Cellular and Molecular Life Sciences

Variants in *SNAP25* are targets of natural selection and influence verbal performances in women

Rachele Cagliani · Stefania Riva · Cecilia Marino · Matteo Fumagalli ·
Maria Grazia D'Angelo · Valentina Riva · Giacomo P. Comi · Uberto Pozzoli ·
Diego Forni · Mario Cáceres · Nereo Bresolin · Mario Clerici · Manuela Sironi

Abstract Descriptions of genes that are adaptively evolving in humans and that carry polymorphisms with an effect on cognitive performances have been virtually absent. *SNAP25* encodes a presynaptic protein with a role in regulation of neurotransmitter release. We analysed the intra-specific diversity along *SNAP25* and identified a region in intron 1 that shows signatures of balancing selection in humans. The estimated TMRCA (time to the most recent common ancestor) of the SNAP25 haplotype phylogeny amounted to 2.08 million years. The balancing selection signature is not secondary to demographic events or to biased gene conversion, and encompasses rs363039. This SNP has previously been associated to cognitive performances with contrasting results in different populations. We analysed this variant in two Italian cohorts in different age ranges and observed a significant genotype effect for rs363039 on verbal performances in females alone. Post hoc analysis revealed that the effect is driven by differences between heterozygotes and both homozygous genotypes. Thus, heterozygote females for rs363039 display higher verbal performances compared to both homozygotes. This finding was replicated in a cohort of Italian subjects suffering from neuromuscular diseases that do not affect cognition. Heterozygote advantage is one of the possible reasons underlying the maintenance of genetic diversity in natural populations. The observation that heterozygotes for rs363039 display higher verbal abilities compared to homozygotes perfectly fits the underlying balancing selection model. Although caution should be used in inferring selective pressures from observed signatures, *SNAP25* might represent the first description of an adaptively evolving gene with a role in cognition.

Keywords *SNAP25* · Balancing selection · Intelligence quotient · Verbal performances

Abbreviations
SNP   Single nucleotide polymorphism
EAS   East Asians

R. Cagliani · S. Riva · C. Marino · M. Fumagalli ·
M. G. D'Angelo · U. Pozzoli · D. Forni · N. Bresolin ·
M. Sironi (✉)
Bioinformatic Lab, Scientific Institute IRCCS E. Medea,
Via don L. Monza 20, 23842 Bosisio Parini, LC, Italy
e-mail: manuela.sironi@bp.lnf.it

V. Riva
The Academic Centre for the Study of Behavioural Plasticity,
Vita-Salute San Raffaele University, Milan, Italy

G. P. Comi · N. Bresolin
Department of Neurological Sciences, Dino Ferrari Centre,
University of Milan, Fondazione Ca' Granda IRCCS Ospedale
Maggiore Policlinico, 20122 Milan, Italy

M. Cáceres
Institut de Biotecnologia i de Biomedicina, Universitat
Autònoma de Barcelona, 08193 Bellaterra, Spain

M. Cáceres
Institució Catalana de Recerca i Estudis Avançats (ICREA),
08010 Barcelona, Spain

M. Clerici
Chair of Immunology, Department of Biomedical Sciences and
Technologies LITA Segrate, University of Milan, 20090 Milano,
Italy

M. Clerici
Fondazione Don C. Gnocchi, IRCCS, 20148 Milano, Italy

CEU	Europeans
YRI	Yoruba
TMRCA	Time to the most recent common ancestor
ASF	Allele frequency spectrum
MLHKA	Maximum-likelihood HKA
BGC	Biased gene conversion
IQ	Intelligence quotient

Introduction

It is commonly believed that human brain function, cognitive abilities and linguistic skills have emerged as a result of natural selection [4, 37, 38, 56], although the specific genes and genetic structures underlying this process have remained elusive. Intelligence is a highly heritable trait in humans with heritability estimates reaching 66% in young adulthood [24]. Therefore, it is conceivable that variants affecting cognition are still evolving adaptively in anatomically modern humans [4, 38], although no specific evidence for this has been reported to date. While several genes involved in neural development and cognitive performances have been identified as targets of natural selection in modern human populations [14, 25, 30, 53, 64], none of their polymorphisms has been reliably shown to influence intelligence, the only exception being a nonsynonymous SNP in *BDNF* [36]. Still, in this latter case the Met66 allele, which is thought to have been driven to high frequency by selection, has a negative impact on cognitive abilities [12].

SNAP25 (synaptosomal associate protein 25 kDa) encodes a central component of the exocytotic machinery necessary for vesicle docking and fusion, and regulation of neurotransmitter release [42, 46]. In neurons, SNAP25 is located at the presynaptic plasma membrane where it also modulates the activity of voltage-gated calcium channels [8, 39, 58]. Polymorphisms in *SNAP25* have been associated with different neuropsychiatric conditions such as ADHD (reviewed in [9]), schizophrenia, bipolar disorders, and autism [9, 13, 23], and variants in the gene were reported to modulate cognitive performances [19, 20, 45]. In particular, SNPs located in a $\sim$14 kb region in intron 1 were associated with the intelligence quotient (IQ) in Dutch cohorts [19, 20]. A subsequent report in a Swedish sample replicated the association of one of these variants (rs363039, A/G) with cognition by measuring working memory capacity [45]. Yet, this study indicated the A allele as increasing cognitive ability, a situation opposite to the one observed in the Dutch cohorts where the G allele was associated with higher IQ [45].

We performed a population genetic analysis of *SNAP25* and association with cognitive performances in different Italian populations. Data herein indicated that the region carrying rs363039 has been a target of balancing selection in human populations and that heterozygosity at this variant confers higher verbal skills in female individuals.

Materials and methods

HapMap subject resequencing

All analysed regions were PCR amplified and directly sequenced; primer sequences are available upon request. PCR products were treated with ExoSAP-IT (USB Corp., Cleveland, OH, USA), directly sequenced on both strands with a Big Dye Terminator sequencing Kit (v3.1 Applied Biosystems) and run on an Applied Biosystems ABI 3130 XL Genetic Analyzer (Applied Biosystems). Sequences were assembled using AutoAssembler version 1.4.0 (Applied Biosystems) and inspected manually by two distinct operators.

Data retrieval, linkage analysis, and haplotype inference

Data from the Pilot 1 phase of the 1000 Genomes Project were retrieved from the dedicated website (http://www.1000genomes.org/) [1]. Sliding window analysis was performed on overlapping 20-SNP windows moving with a step of 3 SNPs. For each window we calculated the estimate of the nucleotide diversity for each population [59] divided by the total number of fixed differences with *Pan troglodytes*. As a control set, genotype data for 5-kb regions from 238 resequenced human genes were derived from the NIEHS SNPs Program web site (http://egp.gs.washington.edu). In particular we selected genes that had been resequenced in populations of defined ethnicity including CEU, YRI, and EAS (NIEHS panel 2). Haplotypes were inferred using PHASE version 2.1 [47, 48]. Linkage disequilibrium analyses were performed using Haploview (v. 4.1) [3]. Data for LD analysis were derived from HapMap.

