

UC Davis

UC Davis Previously Published Works

Title

Population frequencies of the Triallelic 5HTTLPR in six Ethnically diverse samples from North America, Southeast Asia, and Africa.

Permalink

<https://escholarship.org/uc/item/54b2355c>

Journal

Behavior Genetics, 45(2)

Authors

Haberstick, Brett
Smolen, Andrew
Williams, Redford
[et al.](#)

Publication Date

2015-03-01

DOI

10.1007/s10519-014-9703-5

Peer reviewed



Published in final edited form as:

Behav Genet. 2015 March ; 45(2): 255–261. doi:10.1007/s10519-014-9703-5.

POPULATION FREQUENCIES OF THE TRIALLELIC 5HTTLPR IN SIX ETHNICALLY DIVERSE SAMPLES FROM NORTH AMERICA, SOUTHEAST ASIA, AND AFRICA

Brett C. Haberstick^{1,*}, Andrew Smolen¹, Redford B. Williams², George D. Bishop^{3,4}, Vangie A. Foshee⁵, Terence P Thornberry⁶, Rand Conger⁷, Ilene C. Siegler², Xiaodong Zhang^{2,8,9}, Jason D Boardman¹⁰, Zygmunt Frajzyngier¹¹, Michael C. Stallings¹, M. Brent Donnellan¹², Carolyn T. Halpern¹³, and Kathleen Mullan Harris^{14,15}

¹Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, Colorado, USA

²Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, North Carolina, USA

³Yale-NUS College, Singapore, Singapore

⁴Department of Psychology, National University of Singapore, Singapore

⁵Department of Human Behavior, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

⁶Department of Criminology & Criminal Justice, University of Maryland, College Park, Maryland, USA

⁷Department of Human Ecology, University of California, Davis, Davis, California, USA

⁸Neuroscience & Behavioral Disorders Program, Duke-NUS Graduate Medical School Singapore, Singapore

⁹Department of Physiology, National University of Singapore, Singapore

¹⁰Institute of Behavioral Science, University of Colorado Boulder, Boulder, Colorado, USA

¹¹Department of Linguistics, University of Colorado Boulder, Colorado, USA

¹²Department of Psychology, Michigan State University, East Lansing, MI, USA

¹³Department of Maternal and Child Health, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

¹⁴Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Corresponding Author: Brett C. Haberstick, Institute for Behavioral Genetics, University of Colorado Boulder, Campus Box 447, Boulder, Colorado 80309-0447, Telephone: 303-735-5388, Fax: 303-492-8063, Brett.Haberstick@Colorado.edu.

Financial Disclosures & Conflict of Interest Statement: Redford Williams holds a U.S. patent on 5HTTLPR L allele as a risk marker for CVD in persons exposed to chronic stress. Redford Williams is a founder and major stockholder in Williams LifeSkills, Inc., a company that develops, tests and markets behavioral products for stress and anger management.

Informed Consent: All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all participants for being included in the study.

¹⁵Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Abstract

Genetic differences between populations are a potentially an important contributor to health disparities around the globe. As differences in gene frequencies influence study design, it is important to have a thorough understanding of the natural variation of the genetic variant(s) of interest. Along these lines, we characterized the variation of the 5HTTLPR and rs25531 polymorphisms in six samples from North America, Southeast Asia, and Africa (Cameroon) that differ in their racial and ethnic composition. Allele and genotype frequencies were determined for 24,066 participants. Results indicated higher frequencies of the rs25531 G-allele among Black and African populations as compared with White, Hispanic and Asian populations. Further, we observed a greater number of 'extra-long' ('XL') 5HTTLPR alleles than have previously been reported. Extra-long alleles occurred almost entirely among Asian, Black and Non-White Hispanic populations as compared with White and Native American populations where they were completely absent. Lastly, when considered jointly, we observed between sample differences in the genotype frequencies within racial and ethnic populations. Taken together, these data underscore the importance of characterizing the L-G allele to avoid misclassification of participants by genotype and for further studies of the impact XL alleles may have on the transcriptional efficiency of *SLC6A4*.

Keywords

5HTTLPR; Serotonin Transporter; rs25531; Hispanics; Add Health; population genetics

