

# Adaptive evolution of interleukin-3 (IL3), a gene associated with brain volume variation in general human populations

Ming Li<sup>1</sup> · Liang Huang<sup>2</sup> · Kaiqin Li<sup>1</sup> · Yongxia Huo<sup>1</sup> · Chunhui Chen<sup>3</sup> ·  
Jinkai Wang<sup>4</sup> · Jiewei Liu<sup>4</sup> · Zhenwu Luo<sup>5</sup> · Chuansheng Chen<sup>6</sup> · Qi Dong<sup>3</sup> ·  
Yong-gang Yao<sup>1</sup> · Bing Su<sup>4</sup> · Xiong-jian Luo<sup>1</sup>

Received: 30 November 2015 / Accepted: 4 February 2016 / Published online: 13 February 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** Greatly expanded brain volume is one of the most characteristic traits that distinguish humans from other primates. Recent studies have revealed genes responsible for the dramatically enlarged human brain size (i.e., the microcephaly genes), and it has been well documented that many microcephaly genes have undergone accelerated evolution along the human lineage. In addition to being far larger than other primates, human brain volume is also highly variable in general populations. However, the genetic basis underlying human brain volume variation

remains elusive and it is not known whether genes regulating human brain volume variation also have experienced positive selection. We have previously shown that genetic variants (near the *IL3* gene) on 5q33 were significantly associated with brain volume in Chinese population. Here, we provide further evidence that support the significant association of genetic variants on 5q33 with brain volume. Bioinformatic analyses suggested that rs31480 is likely to be the causal variant among the studied SNPs. Molecular evolutionary analyses suggested that *IL3* might have undergone positive selection in primates and humans. Neutrality tests further revealed signatures of positive selection of *IL3* in Han Chinese and Europeans. Finally, extended haplotype homozygosity (EHH) and relative EHH analyses showed that the C allele of SNP rs31480 might have experienced recent positive selection in Han Chinese. Our results suggest that *IL3* is an important genetic regulator for human brain volume variation and implied that *IL3* might have experienced weak or modest positive selection in the evolutionary history of humans, which may be due to its contribution to human brain volume.

Ming Li, Liang Huang, Kaiqin Li and Yongxia Huo contributed equally to this work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00439-016-1644-z) contains supplementary material, which is available to authorized users.

✉ Xiong-jian Luo  
luoxiongjian@mail.kiz.ac.cn

- <sup>1</sup> Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, Yunnan, China
- <sup>2</sup> First Affiliated Hospital of Gannan Medical University, Ganzhou 341000, Jiangxi, China
- <sup>3</sup> State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China
- <sup>4</sup> State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, Yunnan, China
- <sup>5</sup> Department of Microbiology and Immunology, Medical University of South Carolina, Charleston, SC 29425, USA
- <sup>6</sup> Department of Psychology and Social Behavior, University of California, Irvine, CA 92697, USA

## Introduction

As one of the most characteristic features of humans, dramatically increased brain volume is the basis of our highly developed cognitive capabilities. Though significant progress has been made during the past decades, the genetic mechanisms underlying our greatly expanded brain volume remain largely unknown. Fortunately, recent studies have revealed that microcephaly genes may contribute to the evolutionary expansion of human brain volume (Bond et al. 2002, 2003; Jackson et al. 1998, 2002; Jamieson et al. 1999, 2000; Moynihan et al. 2000; Pattison et al. 2000;

Roberts et al. 1999). Mutations of microcephaly genes caused severe neurodevelopmental defects and patients with microcephaly disease have a brain volume of around 400 cm<sup>3</sup>, which is much smaller than the 1200–1600 cm<sup>3</sup> of a normal adult brain volume (Mochida and Walsh 2001; Woods et al. 2005). Since microcephaly genes are major determinants of human brain volume, it is hypothesized that microcephaly genes might have undergone Darwinian positive selection during the evolutionary history of human beings. Consistent with this notion, multiple studies have showed adaptive (or accelerated) evolution of microcephaly genes along lineages leading to humans (Evans et al. 2004; Kouprina et al. 2004; Mekel-Bobrov et al. 2005; Wang and Su 2004; Zhang 2003).

Brain volume is a complex quantitative trait with a strong genetic component (Peper et al. 2007; Posthuma et al. 2002; Thompson et al. 2001, 2002). In addition to being far greater than most other species, the volume of the human brain exhibits considerable variation in the general population (Rushton 1992). To uncover the genetic basis underlying the human brain volume variation, several genome-wide association studies (GWAS) have been performed in populations of European ancestry (Ikram et al. 2012; Stein et al. 2012; Taal et al. 2012). Though these GWAS have identified several promising candidate variants/genes for brain volume variation, it is not known how these genes regulate brain development and brain volume. Furthermore, a proportion of the genetic variants identified in European populations were monomorphic (fixed) in the Chinese population, suggesting that distinct genetic variants may contribute to brain volume variation in the Chinese population.

To identify the potential genetic variants that are associated with brain volume variation in the Chinese population, we previously investigated the association between genetic variants on chromosome 5q22–33 and brain volume in a large healthy Chinese sample ( $N = 1013$ ) and our results indicated that genetic variations encompassing *IL3* gene were significantly associated with brain volume variation in Chinese. We systematically investigated the role of *IL3* in the central nervous system and our findings demonstrated that *IL3* contributes to human brain volume variation by regulating proliferation and survival of neural progenitors (Luo et al. 2012).

In this study, we further detailed the associations between genetic variants on 5q33.1 and brain volume using a dense fine-mapping strategy. Our findings supported that multiple highly linked SNPs on 5q33 are significantly associated with brain volume in Chinese population. Consistent with our previous findings, bioinformatics analyses implied that rs31480 may represent the causal variant among the studied SNPs. Considering that brain volume plays a crucial role in human evolution and positive selection

frequently operated on genes associated with brain volume and cognitive function (Woods et al. 2005), we speculated that genes contributing to human brain volume variation might have also experienced positive selection in the human lineage and modern human populations. We therefore systematically characterized the molecular evolutionary pattern of rs31480 and *IL3* in primates and human populations. Phylogeny-based maximum likelihood methods and neutrality tests suggested that *IL3* might have undergone weak adaptive evolution in human and nonhuman primates. Moreover, EHH and REHH analyses also revealed that rs31480 might have experienced modest recent positive selection in the Chinese population. Collectively, our results implied that *IL3* might experience weak or modest adaptive evolution in the evolutionary history of humans, which may be due to its contribution to brain volume.

## Results

### Fine mapping of 5q33 supports significant association of rs31480 and rs3916441 with brain volume in the Chinese population

We previously explored the association between SNPs located on 5q22–33 and brain volume in the Chinese population and identified several SNPs that showed significant associations with brain volume in females (Luo et al. 2012). Among the significant SNPs, rs3916441 showed the most significant association with brain volume. To further refine the association signals, we utilized a dense fine-scale mapping strategy by genotyping another 15 SNPs surrounding rs3916441 (Table 1; Fig. 1a). Seven of the 15 newly selected SNPs were located upstream of rs3916441 and 8 of them resided downstream of rs3916441 (Fig. 1a). Quantitative trait genetic association analyses indicated that 11 of the 15 selected SNPs showed significant association with brain volume in females (Bonferroni corrected  $P$  value  $< 0.01$ ) (Table 1; Fig. 1a; Supplementary Table S1). Comparing the  $P$  values of the 15 newly genotyped SNPs with our previous results showed that rs3916441 had the smallest  $P$  value (Table 1). These results further confirmed the association between genetic variants located on 5q33 and brain volume in the Chinese population.

There are several genes near the targeted region, including *ACSL6*, *IL3* and *CSF2*. To identify the most likely gene responsible for brain volume variation, we further investigated the linkage disequilibrium (LD) relationships of the studied SNPs and found that the significant SNPs are highly linked in the Chinese (Fig. 1b). Of note, rs31480 and rs3916441 showed a high degree of LD ( $r^2 = 0.89$ ). Most of the significant SNPs (corrected  $P < 0.05$ ) (Supplementary Table S1) were located upstream of the *IL3* gene and

**Table 1** Average cranial volumes of female subjects with three different genotypes at each of the genotyped SNPs

SNP ID <sup>a</sup>	Polymorphism (MAF) <sup>b</sup>	Average brain volume $\pm$ SD			<i>P</i> value <sup>c</sup>	<i>P</i> value <sup>d</sup>
<b>rs247011</b>	A/G (A:0.41)	AA (1216 $\pm$ 90.2)	AG (1222 $\pm$ 92.4)	GG (1247 $\pm$ 93.8)	0.0673	0.359
<b>rs3846727</b>	C/T (C:0.48)	TT (1257 $\pm$ 87.2)	TC (1223 $\pm$ 91.5)	CC (1211 $\pm$ 97.1)	0.0022	0.154
<b>rs6596051</b>	G/A (G:0.49)	AA (1255 $\pm$ 84.2)	AG (1224 $\pm$ 93.9)	GG (1212 $\pm$ 94.8)	0.0035	0.204
<b>rs17132324</b>	A/G (A:0.49)	AA (1214 $\pm$ 96.9)	AG (1222 $\pm$ 89.8)	GG (1260 $\pm$ 89.5)	0.0013	0.260
<b>rs3763116</b>	G/T (G:0.49)	TT (1259 $\pm$ 85.3)	TG (1221 $\pm$ 91.1)	GG (1218 $\pm$ 97.9)	0.0046	0.161
<b>rs3763114</b>	T/A (G:0.49)	TT (1217 $\pm$ 94.3)	TA (1220 $\pm$ 91.6)	AA (1257 $\pm$ 89.4)	0.0067	0.196
<b>rs12656759</b>	C/T (C:0.49)	TT (1257 $\pm$ 86.7)	TC (1222 $\pm$ 93.4)	CC (1215 $\pm$ 94.0)	0.0062	0.201
<i>rs3916441</i>	T/C (T:0.49)	TT (1209 $\pm$ 93.7)	TC (1226 $\pm$ 91.2)	CC (1256 $\pm$ 91.0)	0.0009	0.297
<b>rs10074987</b>	G/A (G:0.49)	AA (1259 $\pm$ 88.8)	AG (1219 $\pm$ 89.9)	GG (1219 $\pm$ 97.2)	0.0062	0.204
<b>rs6868554</b>	G/T (G:0.49)	TT (1257 $\pm$ 86.7)	TG (1223 $\pm$ 92.4)	GG (1212 $\pm$ 94.8)	0.0017	0.220
<b>rs2662406</b>	G/A (G:0.48)	AA (1257 $\pm$ 84.4)	AG (1223 $\pm$ 93.5)	GG (1214 $\pm$ 95.4)	0.0041	0.204
<b>rs246755</b>	T/G (T:0.48)	TT (1217 $\pm$ 95.9)	TG (1220 $\pm$ 92.3)	GG (1258 $\pm$ 86.3)	0.0044	0.153
<b>rs246756</b>	A/G (A:0.49)	AA (1257 $\pm$ 87.6)	AG (1224 $\pm$ 92.4)	GG (1214 $\pm$ 94.1)	0.0054	0.104
<b>rs31401</b>	T/G (T:0.48)	TT (1219 $\pm$ 95.4)	TG (1222 $\pm$ 94.4)	GG (1255 $\pm$ 81.8)	0.0356	0.327
<b>rs2073506</b>	T/C (T:0.19)	TT (1256 $\pm$ 87.9)	TC (1232 $\pm$ 90.9)	CC (1226 $\pm$ 96.0)	1.0	0.152
<b>rs181781</b>	A/G (A:0.30)	AA (1269 $\pm$ 100)	AG (1234 $\pm$ 86.1)	GG (1219 $\pm$ 98.3)	0.0736	0.390
<i>rs31480</i>	C/T (G:0.49)	TT (1257 $\pm$ 92.1)	TC (1224 $\pm$ 89.2)	CC (1216 $\pm$ 98.1)	0.0070	0.223

<sup>a</sup> SNPs genotyped in this study are shown in bold

<sup>b</sup> Minor allele, and the frequency of each marker is listed in brackets