Population genetic analyses

Tajima's D [49], Fu and Li's D* and F* [15] statistics, as well as diversity parameters θ_W [59] and π [33] were calculated using *libsequence* [51]. Calibrated coalescent simulations were performed using the *cosi* package [44] and its best-fit parameters for YRI, CEU, and EAS populations with 10,000 iterations. Coalescent simulations were conditioned on mutation rate and recombination rate. The maximum-likelihood-ratio HKA test was performed using the MLHKA software [65], as previously proposed [16]. In

all analyses the chimpanzee sequence was used as the out-group. The median-joining network to infer haplotype genealogy was constructed using NETWORK 4.5 [2]. Estimate of the time to the most recent common ancestor (TMRCA) was estimated through the application of a maximum-likelihood coalescent method implemented in GENETREE [21, 22]. The method assumes an infinite-site model without recombination; therefore, haplotypes and sites that violate these assumptions need to be removed: for the analysis of *SNAP25* we removed eight variants. Again, the mutation rate μ was obtained on the basis of the divergence between human and chimpanzee and under the assumption both that the species separation occurred 6 million years (MY) ago [18], and of a generation time of 25 years. The migration matrix was derived from previous estimated migration rates [44]. Using this μ and θ maximum likelihood (θ_{ML}), we estimated the effective population size parameter (N_e), which was equal to 12,923. With these assumptions, the coalescence time, scaled in $2N_e$ units, was converted into years. For the coalescence process, 10^6 simulations were performed.

Child cohort

Subjects were drawn from a still-ongoing study on genetic and environmental influences upon language development in a general population sample. Children ($n = 368$) were recruited in Italy with the help of elementary schools and kindergartens in the greater Milan area and in the province of Lecco. The inclusion criteria were a mean score between vocabulary/information and block design sub-tests above 4 (i.e., >-2 SD), no disability certification, Caucasian ethnicity, and a self report of Italian descent of at least one generation. Participants were grouped in two age groups: 3–8 ($n = 185$, mean age: 6.18, SD: 1.57) and 9–11 ($n = 183$, mean age: 9.68, SD: 0.68) years. Two sub-tests of the age-appropriate Wechsler Intelligence Scale (WISC-R; WPPSI-R) [60, 61] were used to provide a proxi measure of verbal and performance IQ, i.e. Vocabulary/Information and Block Design, which are the most representative of, respectively, verbal IQ ($r = 0.82/0.73$) and performance IQ ($r = 0.70/0.59$) (WISC-R/WPPSI-R) [60, 61]. Scaled scores were compiled following manual instructions and were used in the statistical analyses. Mouth-wash samples for DNA extraction were obtained from each subject. Ethics approval for this study was received from Eugenio Medea Scientific Institute Ethical Committee.

Neuromuscular cohort

Seventy patients with different types of neuromuscular disorders not associated to mental retardation were recruited. Details concerning disease type, male/female ratio, and age are available as Electronic Supplementary Material, Table 1. Inclusion criteria was Italian Caucasian origin, age below 60 years, and FIQ ≥ 70. All subjects had a confirmed clinical, histological, and/or molecular diagnosis of LGMD, FSHD, congenital myopathy, SMA, and hereditary peripheral neuropathies, according to international criteria [5, 6, 32, 34, 35]. Cognitive performance was assessed using the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), or the Wechsler Intelligence Scale for Children-Revised (WISC-R) and For Adult-Revised (WAIS-R), according to age [60–62]. FIQ, VIQ, and PIQ did not significantly differ in males and females. Genomic DNA was extracted from peripheral blood using standard procedures. Ethics approval for this study was received from Eugenio Medea Scientific Institute Ethics Committee.

### SNP genotyping and statistical analyses

Genotyping of all SNPs was performed by direct resequencing, as described above; primer pairs are available upon request. All SNPs were in Hardy-Weinberg equilibrium in all cohorts, as assessed by application of Exact Tests [63] as implemented in PLINK [40]. One-way ANOVA was used to assess the difference in cognitive performances. A conservative Bonferroni correction for multiple tests was applied to account for the analysis of multiple SNPs. Post hoc tests were performed using the Tukey's Honest Significant Difference method. All calculations were carried out in the R environment [41].

Results

Evolutionary analysis of *SNAP25*

In order to obtain an overall picture of *SNAP25* intra-specific genetic diversity, we exploited data from the 1000 Genomes Pilot project deriving from the low-coverage whole genome sequencing of 179 individuals with different ancestry [1]. In particular, individuals from three distinct ethnic groups have been analysed: Europeans (CEU), Yoruba from Nigeria (YRI), and Japanese plus Chinese (East Asian, EAS). Using these data we calculated θ_W [59], an estimate of the expected per site heterozygosity, and human-chimpanzee divergence in sliding windows moving along *SNAP25*. As shown in Fig. 1, for the three populations a peak in the ratio of θ_W over divergence was observed in an intron 1 region where an SNP (rs363039) previously associated with cognitive performances is located. The low-coverage 1000 Genome Project approach is estimated to have relatively low power to detect singleton SNPs or rare variants [1]; thus, we used a Sanger

