

Research article

Open Access

Sequence variation and linkage disequilibrium in the GABA transporter-I gene (*SLC6A1*) in five populations: implications for pharmacogenetic research

Rungnapa Hirunsatit^{1,2,3}, Risto Ilomäki⁴, Robert Malison¹, Pirkko Räsänen⁴, Essi Ilomäki⁴, Henry R Kranzler⁵, Thomas Kosten^{1,2}, Atapol Sughondhabrom⁶, Nuntika Thavichachart⁶, Sookjaroen Tangwongchai⁶, Jennifer Listman⁷, Apiwat Mutirangura⁸, Joel Gelernter^{1,2} and Jaakko Lappalainen^{*1,2}

Address: ¹Yale University School of Medicine, Department of Psychiatry, New Haven, CT, USA, ²VA Connecticut Healthcare System, West Haven, CT, USA, ³Chulalongkorn University, Inter-Department Program of Biomedical Science, Faculty of Graduate School, Bangkok, Thailand, ⁴University of Oulu, Department of Psychiatry, Oulu, Finland, ⁵University of Connecticut School of Medicine, Department of Psychiatry, Farmington, CT, USA, ⁶Chulalongkorn University, Department of Psychiatry, Bangkok, Thailand, ⁷New York University, Department of Anthropology, New York, USA and ⁸Chulalongkorn University, Department of Anatomy, Bangkok, Thailand

Email: Rungnapa Hirunsatit - rungnapa.h@student.chula.ac.th; Risto Ilomäki - rilomaki@mail.student.oulu.fi; Robert Malison - robert.malison@yale.edu; Pirkko Räsänen - pirkko.rasanen@oulu.fi; Essi Ilomäki - epalomak@paju.oulu.fi; Henry R Kranzler - Kranzler@PSYCHIATRY.UCHC.EDU; Thomas Kosten - Kosten@bcm.tmc.edu; Atapol Sughondhabrom - atapol.s@gmail.com; Nuntika Thavichachart - fmednta@md.chula.ac.th; Sookjaroen Tangwongchai - chstw@hotmail.com; Jennifer Listman - jbl245@nyu.edu; Apiwat Mutirangura - Apiwat.M@Chula.ac.th; Joel Gelernter - joel.gelernter@yale.edu; Jaakko Lappalainen^{*} - Jaakko.lappalainen@astrazeneca.com

^{*} Corresponding author

Published: 17 October 2007

BMC Genetics 2007, 8:71 doi:10.1186/1471-2156-8-71

Received: 17 January 2007

Accepted: 17 October 2007

This article is available from: <http://www.biomedcentral.com/1471-2156/8/71>

© 2007 Hirunsatit et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: GABA transporter-I (GAT-I; genetic locus *SLC6A1*) is emerging as a novel target for treatment of neuropsychiatric disorders. To understand how population differences might influence strategies for pharmacogenetic studies, we identified patterns of genetic variation and linkage disequilibrium (LD) in *SLC6A1* in five populations representing three continental groups.

Results: We resequenced 12.4 kb of *SLC6A1*, including the promoters, exons and flanking intronic regions in African-American, Thai, Hmong, Finnish, and European-American subjects (total n = 40). LD in *SLC6A1* was examined by genotyping 16 SNPs in larger samples. Sixty-three variants were identified through resequencing. Common population-specific variants were found in African-Americans, including a novel 21-bp promoter region variable number tandem repeat (VNTR), but no such variants were found in any of the other populations studied. Low levels of LD and the absence of major LD blocks were characteristic of all five populations. African-Americans had the highest genetic diversity. European-Americans and Finns did not differ in genetic diversity or LD patterns. Although the Hmong had the highest level of LD, our results suggest that a strategy based on the use of tag SNPs would not translate to a major improvement in genotyping efficiency.

Conclusion: Owing to the low level of LD and presence of recombination hotspots, *SLC6A1* may be an example of a problematic gene for association and haplotype tagging-based genetic studies. The 21-bp promoter region VNTR polymorphism is a putatively functional candidate allele for studies focusing on variation in GAT-I function in the African-American population.

Background

γ -aminobutyric acid (GABA) is the most ubiquitous inhibitory neurotransmitter in brain. Abnormal function of the GABA system has been implicated in almost every common neurological and psychiatric disorder. Augmentation of brain GABA function is the presumed therapeutic mechanism of several classes of medications, including the benzodiazepines, gabapentin, and pregabalin, which are used in treatment of many psychiatric and neurological disorders, addictions, and pain [1,2]. Identifying genetic factors responsible for variation in clinical response holds promise as a way to improve the clinical application of these medications; it may be possible to identify those who are most likely to respond to treatment and those who are at risk to develop adverse effects. For these kinds of studies, it will be important to identify polymorphisms in genes encoding the components of the brain GABA system. Although the HapMap project [3] will identify the major basic blocks of linkage disequilibrium in these genes, detailed patterns of genetic variation, including the discovery of novel functional variants, can only be accomplished through resequencing.

A novel class of medications was recently developed that act by blockade of GABA transporters (GAT), thereby inhibiting GABA uptake. Four main subtypes of GABA transporters, the GAT-1, GAT-2, GAT-3 and Betaine/GABA transporter-1 (BGT-1), have been identified through molecular cloning. The first medication in the GABA reuptake inhibitor class to become available for clinical use was tiagabine [(-)-(R)-1-(4,4-Bis(3-methyl-2-thienyl)3-butenyl) nipecotic acid hydrochloride], which selectively blocks GAT-1 sites [4]. Tiagabine was developed for treatment of seizure disorders but is being used by many psychiatrists to treat a variety of conditions in which augmentation of brain GABA function is thought to be desirable to alleviate clinical symptoms. Preliminary studies suggest that tiagabine is effective in treatment of anxiety [5], sleep disorders [6], depression [7] and addictions [2,8,9]. Animal studies suggest that most of the tiagabine's common adverse effects, such as muscle twitching and sedation, arise from specific inhibition of GAT-1 [10]. Tiagabine has a rare, but serious adverse effect – non-convulsive status epilepticus – which has raised concerns about its off-label use in treatment of psychiatric disorders [11]. Pharmacogenetic study could plausibly help to identify those subjects who are at greatest risk for developing side-effects to tiagabine and other medications that inhibit GAT-1.

The *SLC6A1* gene, which encodes the GAT-1 protein, is an obvious candidate for pharmacogenetic studies of tiagabine and other GAT-1 inhibitors [12]. In anticipation of larger pharmacogenetic studies, we aimed to identify novel genetic variation and examine linkage disequilibrium

in the *SLC6A1* gene in five populations representing three major continental groups [European (EA and Finnish), African (AA) and Asian (Thai and Hmong)]. Populations considered isolated, Finnish and Hmong, and two mixed populations, EA and AA, were examined, to understand how population differences should influence planning for pharmacogenetic studies. By comparing isolated and mixed populations, we hoped to shed further light on the purported benefits of isolated populations in mapping complex genetic traits [13,14]. Furthermore, our goal was to identify a set of haplotype tagging markers for studies focusing on response to GAT-1 inhibition. However, very low levels of LD was discovered in *SLC6A1* gene suggesting that traditional haplotype-based approaches of examining this gene in pharmacogenetic studies would be of limited utility.

