

ARTICLE

A summary statistic approach to sequence variation in noncoding regions of six schizophrenia-associated gene loci

Jane Winantea^{1,4}, My Ngo Hoang^{1,4}, Stefanie Ohlraun², Marcella Rietschel², Sven Cichon^{1,3}, Peter Propping¹, Markus M Nöthen^{1,3}, Jan Freudenberg^{*,1,5} and Yun Freudenberg-Hua¹

¹Institute of Human Genetics, University of Bonn, Bonn, Germany; ²Division of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Mannheim, Germany; ³Department of Genomics, Life and Brain Center, Bonn, Germany

In order to explore the role of noncoding variants in the genetics of schizophrenia, we sequenced 27 kb of noncoding DNA from the gene loci RAC-alpha serine/threonine-protein kinase (*AKT1*), brain-derived neurotrophic factor (*BDNF*), dopamine receptor-3 (*DRD3*), dystrobrevin binding protein-1 (*DTNBP1*), neuregulin-1 (*NRG1*) and regulator of G-protein signaling-4 (*RGS4*) in 37 schizophrenia patients and 25 healthy controls. To compare the allele frequency spectrum between the two samples, we separately computed Tajima's *D*-value for each sample. The results showed a smaller Tajima's *D*-value in the case sample, pointing to an excess of rare variants as compared to the control sample. When randomly permuting the affection status of sequenced individuals, we observed a stronger decrease of Tajima's *D* in 2400 out of 100 000 permutations, corresponding to a *P*-value of 0.024 in a one-sided test. Thus, rare variants are significantly enriched in the schizophrenia sample, indicating the existence of disease-related sequence alterations. When categorizing the sequenced fragments according to their level of human–rodent conservation or according to their gene locus, we observed a wide range of diversity parameter estimates. Rare variants were enriched in conserved regions as compared to nonconserved regions in both samples. Nevertheless, rare variants remained more common among patients, suggesting an increased number of variants under purifying selection in this sample. Finally, we performed a heuristic search for the subset of gene loci, which jointly produces the strongest difference between controls and cases. This showed a more prominent role of variants from the loci *AKT1*, *BDNF* and *RGS4*. Taken together, our approach provides promising strategy to investigate the genetics of schizophrenia and related phenotypes. *European Journal of Human Genetics* (2006) 14, 1037–1043. doi:10.1038/sj.ejhg.5201664; published online 31 May 2006

Keywords: sequence conservation; gene control region; nucleotide diversity; single nucleotide polymorphism; schizophrenia

*Correspondence: Dr J Freudenberg, Institute of Human Genetics, University of Bonn, Wilhelmstrasse 31, Bonn D-53111, Germany.
Tel: +49 228 287 2581; Fax: +49 228 287 2380;
E-mail: jan.freudenberg@uni-bonn.de
All SNPs and genotypes were submitted to dbSNP (ss51505044-ss51505151).

⁴These authors contributed equally to this work.

⁵Current address: Laboratories of Neurogenetics, UCSF, 1550 4th Street, San Francisco, CA 94143-2922, USA
Received 18 October 2005; revised 18 April 2006; accepted 25 April 2006; published online 31 May 2006

Introduction

During the past years, several genomic loci were reproducibly found to be associated with schizophrenia.¹ However, the exact identification of the underlying causative mutations remained elusive.² Owing to the absence of coding disease mutations, an important role of noncoding variants can be hypothesized. In order to approach this hypothesis, we applied a sequencing strategy to search for noncoding

mutations among affected individuals. As complete non-coding regions are prohibitively large, we focused on gene control regions, that is fragments located around transcription start sites. This choice was made under the assumption that functional noncoding variants may affect regulation of gene expression. Within gene control regions, preference was given to genomic regions that are highly conserved between human and rodents. Studies of human single nucleotide polymorphisms (SNPs) showed that these regions display a signature of purifying selection,³ what makes them first choice targets in the search for disease mutations. Although some examples for phenotypic effects of sequence variation in conserved noncoding regions exist,⁴ according to our knowledge no sequencing study so far has focused on nucleotide diversity of these regions among individuals affected with a complex disease phenotype.

