.

Cell **158**,

et al.

Science

324,

et al.

Nat. Biotechnol. **24**,

Science **343**,

et al.

Genome Res. **19**,

Proc. Natl. Acad. Sci. U.

S. A. **110**,

Nature **225**,

Trends Genet. **27**,

J. Theor. Biol. **272**,

et al.

Cell **152**,

Nat. Commun. **4**,

Philos. Trans. R. Soc. Lond. B. Biol. Sci. **368**,

et al.
Cell **160**,

Nat. Protoc. **4**,

Bioinformatics **27**,

281,

J. Mol. Biol.

et al.

Nat. Biotechnol. **28**,

Bioinformatics **26**,

Biometry

Sewall Wright and Evolutionary Biology

The Adaptive Landscape in Evolutionary Biology

et al.

Nat. Genet. **43**,

Science **347**,

Science

352,

et al.

Science **352**,

Nat. Commun. **7**,

et al.

PLoS Genet. **7**,

Mol. Biol. Evol. **33**,

Nature **490**,

Nature **497**,

Nucleic Acids Res. **29**,

Nucleic

Acids Res. **30**,

Elife **4**,

Phys. Rev. Lett. **86**,

Proc. Natl.

Acad. Sci. U. S. A. **83**,

Trends Biochem. Sci. **13**,

Proc. Natl. Acad. Sci. U. S. A. **99**,

Proc. Natl. Acad. Sci. U. S. A. **99**,

(
PLoS Genet. **11**,

Acknowledgments

Author Contributions

Competing financial interests:

Data availability:

Figures

Figure 1. Adaptive landscapes of transcription factor binding affinity. A

E



Figure 2. The navigability of adaptive landscapes of transcription factor binding affinity. A

B

P #

C

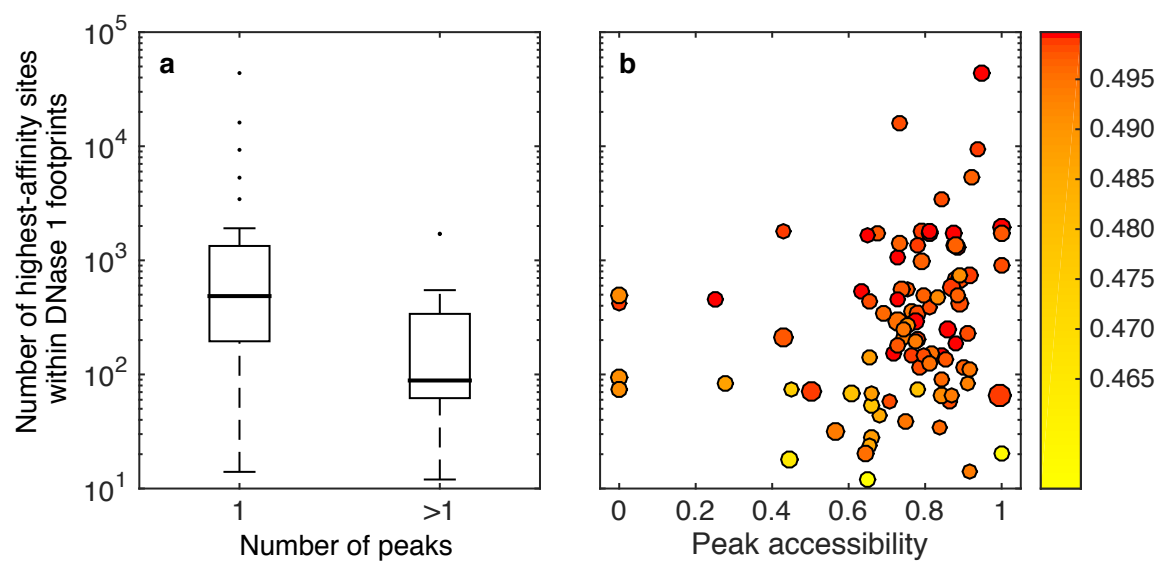


Figure 3. *In vivo* binding site abundance correlates with landscape navigability.

M. musculus

A

P

#

B

P-

#

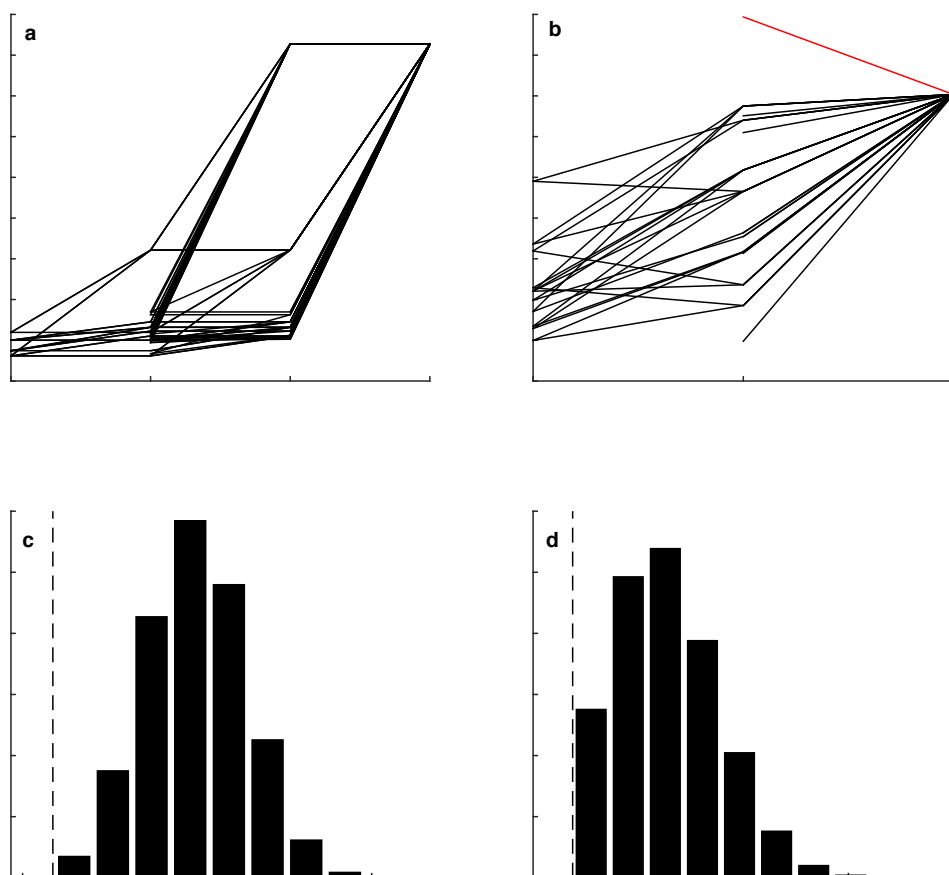


Figure 4. Gene expression increases along accessible mutational paths and reflects landscape topography. A, B

A

B

B

δ

δ

δ

B

C, D

C)

D

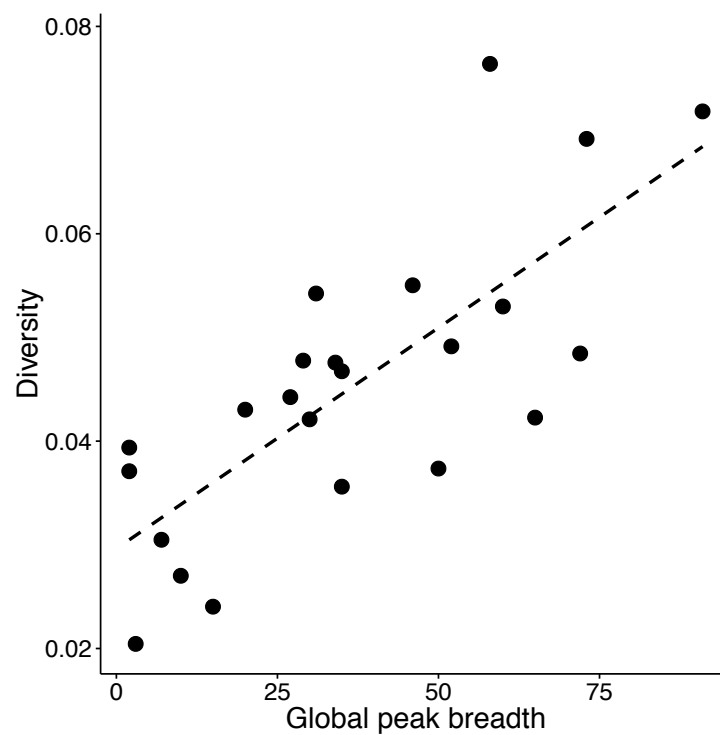
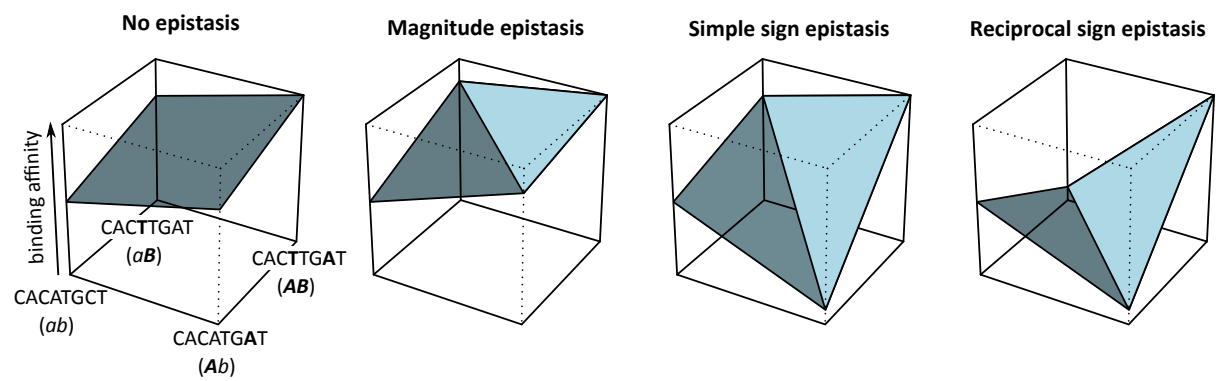


Figure 5 Global peak breadth influences the diversity of TF binding sites in the yeast genome.

S. cerevisiae

P-

#



Box 1. Epistasis.

AB ***Ab*** ***aB***

a ***A***

b ***B*** *Magnitude*

epistasis

Simple sign epistasis

Reciprocal sign epistasis

δ

sign epistasis

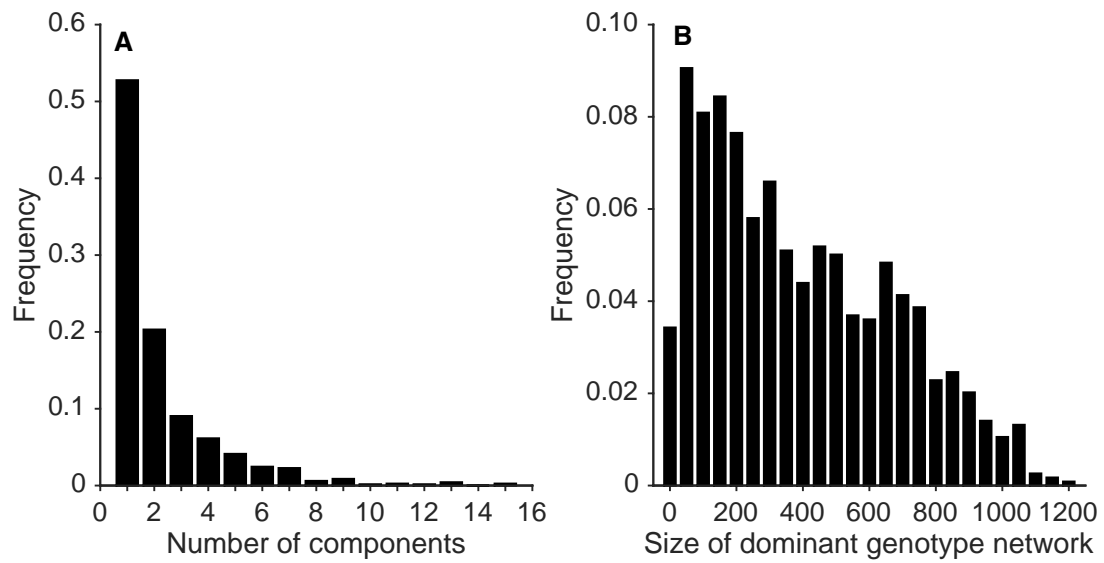
et al.