Results

Algorithmic promoter prediction

The sequence for *SLC6A1*, located in contig NT_022517, which contains the sequence for human chromosome 3, was used as a reference. ElDorado software [15] located two putative *SLC6A1* promoter regions in the contig at positions 10973950–10974553 and 10998399–10998999 (+ strand). The first region is located in an area surrounding the boundary between the 5' region and exon 1 of *SLC6A1* and is estimated to be 604 bp long. Hereafter, we refer to this region as the upper promoter region. The second putative promoter region of 601 bp was located in the area flanking the boundary between intron 2 and exon 3. This region extends 70 bp across the *SLC6A1* start codon in the 3' direction in exon 3. We refer to this region as the lower promoter region.

Nucleotide diversity in *SLC6A1* in five populations

The *SLC6A1* promoter regions, 16 exons, and the flanking intronic regions were amplified and resequenced in 7–9 samples from each of the five populations. A total of 61 SNPs were found in the sequencing screening sample, including 5 SNPs in the promoter regions, 3 synonymous coding sequence SNPs, and 53 non-coding region SNPs. No non-synonymous SNPs were found. The identified *SLC6A1* variants, including frequency in the screening sample, location, and flanking sequence are provided in the additional data file 1.

In the screening sample, 33 of the 61 SNPs discovered (54%) were present in more than one population. More SNPs unique to a single population or continental group were found in the AAs than in all of the other populations combined (counting Finns and EAs, and Thai and Hmong, together). Sixteen population-specific SNPs were observed in the AAs, one in the EAs, 6 in the Finnish, 4 in the Thai, and 2 in the Hmong (see additional data file 1); i.e., a total of 13 in the populations other than AAs. Three

common population-specific SNPs were found in AA population (-24794A/G, -24126G/T and 17885 A/C, minor allele frequencies = 0.39, 0.33 and 0.22) (These estimates of population allele frequencies have to be interpreted with caution since the data were derived from sequencing only 14–18 chromosomes per population). Analysis of nucleotide diversity based on the sequencing data indicated that AAs had the greatest number of polymorphic sites ($n = 41$), highest nucleotide diversity per bp (2.77×10^{-4}) and highest Watterson's estimator of theta (θ) (2.55×10^{-4}). Inspection of nucleotide diversity in the Hmong, Finnish, EA, and Thai populations revealed no marked differences among these four populations (Table 1). As expected, nucleotide diversity was lower in exons as compared to intronic regions ($\pi_{EX} = 5.5-7.0 \times 10^{-5}$, $\pi_{INT} = 1.1-2.3 \times 10^{-4}$). Interestingly, nucleotide diversity was not higher in the AAs when exonic sequence only was considered ($\pi_{AA} = 5.9 \times 10^{-5}$, $\pi_{EA} = 6.4 \times 10^{-5}$, $\pi_{Finn} = 7.1 \times 10^{-5}$, $\pi_{Thai} = 5.6 \times 10^{-5}$, $\pi_{Hmong} = 5.6 \times 10^{-5}$). The average number of heterozygous SNPs per person in the approximately 12.4 kb of the *SLC6A1* sequence was the highest in AAs (12.22) and was lowest in the Hmong population (3.75). In Finns, the mean number of heterozygous SNPs was 8.63, in EAs it was 6.57, and in Thai it was 8.38 (Table 1). The number of heterozygous SNPs differed significantly by population (ANOVA $p < 0.0001$). *Post hoc* comparisons showed that the Hmong had fewer heterozygous SNPs than the other populations (each comparison $p < 0.009$), with the exception of EAs ($p = 0.11$). AAs had more heterozygous SNPs than each of the other populations (each comparison $p < 0.032$).

Novel length polymorphisms in the promoter region

Two novel length polymorphisms were identified in the upper promoter of *SLC6A1*. The largest, and perhaps the more interesting polymorphism, is a 21-bp VNTR polymorphism, which is present either as a single element or as two tandem repeats, hereafter referred to as the *SLC6A1* short and *SLC6A1* long allele, respectively. The more common *SLC6A1* short allele contains an A/G SNP in its 16th base (-24794 A/G) (Figure 1). In our sample, the A allele of the -24794 SNP is in perfect linkage disequilibrium with the long allele, which contains G in its 16th base. In

other words, the A allele always coincides with the long allele.

A novel two-base-pair deletion (GG/-GG) at positions (-24780 to -24781) was present in all 5 populations studied. This polymorphism is located only 8 bp 3' to the last base pair of the short/long polymorphism. The -GG allele occurs only in the context of the short allele.

The *SLC6A1* short/long and GG/-GG polymorphisms were genotyped in 46 EA, 60 AA, 59 Thai, 47 Finnish and 48 Hmong individuals. The allele frequency of the long allele was 0.39 in the AA sample but it was not found in other populations. The GG/-GG polymorphism occurred in all 5 populations. The -GG allele frequency was 0.30 in EA, 0.23 in AA, 0.33 in Finnish, 0.22 in Thai and 0.16 in Hmong.

Linkage disequilibrium in the *SLC6A1* gene

For analysis of linkage disequilibrium in *SLC6A1*, a total of 16 SNPs were genotyped in the AA, Thai, Hmong, EA, and Finnish samples. The SNPs were selected with the goal of encompassing the *SLC6A1* gene with SNPs with allele frequencies $> 10\%$ in most populations studied in order to allow comparison of LD patterns between populations. Twelve of the 16 SNPs studied met this criterion. SNP rs1710879 was virtually monomorphic in AA population. The allele frequencies of all 16 SNPs were in Hardy-Weinberg equilibrium (HWE). The LD structure of the *SLC6A1* gene, as detected by this set of SNPs, was evaluated by calculating D' and r^2 using the HAPLOVIEW program [16]. The LD structure of the *SLC6A1* gene is presented in Figure 2. Fragmentation of LD into several poorly-defined blocks was noted in all 5 populations studied (Figure 2). There were two short blocks of LD ($D' = 0.8-1$) observed in all five populations. The first block is located between the markers -29477 (Marker 1, Figure 2) and -24321 (Marker 2, Figure 2). The second LD block, which is located between markers -17590 (Marker 3, Figure 2) and -9765 (Marker 5, Figure 2), was also found in all populations studied, although the level of LD was lower in AAs. The third LD block was observed in the EA and Finnish populations between markers -1529 (Marker 6, Figure 2) and 3164 (Marker 8, Figure 2).

Table 1: Indices of *SLC6A1* nucleotide diversity in five populations.

