SUPPLEMENTARY MATERIAL

Analysis of Genetic Variation Indicates DNA Shape Involvement in Purifying Selection

Xiaofei Wang[1,#], Tianyin Zhou[1,#,&,*], Zeba Wunderlich[2], Matthew T. Maurano[3], Angela H. DePace[4], Sergey V. Nuzhdin[1], and Remo Rohs[1,5,*]

[1]Molecular and Computational Biology Program, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA

[2]Department of Developmental and Cell Biology, University of California Irvine, Irvine, CA 92697, USA

[3]Institute for Systems Genetics, New York University Medical Center, New York, NY 10016, USA
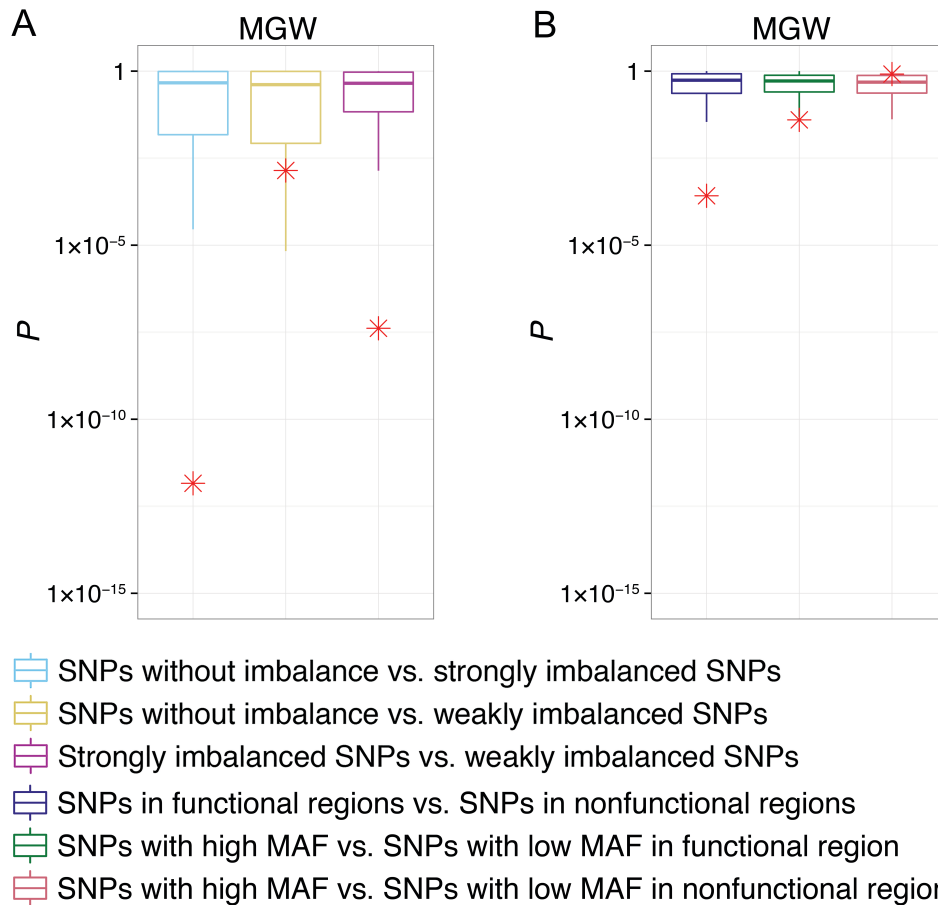
[4]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

[5]Departments of Chemistry, Physics & Astronomy, and Computer Science, University of Southern California, Los Angeles, CA 90089, USA

[&]Present address: Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA

[#]These authors contributed equally to this work.

*Corresponding authors: tianyinz@alumni.usc.edu (T.Z.) and rohs@usc.edu (R.R.)

A MGW   **B** MGW

☐ SNPs without imbalance vs. strongly imbalanced SNPs
☐ SNPs without imbalance vs. weakly imbalanced SNPs
☐ Strongly imbalanced SNPs vs. weakly imbalanced SNPs
☐ SNPs in functional regions vs. SNPs in nonfunctional regions
☐ SNPs with high MAF vs. SNPs with low MAF in functional region
☐ SNPs with high MAF vs. SNPs with low MAF in nonfunctional region

