SCIENCE CHINA
Life Sciences

CUSBEA ARTICLE SERIES
• RESEARCH PAPER •

September 2014 Vol.57 No.9: 876–888
doi: 10.1007/s11427-014-4704-4

Scoring the collective effects of SNPs: association of minor alleles with complex traits in model organisms

YUAN DeJian[1†], ZHU ZuoBin[1†], TAN XiaoHua[1], LIANG Jie[1], ZENG Chen[1], ZHANG JieGen[2], CHEN Jun[2], MA Long[1], DOGAN Ayca[3], BROCKMANN Gudrun[3], GOLDMANN Oliver[4], MEDINA Eva[4], RICE Amanda D.[5], MOYER Richard W.[5], MAN Xian[1], YI Ke[1], LI YanKe[1], LU Qing[1], HUANG YiMin[1] & HUANG Shi[1*]

[1]*State Key Laboratory of Medical Genetics, Central South University, Changsha 410078, China;*
[2]*High Performance Computing Center, Modern Educational Technology Center, New Campus, Central South University, Changsha 410083, China;*
[3]*Department of Crop and Animal Sciences, Faculty of Agriculture and Horticulture, Humboldt-Universität zu Berlin, Invalidenstraße 42, Berlin 10115, Germany;*
[4]*Infection Immunology Group, Helmholtz Centre for Infection Research, Inhoffenstraße 7, Braunschweig 38124, Germany;*
[5]*Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL 32610, USA*

Received April 15, 2014; accepted June 6, 2014; published online August 6, 2014

It has long been assumed that most parts of a genome and most genetic variations or SNPs are non-functional with regard to reproductive fitness. However, the collective effects of SNPs have yet to be examined by experimental science. We here developed a novel approach to examine the relationship between traits and the total amount of SNPs in panels of genetic reference populations. We identified the minor alleles (MAs) in each panel and the MA content (MAC) that each inbred strain carried for a set of SNPs with genotypes determined in these panels. MAC was nearly linearly linked to quantitative variations in numerous traits in model organisms, including life span, tumor susceptibility, learning and memory, sensitivity to alcohol and anti-psychotic drugs, and two correlated traits poor reproductive fitness and strong immunity. These results suggest that the collective effects of SNPs are functional and do affect reproductive fitness.

collective effects, complex traits, minor alleles, SNPs, recombinant inbred lines, minor allele content (MAC)

Citation: Yuan DJ, Zhu ZB, Tan XH, Liang J, Zeng C, Zhang JG, Chen J, Ma L, Dogan A, Brockmann G, Goldmann O, Media E, Rice AD, Moyer RW, Man X, Yi K, Li YK, Lu Q, Huang YM, Huang S. Scoring the collective effects of SNPs: association of minor alleles with complex traits in model organisms. Sci China Life Sci, 2014, 57: 876–888, doi: 10.1007/s11427-014-4704-4

Past studies of complex traits have met with great success in identifying a number of significant genetic variants, although such variants usually account for only a small fraction of the total trait variation and their functional roles typically remain unclear [1–8]. The focus on searching for a few major effect variants is under the null hypothesis in the field of population genetics that the majority of genetic variations are neutral. This hypothesis, however, is at best incomplete as the question of what determines genetic diversity has long remained unsolved [9]. The neutral hypothesis in fact was mistaken right from its inception and never really explained the first and most remarkable result in molecular evolution, the genetic equidistance result [10]. While the assumption of neutrality has often passed tests by sequence-alignment based informatics approaches, such methods usually have their own set of assumptions, includ-

†Contributed equally to this work
*Corresponding author (email: huangshi@sklmg.edu.cn)

ing assuming certain DNAs to be neutral such as synonymous (syn) sites and transposon-element derived sequences, and are therefore not truly conclusive tests free of neutral or uncertain assumptions [11–13]. In contrast, experimental science found little evidence for the neutral assumption. A majority of the noncoding parts of the human genome are transcribed [14], and numerous experimental researchers have now recognized an important functional role for these non-coding RNAs (for a review see [13,15]). Furthermore, we have recently proposed a more complete theory to supersede the neutral hypothesis and a key prediction of our theory is that the majority of a genome are functional [16,17].

SNPs typically have just two alleles in a population and the minor allele (MA) has frequency (MAF) <0.5. Unlike a rare variant, however, a common MA typically has MAF >0.1. In theory, neutral could have two contrasting meanings. A minor allele could be either neutral because it is non-functional and under no Darwinian natural selection or only seemingly neutral or nearly neutral or slightly deleterious because it is both beneficial and deleterious and under both positive and negative selection. Such opposite meanings of neutral may not be easily distinguishable by popular statistical tests for detecting selection. MAs were often found to be disease risk alleles [18]. However, most studies only looked at a few SNPs. The question whether the collective effects of common SNPs' MAs are neutral has yet to be addressed.

While too little genetic variations are known to hurt adaptive capacities, it is much less appreciated whether too much may exceed an organism's maximum level of tolerable disorder or entropy, given that mutations are after all random and disorderly in origin. Entropy is generally defined as the logarithm of the number of ways the microstate can rearrange itself without grossly affecting the macrostate. An organism as a macrostate can accommodate certain limited amounts of microstates at the level of DNA rearrangements or variations. Entropy is thus related to normal genetic diversity as measured by the number of common SNPs. Genetic diversity at the maximum or optimum tolerable level would be adaptive or beneficial but would be deleterious if it is either above or below that level. It is therefore a priori expected that genetic diversity should be under both positive and negative selection and always at optimum level if time is long enough for equilibrium to be reached.

We here used a novel method to test whether the total amount of SNPs carried by an individual is at an optimum level. We made use of multiple panels of genetic reference populations or recombinant inbred lines (RILs) that provide a powerful means to study the genetic basis of complex phenotypes [19–26]. The RIL panels are derived from breeding of parental strains differing in phenotypes and genotypes. The F1 and F2 or up to F10 progenies are intercrossed to maximize random recombination and hence allelic diversity in the offspring, which were then randomly selected for inbreeding up to 20 generations to generate the final panel of RILs homozygous for almost all variants or SNPs. During the random mating and subsequent inbreeding process, there are ample opportunities for neutral variants to drift and for non-neutral variants to be selected. Immunity against pathogens is essential for survival and depends on allelic diversity, which would positively select for enrichment of variants. On the other hand, individuals may die or be aborted before birth due to deleterious variants. While the population size of a RIL panel is small, the actual size of the offspring population of the original parents is much greater and includes many that died because of negative selection.

If a trait is determined by multiple loci and robust to minor perturbations, one may expect that the trait may be genetically affected in two mutually non-exclusive ways. One is a major effect mutation in one of these loci that alters a component of a multi-component pathway. Alternatively, it may take a large amount of mutations to harm the trait, while such mutations individually or in small amounts may have few discernable effects or even beneficial effects. Furthermore, variations in the amount of mutations may account for quantitative variations commonly found in complex traits. For example, the more the variants the better the adaptive immunity up to a point when too much variants may start to hurt other traits or be cancer prone.

For any given panel of RILs, most SNPs would show MAs that are carried by less than half of the strains in the panel and the strains would differ in the contents of MAs that each carries. We defined "MA contents (MAC)" as the number of MAs in an individual divided by the number of SNPs scanned. Different from MAF, MAC is an individual measure. One predicts that strains with higher MAC should be similar to those with lower if the neutrality assumption is true. We here tested this by performing new trait analysis experiments as well as by using the large collection of data accumulated in the past several decades for genetic reference populations.

