Population differences in the International Multi-Centre ADHD Gene Project

Neale, B.; Sham, P.; Purcell, S.; Banaschewski, T.; Buitelaar, J.; Franke, B.; Sonuga-Barke, E.J.S.; Ebstein, R.; Eisenberg, J.; Mulligan, A.; Gill, M.; Manor, I.; Miranda, A.; Mulas, F.; Oades, R.D.; Roeyers, H.; Rothenberger, A.; Sergeant, J.A.; Steinhausen, H.C.; Taylor, E.

published in
Genetic Epidemiology
2008

DOI (link to publisher)
10.1002/gepi.20265

document version
Publisher's PDF, also known as Version of record

Genetic Epidemiology 32: 98–107 (2008)

Population Differences in the International Multi-Centre ADHD Gene Project

Benjamin M. Neale[1]* Pak C. Sham,[16,17] Shaun Purcell,[18,19] Tobias Banaschewski,[4] Jan Buitelaar[5] Barbara Franke,[5,6] Edmund Sonuga-Barke,[8] Richard Ebstein,[10] Jacques Eisenberg,[11] Aisling Mulligan,[12] Michael Gill,[12] Iris Manor,[10] Ana Miranda,[7] Fernando Mulas,[7] Robert D. Oades,[3] Herbert Roeyers[13] Aribert Rothenberger,[4] Joseph Sergeant,[9] Hans-Christoph Steinhausen,[14] Eric Taylor,[3] Margaret Thompson,[8] Kaixin Zhou,[1] Philip Asherson[1] and Stephen V. Faraone[2,15]

[1]*MRC Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, United Kingdom*
[2]*Department of Psychiatry, SUNY Upstate Medical University, Syracuse, New York*
[3]*University Clinic for Child and Adolescent Psychiatry, Essen, Germany*
[4]*Child and Adolescent Psychiatry, University of Göttingen, Germany*
[5]*Department of Psychiatry, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands*
[6]*Department of Human Genetics, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands*
[7]*Department of Developmental and Educational Psychology, University of Valencia, Spain*
[8]*Department of Psychology, King's College, University of Southampton, Highfield, Southampton, United Kingdom*
[9]*Department of Neuropsychology, Vrije Universiteit, Amsterdam, The Netherlands*
[10]*ADHD Clinic, Geha Mental Health Center, Petak Tikvah, Israel*
[11]*S. Herzog Memorial Hospital, Jerusalem, Israel*
[12]*Department of Psychiatry, Trinity Centre for Health Sciences, St. James's Hospital, Dublin, Ireland*
[13]*Department of Experimental and Clinical Psychology, Ghent University, Belgium*
[14]*Department of Child and Adolescent Psychiatry, University of Zurich, Switzerland*
[15]*Department of Neuroscience and Physiology, SUNY Upstate Medical University*
[16]*Department of Psychiatry, Hong Kong University, China*
[17]*Genome Research Centre, Hong Kong University, China*
[18]*Center for Human Genetic Research, Massachusetts General Hospital*
[19]*Program in Medical and Population Genetics, Broad Institute of Harvard and MIT*

The International Multi-Centre ADHD Gene sample consists of 674 families from eight countries (Belgium, England, Germany, Holland, Ireland, Israel, Spain, and Switzerland) ascertained from clinics for combined-type attention definity hyperactivity disorder in an offspring. 863 SNPs were successfully genotyped across 47 autosomal genes implicated in psychiatric disorders yielding a single nucleotide polymorphism (SNP) density of approximately one SNP per 2.5 kb. A global test of heterogeneity showed 269 SNPs nominally significant (expected 43). Inclusion of the Israeli population accounted for approximately 70% of these nominally significant tests. Hardy-Weinberg equilibrium tests suggest that combining all these populations would induce stratification, but that the Northern European populations (Belgium, England, Germany, Holland, and Ireland) could be appropriate. Tag SNPs were generated using pair-wise and aggressive tagging from Carlson et al. [2004] and de Bakker et al. [2005], respectively, in each population and applied to the other populations. Cross-population performance across Northern Europe was consistent with within population comparisons. Smaller sample size for each population tended to yield more problems for the generation of aggressive tags and the application of pair-wise tags. Any case-control sample employing an Israeli sample with Northern Europeans must consider stratification. A Northern European tag set, however, appears to be appropriate for capturing the variation across populations. *Genet. Epidemiol.* 32:98–107, 2008. © 2007 Wiley-Liss, Inc.

Key words: linkage disequilibrium; tagging; HapMap

Contract grant sponsor: NIH; Contract grant number: R01 MH62873.
*Correspondence to: Benjamin M. Neale, MRC Social Genetic Developmental and Psychiatry Centre, Institute of Psychiatry, London, United Kingdom. E-mail: b.neale@iop.kcl.ac.uk
Received 1 August 2007; Accepted 1 August 2007
Published online 14 September 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/gepi.20265

INTRODUCTION

Population differences have generated a great deal of confusion in case-control association studies. In particular, population stratification can yield both positive and negative bias in association findings. As a result, family-based designs such as the Transmission Disequilibrium Test were developed

[Spielman et al., 1993]. For case-control analyses, geneticists have created methods based on background markers such as genomic control, structure, and L-POP [Devlin and Roeder, 1999; Pritchard et al., 2000; Devlin et al., 2001; Purcell and Sham, 2004]. At the same time, association analysis has become more sophisticated with researchers now routinely employing haplotype-based methods, tagging single nucleotide polymorphisms (SNPs), and general patterns of linkage disequilibrium (LD) [Daly and Altshuler, 2005; Kruglyak, 1999; Palmer and Cardon, 2005; Risch, 2000]. These methods aim to improve efficiency by removing redundant genotyping and gather additional association information indirectly via LD, although the latter has also raised questions about the nature of population differences such as the consistency of LD structures across populations and the similarity of observed haplotypes [van den Oord and Neale, 2003; Evans and Cardon, 2005; Gonzalez-Neira et al., 2006].