Fig. 1 Sliding window analysis of nucleotide diversity along *SNAP25*. The ratio of θ_W over human/chimpanzee divergence was calculated in sliding windows moving along the gene for the three different populations resequenced by the 1000 Genomes Pilot Project, namely YRI (*red*), CEU (*blue*), and EAS (*green*). The exon-intron structure of SNAP25 is also shown, and *shading* indicates the three regions we resequenced. The LD plot in CEU for the region where rs363039, rs363016, and rs363043 are located is also shown (r^2 data were derived from the HapMap website)
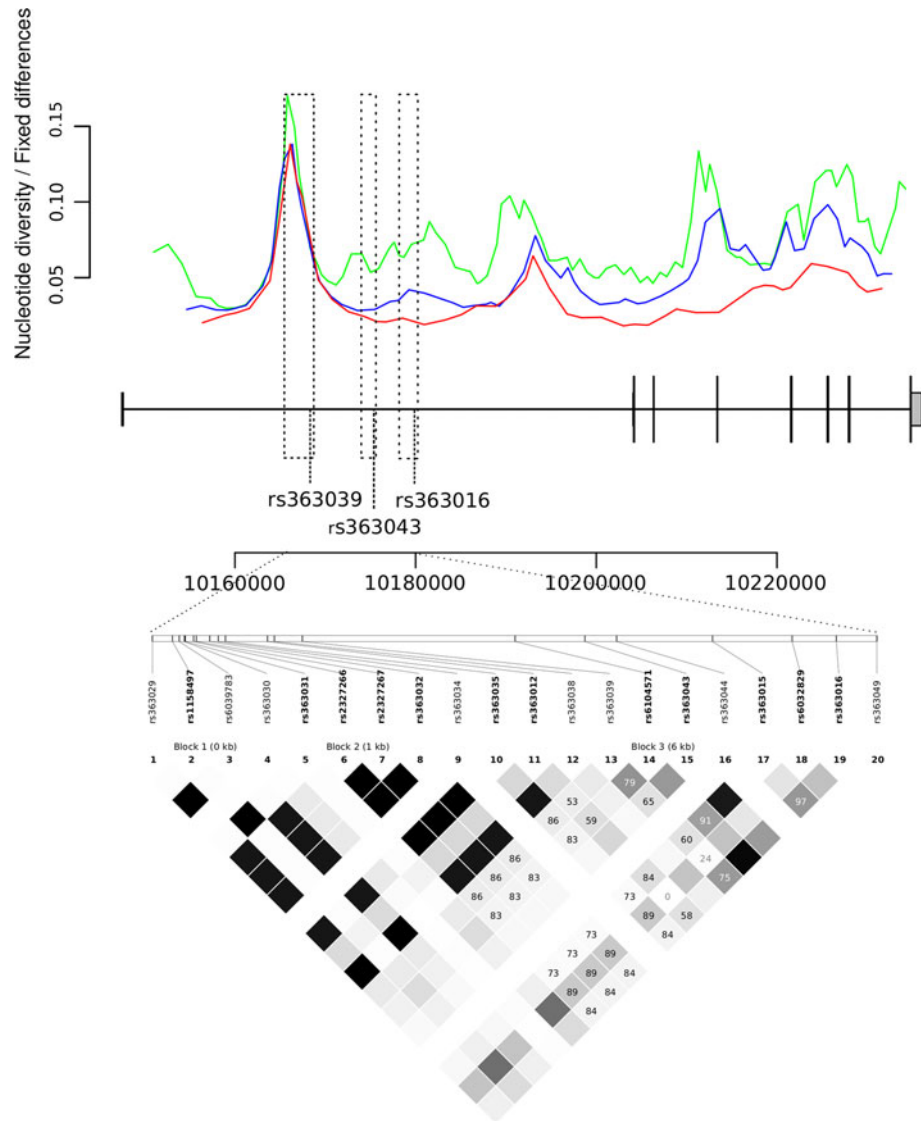
resequencing strategy to verify the sliding-window data and to apply classic population genetic tests. For this purpose we selected three regions: the one corresponding to the diversity peak (3.4 kb), which also carries rs363039, and two additional gene segments in intron 1 (2 kb each, Fig. 1), where two other SNPs associated with cognitive abilities (rs363043 and rs363016) have been described (Fig. 1) [19, 20]. We resequenced these regions in 20 HapMap CEU individuals and recalculated θ_W, as well as π [33], this latter corresponding to the average number of pairwise sequence nucleotide differences between haplotypes. In order to compare the values we obtained for the *SNAP25* regions, we calculated θ_W and π for 5 kb windows (thereafter referred to as reference windows) derived from 238 genes resequenced (Sanger method) by the NIEHS program in the same population; the percentile rank corresponding to the *SNAP25* regions in the distribution of reference windows is reported in Table 1. No exceptional

nucleotide diversity was observed for the gene portions encompassing rs363043 and rs363016. Conversely, a value of θ_W corresponding to the 98th percentile was observed for the region where rs363039 is located, confirming the sliding-window results. π was also high in this region, although it did not reach statistical significance.

High nucleotide diversity might be suggestive of balancing selection, as neutral variants tend to be maintained with the selected alleles. Data in Fig. 1 suggest that the excess of nucleotide diversity we observe is not due to higher local mutations rates as, in this case, an increase in human-chimpanzee divergence would also be expected. In order to formally rule out this possibility we applied a maximum likelihood HKA (Hudson-Kreitman-Aguadè) test [65]. This is based on the concept whereby, under neutral evolution, the amount of within- and between-species diversity is expected to be similar at all loci in the genome [27]. Therefore, the test compares polymorphism

Table 1 Nucleotide diversity and summary statistics for the three SNAP25 gene regions we resequenced

Region	Pop.[a]	S[b]	θ_w ($\times 10^{-4}$)	Rank[c]	π ($\times 10^{-4}$)	Rank[c]	Tajima's D	Rank[c]	p[d]	Fu & Li's D*	Rank[c]	P[d]	Fu & Li's F*	Rank[c]	p[d]
SNAP25 rs363039	YRI	28	19.28	0.98	30.70	>0.99	2.03	>0.99	>0.001	0.72	0.90	0.076	1.8	0.99	0.0072
	CEU	24	16.53	0.98	12.37	0.90	−0.85	0.24	0.82	1.42	0.96	0.022	0.78	0.75	0.20
	EAS	23	15.84	0.99	17.69	0.97	0.39	0.63	0.39	1.71	>0.99	>0.001	1.50	0.96	0.049
SNAP25 rs363016 (2059 bp)	CEU	3	3.43	0.21	4.89	0.43	0.93	0.79	n.d.	0.92	0.85	n.d.	1.07	0.83	n.d.
SNAP25 rs363043 (1771 bp)	CEU	3	3.98	0.28	5.71	0.53	0.95	0.80	n.d.	−0.35	0.46	n.d.	0.039	0.56	n.d.

[a] Population

[b] Number of segregating sites

[c] Percentile rank relative to the distribution of values obtained in reference windows

[d] p value calculated through the use of coalescent simulations

and divergence levels for a region of interest with those calculated for other neutrally evolving genomic segments. The results for the *SNAP25* region carrying rs363039 indicated that a significant excess of polymorphism compared to divergence is observed (selection parameter $k = 4.46$; p value $= 4.74 \times 10^{-4}$; Electronic Supplementary Material, Table 2).

Natural selection acting on specific gene regions can determine a distortion in the allele frequency spectrum (AFS). Common neutrality tests based on the AFS include Tajima's D (D_T) [49], and Fu and Li's D* and F* [15]. Since population history, in addition to selective processes, is known to affect the AFS, we evaluated the significance of neutrality tests by performing coalescent simulations that incorporate demographic scenarios (see "Methods") [44]. As explained above, we also applied an empirical comparison by calculating the percentile rank of D_T, F*, and D* for the three *SNAP25* gene regions relative to reference windows. Neutrality tests failed to reject the null hypothesis for the regions covering rs363043 and rs363016. Conversely, significant values of D* were obtained for the gene segment where rs363039 is located (Table 1). In order to gain further insight into the evolutionary history of this gene region, we extended resequencing analysis to two additional HapMap populations, namely YRI and EAS. As reported in Table 1, high nucleotide diversity was observed in both populations, and most neutrality tests rejected the null hypothesis of selective neutrality in both YRI and EAS. Similarly, the MLHKA test indicated a significant excess of polymorphism compared to divergence in these populations (both p values $<10^{-3}$; Electronic Supplementary Material, Table 2).

