American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 147B:1070–1075 (2008)

Worldwide Genetic Variation in Dopamine and Serotonin Pathway Genes: Implications for Association Studies

Michelle Gardner,[1,2] Jaume Bertranpetit,[1,2] and David Comas[1,2]*

[1]*Unitat de Biologia Evolutiva, Universitat Pompeu Fabra, Doctor Aiguader, Barcelona, Spain*
[2]*CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain*

The dopamine and serotonin systems are two of the most important neurotransmitter pathways in the human nervous system and their roles in controlling behavior and mental status are well accepted. Genes from both systems have been widely implicated in psychiatric and behavioral disorders, with numerous reports of associations and almost equally as numerous reports of the failure to replicate a previous finding of association. We investigate a set of 21 dopamine and serotonin genes commonly tested for association with psychiatric disease in a set of 39 worldwide populations representing global genetic diversity to see whether the failure to replicate findings of association may be explained by population based differences in allele frequencies and linkage disequilibrium (LD) in this gene set. We present results demonstrating a surprising homogeneity of the allele frequencies across worldwide populations in these genes. LD both for populations within continent groupings and across continental regions also showed a remarkable similarity. These findings taken together suggest that ethnic differences in these parameters are not major generators of artifacts in genetic association studies of psychiatric disorders with genes from this set. Therefore, factors other than ethnic differences in genetic variation may explain the discrepancies reported among genetic association studies with this set of genes to date. The transferability of tagSNPs defined in the HapMap populations to other worldwide populations was also investigated and found to be high. A list of tagSNPs per gene and continental region is proposed providing a guide for future association studies with these genes. **© 2008 Wiley-Liss, Inc.**

KEY WORDS: association studies; dopamine pathway genes; serotonin pathway genes; tagSNPs; genetic polymorphisms

Please cite this article as follows: Gardner M, Bertranpetit J, Comas D. 2008. Worldwide Genetic Variation in Dopamine and Serotonin Pathway Genes: Implications for Association Studies. Am J Med Genet Part B 147B:1070–1075.

INTRODUCTION

The dopamine and serotonin systems are two of the major neurotransmitter systems in humans. Dopamine affects brain processes that control both motor and emotional behavior and it is known to have a role in the brain's reward mechanism [Schultz, 2002]. Serotonin has a critical role in temperature regulation, sensory perception, locomotion, and sleep and mood control. It is involved in the pharmacology of depression and psychosis [Cravchik and Goldman, 2000]. Genetic association studies continue to explore the contribution of gene variants in the dopaminergic and serotinergic systems to psychiatric, neurological and behavioral disorders with varying success.

Most psychiatric diseases and behavioral disorders are complex genetic traits with both genetic and environmental factors influencing an individual's risk of being affected. The heritability estimate for some of the psychiatric disorders with an accepted genetic component ranges from 0.28 to 0.90 [Merikangas and Risch, 2003] in addition to there being variability in the estimation of the genetic component within each condition. It is also suggested that at least some of the same predisposition genes may form the basis of susceptibility to a spectrum of psychiatric disorders not just a single disease entity, for example the spectrum that encompasses bipolar disorder to schizophrenia is suggested to have at least some of the same genes giving rise to susceptibility [Craddock et al., 2001; Owen et al., 2007].

This complex pattern of inheritance complicates efforts to identify the genes contributing to the genetic component of an individual's susceptibility to the disease. Despite the numerous association studies to date for each of the dopamine and serotonin system genes and psychiatric illness, there has been trouble reaching consensus in the findings. In particular there has been marked problems in the replication of positive findings in an association study in further studies [Cardon and Bell, 2001; Lohmueller et al., 2003]. This reflects the problem of pinpointing genes involved in complex diseases in general and it is also complicated by specific characteristics of psychiatric disease such as the definition of clinical phenotypes [Merikangas and Risch, 2003]. In fact, very few examples of an accepted link between a specific gene variant and a psychiatric disorder have been described [Owen et al., 2000; Hirschhorn et al., 2002; Thiselton et al., 2004].