A number of studies have examined the consistency of LD information in European samples. A key component of such work is tag selection from the LD information of the Centre d'Etude du Polymorphisme Humain (CEPH) sample. Mueller et al. densely genotyped four gene regions (*PLAU, SNCA, LMNA,* and *FKBP5*) at a frequency of one SNP per ~2–4 kb in strictly European populations to gauge populations differences [Mueller et al., 2005]. In general, mild allele frequency differences were detected. Aside from basic allele frequency differences, Mueller also looked at haplotype frequency and block boundary position as outcome measures, broadly noting that geographical distance correlated quite strongly with observed differences between the two populations. From a tagging perspective, CEPH samples captured variation accurately in two genes, and performed poorly in the other two, indicating variability in the patterns of tagging. de Bakker and colleagues recently examined the extent to which population differences exist in a comparison of various populations. For cross-European tagging, the tags from the CEPH sample were assessed in a Finish sample from the multiethnic cohort study. Generally, the power to detect association from tag SNPs is maintained in the Finnish population [de Bakker et al., 2006]. Nejentsev et al. examined the vitamin D receptor gene in four European populations and an African population. All of the European populations showed dramatically similar LD properties in the gene, and that tags generated from the CEPH sample performed reasonably well. The authors conclude that the similarity of LD is consistent with the model predicting no major founder effect for the Europeans [Nejentsev et al., 2004]. Recently, Ribas et al. published an examination of population differences between a Spanish sample and the CEPH sample. The correlations for allele frequency and LD were 0.91 and 0.95, respectively. Baucet and colleagues explore the population structure of Europeans as compared to the CEPH and Coriell samples. Broadly, Europe splits into two main sets, a northern (Polish, Irish, English, Germans, and some Italians) and a southeastern (Greeks, Armenians, Jews, and some Italians) [Bauchet et al., 2007]. Finally, Evans and Cardon compared the recombination rates across a region in chromosome 20 in the CEPH and a UK sample. The results were quite similar for the two populations, so the close ancestry appears to predict similarity in LD structure [Evans and Cardon, 2005].

In this paper, we present analysis of the similarities and differences of populations across Europe and Israel from the International Multi-Centre ADHD Gene (IMAGE) sample, exploring 47 autosomal genes, many of which have been implicated in psychiatric disorders.

METHODS

The IMAGE sample used in this study has been detailed elsewhere [Brookes et al., 2006]. In brief, the sample is composed of 674 families collected in eight countries from 11 clinical centers: Belgium, Germany, Holland, Ireland, Israel, Spain, Switzerland, and United Kingdom. We considered our sample by looking at each ascertainment center separately. As Germany, Holland, and Israel each had two centers, for some analyses we combined these sets of families, yielding eight sets of data (with number of families in brackets): Swiss (24), German (29+80 = 109), Spanish (52), Israeli (26+97 = 123), Dutch (84+110 = 194), Belgian (22), England (98), and Ireland (52).

Ethical approval for the study was obtained from Ethical Review Boards within each country and informed consent obtained for the use of the samples for analyses relating to the genetic investigation of attention definit hyperactivity disorder (ADHD). All ADHD probands and their siblings were aged 5–17 years at the time of entry into the study and access was required to one or both biological parents for DNA collection. Diagnosis of ADHD was standardized at the outset of the sample collection, so all probands are based on the same phenotypic definition. Exclusion criteria applying to both probands and siblings included autism, epilepsy, IQ < 70, brain disorders, and any genetic or medical disorder associated with externalizing behaviors that might mimic ADHD. DNA was available for both parents in 598 families (88.7%) and from one parent in 73 families (10.8%). 664 (85.6%) of the ADHD cases were males. The age range for both probands and

siblings was 5–17 years with a mean age of 11.2 years (SD = 2.7) for probands and 11.2 years (SD = 3.1) for siblings.

Gene selection focused on ''biological systems'' by nominating 45 genes that were likely to exert an effect through regulation of dopamine, serotonin, and norepinephrine neurotransmission. Six additional genes were selected for their role in circadian rhythm. Of these, four genes were sex-linked and excluded from this analysis.