It has recently been shown that biased gene conversion (BGC) affects neutral substitution patterns, possibly mimicking the effects of natural selection (reviewed in [11]). The recombination rate is relatively high in the region where rs363039 is located (1.65 cM/Mb averaged over the three populations); therefore, we analysed the frequency of A/T → G/C substitutions in the region: out of a total of 36 segregating sites, only 38% were accounted for by A/T → G/C mutations, suggesting that BGC does not play a major role in shaping nucleotide variability at *SNAP25* intron 1.

Additional insight into the evolutionary history of genomic regions can be gained by the analysis of haplotype genealogies. In order to estimate the coalescence time (time to the most recent common ancestor, TMRCA) of haplotypes in the *SNAP25* region carrying rs363039, we used GENETREE, which is based on a maximum-likelihood coalescent method. The estimated TMRCA amounted to 2.08 million years (MY) (SD: 0.52 MY) (Electronic Supplementary Material, Fig. 1). We next constructed a median-joining network of the haplotype genealogy (Fig. 2) that shows the presence of two major haplotype clades both containing chromosomes from all three population samples. In agreement with the relatively high recombination rate in the region, the network shows some recurrent mutations that may result from recombination or gene conversion events.

Association of *SNAP25* variants with cognitive performances

In order to analyse the role of *SNAP25* variants in the modulation of cognitive performance, we enrolled two population of children divided in age ranges: 3–8 ($n = 185$) and 9–12 ($n = 183$) years. These subjects underwent two sub-tests of the Wechsler Intelligence Scale for Children (WISC), namely Vocabulary/Information and Block design. Five SNPs were selected for genotyping: in addition to rs363039, we analysed two variants (rs6039787 and rs3787297) located in the balancing selection region that define specific haplotypes (Fig. 2). Although no natural selection signature was identified in the regions where rs363016 and rs363043 are located, we included these two variants, as they were previously associated with IQ scores [19, 20]. Analysis was performed separately depending on gender. No genotype effect was observed for rs6039787,

Fig. 2 Network analysis for the *SNAP25* region where rs363039 is located. Each node represents a different haplotype, with the size of the circle proportional to frequency. Nucleotide differences between haplotypes are indicated on the branches of the network. The position of SNPs we used in association analysis is shown. *Circles* are colour-coded according to population (*green*: YRI, *blue*: CEU, *red*: EAS). The most recent common ancestor (MRCA) is also shown (*black circle*). The relative position of mutations along a branch is arbitrary. For all haplotypes observed in CEU the allelic status at rs363039 (A/G) is reported
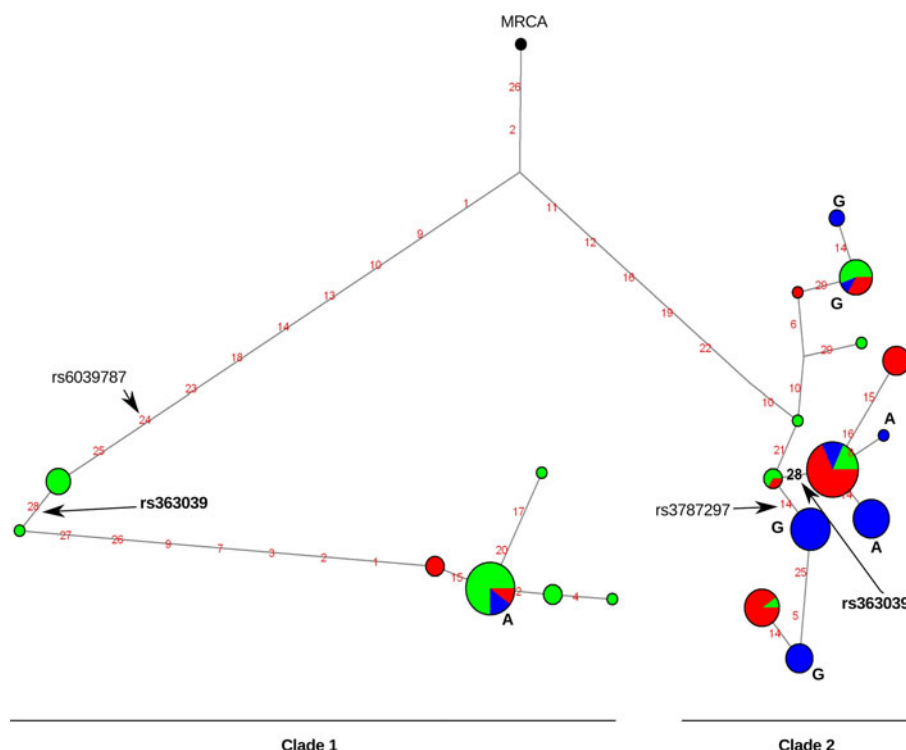
rs3787297, or rs363043 on either Vocabulary/Information or Block design scores (Electronic Supplementary Material, Table 3). Conversely, in both age groups a significant genotype effect was observed for rs363039 on Vocabulary/Information scores in females (Table 2), but not on Block design in either sex. Analysis of the combined cohort showed a strong effect of this SNP on verbal performances (one-way ANOVA, Bonferroni corrected $p = 0.00029$) in the female sample. Tukey post hoc analysis indicated that in both age cohorts the genotype effect is driven by differences between heterozygotes and both homozygous genotypes (3–8 year group: AA vs. AG, $p = 0.017$; AG vs. GG, $p = 0.032$; 9–12 year group: AA vs. AG, $p = 0.008$; AG vs. GG, $p = 0.044$). Thus, heterozygote females for rs363039 display, on average, higher verbal performances compared to both homozygotes, while this effect is absent in males (Fig. 3). As for rs363016, a significant genotype effect in females was observed in the 3–8 age group, but was not confirmed in older children (Table 2). The combined cohort yielded a significant difference in the female cohort, but resulted in weaker association compared to rs363039 (Table 2). Therefore, the genotype effect for rs363016 is most likely accounted for by linkage disequilibrium with rs363039 ($r^2 = 0.49$, D' = 0.83).

In order to replicate these findings in a third independent cohort, we recruited 70 subjects suffering from neuromuscular disorders that do not affect intellectual abilities (see Electronic Supplementary Material, Table 1 for demographic and clinical details). Cognitive performances were evaluated by recording FIQ, VIQ, and PIQ, and rs363039 was genotyped in all subjects. No genotype effect was observed for FIQ and PIQ (Table 2). As for VIQ, females showed a pattern for the three genotypes that recapitulates that observed for Vocabulary/Information scores in the child cohort (Fig. 3), although the difference only showed borderline significance (one-way ANOVA, $p = 0.050$, Table 2). Given the small sample size and the data obtained in the post hoc analysis above (showing a higher Vocabulary/Information score for heterozygotes compared to both homozygote genotypes), we pooled together AA and GG homozygotes and compared their VIQ with that obtained from heterozygotes (Fig. 3): a significant difference with higher VIQ for heterozygotes was observed in females but not males (one-way ANOVA, $p = 0.014$, Table 2).

Thus, the results of these three independent cohorts are consistent in showing that heterozygosity for rs363039, which is located in the balancing selection region, confers higher verbal performances to female subjects.