1 Materials and methods

1.1 MAC calculation and statistical methods

SNPs datasets for the genetic reference populations were obtained from the literature and public databases. All analyses were done with autosomal SNPs. Phenotypes data were from the literature and GeneNetwork. The number of strains in each RIL panel is given in Table S1 in Supporting Information.

The allele frequency of each SNP in a RIL panel or a control cohort was calculated by SNP Tools for Microsoft Excel and PLINK [27,28]. We excluded non-informative SNPs from MAC calculation that have frequency 0 for one of the alleles in both cases and controls or in a RIL panel or have frequency 0.5 in controls or a RIL panel.

A novel software Nucleotide Diversity 1 (ND1) was developed to count the number of mismatches between individual genotypes as well as the number of heterozygous (Het) SNPs of each individual. Genetic distance is defined as the number of mismatches divided by the number of SNPs studied. Pairwise genetic distance (PGD) of a population panel is defined as the average of all non-repetitive pair within a population. The number of genotype mismatches is counted as follows. For homozygous (Hom) vs. Hom mismatch, a difference of 1 was scored. Hom vs. Het was scored 0.5. Het vs. Het was scored 0.5 since half of such cases are expected to be A/B vs. A/B with a difference of 0 whereas the other half are expected to be A/B vs. B/A with a difference of 1 (A/B means the generic two-allele genotype of a SNP with A or B representing one of the four nucleotides and the forward slash separating the two alleles). On a genome wide scale, the number of A/B vs. A/B match due to IBD (identical by descent) is expected to be similar to the number of A/B vs. B/A mismatch. We verified this approach by comparing the PGD in X chromosome for CEU females vs. CEU males using HapMap SNP data and found them to be similar as expected. In contrast, a software based on IBS (identical by status) such as PEAS that scores Het vs. Het as 0 showed the males to have much greater PGD in X than females [29]. For the missing genotypes N/N, N/N vs. Hom was scored as 0 and N/N vs. Het as 0.5.

The following are the detailed steps in calculating MAC and the average distance to the MA set:

(i) Obtain SNP genotype dataset of a RIL panel or a human population panel.

(ii) Calculate allele frequency of each SNP's two alleles in the panel and assign MA status to the allele with the smaller frequency.

(iii) Use the ND1 software to count the number of mismatched SNPs between a sample and the set of MAs assigned in step two, and obtain the average mismatch # per sample. For this counting, the MA set has MAs in homozygous form.

(iv) Number of MAs in a sample=# SNPs free of N/N–# mismatch with MAs.

(v) MAC of each sample = # MAs/# SNPs free of N/N.

(vi) Correction for N/N genotypes in counting the average # mismatch from step 3: corrected average mismatch #=pre-correction #×#SNPs/(#SNPs–# N/N).

(vii) Average distance to the MA set per sample=corrected # mismatch/# SNPs.

The following exemplary data table illustrates the above procedures (Table 1). Shown in the table is a population panel with five samples listed in row 1 with each genotyped for a total of three SNPs as listed in column 1. The genotypes of each SNP of each sample are in columns 3 to 7, with sample 5 having one N/N missing genotype. As an example, for SNP rs12345 in row 2, the A allele has a frequency of 3/10=0.3 and is hence assigned as the MA. The set of MAs thus identified for all three SNPs is listed in column 2 that has the MAs in hom form. The MA set is therefore equivalent to an imagined sample who is homozygous for the MAs of all the SNPs studied (MAC=1). Next, the number of mismatches between each sample and the MA set is determined using ND1 software as shown in row 5, with the average per sample shown in row 8. The number of MAs of each sample is shown in row 6, with MAC value in row 7. The corrected average mismatch number is shown in row 9. The average distance to the MA set is shown in the last row.

The correlation between genotype and phenotype was analyzed by linear and multivariate regression analysis using GraphPad's statistics software Prism 5 and InStat3 and the software Significance Analysis of Microarrays. For multivariate regression analysis of the 3664 traits in BXD panel, most traits were unsuitable for analysis because of missing data. After removing these, there were 21 traits left, from which 13 were filtered out because of non-independent nature based on multivariate analysis. The remaining nine traits were then analyzed by multivariate regression using InStat3. Other statistical methods used include Student's *t* test, two tailed, chi-square test, two tailed, linear and multivariate regression, and Pearson/Spearman correlation analysis. Since the sample size is often large in our data sets, the Whitney-Mann test gave similar results as the *t* test and only *t* test data were presented.

Table 1 Exemplary data table for illustration of MAC calculation

SNP rsID	MA set	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
rs12345	A/A	G/A	G/G	G/G	A/A	G/G
rs12346	T/T	T/T	G/T	G/G	G/G	G/G
rs12347	C/C	C/T	T/T	T/T	T/T	N/N
# mismatch with MA set		1	2.5	3	2	2
Number MAs	3	2	0.5	0	1	0
MAC	1	0.67	0.17	0	0.33	0
Ave # mismatch with MA set: 2.1						
Corrected Ave # mismatch: 2.1×3/(3−1/5)=2.25						
Ave distance to the MA set: 2.25/3=0.75						

1.2 Identification of trait specific set of MAs

A number of traits showed correlation with other traits (Table 2). We examined four traits among these, blood ethanol concentration (BEC, trait 3), its strongly correlated trait resistin level (trait 11), and its two non-correlated traits pain response (trait 4) and open field rearing behavior (trait 5) (Table 2). We picked out 61 strains with BEC data and wanted to select a subset of SNPs that could separate the BEC trait and its related traits from the non-related traits better than what the original random set of 51K SNPs had done.

Each SNP genotype in an Excel data matrix like the above exemplary data table was converted into a MA score of 0, 0.5, or 1, depending on its MA content. If a genotype has no MA, it is scored 0; if it is het, it is 0.5; if it is hom for the MA, it is 1. So, the above exemplary data Table 1 can be converted to the MA score table (Table 3).

Next, the 61 RIL strains were sorted based on their BEC value and divided into three groups of 20 strains each, with group 1 lowest and group 3 highest in BEC value. For each SNP, the average MA score in group 1 and 3 was deter-

Table 2 Correlations among selected traits linked with higher MAC in BXD mice

GN ID[a]	Sample#	Spearman		Trait	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Trait description
		r	*P*																					
10145	25	0.51	0.010	1						**		*[b]											*	Maxi-threshold to ethanol induced ataxia
10169	25	0.61	0.001	2			*	*				*												Methamphet. induced temp. change
11453	61	0.49	<0.0001	3								*		*	***					**				Blood ethanol concentration for males
11307	60	0.44	0.001	4					**								**	*	*					Hargreaves' test for males
11672	57	0.41	0.001	5															*				*	Open field rearing activity from 10–15 min
10022	20	0.61	0.005	6							**		**								**			Saccharin preference versus water ratio
10493	25	0.46	0.020	7									*											Cocaine induced difference in locomotion
10301	23	0.38	0.072	8															*			*		Cocaine, nose pokes in hole board
10494	24	0.39	0.061	9																*				Ethanol induced difference in locomotion
10917	15	0.53	0.041	10												**								Anxiety, transitions between light and dark
14220	28	0.52	0.005	11													*			**		*		Resistin level after high fat diet
12540	22	0.50	0.019	12																				Transferrin saturation fed 3 ppm iron diet
11725	57	0.57	0.002	13																			*	Gain in weight between 8 and 9 weeks
12886	16	0.61	0.012	14															*					Oxygen consumption males
12568	38	0.63	<0.0001	15																		**		Deoxycorticosterone in cerebral cortex
12852	16	0.60	0.015	16																		*		Food intake of 13-week old females
12554	24	0.39	0.056	17																				Depression assay, duration of immobility
14226	28	0.46	0.015	18																				IL17 level after high fat diet
12396	64	0.30	0.016	19																				Time in open quadr. in elevated 0 maze

a) GeneNetwork identification number. b) Symbols *, **, and *** represent P<0.05, 0.01, 0.001 respectively, from Spearman analysis.