<sup>c</sup> Bonferroni corrected *P* values are from female subjects only

<sup>d</sup> Association statistics (uncorrected *P* values) of the studied SNPs in male subjects are also listed. It should be noted that the average cranial volumes listed in this table are from females only

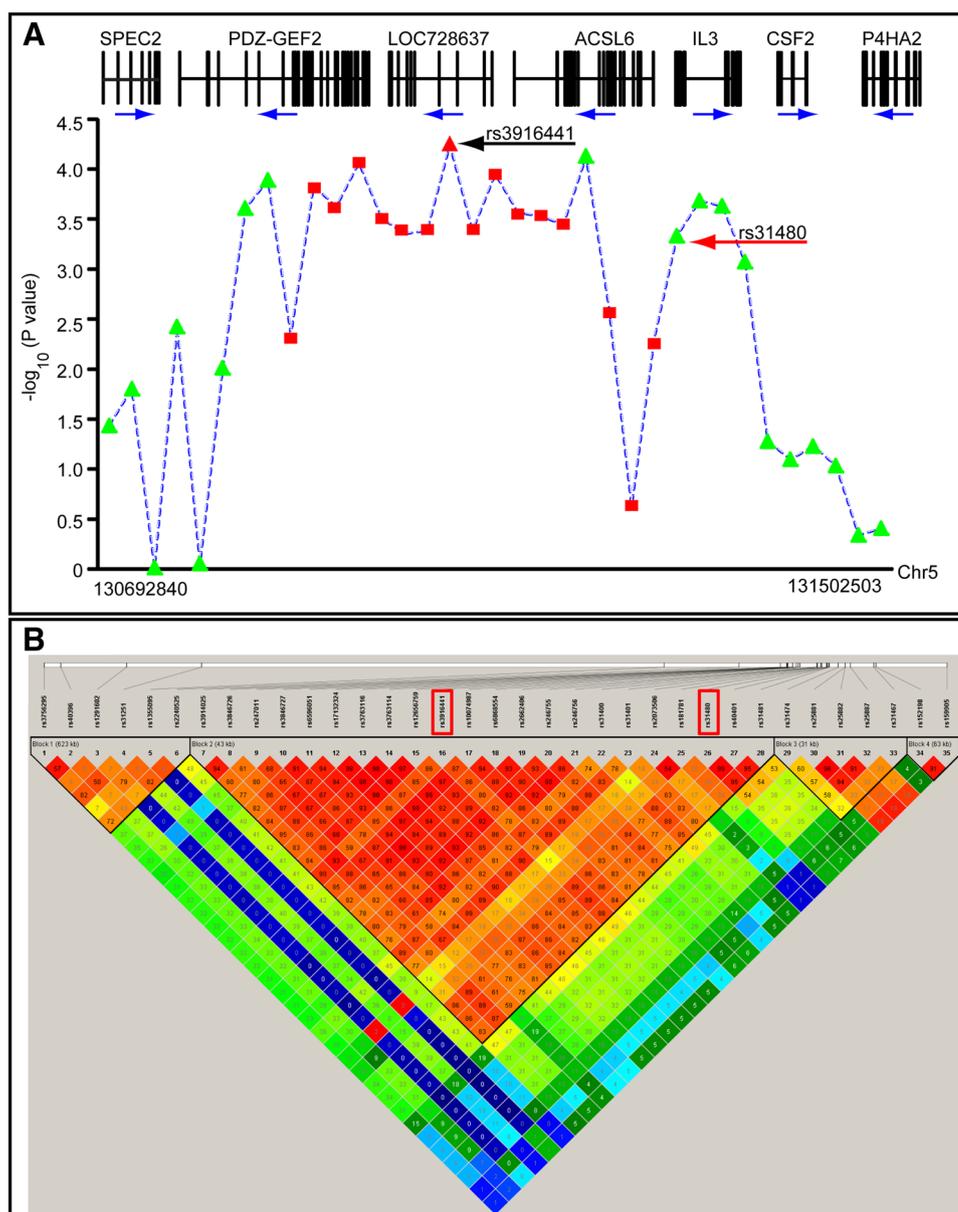
three SNPs (rs31480, located in the core promoter region of *IL3* [16 bp upstream of the transcription start site (TSS)], rs40401, located in exon 1 of *IL3*, and rs31481, located in intron 2 of *IL3*) were located in regulatory or coding region of *IL3*, suggesting that *IL3* may represent the responsible gene. In addition, our previous study showed that *IL3* was continuously expressed in developing mouse brain and can regulate proliferation and survival of neural progenitor cells. Consistently, previous investigations also revealed the important roles of *IL3* in the central nervous system. *IL3* acts as a pivotal neurotrophic factor (Kamegai et al. 1990) and neuroprotectant (Wen et al. 1998; Zambrano et al. 2007) for microglia cells and neurons. It could promote the growth and proliferation of microglial cells (Frei et al. 1985, 1986), enhance survival and differentiation of different neuronal populations (Moroni and Rossi 1995), and prevent delayed neuronal death in the hippocampus (Wen et al. 1998). The study of Cockayne et al. provided further evidence for the involvement of *IL3* in brain function. They showed that disruption of *IL3* production in brain led to neurologic dysfunction (Cockayne et al. 1994). Interestingly, Chen et al. (2007) also found that *IL3* is associated with schizophrenia in females, which further supports that *IL3* may play important roles in the central nervous system. Of note, both Chen et al. and our findings revealed the sex-specific association (that is, the association is only found

in females) of *IL3* with schizophrenia and brain volume. These sex-specific associations are quite intriguing and suggest that the interaction between *IL3* and gender may play important roles in brain development and disease susceptibility. By contrast, the role of *ACSL6* and *CSF2* in the central nervous system remains largely unknown. Though we could not exclude *ACSL6* and *CSF2* completely, current evidence suggests that *IL3* is more likely to be the responsible candidate gene in this region that contributes to brain volume variation in Chinese. We therefore focused on *IL3* in this study. Further work is needed to clarify if *IL3* is the true causal gene in this region.

#### Functional prediction supports that rs31480 may represent a functional SNP

In addition to rs31480, multiple other SNPs (e.g., rs3916441 and rs31400) also showed significant association with brain volume (corrected *P* value <0.05) (Fig. 1a). To test if these significant SNPs have potential functional consequences, we performed functional prediction using GWAVA ([http://www.sanger.ac.uk/sanger/StatGen\\_Gwava](http://www.sanger.ac.uk/sanger/StatGen_Gwava)) (Ritchie et al. 2014), SNPinfo (<http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm>) (Xu and Taylor 2009) and RegulomeDB (<http://www.regulomedb.org>) (Boyle et al. 2012). Consistent with our previous results, GWAVA analyses

**Fig. 1** Association significance between genetic variants on 5q22-33 and brain volume variation in the Chinese population. **a** The genomic locations and the  $P$  values of the studied SNPs. The newly genotyped SNPs are marked by *red* and the SNPs genotyped in our previous study (Luo et al. 2012) are marked by *green*. **b** The linkage disequilibrium pattern of the studied SNPs. LD values [correlation coefficient ( $r^2$ )] are shown and LD blocks were defined by the method of Gabriel et al. (2002). The orientation of *arrows* indicate the direction of gene transcription. rs31480 and rs3916441 are marked by *red boxes*. Note that the significant SNPs are highly linked

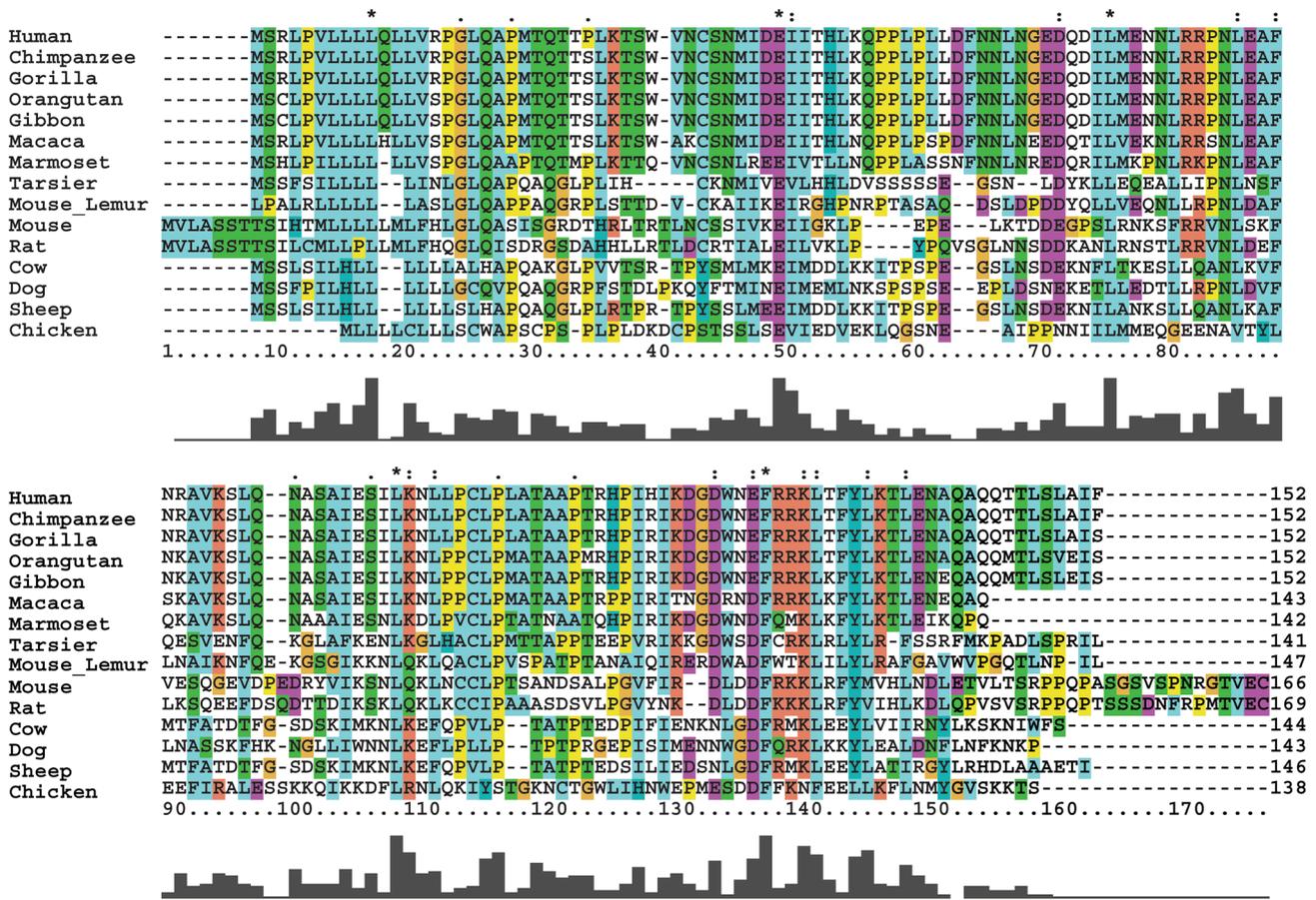


revealed that rs31480 has the highest score (Supplementary Table S2), which further support that rs31480 is likely to be the authentic functional variant among these SNPs. To further validate this result, we used another tool (SNPinfo) to predict the potential functional consequences of the studied SNPs. Again, we found that rs31480 has the highest regulatory potential among the 35 studied SNPs (Supplementary Table S3). In addition, RegulomeDB analyses also support that rs31480 is more likely to be a functional variant than rs3916441 (Supplementary Table S4). Finally, rs31480 is located in the promoter region, an active regulatory region that is crucial for expression regulation. We have shown that rs31480 is likely to be a functional SNP in our previous study (Luo et al. 2012). Rs31480 is located

in a putative transcription factor (SP1) binding site of *IL3* promoter and allelic difference at rs31480 could change SP1 binding affinity and influence *IL3* expression. Taken together, these convergent lines of evidence support that rs31480 may represent a functional SNP.