In the present study, we sequenced 27 kb of genomic DNA from six schizophrenia-associated gene loci in 37 schizophrenic and 25 healthy individuals. Assuming that variants of higher penetrance are more likely to lead to several affected family members, exclusively patients with at least one affected first-grade relative were included. We targeted fragments belonging to the gene loci neuregulin-1 (*NRG1*), dystrobrevin binding protein-1 (*DTNBP1*), regulator of G-protein signaling-4 (*RGS4*), dopamine receptor-3 (*DRD3*), RAC-alpha serine/threonine-protein kinase (*AKT1*) and the brain-derived neurotrophic factor (*BDNF*). The schizophrenia-associated gene loci *NRG1*⁵⁻⁹ *DTNBP1*¹⁰⁻¹⁴ *RGS4*¹⁵⁻¹⁸ encode proteins, which play a role in synaptic function. Among other functions, *NRG1* participates in glutamatergic signaling by regulating the N-methyl-D-aspartate (NMDA) receptor through the interaction of the NRG1 protein and its receptors.⁵ Also *DTNBP1* appears to influence schizophrenia risk through effects on glutamate function.^{19,20} *RGS4* appears to modulate activity of some serotonergic and metabotropic glutamatergic receptors,²¹ whereas its own expression is modulated by dopaminergic signaling.²² The involvement of the dopamine receptor *DRD3* in schizophrenia is supported by a meta-analysis of association studies.²³ and its role as a drug target. Supported by similar lines of evidence is *AKT1*, which is a target of lithium and associated with schizophrenia in several populations.²⁴⁻²⁶ The potential role of *BDNF* in schizophrenia has recently been reviewed²⁷ and it was reported to increase the genetic risk for schizophrenia in a Scottish population.²⁸ All these associations suggest the existence of common variants underlying schizophrenia at the respective gene loci. However, these associations also make these loci to excellent candidate genes in the search for rare mutations that could play a role in the etiology of schizophrenia. To measure the relative excess of rare, presumably deleterious, variants among schizophrenia patients, we use a summary statistic based on Tajima's *D*.²⁹

Materials and methods

Patient samples

All individuals included in the study were of German descent and ascertained at the Department of Psychiatry at the University of Bonn. Written informed consent was obtained from all patients and controls. All patients had been interviewed by experienced psychiatrists using the Structured Clinical Interview for DSM-IV Disorders. Lifetime 'best estimate' diagnoses according to DSM-IV criteria were based on multiple sources of information, including personal structured interview (SCID I), medical records, and family history method. Consensus diagnoses were performed by two psychiatrists, and whenever necessary, more psychiatrists were included in the decision process. The control group was matched by age, sex and ethnic origin. All samples were checked prior to the sequencing experiments for good DNA quality.

DNA sequencing and sequence annotation

Double-stranded DNA sequencing using the chain-termination method³⁰ was performed commercially (AGOWA, Berlin, Germany). Noncoding regions were targeted for sequencing, if annotated in the Ensembl database (version 17.33) as highly conserved between human and rodents (*Mus Musculus* or *Rattus Norvegicus*) and located within 10 kb upstream or downstream of the transcription start site. Conserved regions were extended into flanking regions in both directions that fragments of the size of about 500 bp were targeted for sequencing.

Evaluation of sequencing traces was performed with the software packages Phred and Phrap,³¹ Polyphred³² and Consed.³³ It has been shown that Phred's base calling is highly accurate.³¹ To minimize base calling errors, a Phred quality score of 30, that is 99.9% accuracy of the base call, was used for polymorphism detection. All SNPs detected by Polyphred were manually checked by two human experts to detect false positive SNPs. In addition, a SNP had to be observed in both forward and reverse reads to be considered as a true polymorphism.

Consensus sequences of sequenced fragments were aligned to the human genome using Megablast,³⁴ allowing for the inclusion of human genome annotations into the data analysis process. Data analysis was based on the Ensembl database (ftp://ftp.ensembl.org/pub/release-27/multi-species-27.1/data/mysql/ensembl_compara_27_1, ftp://ftp.ensembl.org/pub/release-28/human-28.35a/data/mysql/homo_sapiens_core_28_35a). Information on non-coding conservation was read from the database tables *genomic_align.txt.table* and *method_link.txt.table*. Highly conserved regions are those identified by the BLASTZ-NET TIGHT method. This definition of highly conserved regions is based on a score that considers an estimate of the neutral mutation rate at the particular locus and roughly corresponds to at least 80% sequence identity over at least 100 bps.