“A thousand empirical adaptive landscapes and their navigability”

Contents

1. Supplementary figures.....	33
2. Supplementary tables	66
3. Supplementary results.....	67
3.1. Summary statistics of genotype networks.....	67
3.2. Global peaks are usually organized into broad plateaus.....	67
3.3. Why epistasis occasionally appears in the additive null model.....	68
3.4. Sign epistasis preferentially occurs among nucleotides that are near one another in the binding site	68
3.5. Peak accessibility decreases when unbound sequences are included	69
3.6. Sensitivity analyses.....	70
<i>3.6.1. Our observations are insensitive to broadly varying thresholds for noise filtering</i>	
<i>3.6.2. Our observations are insensitive to broadly varying affinity thresholds for delineating bound from unbound sequences</i>	
<i>3.6.3. Our observations are consistent across DNA binding domains</i>	
<i>3.6.4. Our observations are consistent across TFs that bind shorter or longer sequences than eight nucleotides</i>	
<i>3.6.5. Peak breadth is sensitive to the use of E-scores as a quantitative phenotype</i>	
3.7. The <i>in vivo</i> relationship between landscape navigability and the abundance of binding sites is not driven by binding affinity or by information content	75
3.8. Our measures of epistasis for bound sequences are conservative	76
4. Supplementary discussion	77
4.1. Caveats.....	79
5. Supplementary references	81

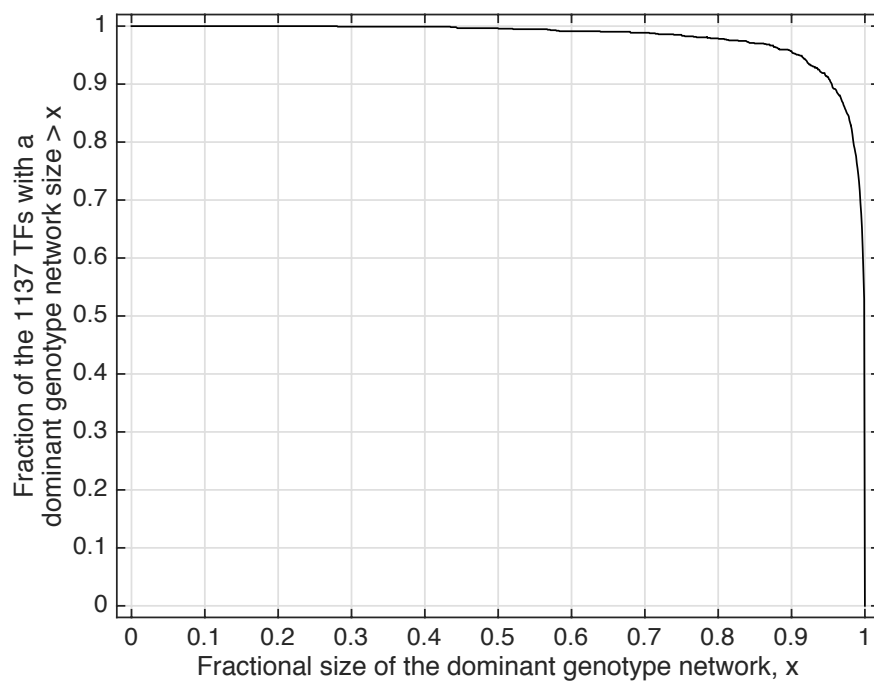
1. Supplementary figures



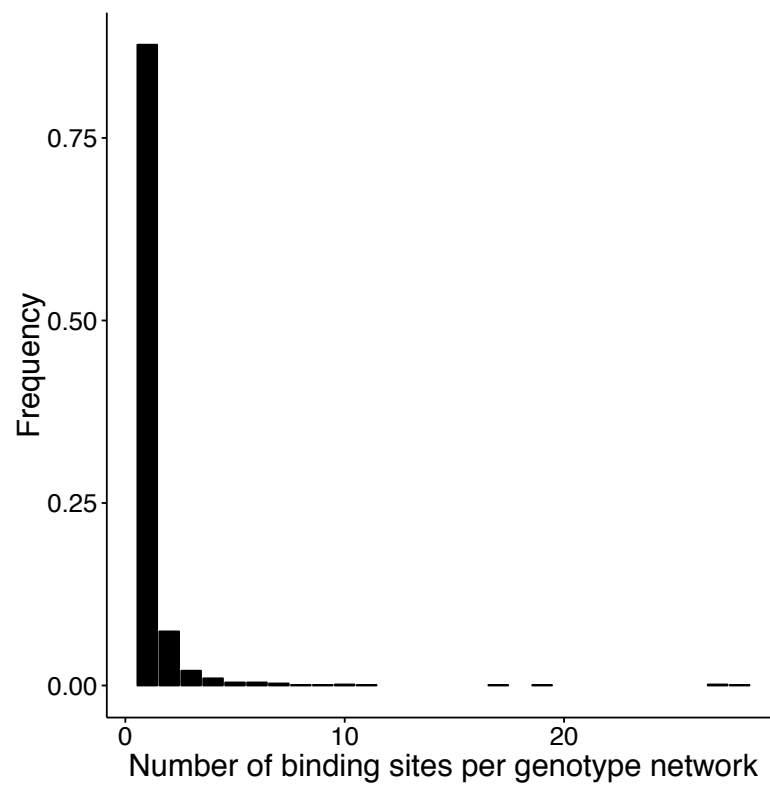
Supplementary Figure 1. Structural properties of the genotype networks of 1,137 TFs.

A

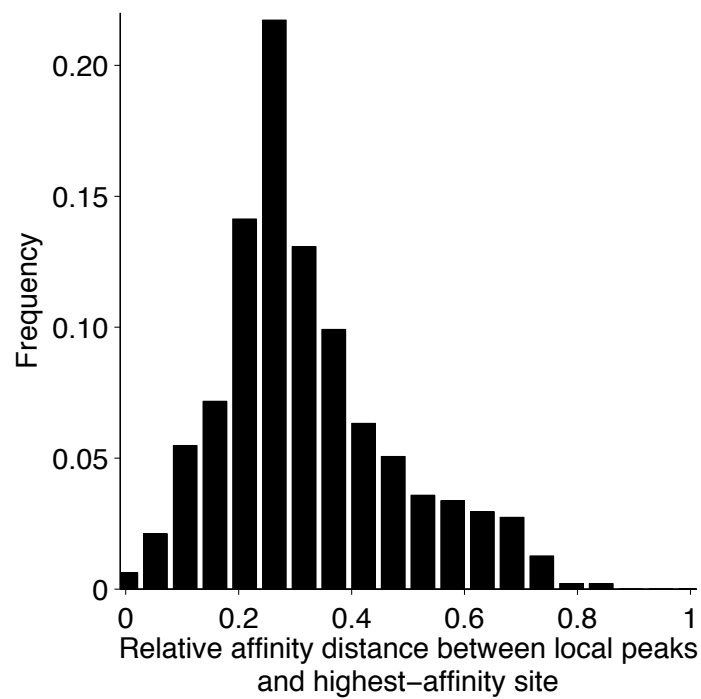
B



Supplementary Figure 2. Cumulative distribution of the fraction of TFs in our dataset that have a dominant genotype network comprising at least $x\%$ of the bound sequences.



Supplementary Figure 3. The number of sequences per non-dominant genotype network is very small.

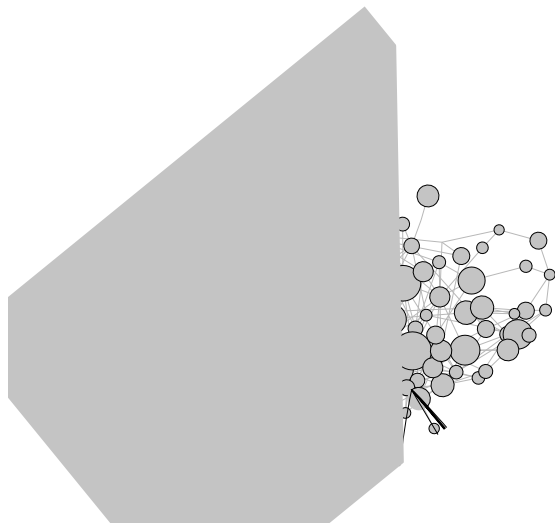


Supplementary Figure 4. Multiple peaks of unequal height.

$$D_i = \frac{1}{N_i} \sum_j E_{ij} = \frac{E_i}{N_i}$$

!

#



Supplementary Figure 5. Visualization of adaptive landscapes that have different numbers of peaks.

A

B

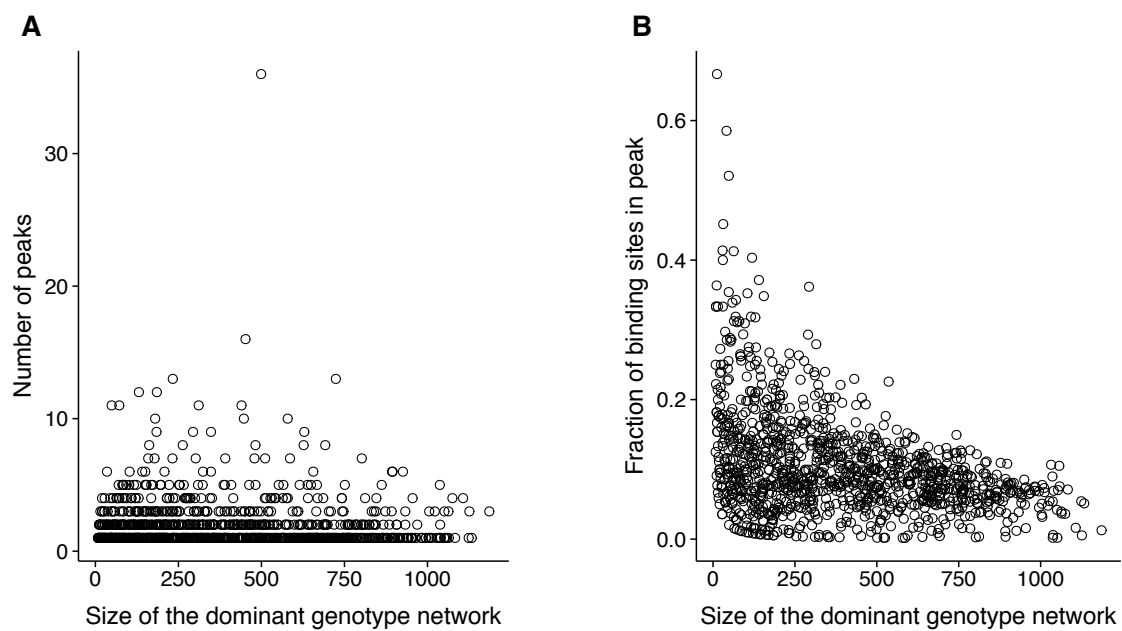
C

D

$$E_g - \delta$$

$$E_g$$

$$\delta$$



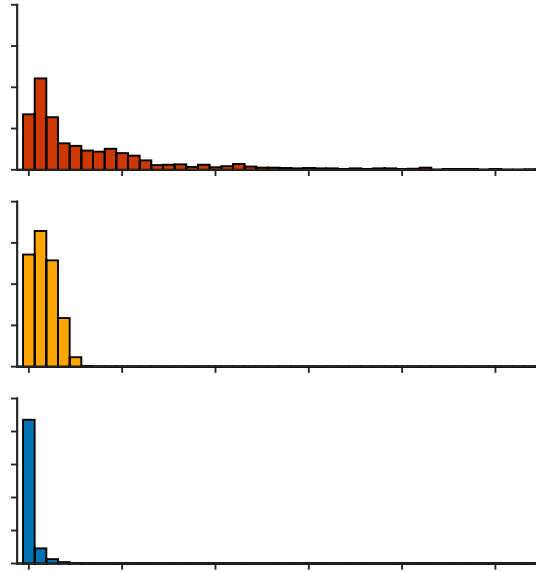
Supplementary Figure 6. Small landscapes are not necessarily more navigable than large landscapes.