Populations	Number of polymorphic sites	Nucleotide diversity per bp (π) (10^{-4})	Watterson's estimator of theta (θ) (10^{-4})	Mean number of heterozygous SNPs per person
European-American	21	1.72	1.41	6.57
African-American	41	2.77	2.55	12.22
Finnish	29	2.11	1.87	8.63
Thai	31	2.14	2.00	8.38
Hmong	24	1.62	1.55	3.75

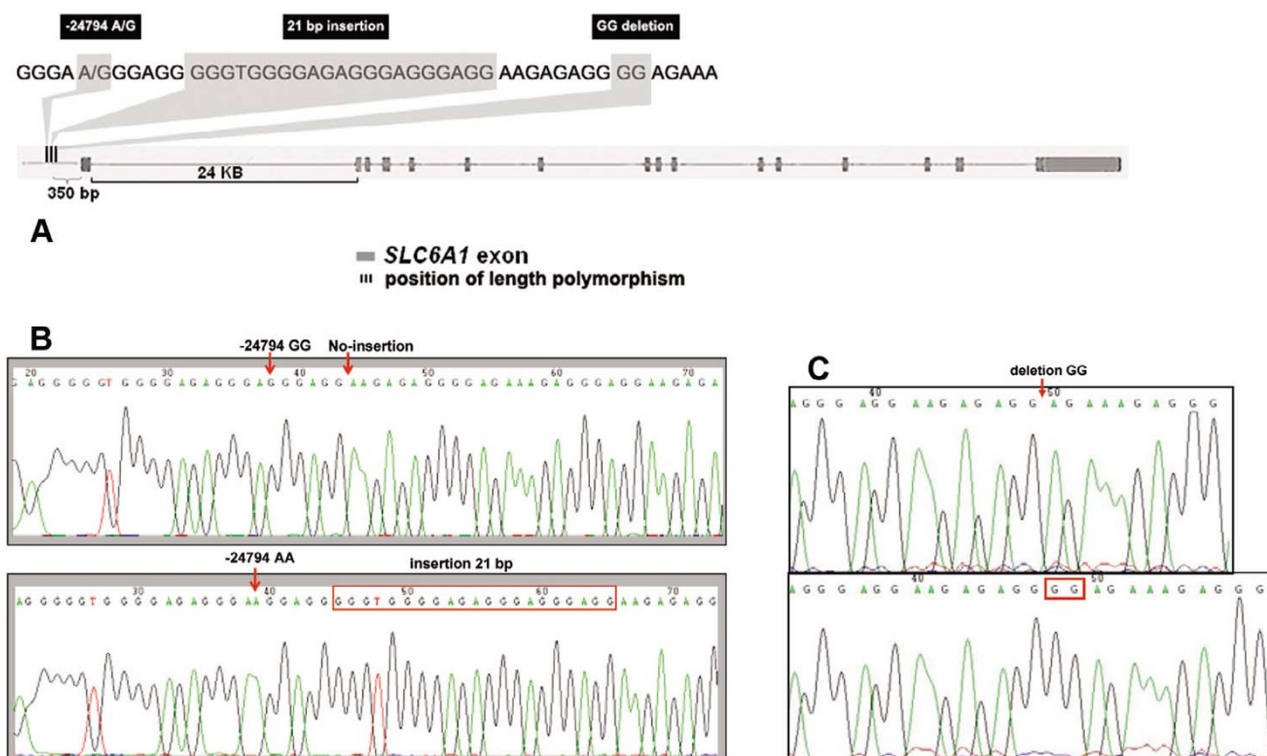
**Figure 1**

Illustration of the 21 bp insertion/deletion polymorphism (*SLC6A1* long and short alleles) and the insertion/deletion GG allele in the *SLC6A1* gene. Picture (A) shows position of -24794 A/G SNP, 21 bp insertion and GG deletion. Picture (B) shows the sequence for the *SLC6A1* short and long alleles and the -24794 A/G SNP. The A allele of the -24794 always coincided with the long allele in our populations. Sequence for the GG insertion/deletion polymorphism is presented in picture (C).

In accordance with low levels of LD, *Tagger* identified few haplotype tagging SNPs [17]. A SNP tagging approach would have allowed omission of three SNPs each in the EA, Finnish, and Thai populations and omission of two SNPs in the Hmong population. None of the 16 SNPs examined in the AA population was identified as a haplotype tagging SNP. We estimated the span of LD in different populations using $r^2/\text{distance}$ as an index. The index of LD span was about twofold higher in the Hmong population than in any of the other populations. The differences were greatest in the distance bins < 10 kb and 30–40 kb, where the median $r^2/\text{distance}$ value was two-to-three fold higher in the Hmong than in the other populations (Figure 3).

Haplotype diversity in *SLC6A1*

To further elucidate the *SLC6A1* haplotype structure, we used PHASE [18,19] to estimate haplotype frequencies in the five populations. Consistent with the observation of low levels of LD, no common *SLC6A1* haplotypes spanning the entire gene were identified in any of the populations. Haplotype dispersion varied depending on the

number of SNPs included in the haplotype. When all 16 SNPs were included in the analysis, no common haplotypes were observed. The most common haplotype had a frequency of 0.048 in the EA, 0.039 in the AA, 0.076 in the Finnish, 0.049 in the Thai and 0.097 in the Hmong populations. Common haplotypes were observed when the analysis was restricted to narrower segments of the gene, i.e., when there were computationally fewer possibilities. We noted that there were differences between populations in the composition and number of common *SLC6A1* haplotypes. To examine this further, we reconstructed each consecutive three-SNP haplotype in each population using a sliding window analysis across the panel of 16 SNPs. The four most common three-SNP haplotypes in each window and in each population were identified. We then calculated how many times each of the common three-SNP haplotypes in each window was disjoint (i.e., not shared) between the populations. A summary pairwise score was calculated for each of the populations, which is presented in Table 2. For example, of all "top-four" three-SNP *SLC6A1* haplotypes, 26.8% were disjoint between AAs and EAs.