### Positive selection of *IL3* gene in primates and humans

Our fine-mapping results support that genetic variants near *IL3* gene contribute to brain volume variation in humans. As a matter of fact, our previous work has shown the important role of *IL3* in the central nervous system. We found that *IL3* and its receptors were continuously expressed in developing mouse brain and *IL3* could promote proliferation and

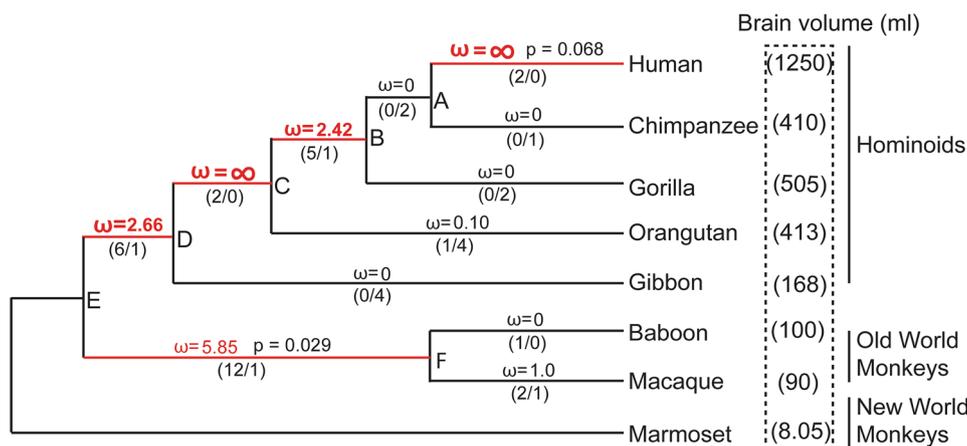


**Fig. 2** Protein sequence alignment of IL3. The IL3 protein sequence is highly divergent, with only five amino acid sites (marked with asterisks) identical among the 15 analyzed species

survival of neural progenitor cells (Luo et al. 2012). Based on the pivotal role of brain size in the origin and evolution of human species, we conjectured that *IL3* might have experienced positive selection. To test whether positive selection acted on *IL3*, we first analyzed the conservation of *IL3* protein sequence through multiple sequence alignment. We found that the protein sequence of *IL3* is highly divergent among the analyzed species, with only five amino sites being identical among the 15 analyzed species (Fig. 2). Positive selection or random mutations could result in this highly divergent protein sequence. To test if the dramatic diversity of protein sequence is attributed to positive selection or random mutations, we investigated the molecular evolutionary pattern of *IL3* in primates.

The phylogenetic relationships of primates are well established (Enard and Paabo 2004; Goodman et al. 1998; Siepel 2009). With the use of the well-characterized phylogeny of primates and the coding sequence of *IL3* from eight primates, we calculated the ratio of non-synonymous ( $K_a$ ) over synonymous ( $K_s$ ) nucleotide substitutions (referred to as  $\omega$ ) (Librado and Rozas 2009). We found

that the  $\omega$  value is highly variable among the studied primates (Fig. 3). For example, 12 non-synonymous substitutions were observed in the common ancestor of baboon and macaque (node F in Fig. 3). However, we noticed that there is only one synonymous substitution on this lineage. The  $\omega$  ratio (5.85) of this lineage is far greater than one ( $P < 0.05$ ). In addition, we observed high  $\omega$  ratios between nodes D and E, nodes D and C, and nodes C and B (Fig. 3). By using the  $Z$  test implemented in MEGA5 (Tamura et al. 2011), we tested if the positive selection acted on the different evolutionary lineages (Fig. 3). The  $\omega$  ratio of the common ancestor of baboon and macaque is significantly greater than one ( $P = 0.0029$ ), indicating potential positive selection on this lineage. In addition, we detected two non-synonymous substitutions and zero synonymous substitution on the human lineage, which leads to an infinite  $\omega$  ratio on human lineage. Further  $Z$  test suggested a trend of significant positive selection on the human lineage ( $P = 0.068$ ). Interestingly, we noticed that except for the common ancestor of human and chimpanzee, the other common ancestors leading to human lineage have exceptionally large  $\omega$  ratios (e.g.,  $\omega$



**Fig. 3** The ratios of non-synonymous ( $Ka$ ) over synonymous ( $Ks$ ) nucleotide substitutions ( $\omega$ ) in primates. The coding sequences of *IL3* were downloaded and aligned with ClustalX (Larkin et al. 2007). Only the coding sequences were used for  $Ka/Ks$  calculation with Pamilo–Bianchi–Li’s method or DnaSP (Li 1993; Librado and Rozas

2009; Pamilo and Bianchi 1993). We used the one-tailed Z test to test if the  $Ka/Ks$  ratios deviated from neutral expectation ( $Ka/Ks = 1$ ) (Tamura et al. 2011). The numbers of non-synonymous and synonymous substitution were shown in parentheses. The labels A–F are inferred internal nodes

**Table 2** Likelihood ratio tests for positive selection on the *IL3* gene

Model	Likelihood	df	$\chi^2$	Models compared	P value	Positively selected sites
One ratio	−1089.32					
Free ratio	−1070.94	12	36.74	One ratio vs. free ratio	<b>0.0002</b>	
One ratio	−1089.32					
Two ratio <sup>a</sup>	−1086.16	1	6.32	One ratio vs. two ratio	<b>0.012</b>	
Two ratio	−1088.54			Fix omega = 1; omega = 1		
Two ratio <sup>a</sup>	−1086.00	1	5.084	Fix omega = 0; omega = 1.5	<b>0.024</b>	
M1	−1087.44					
M2	−1083.90	2	7.08	Site model (M1 vs. M2)	<b>0.029</b>	3R*, 64D*
M0	−1089.32					
M3	−1083.90	4	10.84	Site model (M0 vs. M3)	<b>0.028</b>	

The numeric numbers corresponding to the following primate species: 1-human; 2-chimpanzee; 3-gorilla; 4-orangutan; 5-gibbon; 6-baboon; 7-macaca; 8-marmoset. Please refer to Fig. 3 for the phylogenetic tree

\* Denotes posterior probabilities >90 %. Significant P values are shown in bold

<sup>a</sup> Tree labels of the two-ratio model are as follows: (((((1,2),3)#1,4),5)#1,(6,7)#1,8))

ratio between nodes E and D, nodes D and C, and nodes C and B) (Fig. 3). These results implied possible adaptive evolution of *IL3* in primates and humans.

The highly divergent protein sequences and variable  $\omega$  ratios among different primate lineages suggested the potential positive selection of *IL3* in primates and humans. To further validate the possible adaptive evolution of *IL3* in primates, we performed codon-based neutrality tests developed by Yang et al. (1998). We first tested if there were variable  $\omega$  ratios among the eight studied primates. The likelihood values of one-ratio model (assumes that the  $\omega$  ratio is the same among all of the studied species) and free-ratio model (which assumes a different  $\omega$  ratio

for each lineage in the tree) are −1089.32 and −1070.94, respectively (Table 2). Likelihood ratio test indicated that the  $\omega$  ratios are indeed significant different among lineages ( $\chi^2 = 36.74$ ,  $P = 0.0002$  with  $df = 12$ ) (Table 2). Considering that the excess of non-synonymous substitutions over synonymous among the common ancestor of baboon and macaque (12:1), the nodes between E and D (6:1), the nodes between D and C (2:0), the nodes between C and B (5:1), and the human lineage (2:0), we further tested whether these branches have different  $\omega$  ratios. Comparison of the one-ratio model with the two-ratio model revealed that the  $\omega$  ratio of these branches are indeed different from other branches ( $P = 0.012$ ) (Table 2). We

**Table 3** Neutrality tests of the sequenced region [5276 base pairs were sequenced, including entire *IL3* gene (2675 bp), upstream region (1435 bp), and downstream region (1166 bp)] in human populations

IL3 (5276 bp)	Nucleotide diversity	Tajima's D	Fu and Li's D	Fu and Li's F	Fu and Li's D*	Fu and Li's F*	Fay and Wu's H
CHB (332)	0.00045	-1.717 <sup>#</sup>	-5.668**	-5.668**	-5.362**	-5.362**	0.565
CEU (176)	0.00022	-1.686 <sup>#</sup>	-4.727**	-4.261**	-4.874**	-4.361**	-0.245
YRI (176)	0.00066	-1.441	0.1492	-0.6726	-0.1127	-0.8266	-3.062
CHB + CEU (508)	0.00040	-1.968*	-7.766**	-6.084**	-7.450**	-5.912**	0.670

The sample size refers to the number of chromosomes sequenced

<sup>#</sup>  $P < 0.10$ , \*  $P < 0.05$ , \*\*  $P < 0.02$ , \*\*\*  $P < 0.01$

further examined whether the  $\omega$  ratio of these branches (i.e., the common ancestor of baboon and macaque, the nodes between E and D, the nodes between D and C, the nodes between C and B, and the human lineage) is significantly greater than one. Likelihood ratio test showed that the  $\omega$  ratio of these branches is significantly larger than one ( $P = 0.024$ ) (Table 2). Random effects likelihood (branch-site REL) method (Kosakovsky Pond et al. 2011) was also used to detect potential positive selection. Similarly, the results of branch-site REL showed that the common ancestor of baboon and macaque (node F), node E, and human lineage might have undergone positive selection (Supplementary Figure S1). Statistical tests suggested that marmoset and node F (the common ancestor of baboon and macaque) might have experienced significant positive selection (uncorrected  $P < 0.0001$  for marmoset and  $P = 0.01$  for node F). Exceptionally high  $Ka/Ks$  ratio in the common ancestor of baboon and macaque was observed. Considering the small brain size of the baboon and macaque, the high  $Ka/Ks$  ratio may be related to their relative brain size. We therefore examined the encephalization quotient (EQ, an indicator of relative brain size) of the studied species using published literatures (Aiello and Dean 1990; Roth and Dicke 2005). Humans have the highest EQ (Supplementary Table S5). However, other studied primate species have similar EQ, suggesting that the high  $Ka/Ks$  ratio in the common ancestor of baboon and macaque is not related to their relative brain size. Finally, we conducted codon-based neutrality tests to detect positive selection at individual amino acid site of *IL3* (Yang et al. 2000; Yang and Swanson 2002). Likelihood ratio test indicated the potential positive selection among the amino acid sites (comparing model M1 with M2 resulted in a  $P$  value of 0.029) (Table 2). Bayes empirical Bayes inference (Yang et al. 2005) suggested that positive selection acted on 3R and 64D amino acid sites (posterior probabilities that positive selection acted on 3R and 64D are 0.91 and 0.92, respectively) (Table 2). Collectively, these results suggested that *IL3* gene might have undergone positive selection in primates and humans.

### Positive selection of *IL3* gene in human populations

Phylogenetic-based neutrality tests revealed potential positive selection of *IL3* in humans and primates. We next asked whether *IL3* experienced positive selection in human populations through studying the nucleotide variation near the *IL3* gene in modern human populations. We sequenced a 5276 bp region encompassing the entire *IL3* gene in three representative world populations, including 88 Europeans, 166 Chinese, and 88 Africans. Frequency spectrum-based neutrality tests (e.g., Tajima's D, Fu and Li's D, F, Fu and Li's D\*, F\*, Fay and Wu's H) (Fay and Wu 2000; Fu 1997; Fu and Li 1993; Tajima 1989) were then conducted to test the deviation from neutral evolution. The values of Tajima's D, Fu and Li's D, F, D\*, F\*, Fay and Wu's H were summarized in Table 3. Using coalescent simulations (Schaffner et al. 2005), we further assessed whether these population statistics significantly deviated from neutral expectations. A negative value of Tajima's D was observed in three studied populations (Table 3), suggesting potential positive selection. In addition to Tajima's D, we found that the values of Fu and Li's D, F, D\*, F\* were negative and significantly deviated from neutrality in Chinese and Europeans (Table 3), suggesting that *IL3* gene might have experienced potential positive selection. We further characterized the non-synonymous and synonymous mutations observed in the *IL3* coding region among the sequenced populations (Table 4). The number of non-synonymous mutations observed in the Chinese, European, and African populations are 2, 1, 2. No synonymous substitutions were observed in the Chinese and European populations and one synonymous substitution was detected in the African population (Table 4).