A

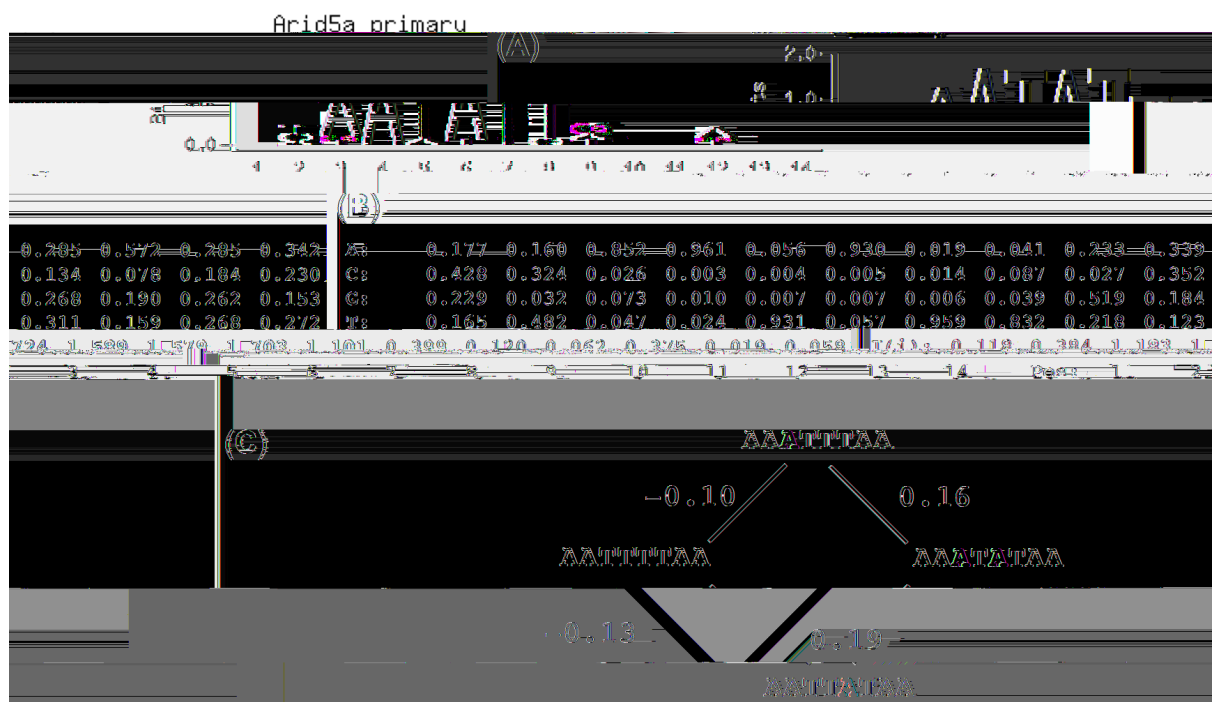
P - # B

P -

#



Supplementary Figure 7. Global peaks are usually organized into broad plateaus



Supplementary Figure 8. Why epistasis occasionally appears in the additive null model. A

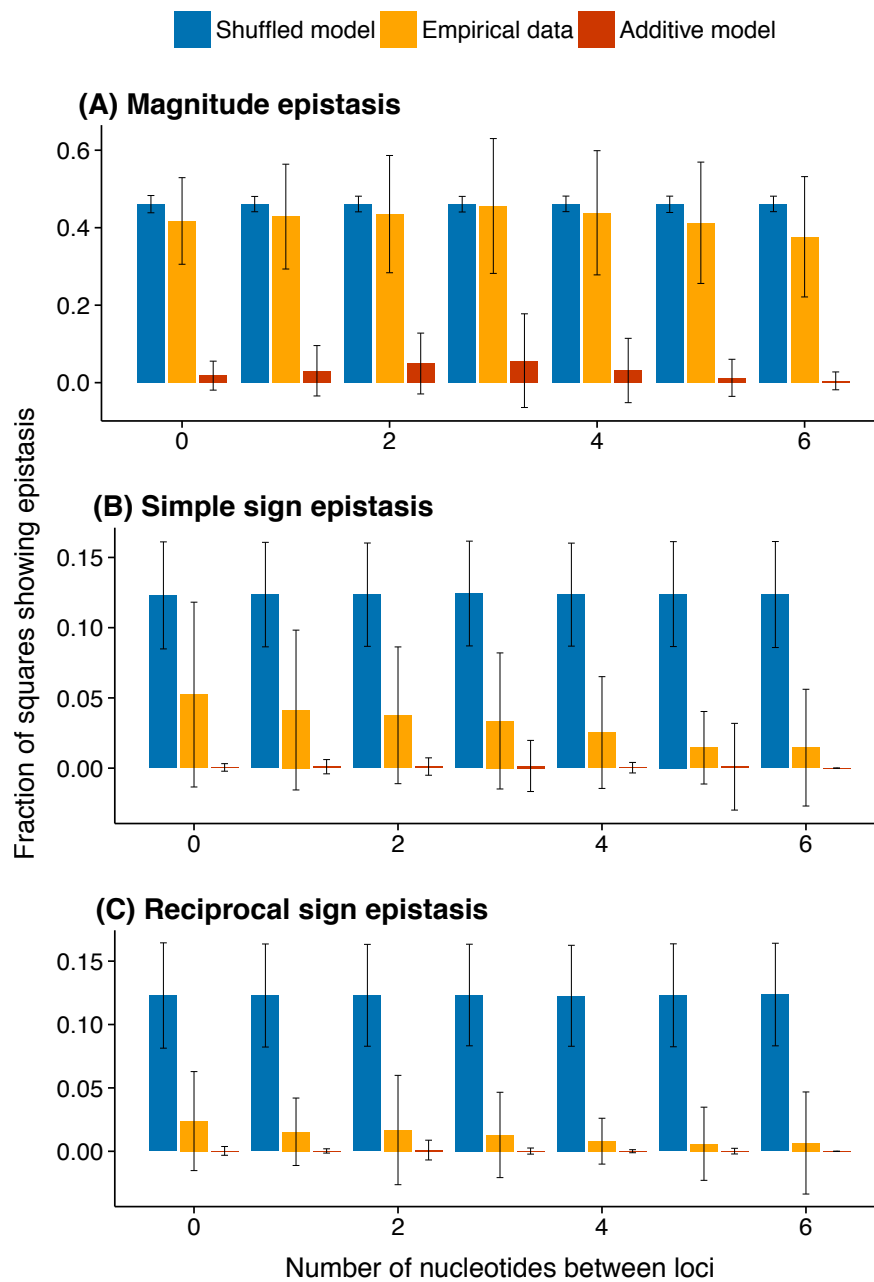
B

C

AATTTTAA

δ

ε



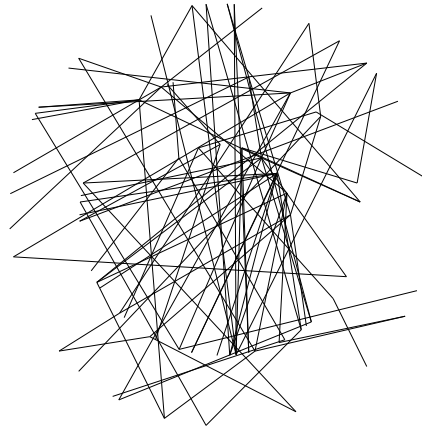
Supplementary Figure 9. Sign epistasis preferentially occurs among nucleotides that are near one another in the binding site.

A

B

C

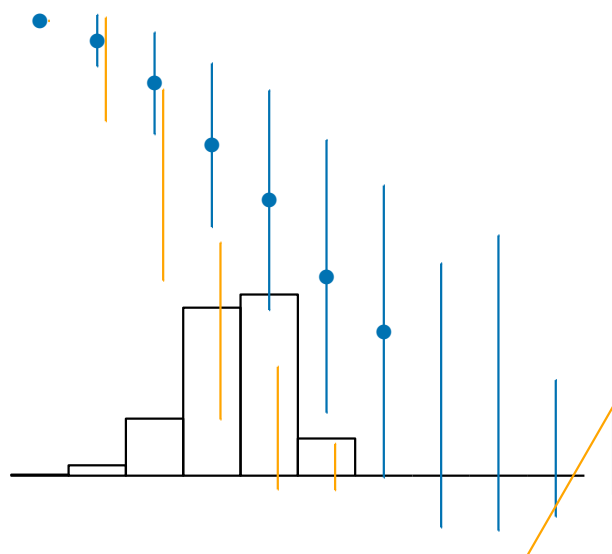
et al.



Supplementary Figure 10. Visualization of two global peaks that vary in their accessibility.