Introduction

Because of its wide distribution throughout the brain, the neurotransmitter serotonin (5HT) has been implicated in a number of neuropsychiatric phenotypes. Through its removal of 5HT from the synapse, the serotonin transporter (*5HTT*, Gene Symbol: *SLC6A4*) plays an important role in the extent and duration of serotonergic signaling. The 5HTT mRNA is encoded by a single gene consisting of 15 exons and is located on chromosome 17 at 17q11.2 (Ramamoorthy et al, 1993). The reported transcriptional differences of the 5HTT gene (Lesch et al, 1996) have been ascribed to genetic variants (Heils et al, 1996; Hu et al 2006). The serotonin transporter linked polymorphic region (5HTTLPR) located 1 kb from the transcription start site consists of a number of 20- to 23- base pair (bp) repeat units that can vary from 13 to 22 units. Within the 5HTTLPR is a 43 bp insertion/deletion (ins/del; Heils et al, 1996) which gives rise to the most common 14-repeat (14R, Short) and 16R (Long) alleles. The ins/del in 5HTTLPR is associated with variations in transcriptional activity: the long promoter variant has approximately three times the expression of the short promoter with the deletion (Lesch et al, 1996). The 14R and 16R alleles account for the majority of the 5HTTLPR alleles, although the distribution of the 17R, 18R, 19R, 20R, and 22R alleles ('XL-alleles') vary considerably among populations (Nakamura et al, 2000; Odgerel et al, 2013; Gelernter et al., 1997; Goldman et al, 2010; Murdoch et al, 2013).

The single nucleotide polymorphism (SNP) rs25531 is also thought to contribute to transcriptional differences of the 5HTT. The A → G transition results in two forms of the L-allele (16R) denoted L-A and L-G and two corresponding S-A and S-G alleles, though the S-G is infrequently observed. This SNP is located either within (Hu et al, 2006, Wray et al, 2009) or immediately outside (Nakamura et al, 2000; Murdoch et al, 2013; Wendland et al, 2006) the 5HTTLPR and was found to segregate nearly perfectly with the L-allele. The occurrence of a G in the L-allele results in a reduction in the transcriptional efficiency of the 5HTT gene similar to that of the short allele (Hu et al, 2006). Until recently, the impact of this allele was not accounted for in genetic association studies and this omission may contribute to the heterogeneity of findings in the 5HTTLPR literature.

Recently, Murdoch et al (2013) published a timely and valuable summary of the population variation at *SLC6A4* (9). These investigators characterized the population-specific distributions of several polymorphisms in *SLC6A4* with focused consideration of the 5HTTLPR and rs25531. In their study of approximately 2,500 individuals around the globe, they observed, among other things: 1) an absence of the G-allele in Native Americans, 2) frequency differences in the 14R that may reflect migration patterns from Africa and Asia into Europe and the Americas, and 3) a total of 10 rare 5HTTLPR allele size classes that included 13R, 15R, 17R, 18R, 19R 20R, and 22R alleles.

Given the potential importance of population differences in the 5HTTLPR and rs25531 for genetic and pharmacogenetic association studies, our group recently characterized both polymorphisms in four samples from North America, one from Southeast Asia, and one from rural villages in Cameroon, Africa. In total we present information for 24,066 individuals. With these data, we aimed to 1) determine allele frequencies in the SNP rs25531 across diverse racial and ethnic populations; 2) assess the evidence for population differences in the distribution in 5HTTLPR genotype as a function of rs25531 status; and 3) evaluate frequency of extra-long ('XL') alleles in the sample populations.

Methods

Participants

A total of 24,066 individuals were drawn from six ongoing studies within the United States, Asia, and Africa. These studies included the: 1) National Longitudinal Study of Adolescent to Adult Health (Add Health; n = 14,784; Harris et al, 2013; Haberstick et al, 2014); 2) Genes in Context (GIC; n = 1,368; Foshee et al, 2014); 3) Rochester Youth and Development Study (RYDS; n = 784; Thornberry et al, 2013); 4) Singapore Cardiovascular Cohort Study (SCCS II; n = 4222; Kaur and Bishop, 2013); 5) Family Transitions Project (FTP; n = 2379; Conger et al, 2012); 6) Cameroon Language and Genetics Study (CLGS; n = 529). Further information on each of these studies is provided in the online Supplement.

Genotyping

National Longitudinal Study of Adolescent to Adult Health (Add Health), Genes In Context (GIC), Rochester Youth Development Study (RYDS), Cameroon Sample, Family in Transitions Project—5HTTLPR genotype was

determined using polymerase chain reaction [PCR] (Haberstick et al, 2014) using DNA collected using the Oragene system (DNAgenotek, Kanata, Ontario, Canada). PCR reactions contained one μ l of DNA [20 ng or less], 1 \times Buffer II [Life Technologies (Life Tech), Grand Island, NY], 1.8 mM MgCl₂, 180 μ M each deoxynucleotide [dNTP, NEB], with 7'-deaza-2'-deoxyGTP (deaza-GTP, Roche Applied Science, Indianapolis, Indiana) substituted for one-half of the dGTP, 10% dimethylsulfoxide (Sigma-Aldrich, St. Louis, MO) forward (fluorescently labeled) and reverse primers, and one unit of AmpliTaq Gold[®] polymerase (Life Tech) in a total volume of 20 μ l. Forward and reverse primer sequences were: NED-ATG CCA GCA CCT AAC CCC TAA TGT (concentration: 600 nM) and GGA CCG CAA GGT GGG CGG GA (concentration: 600 nM), respectively. Amplifications were performed using a modified (Anchordoquy et al, 2003) touchdown PCR method (Don et al, 1992). A 95°C incubation for 10 minutes was followed by two cycles of 95°C for 30 seconds, 65°C for 30 seconds, and 72°C for 60 seconds. The annealing temperature was decreased every two cycles from 65°C to 57°C in 2°C increments (10 cycles total), followed by 30 cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 60 seconds, a final 30-minute incubation at 72°C and a hold at 4°C.