One reason for inconsistencies in replication studies may be variability in association among different populations [Colhoun et al., 2003]. This could be due to real variance in disease incidence among populations, genetic heterogeneity, or

Grant sponsor: Fundación Genoma España (proyectos piloto CeGen 2004–2005); Grant sponsor: Dirección General de Investigación, Ministerio de Educación y Ciencia, Spain BFU2006-01235/BMC; Grant sponsor: Direcció General de Recerca, Generalitat de Catalunya 2005SGR/00608; Grant sponsor: CIBERESP, Spain 2004FI-IQUC2/0047.

*Correspondence to: David Comas, Unitat de Biologia Evolutiva, Universitat Pompeu Fabra, Doctor Aiguader 88, 08003 Barcelona, Spain. E-mail: david.comas@upf.edu

Received 9 October 2007; Accepted 17 December 2007

DOI 10.1002/ajmg.b.30717

Published online 12 February 2008 in Wiley InterScience (www.interscience.wiley.com)

© 2008 Wiley-Liss, Inc.

the combined role of population differences in the susceptibility alleles for different psychiatric diseases and complex phenotypes. Crucially, population differences in linkage disequilibrium (LD) and allele frequencies (of marker or disease susceptibility variant) as a result of the distinct population history of the samples under study (genetic drift, migration and bottlenecks) may play a role [Zondervan and Cardon, 2004; Gardner et al., 2006].

LD has been found to be variable within and among loci and populations [Pritchard and Przeworski, 2001; Gabriel et al., 2002], and this in the context of association studies potentially means differences in LD between causal variant and marker SNP in different populations. Since most association studies rely on the LD between marker and disease causing variant to detect association, this is a possible contributing factor to failure of replication in different study populations. Variability in disease or marker allele frequencies between populations is another possible confounding factor, as mismatches between marker and disease allele frequencies may reduce the statistical power to detect associations between the complex trait and the candidate gene [Zondervan and Cardon, 2004].

Analysis of common SNPs now forms the basis of current approaches to identify the genetic basis of complex disease. The widely accepted 'common variant, common disease' hypothesis [Lohmueller et al., 2003] holds that high frequency (>1%) low penetrant alleles underlie the genetic component of complex diseases such as psychiatric disorders. There are estimated to be more than 9–10 million common SNPs (MAF $\geq$ 0.05) in the human genome [Frazer et al., 2007]. Systematic studies of these genetic variants for association with disease are facilitated by a certain redundancy in SNP genotyping due to LD, that is the tendency of certain alleles to be inherited together. Recent large scale human genetics projects, such as HapMap, aim to quantify LD across the genome and in different populations. The HapMap project, which includes individuals from four populations (CEPH, European Ancestry; Yoruba, Nigeria; Japanese; and Han Chinese), has as one of its main aims the aiding of the reduction of the number of SNPs necessary for typing in association studies [International HapMap Consortium, 2005]. These so called 'tagSNPs' can reduce greatly the cost and time involved in genetic association studies.

Whole genome scans are emerging as a strong alternative to the candidate genes or pathway approach. Recent results of whole genome scans in bipolar disorder however have shown a lack of power for conditions such as psychiatric disorders [The Welcome Trust Case Control Consortium, 2007] in which phenotypes are complex and likely to have a heterogeneous molecular basis. As an example the only clear result of the recent huge Welcome Trust Case Control Consortium (WTCCC) effort, at chromosome 16p12 for bipolar disorder has not been replicated in further analyses (P. Donnelly, personal communication). Genome wide association studies (GWA) remain costly however and some problems persist such as the difficulty in detecting genes of small effect [The Welcome Trust Case Control Consortium, 2007]. In addition many of the genotyping chips used for these GWA studies are focused on specific gene sets (for example cancer genes) or attempt to have whole genome coverage but in fact fail to achieve this in practice [The Welcome Trust Case Control Consortium, 2007]. For these reasons it is likely that beyond huge genome scans the candidate pathway approach based on a solid rational still has a future in psychiatric genetics.