A

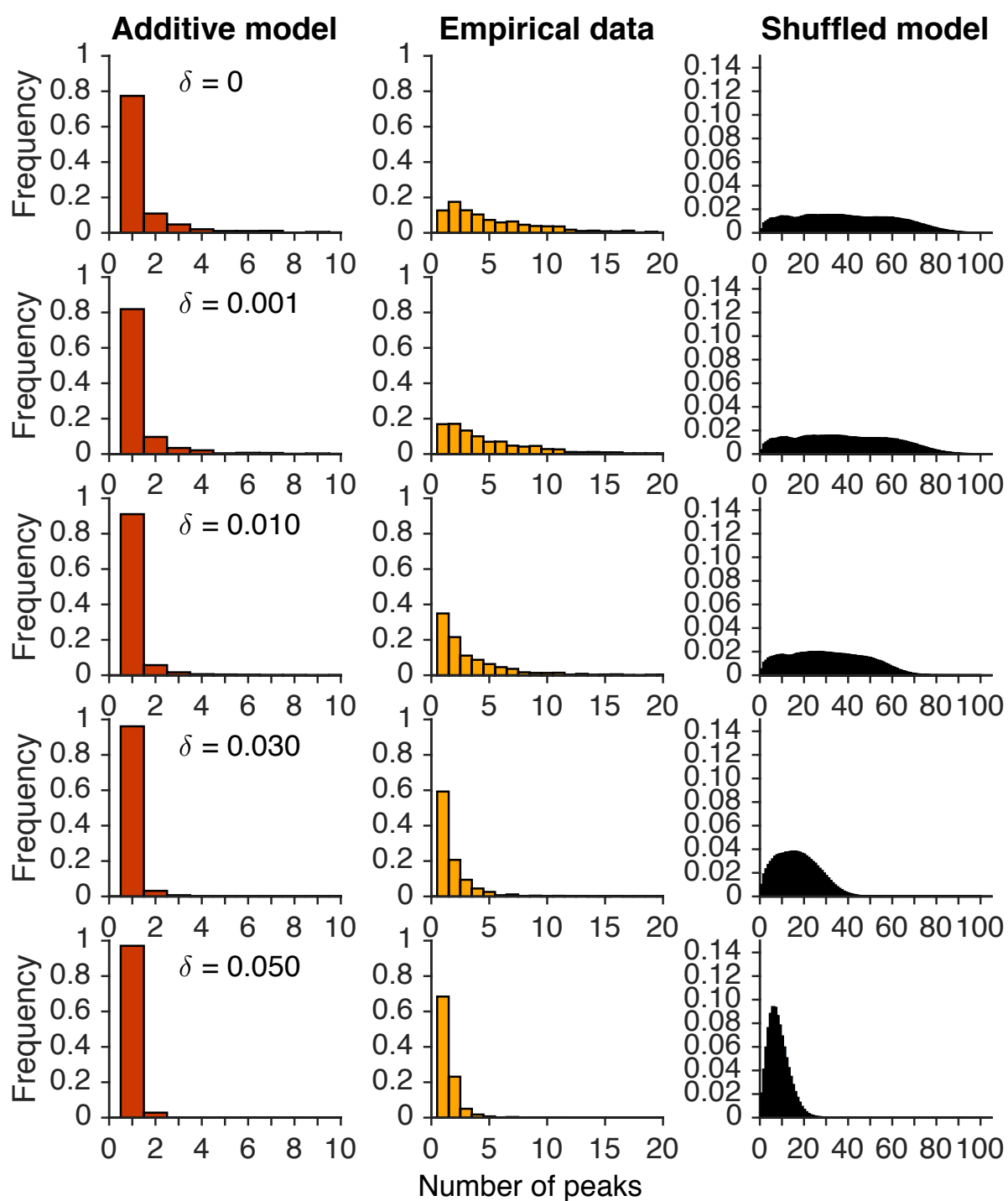
B



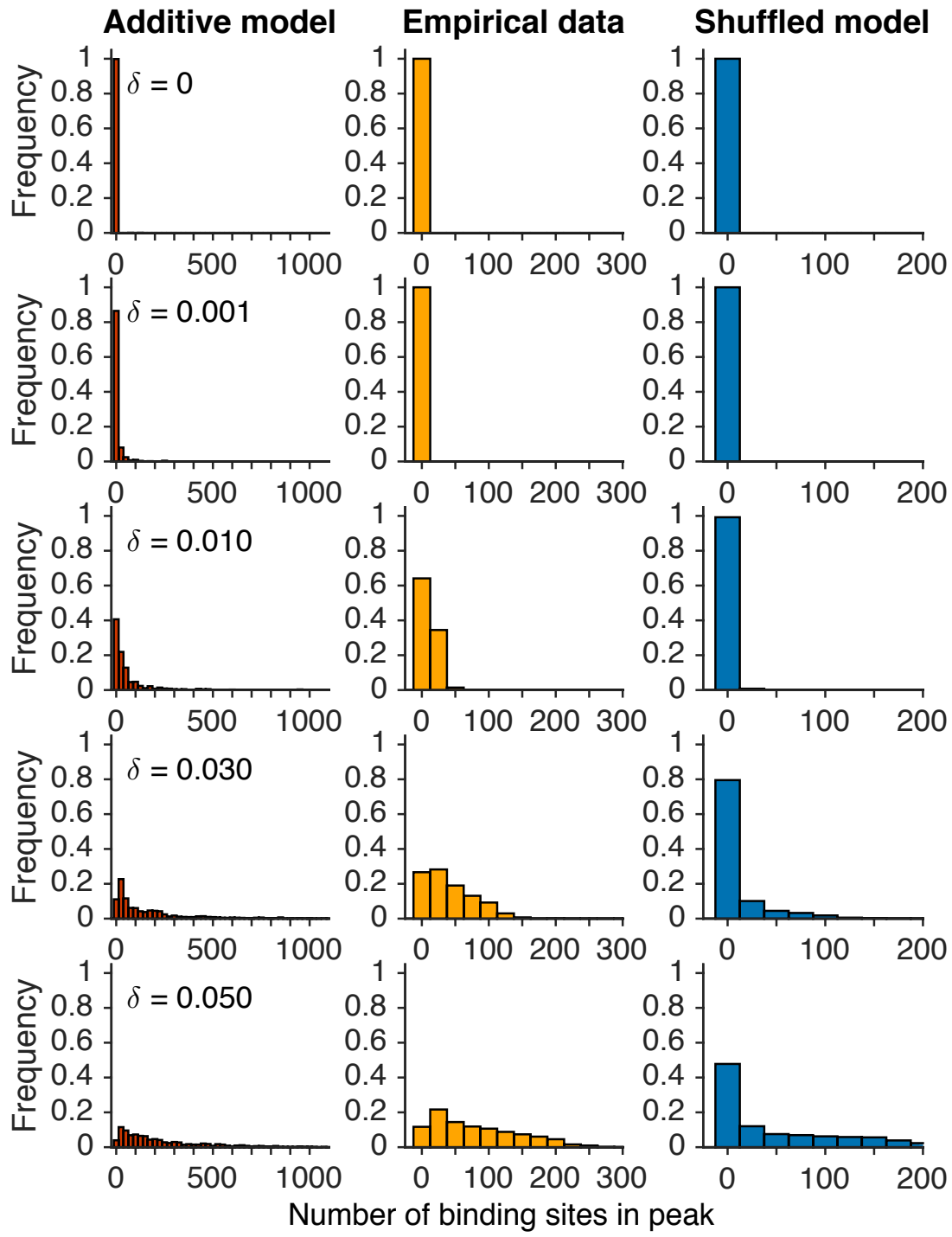
Supplementary Figure 11. Peak accessibility is reduced by the inclusion of unbound sequences.

!

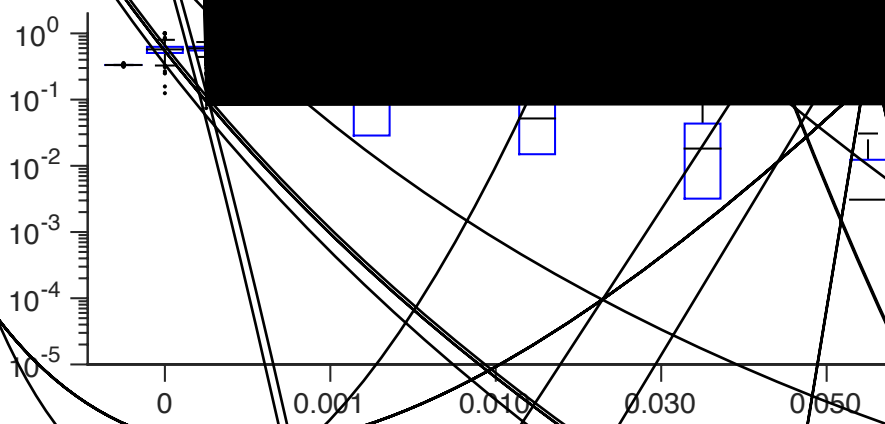
! &



Supplementary Figure 12. The number of peaks per landscape decreases as the noise threshold increases.



Supplementary Figure 13. Global peak breadth increases as the noise threshold increases.



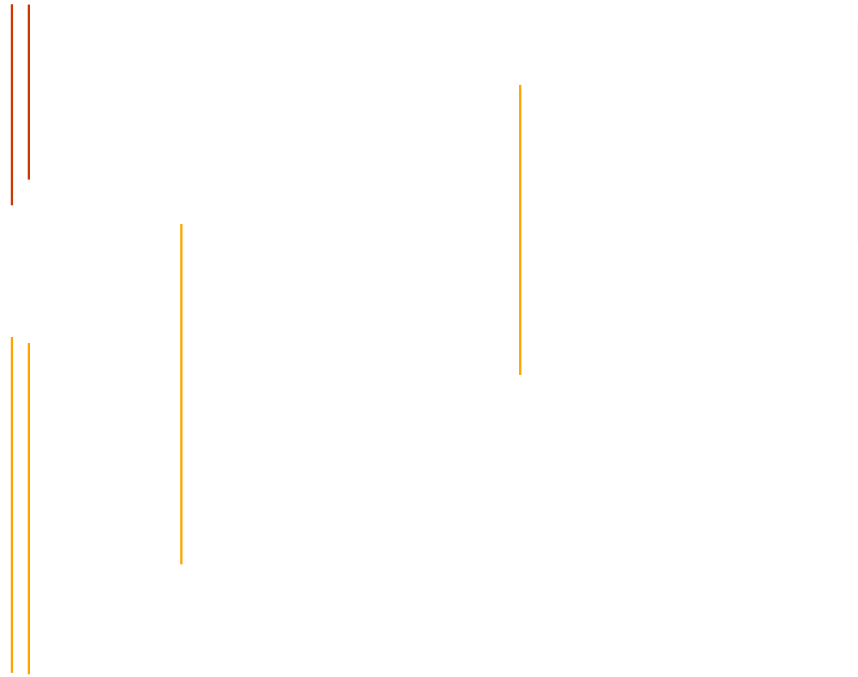
Supplementary Figure 14. Sign epistasis in the empirical data is intermediate to that of the additive and shuffled models for all noise thresholds.

C

A

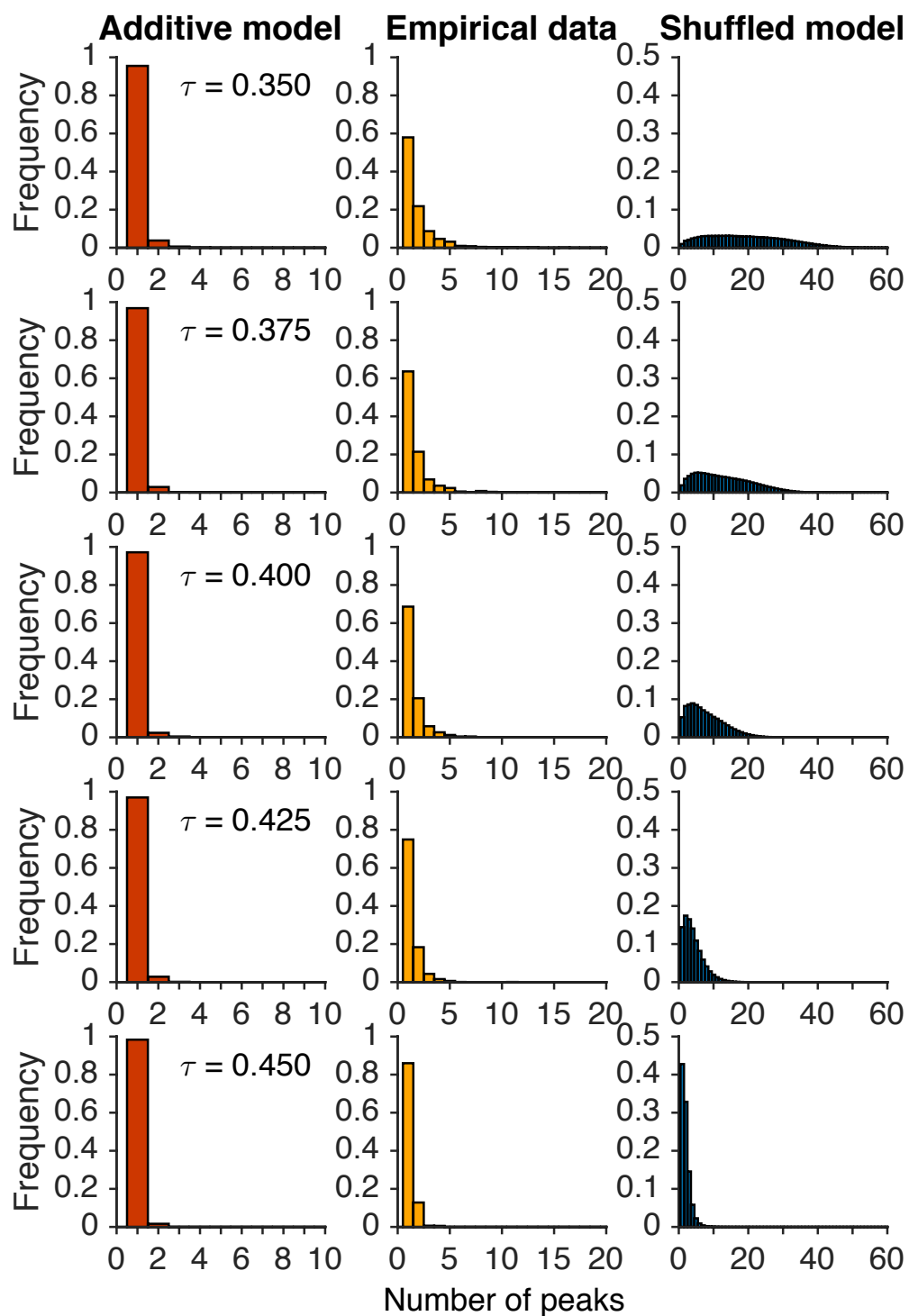
B

!

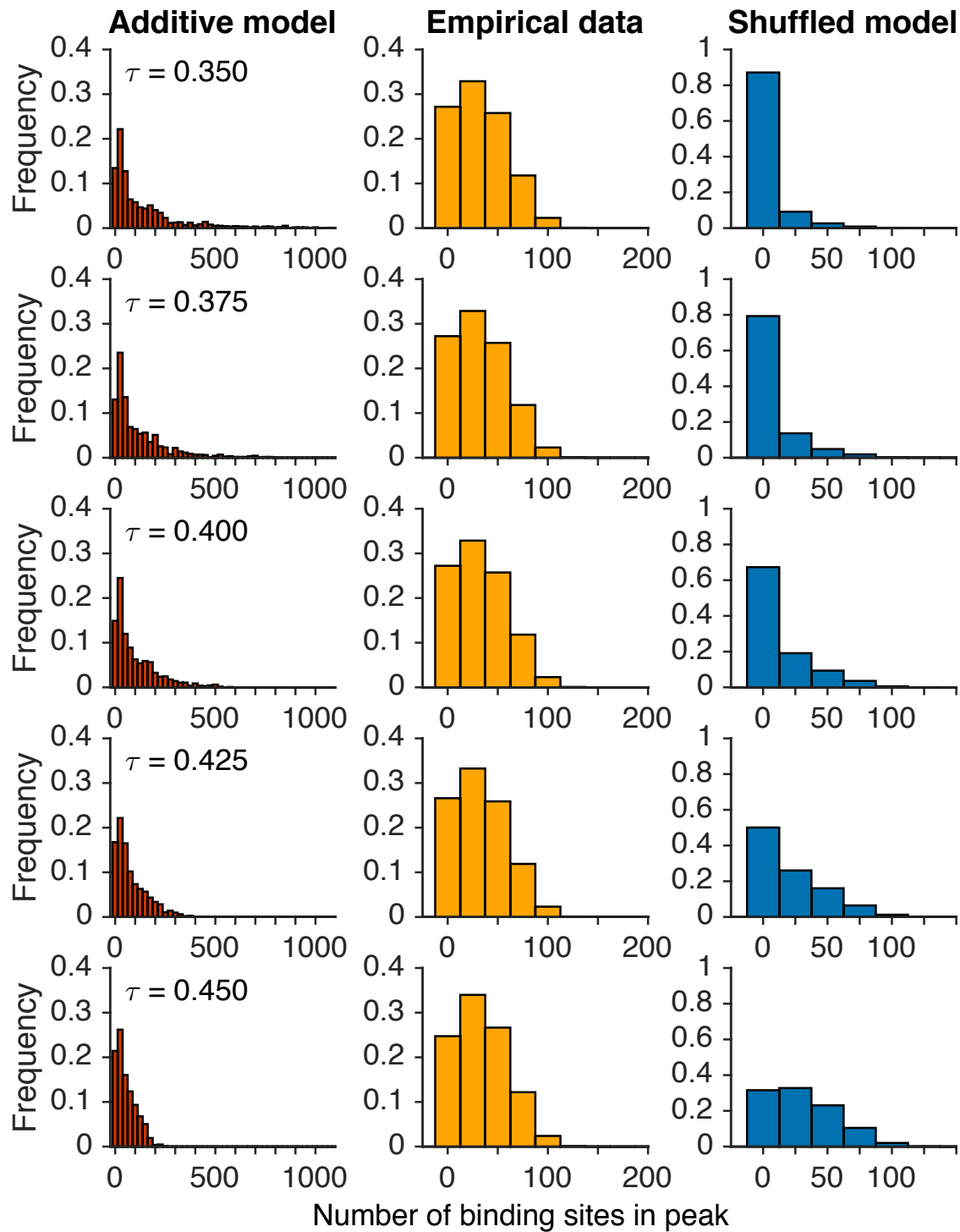


Supplementary Figure 15. Accessibility of the highest affinity site increases as the noise threshold increases.