The SNP rs25531 was assayed in order to determine the L-A and L-G alleles using the PCR method described above with the substitution of primer sequences reported by Hu et al (Hu et al, 2006): 6FAM-GCA ACC TCC CAG CAA CTC CCT GT (500 nM) and GAG GTG CAG GGG GAT GCT GGA A (500 nM), respectively. PCR products were incubated with 5 units of MspI (NEB, Ipswich, MA) for 90 minutes at 37°C (Wendland et al, 2006). A 97 bp restriction digest fragment was indicative of the L-G allele.

Following amplification, PCR products and MspI digests were filter purified using Zymo Research (San Deigo, CA) ZR-96 DNA Sequencing Clean-up Kits following protocols supplied by the manufacturer. An aliquot of PCR products was combined with loading buffer containing size standard [Ro \times 1000, Gel Company, San Francisco, CA) and analyzed with an ABI PRISM[®] 3130 \times 1 Genetic Analyzer (Life Tech) using protocols supplied by the manufacturer. Fragment sizes were analyzed with Genemapper software with the resulting allele sizes independently reviewed by two investigators.

Singapore Cardiovascular Cohort Study (SCCS II)—5HTTLPR genotyping used methods modified from Wray et al (6) and used the Oragene collection system. Briefly, a 50 μ L PCR reaction mix was set up with 20 ng genomic DNA, 1.7 mM MgCl₂, 0.2 μ M forward and reverse primer, respectively, 0.2 mM dNTP, 1.3M betaine (Sigma), 2.5 U GoTaq DNA polymerase (Promega). PCR was run at the following conditions: 1 cycle (95°C for 5 minutes), 40 cycles (95°C for 45 seconds, 55°C for 45 seconds, 72°C for 45 seconds), 1 cycle (72°C for 10 minutes). The forward primer was (5'- CCT TCA CTC CTC GCG GC), and the reverse primer was (5'- GCAG GGG GGA TAC TGC GA), both of which were adapted from Wray et al (2009). PCR products were run on a 2.5% agarose gel. Long-form (L) and short-form were shown by their sizes of 180bp and 137 bp, respectively. In order to distinguish the L-A and L-G forms, 20 μ L PCR products were incubated in a 30 μ L enzyme digestion reaction buffer with 3 μ g BSA, 5 U HPA II (NEB) at 37°C overnight. Digestion pattern was revealed by a 2.5% agarose gel with an uncut 180 bp fragment for the L-A form or a 147 bp fragment for the L-G form.

Analysis

The ethnic group membership was that assigned by each study based on self-reports. Allele frequencies and population genetic parameters were calculated using the PROC ALLELE statement in SAS/GENETICS (Version 9.1; SAS Institute Inc., 2005). PROC ALLELE calculates polymorphic information content (PIC), heterozygosity, Hardy Weinberg Equilibrium (HWE), allelic diversity statistics that can provide an indication of marker informativeness.

Results and Discussion

Single nucleotide polymorphism (SNP) rs25531 in the L-allele (16R) of the 5HTTLPR and the A → G transition results in two forms denoted L-A and L-G. Allele frequencies for these two forms and the S-allele (14R) are provided in Figure 1. Sample sizes within ethnicity and by study are provided in Supplementary Tables 1 to 6. As can be seen, the G-allele is less frequent within all the populations examined than either the L-A form or the S-allele. Between populations, Hispanics, Whites and Native Americans generally have a lower frequency of the G-allele than Asians and Blacks. This observation is consistent with the percentage G-allele identified across ethnic group among these six studies, which ranged between 0.08 (Add Health and GIC) to 0.40 (SCCS II).

Including SNP rs25531 (L-A, L-G) in the biallelic coding of the 5HTTLPR (S, L) results in six two-locus genotypes: L-A/L-A, L-A/L-G, L-A/S, L-G/L-G, L-G/S, and S/S. These six genotypes yield the triallelic 5HTTLPR genotypes denoted as S'/S' (includes: S/S, L-G/S, L-G/L-G), S'/L' (includes: L-A/S, L-A/L-G), and L'/L' (Includes L-A/L-A). Though detected, the single S-G allele observed among a Non-Hispanic White participant in Add Health is not included here. Genotype frequencies and sample sizes within ethnicity by study are provided in Tables 1 through 6 and graphically displayed in Supplemental Figure 1.