For SNP selection, we aimed at a comprehensive analysis of each gene from two perspectives: functionality and tagging. For the functional markers, we targeted SNPs located within coding regions (synonymous and non-synonymous), 5′ and 3′ untranslated regions, intron sequences within 300 bp from intron/exon boundaries, and one SNP per kb covering 5 kb upstream from the start of transcription or known 5′ regulatory regions including the promotor. For tagging we selected a non-redundant set of tagging SNPs that showed $r^2 > 0.80$ with SNPs with minor allele frequency (MAF) $\geq$ 0.05, using the CEPH panel from the HapMap [2003] database. Tagging SNPs were selected using two methods. We used the ''CompleteLink routine'' within CLUSTAG that implements a hierarchical clustering algorithm (hkumath.hku.hk/web/link/CLUSTAG/CLUSTAG.html) [Ao et al., 2005]. The second method used the default algorithm in Haploview (www.broad.mit.edu/mpg/haploview) from Gabriel et al. [2002]. 39.6% (173 out of 437 tSNPs) of the CLUSTAG SNPs and 71.8% (173 out of 241 tSNPs) of the Gabriel method SNPs were shared. 7.0% (77 out of 1,105, functional SNPs) were included in the tag criteria. To avoid redundancy of the marker information, where the two methods recommended selection of two different SNPs that fell within the same cluster defined by CLUSTAG, we preferentially selected the tagging SNP recommended by the CLUSTAG algorithm. If the constraints of the Illumina technology prevented the use of an identified tagging SNPs nominated by CLUSTAG, then an alternative SNP located in the same cluster was selected. At the time of SNP selection, it was unclear what the most effective approach for tagging a region is. As a result, we aimed to err on the side of SNP inclusion to ensure that the variation in each gene was well characterized.

Blood samples were sent to Rutgers University Cell and DNA repository, New Jersey (RUCDR). These were either used to generate lymphocyte cell lines from which DNA was extracted, or DNA was extracted directly from a portion of the blood sample and lymphocytes cryopreserved for future recovery. In a few cases where individuals were not able to supply a blood sample, we used a mouth swab sampling technique and extracted the DNA at the SGDP laboratory in London [Freeman et al., 2003]. DNA stocks for the entire dataset were collated in London where they were stored, organized and plated out for further analysis. Geneservice Ltd Cambridge (UK) performed whole-genome amplification on all samples with $<100\,\mu g$ stock DNA, using the REPLI-g kit (Qiagen Ltd, Crawley, Sussex). DNA samples were arrayed into 96-well plates at a concentration of $50\,ng/\mu L$ and delivered to Illumina Inc. (San Diego, CA).

SNP genotyping was completed using the Illumina high-throughput BeadArray™ technology (http://www.illumina.com). We set a lower limit for MAF of 0.05, but also included a high proportion of non-validated SNPs and SNPs with unknown heterozygosity from ''functional'' regions. Additionally, we excluded variation on the sex chromosomes. For this analysis we finished with a set of 863 SNPs across 47 genes. The average SNP density was approximately one SNP every 2.5 kb.

Ten samples of the total of 2,937 samples ($<0.04\%$) could not be genotyped even after multiple attempts and were removed from the dataset. For the remaining samples the average genotype drop out rate was 0.02% (not including the 34 markers that failed on the whole-genome amplification DNA). The overall genotyping error was estimated to be lower than 0.065% (0.06% + 0.005%). We identified 47 families with potential pedigree errors using PEDCHECK [O'Connell and Weeks, 1998]. Within these families the numbers of Mendelian errors ranged from 16 to 312 and could not be explained by genotyping errors. Forty errors could be corrected by considering different familial relationships but seven of these families exhibited unknown pedigree errors and were removed from the analysis. The final dataset for association analysis included 674 families that contained 156 sporadic Mendelian errors from 987 autosomal markers. The overall detection rate of Mendelian errors is therefore 0.02%, which is consistent with the estimated efficiency of SNP markers [13–75%; Douglas et al., 2002] to detect such errors by identification of Mendelian errors and an estimated overall genotype error rate $<0.065\%$. Pedigree errors may be explained by the complicated procedures inherent within a large international cooperative process. We also imposed a threshold of the Illumina quality control score of 0.5 for all SNPs.

We employed three approaches to determine the level of similarity across the subpopulations; heterogeneity tests, Hardy-Weinberg equilibrium (HWE) tests, and tag SNPs comparisons (the number of tags and the cross-population efficiency). Selecting only markers that were polymorphic in all populations, a χ^2 test was conducted on the allele counts by center site and population group. The χ^2 was conducted

using the program R 2.1.0, and the significance was simulated using 2,000 replicates for each test. Each center site or population group was subsequently dropped from the heterogeneity analysis, to determine the effect on the distribution of results. In addition to global heterogeneity analysis, pair-wise analysis was conducted for every possible pair of populations.

The goal of the heterogeneity analysis is to assess allele frequency differences across the populations. The global test of heterogeneity determines whether the entire set of populations can be grouped together, and by extension whether combined association analysis is appropriate. The pair-wise tests aim to illuminate the relative similarity between populations. In particular, the question of which populations can be grouped together with little cost to association analysis. These pair-wise tests are not attempting to find significant differences, but rather explore how similar these various populations are. Furthermore, r^2 is affected by allele frequency, so allele frequency differences are an important factor for consideration in consistency of LD information. As a convenient way to interpret the results broadly, we report the number of SNPs with global heterogeneity test *P*-value <0.05. The expectation for the number of tests significant at this level is 5%, so for 863 SNPs, we expect to observe approximately 43 differences significant at the 0.05 level.

HWE tests were conducted in each subpopulation and combined within countries. PLINK was used to calculate the HWE test, using the χ^2 test [Purcell et al., 2007]. Consistent deviations from Hardy-Weinberg are a hallmark of population stratification.