Nucleotide diversity estimates

In order to quantify sequence variability, we separately estimated for each sample the two nucleotide diversity parameters π and θ .³⁵ Average nucleotide heterozygosity π is the expected number of nucleotide differences per site between two randomly selected sequences from a population. This can be estimated from the differences d per nucleotide site over all sequence pairs i, j in a sample of n sequences:

$$\hat{\pi} = \sum_{i < j} \frac{d_{ij}}{[n(n-1)/2]}$$

The level of nucleotide polymorphism θ is proportion of polymorphic nucleotide sites that are expected to be observed in a population sample. This can be estimated from the proportion of nucleotide polymorphism S in a sample of n sequences:

$$\hat{\theta} = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}}$$

Furthermore, Tajima's D statistic²⁹ was calculated for the different sequence categories. Tajima's D statistic is the normalized difference between π and θ .

$$D = \frac{\pi - \theta}{\sqrt{\text{Var}(\pi - \theta)}}$$

A negative Tajima's D -value indicates an excess of rare alleles, whereas more positive values indicate an excess of intermediate frequency variants. Calculations were performed by scripts that were implemented by one of the authors (JF) at the University of Bonn as part of a larger software package that had been developed for the analysis of sequencing data in the context of an annotated human genome reference sequence.

Results

Overall analysis of nucleotide diversity

In total, we observed 108 SNPs within the resequenced fragments, 60% of them unknown to dbSNP (version 123). This fraction of unreported SNPs is quite large, when compared to the fraction of SNPs discovered in coding regions on independent datasets and taking into account the strong increase in public SNP database size.³ Potential explanations are the poorer database representation of noncoding SNPs or an excess of disease-related variants in the sequenced sample.

To compare the level of nucleotide diversity between the case and control sample and account for sample size, we next separately estimated for each sample the average heterozygosity π , level of polymorphism θ and Tajima's D statistic. Tajima's D is the normalized difference between π and θ , with smaller values denoting the relative excess of rare variants and higher values denoting the relative excess of common variants (see Materials and methods for

details). Average heterozygosity π displayed similar values in the schizophrenia cases and controls (cases: 7.99×10^{-4} controls: 7.95×10^{-4}). However, the level of polymorphism θ was higher among affected individuals (cases: 7.64×10^{-4} controls: 6.83×10^{-4}), pointing to a relative increase of polymorphic sites in this sample. The consequently smaller Tajima's D among affected individuals (cases: 0.16 controls: 0.58) further indicates more rare variants among cases than among controls.

To assess the significance of this difference in Tajima's D between controls and cases, we randomly permuted the affection status of sequenced individuals. In 2400 out of totally 100 000 random permutations, we found a stronger decrease of Tajima's D in the case sample than in the observed data ($P=0.024$, one-sided test for smaller Tajima's D among patients). Thus, the empirical significance analysis indicates that the observed difference of Tajima's D between controls and cases could go beyond random fluctuations. We feel that the one sided-test for an excess of rare variants in the patient sample is justified here, because one may assume that disease causing mutations are often deleterious and deleterious mutations are often rare in a population. The widespread existence of rare deleterious mutations in coding regions of the human genome had been shown by several earlier sequencing studies.³⁶⁻³⁸

We next repeated the estimation of nucleotide diversity parameters 1000 times with 25 randomly picked individuals from the schizophrenia sample. The average estimates for π , θ and Tajima's D (Table 1) were close to those estimated for the full sample of 37 schizophrenia patients. In none of the 1000 random schizophrenia subsamples we observed a larger Tajima's D -value than in the equally sized control sample. This further supports that the observed differences do not result from the different size of the case and the control sample or only a small subset of case individuals.