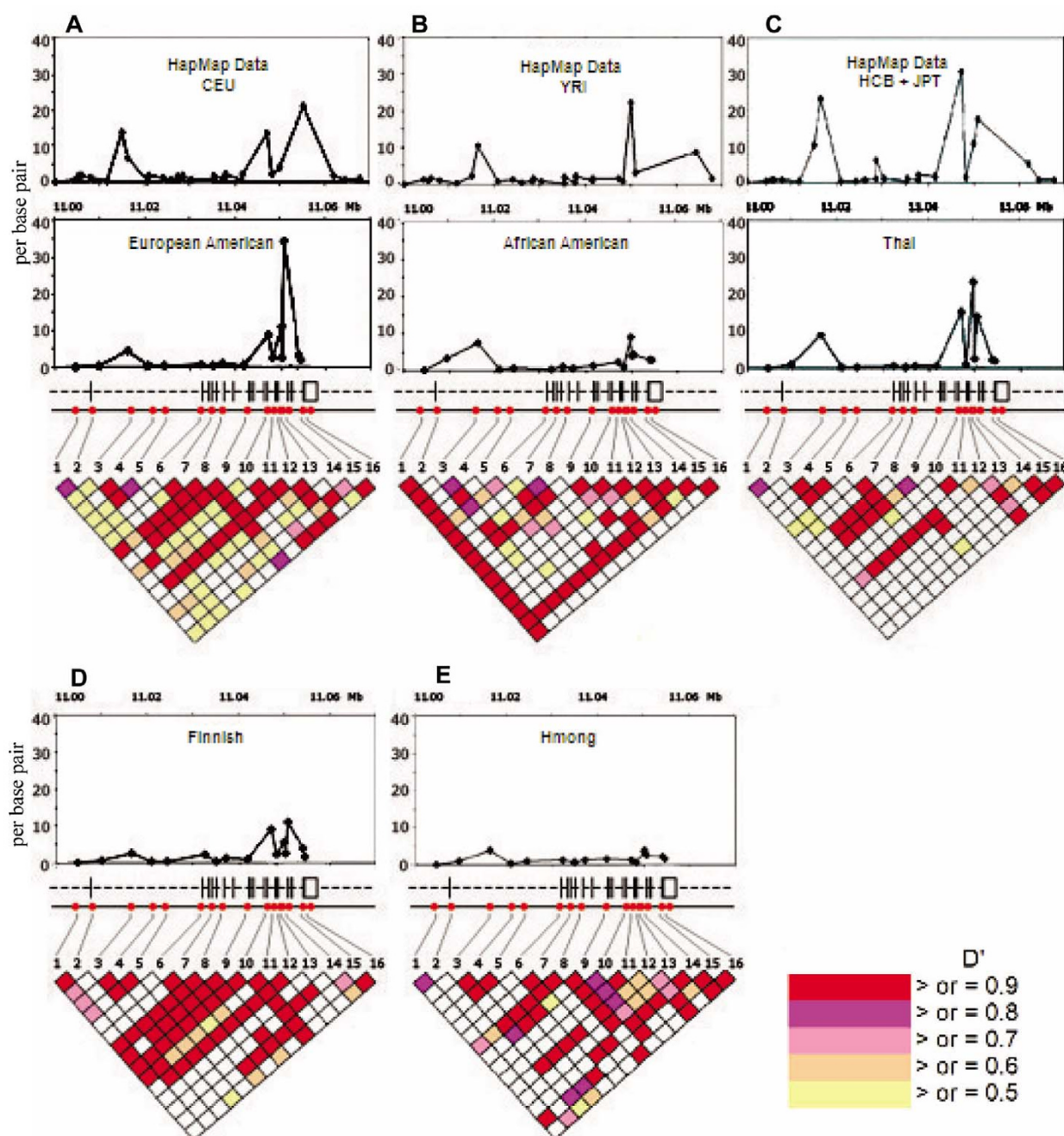
**Figure 2**

Illustration of the *SLC6A1* LD structure and recombination hotspots in the 5 populations. Upper graphs illustrate elevations from the background recombination rates across the *SLC6A1* gene. Y axis represent recombination rate and X axis represents physical distance between the markers [19, 20]. In Figure A, recombination rates for HapMap CEPH Western Europeans (CEU) and European-Americans of the present study are presented. In Figure B, recombination rates for HapMap Yoruban (YRI) and African-Americans of the present study are presented. In Figure C, recombination rates for HapMap combined Han Chinese and Japanese populations (HCB+JPT) and the Thais of the present study are presented. In Figures D and E, recombination rates for the Finns and Hmongs of the present study are presented. In the middle, the exon-intron structure of *SLC6A1* and location of the markers is presented. LD (D') between the SNPs in *SLC6A1* is illustrated in the lower triangular graphs.

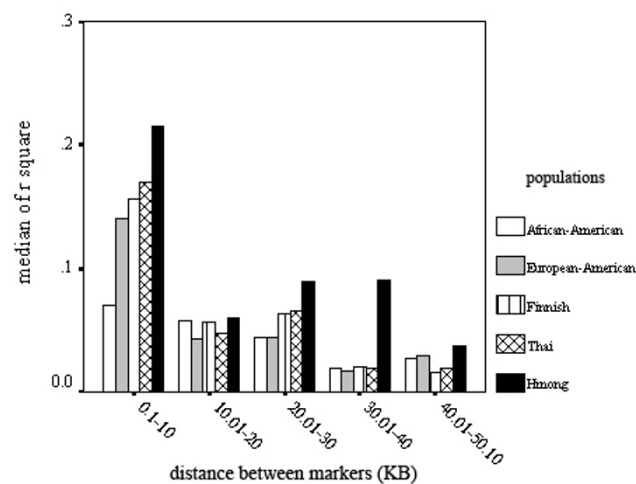


Figure 3
This graph illustrates LD span (expressed as median r^2) in different populations. Median r^2 calculated for SNP pairs in different distance bins (0.1–10 kb, 10.01–20 kb etc) is presented.

Recombination hotspots in SLC6A1

We considered recombination hotspots as an explanation for low level LD in *SLC6A1*. Recombination rates were calculated for the genotype data using the -MR and -X10 options of the PHASE program [20,21] and genotype data for all 16 *SLC6A1* SNPs. Average recombination rates in the five populations are shown in Figure 2. These data show two areas in *SLC6A1* with elevated recombination rates. The first area is demarcated by markers -29477 (Marker 1) and -13071 (Marker 4) and the second one by markers 7772 (Marker 9) and 20172 (Marker 15), hereafter referred to as the upper and lower hotspots, respectively. There were population differences in the amplitude of the hotspots (Figure 2). In the upper hotspot, which is located in the areas of intron 1 and exon 1, an increase in the recombination rate was observed in all five populations. The Hmong and AA populations showed lower rate and width of the lower hotspot, located in the area

Table 2: A summary pairwise percentage score representing the degree to which the four most common haplotypes were disjoint between the five populations.

E					
A	26.8				
F	16.1	26.8			
T	12.7	23.6	16.3		
H	26.8	23.2	30.3	17.9	
	E	A	F	T	H

E = European-American, A = African-American, F = Finnish, T = Thai, H = Hmong. For example, of all consecutive "top-four" 3-SNP *SLC6A1* haplotypes, 26.8 % were disjoint between African-Americans and European-Americans.

between exon 8 and 16, than did the EAs, Finns or Thai (Figure 2). To further elucidate these findings, we analyzed *SLC6A1* genotypes available through HapMap using PHASE, as described above. Genotyping data for the combined group of Japanese and Han Chinese, Yoruba, and Western Europeans were analyzed and compared to our results. We compared genotype data from our 16 SNPs and genotype data from 27–31 SNPs in HapMap. Two sets of SNPs, our and the HapMap SNPs, are not exactly the same but they are all located within the *SLC6A1* gene (one dot on X-axis in Figure 2 represents one SNP). It appears that genotype data from the two samples (the present study and HapMap) are consistent, both showing evidence for two areas of increase in recombination rate within *SLC6A1*. These results support our findings (Figure 2) regarding two recombination hotspots in *SLC6A1*.