Genes from the dopamine and serotonin pathways represent some of the genes most commonly tested for association to psychiatric and behavioral disorders. In this study we aim to describe population based genetic variability in allele frequencies and LD across a set of worldwide populations in a set of 21 genes from these pathways. In addition, we investigate the applicability of the tagSNPs selected in HapMap populations to the more diverse set of human populations of the present study. In this way, we provide a guide for future association studies based on tagSNP selection from HapMap populations applied to other worldwide populations. The LD quantity of the genes in worldwide populations along with allele frequency differences are also considered and discussed.

METHODS

Samples

The sample set consists of the Human Genome Diversity Panel (HGDP) [Cann et al., 2002] a set of 1,064 purified DNA samples from worldwide populations covering most of the whole human genetic diversity [Rosenberg et al., 2002]. From the original panel, several duplicated samples were used as internal controls and some atypical individuals reported previously [Rosenberg, 2006] were omitted. Thus, the panel used is the H1048 according to Rosenberg [2006] which includes 1,048 individuals. Samples were regrouped into 39 populations based on geographic and ethnic criteria to avoid small sample size. Populations were also categorized into seven broad continental regions: Sub-Saharan Africa (SSAFR), Middle East/North Africa (MENA), Europe (EUR), Central/South Asia (CSASIA), East Asia (EASIA), Oceania (OCE), and America (AME), see Gardner et al. for the classification of populations into continental groupings.

Gene Choice, SNP Selection and Genotyping

We present a study which includes many of the important dopamine and serotonin system genes that have been associated with psychiatric disease/behavioral traits, a total of 21 genes. In many cases the gene had been associated with more than one psychiatric disease or behavioral trait. SNPs were chosen at a density of every 5–10 kb within the genes. Additionally, a number of SNPs were selected up to 30 kb flanking 3′ and 5′ of each gene, if possible at 30, 20, 10, and 5 kb flanking either side. All the SNPs were chosen on the basis of being involved in the HapMap project, having a high Illumina genotyping score (0.6 or higher) and also a minimum minor allele frequency of at least 0.1 in one of the HapMap populations to increase the probabilities of the SNPs being polymorphic in all of the worldwide populations. In addition to the above, any coding SNPs with a reasonable Illumina score were chosen regardless of the other criteria applied to non-coding SNPs.

Genotyping was carried out using the Beadarray® Platform (Illumina Inc., San Diego, CA) [Oliphant et al., 2002] according to the manufacturer's protocol.

Data Analysis

The F_{ST} (a measure of population differentiation based on allele frequencies) was calculated using the Arlequin package [Excoffier and Schneider, 2005] for each of the 21 genes at the population and continental level. The average of the F_{ST} values for the 18 autosomal genes (excluding the genes in the X chromosome in order to avoid genome biases) was compared to averages for other F_{ST} value distributions (all autosomal genes). The F_{ST} value distributions were obtained from the following datasets: the Marshfield Indel set (210 biallelic anonymous markers), the 427 gene based SNPs in the ALFRED database [Cheung et al., 2000] (a similar set of 38 worldwide populations as the ones used in the present analysis), and a set of 121 SNPs from a 'neutral' gene free region of Chromosome 22 [Gonzalez-Neira et al., 2004]. In order to provide a graphical view of the population differences, F_{ST} genetic distances [Reynolds and Cockerham, 1983] were

calculated for all autosomal SNPs for each population and represented in a multidimensional scaling (MDS) plot with the STATISTICA program.

SNPator, a SNP data management program (Morcillo et al., in press) was used for SNP data handling and basic data analysis such as calculation and filtering of genotype failures and allele frequency and related calculations including Hardy–Weinberg equilibrium. The phasing of the genotypes using the Bayesian algorithm based PHASE program [Stephens et al., 2001] for haplotype reconstruction was carried out within the framework of the SNPator program. The default settings for PHASE were used and the number of iterations was set to 1,000. In addition, independently of SNPator the Haploview program [Barrett et al., 2004] was used for visualization of LD and haplotype block structure, loading the phased haplotypes. The r^2 values for marker pairs were obtained from Haploview. The quantity of LD was compared between continental regions for the adjacent marker pair r^2 values, using the Friedman test (between all continental groups together) and the Wilcoxon test (pair wise comparisons) within the SPSS package. Variability in the quantity of LD across populations within each continental region was also measured using the Friedman test.