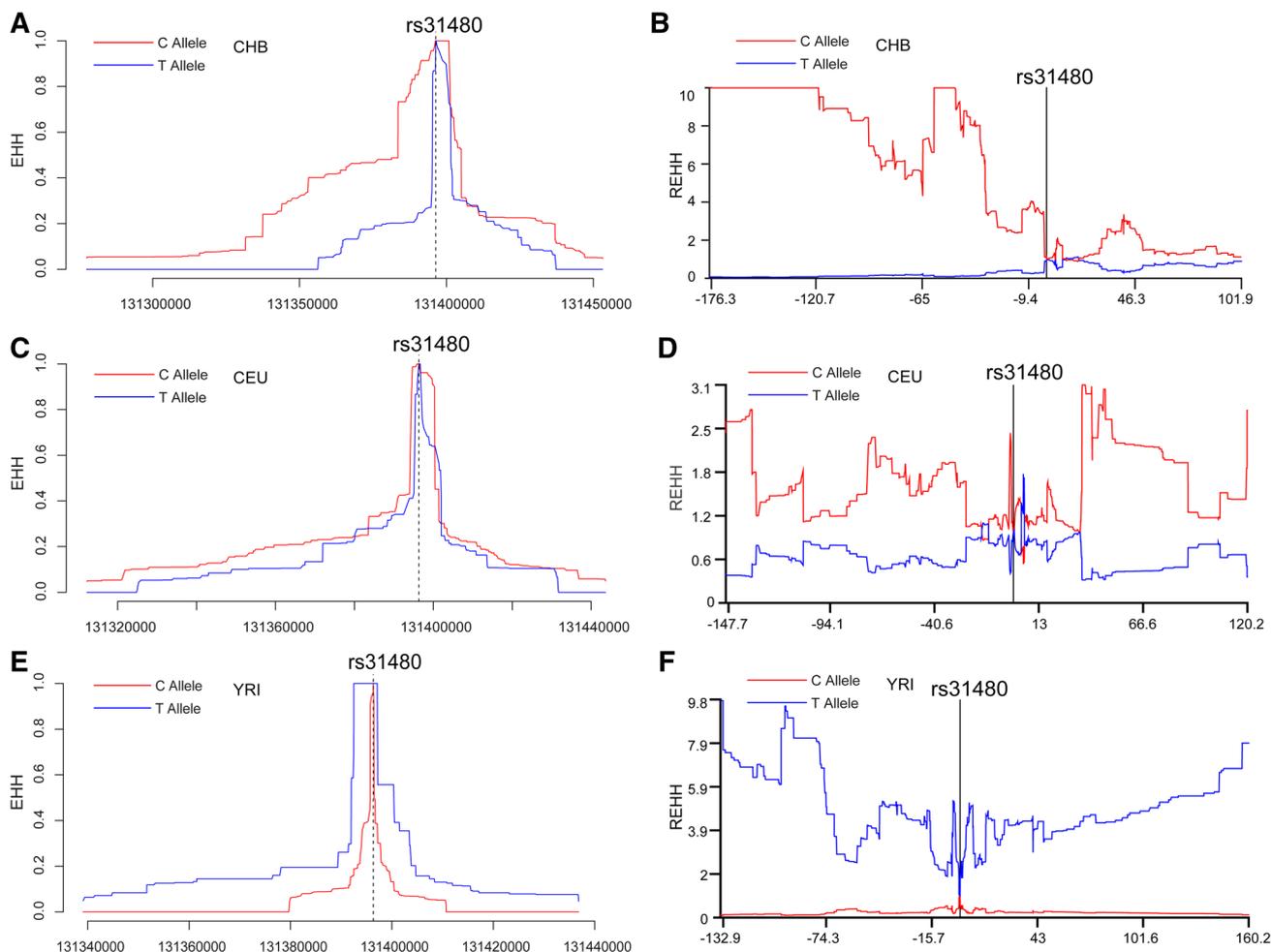
### Recent positive selection of rs31480 in the Chinese population

Rs31480 is a functional SNP that is significantly associated with brain volume variation in humans, suggesting that it may play an important role in human evolution. To test whether rs31480 experienced recent positive selection, we

**Table 4** The number of non-synonymous and synonymous (silent) mutations observed in *IL3* coding region in human populations

Mutation type	CHB ( <i>N</i> = 332)	CEU ( <i>N</i> = 176)	YRI ( <i>N</i> = 176)
Number of non-synonymous	2 (27 <sup>Ser→Pro</sup> , 128 <sup>Arg→Met</sup> ) <sup>a</sup>	1 (27 <sup>Ser→Pro</sup> )	2 (3 <sup>Arg→Cys</sup> , 27 <sup>Ser→Pro</sup> )
Number of synonymous	0	0	1
Number of haplotypes	3	2	4
Haplotype diversity	0.504	0.320	0.445

<sup>a</sup> The number in the parentheses shows the position of the non-synonymous mutations observed in *IL3* coding region, amino acids replacements are also showed

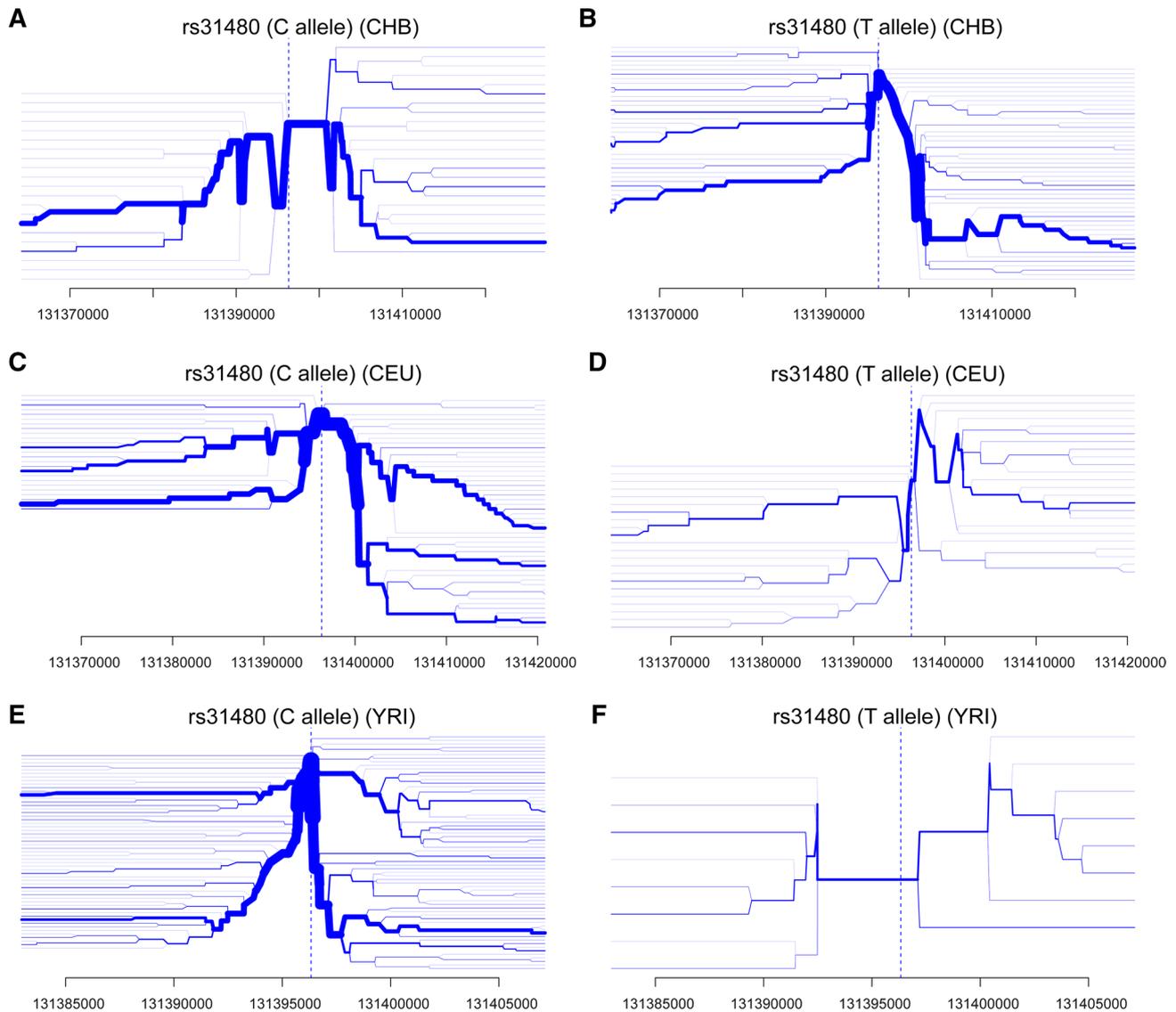


**Fig. 4** EHH and REHH analysis of rs31480 in Chinese, Europeans and Africans. Decay of extended haplotype homozygosity (EHH) for the C allele of rs31480 in three representative world populations (**a** CHB, Han Chinese; **c** CEU, Europeans; **e** YRI, Africans) over physical distance. In the Chinese population, the decay of haplotype homozygosity for the C allele (marked with red) occurs much more slowly than for the T allele (blue), suggesting potential positive selection acting on haplotypes containing this allele. In Europeans and Africans, the decay of haplotype homozygosity for the C allele is

similar to the T allele (**c**, **e**). **b**, **d**, **f** REHH plots of the Chinese population (**b**), European population (**d**), and African population (**f**). The decay of EHH for the C allele and T allele is not significant in the three populations ( $P > 0.05$ ). However, REHH results revealed that the haplotypes with the C allele decay significantly slower than the haplotypes with the T allele ( $P = 0.033$ ) in the Chinese population. REHH results are not statistically significant for the C allele and T allele in Europeans and Africans

conducted extended haplotype homozygosity (EHH) analyses using genotype data from the 1000 human genomes project (The 1000 Genomes Project Consortium 2010).

The EHH tests showed that the C allele of rs31480 might have undergone recent positive selection in Chinese population (Fig. 4a). The rate of EHH decay on the C allele is



**Fig. 5** Bifurcation plot of rs31480 in Chinese, Europeans and Africans. Haplotype bifurcation graphs show that the bifurcation of haplotypes is also slower for the C allele (**a**) of rs31480 compared with the T allele (**b**) in the Chinese population. In Europeans (CEU, **c, d**) and Africans (YRI, **e, f**), bifurcation of haplotypes for the C allele and T allele are comparable, indicating no recent positive selection on rs31480 in Europeans and Africans. The haplotype thicknesses rep-

resent the haplotype frequency. Haplotypes with unusually high EHH and a high population frequency (thicknesses) suggest the presence of recent positive selection that drives rapid increase in frequency of new variants or haplotypes in the population. Unusually long and thick haplotypes are observed for the C allele of rs31480 in Chinese (**a**), but not for the T allele, indicating recent positive selection of the C allele

much slower than the T allele, implying potential positive selection. We also conducted REHH tests and observed similar results (Fig. 4b). The EHH results showed that the haplotypes with rs31480-C allele decay slower than the haplotypes with T allele (Fig. 4a) in the Chinese population, though the difference is not statistically significant ( $P = 0.084$ ). However, REHH results clearly revealed that the haplotypes with C allele decay significantly slower than the haplotypes with T allele in Chinese ( $P = 0.033$ ). The rate of EHH decay on the rs31480-C allele and T allele is

similar in Europeans and Africans ( $P > 0.05$ ) (Fig. 4c, e). Statistical tests indicate that the decay of EHH and REHH on the C allele is not statistically different from the T allele in Europeans and Africans ( $P > 0.05$ ) (Fig. 4d, f). We further plotted the haplotype bifurcation diagram and observed similar results (Fig. 5). The results of haplotype bifurcation plot demonstrated clear long-range LD (as seen by the predominance of one thick branch in the haplotype carrying derived allele) for the haplotypes carrying the derived allele compared with the ancestral allele in the Chinese

population (Fig. 5a, b). Intriguingly, we noticed that EHH and haplotype bifurcation decays at similar rates for both C and T alleles in Europeans and Africans, indicating that the C allele of rs1480 did not undergo recent positive selection in Europeans and Africans (Fig. 5c–f). Collectively, these results implied that the C allele of rs1480 might have experienced weak or modest recent positive selection in the Chinese population.