Discussion

The goal of the present study was to comprehensively analyze sequence variation and linkage disequilibrium in the *SLC6A1* gene in anticipation of larger pharmacogenetic studies of tiagabine and other GAT-1 inhibitors. Resequencing 12.4 kb of the *SLC6A1* gene, including all 16 exons and the two putative promoter regions, revealed numerous novel genetic variants. Perhaps the most interesting polymorphism identified was a 21 bp VNTR polymorphism located in the upper promoter sequence of the *SLC6A1* gene (Figure 1). We have termed the alleles as the "SLC6A1 short," which has one copy of the allele and "SLC6A1 long," in which the allele is duplicated (Figure 1). Interestingly, the long allele was common in AAs (39%), while in the other populations, it was absent. A likely explanation for the lack of this allele in non-African populations is genetic drift. However, other explanations, such as natural selections, are also possible. Functional studies focusing on understanding whether the short and long allele lead to differential expression of the GAT-1 protein are clearly warranted; these studies may also help in elucidating whether the allele frequency discrepancy between African-Americans and other populations is due to genetic drift or selection. Previously, a comparable promoter region VNTR polymorphism was described in the serotonin transporter gene, which is known to influence the expression of the gene [22]. Several studies have shown that this polymorphism partially accounts for differences in the therapeutic response to serotonin selective reuptake inhibitors (SSRIs) [23-26], susceptibility to depression [27,28] and alcohol dependence [29]. The *SLC6A1* short/long is a candidate polymorphism for moderating the response to tiagabine and susceptibility to neuropsychiatric disorders in which GABA dysfunction may play a role, albeit only in AA populations (and any other populations where this variant is present).

Genetic diversity in *SLC6A1* revealed through *SLC6A1* resequencing showed interesting results. Although a limited number of chromosomes was examined, certain trends were found. First, no non-synonymous SNPs were discovered in the 80 chromosomes sequenced, suggesting that the coding sequence of *SLC6A1* has been conserved against common amino-acid altering substitutions through active background selection. Consistent with these results, the nucleotide diversity was much lower in the *SLC6A1* exons compared to the intronic regions examined [30-32]. Comparison of the sequence data between populations revealed higher nucleotide diversity in AAs than in other populations (Table 1). In accordance with these data, the only population in which we found common population specific SNPs, were AAs. In exons, however, nucleotide diversity was no higher in the AA population than in the other populations. In other populations, the SNPs observed in only one population were rarer. Although Finns and Hmong are considered to be isolated populations, nucleotide diversity in these two populations was not different from that observed in EAs or Thais. Overall, no major differences in nucleotide diversity were observed among the Hmong, Thai, EA, and Finnish populations. Together, these findings most likely reflect the older age of the African population relative to the other populations, which had allowed more intronic variation to accumulate in this population, founder effects in non-African populations, and selection pressure conserving the *SLC6A1* exonic sequence. The extent of conservation of GAT-1 amino acid sequence suggests an important role of this protein in normal brain function. The Hmong had a significantly lower number of heterozygous SNPs in comparison to the other populations. One explanation for this finding is that the Hmong subjects may have been distantly related. We postulate that differences in the degree and age of population bottlenecks between the Hmong and the other populations are less likely explanations for the lower heterozygosity observed in the Hmong; we feel that this explanation is less likely because all non-African populations had a low observed frequency of population-specific SNPs and because Hmong nucleotide diversity was not significantly lower than that of the other non-African populations [33]. A caveat of this study is that the sample size per ethnicity was small. Consequently, rare non-synonymous SNPs, specific to a population, would have not been identified. It may be useful to resequence larger samples to identify these kinds of variations, at least in the primary target populations of clinical trials. Laboratory methods established for the present study should facilitate such analyses. Another limitation of the study is that the resequencing effort focused on exons. If deep intronic variation in *SLC6A1* contributes to functional variation at the protein level, those variants would have been missed in the present study. A low level of LD in *SLC6A1* was

observed in all five populations (Figure 2). Consistent with these results, in the EA, Thai, Hmong, and Finnish populations, only two or three haplotype tagging SNPs in the areas of preserved LD were identified. In the AAs, no haplotype tagging SNPs were found in the 16 SNPs genotyped. These results suggest that very dense SNP panels would be required to capture common variation in this gene. Using r^2 /distance as the index, a longer LD span was observed in the Hmong population than in the other populations (Figure 3). However, higher LD probably would not translate to significant practical improvement in genotyping efficiency overall, because there were no major differences in the number of haplotype blocks in Hmong than in the Finnish, Thai and EA populations. Considering the low level of LD, *SLC6A1* may pose special challenges for association studies both in isolated and mixed populations. The common *SLC6A1* 3-SNP haplotypes were largely the same in the five populations (Table 2), but were not completely overlapping. These results suggest a certain degree of, but not absolute, portability of SNP genotyping sets between the populations. It would be interesting to study larger EA, AA, Hmong, Thai and Finnish population samples to further refine the structure and frequencies of the common overlapping and population-specific haplotypes [34]. In addition, it will be interesting to study whether haplotype and SNP profile characteristics, such as absence of common non-synonymous substitutions extends to patient populations suffering from various neuropsychiatric disorders, which were not studied here. The present study primarily focused on non-clinical samples and therefore no data were available to assess whether disease associated variants are present in human populations.

Intrigued by these findings, we examined whether recombination hotspots could explain low levels of LD in *SLC6A1*. Two hotspots were identified using PHASE; the first is located in the areas of exon 1 and intron 1. The second hotspot is located in the area demarcated by exons 8 and 16. As expected, within the hotspots, D' fell off rapidly. For example, in the Finns, in the area of the distal hotspot, D' was only 0.184 between markers 12 and 13, which are spaced 107 bp apart and D' was 0.044 between 13 and 14, which are spaced 489 bp apart. Areas of high recombination, such as seen in *SLC6A1*, potentially limit large-scale association studies, as it would be exceptionally difficult to find risk alleles relying on linkage disequilibrium if the alleles were located inside a recombination hotspot.

Conclusion

SLC6A1 is a complicated target for pharmacogenetic studies because low levels of LD and recombination hotspots would make it difficult to establish an association between genetic markers and response to GAT-1 modula-

tion. Furthermore, our study suggests that focusing on isolated populations, such as Finns or Hmong, would not provide major benefits to genotyping efficacy of *SLC6A1*. An interesting 21-bp VNTR polymorphisms (*SLC6A1* short and *SLC6A1* long) was discovered in the promoter sequence of the *SLC6A1* gene, which is common in African-Americans. This new polymorphism is a novel candidate for studies focusing on genetically influenced differences in GAT-1 expression, and hence, response to medications that inhibit GAT-1 function.