Using the Tagger program implemented in Haploview, we determined the number of tag SNPs to capture all information across the region at an r^2 of 0.8 or better [Carlson et al., 2004; Patil et al., 2001]. We employed both ''pair-wise tagging'' from Carlson et al. and the ''aggressive tagging'' from de Bakker et al. considering all variation [Carlson et al., 2004; de Bakker et al., 2005]. We examined the average number of tags necessary across all populations. We elected to use the number of tag SNPs rather than specific tag SNPs as mild variation in the observed r^2 can cause different SNP selection. Additionally, the number of SNPs rather than specific SNPs ought to serve as an indication of the amount of LD across each region. The efficacy of the tags selected is measured by the cross-population tagging performance. Each population was used as a training set to generate tags and then those tags were tested in every other population. Both tagging approaches outlined above, pair-wise and aggressive, were employed.

RESULTS

Comparing all population centers jointly yielded 269 out of 863 SNPs with a *P*-value <0.05. A kernel density estimator plot of the distribution of *P*-values can be found in Figure 1. Clearly, the overall distribution of tests is biased heavily toward smaller *P*-values. Such a result is an indication of major population differences in allele frequencies. Allele frequency differences are a major component of population stratification, and important in the consideration of association evidence.

Table I shows the number and percentage of tests significant at a nominal 0.05 *P*-value level for all pair-wise population comparisons and for the global test conditional on dropping each center. In

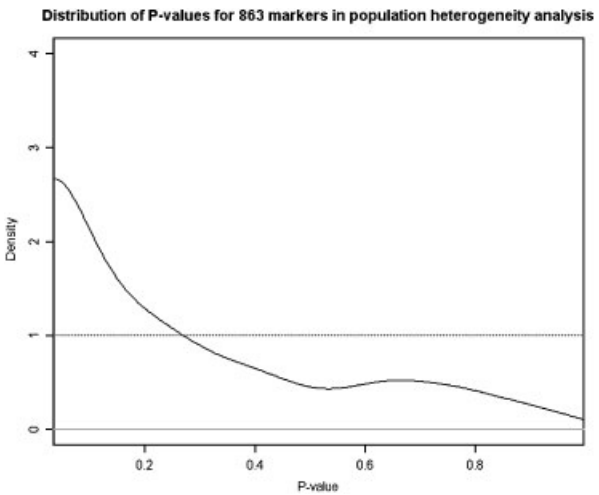

Fig. 1. Distribution of *P*-values for 863 markers in population heterogeneity analysis. The solid line represents the distribution of *P*-values from the full heterogeneity analysis. The dotted line represents the expected uniform distribution of *P*-values under the null model of no allele frequency differences.

TABLE I. The number and percentage of markers nominally significant for a simple χ^2 heterogeneity test

	BE	CH	DE1	DE2	EN	ES	IL1	IL2	IR	NL1	NL2
BE	290										
	(0.336)										
CH	39	288									
	(0.045)	(0.334)									
DE1	21	28	285	**295**							
	(0.024)	(0.032)	(0.33)	**(0.342)**							
DE2	23	51	37	283							
	(0.027)	(0.059)	(0.043)	(0.328)							
EN	30	69	30	31	273						
	(0.035)	(0.08)	(0.035)	(0.036)	(0.316)						
ES	37	42	56	69	63	278					
	(0.043)	(0.049)	(0.065)	(0.08)	(0.073)	(0.322)					
IL1	99	89	91	119	137	106	253	**81**			
	(0.115)	(0.103)	(0.105)	(0.138)	(0.159)	(0.123)	(0.293)	**(0.094)**			
IL2	138	114	119	253	257	166	26	142			
	(0.16)	(0.132)	(0.138)	(0.293)	(0.298)	(0.192)	(0.03)	(0.165)			
IR	47	57	47	78	67	71	155	266	259		
	(0.054)	(0.066)	(0.054)	(0.09)	(0.078)	(0.082)	(0.18)	(0.308)	(0.3)		
NL1	61	54	61	46	66	96	150	279	87	269	**243**
	(0.071)	(0.063)	(0.071)	(0.053)	(0.076)	(0.111)	(0.174)	(0.323)	(0.101)	(0.312)	**(0.282)**
NL2	46	59	43	51	53	105	169	309	88	35	256
	(0.053)	(0.068)	(0.050)	(0.059)	(0.061)	(0.122)	(0.196)	(0.358)	(0.102)	(0.041)	(0.297)

BE: Ghent, Belgium; CH: Zurich, Switzerland; DE1: Essen, Germany; DE2: Goettingen, Germany; EN: London, England; ES: Valencia, Spain; IL1: Jerusalem, Israel; IL2: Tel Aviv, Israel; IR: Dublin, Ireland; NL1: Amsterdam, the Netherlands; NL2: Nijmegen, the Netherlands.The cells at the cross of the same population, refer to the number of tests that are nominally significant when excluding that population from the analysis (e.g. the 1st cell shows the exclusion the BE population from the heterogeneity analysis, with 290 tests nominally significant). The numbers in bold refer to dropping both population centers (DE, IL, and NL) from a country (e.g. removing both DE populations leaves 295 nominally significant tests). The rest of the table contains the number of nominally significant tests from pair-wise population comparisons.The number of tests nominally significant considering all populations is 269.

instances of multiple centers in a country, these centers were combined and dropped.

Figure 2 shows the distribution of HWE tests for each population, and country. Additionally, global and Northern European (Belgium, England, Holland, Germany, Ireland) sets were tested. Extensive data cleaning has already taken place (see above), so some evidence for differences may have already been removed.

