Molecular Psychiatry (2011), 1–2

www.nature.com/mp

LETTER TO THE EDITOR

Identification of tag haplotypes for 5HTTLPR for different genome-wide SNP platforms

Molecular Psychiatry advance online publication, 14 June 2011; doi:10.1038/mp.2011.68

The length polymorphism repeat (5HTTLPR) in the promoter region of the serotonin transporter gene (SLC6A4, also known as 5HTT) is extensively studied in the context of psychiatric phenotypes, particularly in major depressive disorder. However, investigation of this polymorphism in the context of the current generation of large-scale genome-wide association studies is precluded, as the genotyping technology is limited to single-nucleotide polymorphisms (SNPs). Using genome-wide and 5HTTLPR genotype data from a total of 2823 unrelated individuals, we show that no single SNP is in high linkage disequilibrium (LD) with 5HTTLPR but some two-SNP haplotypes provide reasonable predictors. Hence, two-SNP haplotypes can be used as proxies for 5HTTLPR in genome-wide association studies. Analyses are repeated for sets of SNPs that are included in different genome-wide SNP platforms.

The 5HTTLPR is defined by a length variation of a repetitive sequence with a short (484 base pairs, 14 repeat units) and a long allele (528 base pairs, 16 repeat units) on chromosome 17. The basal activity of the long allele transcript is about threefold higher than that of the short allele, resulting in induced expression and function of the SLC6A4 gene.[1] Numerous studies have investigated the association between 5HTTLPR and anxiety- and depression-related traits, particularly in the context of interaction with stressful life events, with conflicting[2] results. The era of genome-wide association studies has generated much larger sample sizes but investigation of 5HTTLPR has been precluded, as the technology developed for high-throughput genotyping is limited to SNPs and assay of the length polymorphism is expensive and technically challenging. Likewise, 5HTTLPR genotypes are not available for samples included in the International HapMap project so that selection of SNPs that tag 5HTTLPR is not feasible from this publicly available database. Here, we use 2823 samples with both genome-wide SNP and 5HTTLPR genotypes to identify SNPs and/or SNP haplotypes that tag 5HTTLPR polymorphism.

Characteristics of the samples and genotyping quality control are extensively described elsewhere,[3–5] but subsets of samples were genotyped on the Illumina 317 array, Illumina Human370 CNV quad, Illumina HumanHap610 quad and Affymetrix 6.0 platforms. For 5HTTLPR genotyping, we used an assay[3] that is less prone to bias towards short allele calling compared with the original assay[1] and all assays were performed in triplicate to maximise accuracy. We explored the region around 5HTTLPR to examine the extent to which SNPs included in different genome-wide SNP platforms tag 5HTTLPR. We first selected all markers within an ~1000 kilobase region around 5HTTLPR and identified the LD pattern for markers genotyped on this platform. Based on the LD pattern, we selected a narrower region around 5HTTLPR (~25 kb downstream and ~155 kb upstream of 5HTTLPR) for detailed analyses. The Tagger[6] option within Haploview[7] was used to see if any one, two or three marker combinations could be used to predict 5HTTLPR genotype. LD measures (D' and r^2) for all marker combinations within the selected region are provided in the online supplement (Supplementary Tables S1 and S2 for Illumina 610 and Affymetrix 6.0 platfoms, respectively). The highest r^2 between any single SNP and 5HTTLPR was $r^2 = 0.50$, for both rs7214014 genotyped on Illumina HumanHap610 quad and its proxy ($r^2 = 1$) rs8072345 genotyped on Affymetrix 6.0. For Illumina platforms, we found several two-SNP haplotypes that had $r^2 > 0.75$ with 5HTTLPR. The best two-SNP proxy is provided in Table 1 with the TA haplotype of rs2129785 and rs11867581 that tags the short allele. Other two-SNP haplotypes comprised rs2129785 and other perfect or near perfect proxies of rs11867581 (Supplementary Table S5). No haplotypes from the Affymetrix 6.0 platform tagged the 5HTTLPR, although a good proxy for rs2129785 is included on the Affymetrix Axiom chip[8] (Supplementary Table S5).