In general, the six-two locus genotypes show a complex frequency distribution that varies more between ethnic groups than within. As compared with Whites, Asians have the lowest frequency of any 'long-allele' whereas Blacks and Africans have a relatively low frequency of any 'short-allele'. A similar pattern was found among those who self-reported an Asian- or Black-Hispanic ethnicity. There were significant frequency differences between Non-Hispanic Whites and Non-Hispanic Blacks ($p < 0.001$). Interestingly, the genotypic distribution of the six two-locus genotypes in the Cameroon sample did not differ from those observed among North American Blacks ($p = 0.94$), possibly reflecting historical migration patterns.

Three other novel observations can be made from these data. First, as compared with Blacks and Africans, the absence or near absence of the L-G allele among Native Americans, Caucasians, and Hispanics appears to reflect a genetic distinction between Hispanics in the United States having a predominantly Caucasian and African ancestry and those having a predominantly Caucasian and Native American ancestry (Manichaikul et al, 2012; Bryc et al, 2010; Genovese et al, 2013). Second, in most samples using rs25531 to define the high activity L' allele, presence of the minor G allele results in a relatively small (< 10%)

decrease in frequency of the L' allele in White Non-Hispanics but a larger (> 20%) decrease in Black Non-Hispanics, suggesting that use of the G allele will have a larger impact on genetic association results in Black Non-Hispanics. Third, across all of the samples, we identified a total of 371 'XL-alleles' (rare allele size classes: 17R, 18R, 19R, 20R, 22R), the vast majority of which found as heterozygotes. Interestingly, the 'XL-alleles' were predominantly identified among Asians, Non-Hispanic Blacks and Hispanic Blacks from North America. Taken together with the absence of 'XL-alleles' among European Whites, Native-American and Non-European Hispanic, Native Americans, and the majority of those who self-reported 'Mixed' ethnicity, the presence of an 'XL-allele' may be a useful indicator of population origin in addition to other variants with allele frequency differences (Reich et al, 2012; Wang et al, 2008).

In sum, we document wide variation in the allelic and genotypic frequencies of the 5HTTLPR alleles. Specifically, we found greater numbers of 'XL-alleles' than observed in other studies as well as wide variation in the minor (G) allele of rs25531 among racial groups. The effect of these on *5HTT* transcription (Hu et al, 2006; Vilavendran et al, 2012) encourages their characterization in large, ethnically diverse samples in order to enhance the opportunity for association studies to elucidate the contribution of this polymorphism to neuropsychiatric and health related phenotypes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research uses data from the National Longitudinal Study of Adolescent to Adult Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and K.M.H. at the University of North Carolina at Chapel Hill, and funded by grant PO1 HD031921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the National Longitudinal Study of Adolescent Health data files is available on the National Longitudinal Study of Adolescent to Adult Health Web site (<http://www.cpc.unc.edu/addhealth>). This research was also supported by grants R01-HD057222 (Genes in Context Study); R01-DA020195 (Rochester Youth Development Study); R-581-000-117-101, R-581-000-099, R-581-000-090-101, R-581-000-083-101 and R-581-000-062-112 (Singapore Cardiovascular Study); P01-HD064687 (Family Transitions Project); 2P01 –HL036587 to RBW and ICS; and the University of Colorado Butcher Foundation (Cameroon Sample).

Literature Cited

- Anchordoquy HC, McGeary C, Liu L, Krauter KS, Smolen A. Genotyping of three candidate genes after whole-genome preamplification of DNA collected from buccal cells. *Behav Genet.* 2003; 33:73–78. [PubMed: 12645824]
- Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, Hammer M, Bustamante CD, Ostrer H. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci USA.* 2010; 107(S2):8954–8961. [PubMed: 20445096]
- Don RH, Cox RT, Wainwright BJ, Baker K, Mattick JS. "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* 1992; 19:4008. [PubMed: 1861999]
- Foshee VA, Benefield TS, Puvanesarajah S, Reyes HL, Haberstick BC, Smlen A, Ennett ST, Suchindran C. Self-regulatory failure and the perpetration of adolescent dating violence: examining an alcohol use by gene explanation. *Aggress Behav.* 2014 Epub ahead of print.