Supplementary Fig. S1. Boxplots for Mann-Whitney *P*-values in shuffling tests.
All of the Mann-Whitney tests shown in **Fig. 3** and **Fig. 5** were repeated by using 1,000 arbitrarily shuffled MGW predictions. Base-10 logarithms of the corresponding 1,000 Mann-Whitney *P*-values were plotted with box plots. Mann-Whitney *P*-values obtained by using DNAshape-derived MGW patterns are marked with a red asterisk. Tests for A) human and B) *Drosophila* data.
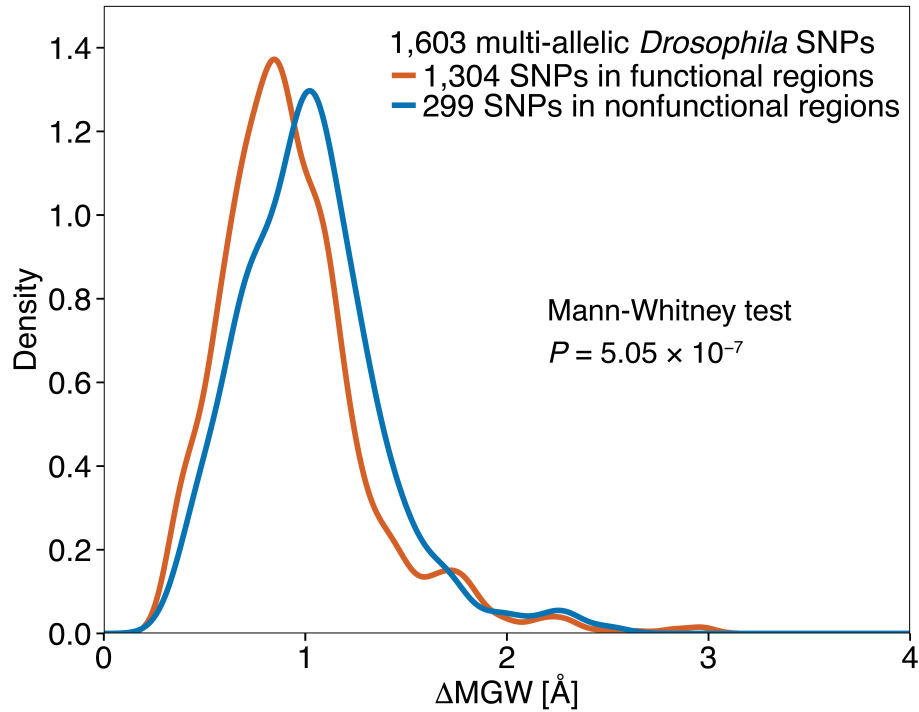
Supplementary Fig. S2. Distribution of ΔMGW values for SNPs in functional and nonfunctional regions, obtained by using DNAshape-derived MGW in multi-allelic analysis. Compared to the distribution for functional regions (red plot), the distribution for nonfunctional regions (blue plot) was significantly shifted towards the right. This trend indicates that the change in MGW values induced by SNPs in nonfunctional regions was greater than the change induced by SNPs in functional regions. Sample sizes for all groups are listed in the legend.

Supplementary Table 1. TFs used to determine functional regions.

TF	Symbol	DNA binding domain	Regulatory class
Bicoid	bcd	homeodomain	A/P Maternal
Caudal	cad	homeodomain	A/P Maternal
forkhead	fkh	forkhead domain	A/P Zygo gap/term
giant	gt	b-zip domain	A/P Zygo gap/term
huckebein	hkb	TFIIIA Zn finger	A/P Zygo gap/term
hunchback	hb	TFIIIA Zn finger	A/P Zygo gap/term
knirps	kni	receptor Zn finger	A/P Zygo gap/term
knirps like	knil	receptor Zn finger	A/P Zygo gap/term
Kruppel	Kr	TFIIIA Zn finger	A/P Zygo gap/term
orthodenticle	oc	homeodomain	A/P Zygo-gap/term
tailless	tll	receptor Zn finger	A/P Zygo-gap/term
Dichaete	D	HMG/SOX class	A/P Zygo-pair rule
even skipped	eve	homeodomain	A/P Zygo-pair rule
ftz	ftz	homeodomain	A/P Zygo-pair rule
hairy	h	bHLH	A/P Zygo-pair rule
odd paired	opa	TFIIIA Zn finger	A/P Zygo-pair rule
paired	prd	homeo & Prd domain	A/P Zygo-pair rule
runt	run	runt domain	A/P Zygo-pair rule
sloppy paired 1	slp1	forkhead domain	A/P Zygo-pair rule
sloppy paired 2	slp2	forkhead domain	A/P Zygo-pair rule
Stat92E	Stat92E	Stat domain	A/P Zygo-pair rule
sis of bowl & odd	sob	Zn finger	A/P Zygo-pair rule
odd-skipped	odd	Zn finger	A/P Zygo-pair rule
bowl	bowl	Zn finger	A/P Zygo-pair rule
Daughterless	da	bHLH	D/V Maternal
dorsal	dl	NFkB/rel	D/V Maternal
brinker	brk	novel	D/V Zygo
Mad	Mad	SMAD-MH1	D/V Zygo
Medea	Med	SMAD-MH1	D/V Zygo
schnurri	shn	TFIIIA Zn finger	D/V Zygo
snail	sna	TFIIIA Zn finger	D/V Zygo
twist	twi	bHLH	D/V Zygo
zerknult 1	zen 1	homeodomain	D/V Zygo
zerknult 2	zen 2	homeodomain	D/V Zygo