Figure 3 shows the HWE tests for the Northern European set. These two figures show the radically different amount of deviation from HWE observed in the combined sample, versus the Northern European sample. Deviation from HWE is a hallmark of population stratification, and so these results further confirm the relative difference of the Israeli populations from the Northern Europeans.

Table II shows the number of SNPs necessary to characterize the entire set of markers using both the pair-wise and aggressive tagging approaches. For both of these approaches, a minimum R^2 threshold of 0.8 for all markers in the region was set. The correlations between the number of tag SNPs and sample size are 0.28 and -0.81 for pair-wise and aggressive tagging, respectively.

Tables III contains the cross-tagging performance in terms of percentage of SNPs captured by the pre-defined tagging set. Across the datasets, we have a range of sample sizes. In general, larger populations show more pair-wise tags and fewer aggressive tags. The correlation between tag number and sample size is considerably stronger in the aggressive tagging approach.

DISCUSSION AND CONCLUSION

As shown in Table I, the allelic heterogeneity analysis shows clear allele frequency differences between the Israeli populations and the rest of the European populations. Furthermore, allele frequency comparisons between Northern European populations show considerably less heterogeneity. The Northern European populations may still hold population differences at a finer level which are not picked up, given the sample size and subset of all genome variation genotyped in our analysis. As indicated by Bauchet et al. Southeastern and Jewish populations ought not be grouped together with Northern European populations as there are significant differences [Bauchet et al., 2007].

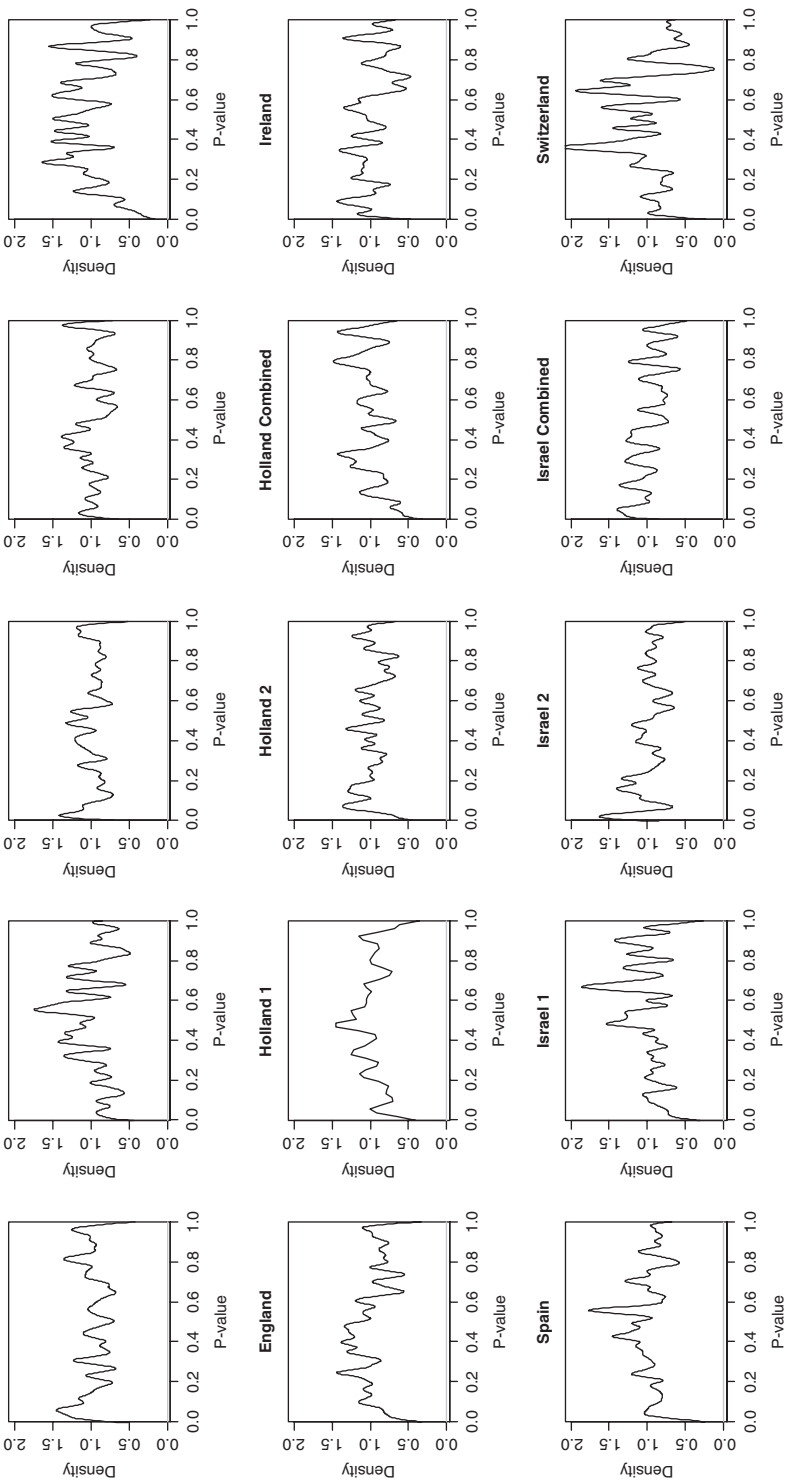

Fig. 2. Centre and country Hardy-Weinberg test distributions. Each centre's distribution of Hardy-Weinberg tests. For Germany, Holland, and Israel the combined plot refers to pooling the samples and recalculating all the Hardy-Weinberg tests.