## Discussion

Human evolution is characterized by greatly expanded brain volume and highly developed cognitive abilities. It is well established that dramatically enlarged brain volume plays a key role in the origin of human-specific traits, including wisdom, intelligence, and creativity. As a complex quantitative trait with high heritability, human brain volume is also highly variable in general human populations. Yet, the genetic mechanisms underlying human brain volume variation are largely unknown. Considering that human brain volume is correlated with cognitive function, including general intelligence, working memory, and processing speed, it is important to uncover the genetic mechanisms underlying human brain volume variation. We previously showed that genetic variants on 5q33 were significantly associated with brain volume in the Chinese population. Of note, we have successfully replicated the association between genetic variants on 5q33 and brain volume in two independent samples (European populations) in our previous study. Fine-mapping and functional assays demonstrated that *IL3* was likely the responsible gene. *IL3* and its receptors are continuously expressed in developing mouse brain and *IL3* could promote proliferation and survival of neural progenitors (Luo et al. 2012). In this study, we further detailed the association between genetic variants on 5q33 and human brain volume through genotyping additional 15 SNPs and our results are consistent with previous findings.

*IL3* encodes interleukin-3, a 152 amino acids protein that plays a pivotal role in the development of multiple cell types, including multipotent hematopoietic stem cells and a variety of cell types originating in the bone marrow (Dorssers et al. 1987; Yang et al. 1986). *IL3* exerts its biological effects through binding to its receptors. Ligand-specific alpha subunit (IL3RA) and signal transducing beta subunit (IL3RB) constitute the receptor complex of IL3 (Stomski et al. 1996). Three signaling pathways are activated by IL3, the mitogen-activated protein kinases (MAKP), the Janus kinase/signal transducers and activators of transcription (JAK/STAT), and the phosphatidylinositol 3-kinase-Akt (PI3K-AKT) signaling pathways (Martinez-Moczygemba and Huston 2003). Recent studies have revealed that *IL3*

also has a crucial role in the central nervous system (Tabira et al. 1998). *IL3* was found to be expressed in the hippocampal and cortical neurons of mouse brain (Konishi et al. 1994) and could promote the growth and proliferation of microglial cells (Frei et al. 1985, 1986), a type of glial cells that act as the first and main form of active immune defense in the central nervous system. In addition to its important role in glial cells, *IL3* also plays a pivotal role in neurons (Kamegai et al. 1990; Wen et al. 1998; Zambrano et al. 2007). It can promote the extension of process in cultured cholinergic neurons and the formation of neuronal network (Moroni and Rossi 1995). Consistent with the role of *IL3* in glial cells and neurons, loss of *IL3* resulted in neurologic disorder (Cockayne et al. 1994). In addition to these reported functions, we demonstrated a novel role of *IL3* in the central nervous system. We found that *IL3* could promote the neuronal survival and proliferation of neural progenitors (Luo et al. 2012).

To the best of our knowledge, this is the first study that systematically investigated the molecular evolutionary pattern of the *IL3* gene. Our results demonstrated that *IL3* might have experienced weak or modest positive selection in the evolution history of human beings. First, multiple protein sequence alignment showed that of *IL3* was highly variable among the studied species. Second, codon-based neutrality tests indicated that the ratio of non-synonymous to synonymous substitution ( $\omega$  ratios) is different among the studied primates, suggesting potential positive selection. Third, we also found variable  $\omega$  ratios among different amino sites, an indicator of positive selection. Fourth, frequency spectrum-based neutrality tests showed a significant departure from neutral substitution patterns, indicating a possible positive selection. Collectively, these consistent results suggest that *IL3* might have undergone weak or modest adaptive evolution in humans.

The results of our molecular evolutionary analyses revealed that *IL3* might have undergone positive selection in primates and humans. The association between *IL3* and brain volume (or the important role of *IL3* in the central nervous system) is one of the possible driving forces for the positive selection detected on *IL3*. A growing body of evidence strongly suggests that *IL3* is a multi-faceted factor that plays crucial roles in the central nervous system. As a matter of fact, previous studies have consistently shown accelerated evolution of the nervous system genes in the origin of humans (Dorus et al. 2004; Evans et al. 2004; Montgomery et al. 2011). However, considering that the signal of positive selection is widespread in apes and old world monkeys (Fig. 3), we could not exclude other factors completely. As an important member of the cytokine family, *IL3* also plays crucial roles in the immune system (Liew et al. 1989; Nishinakamura et al. 1995; Willinger et al. 2011). Of note, a recent study found that *IL3*

amplifies acute inflammation and is a potential therapeutic target in sepsis (Weber et al. 2015). Immune-related genes are important targets for positive selection and previous investigations have frequently reported adaptive evolution of immune-related genes (Bamshad and Wooding 2003; Barreiro and Quintana-Murci 2010; Karlsson et al. 2014). Therefore, the important roles of *IL3* in immune defense is another possible driving force for the positive selection detected on *IL3*. More work is needed to further clarify the most possible driving force for the positive selection detected on *IL3*.

EHH and REHH results suggested that *IL3* might have experienced recent positive selection in Asian populations. However, we did not detect signals of recent positive selection in Europeans and Africans. A possible explanation for this population-specific positive selection is that *IL3* may not be associated with brain volume in non-Chinese populations (e.g., Africans). That is, *IL3* may only be associated with brain volume in the Chinese population. As a matter of fact, we noticed that a proportion of brain volume-associated genetic variants identified in the European populations (Ikram et al. 2012; Stein et al. 2012) are monomorphic (fixed) in the Chinese population, suggesting that distinct genetic variants may contribute to brain volume variation in the Chinese population. In addition, other environmental factors such as diet, host–pathogen interactions (*IL3* also plays crucial role in immune defense), and living environment (e.g., temperature, humidity, and altitude) may also result in population-specific positive selection. Further work is needed to elucidate the underlying mechanisms of the Chinese-specific positive selection.

We observed sex-specific associations between genetic variants on 5q33 and brain volume variation, suggesting that genetic variants on 5q33 may have different effects on brain volume in females and males. To explore the possible reasons behind this sex-specific association, we previously investigated the interaction between *IL3* and estrogen, a hormone with higher concentration in females than in males. We found that the expression level of *IL3* was not directly regulated by estrogen. Nevertheless, we found that *IL3* can activate estrogen receptors. Of note, *IL3* and estrogen can operate synergistically to further enhance the activity of estrogen receptor beta. As the concentration of estrogen and its receptors in females are higher than in males, *IL3* expression level may have different effects on brain development (e.g., proliferation and survival of neural progenitors) in females and males. In males, though individuals with TT genotype at rs31480 have higher *IL3* expression than CC carriers, the concentration of estrogen and estrogen receptors are relatively low; therefore, the effects of *IL3* expression level on brain development are relative weak. By contrast, due to the high level of estrogen and its receptors, the synergistic interaction between *IL3* and

estrogen can amplify the effects of *IL3* greatly, resulting in differential activation of signaling pathways mediated by *IL3* and eventually leading to brain volume variation for individuals with different genotypes at rs31480 in females (Luo et al. 2012).

It should be noted that our EHH and REHH analyses detected no recent selection signal around rs31480 in Europeans and Africans. In Chinese, haplotypes with C allele decay slowly than the haplotypes with T allele, suggesting that recent positive selection might act on the C allele. These results imply that the selection on *IL3* is mild. Consistent with this, we found the selected C allele is also presented in the Neanderthal genome (Green et al. 2010). The absolute human brain volume increased dramatically over the past several million years after the separation of the human and chimpanzee species. It is possible that *IL3* experienced positive selection during the early stage of human evolution. That is, the positive selection might happen before the modern humans migrated out of Africa. Another explanation for the EHH results is that human brain volume might be under weak negative selection in recent human history (i.e., after the modern humans migrated out of Africa). Despite the dramatic increase of human brain volume after the separation of the human and chimpanzee species, fossil records implied that human brain volume decreased over the past 35,000 years (Henneberg 1988; Ruff et al. 1997). In addition, recent studies also reported the association between bigger brains and autism and fragile X syndrome (Hazlett et al. 2012; Nordahl et al. 2011), implying that the larger brain-associated genetic variants may also have potential deleterious effects. In summary, our findings revealed that genetic variants near *IL3* are associated with human brain volume variation and *IL3* might have experienced weak or modest positive selection in the evolutionary history of humans, which may be due to its contribution to human brain volume.

## Methods

### Measurement of brain volume

Brain volume measurements and calculation were performed as previously described (Wang et al. 2008). Briefly, 1013 unrelated healthy subjects were recruited in this study, including 460 males and 553 females. All of the participants were from southwestern China (Yunnan province), with the age range from 19 to 28 years. Informed consents were provided by all of the individuals. This study was approved by the internal review board of Kunming Institute of Zoology, Chinese Academy of Sciences. All the methods described here were conducted in accordance with the guidelines and regulations approved by the institutional

review board of Kunming Institute of Zoology, Chinese Academy of Sciences. More detailed information about the studied subjects, brain volume measurements, and calculation can be found in our previous studies (Wang et al. 2008).

### SNP selection and genotyping

Our previous study revealed that rs3916441 showed the most significant association with brain volume. To further explore if other SNPs near rs3916441 are associated with brain volume more significantly, we adopted a dense fine-scale mapping strategy. In total, we selected 15 SNPs. Seven of the newly selected SNPs were located upstream of rs3916441 and eight of them downstream of rs3916441. The detailed information of the newly selected SNPs is shown in Supplementary Table S1. SNP genotyping was performed using the SNaPshot method as described in our previous study (Luo et al. 2012). In brief, the region containing the specific SNP was amplified by polymerase chain reaction (PCR). The PCR products were purified and treated with shrimp alkaline phosphatase (SAP) and Exonuclease I. Then the SNP was genotyped by the single-base extension method, which was finished using 3730 DNA analyzer (Applied Biosystems). The GeneMapper software was used to read the genotyping results.

### Quantitative trait genetic association analysis

PLINK (Purcell et al. 2007) was utilized to perform the genetic association between the genotyped SNPs and brain volume. Briefly, for each SNP, 1 and 2 represent the homozygotes, respectively, 1.5 represents the heterozygotes, and cranial volume is considered to be dependent. Additive genetic model was used and linear regression was conducted as described previously (Wang et al. 2008). The association significance ( $P$  values) between the studied SNPs and brain volume was corrected using Bonferroni multiple testing correction, and the original and adjusted  $P$  values are listed in Supplementary Table S1.

### Linkage disequilibrium analysis

Linkage disequilibrium (correlation coefficient,  $r^2$ ) between the studied SNPs was analyzed using Haploview (Barrett et al. 2005). We defined the haplotype blocks using the criteria of Gabriel et al. (2002).

### Functional prediction of the studied SNPs

We previously have shown that rs31480 is likely to be a functional SNP using both bioinformatics analyses and functional experiments (i.e., electrophoretic mobility shift

and reporter gene assays). To further investigate if other SNPs have potential functional consequences, we conducted bioinformatics predictions using three different methods. We first performed functional prediction of the studied SNPs using Genome-Wide Annotation of Variants (GWAVA) ([http://www.sanger.ac.uk/sanger/StatGen\\_Gwava](http://www.sanger.ac.uk/sanger/StatGen_Gwava)) (Ritchie et al. 2014). GWAVA predicts the functional consequence of non-coding genetic variants based on a wide range of annotations from ENCODE/GENCODE, along with genome-wide properties such as evolutionary conservation and GC content. The GWAVA score ranges 0–1, with higher score indicating variants predicted as more likely to be functional. We then used another tool (SNPinfo) (<http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm>) (Xu and Taylor 2009) to predict the potential functional consequences of the studied SNPs. Finally, we predicted the potential functional consequences of rs3916441 and rs31480 using RegulomeDB (<http://www.regulomedb.org>) (Boyle et al. 2012). Multiple types of data (e.g., ChIP-seq, DNase-seq and eQTLs) from the Encyclopedia of DNA Elements (ENCODE) project and other sources were integrated into the RegulomeDB to estimate the possible function of the candidate SNPs.