Table 3 Exemplary data table for MA score calculation

SNP rsID	MA set	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
rs12345	A/A	0.5	0	0	1	0
rs12346	T/T	1	0.5	0	0	0
rs12347	C/C	0.5	0	0	0	N/N

mined and the significance of the difference between them was determined by *t* test. Each SNP was thus given a *P* value from the *t* test. Next the whole set of SNPs was ranked by the *P* values and divided into 10 groups based on the *P* values. These 10 groups were designated as *P*<0.05, 0.05–0.11, 0.11–0.2, <0.2, 0.2–0.5, <0.5, 0.5–0.75, <0.75, 0.75–1, or 0–1. For example, the SNPs in the <0.05 group all have *P* <0.05; the SNPs in the 0.05–0.11 group all have *P* value between 0.05 and 0.11. Finally, each group of SNPs was used to calculate MAC of each RIL strain. The same strain typically had different MAC values for different groups of SNPs.

From the 61 RIL strains, we identified 24 strains that have phenotype data for all four traits concerned. Using each of the above 10 groups SNPs, we calculated the MAC of each strain. So each strain has 10 different MAC values corresponding to the 10 groups of SNPs. We then tested for correlation between MAC and the four traits of concern in the 24 strains to see which of the 10 groups of SNPs gave the best result. A strong correlation with BEC and its related traits but a poor one with unrelated traits was considered a good specificity profile for a group of SNPs. The group designated as <0.75 was found to have the best specificity profile and has 30336 SNPs.

1.3 Animal experiments

The study performed animal experiments and the animals' care was in accordance with institutional guidelines. The Institutional Animal Care and Use Committee of the Central South University has approved this study.

For brood size measurement, all lines were synchronized by transferring five adult nematodes to fresh dishes and allowing them to lay eggs for 3–4 h, after which the nematodes were removed. Twenty L4 individuals from each line were picked into 20 dishes and were allowed to lay eggs each day into a new dish for a total 8 d or until no more eggs were laid. The eggs in each dish were allowed to develop for 2 d before being counted.

See extended experimental procedures in Supporting Information for experiments on immune responses and high fat diet-induced obesity in BXD mice.

2 Results

2.1 MA distribution profiles in genetic reference populations

We calculated MAF for each scanned SNP in a panel of genetic reference population. We then calculated MAC for each strain of a panel and plotted the MAC distribution curve (Figure 1; Table S1 in Supporting Information for strain descriptions). Great variations in MAC (~0.2 to ~0.7) were observed for *C. elegans* RILs from either the Kruglyak or the Kammenga laboratory (Figure 1A and B) [19,20], a yeast segregants panel analogous to a RIL panel in animals (Figure 1C) [21,22], and the BXD mouse RIL panel (Figure 1F) [24,25]. Relative to these RILs, *D. melanogaster* inbred panel derived from the wild showed lower MAC and smaller variation range (Figure 1D) [23]. RILs that were only partially inbred such as the collaborative cross (CC) G2F7 mouse panel that has been inbred for only seven generations also showed small variation range in MAC (Figure 1E) [26]. For certain panels with large variations, an abrupt turn at the ends of the distribution curve, especially the higher end, was apparent, indicating an under-representation and hence lower survival success of strains with low or high MAC (Figure 1A–D and F). The population distribution of MAC showed a bell curve as expected (Figure S1 in Supporting Information). For calculating MAC, the number of informative SNPs used for each panel ranged from ~120 to ~151000. Since the SNPs used here are largely selected in a non-biased way, the number of SNPs used should not significantly affect the calculation of MAC. Indeed as shown for the BXD mouse panel, MAC calculated from ~51000 SNPs were highly similar to those calculated using two different non-overlapping sets of 1000 SNPs randomly selected from the ~51000 (Figure S2 in Supporting Information).

2.2 MAC correlates with quantitative variations in complex traits in model organisms

To determine whether MAC may affect reproductive fitness, we examined brood size of 42 *C. elegans* RILs from the Kruglyak laboratory with Hawaii (HW) *npr-1* genotype and 62 RILs with N2 *npr-1* genotype (Figure 2A and B). Their parental strains Hawaii CB4856 and Bristol N2 differ in *npr-1* by one major effect SNP (F215V). Higher MAC was linked with lower brood size in a nearly linear fashion, with its effect stronger in HW *npr-1* background (Figure 2A and B). The deleterious effect of higher MAC on reproductive traits was confirmed in three other RIL panels in mouse and rat, BXD, CC (G2F6), and BXHHXB (Table 4). In addition, higher MAC was linked with lower life span in *C. elegans* and mouse (Table 4), less startle response (the ability to respond rapidly to harmful changes in the environment) in *D. melanogaster* (for males, Spearman *r*=–0.23, *P*=0.004) [23], and more chill coma response in *D. melanogaster* (for females, Spearman *r*=0.22, *P*=0.007) [23].

There are 3664 traits for the BXD mouse panel of 89 strains characterized by numerous studies with data archived at GeneNetwork [30,31]. Fifteen traits were found linked with MAC by Pearson analysis and 15 by Spearman analysis (*P*<0.0001), including BEC, higher deoxycorticosterone level in cerebral cortex, and higher adrenal weight (Table S2 in Supporting Information). In comparison, assigning an arbitrary numerical value to each RIL strain did not produce any trait correlation with *P*<0.0001 in any of the 100 tests we did.