tagSNP Transferability

We examined the efficiency of the set of tagSNPs defined in each of the three HapMap populations (CEPH, EUR; Japanese and Chinese analyzed together as Asian; and Yoruban, Africa) as applied to the populations of the HGDP. The tagSNPs were defined in each of these three HapMap populations from the set of SNPs genotyped in the present study. The tagSNP set defined in the HapMap CEPH population was then applied to the populations of the EUR, MENA and CSASIA continental groupings of the HGDP; the HapMap Asian tagSNPs were applied to EASIA, OCE and AME groups, and finally the Yoruban HapMap tagSNP set was applied to the SSAFR group.

The Tagger program [de Bakker et al., 2005] as implemented in the Haploview program was used for tagSNP selection. Tagger selects tagSNPs according to the Carlson et al. [2004] algorithm. The genotyped SNPs in each gene were included as the pool of SNPs from which the tagSNP set would be defined. A minimum minor allele frequency of 0.001 was applied (the default value) and the r^2 threshold was set to 0.8. The obtained tagSNP set for each of the HapMap populations was applied to the comparable HGDP populations (see above) in the following manner: each of the 39 populations for a gene was uploaded into Haploview in turn. The HapMap tagSNP sets obtained were applied in turn to each HGDP population. Thus, we obtained the mean r^2 value of the HapMap tagSNPs applied to each population, the percentage of SNPs with an $r^2 \geq 0.8$, and the number of tagSNPs.

RESULTS

A total of 303 SNPs (including 17 coding SNPs) were successfully typed with more than 50% success in all populations. The final mean spacing across the 21 genes was 8.36 kb (Table I). No particular SNP or population showed significant deviation from Hardy–Weinberg after correcting for multiple testing.

Genetic Variation Worldwide

Differentiation based on SNP allele frequencies at the population and continental level was tested with the F_{ST} statistic. In general, the F_{ST} values across the dataset were low as observed by a mean F_{ST} value of 0.106 (18 autosomal genes only), indicating low differentiation between populations. This value represents the lowest mean F_{ST} for any of the three F_{ST} distributions used for comparison, consisting of sets of SNPs from genes involved in other genomic pathways or from other genomic regions. The average F_{ST} for X-chromosome genes was higher than for autosomes (0.146), reflecting the lower effective population size (two copies in females and one in males) of this chromosome [International HapMap Consortium, 2005]. We also observed a higher F_{ST} in coding SNPs (mean F_{ST} value of 0.135) compared to non-coding SNPs, that is, coding SNPs in genes exhibited a marked increase in population differentiation across populations, which concurs with previous reports [Hinds et al., 2005].

To view graphically the genetic relationships between populations for the serotonin and dopamine autosomal genes, a MDS plot based on genetic distances was drawn (Fig. 1). The

TABLE I. Genes in Dopamine and Serotonin Pathways Analyzed in Worldwide Populations

Gene	Pathway	Chromosome	Gene length (kb)	N SNPs	SNP spacing (kb)	N Coding SNP (non-syn)
COMT	Dopamine	22q11.21	27,254	16	4.89	1 (1)
DBH	Dopamine	9q34	22,969	16	4.81	4 (3)
DDC	Dopamine	7p11	102,615	21	7.74	2 (1)
DRD1	Dopamine	5q35.1	3,126	8	8.40	1 (1)
DRD2	Dopamine	11q23	65,575	16	7.40	—
DRD3	Dopamine	3q13.3	50,199	14	8.00	1 (1)
DRD4	Dopamine	11p15.5	3,398	6	8.36	—
DRD5	Dopamine	4p16.1	2,029	7	8.09	—
MAOA	Dopamine	Xp11.3	90,601	12	7.86	1
MAOB	Dopamine	Xp11.23	115,835	10	9.91	—
PPP1R1B	Dopamine	17q12	9,696	7	6.36	—
SLC6A3	Dopamine	5p15.3	52,636	16	6.93	2
TH	Dopamine	11p15.5	7,875	9	8.43	1 (1)
HTR1A	Serotonin	5q11.2-q13	1,268	6	10.49	—
HTR1B	Serotonin	6q13	1,259	10	17.25	2 (2)
HTR2A	Serotonin	13q14-q21	62,665	18	9.91	—
HTR2C	Serotonin	Xq24	326,073	39	6.14	1 (1)
HTR4	Serotonin	5q31-q33	203,037	32	7.23	—
SLC6A4	Serotonin	17q11.1-q12	37,799	14	9.72	—
TPH1	Serotonin	11p15.3-p14	21,113	9	10.01	—
TPH2	Serotonin	12q21.1	93,595	17	7.66	1