Discussion

The search for signatures of natural selection acting on human genes serves a double purpose: first it helps to clarify the evolutionary history of our species, especially in relation to complex phenotypic traits such as cognition; second, the identification of gene regions subject to non-neutral evolution can be exploited to recognise functional

Table 2 Association analysis of rs363039 and rs363016 with cognitive performance in the child and neuromuscular cohorts

| SNP	Population	Sex (n)	Test	Score (SD)			F	p	p^c
rs363039	Children			AA	AG	GG			
	3–8 years	Males (81)	V/I[a]	10.00 (3.20)	9.66 (2.45)	10.69 (3.23)	1.03	0.36	>0.05
			Block[b]	13.80 (1.75)	11.23 (2.83)	11.10 (3.95)	2.89	0.062	>0.05
	3–8 years	Females (104)	V/I[a]	8.23 (2.24)	10.71 (2.55)	9.18 (3.04)	**5.19**	**0.0072**	**0.036**
			Block[b]	11.08 (2.93)	11.58 (3.18)	10.92 (3.21)	0.48	0.62	>0.05
	9–12 years	Males (108)	V/I[a]	8.25 (2.43)	8.02 (3.47)	8.67 (3.14)	0.46	0.63	>0.05
			Block[b]	10.38 (2.77)	9.82 (2.93)	10.70 (4.27)	0.71	0.50	>0.05
	9–12 years	Females (75)	V/I[a]	5.67 (1.75)	10.12 (3.78)	8.14 (2.82)	**5.97**	**0.0042**	**0.021**
			Block[b]	9.67 (1.37)	11.31 (2.68)	10.68 (2.38)	1.25	0.29	>0.05
	Combined	Males (189)	V/I[a]	9.22 (2.94)	8.70 (3.18)	9.49 (3.31)	1.18	0.31	>0.05
			Block[b]	12.28 (2.80)	10.40 (2.96)	10.86 (4.12)	2.18	0.12	>0.05
	Combined	Females (179)	V/I[a]	7.42 (2.39)	10.47 (3.09)	8.73 (2.98)	**10.34**	$\mathbf{5.8 \times 10^{-5}}$	**0.0003**
			Block[b]	10.63 (2.59)	11.47 (2.97)	10.81 (2.87)	1.17	0.31	>0.05
rs363016	Children			CC	CT	TT			
	3–8 years	Males (81)	V/I[a]	7.79 (2.57)	10.18 (2.81)	10.19 (3.21)	0.13	0.88	>0.05
			Block[b]	12.68 (2.67)	11.74 (3.10)	11.00 (4.05)	1.38	0.26	>0.05
	3–8 years	Females (104)	V/I[a]	8.88 (2.78)	10.63 (2.58)	8.72 (2.91)	**5.90**	**0.0038**	**0.019**
			Block[b]	11.24 (3.29)	11.22 (3.25)	11.56 (3.33)	0.12	0.89	>0.05
	9–2 years	Males (108)	V/I[a]	8.54 (4.18)	8.05 (2.76)	9.65 (3.50)	1.85	0.16	>0.05
			Block[b]	10.69 (2.56)	10.13 (2.98)	12.19 (3.75)	3.24	0.045	>0.05
	9–12 years	Females (75)	V/I[a]	7.57 (3.46)	9.00 (3.50)	8.30 (2.84)	0.70	0.50	>0.05
			Block[b]	9.71 (2.43)	11.26 (2.53)	10.56 (2.42)	1.34	0.27	>0.05
	Combined	Males (189)	V/I[a]	9.28 (3.31)	9.10 (2.96)	9.92 (3.34)	1.16	0.32	>0.05
			Block[b]	11.88 (2.77)	10.89 (3.12)	11.60 (3.91)	1.22	0.30	>0.05
	Combined	Females (179)	V/I[a]	8.50 (2.98)	10.00 (3.06)	8.54 (2.87)	**5.10**	**0.0071**	**0.036**
			Block[b]	10.79 (3.09)	11.24 (2.98)	11.13 (3.00)	0.20	0.82	>0.05
rs363039	Neuromusc.			AA	AG	GG			
		Males (36)	FIQ	102.50 (0.71)	101.82 (16.24)	93.33 (13.93)	1.26	0.30	n.p.
			VIQ	102.00 (7.07)	102.09 (14.70)	92.25 (14.97)	1.81	0.18	n.p.
			PIQ	103.50 (4.95)	101.55 (20.77)	96.00 (14.21)	0.39	0.68	n.p.
		Females (34)	FIQ	99.20 (5.40)	104.36 (10.90)	96.78 (10.94)	1.82	0.18	n.p.
			VIQ	93.40 (11.74)	106.45 (10.78)	95.39 (13.23)	**3.30**	**0.050**	n.p.
			PIQ	101.00 (7.00)	102.27 (15.07)	102.00 (10.69)	0.02	0.98	n.p.
rs363039	Neuromusc.			AA and GG	AG				
		Males (36)	FIQ	94.64 (13.24)	101.82 (16.24)		1.92	0.18	n.p.
			VIQ	93.64 (14.36)	102.09 (14.70)		2.88	0.099	n.p.
			PIQ	97.07 (13.42)	101.55 (20.77)		0.51	0.48	n.p.
		Females (34)	FIQ	97.30 (9.94)	104.36 (10.90)		3.53	0.069	n.p.
			VIQ	94.96 (12.69)	106.45 (10.78)		**6.69**	**0.014**	n.p.
			PIQ	101.78 (9.87)	102.27 (15.07)		0.013	0.91	n.p.

[a] Vocabulary/information

[b] Block design

[c] Bonferroni corrected *p* value; significant *p* values are in bold

variants, which are a prerequisite for selection to act. In particular, balancing selection signatures, due to the combined effects of mutation and recombination, are expected to extend over relatively short genomic regions [7] where the functional selection target(s) is located. Indeed, as shown in the sliding-window analysis (Fig. 1), the high-diversity peak is restricted to a relatively narrow region in *SNAP25*, which encompasses rs363039. Application of