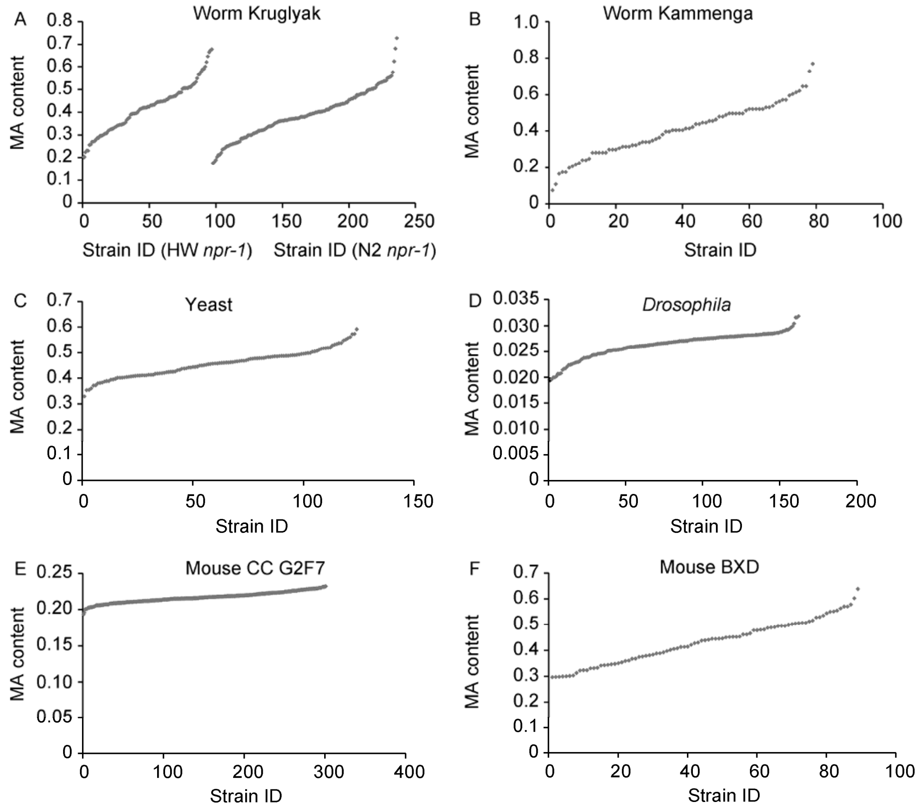


Figure 1 Distribution profile of MAC of each strain in a panel of genetic reference population. A, *C. elegans* RIL strains from the Kruglyak laboratory separated by the npr-1 F215V mutation into the HW and N2 types. B, *C. elegans* RIL strains from the Kammenga laboratory. C, Budding yeast segregants panel. D, *D. melanogaster* inbred strain panel derived from randomly selected individuals in the wild. E, Mouse RIL panel from Collaborative Cross at 7th generation of inbreeding. F, Mouse BXD RIL panel. Strain ID numbers are arbitrary in order to fit space for the Figure.

There were 297 traits with *P*<0.05 (Table S2 in Supporting Information). Random sorting tests suggest that 60% of these correlations may be false positive at this *P* value. A few examples include lower maximum threshold to ethanol induced ataxia (Figure 3A) and lower blood ethanol concentration in males 20 min after ethanol injection (Figure 3B). A number of related neurological traits were linked with higher MAC, including smaller methamphetamine-induced body temperature change, slower reversal learning, and higher sensitivity to pain (Table S2 in Supporting Information, Table 4).

Most of the 3664 traits in Table S2 in Supporting Information were scored for less than half of the panel and different sets of strains were often used for scoring different traits. After filtering out traits and strains with too many missing data, we were able to perform multivariate regression analysis on nine traits, which identified three significant associations, including BEC, adrenal zona fasciculata width, and hair coat color (Table S3 in Supporting Information).

In addition to reproductive fitness as mentioned above, a number of other traits were repeatedly found linked with MAC in different panels of RILs (Table 4). One was tumor susceptibility (Figure 4, Table 4). The effect of MAC in urethane induced lung tumor was only apparent when *kras2* oncogene was wild type (Figure 4A vs. 4B) [32]. There were also traits such as blood pressure that were repeatedly not found associated with MAC (Table 4). MAC also consistently associated with traits linked to obesity and type 2 diabetes. In BXHHXB rat, more MAs were linked with higher glucose level after high fructose diet (Figure 5A) and lower serum dopamine level (Figure 5B). In high fat diet fed BXD mice, higher MAC correlated with higher resistin level and more body weight increase (Table 2; Table S2 in Supporting Information).

We next examined the hypothesis that greater genetic diversity may enhance adaptive immunity. A number of immunity traits in two panels of mouse RILs (BXD and BXH) were significantly associated with MAC (Table 5). Importantly, higher MAC was uniformly associated with stronger immune responses.

The yeast segregant panel mentioned above in Figure 1C has been stably grown for many generations so that segregants with excess deleterious SNPs would have failed to grow to be included in the panel. This yeast segregant panel has been used to identity SNPs involved in 316 response profiles to 92 drugs and chemical compounds [21]. From this published dataset, we identified 12 MAC linked traits at

Table 4 Repeated tests of associations between traits and MAC

RIL panel	Correlation	Reprod. fit.[a]			Life span			Alcohol sens.[b]			Cancer[c]			Blood pressure		
		r	*P*	#	*r*	*P*	#	*r*	*P*	#	*r*	*P*	#	*r*	*P*	#
BXD	Pearson	–0.23	ns	61	–0.12	ns	20	–0.27	0.04	25	0.22	ns	22	–0.27	ns	21
	Spearman	–0.25	0.05		–0.16	ns		–0.25	0.05		0.13	ns		–0.28	ns	
AXBBXA	Pearson										*t* test, *P*<0.05		17			
	Spearman										Lung tumor					
CXB	Pearson										0.69	ns	7			
	Spearman										0.82	0.03				
LXS	Pearson				–0.25	ns	43	–0.23	0.05	74						
	Spearman				–0.29	0.06		–0.27	0.02							
CC (F6)	Pearson	–0.14	0.04	245												
	Spearman	–0.13	ns													
BXHHXB	Pearson	–0.49	0.02	24										–0.02	ns	32
	Spearman	–0.50	0.01											–0.06	ns	
Worm Kruglyak	Pearson	–0.45	0.002	42												
	Spearman	–0.45	0.002													
Worm Kammenga	Pearson				–0.27	ns	35									
	Spearman				–0.35	0.04										

a) Reproductive fitness includes uterus horn length (BXD), litter/brood size (CC, Worm), and fetal weights in left horn of uterus (BXHHXB rat). b) Alcohol sensitivity was assayed by distance traveled after ethanol injection. c) Cancer includes DEN induced liver tumor (BXD), urethane induced lung tumor (AXBBXA), and virus induced lymphoma (CXB).

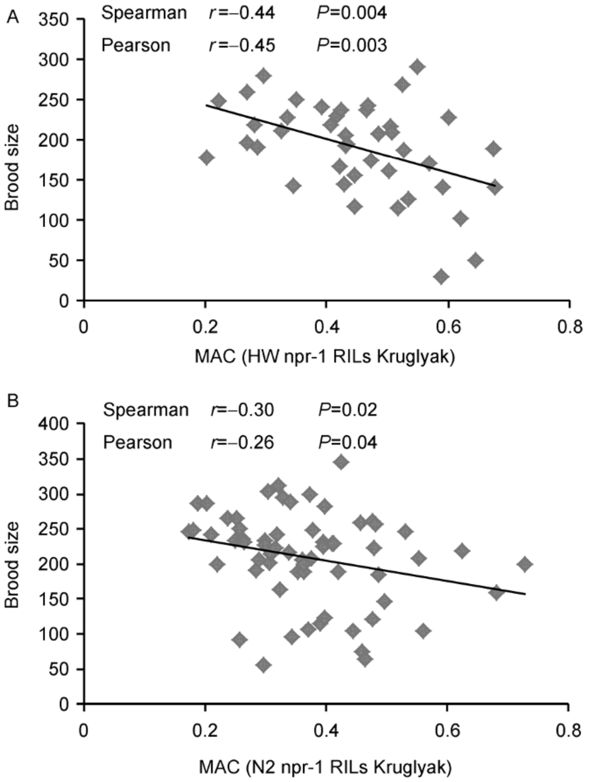


Figure 2 MAC on reproductive fitness in *C. elegans*. A, Brood size in Kruglyak RIL strains with HW npr-1 genotype. B, Brood size in Kruglyak RIL strains with N2 npr-1 genotype.