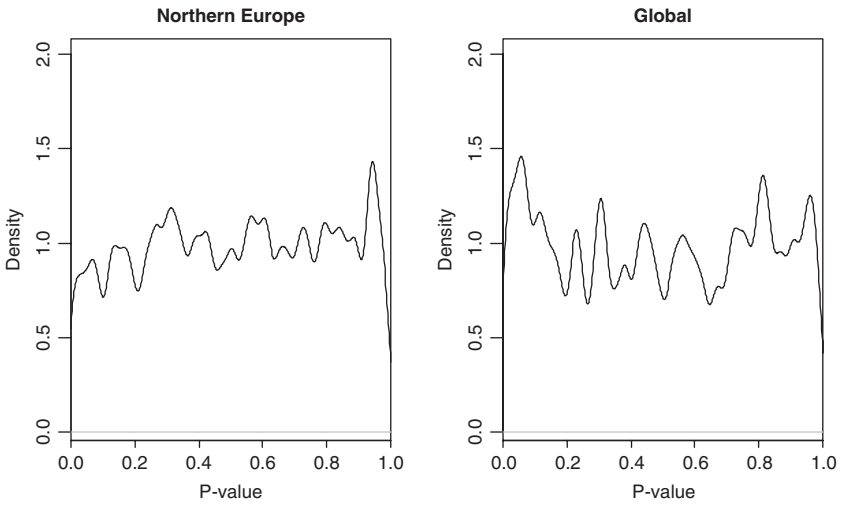

Fig. 3. Distribution of Hardy-Weinberg tests in the Northern European and all populations. The distribution of Hardy-Weinberg tests across the 863 markers for Northern Europe and all populations combined. Northern Europe is composed of Belgium, Germany 1 and 2, Holland 1 and 2, England, and Ireland.

TABLE II. Number of tag SNPs necessary to characterize each population

Population	No. of families	No. of pair-wise tags	No. of aggressive tags
BE	22	534	485
CH	24	530	488
DE1	29	534	483
DE2	80	540	461
EN	98	537	439
ES	52	544	469
IL1	26	543	494
IL2	97	555	467
IR	52	527	441
NL1	84	526	449
NL2	110	538	435
Total	674	542	*

BE: Ghent, Belgium; CH: Zurich, Switzerland; DE1: Essen, Germany; DE2: Goettingen, Germany; EN: London, England; ES: Valencia, Spain; IL1: Jerusalem, Israel; IL2: Tel Aviv, Israel; IR: Dublin, Ireland; NL1: Amsterdam, the Netherlands; NL2: Nijmegen, the Netherlands.

Pair-wise tagging refers to the Carlson et al., 2004 method. Aggressive tagging refers to the de Bakker et al., 2005 method. Both of these tags algorithms were implemented using Haploview and Tagger.

Mean number of tags across the centres is 537.0 for pair-wise tagging and 464.6 for aggressive tagging.

The correlation between the number of families and number of tags across centres is 0.28 for pair-wise tagging and −0.82 for aggressive tagging.

*Computer memory restrictions prevented a global analysis of aggressive tagging.

The HWE tests presented in Figure 2 show that each country is largely behaving in an appropriate fashion from a genotype distribution perspective. Additionally, the HWE tests suggest that the Northern Europeans are largely similar given the relative uniform nature of the *P*-value distribution.

For the cross-tagging performance, the three countries with two centers each can be considered a convenient benchmark for comparison. Accordingly, the Belgian, English, German, Dutch, Irish, and to a lesser extent Swiss appear similar for pair-wise tagging, with the Spanish, and Israelis more distinct. Other work in this vein indicates that the CEPH sample, which is of Northern European ancestry, provides a good tagging set [de Bakker et al., 2006; Gonzalez-Neira et al., 2006; Nejentsev et al., 2004]. However, the sample size appears to be the distinguishing factor in terms of performance when using aggressive tagging. Considering the two centers in each population, the smaller populations perform as better training sets for the larger populations for pair-wise tagging, than vice versa. One possible reason is that the error for the estimate of the allele frequency is greater in smaller populations. As a result, rare SNPs are more likely to be included and more common SNPs are more likely to be excluded from the smaller populations. Aggressive tagging suffers more in cross-population comparisons because of the increased importance of LD information.

Work on population differences using measures of similarity such as heterogeneity tests and the number of tagging SNPs may be colored by phenotypic considerations. Such measures could potentially differ because of ascertainment biases, allelic or locus heterogeneity. Even with that qualification, results in this paper may be confounded with association signals or imposed selection.