Fig. 3 Boxplot of Vocabulary/
Information and VIQ scores.
Vocabulary/Information and
VIQ scores are shown according
to rs363039 genotype. Results
refer to two age cohorts
combined for children (**a**, **b**).
For the neuromuscular cohort,
VIQ scores are reported based
on genotype status at rs363039
with homozygous genotypes
analysed separately (**c** and **d**) or
together (**e** and **f**)
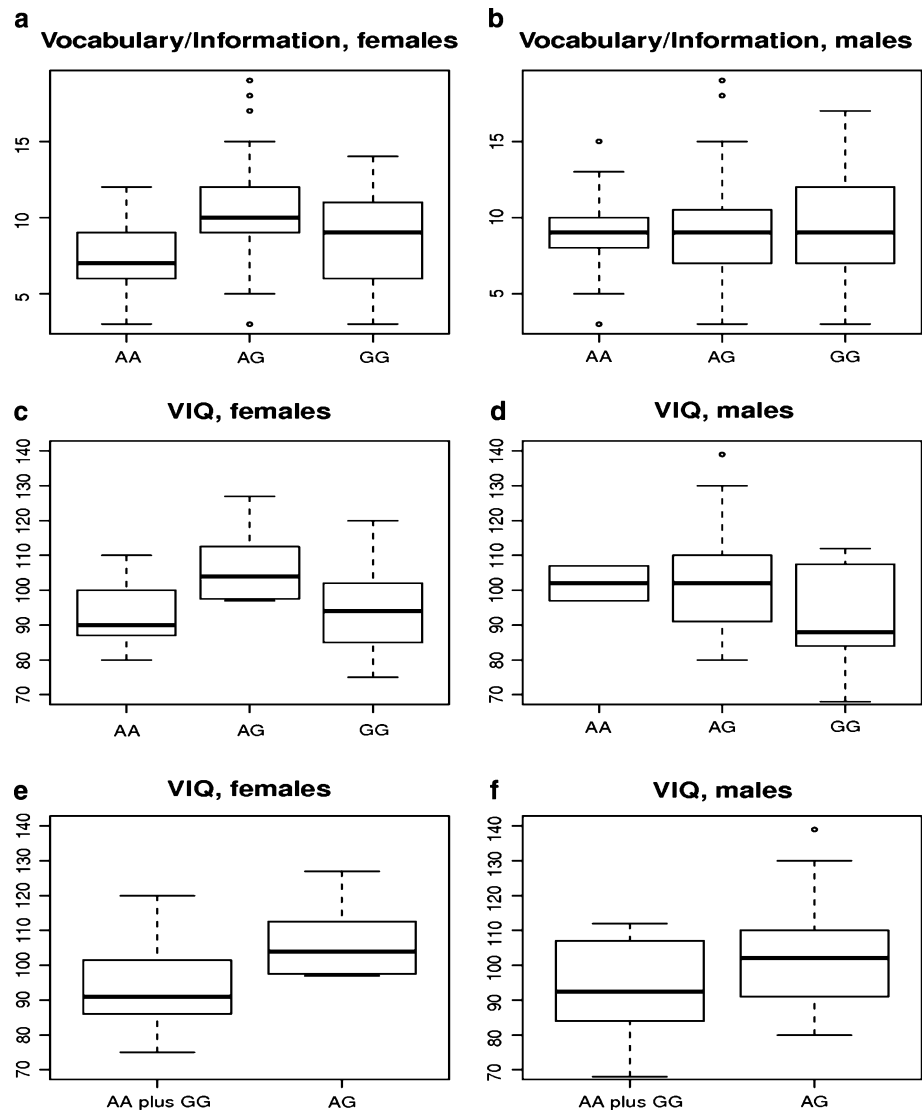
population genetic tests based on the allele frequency spectrum in most cases allowed rejection of the null hypothesis of selective neutrality for this gene region. It is worth noting that the failure to reach statistical significance for some of these tests is likely due to the fact that in non-African populations, the two major haplotype clades have extremely different frequencies, resulting in no marked skew towards intermediate frequency variants (to which these tests are sensitive). Support of the notion whereby the region carrying rs363039 represents a selection target also comes from the estimation of the coalescent time: we obtained a TMRCA dating back to 2 MY, which is deeper than the same estimates obtained for most autosomal neutrally evolving genes, these ranging from 0.8 to 1.5 MY [52]. However, some caution should be used in interpreting this result, as inference of phylogenies in the presence of recombination is not extremely robust. This observation also leads to the question as to whether rs363039 represents

the selection target or not, and, in turn, if increased verbal abilities represent the selected trait. On one hand, theory predicts that the selected variant(s) should be located on the basal branches separating the major haplotype clades, and this is not the case for rs363039. On the other hand, we selected variants to be genotyped depending on their location in the haplotype genealogy (so as to allow association analysis of major haplotypes in Europeans) but only rs363039 yielded significant association with IQ scores, suggesting that it represents the causal variant in the region. Different explanations can reconcile these apparent conflicting results. Selection may have acted to maintain heterozygosity at rs363039 (due to its conferring higher verbal abilities), and historical recombination (or gene conversion) events have resulted in the observed phylogeny. Alternatively, a variant different from, but in proximity to rs363039 and a trait unrelated to cognition may have been targeted by selection. *SNAP25* is a central

component of the exocytosis machinery in pancreatic beta islets, where it interacts with several components, including syntaxin A1 (*STX1A*) and calpain-10 (*CAPN10*) to modulate insulin release [29, 50]. In line with this function, SNPs in both *CAPN10* and *STX1A* have been associated with the predisposition to type 2 diabetes [43, 54, 55]. Genes associated with metabolism in general, and type 2 diabetes in particular, are thought to have represented selection targets during human evolutionary history, as a result of changes in diet following the advent of agriculture. This possibility has been verified in the case of *CAPN10* [57]. Therefore, the balancing selection signature we identified in *SNAP25* may be related to its role in insulin secretion and metabolism. Nonetheless, these two alternatives are not mutually exclusive as different selective forces, occurring across different time periods, may have affected the evolutionary fate of the genomic region.

Despite the generally held view that intelligence and verbal skills should have represented selected traits [4, 37, 38, 56], the description of genes/gene regions that are adaptively evolving in humans and that carry polymorphisms with an effect on cognitive performances have been virtually absent. Even the signatures of strong and recent positive selection at *MCPH1* and *ASPM* [14, 30], two genes that result in microcephaly and cognitive impairment when mutated, seem to be driven by pressures unrelated to cognition [31]. In fact, both genes are expressed in tissues other than brain, and their variants may have pleiotropic effects. Similarly, the selective force responsible for the increase of the Met66 allele of *BDNF* in North African, European, and Asian populations is presently unknown [36], and this allele is associated with lower rather than higher cognitive abilities [12].

One possibility for this failure to identify selection signature for variants that affect intelligence is that most large-scale analyses of selective patterns in the human genome have focussed on relatively recent, positive selection events. This is mainly due to the difficulty in identifying long-standing balancing selection signatures at the genome-wide level. However, older (and possibly weaker) selective events may have operated over a time frame that is more relevant to the evolutionary history of cognitive abilities given that, as noticed by Vallender [56], anatomically modern humans emerged between 100,000 and 200,000 years ago, and the "human" brain was already established by then. In this respect it is worth noting that heterozygote advantage is one of the possible reasons underlying the maintenance of genetic diversity in natural populations. Thus, the observation that heterozygotes for rs363039 display higher verbal abilities compared to homozygotes perfectly fits the underlying balancing selection model. Besides this evolutionary consideration, the fact that heterozygosity at this variant is associated with

higher cognitive performances may have a biochemical explanation, as well. Typically, the first intron of genes contains sequences important for gene regulation, and different transcription factors (TF) have been predicted to bind the region where rs363039 is located, depending on its allelic status [19, 45]. Albeit these predictions should be taken with caution because of the largely degenerated consensuses for TF binding sites, it may be speculated that heterozygotes display a wider *SNAP25* expression flexibility in terms of either response to specific stimuli or cell-type specificity.