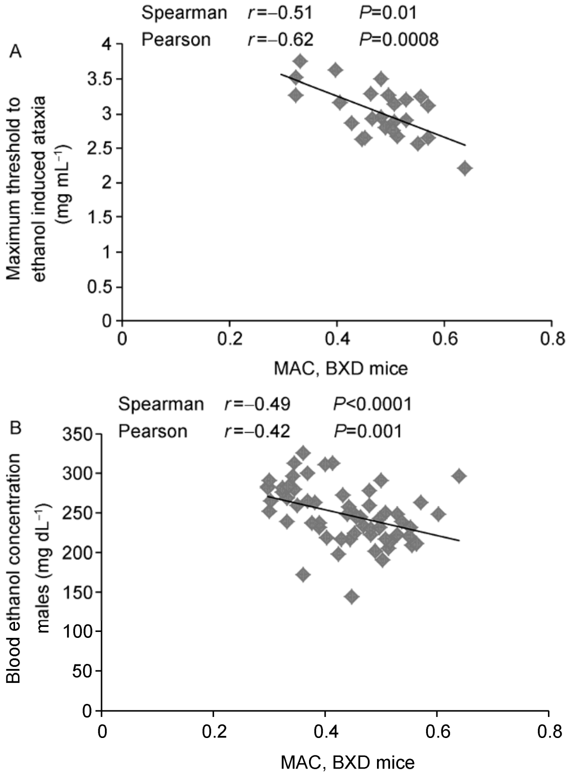
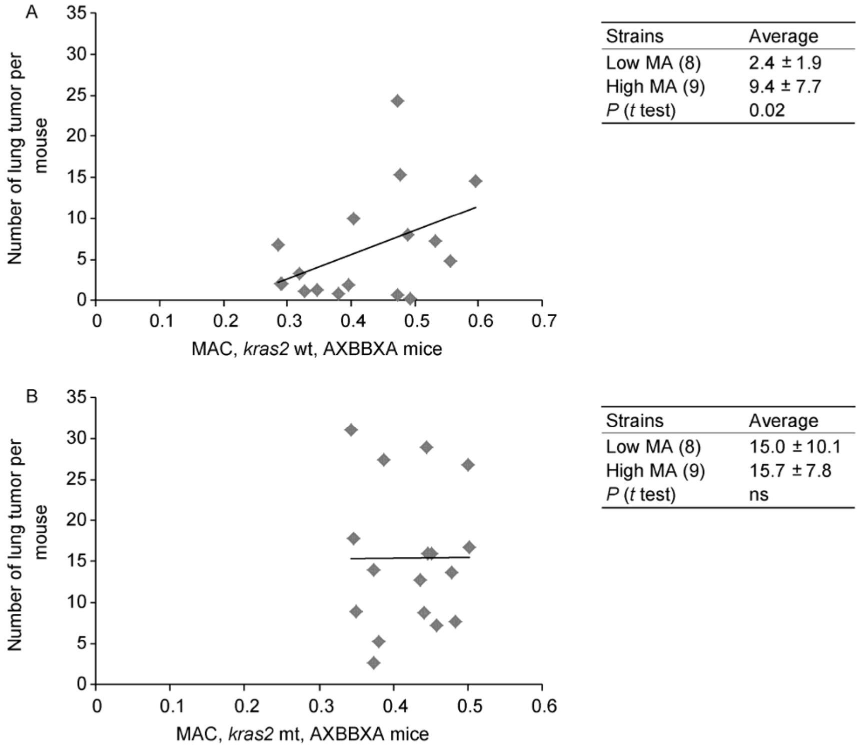


Figure 3 MAC on ethanol traits in BXD mice. A, Maximum threshold to ethanol induced ataxia. B, Blood ethanol concentration in males.

Table 5 MAC on immune responses in mouse RIL panels

Mice	GN ID[a)]	Sample #	Trait description	Pearson		Spearman	
				r	*P*	*r*	*P*
BXD	13969	24	CFU in liver 48 h post i.v. *S. aureus* infection	–0.57	0.003	–0.66	0.0005
	14313	6	White blood cell count after *C. albicans* i.v. infection	0.93	0.01	0.83	0.04
	10779	18	IgG1 anti-cF.IX (coagul. Factor IX) after cF.IX injection	0.6	0.01	0.48	0.04
	10695	21	Pulmonary granulomatous inflammation by BCG	–0.47	0.03	–0.39	ns
	12668	33	Formation of dermal lesions, ECTV footpad	–0.36	0.04	–0.32	ns
	10663	20	Cytotoxicity in spleen T cells post AdLacZ i.v. injection	–0.43	ns	–0.55	0.01
	10806	25	Mortality after i.p. *C. psittaci* infection	–0.39	0.05	–0.43	0.03
BXH	10115	10	Survival times of allograft	–0.68	0.02	–0.6	0.04

a) GeneNetwork identification number.



Strains	Average
Low MA (8)	2.4 ± 1.9
High MA (9)	9.4 ± 7.7
P (*t* test)	0.02

Strains	Average
Low MA (8)	15.0 ± 10.1
High MA (9)	15.7 ± 7.8
P (*t* test)	ns

Figure 4 MAC on tumorigenesis in mouse RILs. A, The number of lung tumors induced by urethane in *kras2* wild type AXBBXA strains. B, The number of lung tumors induced by urethane in *kras2* mutant AXBBXA strains. Also shown are average tumor values of top and bottom half in MAC and *t* test *P* values.

zero false discovery rate, which all showed more growth inhibition in strains with higher MAC, indicating a link between lower reproductive fitness and higher MAC. Seven among these involved four drugs that are FDA-approved antipsychotic and antidepression drugs (sertraline, trimeprazine, chlorpromazine, and trifluoperazine), and one involved the FDA-approved cancer drug Tamoxifen (Table 6).

2.3 Trait specific set of MAs

We found that certain traits were correlated with certain other traits (Table 2). To confirm among the MAC-linked traits that correlated traits share more MAs than non-correlated traits (Table 2), we developed an approach to identifying trait specific set of MAs as described in the Methods. From 51469 SNPs originally used for calculating MAC for the BXD panel, we identified a BEC-specific set of 30336 SNPs. When the MAC value of each RIL strain was calculated using the BEC specific set, the BEC trait was strongly linked with MAC (Spearman *r*=–0.66, *P*=0.0004) and so was its related trait resistin level (*r*=0.53, *P*=0.008). In comparison, relatively weak association was noted for the two BEC-nonrelated traits, rearing behavior (*r*=–0.4, *P*=0.06) and pain (*r*=–0.46, *P*=0.03). In contrast, two non-related traits, resistin level and pain response,

Table 6 MAC correlation with yeast growth in the presence of a compound using 105 segregants

Compounds	Pearson		Spearman	
	r	*P*	*r*	*P*
diphenyleneiodonium 64 h 16 μmol L^{-1}	–0.44	<0.0001	–0.39	<0.0001
sertraline 68 h 20.9 μmol L^{-1}[a)]	–0.36	0.0004	–0.36	0.0003
sertraline 52 h 20.9 μmol L^{-1}	–0.30	0.003	–0.30	0.004
sertraline 78 h 20.9 μmol L^{-1}	–0.31	0.002	–0.33	0.001
diphenyleneiodonium 64 h 16 μmol L^{-1}	–0.30	0.003	–0.29	0.004
alverine 118 h 105.5 μmol L^{-1}	–0.29	0.004	–0.31	0.002
tamoxifen 70 h 13.5 μmol L^{-1}	–0.33	0.001	–0.28	0.007
trimeprazine 80 h 83.8 μmol L^{-1}	–0.33	0.001	–0.33	0.001
chlorpromazine 70 h 15.7 μmol L^{-1}	–0.31	0.003	–0.27	0.009
sertraline 90 h 20.9 μmol L^{-1}	–0.31	0.002	–0.32	0.002
cinnarazine 68 h 33.9 μmol L^{-1}	–0.27	0.009	–0.22	0.03
trifluoperazine 90 h 26 μmol L^{-1}	–0.27	0.007	–0.27	0.007

a) FDA-approved drugs in italics.