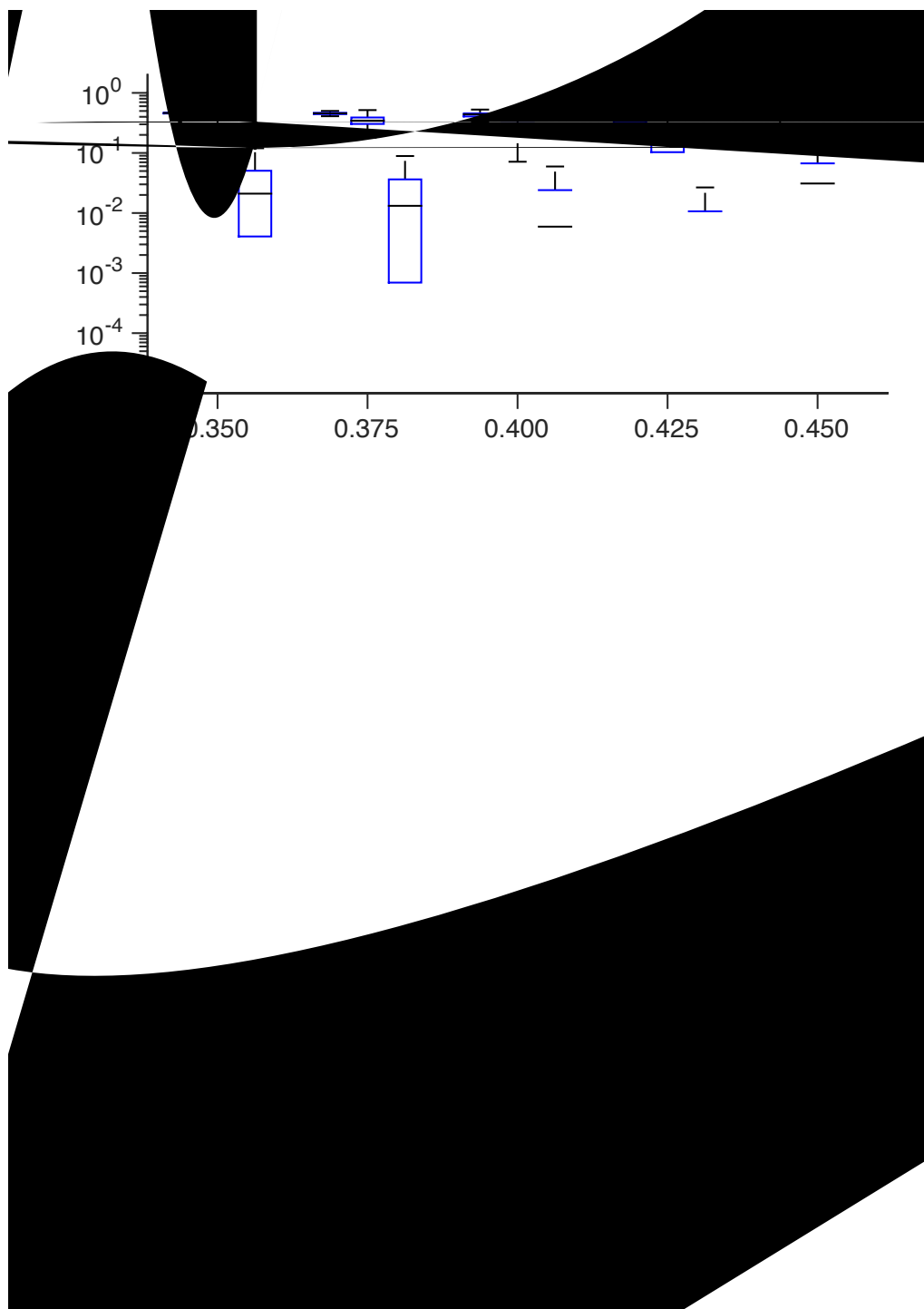
δ



Supplementary Figure 16. The number of peaks per landscape decreases as the binding affinity threshold increases.



Supplementary Figure 17. The empirical distribution of global peak breadth does not vary with the binding affinity threshold.



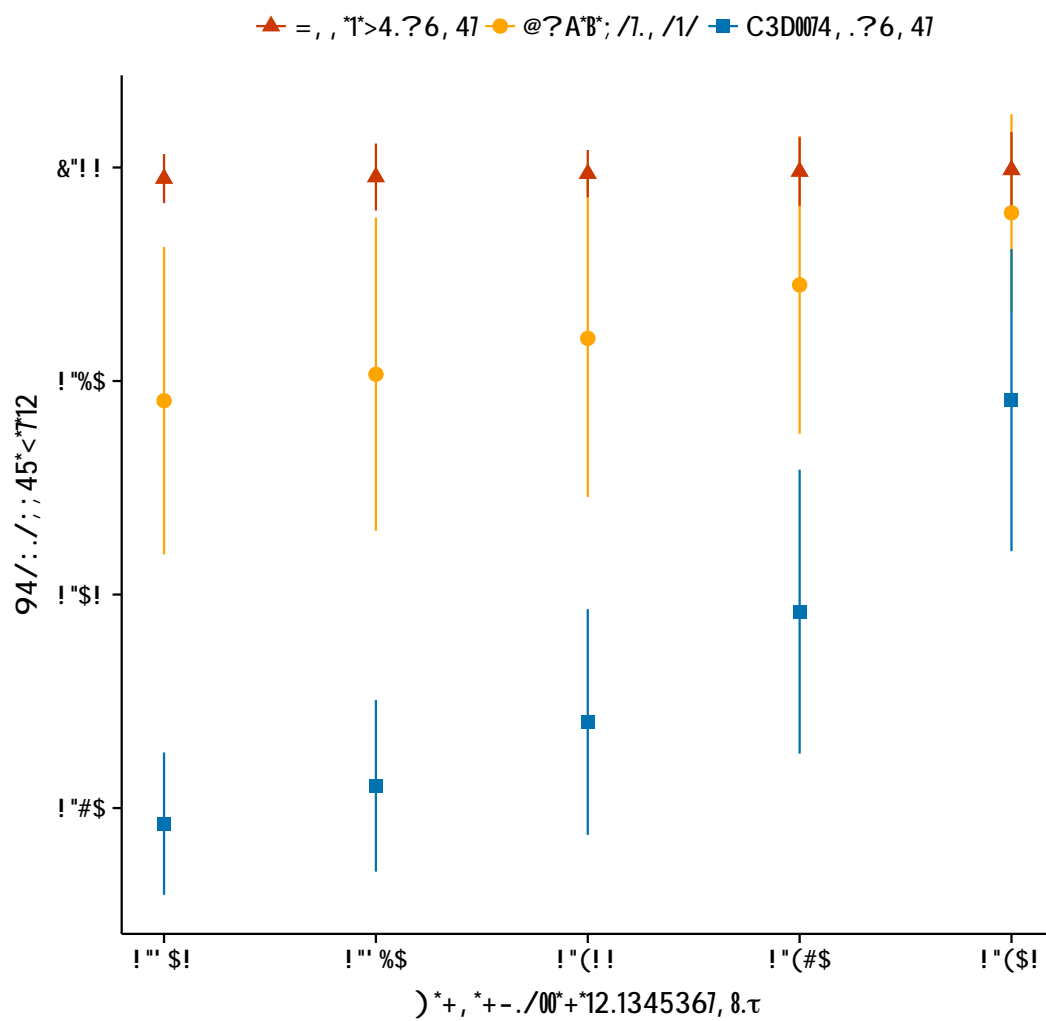
Supplementary Figure 18. Epistasis in the empirical data is intermediate to that of the additive and shuffled models for all affinity thresholds.