N SNPs, number of SNPs genotyped in each gene; N Coding SNP, number of coding SNPs (in parentheses, the number of non-synonymous coding SNPs).

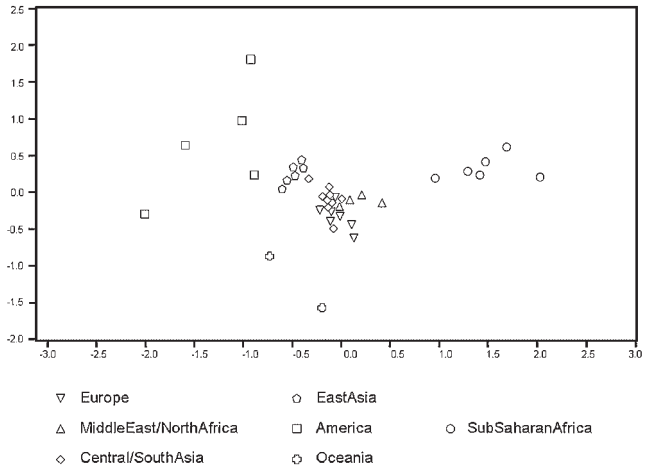

Fig. 1. Multidimensional scaling (MDS) plot of the F_{ST} population values for the serotonin and dopamine genes. Each point corresponds to the average F_{ST} values for a population.

plot shows populations grouped according to continental region, with slightly higher intra-continental differences between AME, OCE and African populations, whereas Europeans and Asians present high genetic homogeneity for serotonin and dopamine genes.

In the comparisons of LD between continental regions, a significant ($P < 0.05$) result comparing all continental groups (by means of the Friedman test) or performing pair wise comparisons (Wilcoxon test) would imply a difference in the amount of LD. Eight out of 21 of the genes showed significant differences in quantity of LD between continental regions ($P < 0.05$, Friedman test) after correction for multiple testing. This was mainly explained by the differences exhibited by the *HTR4* and *HTR2C* genes in LD amount as seen with the Wilcoxon pair wise analysis (significant after correction for multiple testing). For *HTR2C* this was due to differences between the SSAFR compared to the EUR and MENA continental groups and for *HTR4* there were differences between SSAFR and other continental groupings and also between EUR and the other continental groupings. In addition, the *COMT* gene also showed significant differences between SSAFR populations and the rest of the world populations. The contribution of the SSAFRs to these differences was confirmed by looking at the LD structure, where there was a lower amount of LD than in other continents. The *HTR2C* gene, located in the X chromosome, is almost a contiguous stretch of LD in most continental groups except for SSAFR. This result should be interpreted with caution however, as the X chromosome in general is known to have higher LD [Altshuler et al., 2005] due to its lower effective population size.

A significant result in the Friedman test for quantity of LD across populations within a continental group would denote differences in LD within that continental group. On the whole, LD levels were relatively homogeneous within continents with just *HTR2C* in the CSASIA and MENA continental grouping, the *HTR4* gene in SSAFRs, and the *TPH2* gene in CSASIAs showing significant differences in LD after correction for multiple testing.

tagSNP Transferability

Of the 303 SNPs genotyped, the number of tagSNPs for each HapMap sample was 173 for Europeans (CEPH sample), 167 for Asians (Chinese and Japanese) and 213 for Africans (Nigerian Yoruba; Supplementary Information 1, http://www.upf.edu/cexs/recerca/bioevo/softanddata.htm). This greater number of tagSNPs in Africans is in agreement with previous results [Carlson et al., 2004; Gonzalez-Neira et al., 2006; Gu et al., 2007] due to the lower LD in Africa. Thus, the average "tagging efficiency" (total number of SNPs genotyped divided by the number of tagSNPs in each sample) is lower in African samples (average 1.4), compared to Europeans and Asians (average ~1.8; Table II). The tagging efficiency differs substantially depending on the gene analyzed, ranging from 1 (the lowest efficiency where all the SNPs are tagSNPs) in the case of *TH* gene to 9.75 in the case of *HTR2C* in Europeans, where only four SNPs would be needed to capture the 39 SNPs genotyped in the gene with an $r^2 \geq 0.8$.