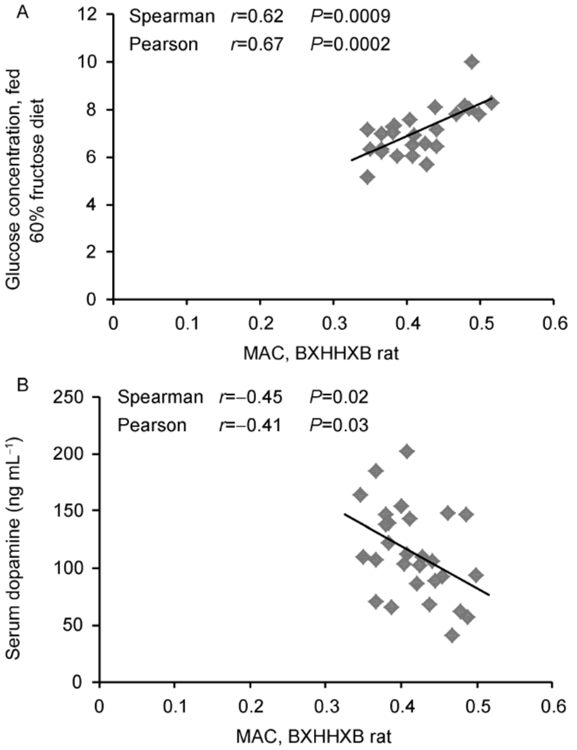


Figure 5 MAC on glucose and dopamine levels in rat RILs. A, Glucose concentration in 10-week-old BXHHXB male rats fed a diet with 60% fructose from 8 to 10 weeks. B, Serum dopamine level in 6-week-old male BXHHXB rats.

scored the best correlation with MAC among the four traits of concern if MAC value was calculated using the original random set of 51469 SNPs.

We next asked whether poor reproductive fitness and strong immunity are correlated, which would indicate some sharing of SNPs. The MAC-linked reproductive trait in BXD mice is uterus horn length at maturity as shown in Table 4. There was a correlation between this trait and formation of secondary dermal lesions upon ectromelia virus infection of footpad (Pearson *r*=0.40, *P*=0.03 for mixed sexes).

3 Discussion

3.1 The collective effects of SNPs are not neutral

Our results suggest a non-neutral role for the collective effects of most SNPs. MAC was linked with poorer rather than better performance in adaptive traits. Lower reproductive fitness may be sufficient to explain the lower frequency of some of these MAs. Negative selection *en utero* may also do so, and the effects of these MAs on some adult traits may reflect pleiotropy. In contrast, the link between higher MAC and better immunity and the inverse correlation between immunity and reproduction indicate simultaneous positive selection of the negatively selected MAs and explain why a common MA should be common rather than rare. While we have yet to obtain direct evidence for a functional role of the collective effects in any traits, the most parsimonious explanation for all the results here is natural selection. These results represent the first formal test of the neutral theory with regard to the collective effects of SNPs and have dramatically restricted the relevance of the infinite sites model.

This study analyzed numerous mouse traits for associations with MAC. About a dozen traits showed highly significant association with MAC (*P*<0.0001) when no such association was observed for 100 random sorting simulations. These traits are therefore sufficient to support the conclusion that MAC is linked with certain traits. Furthermore, our direct experimental test of the hypothesis of MAC association with reproductive fitness in *C. elegans* produced highly significant result, which is also sufficient to support a functional role of the collective effects of SNPs.

Among the numerous traits in BXD mice examined here, hundreds passed the weak significance value of *P*<0.05. There is a high possibility of false positives here. Most of these should therefore be considered as results of an exploratory study needing future verification. Using a multiplicity adjustment method such as the Bonferronni correction here may be of limited value since the method is widely known to be controversial and even absurd [33]. Such corrections are favored by researchers adhering to the neutral framework as they help to artificially reduce the amount of functional SNPs. Indeed, certain associations that would have been found by Bonferronni correction as false positives were in fact real because they could be repeatedly observed in independent studies as shown in Table 4. Most animal experiments have small sample sizes due to practical and financial reasons. The value of our study is to give the community a select list of MAC-linked traits worthy of future confirmatory studies.

Do the results here mean an additive effect of large numbers of MAs in MAC action and hence non-neutrality of most MAs? Many major effect risk alleles of diseases are known to be minor alleles [18], which may plausibly imply that the effect of MAC may be mediated by a few known major effect risk alleles rather than large numbers of minor effect MAs. But this may not necessarily be the case. The effect of MAC was in fact abolished or weakened by major effect MAs such as *kras2* mutation in lung cancer or *npr-1* mutation in brood size as found here. Furthermore, MAC preferentially affects traits with larger number of known additive QTLs [34]. Obviously, the more the number of QTLs involved in a trait, the less the individual effect of each QTL on the trait. Thus, MAC-linked traits are expected to have more additive minor effect SNPs as risk alleles than those not linked to MAC. The individual effect of such SNPs may not be possible for existing methods like GWAS to detect. Thus, the concept and methods of MAC here may help solve the “missing heritability” problem of some complex traits [1].

The results here suggest insights into the pathogenesis of certain diseases. It is well established that accumulation of somatic mutations causes cancer, although most such mutations are assumed to be neutral or “passengers” rather than “drivers” [35,36]. That RIL strains with more germline SNPs or MAC have higher lung cancer incidences suggests an oncogenic role for too much genetic variations. This makes good sense since cells with more random variations or SNPs should have more entropy, which would make growth control less precise and stable. Alcohol addiction in humans is associated with lower initial sensitivity to alcohols/drugs and strong alcohol and sweet preference and consumption [37]. Strains with high MAC showed these phenotypes and may thus serve as models of human alcoholism. High MAC in mice were linked with increase in resistin and insulin level and decrease in IL-17 level. Such alterations have been implicated in mouse and human obesity and T2D [38,39].

The mechanism of action of MAC in complex diseases/traits may be hard to delineate precisely and usefully, since the defining characteristic of complexity may be the breakdown of causality. As simply put by Goldenfeld and Woese, “complex systems are ones for which observed effects do not have uniquely definable causes, due to the huge nature of the phase space and the multiplicity of paths” [40]. Thus, holistic system or architectural plan approaches may be more productive in studying MAC action.

3.2 Optimum genetic diversity

Genetic diversity may increase with time but will eventually reach an optimum limit due to negative selection of the deleterious effects of too many random mutations. Complex diseases such as cancer and lower reproductive fitness as linked with higher MAC may be the price paid for maintaining the optimum or maximum limit. On the other hand, adaptive immunity may suffer if MAC is too low. So, it is optimum MAC level rather than either low MAC or high MAC that is adaptively most advantages.