- Genovese G, Handsaker RE, Li H, Kenny EE, McCarroll SA. Mapping the human reference Genome's missing sequence by three-way admixture in Latino genomes. *Am. J. Hum Genet.* 2013; 93:411–421. [PubMed: 23932108]
- Gelernter J, Kranzler H, Cubells JF. Serotonin transporter protein (SLC6A4) allele and haplotype frequencies and linkage disequilibrium in African- and European-American and Japanese populations and in alcohol-dependent subjects. *Hum Genet.* 1997; 101:243–246. [PubMed: 9402979]
- Goldman N, Gleib DA, Lin YH, Weinstein M. The serotonin transporter polymorphism (5-HTTLPR): allelic variation and links with depressive symptoms. *Depression and Anxiety.* 2010; 27:260–269. [PubMed: 20196101]
- Haberstick BC, Smolen A, Stetler GL, Tabor JW, Roy T, Casey RH, Roy F, Ryals LA, Hewitt C, Whitsel EA, Halpern CT, Killea-Jones LA, Lessem JM, Hewitt JK, Harris KM. Simple sequence repeats in the National Longitudinal Study of Adolescent Health: an ethnically diverse resource for genetic analysis of health and behavior. *Behav Genet.* 2014; 44:487–497. [PubMed: 24890516]
- Harris KM, Halpern CT, Haberstick BC, Smolen A. The National Longitudinal Study of Adolescent Health (Add Health) sibling pairs data. *Twin Res Hum Genet.* 2013; 16:391–398. [PubMed: 23231780]
- Heils A, Teufel A, Petri S, Stober G, Riederer P, Bengel D, Lesch KP. Allelic Variation of the Human Serotonin Transporter Gene Expression. *J Neurochem.* 1996; 66:2621–2624. [PubMed: 8632190]
- Hu X, Lipsky RH, Zhu G, Akhtar LA, Taubman J, Greenberg BD, Xu K, Arnold PD, Richter MA, Kennedy JL, Murphy DL, Goldman D. Serotonin Transporter Promoter Gain-of-Function Genotypes Are Linked to Obsessive-Compulsive Disorder. *Am J Hum. Genet.* 2006; 78:815–826. [PubMed: 16642437]
- Kaur D, Bishop GD. Cardiovascular responses to stress in Singapore and India. *Int J Psychophysiol.* 2013; 87(2):130–140. [PubMed: 23206971]
- Lesch KP, Bengel D, Heils A, Sabol SZ, Greenberg BD, Petri S, Benjamin J, Muller CR, Hamer DH, Murphy DL. Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science.* 1996; 274:1527–1531. [PubMed: 8929413]
- Manichaikul A, Palmas W, Rodriguez CJ, Peralta CA, Divers J, Guo X, Chen WM, Wong Q, Williams K, Kerr KF, Taylor KD, Tsai MY, Goodarzi MO, Sale MM, Diez-Roux AV, Rich SS, Rotter JJ, Mychaleckyj JC. Population structure of Hispanics in the United States: The Multi-Ethnic Study of Atherosclerosis. *PloS Genetics.* 2012; 8(4):e1002640. [PubMed: 22511882]
- Murdoch JD, Speed WC, Pakstis AJ, Heffelfinger CE, Kidd KK. Worldwide population variation and haplotype analysis at the serotonin transporter gene SLC6A4 and implications for association studies. *Biol Psychiatry.* 2013; 74(12):879–889. [PubMed: 23510579]
- Nakamura M, Ueno S, Sano A, Tanabe H. The human serotonin transporter gene linked polymorphism (5HTTLPR) shows ten novel allelic variants. *Mol Psychiatry.* 2000; 5:32–38. [PubMed: 10673766]
- Odgerel Z, Talati A, Hamilton SP, Levinson DF, Weissman MM. Genotyping serotonin transporter polymorphisms 5-HTTLPR and rs25531 in European- and African-American subjects from the National Institute of Mental Health's Collaborative Center for Genomic Studies. *Transl Psychiatry.* 2013; 3:e307. [PubMed: 24064711]
- Ramamoorthy S, Bauman AL, Moore KR, Han H, Yang-Feng T, Chang AS, Ganapathy V, Blakely RD. Antidepressant- and cocaine-sensitive human serotonin transporter: molecular cloning, expression, and chromosomal localization. *PNAS.* 1993; 90:2542–2546. [PubMed: 7681602]
- Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, et al. Reconstructing Native American population history. *Nature.* 2012; 488:370–374. [PubMed: 22801491]
- SAS Institute Incorporated. SAS/GENETICS 9.1 User's Guide. Cary, NC, USA: SAS Institute; 2005.
- Thornberry TP, Henry KL. Intergenerational continuity in maltreatment. *J Abnorm Child Psychol.* 2013; 41(4):555–569. [PubMed: 23192742]
- Wendland JR, Martin BJ, Kruse MR, Lesch KP, Murphy DL. Simultaneous genotyping of four functional loci of human SLC6A4, with a reappraisal of 5-HTTLPR and rs25531. *Mol Psychiatry.* 2006; 11:224–226. [PubMed: 16402131]