The dopamine and serotonin tagSNPs defined in the HapMap populations (CEPH-Europeans, Chinese/Japanese, and Yoruba) applied to HGDP populations grouped into three equivalent groups [see Gardner et al., 2006 for the grouping categories] capture, as an average, more than 80% of the SNPs genotyped: 85.6% of the SNPs in the European group, 82.9% in the Asian group, and 83.4% in the African group with an $r^2 \geq 0.8$. Furthermore, the results obtained for populations within each continental region were extremely similar (data not shown) indicating that the tagSNPs defined in the HapMap populations can be applied to any of their corresponding population samples. For the entire list of the tagSNPs proposed for each individual gene in each continental region see Supplementary Information 2 (http://www.upf.edu/cexs/recerca/bioevo/softanddata.htm).

DISCUSSION

One of the principal aims of the present study was to shed light on the population related reasons underlying the lack of consistent replication of association findings in the context of the dopamine and serotonin genes with psychiatric disease. The results found analyzing 303 SNPs in 21 genes in 39 worldwide samples demonstrate that the dopamine and serotonin SNP frequencies across populations are no more differentiated across populations than other regions of the

TABLE II. Average Tagging Efficiency[a] of HapMap Populations in Dopamine and Serotonin Genes

	European	Asian	African
COMT	1.45	1.45	1.23
DBH	1.33	1.78	1.33
DDC	1.31	1.91	1.40
DRD1	1.60	1.33	1.33
DRD2	1.45	1.60	1.23
DRD3	1.14	1.56	1
DRD4	1.20	1.20	1.20
DRD5	2.33	1.17	1.40
HTR1A	6	2	2
HTR1B	1.67	1.25	2
HTR2A	1.16	1.38	1.29
HTR2C	9.75	4.88	2.05
HTR4	2.46	1.68	1.52
MAOA	4	6	2.40
MAOB	1.43	1.67	1.43
PPP1R1B	1.40	7	1.40
SLC6A3	1.45	2.29	1.33
SLC6A4	2	2.33	1.27
TH	1	1	1
TPH1	1.50	1.29	1.29
TPH2	1.70	1.55	1.42

[a]Total number of SNPs genotyped divided by the number of tagSNPs in each sample.

genome, as shown by the average F_{ST} values. However, coding SNPs showed higher variance in allele frequencies across populations and continents, in agreement with previous reports [Hinds et al., 2005], as did SNPs on the X chromosome. In addition, none of the populations studied was an outlier of note with regard to allele frequencies, and populations showed a fair amount of genetic homogeneity within continental groups as seen in the MDS plot of the average F_{ST} values per population.

Another remarkable result of the present study is the population homogeneity in dopamine and serotonin LD within geographical regions, with LD being lower in SSAFR populations in agreement with previous reports [Reich et al., 2001; Gonzalez-Neira et al., 2004; Hinds et al., 2005]. One of the possible causes behind the lack of replication of association studies is the variation of LD among the different populations studied. The LD approach to gene mapping assumes that the associated variant is not causal but is in high LD with such a disease causing variant. Our results show that the failure to replicate a finding of association involving dopamine and serotonin genes between populations of the same geographical region to be due factors (such as misclassification of phenotypes, lack of power, false positives or chance) other than population differences in LD between the markers analyzed and the causal variant. This LD homogeneity in serotonin and dopamine genes within continental regions is in agreement with previous results in other genomic regions [Gonzalez-Neira et al., 2004, 2006].