The two extremes in the quantitative values of a trait, such as either too high or too low level in a hormone such as deoxycorticosterone as found here, often represent suboptimum population minorities, and are less desirable relative to the optimum level. The two extreme values would be represented by high and low MAC (Figure 6). Thus, both high and low MAC would be associated with less desirable traits. Here, lower survival rate of individuals with low MAC that are below optimum level would be in effect equivalent to higher survival rate for individuals with greater MAC. So, the nearly linear association of MAC with quantitative variations of traits automatically insures positive selection for MAs, in addition to negative selection. Such dual selection may explain why a MA is not too rare in frequency.

At the optimum level of allelic diversity, the overall slightly deleterious nature of MAs would be in homeostasis

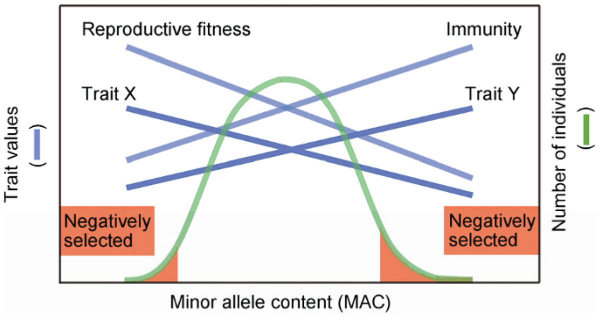


Figure 6 General schemes for both positive and negative selections of minor alleles. A complex trait typically shows quantitative variations in a population, with suboptimum values (either too high or too low) in the tail ends of the population bell curve as schematically shown here. Association of low MAC with suboptimum values of traits would result in negative selection of low MAC, which is equivalent to positive selection for higher MAC. On the other hand, association of high MAC with suboptimum values of traits also results in negative selection of high MAC. Thus, the nearly linear association of MAC with quantitative variations of complex traits automatically insures both positive and negative selection for MAs.

with the slightly beneficial nature of major alleles. The optimum concept here means a Pareto optimum or simply the best that can be achieved due to a balance between positive and negative selection at a particular time point under a specific level of epigenetic or organismal complexity [41]. As time and complexity changes, the optimum level of nucleotide diversity will also change.

What determines genetic diversity has been a long-standing unsolved puzzle [9]. The key to solve this puzzle may be to recognize two kinds of diversity, optimum and liner time dependent, as we proposed in the maximum genetic diversity theory [16,17]. The results here provide evidence for a critical role of physiology or system construction requirements in optimum genetic diversity as suggested by the maximum genetic diversity theory [16,17]. Genome compositional constraints may also play a role [42]. Optimum genetic diversity is a novel concept that may help solve more puzzles in genetics in the years to come.

We thank E. Andersen, L. Kruglyak (Princeton University, USA), A. Bendesky, P. McGrath, C. Bargmann (Rockefeller University, USA), E. Chesler (Jackson Lab, USA), M. Rockman (New York University, USA), R. Brem, E. Smith (University of California, Berkeley, USA), L. Xia, J. Li (Central South University, USA), B. Chen (Bioinformatic.org), L. Zhou, Y. Zheng (Central South University, USA), R. Mott (Wellcome Trust Center for Human Genetics, UK), X. Zhou, and R. Williams (University of Tennessee, USA) for research materials or technical assistance, and A. Bendesky and C. Bargmann for critical reading of the manuscript. This work was supported by the National Natural Science Foundation of China (81171880), the National Basic Research Program of China (2011CB51001 to S. Huang), and the GeNeSys Consortium (to O. Goldmann and E. Medina).

1 Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature, 2009, 461: 747–753

2 Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES; International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Johnson TA, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. Genome-wide detection and characterization of positive selection in human populations. Nature, 2007, 449: 913–918

3 Conrad DF1, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J; Wellcome Trust Case Control Consortium, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME. Origins and functional impact of copy number variation in the human genome. Nature, 2010, 464: 704–712

4 Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature, 2009, 460: 748–752

5 Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, Dudbridge F, Holmans PA, Whittemore AS, Mowry BJ, Olincy A, Amin F, Cloninger CR, Silverman JM, Buccola NG, Byerley WF, Black DW, Crowe RR, Oksenberg JR, Mirel DB, Kendler KS, Freedman R, Gejman PV. Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature, 2009, 460: 753–757

6 Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Pietiläinen OP, Mors O, Mortensen PB, Sigurdsson E, Gustafsson O, Nyegaard M, Tuulio-Henriksson A, Ingason A, Hansen T, Suvisaari J, Lonnqvist J, Paunio T, Børglum AD, Hartmann A, Fink-Jensen A, Nordentoft M, Hougaard D, Norgaard-Pedersen B, Böttcher Y, Olesen J, Breuer R, Möller HJ, Giegling I, Rasmussen HB, Timm S, Mattheisen M, Bitter I, Réthelyi JM, Magnusdottir BB, Sigmundsson T, Olason P, Masson G, Gulcher JR, Haraldsson M, Fossdal R, Thorgeirsson TE, Thorsteinsdottir U, Ruggeri M, Tosato S, Franke B, Strengman E, Kiemeney LA; Genetic Risk and Outcome in Psychosis (GROUP), Melle I, Djurovic S, Abramova L, Kaleda V, Sanjuan J, de Frutos R, Bramon E, Vassos E, Fraser G, Ettinger U, Picchioni M, Walker N, Toulopoulou T, Need AC, Ge D, Yoon JL, Shianna KV, Freimer NB, Cantor RM, Murray R, Kong A, Golimbet V, Carracedo A, Arango C, Costas J, Jönsson EG, Terenius L, Agartz I, Petursson H, Nöthen MM, Rietschel M, Matthews PM, Muglia P, Peltonen L, St Clair D, Goldstein DB, Stefansson K, Collier DA. Common variants conferring risk of schizophrenia. Nature, 2009, 460: 744–747

7 O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, Nikolov I, Hamshere M, Carroll L, Georgieva L, Dwyer S, Holmans P, Marchini JL, Spencer CC, Howie B, Leung HT, Hartmann AM, Möller HJ, Morris DW, Shi Y, Feng G, Hoffmann P, Propping P, Vasilescu C, Maier W, Rietschel M, Zammit S, Schumacher J, Quinn EM, Schulze TG, Williams NM, Giegling I, Iwata N, Ikeda M, Darvasi A, Shifman S, He L, Duan J, Sanders AR, Levinson DF, Gejman PV, Cichon S, Nöthen MM, Gill M, Corvin A, Rujescu D, Kirov G, Owen MJ, Buccola NG, Mowry BJ, Freedman

R, Amin F, Black DW, Silverman JM, Byerley WF, Cloninger CR; Molecular Genetics of Schizophrenia Collaboration. Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nat Genet, 2008, 40: 1053–1055

8 Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, König IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens AC, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllensten U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Döring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burtt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA Jr, Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S. Biological, clinical and population relevance of 95 loci for blood lipids. Nature, 2010, 466: 707–713

9 Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, Venkat A, Andolfatto P, Przeworski M. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol, 2012, 10: e1001388

10 Huang S. The overlap feature of the genetic equidistance result, a fundamental biological phenomenon overlooked for nearly half of a century. Biological Theory, 2010, 5: 40–52

11 Fay JC, Wyckoff GJ, Wu CI. Testing the neutral theory of molecular evolution with genomic data from Drosophila. Nature, 2002, 415: 1024–1026