Methods

DNA samples

For re-sequencing of the *SLC6A1* gene, 40 genomic DNA samples were collected from unrelated individuals representing 5 different populations: EA (n = 7), AA (n = 9), Finnish (n = 8), Thai (n = 8), and Hmong (n = 8). The Finnish subjects were unrelated parents of adolescent subjects who were participating in an epidemiological study focusing on the identification of risk factors for early-onset mental illness and substance dependence in Finland [35]. The Thai and Hmong populations were collected in Thailand as part of an ongoing genetic association and population genetic study. The Thais selected for re-sequencing had grandparents and parents of Thai ancestry (Thai-Thai) or had mixed Thai and Chinese ancestry (Thai-Chinese), by subject report. These samples were obtained from a blood drive in Bangkok, Thailand. The Hmong subjects were recruited in a Hmong village in the northern part of Thailand. The AA and EA samples have been described earlier elsewhere [36]. Both EA and AA samples were self-identified and confirmed as such by Bayesian marker clustering [36]. All subjects provided informed consent as approved by the appropriate institutional review boards. In addition, 46 EA, 60 AA, 59 Thai, 47 Finnish and 48 Hmong individuals were genotyped for 16 *SLC6A1* SNPs to examine linkage disequilibrium (LD) in this gene. The Thai subjects selected for examination of linkage disequilibrium were Thai-Thai. Recruitment and population characteristics of subjects selected for *SLC6A1* genotyping were identical of subjects selected for re-sequencing [35,36]. The participants were recruited from non-clinical populations. No detailed medical information was available for most of the participants. Therefore biases deriving from undetected medical conditions of the control sample could not be controlled. All studies described in this article were conducted according to the Declaration of Helsinki. The studies were approved by the institutional review boards of Yale University School of Medicine, West Haven VA Hospital, Northern- Ostrobothnia Hospital District (University of Oulu, Finland) and Chulalongkorn University (Bangkok, Thailand). All subjects signed a written informed consent for participation in this study.

Promoter prediction

The ElDorado program of the Genomatix software package was used to predict the location of the *SLC6A1* promoter region [15]. The sequence of the *SLC6A1* gene submitted to the promoter region analysis was obtained from the National Center for Biotechnology Information (NCBI) [37].

Amplification and sequencing

For sequencing of *SLC6A1*, the upper and lower promoter regions, all 16 *SLC6A1* exons (total of 4.4 kb) and 7.3 kb of flanking intronic regions were amplified. About 70 bp of the predicted 601 bp of the lower promoter region were not included in the sequence analysis. Approximately 12.4 kb of the *SLC6A1* gene was amplified, corresponding to about 25% of the total length of the gene. All primers were designed with the PRIMER3 software [38]. Primers were obtained from Invitrogen (Carlsbad, CA). PCR amplification was optimized before sequencing by testing different cycling conditions. Betaine (Sigma Aldrich, St. Louis, MO) at 0.5–1 M final concentrations was added to the reactions, as needed, to enhance specificity and yield of PCR amplification. PCR reactions were carried out in 15 µl volumes containing 20 ng genomic DNA, 200 µM of dNTPs mix (Stratagene, La Jolla, CA), 1 µM of mixed primers forward and reverse, 1X PC2 buffer, 0.75 U of KlenTaq1™ (Ab Peptides, St Louis, MO) and 0.5–1 M betaine when needed. Thermocycling conditions consisted of an initial denaturation step at 95°C for 5 min, 30 cycles of denaturation step at 95°C for 30 sec, an annealing step at 60–65°C 30 sec, and an extension step at 72°C. The duration of the extension step varied from 30 sec to 2 min depending on the length of the amplicon. After optimization, genomic DNA samples from each population were PCR amplified followed by purification with MinElute PCR purification columns (Qiagen, Valencia, CA) or the reaction mixtures were treated with ExoSAP-IT (USB, Cleveland, OH) to remove excess nucleotides and primers. Purified PCR samples were sequenced in the forward and reverse directions at Yale University W.M Keck Foundation Biotechnology Resource Laboratory. Sequencing reactions were conducted using the BigDye Terminator v3.1 cycle sequencing kit and an ABI 9800 Thermocycler (Applied Biosystem, Foster city, CA). Sequencing reactions were analyzed on an ABI 3730 xl DNA Analyzer (Applied Biosystems, Foster city, CA). Owing to technical problems, approximately 300 bp in exon 7 and intron 8 (2.4% from total 12.4 kb sequenced region) is missing in the sequencing data in Thai and Hmong populations (see additional data file 1).

Owing to the repeat elements contained in the identified upper promoter sequence and homologous sequences within the *SLC6A1*, the upper promoter region and parts of exon 1 were amplified using nested PCR. In addition,

for amplification of a 180 bp fragment located in the junction of the 5' upstream region and exon 1, a region which is very high in CG content, 7-deaza-dGTP (New England BioLabs, Beverly, MA) was added to the reactions.

Genotyping and linkage disequilibrium study

A total of 16 SNPs were chosen for genotyping in population samples to examine haplotype structure of the *SLC6A1* gene. Nine SNPs chosen for genotyping were identified through resequencing: -24321A/C, -1529A/G, 949A/G, 3164C/T, 14351A/G, 16009A/G, 16116C/T, 20172C/T and 20622A/G. The remaining seven SNPs, -29477C/T, -17590C/T, -13071A/G, -9765C/T, 7772A/G, 13269C/T, and 16605C/T, were chosen from the NCBI dbSNP [39] collection. Of the 16 SNPs studied, 14 were available through Applied-Biosystem's Assay-On-Demand service (Applied Biosystems, Foster city, CA). One assay was custom designed and obtained through the ABI's Assay-by-Design service (Applied Biosystems, Foster city, CA). PCR amplification of the 5' nuclease assays were conducted using 1 ng of DNA, 1X TaqMan universal PCR master mix (Applied Biosystems, Foster city, CA), 0.5X SNP genotyping assay mix [Applied Biosystems, Foster city, CA]. PCR conditions were as follows: denaturation step of 95°C for 10 min, followed by 50 cycles of 95°C for 15 sec and 60°C for 1 min. Amplification was performed on PTC-200 cyclers (MJ Research, Hercules, CA) and data were analyzed using the ABI Prism 7900HT Sequence Detector System and software version 2.1 (Applied Biosystem, Foster city, CA). All samples were run in duplicate for quality control purposes. Based on comparison of the duplicate runs, we estimated the genotyping error rate to be less than 0.05%. The -24321A/C SNP was genotyped using 7-deaza-dGTP sequencing because its location inside a GC-rich region made it very difficult to design a 5' nuclease assay for this SNP.

Genotyping of the length polymorphisms

Amplification of the region containing the 21 bp short/long VNTR and 2 bp GG/-GG insertion/deletion polymorphisms was accomplished using primers 5'AAGGAGAGAGATTGGAGCG 3' and 5'CTTCTTTCCTCTCGCATTC 3' (Invitrogen, Carlsbad, CA). PCR reactions were conducted in 15 µl volumes containing 20 ng genomic DNA, 200 µM of dNTPs mix (Stratagene, La Jolla, CA), 1 µM of mixed reverse and forward primers, 1X PC2 buffer, 0.75 U of KlenTaq1™ (Ab Peptides, St Louis, MO) and 1 M Betaine. The thermocycling conditions consisted of an initial step at 95°C for 5 min, 30 cycles of denaturation at 95°C for 30 sec, annealing 60°C 30 sec, and extension 72°C 30 sec. The lengths of the PCR products corresponding to the long and short alleles are 166 bp and 145 bp. The long and short alleles were separated using 3% metaphore agarose and gel electrophoresis (ISC BioExpress, Kaysville, UT). The GG/-GG insertion/deletion polymorphism was

genotyped using direct sequencing of the PCR product as described in above.