The International HapMap project had as one of its main aims the characterization of LD across different genomic regions towards the end of facilitating genetic association studies [International HapMap Consortium, 2005]. A definitive conclusion on the applicability of the findings from the three HapMap populations (a European, African and Asian population) to other worldwide populations has not been yet reached although the existing analyses suggest it is promising, [Gonzalez-Neira et al., 2006; Ribas et al., 2006; de Bakker et al., 2006; Gu et al., 2007]. The utilization of tagSNPs in association studies takes advantage of LD to reduce the number of SNPs typed, allowing a reduction of cost and time. We addressed this issue in the context of a set of dopamine and serotonin genes commonly examined for association to psychiatric disease and with many reported associations including a few confirmed associations. Our results show differences in tagging efficiency depending on the amount of LD. African populations as compared to European or Asians show a lower tagging efficiency in the present set of genes, and therefore, more SNPs should be typed for these populations when performing an association study with dopamine and serotonin genes. Furthermore, the LD differences across genomic regions cause tag efficiency differences depending on the gene analyzed: the genotyping effort in *HTR2C* can be reduced in all populations due to the presence of high LD, whereas other genes such as *TH* exhibit a low efficiency and therefore more SNPs are necessary to ensure sufficient coverage.

The transferability of tagSNPs defined in the HapMap samples to other populations not represented in HapMap is a crucial issue. Initial studies of transferability of tagSNPs defined in HapMap to the HGDP worldwide population dataset for other genomic regions have indicated that HapMap data could be used as a good proxy to the diversity found in other populations [Gonzalez-Neira et al., 2006; Gu et al., 2007]. Our results for serotonin and dopamine genes reinforces this transferability of HapMap data to other population samples. Association studies containing European, Middle Eastern, North African, and South/Central Asian samples may therefore use the HapMap CEPH-European tagSNPs in dopamine and serotonin pathway genes since our present results show a clear homogeneity of these samples in allele frequencies and LD. In the same way, association studies involving EASIA, Amerindian, and OCE samples may use the HapMap Asian tagSNPs, and finally HapMap Yoruban tagSNPs may be used in association studies containing African samples.

WEBSITES

SNPator: http://bioinformatica.cegen.upf.es.
HapMap: www.hapmap.org.
Haploview: http://www.broad.mit.edu/mpg/haploview/.

ACKNOWLEDGMENTS

We are grateful to Anna González-Neira in the Centro Nacional de Genotipación (CeGen), Madrid for help with the SNP genotyping; and Howard Cann, from CEPH, Paris, for his effort in providing the HGDP samples. We would like to thank Francesc Calafell (UPF) for suggestions on the analysis and Araceli Rosa (UPF) for critical reading of the manuscript. We thank Ferran Casals, Elena Bosch, Arcadi Navarro, Carlos Morcillo, Josep Alegre, Ricard Sangrós, Tomàs Marquès, Andrés Moreno and Anna Ferrer (UPF) for technical help. The present study was supported by "Fundación Genoma España" (proyectos piloto CeGen 2004–2005), the Dirección General de Investigación, Ministerio de Educación y Ciencia, Spain (BFU2006-01235/BMC), Direcció General de Recerca, Generalitat de Catalunya (2005SGR/00608). M.G. received a fellowship (2004FI-IQUC2/0047) from the International Graduate School of Catalunya (Generalitat de Catalunya) and is now part of the CIBERESP, Spain.

REFERENCES

Altshuler D, Brooks LD, Chakravarti A, Collins AS, Daly MJ, Donnelly P. 2005. A haplotype map of the human genome. Nature 437(7063):1299–1320.

Barrett JC, Fry B, Maller J, Daly MJ. 2004. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265.

Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, et al. 2002. A human genome diversity cell line panel. Science 296(5566):261–262.

Cardon LR, Bell JI. 2001. Association study designs for complex diseases. Nat Rev Genet 2(2):91–99.

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(1):106–120.

Cheung KH, Osier MV, Kidd JR, Pakstis AJ, Miller PL, Kidd KK. 2000. ALFRED: An allele frequency database for diverse populations and DNA polymorphisms. Nucleic Acids Res 28:361–363.