Nucleotide diversity of noncoding conserved regions

Following human genome annotations provided by the Ensembl database (version ensembl core 28.35a, ensembl compara 27.1),³⁹ we now categorized the resequenced noncoding regions into those highly conserved between humans and rodents and those not fulfilling this criterion. These conserved regions have experienced reduced evolutionary change over the past 80 million years, either due to chance, lower mutation rate or evolutionary constraint. We find that the 37 SNPs located in those regions more often were rare SNPs (minor allele frequency <5%) than SNPs not located in such regions ($\chi^2=13.8$, $df=2$, $P=0.001$) (Figure 1). The average minor allele of SNPs located in conserved regions was significantly lower both in the case ($P<0.001$ by two-sided Mann-Whitney U -test) and the control sample ($P=0.004$ by two-sided Mann-Whitney U -test), pointing to the influence of purifying

Table 1 Overall nucleotide diversity of noncoding regions

Sequence category	bps	Schizophrenia sample			Control sample			Joint sample		
		$\pi \times 10^{-4}$	$\theta \times 10^{-4}$	Tajima's <i>D</i>	$\pi \times 10^{-4}$	$\theta \times 10^{-4}$	Tajima's <i>D</i>	$\pi \times 10^{-4}$	$\theta \times 10^{-4}$	Tajima's <i>D</i>
noncds	27179	7.93	7.58	0.16	7.95	6.83	0.58	8.03	7.28	0.34
hcons	10777	4.02	6.12	-1.14	3.64	4.86	-0.82	3.89	6.09	-1.12
nhcons	16402	10.44	8.54	0.78	10.72	8.14	1.11	10.65	8.07	1.04

Estimates of nucleotide diversity parameters for different sequence categories in the case, control and joint sample. For each sequence category the number of base pairs resequenced, the estimated average nucleotide heterozygosity π , level of nucleotide polymorphism θ and Tajima's *D*-value are given. Human genome sequence annotations were transferred from the Ensembl database to the sequenced fragments. Noncoding genomic regions were categorized as *hcons* that is highly conserved between human and rodents (either mouse or rat) or as *nhcons*, which not fulfill this criterion.

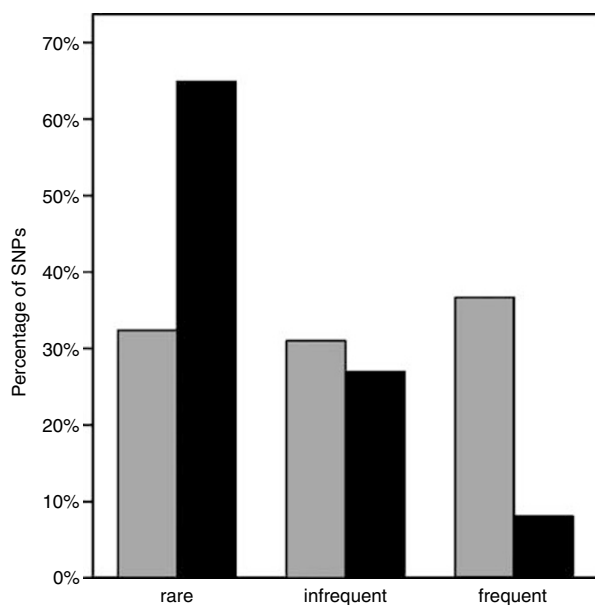


Figure 1 Percentage of SNPs from different frequency categories (minor allele frequency: rare <5%, infrequent <20%, common $\geq 20\%$) that were observed in noncoding conserved (black bars) and noncoding nonconserved (grey bars) regions.

selection on the frequency of alleles in conserved noncoding regions both among patients and among controls.

Within conserved regions, we observed a markedly decreased diversity, as measured by π and θ , both in the case and the control sample (Table 1). Interestingly, both among affected and unaffected individuals, Tajima's *D* was negative within conserved regions (cases: -1.22 controls: -0.82), whereas it was positive within nonconserved regions (cases: 0.86 controls: 1.11). The positive Tajima's *D* of the presumably neutrally evolving nonconserved regions could result from population demographic events. The negative Tajima's *D* of the conserved regions supports purifying selection as one cause of interspecies sequence conservation. Consistent with disease susceptibility as a reason of purifying selection, Tajima's *D*-value is smaller in the case sample both in conserved and nonconserved

noncoding regions. However, the case-control comparison did not reach nominal significance after the restriction to only one of these sequence categories, presumably due to reduced statistical power resulting from the smaller number of SNPs in each subcategory.