As noted above, the association of rs363039 with cognitive skills has been reported previously, although different alleles have been associated with higher performances. The reason for this inconsistency has remained unexplained, although our data suggest that stratification by sex and application of non-allelic models might help reconcile those data. Indeed, we analysed three independent populations and obtained the same pattern in all cases, with heterozygote females displaying higher verbal performances compared to homozygotes. Thus, we consider this finding to be robust, although it would certainly benefit from replication in larger cohorts. Similarly, the possibility exists that the neuromuscular sample does not perfectly reflect the IQ score variation in healthy subjects. Nonetheless, we envisage no reason why allelic status at rs363039 should have a sex-driven interaction with any of these disorders, and we included a similar number of male and female subjects for each disease (Electronic Supplementary Material, Table 1). Moreover, the average FIQ, VIQ, and PIQ were not significantly different between male and female neuromuscular patients, and the VIQ pattern in relation to rs363039 perfectly paralleled the Vocabulary/Information scores obtained in the healthy children cohorts. As for the remaining SNPs, we failed to replicate the association, previously reported by Gosso and co-workers [19] of rs363016 and rs363043 with intelligence. Yet, it should be noticed that Gosso et al. used a family-based approach, which might have resulted in the strongest power to detect associations, and large subject cohorts from the Netherlands may display higher genetic homogeneity and linkage disequilibrium compared to our Italian sample.

Finally, the reason why heterozygosity at rs363039 affects verbal performances in female subjects alone remains to be clarified. A sex-specific difference has been described in the brain expression of *SNAP25*, although few samples were used for the analyses and no genotypic information was incorporated in the study [10]. Previous works have also indicated that estrogen regulates *SNAP25* expression in rat brain areas and in the pituitary [26, 28]. In this respect it is worth mentioning that several neuropsychiatric disorders that have been associated with variants in *SNAP25* (e.g. ADHD, schizophrenia, autism) display

well-known gender prevalence biases. Whether these differences also depend on a differential expression of *SNAP25* in males and females [17] remains to be elucidated, and might provide further insight into the role of this protein in both normal cognitive development and in neuropsychiatric disease.

References

1. 1000 Genomes Project Consortium, Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073

2. Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

3. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265. doi:10.1093/bioinformatics/bth457

4. Bishop DV (2009) Genes, cognition, and communication: insights from neurodevelopmental disorders. Ann NY Acad Sci 1156:1–18

5. Bruno C, Minetti C (2004) Congenital myopathies. Curr Neurol Neurosci Rep 4:68–73

6. Bushby K (2009) Diagnosis and management of the limb girdle muscular dystrophies. Pract Neurol 9:314–323

7. Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet 2:e64

8. Condliffe SB, Corradini I, Pozzi D, Verderio C, Matteoli M (2010) Endogenous SNAP-25 regulates native voltage-gated calcium channels in glutamatergic neurons. J Biol Chem 285:24968–24976

9. Corradini I, Verderio C, Sala M, Wilson MC, Matteoli M (2009) SNAP-25 in neuropsychiatric disorders. Ann NY Acad Sci 1152:93–99

10. Downes EC, Robson J, Grailly E, Abdel-All Z, Xuereb J, Brayne C, Holland A, Honer WG, Mukaetova-Ladinska EB (2008) Loss of synaptophysin and synaptosomal-associated protein 25-kDa (SNAP-25) in elderly Down syndrome individuals. Neuropathol Appl Neurobiol 34:12–22

11. Duret L, Galtier N (2009) Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet 10:285–311

12. Egan MF, Kojima M, Callicott JH, Goldberg TE, Kolachana BS, Bertolino A, Zaitsev E, Gold B, Goldman D, Dean M, Lu B, Weinberger DR (2003) The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. Cell 112:257–269

13. Etain B, Dumaine A, Mathieu F, Chevalier F, Henry C, Kahn JP, Deshommes J, Bellivier F, Leboyer M, Jamain S (2010) A SNAP25 promoter variant is associated with early-onset bipolar disorder and a high expression level in brain. Mol Psychiatry 15:748–755

14. Evans PD, Gilbert SL, Mekel-Bobrov N, Vallender EJ, Anderson JR, Vaez-Azizi LM, Tishkoff SA, Hudson RR, Lahn BT (2005) Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. Science 309:1717–1720

15. Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

16. Fumagalli M, Cagliani R, Pozzoli U, Riva S, Comi GP, Menozzi G, Bresolin N, Sironi M (2009) Widespread balancing selection and pathogen-driven selection at blood group antigen genes. Genome Res 19:199–212

17. Ghezzo A, Guerini FR, Bolognesi E, Matteoli M, Manca S, Sotgiu S, Bejor M, Clerici M, Chiappedi M (2009) Neuropsycological gender differences in healthy individuals and in pediatric neurodevelopmental disorders. A role for SNAP-25. Med Hypotheses 73:978–980

18. Glazko GV, Nei M (2003) Estimation of divergence times for major lineages of primate species. Mol Biol Evol 20:424–434

19. Gosso MF, de Geus EJ, Polderman TJ, Boomsma DI, Heutink P, Posthuma D (2008) Common variants underlying cognitive ability: further evidence for association between the SNAP-25 gene and cognition using a family-based study in two independent Dutch cohorts. Genes Brain Behav 7:355–364

20. Gosso MF, de Geus EJ, van Belzen MJ, Polderman TJ, Heutink P, Boomsma DI, Posthuma D (2006) The SNAP-25 gene is associated with cognitive ability: evidence from a family-based study in two independent Dutch cohorts. Mol Psychiatry 11:878–886

21. Griffiths RC, Tavare S (1995) Unrooted genealogical tree probabilities in the infinitely-many-sites model. Math Biosci 127:77–98

22. Griffiths RC, Tavare S (1994) Sampling theory for neutral alleles in a varying environment. Philos Trans R Soc Lond B Biol Sci 344:403–410

23. Guerini FR, Bolognesi E, Chiappedi M, Manca S, Ghezzo A, Agliardi C, Sotgiu S, Usai S, Matteoli M and Clerici M (2011) SNAP-25 single nucleotide polymorphisms are associated with hyperactivity in autism spectrum disorders (in press)

24. Haworth CM, Wright MJ, Luciano M, Martin NG, de Geus EJ, van Beijsterveldt CE, Bartels M, Posthuma D, Boomsma DI, Davis OS, Kovas Y, Corley RP, Defries JC, Hewitt JK, Olson RK, Rhea SA, Wadsworth SJ, Iacono WG, McGue M, Thompson LA, Hart SA, Petrill SA, Lubinski D, Plomin R (2010) The heritability of general cognitive ability increases linearly from childhood to young adulthood. Mol Psychiatry 15:1112–1120