The long allele of 5HTTLPR harbours an SNP rs25531. Long alleles containing the rarer G allele

Table 1 Haplotype frequencies and linkage disequilibrium estimates (r^2) of identified tagging haplotypes for 5HTTLPR

5HTTLPR/rs2129785/rs11867581
L/T/G: 0.422
S/T/A: 0.412
L/C/A: 0.106
L/T/A: 0.039
S/T/G: 0.018
$r^2 = 0.775$

Abbreviations: L, long allele; S, short allele.
The TA haplotype is coupled with the short allele of the 5HTTLPR.

are functionally equivalent to S alleles;[9] and therefore, some association studies of 5HTTLPR have also included association analysis of the SNP. Therefore, we selected long-long 5HTTLPR participants only and investigated possible tagging SNPs or haplotypes for rs25531. However, no SNPs or multi-marker haplotypes could predict this SNP (Supplementary Tables S3 and S4).

These results complement those reported using the same study sample which showed that the CA haplotype of SNPs rs4251417 (minor allele frequency = 0.091) and rs2020934 (minor allele frequency = 0.489) is coupled with the short allele of 5HTTLPR ($r^2 = 0.72$).[3] However, SNP rs2020934 is not included either in commercial genome-wide SNP platforms or in the HapMap project limiting the use of this haplotype. We found that SNPs rs7214014 and rs8072345, the two SNPS that showed highest LD with 5HTTLPR in our data, were only moderate proxies for rs2020934 ($r^2 = 0.63$ and $r^2 = 0.58$, respectively). SNP rs2020934 did not constitute a better predictive haplotype for 5HTTLPR with any of the other SNPs from our genome-wide association studies data, than the haplotypes reported in Table 1. Based on HapMap3 data, we concluded that no additional SNPs from other genotype platforms than those included in this study were in high LD with SNPs from the tagging haplotypes, implying that no further proxies can be added to the tagging haplotypes for 5HTTLPR that are identified here. In a recent study by Handsaker *et al.*,[10] investigation of LD between SNPs and length polymorphisms from the 1000 genomes did not identify single SNPs that tagged 5HTTLPR with $r^2 \geqslant 0.8$.

To conclude, two-SNP haplotypes, but not single markers, can be used as proxies for 5HTTLPR. This means that existing databases that include subjects with genome-wide genotype data can be used to investigate the association between 5HTTLPR and these phenotypic measures.

Conflict of interest

The authors declare no conflict of interest.

AAE Vinkhuyzen[1], T Dumenil[1], L Ryan[1], SD Gordon[1], AK Henders[1], PAF Madden[2], AC Heath[2], GW Montgomery[1], NG Martin[1] and NR Wray[1]
[1]*Queensland Institute of Medical Research, Brisbane, QLD, Australia and* [2]*Washington University School of Medicine, St Louis, MO, USA*
E-mail: anna.vinkhuyzen@qimr.edu.au

References

1 Lesch KP, Bengel D, Heils A, Sabol SZ, Greenberg BD, Petri S *et al. Science* 1996; **274**: 1527–1531.
2 Risch N, Herrell R, Lehner T, Liang KY, Eaves L, Hoh J *et al. JAMA* 2009; **301**: 2462–2471.
3 Wray NR, James MR, Gordon SD, Dumenil T, Ryan L, Coventry WL *et al. Biol Psychiatry* 2009; **66**: 468–476.
4 Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G *et al. Am J Hum Genet* 2009; **85**: 750–755.
5 Wray NR, Pergadia ML, Blackwood DH, Penninx BW, Gordon SD, Nyholt DR *et al. Mol Psychiatry*; advance online publication, 2 November 2010; doi 10.1038/mp.2010.109.
6 de Bakker PIW, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. *Nat Genet* 2005; **37**: 1217–1223.
7 Barrett JC, Fry B, Maller J, Daly MJ. *Bioinformatics* 2005; **21**: 263–265.
8 Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. *Bioinformatics* 2008; **24**: 2938–2939.
9 Wendland JR, Martin BJ, Kruse MR, Lesch KP, Murphy DL. *Mol Psychiatry* 2006; **11**: 224–226.
10 Handsaker RE, Korn JM, Nemesh J, McCarroll SA. *Nat Genet* 2011; **43**: 269–276.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)