Nucleotide diversity of single gene loci

Diversity estimates of the individual gene loci were consistent with the results jointly observed over all loci (Table 2). Estimates of π and θ were above average at the loci *RGS4*, *NRG1*, and *AKT1*, whereas they were below average at *DRD3*, *DTNPB1* and *BDNF*. A higher diversity was measured both by π and θ in the case sample at each locus, except *DRD3*. As measured by Tajima's *D*, rare variants were more common among cases at each gene locus, except *DTNPB1*. This corresponds to a nonsignificant trend towards smaller Tajima's *D*-values among cases ($P=0.086$ one-sided Wilcoxon test of two related samples). To quantify the contribution of single gene loci to the overall difference between controls and cases, we separately repeated the permutation analysis for each gene locus. We measured the strength of the signal at each locus as the *Z*-score of the observed difference of Tajima's *D* between controls and cases on the distribution of this difference in 100 000 randomly permuted samples (Table 2). The difference in the randomly permuted samples was distributed approximately normal with a mean close to zero for all six loci. As the actually observed decrease of Tajima's *D* in the case sample did not belong to the most extreme 5% at any locus, nominally significant differences could not be achieved at any single gene locus.

An interesting question is now to find out, which loci contribute to the significant difference that is observed across all loci. We approached this task by modifying the previously developed SNP set association strategy⁴⁰ to search for the subset of gene loci, which produces the strongest difference between the two samples. This adapted strategy, which might be denoted 'locus set association', first ranks the gene loci with respect to the strength of the observed difference in Tajima's *D* between controls and cases, as measured by the above calculated *Z*-scores. Then, increasing numbers of loci from this ranked list are

Table 2 Noncoding nucleotide diversity at each investigated gene locus

Gene locus	Bp	SNPs	Schizophrenia sample			Control sample			Tajima's D	Z-score
			$\pi \times 10^{-4}$	$\theta \times 10^{-4}$	Tajima's D	SNPs	$\pi \times 10^{-4}$	$\theta \times 10^{-4}$		
AKT1	3026	16	10.51	11.34	-0.22	9	9.66	7.03	1.10	1.57
BDNF	6760	11	2.87	3.36	-0.40	7	2.66	2.38	0.32	1.09
DRD3	4669	19	5.44	8.35	-1.04	21	7.89	10.04	-0.69	0.81
DTNBP1	2409	6	4.67	5.11	-0.20	5	3.38	4.68	-0.68	-0.54
NRG1	7469	34	11.21	9.45	0.60	28	10.86	8.53	0.91	0.82
RGS4	2846	14	15.82	10.15	1.60	12	15.76	9.41	1.99	0.97

Estimates of nucleotide diversity parameters for each investigated gene locus among cases and controls. For each locus the number of base pairs resequenced and aligned to the Human Genome Assembly (NCBI35), the number of SNPs identified as well as the estimates of nucleotide heterozygosity π , the level of nucleotide polymorphism θ and Tajima's *D*-value are shown. Z-scores are based on the distribution of the difference in Tajima's *D* between controls and cases obtained by 100 000 random permutations of the affection status at each gene locus.

included into the case-control comparison, that is the two samples are compared by permutation analysis only at the top locus, the top two loci and so on up to the full number of all gene loci. This gives a list of *P*-values referring to increasing numbers of gene loci and measuring the strength of the case-control signal for the respective subsets of loci. The smallest *P*-value from this list, p_{min} , now indicates the particular subset of loci, which jointly produce the strongest difference between the two samples (Figure 2, Supplementary Table 1). However, due to assertive choice of this particular combination of gene loci, p_{min} should not be interpreted as an error probability. The list of loci jointly producing the strongest signal in our data contains the genes *AKT1*, *BDNF*, *RGS4* and potentially *NRG1*, whereas the inclusion of *DTNBP1* and *DRD3* seems to weaken the overall signal. Thus, rare alleles preferentially located at the former three loci might contribute to the genetic etiology of schizophrenia in our patient sample.