- Vijayendran M, Cutrona C, Beach SRH, Brody GH, Russell D, Phillbert RA. The relationship of the serotonin transporter (SLC6A4) extra long variant to gene expression in an African American sample. *Am J Med Genet B Neuropsychiatric Genet.* 2012; 159B:611–612.
- Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, et al. Geographic patterns of genome admixture in Latin Americans Mestizos. *PLoS Genet.* 2008; 4:e1000037. [PubMed: 18369456]
- Wray NR, James MR, Gordon DS, Dumenil T, Ryan L, Coventry WL, Statham DJ, Pergadia ML, Madden PA, Heath AC, Montgomery GW, Martin NG. Accurate, large-scale genotyping of 5HTTLPR and flanking single nucleotide polymorphisms in an association study of depression, anxiety, and personality measures. *Biol Psychiatry.* 2009; 66:468–476. [PubMed: 19541292]

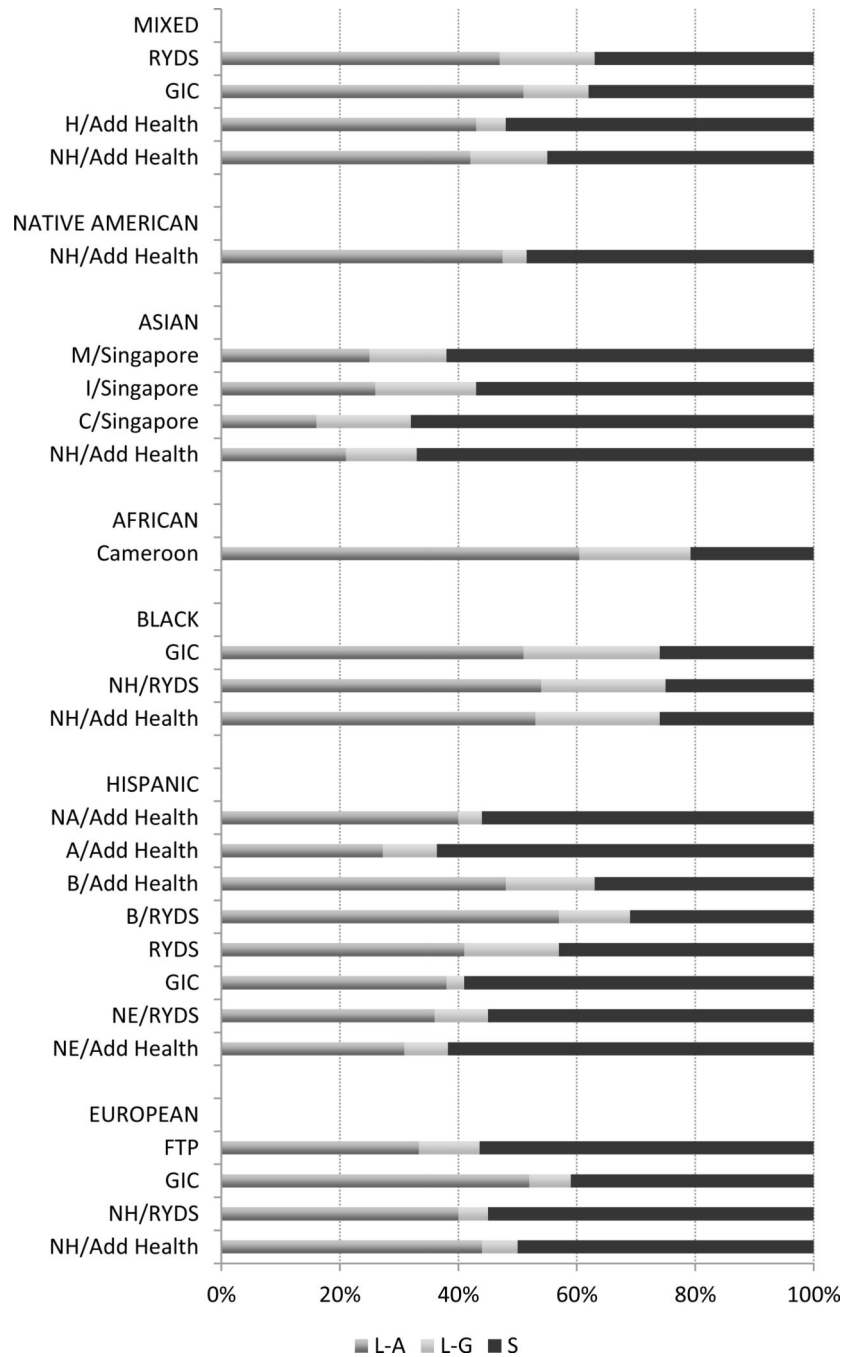


FIGURE 1. Allele frequencies for SNP rs25531 as a function of ethnicity across six samples (N = 23,615)
 Note: H, Hispanic; NH, Non-Hispanic; NA, Native American; NE, Non-White; B, Black; A, Asian; M, Malay; I, Indian; C, Chinese; GIC, Genes In Context; RYDS, Rochester Youth and Development Study; Add Health, National Longitudinal Study of Adolescent Health; FTP, Family Transitions Project; MIXED was a self-report grouping choice without