### Multiple alignments of IL3 protein sequence

The protein sequence of IL3 was downloaded from NCBI (<http://www.ncbi.nlm.nih.gov/>), Ensembl (<http://www.ensembl.org/index.html>), and the UCSC genome browser (<http://genome.ucsc.edu/>). Protein sequence alignment was performed using ClustalX (Larkin et al. 2007).

### Likelihood ratio tests

The *IL3* coding sequences of the eight primates (human, chimpanzee, gorilla, orangutan, gibbon, baboon, macaque, and marmoset) were extracted from the NCBI, Ensembl, and UCSC genome browser. After aligning with ClustalX, the aligned sequences were used to reconstruct the ancestral sequences using *Baseml* program implemented in PAML (v4.7) (Yang 1997). The *Codeml* program in PAML software package was used to conduct the likelihood ratio tests. The branch model was used to detect if the ratio of non-synonymous substitution rate ( $K_a$ ) to synonymous substitution rate ( $K_s$ ) is significantly different among the studied species (Yang 1998). The site model was used to test if there were variable  $\omega$  ratios among the amino acid sites (Anisimova et al. 2002; Nielsen and Yang 1998; Yang et al. 2005). For the branch model, one ratio, two ratios, and free ratio branch models were used. For the site model, nested models M1 vs. M2 and M0 vs. M3 were compared to detect the potential positively selected sites. More detailed information on the likelihood ratio test, and branch

and site models can be found in the original publications (Yang 1998; Yang et al. 2000, 2005; Yang and Swanson 2002).

### Detecting positive selection with random effects likelihood (branch–site REL) method

Recently, Kosakovsky Pond et al. (2011) developed a new method (branch–site REL) to detect positive selection. Compared with the current branch–site model, which assumes that all branches in the tree can be partitioned a priori into two classes, branch–site REL introduced a new class of models in which substitution rates may vary from branch to branch and from site to site. Branch–site REL was used to detect the potential positive selection among primates.

### Population genetics analysis

To test if the evolution of *IL3* deviated from neutral expectation in human populations, we carried out population genetics analyses in three representative populations (Asians, Europeans, and Africans). A total of 342 individuals, including 166 Chinese, 88 Europeans, and 88 Africans, were recruited in this study. The genomic sequence encompassing the entire *IL3* gene (2675 bp), and upstream (1435 bp) and downstream (1166) sequences were sequenced using 3730 DNA analyzer (Applied Biosystems). We utilized the fast-PHASE program (Scheet and Stephens 2006) to infer the haplotypes based on the polymorphisms detected in the human populations. The population genetics parameters, including nucleotide diversity  $\pi$  and segregating sites  $\theta_w$ , were calculated in each population. We used the DnaSP (v5.0) (Librado and Rozas 2009) to conduct the frequency spectrum-based neutrality tests, including Tajima's D, Fu, and Li's D, F, Fu, and Li's D\*, F\*, Fay and Wu's H (Fay and Wu 2000; Fu 1997; Fu and Li 1993; Tajima 1989). For Fay and Wu's H test, coalescent simulations implemented in DnaSP were used to test if the sequenced region encompassing *IL3* significantly deviated from neutral evolution. We first calculated the values of Fay and Wu's H, theta and sample size based on the input sequences from each population. These values were then used for coalescent simulations. Detailed parameters used in coalescent simulations are described below: (1) Chinese population (CHB), Theta: 2.334, sample size: 332, Fay and Wu's H: 0.565; (2) European population (CEU), Theta: 1.153, sample size: 176, Fay and Wu's H: -0.245; (3) African population (YRI), Theta: 4.435, sample size: 176, Fay and Wu's H: -3.062. 10,000 simulations were conducted to obtain the empirical *P* value. Chimpanzee sequence was used

as outgroup for the test of Fu and Li's D, F, Fu and Li's D\*, F\*, Fay and Wu's H.

### Extended haplotype homozygosity (EHH) and relative EHH (REHH) analysis

To test if rs31480 experienced recent positive selection, we carried out EHH and REHH analyses (Sabeti et al. 2002). Genotype data (5:131196332–131596332, hg19) were obtained from the 1000 Genomes project (The 1000 Genomes Project Consortium 2010), including 103 Chinese (CHB), 99 Europeans (CEU), and 108 Africans (YRI). The extracted genotype data were phased with fastPHASE program (V1.2) (Scheet and Stephens 2006), and *Rehh* package (Gautier and Vitalis 2012) in R was used to calculate the EHH of the region surrounding the rs31480. The statistical significance of EHH and REHH was assessed using simulations implemented in Sweep program (V1.1) (Sabeti et al. 2002) (<http://www.broad-institute.org/mpg/sweep/>). We examined if the selected allele of rs31480 (C allele) is presented in archaic human genome using the Neanderthal genome (Green et al. 2010).

**Acknowledgments** XJL was supported by the 100 Talents Program (*BaiRenJiHua*) of the Kunming Institute of Zoology, Chinese Academy of Sciences, and Project of Thousand Youth Talents of China. YGY was supported by the Ministry of Science and Technology of China (2011CB910900) and the Strategic Priority Research Program (B) of the Chinese Academy of Sciences (XDB02020300). Chunhui Chen was supported by the National Natural Science Foundation of China (31100807). QD and Chuansheng Chen were supported by the 111 Project of the Ministry of Education of China (B07008).

### Compliance with ethical standards

**Conflict of interest** The authors declare no competing financial interests.

### References

- Aiello L, Dean C (1990) An introduction to human evolutionary anatomy. Academic Press, London
- Anisimova M, Bielawski JP, Yang Z (2002) Accuracy and power of bayes prediction of amino acid sites under positive selection. *Mol Biol Evol* 19:950–958
- Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. *Nat Rev Genet* 4:99–111
- Barreiro LB, Quintana-Murci L (2010) From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet* 11:17–30
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265
- Bond J, Roberts E, Mochida GH, Hampshire DJ, Scott S, Askham JM, Springell K, Mahadevan M, Crow YJ, Markham AF, Walsh CA, Woods CG (2002) ASPM is a major determinant of cerebral cortical size. *Nat Genet* 32:316–320

- Bond J, Scott S, Hampshire DJ, Springell K, Corry P, Abramowicz MJ, Mochida GH, Hennekam RC, Maher ER, Fryns JP, Alswaid A, Jafri H, Rashid Y, Mubaidin A, Walsh CA, Roberts E, Woods CG (2003) Protein-truncating mutations in ASPM cause variable reduction in brain size. *Am J Hum Genet* 73:1170–1177
- Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 22:1790–1797
- Chen X, Wang X, Hossain S, O'Neill FA, Walsh D, van den Oord E, Fanous A, Kendler KS (2007) Interleukin 3 and schizophrenia: the impact of sex and family history. *Mol Psychiatry* 12:273–282
- Cockayne DA, Bodine DM, Cline A, Nienhuis AW, Dunbar CE (1994) Transgenic mice expressing antisense interleukin-3 RNA develop a B-cell lymphoproliferative syndrome or neurologic dysfunction. *Blood* 84:2699–2710
- Dorssers L, Burger H, Bot F, Delwel R, Geurts van Kessel AH, Lowenberg B, Wagemaker G (1987) Characterization of a human multilineage-colony-stimulating factor cDNA clone identified by a conserved noncoding sequence in mouse interleukin-3. *Gene* 55:115–124
- Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M, Wyckoff GJ, Malcom CM, Lahn BT (2004) Accelerated evolution of nervous system genes in the origin of Homo sapiens. *Cell* 119:1027–1040
- Enard W, Paabo S (2004) Comparative primate genomics. *Annu Rev Genomics Hum Genet* 5:351–378
- Evans PD, Anderson JR, Vallender EJ, Gilbert SL, Malcom CM, Dorus S, Lahn BT (2004) Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. *Hum Mol Genet* 13:489–494
- Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413
- Frei K, Bodmer S, Schwerdel C, Fontana A (1985) Astrocytes of the brain synthesize interleukin 3-like factors. *J Immunol* 135:4044–4047
- Frei K, Bodmer S, Schwerdel C, Fontana A (1986) Astrocyte-derived interleukin 3 as a growth factor for microglia cells and peritoneal macrophages. *J Immunol* 137:3521–3527
- Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133:693–709
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225–2229
- Gautier M, Vitalis R (2012) rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28:1176–1177
- Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP (1998) Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. *Mol Phylogenet Evol* 9:585–598
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspina AS, Jensen JD, Marques-Bonet T, Alkan C, Prufer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Hober B, Hoffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova LV, Laluzza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PL, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Paabo S (2010) A draft sequence of the Neandertal genome. *Science* 328:710–722
- Hazlett HC, Poe MD, Lightbody AA, Styner M, MacFall JR, Reiss AL, Piven J (2012) Trajectories of early brain volume development in fragile X syndrome and autism. *J Am Acad Child Adolesc Psychiatry* 51:921–933
- Henneberg M (1988) Decrease of human skull size in the Holocene. *Hum Biol* 60:395–405
- Ikram MA, Fornage M, Smith AV, Seshadri S, Schmidt R, Debetto S, Vrooman HA, Sigurdsson S, Ropele S, Taal HR, Mook-Kanamori DO, Coker LH, Longstreth WT Jr, Niessen WJ, DeStefano AL, Beiser A, Zijdenbos AP, Struchalin M, Jack CR Jr, Rivadeneira F, Uitterlinden AG, Knopman DS, Hartikainen AL, Pennell CE, Thiering E, Steegers EA, Hakonarson H, Heinrich J, Palmer LJ, Jarvelin MR, McCarthy MI, Grant SF, St Pourcain B, Timpson NJ, Smith GD, Sovio U, Nalls MA, Au R, Hofman A, Gudnason H, van der Lugt A, Harris TB, Meeks WM, Vernooij MW, van Buchem MA, Catellier D, Jaddoe VW, Gudnason V, Windham BG, Wolf PA, van Duijn CM, Mosley TH Jr, Schmidt H, Launer LJ, Breteler MM, DeCarli C (2012) Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nat Genet* 44:539–544
- Jackson AP, McHale DP, Campbell DA, Jafri H, Rashid Y, Mannan J, Karbani G, Corry P, Levene MI, Mueller RF, Markham AF, Lench NJ, Woods CG (1998) Primary autosomal recessive microcephaly (MCPH1) maps to chromosome 8p22-pter. *Am J Hum Genet* 63:541–546
- Jackson AP, Eastwood H, Bell SM, Adu J, Toomes C, Carr IM, Roberts E, Hampshire DJ, Crow YJ, Mighell AJ, Karbani G, Jafri H, Rashid Y, Mueller RF, Markham AF, Woods CG (2002) Identification of microcephalin, a protein implicated in determining the size of the human brain. *Am J Hum Genet* 71:136–142
- Jamieson CR, Govaerts C, Abramowicz MJ (1999) Primary autosomal recessive microcephaly: homozygosity mapping of MCPH4 to chromosome 15. *Am J Hum Genet* 65:1465–1469
- Jamieson CR, Fryns JP, Jacobs J, Matthijs G, Abramowicz MJ (2000) Primary autosomal recessive microcephaly: MCPH5 maps to 1q25-q32. *Am J Hum Genet* 67:1575–1577
- Kamegai M, Nijijima K, Kunishita T, Nishizawa M, Ogawa M, Araki M, Ueki A, Konishi Y, Tabira T (1990) Interleukin 3 as a trophic factor for central cholinergic neurons in vitro and in vivo. *Neuron* 4:429–436
- Karlsson EK, Kwiatkowski DP, Sabeti PC (2014) Natural selection and infectious disease in human populations. *Nat Rev Genet* 15:379–393
- Konishi Y, Kamegai M, Takahashi K, Kunishita T, Tabira T (1994) Production of interleukin-3 by murine central nervous system neurons. *Neurosci Lett* 182:271–274
- Kosakovskiy SL, Murrell B, Fourment M, Frost SD, Delpont W, Scheffler K (2011) A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol* 28:3033–3043
- Kouprina N, Pavlicek A, Mochida GH, Solomon G, Gersch W, Yoon YH, Collura R, Ruvolo M, Barrett JC, Woods CG, Walsh CA, Jurka J, Larionov V (2004) Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. *PLoS Biol* 2:E126
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948
- Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* 36:96–99
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452