Statistical analysis

Indices of sequence variation in *SLC6A1* were calculated using a web application SLIDER [40]. These indices included the number of polymorphic sites, nucleotide diversity per base pair (π) and the Watterson's estimator of theta (θ). Nucleotide diversity per base pair (π) describes the mean number of differences per site between two sequences chosen at random from a sample of sequences. The Watterson's estimator of theta (θ) is the observed number of SNPs adjusted for the sample size and new mutation rate expected to occur in each generation [41]. In addition, for each subject we calculated the number of heterozygous SNPs observed in the sequence data. The number of heterozygous SNPs was compared between populations using ANOVA followed by post hoc Fisher's Least Significant Difference-test.

PHASE software, which implements a Bayesian algorithm for haplotype reconstruction, was used to estimate haplotype frequencies [18,19]. PHASE's options -X10 and -MR were used to estimate recombination rates across *SLC6A1* [20,21]. The value on the Y-axis of Figure 2 shows changes in recombination parameter (ρ) per base pair of *SLC6A1* exceeding the background recombination rate [20,21]. The average recombination rate was estimated based on 1,000 burn-ins and 1,000 iterations. Recombination frequencies at *SLC6A1* were compared visually between our and HapMap data. No statistical analyses were performed. To evaluate haplotype diversity among populations, we studied how often the most common haplotypes were shared or disjoint. The haplotypes were identified using a sliding window analysis across every three consecutive *SLC6A1* SNPs. The four most common three-SNP haplotypes in each window and in each population were identified. The rationale for choosing the four most common haplotypes for this analysis was that visual inspection of the haplotype frequencies told us that in each window and in each population virtually all variation in haplotype diversity was captured by the four most common haplotypes. The average percent of all haplotypes captured by the top four haplotypes was 96%. We then calculated how many times each of the common three-SNP haplotype was disjoint between the populations. A summary pairwise score derived for the populations is presented in Table 2.

LD patterns in *SLC6A1* were visualized using HAPLOVIEW version 3.2 [16]. We used the *Tagger* algorithm, implemented in HAPLOVIEW, to search for haplotype tagging SNPs in *SLC6A1*. [17] We used the default *Tagger* thresholds $r^2 > 0.8$ and LOD score > 3 . POWERMARKER [42] was used to calculate allele frequencies and examination

of Hardy-Weinberg equilibrium (HWE). To illustrate differences in the span of LD in the five populations, r^2 was plotted against physical distance. To do this, r^2 was calculated for all SNP pairs. Because these values were not normally distributed, median values are presented. Physical distance (bp) was divided into distance bins to illustrate population differences in LD span across a range of physical distances. Median r^2 in distance bins (0.1–10 kb, 10.01–20 kb etc) in different populations is presented in Figure 3. No statistical analyses were performed on these data. Software CENSOR was used to search for repeat elements within the recombination hotspots [43,44].

Abbreviations

GABA γ -aminobutyric acid

GAT GABA transporter

BGT-1 betaine/GABA transporter-1

SLC6A1 solute carrier family 6 (neurotransmitter transporter, GABA), member 1

LD linkage disequilibrium

SNPs single nucleotide polymorphisms

ANOVA analysis of variance

-MR recombination model

-X10 run 10 times

π nucleotide diversity

θ Watterson's estimator of zeta

EA European-American

AA African-American

HWE Hardy-Weinberg equilibrium

LOD log of the likelihood odds ratio

Competing interests

The author(s) declares that there are no competing interests.

Authors' contributions

JL (Jaakko Lappalainen) initiated and planned to study SLC6A1 gene. RH did the experiment and drafted the manuscript. Finnish samples in this project were collected by RI, PR and EI. African-American and European-American samples were collected by HK, TK and JG. RM, JG, AS,

NT, ST, JL (Jennifer Listman) and AM participated in collecting samples in Thai and Hmong populations. All authors have read and approved the final version of this manuscript.

Additional material

Additional file 1

List of SNPs discovered in SLC6A1 by resequencing 40 individuals from Thai ($n = 8$), Hmong ($n = 8$), European-American ($n = 7$), African-American ($n = 9$) and Finnish ($n = 8$) populations. The data provided descriptions of all SNPs in SLC6A1 in the population samples.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2156-8-71-S1.doc>]

Acknowledgements

Supported by the National Institutes of Health (K08 AA13732, R01 AA11330, P50 AA12870, M01 RR06192, K05 AA14906, K24 AA13736, R01 AA011321-04, P50-AA03510, K24 MH64122, P50 DA018197), the Alcoholic Beverage Medical Research Foundation (ABMRF), the US Department of Veterans Affairs [VA Alcohol Research Center, VA Mental Illness Research, Education and Clinical Center (VA MIRECC), and VA Veterans Research Enhancement Award Program (VA REAP)] and Thailand Research Fund. Technical help of Greg Dalton-Key, Ann Marie Lacobelle Wantroba and Katarzyna Kanarek is also greatly appreciated. Special thanks to Chupong Ittiwut for help with the illustrations.

References

1. LaRoche SM, Helmers SL: **The new antiepileptic drugs: clinical applications.** *Jama* 2004, **291**(5):615-620.
2. Sofuoglu M, Kosten TR: **Novel approaches to the treatment of cocaine addiction.** *CNS Drugs* 2005, **19**(1):13-25.
3. Thorisson GA, Smith AV, Krishnan L, Stein LD: **The International HapMap Project Web site.** *Genome Res* 2005, **15**(11):1592-1593.
4. Fink-Jensen A, Suzdak PD, Swedberg MD, Judge ME, Hansen L, Nielsen PG: **The gamma-aminobutyric acid (GABA) uptake inhibitor, tiagabine, increases extracellular brain levels of GABA in awake rats.** *Eur J Pharmacol* 1992, **220**(2-3):197-201.
5. Rosenthal M: **Tiagabine for the treatment of generalized anxiety disorder: a randomized, open-label, clinical trial with paroxetine as a positive control.** *J Clin Psychiatry* 2003, **64**(10):1245-1249.
6. Mathias S, Wetter TC, Steiger A, Lancel M: **The GABA uptake inhibitor tiagabine promotes slow wave sleep in normal elderly subjects.** *Neurobiol Aging* 2001, **22**(2):247-253.
7. Carpenter LL, Schecter JM, Tyrka AR, Mello AF, Mello MF, Haggarty R, Price LH: **Open-label tiagabine monotherapy for major depressive disorder with anxiety.** *J Clin Psychiatry* 2006, **67**(1):66-71.
8. Gonzalez G, Sevarino K, Sofuoglu M, Poling J, Oliveto A, Gonsai K, George TP, Kosten TR: **Tiagabine increases cocaine-free urines in cocaine-dependent methadone-treated patients: results of a randomized pilot study.** *Addiction* 2003, **98**(11):1625-1632.
9. Sofuoglu M, Mouratidis M, Yoo S, Culligan K, Kosten T: **Effects of tiagabine in combination with intravenous nicotine in overnight abstinent smokers.** *Psychopharmacology (Berl)* 2005, **181**(3):504-510.
10. Chiu CS, Brickley S, Jensen K, Southwell A, McKinney S, Cull-Candy S, Mody I, Lester HA: **GABA transporter deficiency causes tremor, ataxia, nervousness, and increased GABA-induced tonic conductance in cerebellum.** *J Neurosci* 2005, **25**(12):3234-3245.
11. Flowers CM, Racoosin JA, Kortepeter C: **Seizure activity and off-label use of tiagabine.** *N Engl J Med* 2006, **354**(7):773-774.