Colhoun HM, McKeigue PM, Davey Smith G. 2003. Problems of reporting genetic associations with complex outcomes. Lancet 361(9360):865–872.

Craddock N, Dave S, Greening J. 2001. Association studies of bipolar disorder. Bipolar Disord 3(6):284–298.

Cravchik A, Goldman D. 2000. Neurochemical individuality: Genetic diversity among human dopamine and serotonin receptors and transporters. Arch Gen Psychiatry 57(12):1105–1114.

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(11):1217–1223.

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(11):1298–1303.

Excoffier LGL, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol Bioinformatics Online 1:47–50.

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449(7164):851–861.

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(5576):2225–2229.

Gardner M, Gonzalez-Neira A, Lao O, Calafell F, Bertranpetit J, Comas D. 2006. Extreme population differences across Neuregulin 1 gene, with implications for association studies. Mol Psychiatry 11(1):66–75.

Gonzalez-Neira A, Calafell F, Navarro A, Lao O, Cann H, Comas D, Bertranpetit J. 2004. Geographic stratification of linkage disequilibrium: A worldwide population study in a region of chromosome 22. Hum Genomics 1(6):399–409.

Gonzalez-Neira A, Ke X, Lao O, Calafell F, Navarro A, Comas D, Cann H, Bumpstead S, Ghori J, Hunt S, et al. 2006. The portability of tagSNPs across populations: A worldwide survey. Genome Res 16(3):323–330.

Gu S, Pakstis AJ, Li H, Speed WC, Kidd JR, Kidd KK. 2007. Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations. Eur J Hum Genet 15:302–312.

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. 2005. Whole-genome patterns of common DNA variation in three human populations. Science 307(5712):1072–1079.

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. 2002. A comprehensive review of genetic association studies. Genet Med 4(2):45–61.

International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437(7063):1299–1320.

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33(2):177–182.

Merikangas KR, Risch N. 2003. Will the genomics revolution revolutionize psychiatry? Am J Psychiatry 160(4):625–635.

Morcillo C, Alegre P, Sangros R, Gazave E, de Cid-Ibeas R, Milne R, Amigo J, Ferrer-Amedlla A, Moreno-Estrada A, Gardner M, et al. SNP analysis to results (SNPator): A web-based environment oriented to statistical genomics analyses upon SNP data. Bioinformatics (Submitted for publication).

Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. 2002. Beadarray technology: Enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques 32(Suppl):S56–S61.

Owen MJ, Cardno AG, O'Donovan MC. 2000. Psychiatric genetics: Back to the future. Mol Psychiatry 5(1):22–31.

Owen MJ, Craddock N, Jablensky A. 2007. The genetic deconstruction of psychosis. Schizophr Bull 33(4):905–911.

Pritchard JK, Przeworski M. 2001. Linkage disequilibrium in humans: Models and data. Am J Hum Genet 69(1):1–14.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, et al. 2001. Linkage disequilibrium in the human genome. Nature 411(6834):199–204.

Reynolds J, Weir BS, Cockerham CC. 1983. Estimation for the coancestry coefficient: Basis for a short-term genetic distance. Genetics 105(3):767–779.

Ribas G, Gonzalez-Neira A, Salas A, Milne RL, Vega A, Carracedo B, Gonzalez E, Barroso E, Fernandez LP, Yankilevich P, et al. 2006. Evaluating HapMap SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. Hum Genet 118(6):669–679.

Rosenberg NA. 2006. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. Ann Hum Genet 70(Pt 6):841–847.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic structure of human populations. Science 298(5602):2381–2385.

Schultz W. 2002. Getting formal with dopamine and reward. Neuron 36(2):241–263.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68(4):978–989.

The Welcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447(7145):661–678.

Thiselton DL, Webb BT, Neale BM, Ribble RC, O'Neill FA, Walsh D, Riley BP, Kendler KS. 2004. No evidence for linkage or association of neuregulin-1 (NRG1) with disease in the Irish study of high-density schizophrenia families (ISHDSF). Mol Psychiatry 9(8):729.

Zondervan KT, Cardon LR. 2004. The complex interplay among factors that influence allelic association. Nat Rev Genet 5(2):89–100.