TABLE III. Cross-population tagging performance for pair-wise tagging

		Pair-wise tag generation										
		BE	DE1	DE2	NL1	NL2	IL1	IL2	ES	CH	IR	EN
Tag Application	BE	1.00	0.91	0.90	0.93	0.92	0.88	0.87	0.89	0.92	0.93	0.91
	DE1	0.94	1.00	0.93	0.94	0.93	0.92	0.91	0.92	0.95	0.95	0.93
	DE2	0.97	0.96	1.00	0.96	0.97	0.93	0.92	0.95	0.96	0.97	0.96
	NL1	0.96	0.94	0.93	1.00	0.94	0.93	0.91	0.93	0.95	0.96	0.94
	NL2	0.96	0.94	0.94	0.96	1.00	0.91	0.91	0.93	0.94	0.97	0.94
	IL1	0.93	0.93	0.91	0.93	0.92	1.00	0.92	0.91	0.94	0.93	0.93
	IL2	0.96	0.96	0.95	0.96	0.95	0.96	1.00	0.95	0.96	0.97	0.95
	ES	0.96	0.94	0.95	0.96	0.96	0.94	0.93	1.00	0.95	0.97	0.95
	CH	0.92	0.92	0.90	0.92	0.91	0.90	0.87	0.89	1.00	0.93	0.91
	IR	0.95	0.94	0.92	0.95	0.94	0.91	0.90	0.92	0.94	1.00	0.93
	EN	0.96	0.95	0.95	0.96	0.96	0.94	0.93	0.95	0.96	0.97	1.00
		Agressive tag generation										
	BE	1.00	0.93	0.93	0.94	0.94	0.90	0.91	0.92	0.94	0.94	0.93
Tag Application	DE1	0.93	1.00	0.93	0.94	0.94	0.90	0.91	0.92	0.94	0.95	0.94
	DE2	0.88	0.87	1.00	0.94	0.96	0.85	0.90	0.91	0.88	0.92	0.94
	NL1	0.89	0.88	0.92	1.00	0.94	0.85	0.90	0.90	0.89	0.93	0.92
	NL2	0.86	0.86	0.93	0.93	1.00	0.84	0.87	0.87	0.88	0.92	0.91
	IL1	0.91	0.92	0.91	0.93	0.93	1.00	0.92	0.92	0.94	0.93	0.93
	IL2	0.88	0.90	0.91	0.93	0.93	0.87	1.00	0.90	0.90	0.93	0.92
	ES	0.91	0.91	0.95	0.96	0.95	0.89	0.94	1.00	0.93	0.95	0.95
	CH	0.91	0.92	0.92	0.94	0.94	0.87	0.88	0.91	1.00	0.94	0.92
	IR	0.89	0.90	0.93	0.93	0.94	0.86	0.89	0.90	0.90	1.00	0.93
	EN	0.89	0.88	0.94	0.94	0.96	0.86	0.91	0.92	0.90	0.94	1.00

BE: Ghent, Belgium; CH: Zurich, Switzerland; DE1: Essen, Germany; DE2: Goettingen, Germany; EN: London, England; ES: Valencia, Spain; IL1: Jerusalem, Israel; IL2: Tel Aviv, Israel; IR: Dublin, Ireland; NL1: Amsterdam, the Netherlands; NL2: Nijmegen, the Netherlands.
Tag generation refers to the training set for the tags. Thus, each column refers to the same tag set generated by each centre. Each row refers to the efficacy of tagging in each population in terms of percentage of SNPs that are captured at or above an r^2 of 0.8 or better. For example, the performance of the two German populations (DE1 and DE2) in each other for pair-wise tagging is as follows: For the tag set defined in DE1 and applied in DE2 the percentage of SNPs captured at or above 0.8 is 96%, while the tag set defined in DE2 and applied in DE, the percentage of SNPs captured at or above 0.8 is 93%.

For example, Zaykin et al. [2006] have recently published a method to examine differences in LD structure as evidence for association.

The marker selection approach utilized for tag selection creates an obvious bias toward markers with a more common than average MAF. First, current SNP databases are enriched for more common markers, as they are easier to detect. Additionally, we used a tagging approach with a condition on the MAF of the SNPs necessary to be tagged. More common markers might show a different pattern of variation between populations than rare alleles. Also, private alleles for some subpopulations, such as the Spanish and the Israeli, might be underrepresented in HapMap samples, yielding to a bias away from observable population differences. Finally, the marker selection protocol for the tags at the outset of the project causes bias in the relative tagging efficiency in these regions.

We selected our families based on the presence of a clinical phenotype and looked at genes with a strong functional hypothesis in relation to the selected phenotype. While this may limit the generalizability of some of the conclusions to the general population, most of the samples in genetics are of clinical populations. So, while this is a limitation from a general population perspective, it also is advantageous from the perspective of seeing how much clinical populations vary on heterozygosity and tagging. The pooling of samples from different sites within the same country could potentially deflate the differences that we could observe.

In general, variation between populations tends to diminish as the populations become more homogenous. The German, Dutch, Belgian, Irish, and English cohorts show relatively little difference between each other, while the Spanish, Swiss, and Israeli samples tend to differ more greatly from the Northern Europeans. However, this difference is far smaller than that observed in surveys of more diverse populations (e.g. Africans vs. Caucasian Europeans) [Crawford et al., 2004]. Finally, while

variation changes between populations it appears that the amount of LD structure in each region is relatively similar.

Any case-control association study using these populations together ought to consider using genomic control or a population classification system in conjunction with a statistical technique for combining evidence across groups, such as the Mantel-Haenszel technique. Additionally, the Israeli sample requires more tags than the other populations, after correcting for sample size, indicating that the CEPH sample defined tag set used to define the marker selection did not capture all the variation in the Israeli population. In general, most investigations into LD differences find that Northern Europeans can be grouped together for the purposes of generating and applying tag SNPs [Mueller et al., 2005; Gonzalez-Neira et al., 2006; de Bakker et al., 2006]. Southern Europeans and Israelis tend to form a different population cluster [Bauchet et al., 2007]. Our results are largely consistent with these observations in the context of genes implicated is psychiatric disorders.