A

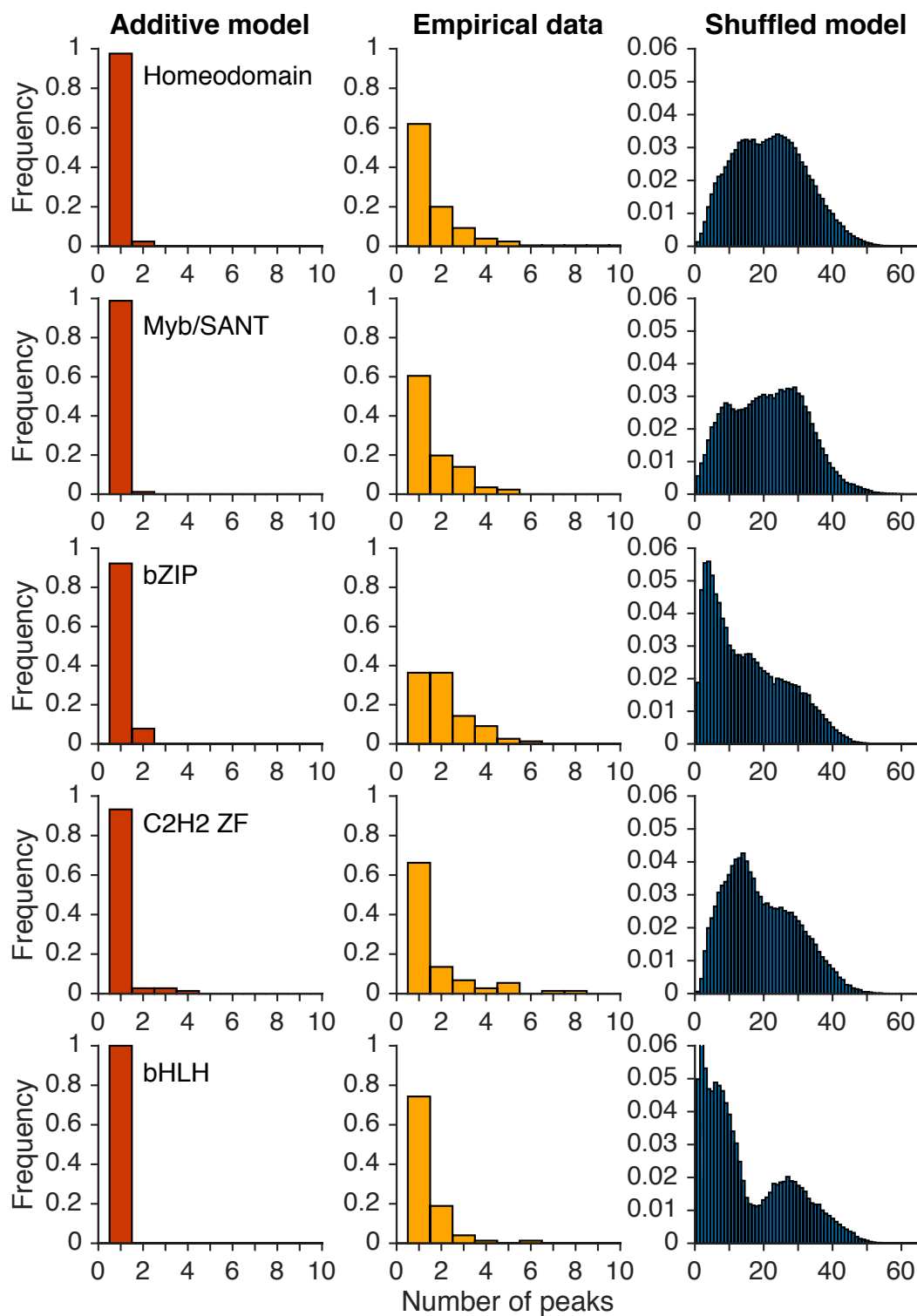
B

C

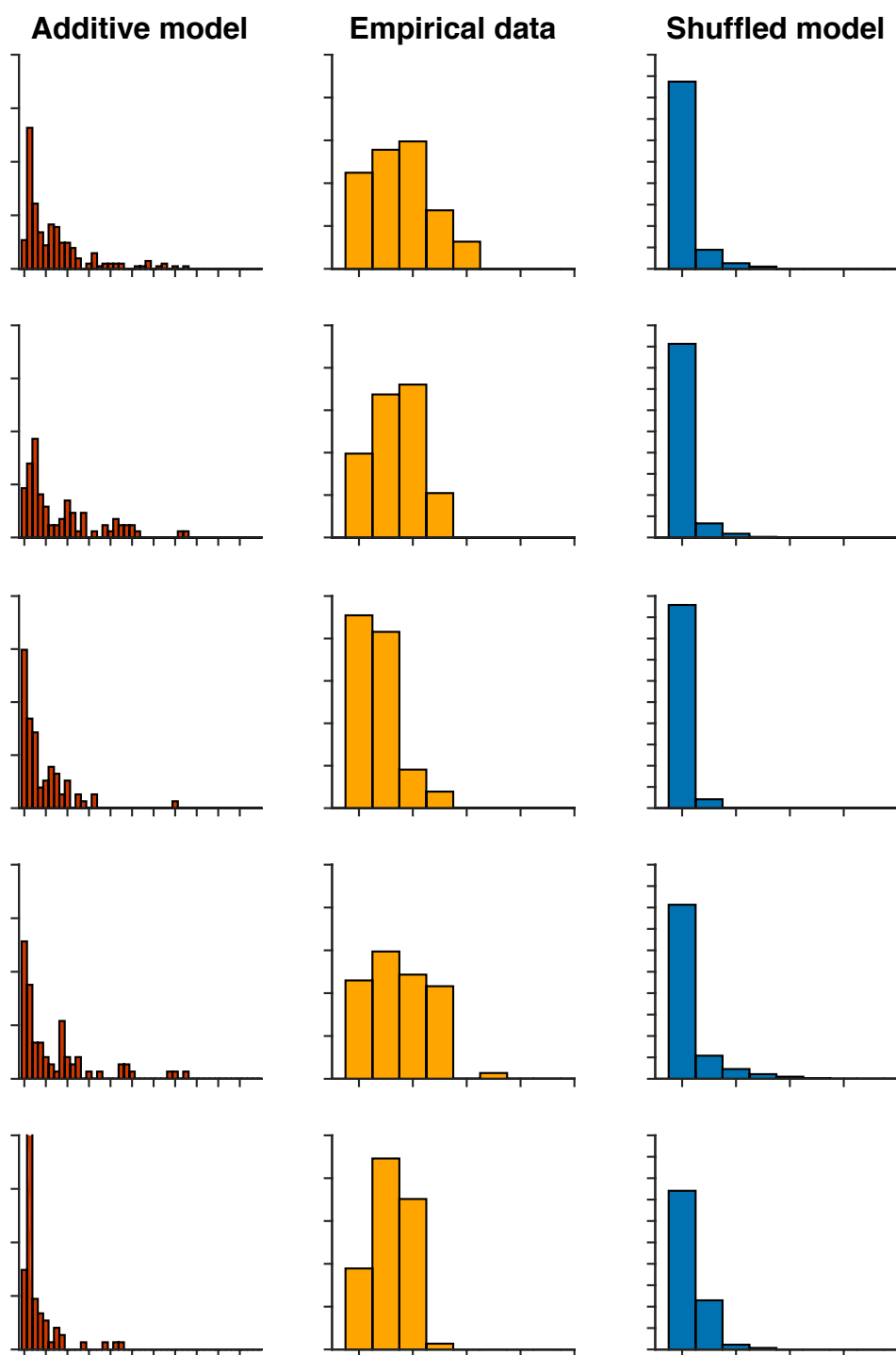
!



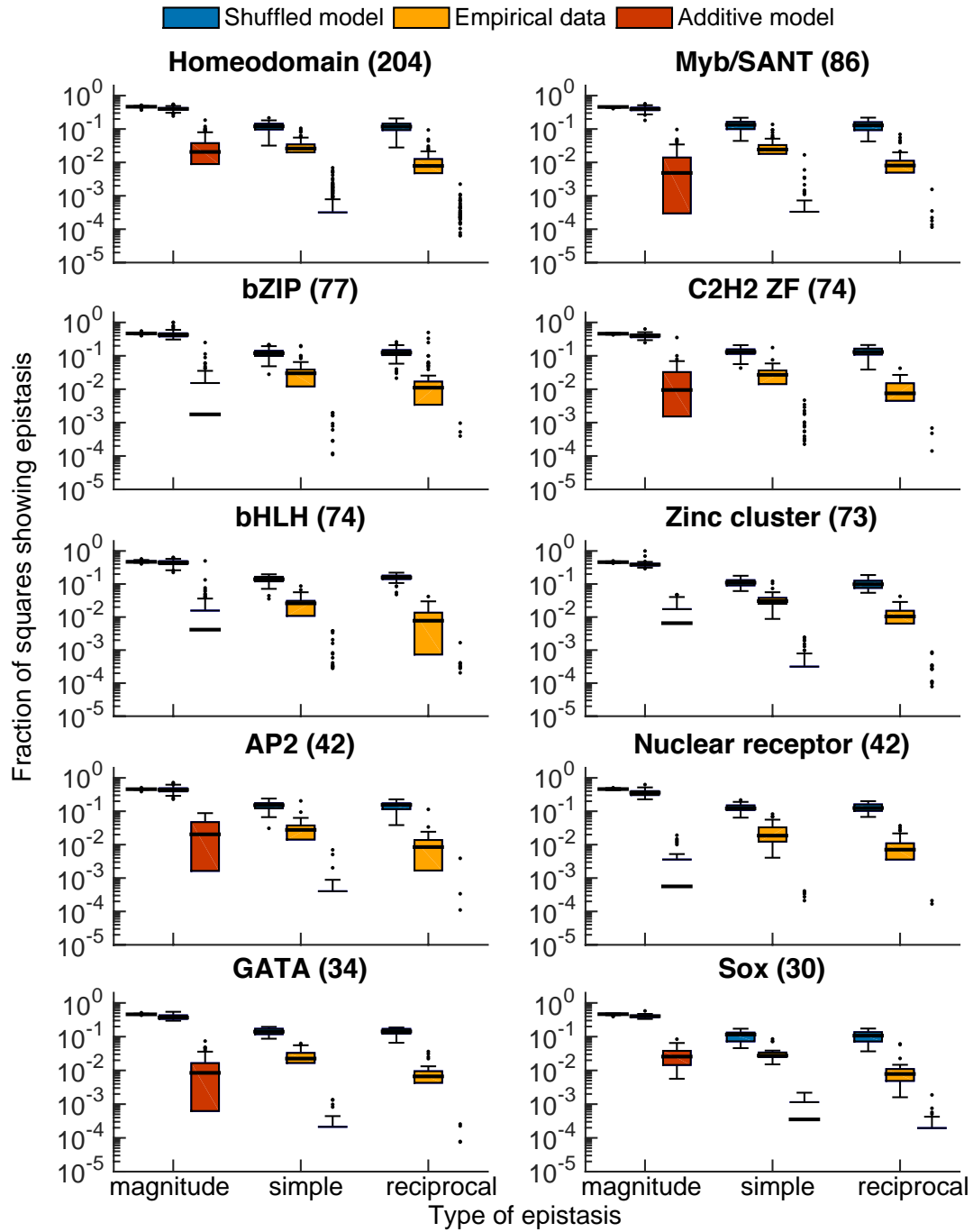
Supplementary Figure 19. Accessibility of the highest-affinity site increases as the binding affinity threshold increases.



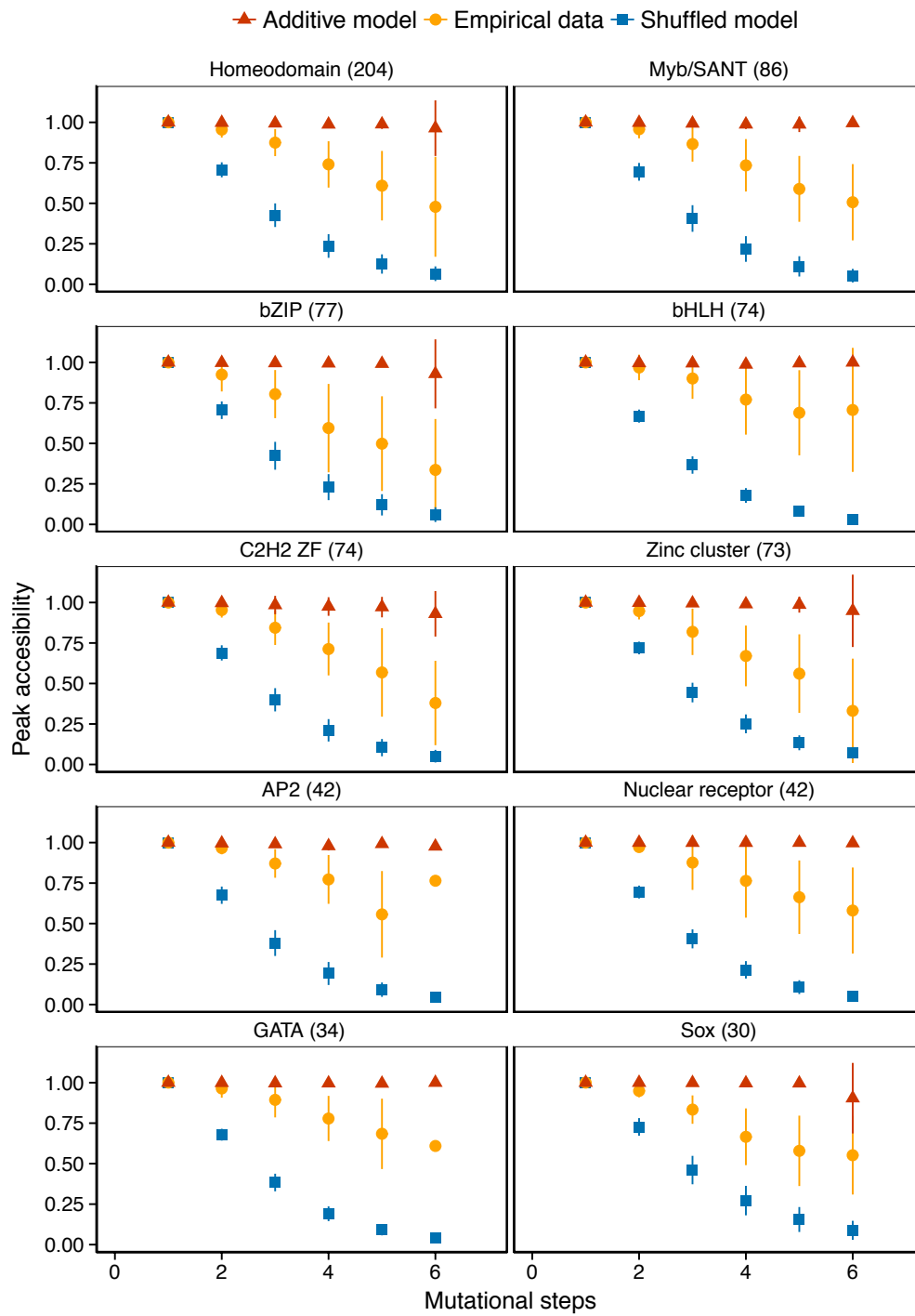
Supplementary Figure 20. The number of peaks is intermediate to that of the additive and shuffled models across DNA binding domains.