12 Ponting CP, Hardison RC. What fraction of the human genome is functional? Genome Res, 2011, 21: 1769–1776

13 Mattick JS, Dinger ME. The extent of functionality in the human genome. HUGO J, 2013, 7, doi:10.1186/1877-6566-1187-1182

14 ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature, 2012, 489: 57–74

15 Pheasant M, Mattick JS. Raising the estimate of functional human sequences. Genome Res, 2007, 17: 1245–1253

16 Hu T, Long M, Yuan D, Zhu Z, Huang Y, Huang S. The genetic equidistance result, misreading by the molecular clock and neutral theory and reinterpretation nearly half of a century later. Sci China Life Sci, 2013, 56: 254–261

17 Huang S. Inverse relationship between genetic diversity and epigenetic complexity. Preprint available at Nature Precedings, 2009, http://dx.doi.org/10.1038/npre.2009.1751.2

18 Park JH, Gail MH, Weinberg CR, Carroll RJ, Chung CC, Wang Z, Chanock SJ, Fraumeni JF Jr., Chatterjee N. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. Proc Natl Acad Sci USA, 2011, 108: 18026–18031

19 Seidel HS, Rockman MV, Kruglyak L. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. Science, 2008, 319: 589–594

20 Vinuela A, Snoek LB, Riksen JA, Kammenga JE. Genome-wide gene expression regulation as a function of genotype and age in *C. elegans*. Genome Res, 2010, 20: 929–937

21 Perlstein EO, Ruderfer DM, Roberts DC, Schreiber SL, Kruglyak L. Genetic basis of individual differences in the response to small-molecule drugs in yeast. Nat Genet, 2007, 39: 496–502

22 Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. Science, 2002, 296: 752–755

23 Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RR, Mackay TF. Systems genetics of complex traits in *Drosophila melanogaster*. Nat Genet, 2009, 41: 299–307

24 Taylor BA. Recombinant Inbred Strains: Use in Gene Mapping. New York: Academic Press, 1978

25 Philip VM, Duvvuru S, Gomero B, Ansah TA, Blaha CD, Cook MN, Hamre KM, Lariviere WR, Matthews DB, Mittleman G, Goldowitz D, Chesler EJ. High-throughput behavioral phenotyping in the expanded panel of BXD recombinant inbred strains. Genes Brain Behav, 2010, 9: 129–159

26 Philip VM, Sokoloff G, Ackert-Bicknell CL, Striz M, Branstetter L, Beckmann MA, Spence JS, Jackson BL, Galloway LD, Barker P, Wymore AM, Hunsicker PR, Durtschi DC, Shaw GS, Shinpock S, Manly KF, Miller DR, Donohue KD, Culiat CT, Churchill GA, Lariviere WR, Palmer AA, O'Hara BF, Voy BH, Chesler EJ. Genetic analysis in the Collaborative Cross breeding population. Genome Res, 2011, 21: 1223–1238

27 Chen B, Wilkening S, Drechsel M, Hemminki K. SNP_tools: a compact tool package for analysis and conversion of genotype data for MS-Excel. BMC Res Notes, 2009, 2: 214

28 Purcell S1, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet, 2007, 81: 559–575

29 Xu S, Gupta S, Jin L. PEAS V1.0: a package for elementary analysis of SNP data. Mol Ecol Resources, 2010, 10: 1085–1088

30 Taylor BA, Wnek C, Kotlus BS, Roemer N, MacTaggart T, Phillips SJ. Genotyping new BXD recombinant inbred mouse strains and comparison of BXD and consensus maps. Mamm Genome, 1999, 10: 335–348

31 Wang J, Williams RW, Manly KF. WebQTL: web-based complex trait analysis. Neuroinformatics, 2003, 1: 299–308

32 Ryan J, Barker PE, Nesbitt MN, Ruddle FH. KRAS2 as a genetic marker for lung tumor susceptibility in inbred mice. J Natl Cancer Inst, 1987, 79: 1351–1357

33 Perneger TV. What's wrong with Bonferroni adjustments. BMJ, 1998, 316: 1236–1238

34 Zhu Z, Lu Q, Yuan D, Li Y, Man X, Zhu Y, Huang S. Role of genetic polymorphisms in transgenerational inheritance of inherent as well as acquired traits in budding yeast. 2013, arXiv:1302.7276

35 Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B. The genomic landscapes

of human breast and colorectal cancers. Science, 2007, 318: 1108–1113

36 Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr., Kinzler KW. Cancer genome landscapes. Science, 2013, 339: 1546–1558

37 Crabbe JC. Genetic contributions to addiction. Annu Rev Psychol, 2002, 53: 435–462

38 Zuniga LA, Shen WJ, Joyce-Shaikh B, Pyatnova EA, Richards AG, Thom C, Andrade SM, Cua DJ, Kraemer FB, Butcher EC. IL-17 regulates adipogenesis, glucose homeostasis, and obesity. J Immunol, 2010, 185: 6947–6959

39 Steppan CM, Bailey ST, Bhat S, Brown EJ, Banerjee RR, Wright CM, Patel HR, Ahima RS, Lazar MA. The hormone resistin links obesity to diabetes. Nature, 2001, 409: 307–312

40 Goldenfeld N, Woese C. Life is physics: evolution as a collective phenomenon far from equilibrium. Annu Rev Cond Matt Phys, 2011, 2: 375–399

41 Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo A, Dekel E, Kavanagh K, Alon U. Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. Science, 2012, 336: 1157–1160

42 Bernardi G, Bernardi G. Compositional constraints and genome evolution. J Mol Evol, 1986, 24: 1–11

Biographical Sketch Dr. Huang Shi was a professor at the State Key Laboratory of Medical Genetics at Central South University in Changsha, China. He grew up in the army compound of the Chinese Academy of Military Medicine in Beijing where his father was a professor. His childhood interest was however not medicine but sports and later fine arts. An unsuccessful effort in the entrance examination of the Chinese Central Academy of Fine Arts in 1978 changed his interest to science. He entered Fudan University in 1979 and graduated in 1983 with a bachelor's degree in genetics. He was a CUSBEA fellow of Class III (1984) and obtained his Ph.D. in biochemistry at the University of California at Davis in 1988. After finishing a postdoctoral training at the University of California at San Diego, he was appointed in 1992 assistant professor at the Sanford-Burnham Institute and promoted to associate professor in 1998. He was appointed professor at Central South University in 2009. The early training in art helped shape his taste in aesthetics and interest in science only as a creative endeavor. His laboratory discovered the RIZ or PRDM family of histone methyltransferases and proposed an epigenetic pathway of carcinogenesis by diet rich in meat and low in vegetables. Since 2003, he initiated study of the relationship between genetics and epigenetics and its role in the evolution of biological complexity. He proposed the maximum genetic diversity hypothesis and has been using it to rewrite evolution and population genetics as well as to solve genetic puzzles of complex traits/diseases about which the existing paradigm is clueless. He was one of the 1993 class of Pew Scholars in the Biomedical Sciences.

Supporting Information

Figure S1 MAC distribution profiles among the strains in a RILs or segregants panel.

Figure S2 MAC value is independent of SNP numbers or random selection of SNPs.

Table S1 MAC values of RIL strains or segregants

Table S2 Correlation between traits and MAC values in BXD mice

Table S3 Multivariate regression analysis of selected traits

The supporting information is available online at life.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.