- Liew FY, Millott S, Li Y, Lelchuk R, Chan WL, Ziltener H (1989) Macrophage activation by interferon-gamma from host-protective T cells is inhibited by interleukin (IL)3 and IL4 produced by disease-promoting T cells in leishmaniasis. *Eur J Immunol* 19:1227–1232
- Luo XJ, Li M, Huang L, Nho K, Deng M, Chen Q, Weinberger DR, Vasquez AA, Rijpkema M, Mattay VS, Saykin AJ, Shen L, Fernandez G, Franke B, Chen JC, Chen XN, Wang JK, Xiao X, Qi XB, Xiang K, Peng YM, Cao XY, Li Y, Shi XD, Gan L, Su B (2012) The interleukin 3 gene (IL3) contributes to human brain volume variation by regulating proliferation and survival of neural progenitors. *PLoS One* 7:e50375
- Martinez-Moczygema M, Huston DP (2003) Biology of common beta receptor-signaling cytokines: IL-3, IL-5, and GM-CSF. *J Allergy Clin Immunol* 112:653–665 (quiz 666)
- Mekel-Bobrov N, Gilbert SL, Evans PD, Vallender EJ, Anderson JR, Hudson RR, Tishkoff SA, Lahn BT (2005) Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. *Science* 309:1720–1722
- Mochida GH, Walsh CA (2001) Molecular genetics of human microcephaly. *Curr Opin Neurol* 14:151–156
- Montgomery SH, Capellini I, Venditti C, Barton RA, Mundy NI (2011) Adaptive evolution of four microcephaly genes and the evolution of brain size in anthropoid primates. *Mol Biol Evol* 28:625–638
- Moroni SC, Rossi A (1995) Enhanced survival and differentiation in vitro of different neuronal populations by some interleukins. *Int J Dev Neurosci* 13:41–49
- Moynihan L, Jackson AP, Roberts E, Karbani G, Lewis I, Corry P, Turner G, Mueller RF, Lench NJ, Woods CG (2000) A third novel locus for primary autosomal recessive microcephaly maps to chromosome 9q34. *Am J Hum Genet* 66:724–727
- Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936
- Nishinakamura R, Nakayama N, Hirabayashi Y, Inoue T, Aud D, McNeil T, Azuma S, Yoshida S, Toyoda Y, Arai K et al (1995) Mice deficient for the IL-3/GM-CSF/IL-5 beta c receptor exhibit lung pathology and impaired immune response, while beta IL3 receptor-deficient mice are normal. *Immunity* 2:211–222
- Nordahl CW, Lange N, Li DD, Barnett LA, Lee A, Buonocore MH, Simon TJ, Rogers S, Ozonoff S, Amaral DG (2011) Brain enlargement is associated with regression in preschool-age boys with autism spectrum disorders. *Proc Natl Acad Sci USA* 108:20195–20200
- Pamilo P, Bianchi NO (1993) Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Mol Biol Evol* 10:271–281
- Pattison L, Crow YJ, Deebble VJ, Jackson AP, Jafri H, Rashid Y, Roberts E, Woods CG (2000) A fifth locus for primary autosomal recessive microcephaly maps to chromosome 1q31. *Am J Hum Genet* 67:1578–1580
- Peper JS, Brouwer RM, Boomsma DI, Kahn RS, Hulshoff Pol HE (2007) Genetic influences on human brain structure: a review of brain imaging studies in twins. *Hum Brain Mapp* 28:464–473
- Posthuma D, De Geus EJ, Baare WF, Hulshoff Pol HE, Kahn RS, Boomsma DI (2002) The association between brain volume and intelligence is of genetic origin. *Nat Neurosci* 5:83–84
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- Ritchie GR, Dunham I, Zeggini E, Flicek P (2014) Functional annotation of noncoding sequence variants. *Nat Methods* 11:294–296
- Roberts E, Jackson AP, Carradice AC, Deebble VJ, Mannan J, Rashid Y, Jafri H, McHale DP, Markham AF, Lench NJ, Woods CG (1999) The second locus for autosomal recessive primary microcephaly (MCPH2) maps to chromosome 19q13.1–13.2. *Eur J Hum Genet* 7:815–820
- Roth G, Dicke U (2005) Evolution of the brain and intelligence. *Trends Cogn Sci* 9:250–257
- Ruff CB, Trinkaus E, Holliday TW (1997) Body mass and encephalization in Pleistocene *Homo*. *Nature* 387:173–176
- Rushton JP (1992) Cranial capacity related to sex, rank, and race in a stratified random sample of 6,325 U.S. Military personnel. *Intelligence* 16:401–413
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837
- Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D (2005) Calibrating a coalescent simulation of human genome sequence variation. *Genome Res* 15:1576–1583
- Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* 78:629–644
- Siepel A (2009) Phylogenomics of primates and their ancestral populations. *Genome Res* 19:1929–1941
- Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM, Toro R, Appel K, Bartecek R, Bergmann O, Bernard M, Brown AA, Cannon DM, Chakravarty MM, Christoforou A, Domin M, Grimm O, Hollinshead M, Holmes AJ, Homuth G, Hottenga JJ, Langan C, Lopez LM, Hansell NK, Hwang KS, Kim S, Laje G, Lee PH, Liu X, Loth E, Lourdasamy A, Mattingdal M, Mohnke S, Maniega SM, Nho K, Nugent AC, O'Brien C, Pampmeyer M, Putz B, Ramasamy A, Rasmussen J, Rijpkema M, Risacher SL, Roddey JC, Rose EJ, Ryten M, Shen L, Sprooten E, Strengman E, Teumer A, Trabzuni D, Turner J, van Eijk K, van Erp TG, van Tol MJ, Wittfeld K, Wolf C, Woudstra S, Aleman A, Alhusaini S, Almasy L, Binder EB, Brohawn DG, Cantor RM, Carless MA, Corvin A, Czisch M, Curran JE, Davies G, de Almeida MA, Delanty N, Depondt C, Duggirala R, Dyer TD, Erk S, Fagerness J, Fox PT, Freimer NB, Gill M, Goring HH, Hagler DJ, Hoehn D, Holsboer F, Hoogman M, Hosten N, Jahanshad N, Johnson MP, Kasperaviciute D, Kent JW Jr, Kochunov P, Lancaster JL, Lawrie SM, Liewald DC, Mandl R, Matarin M, Mattheisen M, Meisenzahl E, Melle I, Moses EK, Muhleisen TW et al (2012) Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet* 44:552–561
- Stomski FC, Sun Q, Bagley CJ, Woodcock J, Goodall G, Andrews RK, Berndt MC, Lopez AF (1996) Human interleukin-3 (IL-3) induces disulfide-linked IL-3 receptor alpha- and beta-chain heterodimerization, which is required for receptor activation but not high-affinity binding. *Mol Cell Biol* 16:3035–3046
- Taal HR, St Pourcain B, Thiering E, Das S, Mook-Kanamori DO, Warrington NM, Kaakinen M, Kreiner-Moller E, Bradfield JP, Freathy RM, Geller F, Guxens M, Cousminer DL, Kerkhof M, Timpson NJ, Ikram MA, Beilin LJ, Bonnelykke K, Buxton JL, Charoen P, Chawes BL, Eriksson J, Evans DM, Hofman A, Kemp JP, Kim CE, Klopp N, Lahti J, Lye SJ, McMahon G, Mentch FD, Muller-Nurasyid M, O'Reilly PF, Prokopenko I, Rivadeneira F, Steegers EA, Sunyer J, Tiesler C, Yaghootkar H, Breteler MM, Decarli C, Debette S, Fornage M, Gudnason V, Launer LJ, van der Lugt A, Mosley TH Jr, Seshadri S, Smith AV, Vernooij MW, Blakemore AI, Chiavacci RM, Feenstra B, Fernandez-Banet J, Grant SF, Hartikainen AL, van der Heijden AJ, Iniguez C, Lathrop M, McArdle WL, Molgaard A, Newnham JP, Palmer LJ, Palotie A, Pouta A, Ring SM, Sovio U, Standl M,

- Uitterlinden AG, Wichmann HE, Vissing NH, van Duijn CM, McCarthy MI, Koppelman GH, Estivill X, Hattersley AT, Melbye M, Bisgaard H, Pennell CE, Widen E, Hakonarson H, Smith GD, Heinrich J, Jarvelin MR, Jaddoe VW (2012) Common variants at 12q15 and 12q24 are associated with infant head circumference. *Nat Genet* 44:532–538
- Tabira T, Chui DH, Fan JP, Shirabe T, Konishi Y (1998) Interleukin-3 and interleukin-3 receptors in the brain. *Ann NY Acad Sci* 840:107–116
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
- The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073
- Thompson PM, Cannon TD, Narr KL, van Erp T, Poutanen VP, Huttenen M, Lonnqvist J, Standertskjold-Nordenstam CG, Kaprio J, Khaledy M, Dail R, Zoumalan CI, Toga AW (2001) Genetic influences on brain structure. *Nat Neurosci* 4:1253–1258
- Thompson P, Cannon TD, Toga AW (2002) Mapping genetic influences on human brain structure. *Ann Med* 34:523–536
- Wang YQ, Su B (2004) Molecular evolution of microcephalin, a gene determining human brain size. *Hum Mol Genet* 13:1131–1137
- Wang JK, Li Y, Su B (2008) A common SNP of MCPH1 is associated with cranial volume variation in Chinese population. *Hum Mol Genet* 17:1329–1335
- Weber GF, Chousterman BG, He S, Fenn AM, Nairz M, Anzai A, Brenner T, Uhle F, Iwamoto Y, Robbins CS, Noiret L, Maier SL, Zonnchen T, Rahbari NN, Scholch S, Klotzsche-von Ameln A, Chavakis T, Weitz J, Hofer S, Weigand MA, Nahrendorf M, Weissleder R, Swirski FK (2015) Interleukin-3 amplifies acute inflammation and is a potential therapeutic target in sepsis. *Science* 347:1260–1265
- Wen TC, Tanaka J, Peng H, Desaki J, Matsuda S, Maeda N, Fujita H, Sato K, Sakanaka M (1998) Interleukin 3 prevents delayed neuronal death in the hippocampal CA1 field. *J Exp Med* 188:635–649
- Willinger T, Rongvaux A, Takizawa H, Yancopoulos GD, Valenzuela DM, Murphy AJ, Auerbach W, Eynon EE, Stevens S, Manz MG, Flavell RA (2011) Human IL-3/GM-CSF knock-in mice support human alveolar macrophage development and human immune responses in the lung. *Proc Natl Acad Sci USA* 108:2390–2395
- Woods CG, Bond J, Enard W (2005) Autosomal recessive primary microcephaly (MCPH): a review of clinical, molecular, and evolutionary findings. *Am J Hum Genet* 76:717–728
- Xu Z, Taylor JA (2009) SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res* 37:W600–W605
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556
- Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15:568–573
- Yang Z, Swanson WJ (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol Biol Evol* 19:49–57
- Yang YC, Ciarletta AB, Temple PA, Chung MP, Kovacic S, Witek-Giannotti JS, Leary AC, Kriz R, Donahue RE, Wong GG et al (1986) Human IL-3 (multi-CSF): identification by expression cloning of a novel hematopoietic growth factor related to murine IL-3. *Cell* 47:3–10
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449
- Yang Z, Wong WS, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22:1107–1118
- Zambrano A, Otth C, Mujica L, Concha II, Maccioni RB (2007) Interleukin-3 prevents neuronal death induced by amyloid peptide. *BMC Neurosci* 8:82
- Zhang J (2003) Evolution of the human ASPM gene, a major determinant of brain size. *Genetics* 165:2063–2070