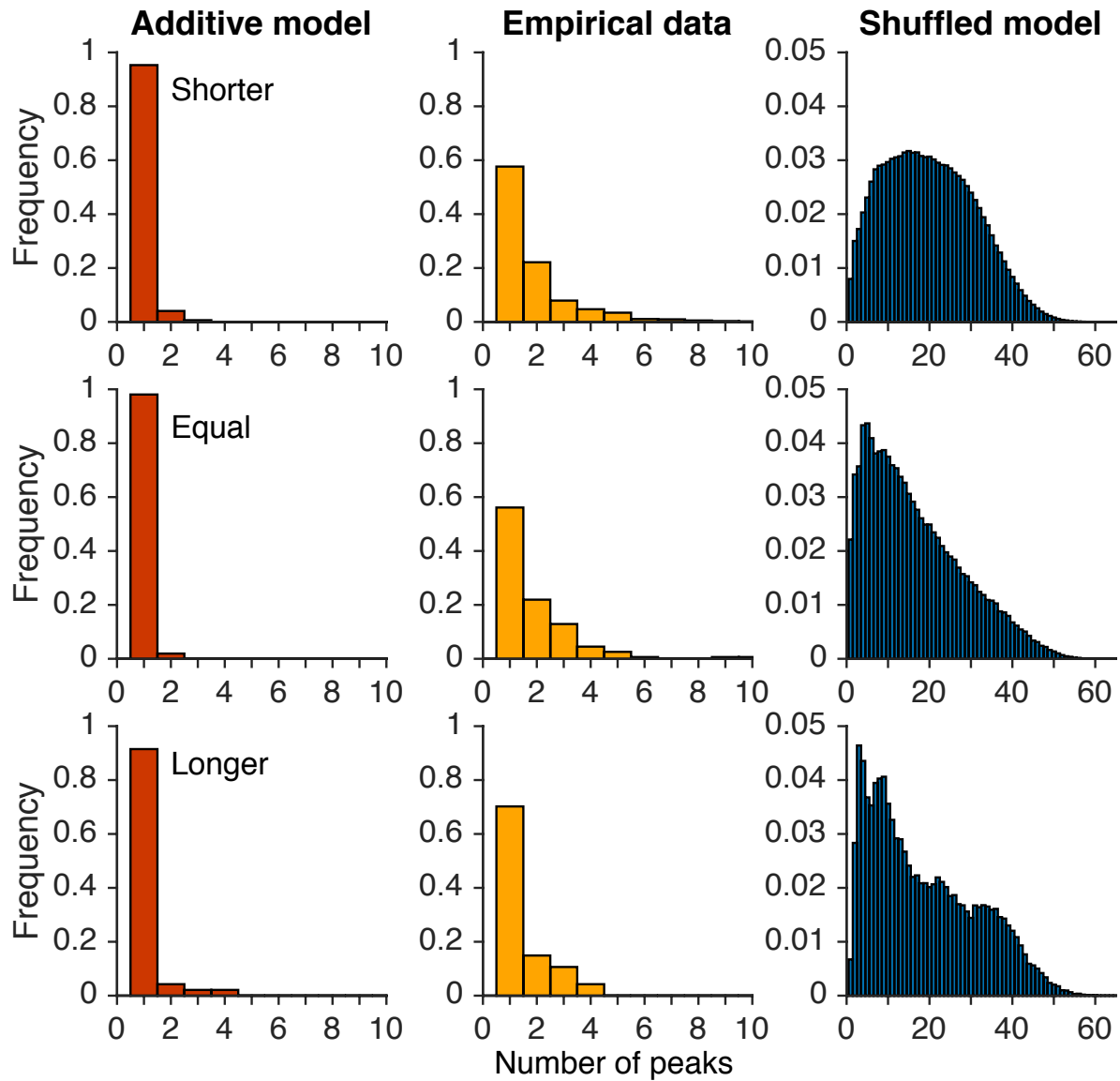
Supplementary Figure 21. The size of the global peak is intermediate to that of the additive and shuffled models across DNA binding domains.



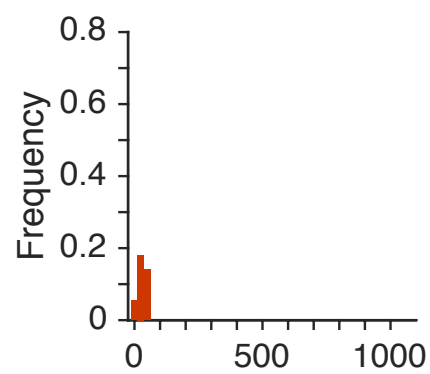
Supplementary Figure 22. Epistasis in the empirical data is intermediate to that of the additive and shuffled models, irrespective of DNA binding domain.



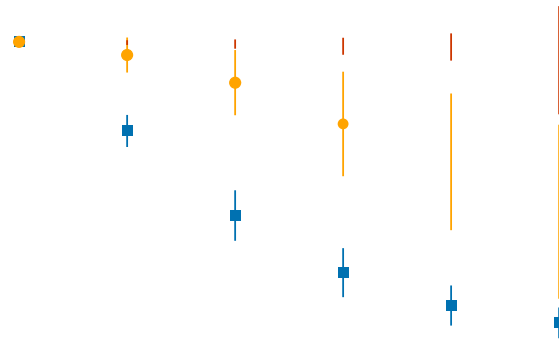
Supplementary Figure 23. Peak accessibility in the empirical data is intermediate to that of the additive and shuffled models, irrespective of DNA binding domain.



Supplementary Figure 24. The number of peaks in the empirical data is intermediate to that of the additive and shuffled models for TFs that bind sequences that are shorter or longer than eight nucleotides.



Supplementary Figure 25. The global peak breadth in the empirical data is intermediate to that of the additive and shuffled models for TFs that bind sequences that are shorter or longer than eight nucleotides.



Supplementary Figure 27. Peak accessibility in the empirical data is intermediate to that of the additive and shuffled models for TFs that bind sequences that are shorter or longer than eight nucleotides.

A

B

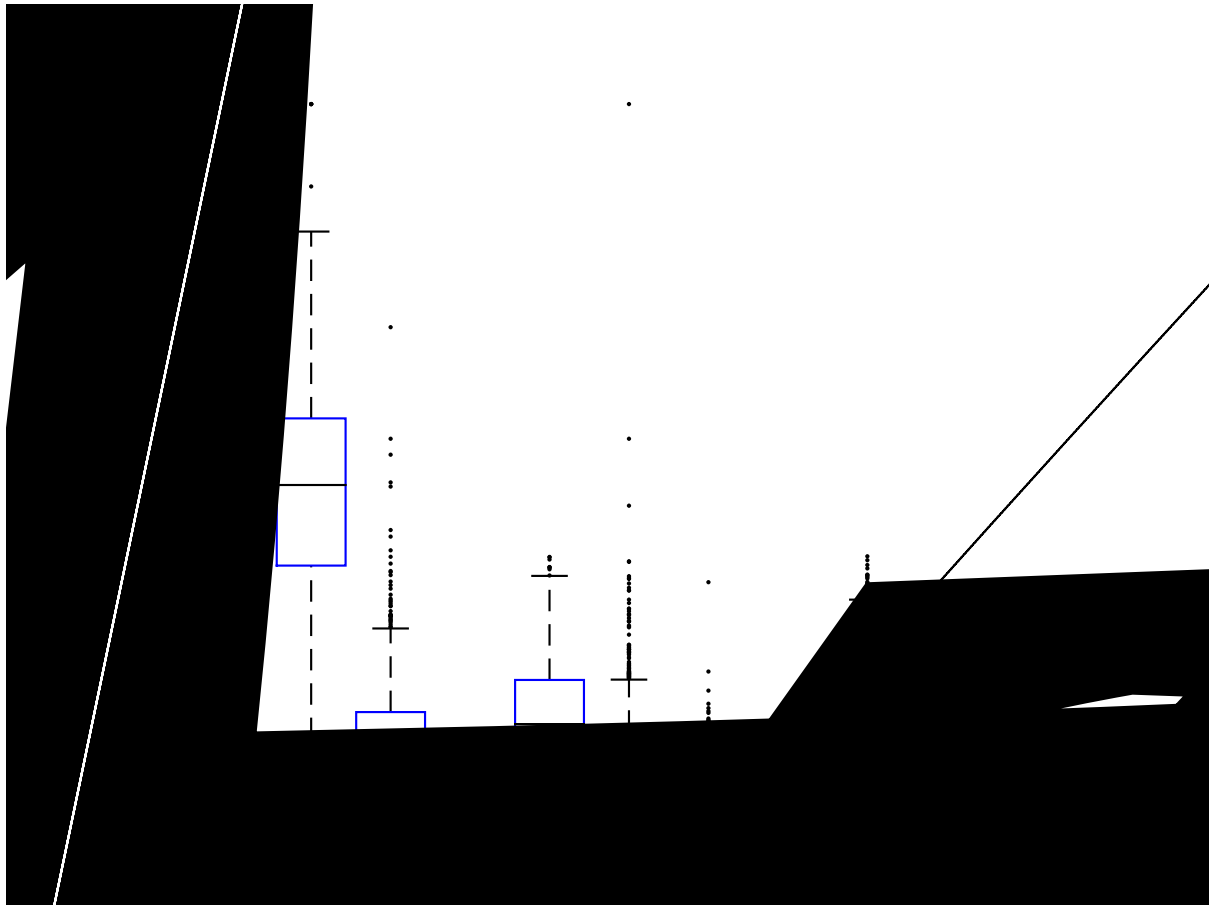
C



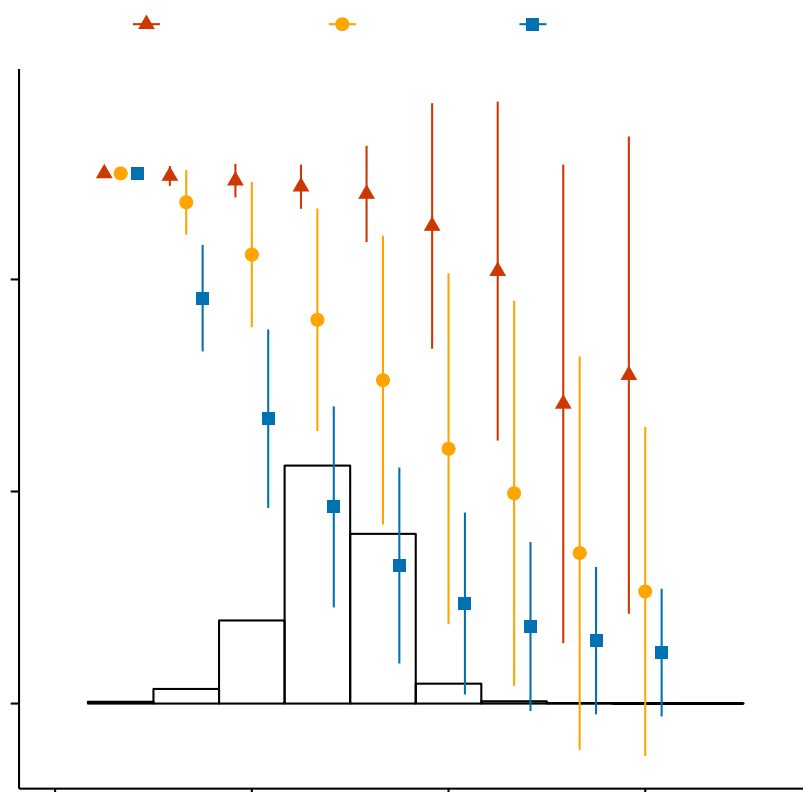
Supplementary Figure 28. The number of peaks in the empirical data is intermediate to that of the additive and shuffled models when using Z-scores as a proxy for binding affinity

A

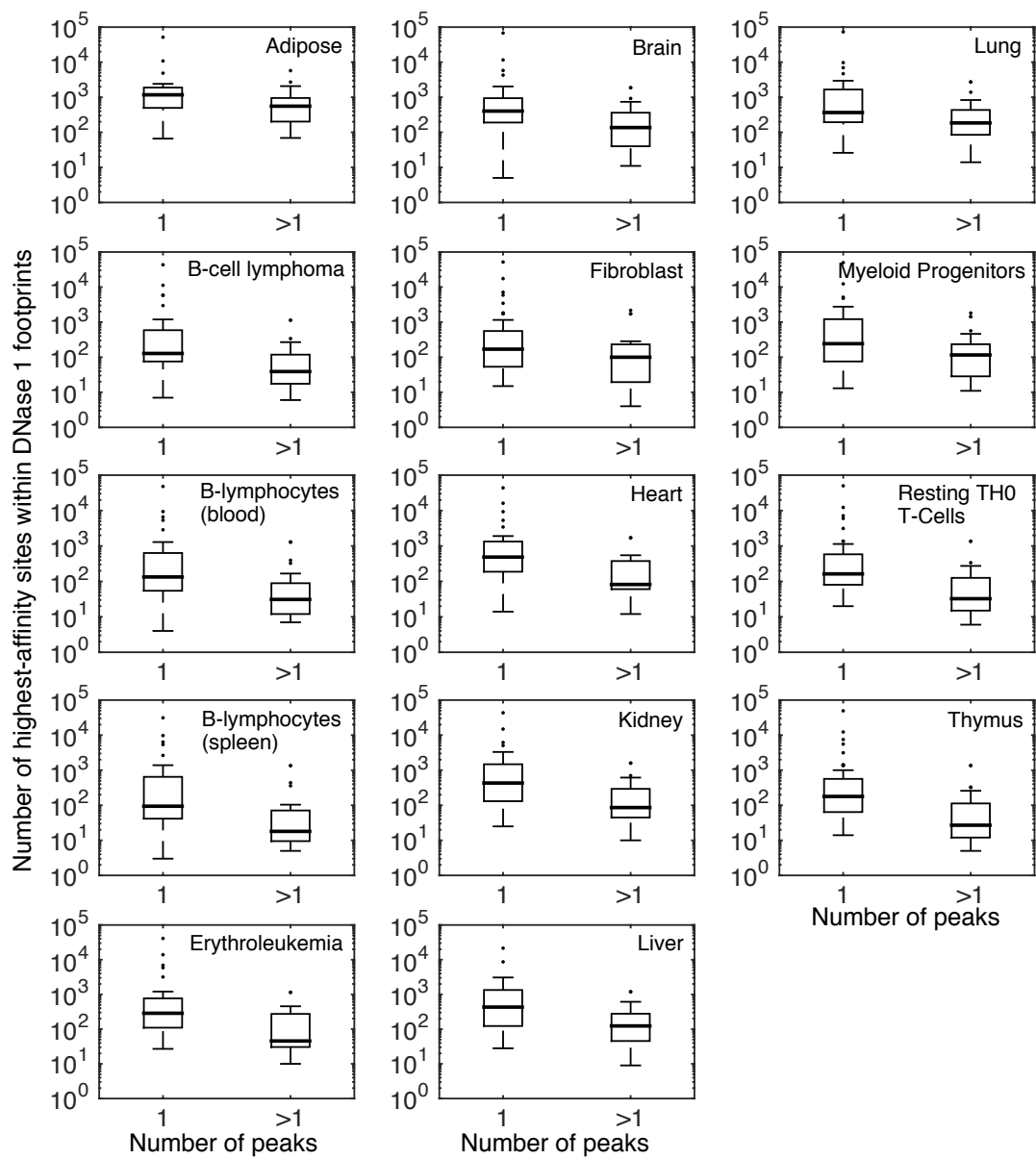
B



Supplementary Figure 29. Epistasis in adaptive landscapes of transcription factor binding affinity when using Z-scores as a proxy for binding affinity.