Discussion

We investigated nucleotide diversity of noncoding fragments around the transcription start site of six schizophrenia-associated gene loci. An excess of rare variants was observed in the sample of schizophrenia patients as compared to a control sample. This observation suggests that rare variants in the sequenced fragments could contribute to the etiology of schizophrenia. The location of the fragments in gene control regions is consistent with an important role for regulatory mutations in the genetics of schizophrenia. However, we are aware of the fact that the observed differences between cases and controls are only borderline significant. It would be, therefore, interesting to see, if the observed differences are part of a more general pattern. Thus, larger case-control sequencing studies at the six investigated gene loci as well as further candidate loci appear to be a worthwhile endeavor to understand the genetics of schizophrenia and related phenotypes.

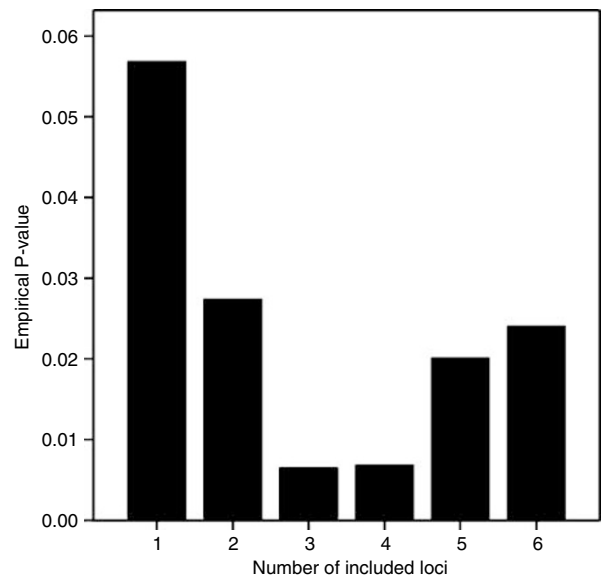


Figure 2 *P*-value estimates based on 100 000 random permutations of the affection status when considering an increasing number of gene loci. The six loci are initially ranked with respect to the strength of the difference in Tajima's *D* between controls and cases. Then increasing numbers of gene loci from the ranked list *AKT1*, *BDNF*, *RGS4*, *NRG1*, *DRD3* and *DTNBP1* are included into the analysis, denoted in the figure by the number of included loci.

The presented use of a summary statistic, such as the difference of Tajima's *D* between the control and case sample, might provide an important step forward in the analysis of case-control sequencing experiments. The significance of the observed difference in Tajima's *D* is tested by a permutation-based strategy, quantifying the strength of the difference in the allele frequency spectrum between the two samples. Although the allele frequency spectrum significantly differs between the samples across all gene loci, not all loci contribute equally to this signal. Consequently, a heuristic search efficiently identifies a subset of loci, which produces a stronger signal than any single locus or the full list of all gene loci.

The choice of investigated gene loci was based on earlier reports of common schizophrenia susceptibility haplotypes. Although some of these earlier reports of common susceptibility haplotypes may turn out to be false positive, rare variants at those loci still could play a role in the etiology of schizophrenia. This idea is supported by fact that all of the six genes are good functional candidates, based on the current understanding of the molecular biology of schizophrenia. Nevertheless, a weaker signal at a gene locus could also indicate a lower impact of deleterious sequence variation at this locus on the schizophrenia phenotype, here selected for patients with positive family history.

An important role in the origin of conserved noncoding regions has been proposed for purifying selection several years ago.⁴¹ More recent studies showed a significantly lower allele frequency for human SNPs from conserved noncoding regions,^{3,42} consistent with the ongoing action of purifying selection in the human lineage. In the present study, we again observe an excess of rare variants in genomic regions that are conserved between human and rodents, as compared to nonconserved regions. This excess of rare variants in conserved regions applies both to the case and the control sample, pointing to the occurrence of deleterious mutations also in the latter. On the other hand, the patient sample displays more rare variants both within conserved and within nonconserved regions. This indicates an increased number of deleterious mutations, which could be due to an increased number of schizophrenia susceptibility mutations.

In conclusion, our investigation of noncoding sequence variability at six schizophrenia-associated gene loci found rare variants more common among patients than among healthy controls. In addition, rare variants were found concentrated within regions that are conserved between human and rodents. These observations are consistent with the existence of deleterious noncoding mutations in the general population, which are enriched among affected individuals and within conserved regions. The overall excess of rare variants in the case sample primarily resulted from differences at the loci *AKT1*, *BDNF* and *RGS4*. The described strategy might provide a useful framework for further medical sequencing studies in the analysis of genetically complex traits.