## Supplementary Material

**Table S1. Marker characteristics and association significance with brain volume in females**

SNP id <sup>a</sup>	Location	Polymorphism	R <sup>2</sup>	T <sup>b</sup>	P value <sup>b</sup> (Original)	P value <sup>c</sup> (Corrected)
<i>rs3756295</i>	130720739	G/C	0.00888	-2.095	0.03665	1
<i>rs40396</i>	130735943	G/C	0.0120	2.423	0.01577	0.55195
<i>rs1291602</i>	130794561	G/A	4.04E-06	-0.0446	0.9644	1
<i>rs31251</i>	130861845	G/A	0.0169	-2.911	0.003772	0.132
<i>rs1355095</i>	131276668	G/A	4.96E-05	-0.1557	0.8763	1
<i>rs2240525</i>	131343783	C/T	0.0137	-2.596	0.00971	0.340
<i>rs3914025</i>	131381184	G/A	0.0277	3.693	0.000247	0.0087
<i>rs3846726</i>	131386898	G/A	0.0299	-3.86	0.000129	0.0045
<b>rs247011</b>	131387513	G/A	0.0157	-2.854	0.004491	0.1572
<b>rs3846727</b>	131387577	G/A	0.0278	3.824	0.000148	0.0052
<b>rs6596051</b>	131391836	C/T	0.0261	3.703	0.000237	0.00828
<b>rs17132324</b>	131391878	G/A	0.0298	-3.959	8.61E-05	0.00301
<b>rs3763116</b>	131391913	G/T	0.0253	3.632	0.00031	0.01083
<b>rs3763114</b>	131392232	A/T	0.0239	-3.535	0.000444	0.01555
<b>rs12656759</b>	131395509	C/T	0.0243	3.555	0.000413	0.01444
<i>rs3916441</i>	131397140	C/T	0.0319	-4.059	5.72E-05	0.00200
<b>rs10074987</b>	131398611	G/A	0.0241	3.555	0.000413	0.01447
<b>rs6868554</b>	131398726	G/T	0.0288	3.894	0.000112	0.00391
<b>rs2662406</b>	131414041	C/T	0.0256	3.665	0.000273	0.00957
<b>rs246755</b>	131414177	A/C	0.0254	-3.644	0.000296	0.01035
<b>rs246756</b>	131414258	C/T	0.0246	3.592	0.00036	0.01259
<i>rs31400</i>	131417406	G/A	0.0317	-3.995	7.45E-05	0.00261
<b>rs31401</b>	131418280	G/T	0.0179	-3.054	0.002373	0.08306
<b>rs2073506</b>	131422637	A/G	0.00344	1.195	0.2327	1
<b>rs181781</b>	131423014	A/G	0.0190	2.828	0.004906	0.1717
<i>rs31480</i>	131424231	G/A	0.0242	3.523	0.000465	0.0163
<i>rs40401</i>	131424377	G/A	0.0271	3.738	0.000207	0.0072
<i>rs31481</i>	131425101	C/T	0.0272	3.706	0.000234	0.0082
<i>rs31474</i>	131432926	G/A	0.0219	-3.357	0.000848	0.0297
<i>rs25881</i>	131439037	G/A	0.00763	1.941	0.05279	1
<i>rs25882</i>	131439395	G/A	0.00626	-1.754	0.0801	1
<i>rs25887</i>	131443960	G/T	0.00715	1.893	0.05893	1
<i>rs31467</i>	131464737	G/A	0.00564	1.687	0.09217	1
<i>rs152198</i>	131466709	G/A	0.00111	0.7448	0.4568	1
<i>rs159905</i>	131530402	G/A	0.00149	0.8606	0.3899	1

<sup>a</sup>SNPs genotyped in our previous study are shown in italic. The 15 newly SNPs genotyped in this study are shown in bold. <sup>b</sup>The test statistics *T* and *P* values were obtained by Plink v1.07 under the additive genetic model. <sup>c</sup>The *P* values were corrected by using Bonferroni multiple testing correction.

**Table S2. Functional prediction of the studied SNPs with GWAVA**  
[http://www.sanger.ac.uk/sanger/StatGen\\_Gwava](http://www.sanger.ac.uk/sanger/StatGen_Gwava)).

SNP ID	Chromosome	Position	GWAVA Prediction Score <sup>a</sup>
rs3756295	5	130,692,840	0.25
rs40396	5	130,708,044	0.18
rs1291602	5	130,766,662	0.39
rs31251	5	130,833,946	0.17
rs1355095	5	131,248,769	0.15
rs2240525	5	131,315,884	0.3
rs3914025	5	131,353,285	0.41
rs3846726	5	131,358,999	0.54
rs247011	5	131,359,614	0.38
rs3846727	5	131,359,678	0.46
rs6596051	5	131,363,937	0.29
rs17132324	5	131,363,979	0.32
rs3763116	5	131,364,014	0.28
rs3763114	5	131,364,333	0.39
rs12656759	5	131,367,610	0.38
<b>rs3916441</b>	<b>5</b>	<b>131,369,241</b>	<b>0.24</b>
rs10074987	5	131,370,712	0.15
rs6868554	5	131,370,827	0.16
rs2662406	5	131,386,142	0.29
rs246755	5	131,386,278	0.24
rs246756	5	131,386,359	0.25
rs31400	5	131,389,507	0.46
rs31401	5	131,390,381	0.54
rs2073506	5	131,394,738	0.46
rs181781	5	131,395,115	0.43
<b>rs31480</b>	<b>5</b>	<b>131,396,332</b>	<b>0.87</b>
rs40401	5	131,396,478	0.18
rs31481	5	131,397,202	0.2
rs31474	5	131,405,027	0.4
rs25881	5	131,411,138	0.3
rs25882	5	131,411,460	0.22
rs25887	5	131,416,061	0.16
rs31467	5	131,436,838	0.34
rs152198	5	131,438,810	0.4
rs159905	5	131,502,503	0.08

<sup>a</sup>GWAVA prediction score, higher score suggests higher possibility that the variant is functional.

**Table S3. Functional prediction of the studied SNPs using SNPinfo**<http://snpinfo.niehs.nih.gov/cgi-bin/snpinfo/snpfunc.cgi>.

SNP ID	Chromosome	Position	Allele	TFBS <sup>a</sup>	Regulatory Potential <sup>b</sup>
rs3756295	5	130720739	G/C	--	0
rs40396	5	130735943	G/C	--	NA
rs1291602	5	130794561	C/T	--	0.289489
rs31251	5	130861845	T/C	--	0.049836
rs1355095	5	131276668	T/C	--	0
rs2240525	5	131343783	A/G	--	0
rs3914025	5	131381184	C/T	--	0
rs3846726	5	131386898	T/C	--	NA
rs247011	5	131387513	G/A	--	NA
rs3846727	5	131387577	T/C	--	0
rs6596051	5	131391836	C/T	--	0.210977
rs17132324	5	131391878	A/G	--	0.176642
rs3763116	5	131391913	C/A	--	0.123661
rs3763114	5	131392232	T/A	--	0
rs12656759	5	131395509	C/T	--	NA
<b>rs3916441</b>	<b>5</b>	<b>131397140</b>	<b>A/G</b>	<b>--</b>	<b>NA</b>
rs10074987	5	131398611	A/G	--	NA
rs6868554	5	131398726	G/T	--	NA
rs2662406	5	131414041	G/A	--	0
rs246755	5	131414177	A/C	--	0
rs246756	5	131414258	C/T	--	0
rs31400	5	131417406	C/T	--	0
rs31401	5	131418280	G/T	--	0
rs2073506	5	131422637	T/C	Y	0.07439
rs181781	5	131423014	A/G	Y	0.169846
<b>rs31480</b>	<b>5</b>	<b>131424231</b>	<b>C/T</b>	<b>Y</b>	<b>0.340306</b>
rs40401	5	131424377	C/T	--	0
rs31481	5	131425101	A/G	Y	0.074596
rs31474	5	131432926	C/T	Y	0.055775
rs25881	5	131439037	C/T	--	0.074454
rs25882	5	131439359	C/T	--	0.062762
rs25887	5	131443960	A/C	--	0.172942
rs31467	5	131464737	C/T	--	0
rs152198	5	131466709	T/C	--	0.083624
rs159905	5	131530402	C/T	--	NA

<sup>a</sup>TFBS, transcription factor binding site. <sup>b</sup>The regulatory potential of the SNP, higher value represents higher possibility that the SNP is functional.

**Table S4. Functional prediction of rs31480 and rs3916441 using Regulome DB**  
[\(<http://www.regulomedb.org>\)](http://www.regulomedb.org).

SNP ID	Chromosome	Coordinate	Regulome DB Score <sup>a</sup>
rs3916441	chr5	131369240	6
rs31480	chr5	131396331	5

<sup>a</sup>Regulome DB Score represents the regulatory potential of the SNP. Of note, Regulome DB Score is different from GWAVA and SNPinfo scores. Smaller Regulome DB Score indicates higher regulatory potential. Detailed information about Regulome DB Score are listed below:

**The scoring scheme refers to the following available datatypes for a single coordinate.**

Score	Supporting data
1a	eQTL + TF binding + matched TF motif + matched DNase Footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase Footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding / DNase peak
2a	TF binding + matched TF motif + matched DNase Footprint + DNase peak
2b	TF binding + any motif + DNase Footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif
4	TF binding + DNase peak
<b>5</b>	<b>TF binding or DNase peak</b>
<b>6</b>	<b>other</b>

Note: This information is from **Regulome DB** website (<http://regulomedb.org/help>).

**Table S5. The Encephalisation Quotients (EQ)<sup>a</sup> of the eight studied primates**

Species	Body weight (g)	Brain weight (mg)	Martin EQ
Human	44000	1250000	6.28
Chimpanzee	36350	410300	2.38
Gorilla	126500	505900	1.14
Orangutan	53000	413300	1.80
Gibbon	6521	112057	2.40
Baboon	17043	168357	1.74
Macaque	7280	90300	1.78
Marmoset	400	7000	1.70

<sup>a</sup>Encephalization quotient (EQ), or encephalization level is a measure of relative brain size defined as the ratio between actual brain mass and predicted brain mass for an animal of a given size. Data are from Roth et al [1] and Aiello et al [2].

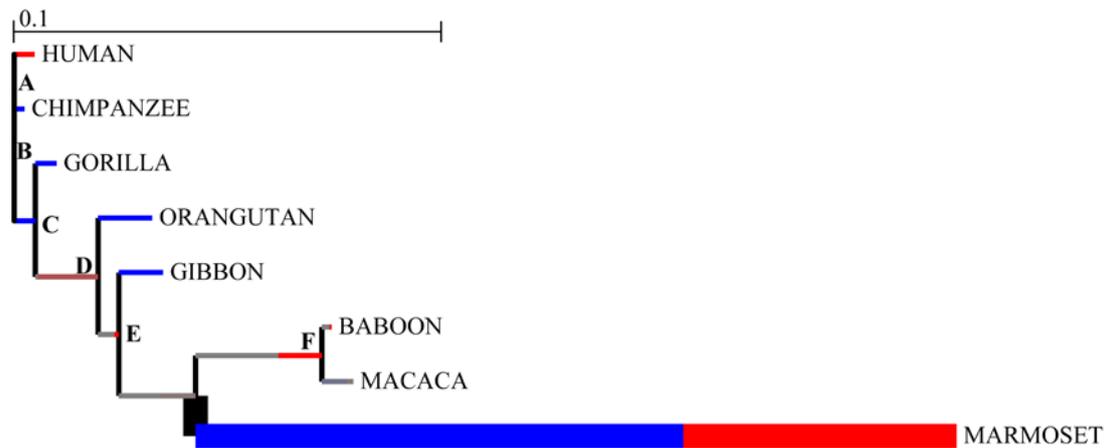


Figure S1. **Detecting of positive selection using Branch-site REL.** The hue of each color indicates strength of selection, with primary red corresponding to  $\omega > 5$ , primary blue to  $\omega = 0$  and grey to  $\omega = 1$ . The width of each color component represents the proportion of sites in the corresponding class. Thicker branches have been classified as undergoing episodic diversifying selection by the sequential likelihood ratio test at corrected  $p \leq 0.05$ . For more details, please refer to the paper of Kosakovsky Pond et al (<http://www.datamonkey.org/>).

## References

- Aiello L, and Dean C. 1990. An introduction to human evolutionary anatomy. Academic Press, London.
- Roth G, and Dicke U. 2005. Evolution of the brain and intelligence. *Trends Cogn Sci* **9**:250-257.