ACKNOWLEDGMENTS

The IMAGE project is a multi-site, international effort supported by NIH Grant R01MH62873 to SV Faraone. Site Principal Investigators are Philip Asherson, Tobias Banaschewski, Edmund Sonuga-Barke, Jan Buitelaar, Richard P Ebstein, Stephen V Faraone, Michael Gill, Ana Miranda, Robert D Oades, Herbert Roeyers, Aribert Rothenberger, Joseph Sergeant, and Hans-Christoph Steinhausen. Senior co-investigators are Margaret Thompson, Pak Sham, Peter McGuffin, Robert Plomin, Ian Craig and Eric Taylor. Chief Investigators at each site are Rafaela Marco, Nanda Rommelse, Fernando Mulas, Wai Chen, Henrik Uebel, Hanna Christiansen, U Mueller, Cathelijne Buschgens, Barbara Franke, and Lamprini Psychogiou. Other investigators are Marieke Altink, Ellen Fliers, Ruud Minderaa, and Alysa Doyle. We thank all the families who kindly participated in this research.

REFERENCES

Ao SI, Yip K, Ng M, Cheung D, Fong PY, Melhado I, Sham PC. 2005. CLUSTAG: hierarchical clustering and graph methods for selecting tag SNPs. Bioinformatics 21:1735–1736.

Bauchet M, McEvoy B, Pearson LN, Quillen EE, Sarkisian T, Hovhannesyan K, Deka R, Bradley DG, Shriver MD. 2007. Measuring European Population Stratification using Microarray Genotype Data. Am J Hum Genet 80.

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. 2004. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 74:106–120.

Crawford DC, Carlson CS, Rieder MJ, Carrington DP, Yi Q, Smith JD, Eberle MA, Kruglyak L, Nickerson DA. 2004. Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations. Am J Hum Genet 74:610–622.

Daly MJ, Altshuler D. 2005. Partners in crime. Nat Genet 37:337–338.

de Bakker PI, Burtt NP, Graham RR, Guiducci C, Yelensky R, Drake JA, Bersaglieri T, Penney KL, Butler J, Young S, et al. 2006. Transferability of tag SNPs in genetic association studies in multiple populations. Nat Genet 38:1298–1303.

de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. 2005. Efficiency and power in genetic association studies. Nat Genet 37:1217–1223.

Devlin B, Roeder K. 1999. Genomic control for association studies. Biometrics 55:997–1004.

Devlin B, Roeder K, Wasserman L. 2001. Genomic control, a new approach to genetic-based association studies. Theor Popul Biol 60:155–166.

Douglas JA, Skol AD, Boehnke M. 2002. Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. Am J Hum Genet 70:487–495.

Evans DM, Cardon LR. 2005. A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations. Am J Hum Genet 76:681–687.

Freeman B, Smith N, Curtis C, Huckett L, Mill J, Craig IW. 2003. DNA from buccal swabs recruited by mail: evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping. Behav Genet 33:67–72.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. 2002. The structure of haplotype blocks in the human genome. Science 296:2225–2229.

Gonzalez-Neira A, Ke X, Lao O, Calafell F, Navarro A, Comas D, Cann H, Bumpstead S, Ghori J, Hunt S, Deloukas P, Dunham I, Cardon LR, Bertranpetit J. 2006. The portability of tagSNPs across populations: a worldwide survey. Genome Res 16: 323–330.

HapMap. 2003. The International HapMap Project. Nature 426: 789–796.

Kruglyak L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22: 139–144.

Mueller JC, Lohmussaar E, Magi R, Remm M, Bettecken T, Lichtner P, Biskup S, Illig T, Pfeufer A, Luedemann J, Schrieber S, Pramstaller P, Pichler I, Romeo G, Gaddi A, Testa A, Wichmann HE, Metspalu A, Meitinger T. 2005. Linkage disequilibrium patterns and tagSNP transferability among European populations. Am J Hum Genet 76:387–398.

Nejentsev S, Godfrey L, Snook H, Rance H, Nutland S, Walker NM, Lam AC, Guja C, Ionescu-Tirgoviste C, Undlien DE, Ronningen KS, Tuomilehto-Wolf E, Tuomilehto J, Newport MJ, Clayton DG, Todd JA. 2004. Comparative high-resolution analysis of linkage disequilibrium and tag single nucleotide polymorphisms between populations in the vitamin D receptor gene. Hum Mol Genet 13:1633–1639.

O'Connell JR, Weeks DE. 1998. PedCheck: a program for identifying genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259–266.

Palmer LJ, Cardon LR. 2005. Shaking the tree: mapping complex disease genes with linkage disequilibrium. Lancet 366: 1223–1234.

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR. 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294:1719–1723.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, 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.

Purcell S, Sham P. 2004. Properties of structured association approaches to detecting population stratification. Hum Hered 58:93–107.

Risch NJ. 2000. Searching for genetic determinants in the new millennium. Nature 405:847–856.

Spielman RS, McGinnis RE, Ewens WJ. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506–516.

van den Oord EJ, Neale BM. 2003. Will haplotype maps be useful for finding genes? Mol Psychiatry 9:227–236.

Zaykin DV, Meng Z, Ehm MG. 2006. Contrasting linkage-disequilibrium patterns between cases and controls as a novel association-mapping method. Am J Hum Genet 78:737–746.