Acknowledgements

We thank the three anonymous reviewers for their helpful and constructive comments on the manuscript. We thank the study participants for their contribution to this study. This study was funded by the German Human Genome Project (DHGP2) and the German National Genome Research Network (NGFN1).

References

- Owen MJ, Williams NM, O'Donovan MC: The molecular genetics of schizophrenia: new findings promise new insights. *Mol Psychiatry* 2004; 9: 14–27.
- Craddock N, O'Donovan MC, Owen MJ: The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *J Med Genet* 2005; 42: 193–204.
- Freudenberg-Hua Y, Freudenberg J, Winantea J *et al*: Systematic investigation of genetic variability in 111 human genes-implications for studying variable drug response. *Pharmacogenom J* 2005; 5: 183–192.
- Dermitzakis ET, Reymond A, Antonarakis SE: Conserved non-genic sequences – an unexpected feature of mammalian genomes. *Nat Rev Genet* 2005; 6: 151–157.
- Stefansson H, Sigurdsson E, Steinthorsdottir V *et al*: Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet* 2002; 71: 877–892.
- Williams NM, Preece A, Spurlock G *et al*: Support for genetic variation in neuregulin 1 and susceptibility to schizophrenia. *Mol Psychiatry* 2003; 8: 485–487.
- Bakker SC, Hoogendoorn ML, Selten JP *et al*: Neuregulin 1: genetic support for schizophrenia subtypes. *Mol Psychiatry* 2004; 9: 1061–1063.
- Corvin AP, Morris DW, McGhee K *et al*: Confirmation and refinement of an 'at-risk' haplotype for schizophrenia suggests the EST cluster, Hs.97362, as a potential susceptibility gene at the Neuregulin-1 locus. *Mol Psychiatry* 2004; 9: 208–213.
- Petryshen TL, Middleton FA, Kirby A *et al*: Support for involvement of neuregulin 1 in schizophrenia pathophysiology. *Mol Psychiatry* 2005; 10: 366–374, 328.
- Straub RE, Jiang Y, MacLean CJ *et al*: Genetic variation in the 6p22.3 gene *DTNBP1*, the human ortholog of the mouse *dysbindin* gene, is associated with schizophrenia. *Am J Hum Genet* 2002; 71: 337–348.
- Schwab SG, Knapp M, Mondabon S *et al*: Support for association of schizophrenia with genetic variation in the 6p22.3 gene, *dysbindin*, in sib-pair families with linkage and in an additional sample of triad families. *Am J Hum Genet* 2003; 72: 185–190.
- Van Den Bogaert A, Schumacher J, Schulze TG *et al*: The *DTNBP1* (*dysbindin*) gene contributes to schizophrenia, depending on family history of the disease. *Am J Hum Genet* 2003; 73: 1438–1443.
- Bray NJ, Preece A, Williams NM *et al*: Haplotypes at the dystrobrevin binding protein 1 (*DTNBP1*) gene locus mediate risk for schizophrenia through reduced *DTNBP1* expression. *Hum Mol Genet* 2005; 14: 1947–1954.
- Funke B, Finn CT, Plocik AM *et al*: Association of the *DTNBP1* locus with schizophrenia in a US population. *Am J Hum Genet* 2004; 75: 891–898.
- Chowdari KV, Mirnics K, Semwal P *et al*: Association and linkage analyses of *RGS4* polymorphisms in schizophrenia. *Hum Mol Genet* 2002; 11: 1373–1380.
- Williams NM, Preece A, Spurlock G *et al*: Support for *RGS4* as a susceptibility gene for schizophrenia. *Biol Psychiatry* 2004; 55: 192–195.
- Chen X, Dunham C, Kendler S *et al*: Regulator of G-protein signaling 4 (*RGS4*) gene is associated with schizophrenia in Irish high density families. *Am J Med Genet B Neuropsychiatr Genet* 2004; 129: 23–26.
- Morris DW, Rodgers A, McGhee KA *et al*: Confirming *RGS4* as a susceptibility gene for schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* 2004; 125: 50–53.
- Numakawa T, Yagasaki Y, Ishimoto T *et al*: Evidence of novel neuronal functions of *dysbindin*, a susceptibility gene for schizophrenia. *Hum Mol Genet* 2004; 13: 2699–2708.
- Talbot K, Eidem WL, Tinsley CL *et al*: *Dysbindin-1* is reduced in intrinsic, glutamatergic terminals of the hippocampal formation in schizophrenia. *J Clin Invest* 2004; 113: 1353–1363.
- De Blasi A, Conn PJ, Pin J, Nicoletti F: Molecular determinants of metabotropic glutamate receptor signaling. *Trends Pharmacol Sci* 2001; 22: 114–120.
- Taymans JM, Kia HK, Claes R, Cruz C, Leysen J, Langlois X: Dopamine receptor-mediated regulation of *RGS2* and *RGS4*