12. Kurland L, Lind L, Melhus H: **Using genotyping to predict responses to anti-hypertensive treatment.** *Trends Pharmacol Sci* 2005, **26(9)**:443-447.
13. Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, Todd JA: **The genetically isolated populations of Finland and sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes.** *Nat Genet* 2000, **25(3)**:320-323.
14. Shifman S, Darvasi A: **The value of isolated populations.** *Nat Genet* 2001, **28(4)**:309-310.
15. **The Genomatix software package** [<http://www.genomatix.de>].
16. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21(2)**:263-265.
17. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D: **Efficiency and power in genetic association studies.** *Nat Genet* 2005, **37(11)**:1217-1223.
18. Stephens M, Smith NJ, Donnelly P: **A new statistical method for haplotype reconstruction from population data.** *Am J Hum Genet* 2001, **68(4)**:978-989.
19. Stephens M, Donnelly P: **A comparison of bayesian methods for haplotype reconstruction from population genotype data.** *Am J Hum Genet* 2003, **73(5)**:1162-1169.
20. Li N, Stephens M: **Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data.** *Genetics* 2003, **165(4)**:2213-2233.
21. Crawford DC, Bhargava T, Li N, Hellenthal G, Rieder MJ, Nickerson DA, Stephens M: **Evidence for substantial fine-scale variation in recombination rates across the human genome.** *Nat Genet* 2004, **36(7)**:700-706.
22. Heils A, Teufel A, Petri S, Stober G, Riederer P, Bengel D, Lesch KP: **Allelic variation of human serotonin transporter gene expression.** *J Neurochem* 1996, **66(6)**:2621-2624.
23. Smeraldi E, Zanardi R, Benedetti F, Di Bella D, Perez J, Catalano M: **Polymorphism within the promoter of the serotonin transporter gene and antidepressant efficacy of fluvoxamine.** *Mol Psychiatry* 1998, **3(6)**:508-511.
24. Zanardi R, Benedetti F, Di Bella D, Catalano M, Smeraldi E: **Efficacy of paroxetine in depression is influenced by a functional polymorphism within the promoter of the serotonin transporter gene.** *J Clin Psychopharmacol* 2000, **20(1)**:105-107.
25. Pollock BG, Ferrell RE, Mulsant BH, Mazumdar S, Miller M, Sweet RA, Davis S, Kirshner MA, Houck PR, Stack JA, Reynolds CF, Kupfer DJ: **Allelic variation in the serotonin transporter promoter affects onset of paroxetine treatment response in late-life depression.** *Neuropsychopharmacology* 2000, **23(5)**:587-590.
26. Yu YW, Tsai SJ, Chen TJ, Lin CH, Hong CJ: **Association study of the serotonin transporter promoter polymorphism and symptomatology and antidepressant response in major depressive disorders.** *Mol Psychiatry* 2002, **7(10)**:1115-1119.
27. Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, McClay J, Mill J, Martin J, Braithwaite A, Poulton R: **Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene.** *Science* 2003, **301(5631)**:386-389.
28. Kaufman J, Yang BZ, Douglas-Palumberi H, Houshyar S, Lipschitz D, Krystal JH, Gelernter J: **Social supports and serotonin transporter gene moderate depression in maltreated children.** *Proc Natl Acad Sci U S A* 2004, **101(49)**:17316-17321.
29. Feinn R, Nellissery M, Kranzler HR: **Meta-analysis of the association of a functional serotonin transporter promoter polymorphism with alcohol dependence.** *Am J Med Genet B Neuropsychiatr Genet* 2005, **133(1)**:79-84.
30. Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF: **DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene.** *Nat Genet* 1998, **19(3)**:233-240.
31. Solus JF, Arietta BJ, Harris JR, Sexton DP, Steward JQ, McMunn C, Ihrie P, Mehall JM, Edwards TL, Dawson EP: **Genetic variation in eleven phase I drug metabolism genes in an ethnically diverse population.** *Pharmacogenomics* 2004, **5(7)**:895-931.
32. Badagnani I, Chan W, Castro RA, Brett CM, Huang CC, Stryke D, Kawamoto M, Johns SJ, Ferrin TE, Carlson EJ, Burchard EG, Giacomini KM: **Functional analysis of genetic variants in the human concentrative nucleoside transporter 3 (CNT3; SLC28A3).** *Pharmacogenomics J* 2005, **5(3)**:157-165.
33. Nielsen R: **Molecular signatures of natural selection.** *Annu Rev Genet* 2005, **39**:197-218.
34. Cheong HS, Lee SO, Choi CB, Sung YK, Shin HD, Bae SC: **MERTK polymorphisms associated with risk of haematological disorders among Korean SLE patients.** *Rheumatology (Oxford)* 2007, **46(2)**:209-214.
35. Makikyro TH, Hakko HH, Timonen MJ, Lappalainen JA, Ilomaki RS, Marttunen MJ, Lakso K, Rasanen PK: **Smoking and suicidality among adolescent psychiatric patients.** *J Adolesc Health* 2004, **34(3)**:250-253.
36. Yang BZ, Zhao H, Kranzler HR, Gelernter J: **Practical population group assignment with selected informative markers: characteristics and properties of Bayesian clustering via STRUCTURE.** *Genet Epidemiol* 2005, **28(4)**:302-312.
37. **The National Center for Biotechnology Information (NCBI)** [<http://www.ncbi.nlm.nih.gov/>].
38. **The PRIMER3 software** [http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi].
39. **The NCBI dbSNP** [<http://www.ncbi.nlm.gov/SNP>].
40. **Web application SLIDER** [<http://genapps.uchicago.edu/index.html>].
41. Bamshad M, Wooding SP: **Signatures of natural selection in the human genome.** *Nat Rev Genet* 2003, **4(2)**:99-111.
42. **The POWERMARKER software** [<http://www.powermarker.net>].
43. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: **Rebase Update, a database of eukaryotic repetitive elements.** *Cytogenet Genome Res* 2005, **110(1-4)**:462-467.
44. **Web application CENSOR** [<http://www.girinst.org>].

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