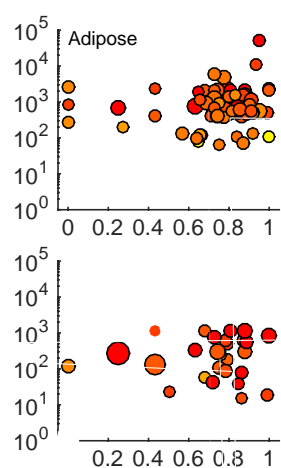


Supplementary Figure 30. Peak accessibility when using Z-scores as proxy for binding affinity.



Supplementary Figure 31. *In vivo* binding site abundance is higher for TFs with single-peaked landscapes than for TFs with multi-peaked landscapes.

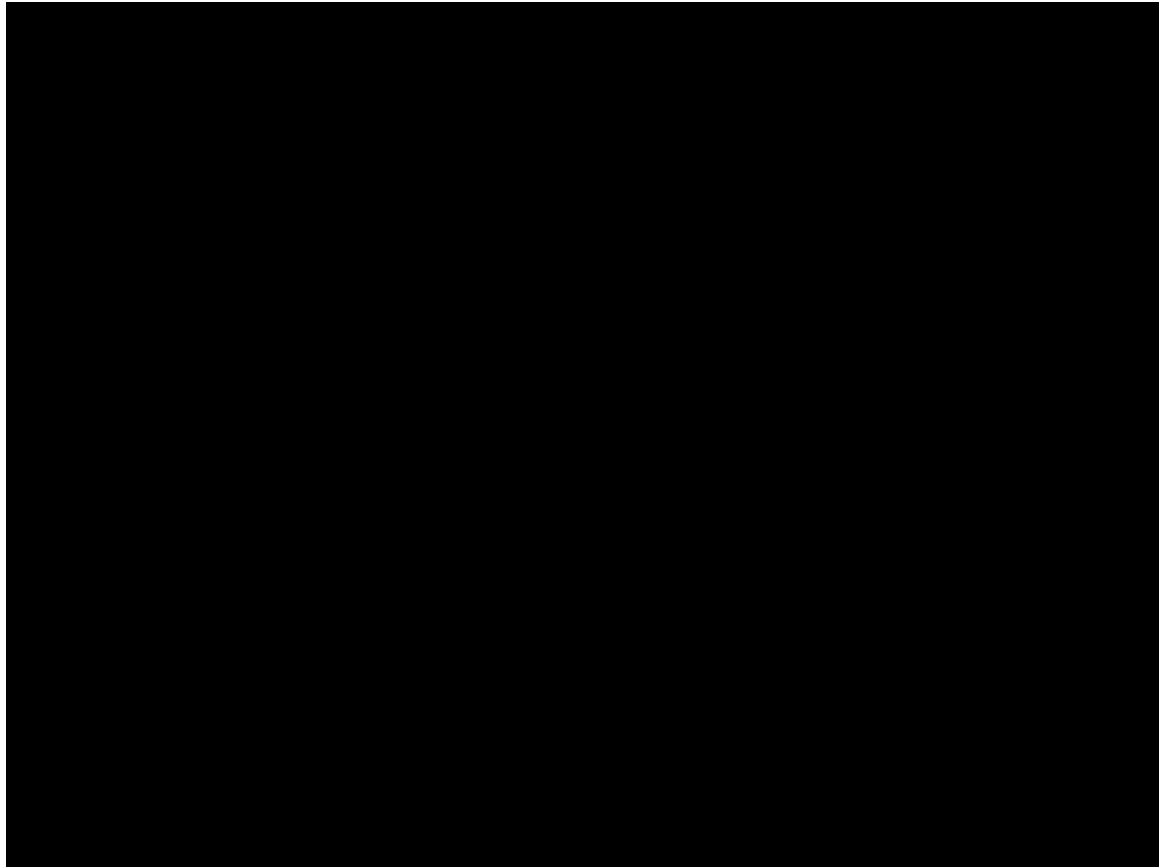
M. musculus



32. *In vivo* binding site abundance correlates with peak accessibility.

us

P



Supplementary Figure 33. Our measures of epistasis based only on bound sequences are conservative.

2. Supplementary tables

Supplementary Table 1.

Supplementary Table 2.

Supplementary Table 3.

Supplementary Table 4.

Supplementary Table 5.

R

Supplementary Table 6.

Supplementary Table 7.

3. Supplementary results

3.1. Summary statistics of genotype networks

dominant

3.2. Global peaks are usually organized into broad plateaus

#

P-

P-

#

E

Z

3.3. Why epistasis occasionally appears in the additive null model

AATTTTAA

AAATATAA

AATTTTAA

AATTTTAA

TATATATA

ATATATAT

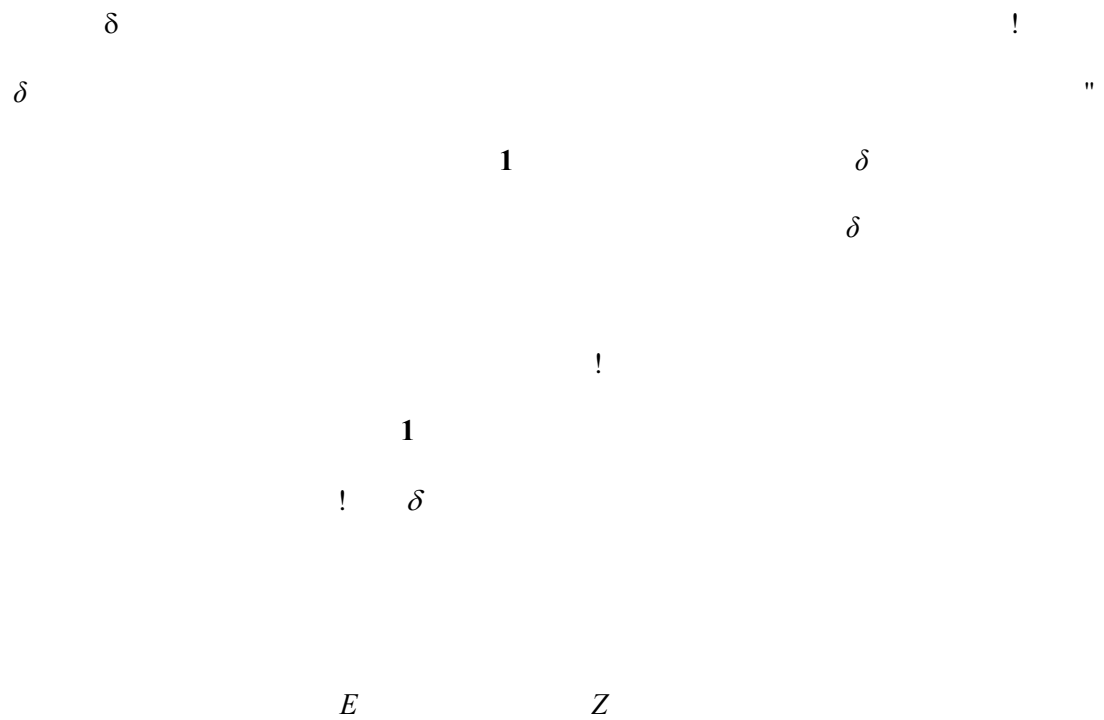
3.4. Sign epistasis preferentially occurs among nucleotides that are near one another in the binding site

3.5. Peak accessibility decreases when unbound sequences are included

!

! &

3.6. Sensitivity analyses



3.6.1. Our observations are insensitive to broadly varying thresholds for noise filtering



δ

!

δ

δ

δ

δ

δ

δ

δ

δ

3.6.2. Our observations are insensitive to broadly varying affinity thresholds for delineating bound from unbound sequences

E

!

!

!

!

E

!

!

!

!

!

!

!

!

!

!

!

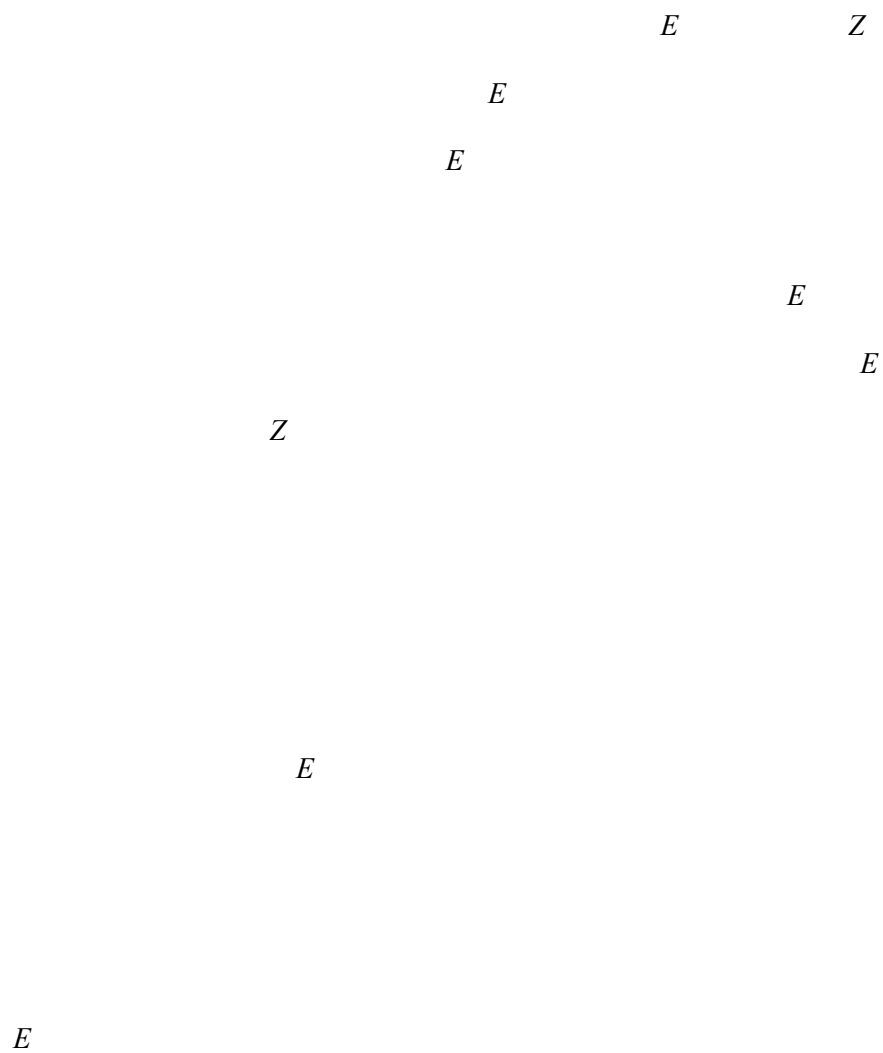
!

!

3.6.3. Our observations are consistent across DNA binding domains

3.6.4. Our observations are consistent across TFs that bind shorter or longer sequences than eight nucleotides

3.6.5. Peak breadth is sensitive to the use of *E*-scores as a quantitative phenotype



3.7. The *in vivo* relationship between landscape navigability and the abundance of binding sites is not driven by binding affinity or by information content

in vivo

P

in vivo

3.8. Our measures of epistasis for bound sequences are conservative

!

4. Supplementary discussion

Escherichia coli

in vitro

in vitro

lac

4.1. Caveats

in vitro

in

vivo

in vivo

in vitro

in vivo

in vivo

! δ

in vivo

5. Supplementary references