- mRNA differentially depends on ascending dopamine projections and time. *Eur J Neurosci* 2004; **19**: 2249–2260.
- 23 Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN: Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003; **33**: 177–182.
- 24 Emamian ES, Hall D, Birnbaum MJ, Karayiorgou M, Gogos JA: Convergent evidence for impaired AKT1-GSK3beta signaling in schizophrenia. *Nat Genet* 2004; **36**: 131–137.
- 25 Ikeda M, Iwata N, Suzuki T *et al*: Association of AKT1 with schizophrenia confirmed in a Japanese population. *Biol Psychiatry* 2004; **56**: 698–700.
- 26 Schwab SG, Hoefgen B, Hanses C *et al*: Further evidence for association of variants in the AKT1 gene with schizophrenia in a sample of European sib-pair families. *Biol Psychiatry* 2005; **58**: 446–450.
- 27 Angelucci F, Brene S, Mathe AA: BDNF in schizophrenia, depression and corresponding animal models. *Mol Psychiatry* 2005; **10**: 345–352.
- 28 Neves-Pereira M, Cheung JK, Pasdar A *et al*: BDNF gene is a risk factor for schizophrenia in a Scottish population. *Mol Psychiatry* 2005; **10**: 208–212.
- 29 Tajima F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; **123**: 585–595.
- 30 Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977; **74**: 5463–5467.
- 31 Ewing B, Hillier L, Wendl MC, Green P: Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998; **8**: 175–185.
- 32 Nickerson DA, Tobe VO, Taylor SL: PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 1997; **25**: 2745–2751.
- 33 Gordon D, Abajian C, Green P: Consed: a graphical tool for sequence finishing. *Genome Res* 1998; **8**: 195–202.
- 34 Zhang Z, Schwartz S, Wagner L, Miller W: A greedy algorithm for aligning DNA sequences. *J Comput Biol* 2000; **7**: 203–214.
- 35 Hartl DL, Clark AG: *Principles of population genetics*. Sunderland, Massachusetts: Sinauer Associates, Inc, 1997.
- 36 Leabman MK, Huang CC, DeYoung J *et al*: Natural variation in human membrane transporter genes reveals evolutionary and functional constraints. *Proc Natl Acad Sci USA* 2003; **100**: 5896–5901.
- 37 Freudenberg-Hua Y, Freudenberg J, Kluck N, Cichon S, Propping P, Nothen MM: Single nucleotide variation analysis in 65 candidate genes for CNS disorders in a representative sample of the European population. *Genome Res* 2003; **13**: 2271–2276.
- 38 Hughes AL, Packer B, Welch R, Bergen AW, Chanock SJ, Yeager M: Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proc Natl Acad Sci USA* 2003; **100**: 15754–15757.
- 39 Birney E, Andrews D, Bevan P *et al*: Ensembl 2004. *Nucleic Acids Res* 2004; **32**: D468–D470.
- 40 Hoh J, Wille A, Ott J: Trimming, weighting, and grouping SNPs in human case-control association studies. *Genome Res* 2001; **11**: 2115–2119.
- 41 Waterston RH, Lindblad-Toh K, Birney E *et al*: Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002; **420**: 520–562.
- 42 Drake JA, Bird C, Nemes J *et al*: Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nat Genet* 2006; **38**: 223–227.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)