clarification for participants in the Add Health, RYDS, and GIC samples. Sample sizes and allele frequencies within ethnicity by study are provided in Supplemental Tables 1 – 6.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1
Genotype Frequencies for triallelic coding of 5HTTLPR and rs25531 by race/ethnicity in the National Longitudinal Study of Adolescent to Adult Health (Add Health).

Race/Ethnicity	N	Genotype (N, %)							
		L-A/L-A	L-A/L-G	L-A/S	L-G/L-G	L-G/S	S/S		
White, Non-Hispanic	7878	1982 (0.25)	537 (0.07)	3379 (0.43)	38 (<0.01)	483 (0.06)	1449 (0.18)		
White, Hispanic	981	207 (0.21)	52 (0.05)	411 (0.42)	--	58 (0.06)	251 (0.26)		
Black, Non-Hispanic	3290	882 (0.27)	769 (0.23)	867 (0.26)	128 (0.04)	326 (0.10)	261 (0.08)		
Black, Hispanic	90	19 (0.21)	15 (0.17)	31 (0.34)	2 (0.02)	7 (0.08)	14 (0.16)		
Asian, Non-Hispanic	911	38 (0.04)	41 (0.05)	231 (0.25)	15 (0.02)	149 (0.16)	405 (0.44)		
Asian, Hispanic	86	41 (0.08)	3 (0.06)	14 (0.26)	--	7 (0.13)	21 (0.40)		
Native American, Non-Hispanic	273	61 (0.22)	10 (0.04)	127 (0.47)	--	12 (0.04)	62 (0.23)		
Native American, Hispanic	126	24 (0.19)	5 (0.04)	49 (0.39)	--	5 (0.04)	43 (0.34)		
Mixed/Other, Non-Hispanic	128	26 (0.20)	18 (0.14)	37 (0.29)	--	14 (0.11)	32 (0.25)		
Mixed/Other, Hispanic	1074	194 (0.18)	47 (0.04)	481 (0.45)	5 (<0.01)	60 (0.06)	285 (0.27)		

Note: Not tabled are genotypes containing 103 *XL-alleles[†]; 57 (0.02) in Black, Non-Hispanic; 32 (0.04) in Asian, Non-Hispanic; 9 (<0.01) European, Non-Hispanic; 2 (0.02) in Black, Hispanic; and 1 each among European-Hispanic, Asian-Hispanic, and Native American-Non-Hispanic.

HWE: White, Non-Hispanic: $\chi^2 = 0.1697$, $df = 2$, $p = 0.9209$; White, Hispanic: $\chi^2 = 2.0234$, $df = 2$, $p = 0.3636$; Black, Non-Hispanic: $\chi^2 = 0.7165$, $df = 2$, $p = 0.6989$; Black, Hispanic: $\chi^2 = 0.1826$, $df = 2$, $p = 0.9127$; Asian, Non-Hispanic: $\chi^2 = 0.6745$, $df = 2$, $p = 0.7137$; Asian, Hispanic: $\chi^2 = 32.1833$, $df = 2$, $p < 0.0001$; Native American, Non-Hispanic: $\chi^2 = 0.0368$, $df = 2$, $p = 0.9817$; Native American, Hispanic: $\chi^2 = 1.8312$, $df = 2$, $p = 0.4002$; Mixed/Other, Non-Hispanic: $\chi^2 = 1.3909$, $df = 2$, $p = 0.4988$; Mixed/Other, Hispanic: $\chi^2 = 0.0489$, $df = 2$, $p = 0.9758$.

Genotype Frequencies for triallelic coding of 5HTTLPR and rs25531 by race/ethnicity in the **Genes in Context Study (GIC)**.

Table 2

Race/Ethnicity	N	Genotypes (N, %)						
		L-A/L-A	L-A/L-G	L-A/S	L-G/L-G	L-G/S	S/S	
White	641	158 (0.25)	39 (0.06)	272 (0.42)	6 (0.01)	41 (0.06)	125 (0.20)	
Black	621	160 (0.26)	170 (0.27)	148 (0.24)	25 (0.04)	67 (0.11)	51 (0.08)	
Hispanic	32	3 (0.09)	1 (0.03)	17 (0.53)	--	1 (0.03)	10 (0.31)	
Mixed	74	19 (0.26)	10 (0.14)	27 (0.34)	--	6 (0.08)	12 (0.16)	

HWE: White, $\chi^2 = 0.5662$, $df = 2$, $p = 0.7534$; Black, $\chi^2 = 0.4129$, $df = 2$, $p = 0.8134$; Hispanic, $\chi^2 = 1.7077$, $df = 2$, $p = 0.4258$; Mixed, $\chi^2 = 0.0278$, $df = 2$, $p = 0.9862$.