25. Haygood R, Fedrigo O, Hanson B, Yokoyama KD, Wray GA (2007) Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. Nat Genet 39:1140–1144

26. Jacobsson G, Razani H, Ogren SO, Meister B (1998) Estrogen down-regulates mRNA encoding the exocytotic protein SNAP-25 in the rat pituitary gland. J Neuroendocrinol 10:157–163

27. Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

28. Lustig RH, Hua P, Wilson MC, Federoff HJ (1993) Ontogeny, sex dimorphism, and neonatal sex hormone determination of synapse-associated messenger RNAs in rat brain. Brain Res Mol Brain Res 20:101–110

29. Marshall C, Hitman GA, Partridge CJ, Clark A, Ma H, Shearer TR, Turner MD (2005) Evidence that an isoform of calpain-10 is a regulator of exocytosis in pancreatic beta-cells. Mol Endocrinol 19:213–224

30. 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

31. Mekel-Bobrov N, Posthuma D, Gilbert SL, Lind P, Gosso MF, Luciano M, Harris SE, Bates TC, Polderman TJ, Whalley LJ, Fox H, Starr JM, Evans PD, Montgomery GW, Fernandes C, Heutink P, Martin NG, Boomsma DI, Deary IJ, Wright MJ, de Geus EJ, Lahn BT (2007) The ongoing adaptive evolution of ASPM and Microcephalin is not explained by increased intelligence. Hum Mol Genet 16:600–608

32. Munsat TL, Davies KE (1992) International SMA consortium meeting. (26–28 June 1992, Bonn, Germany). Neuromuscul Disord 2:423–428

33. Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

34. Padberg GW, Lunt PW, Koch M, Fardeau M (1991) Diagnostic criteria for facioscapulohumeral muscular dystrophy. Neuromuscul Disord 1:231–234

35. Parman Y (2007) Hereditary neuropathies. Curr Opin Neurol 20:542–547

36. Petryshen TL, Sabeti PC, Aldinger KA, Fry B, Fan JB, Schaffner SF, Waggoner SG, Tahl AR, Sklar P (2010) Population genetic study of the brain-derived neurotrophic factor (BDNF) gene. Mol Psychiatry 15:810–815

37. Pinker S (2003) Language as an adaptation to the cognitive niche. In: Christiansen M, Kirby S (eds) Language evolution: states of the art. Oxford University Press, New York

38. Pinker S (2010) Colloquium paper: the cognitive niche: coevolution of intelligence, sociality, and language. Proc Natl Acad Sci USA 107(Suppl 2):8993–8999

39. Pozzi D, Condliffe S, Bozzi Y, Chikhladze M, Grumelli C, Proux-Gillardeaux V, Takahashi M, Franceschetti S, Verderio C, Matteoli M (2008) Activity-dependent phosphorylation of Ser187 is required for SNAP-25-negative modulation of neuronal voltage-gated calcium channels. Proc Natl Acad Sci USA 105:323–328

40. 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

41. R Development Core Team (2008) R: a language and environment for statistical computing. Vienna, Austria

42. Raber J, Mehta PP, Kreifeldt M, Parsons LH, Weiss F, Bloom FE, Wilson MC (1997) Coloboma hyperactive mutant mice exhibit regional and transmitter-specific deficits in neurotransmission. J Neurochem 68:176–186

43. Romeo S, Sentinelli F, Cavallo MG, Leonetti F, Fallarino M, Mariotti S, Baroni MG (2008) Search for genetic variants of the SYNTAXIN 1A (STX1A) gene: the -352 A > T variant in the STX1A promoter associates with impaired glucose metabolism in an Italian obese population. Int J Obes (Lond) 32:413–420

44. 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

45. Soderqvist S, McNab F, Peyrard-Janvid M, Matsson H, Humphreys K, Kere J, Klingberg T (2010) The SNAP25 gene is linked to working memory capacity and maturation of the posterior cingulate cortex during childhood. Biol Psychiatry 68:1120–1125

46. Sorensen JB, Nagy G, Varoqueaux F, Nehring RB, Brose N, Wilson MC, Neher E (2003) Differential control of the releasable vesicle pools by SNAP-25 splice variants and SNAP-23. Cell 114:75–86

47. Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet 76:449–462

48. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989

49. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

50. Takahashi N, Hatakeyama H, Okado H, Noguchi J, Ohno M, Kasai H (2010) SNARE conformational changes that prepare vesicles for exocytosis. Cell Metab 12:19–29

51. Thornton K (2003) Libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics 19:2325–2327

52. Tishkoff SA, Verrelli BC (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet 4:293–340

53. Torri F, Akelai A, Lupoli S, Sironi M, Amann-Zalcenstein D, Fumagalli M, Dal Fiume C, Ben-Asher E, Kanyas K, Cagliani R, Cozzi P, Trombetti G, Strik Lievers L, Salvi E, Orro A, Beckmann JS, Lancet D, Kohn Y, Milanesi L, Ebstein RB, Lerer B, Macciardi F (2010) Fine mapping of AHI1 as a schizophrenia susceptibility gene: from association to evolutionary evidence. FASEB J 24:3066–3082

54. Tsunoda K, Sanke T, Nakagawa T, Furuta H, Nanjo K (2001) Single nucleotide polymorphism (D68D, T to C) in the syntaxin 1A gene correlates to age at onset and insulin requirement in Type II diabetic patients. Diabetologia 44:2092–2097

55. Turner MD, Cassell PG, Hitman GA (2005) Calpain-10: from genome search to function. Diabetes Metab Res Rev 21:505–514

56. Vallender EJ (2008) Exploring the origins of the human brain through molecular evolution. Brain Behav Evol 72:168–177

57. Vander Molen J, Frisse LM, Fullerton SM, Qian Y, Del Bosque-Plata L, Hudson RR, Di Rienzo A (2005) Population genetics of CAPN10 and GPR35: implications for the evolution of type 2 diabetes variants. Am J Hum Genet 76:548–560

58. Verderio C, Pozzi D, Pravettoni E, Inverardi F, Schenk U, Coco S, Proux-Gillardeaux V, Galli T, Rossetto O, Frassoni C, Matteoli M (2004) SNAP-25 modulation of calcium dynamics underlies differences in GABAergic and glutamatergic responsiveness to depolarization. Neuron 41:599–610

59. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

60. Wechsler D (1989) WPPSI-R manual: Wechsler preschool and primary scale of intelligence-revised. The Psychological Corporation, San Antonio, TX

61. Wechsler D (1981) Examiner's manual: Wechsler intelligence scale for children revised. Psychological Corporation, New York

62. Wechsler D (1981) The Psychometric Tradition. Dev Wechsler Adult Intel Scale 6:82–85

63. Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy–Weinberg equilibrium. Am J Hum Genet 76:887–893

64. Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R (2007) Localizing recent adaptive evolution in the human genome. PLoS Genet 3:e90

65. Wright SI, Charlesworth B (2004) The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. Genetics 168:1071–1076