Genotype Frequencies for triallelic coding of 5HTTLPR and rs25531 by ethnicity in the **Rochester Youth and Development Study (RYDS)**.

Table 3

Ethnicity	N	Genotype (N, %)						
		L-A/L-A	L-A/L-G	L-A/S	L-G/L-G	L-G/S	S/S	
White, Non-Hispanic	85	16 (0.19)	3 (0.03)	34 (0.40)	1 (0.01)	3 (0.04)	28 (0.33)	
White, Hispanic	11	1 (0.09)	1 (0.09)	5 (0.45)	--	1 (0.09)	3 (0.27)	
Black, Non-Hispanic	509	140 (0.28)	98 (0.19)	131 (0.26)	32 (0.06)	52 (0.10)	32 (0.06)	
Black, Hispanic	38	10 (0.26)	6 (0.16)	14 (0.37)	--	3 (0.08)	3 (0.08)	
Hispanic	58	11 (0.19)	6 (0.10)	20 (0.34)	2 (0.03)	8 (0.14)	11 (0.19)	
Mixed	83	16 (0.19)	15 (0.18)	28 (0.34)	1 (0.01)	8 (0.10)	13 (0.16)	

Note: Not tabled are containing 27 'XL'-alleles; 23 (0.05) in Black, Non-Hispanic, 2 (0.05) in Black, Hispanic, and 2 (0.02) among those endorsing Mixed ethnicity.

HWE: White, Non-Hispanic, $\chi^2 = 0.8093$, $df = 2$, $p = 0.6672$; White, Hispanic, $\chi^2 = 0.2000$, $df = 2$, $p = 0.9048$; Black, Non-Hispanic, $\chi^2 = 1.5163$, $df = 2$, $p = 0.4685$; Black, Hispanic, $\chi^2 = 0.4560$, $df = 2$, $p = 0.7961$; Hispanic, $\chi^2 = 1.998$, $df = 2$, $p = 0.3681$; Mixed, $\chi^2 = 0.2023$, $df = 2$, $p = 0.9037$.

Genotype Frequencies for triallelic coding of 5HTTLPR and rs25531 by ethnicity in the **Singapore Sample**.

Table 4

Ethnicity	N	Genotype (N, %)						
		L-A/L-A	L-A/L-G	L-A/S	L-G/L-G	L-G/S	S/S	
Chinese	2674	80 (0.03)	107 (0.04)	596 (0.22)	75 (0.03)	581 (0.20)	1235 (0.43)	
Malay	831	51 (0.06)	50 (0.06)	264 (0.30)	20 (0.02)	133 (0.15)	313 (0.35)	
Asian Indian	717	55 (0.08)	43 (0.06)	222 (0.31)	28 (0.04)	138 (0.19)	227 (0.31)	

Note: Not tabled are genotypes containing 233 -XL-alleles: 174 (0.06) in Chinese, 55 (0.06) in Malay and 4 (0.01) in Indian.

HWE: Chinese, $\chi^2 = 9.4248$, $df = 2$, $p = 0.0089$; Malay, $\chi^2 = 0.0228$, $df = 2$, $p = 0.9886$; Asian Indian, $\chi^2 = 1.2661$, $df = 2$, $p = 0.5309$.

Genotype Frequencies for triallelic coding of 5HTTLPR and rs25531 by ethnicity in the Cameroon Language and Genetics Study (CLGS).

Table 5

		Genotype (N, %)					
	N	L-A/L-A	L-A/L-G	L-A/S	L-G/L-G	L-G/S	S/S
Cameroon	529	194 (0.37)	111 (0.18)	128 (0.24)	27 (0.05)	31 (0.06)	30 (0.06)

Note: Not tabled are genotypes containing 8 (0.02) 'XL-alleles' observed in the Cameroon sample.
HWE: $\chi^2 = 0.9175$, $df = 2$, $p = 0.6320$.

Genotype Frequencies for triallelic coding of 5HTTLPR and rs25531 by ethnicity in the **Family Transitions Project**.

Table 6

	Genotype (N, %)						
	N	L-A/L-A	L-A/L-G	L-A/S	L-G/L-G	L-G/S	S/S
White, Non-Hispanic	2379	622 (0.26)	185 (0.08)	1031 (0.44)	7 (< 0.01)	140 (0.06)	394 (0.17)

HWE: $\chi^2 = 1.5794$, $df = 2$, $p = 0.4539$.