2013

Add Health Wave IV Documentation

Candidate Genes

Report prepared by
Andrew Smolen
Eric A. Whitsel
Joyce Tabor
Ley A. Killeya-Jones
Carmen C. Cuthbertson
Jon M. Hussey
Carolyn T. Halpern
Kathleen Mullan Harris

Add Health is supported by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Further information may be obtained by contacting Add Health at addhealth@unc.edu.

Carolina Population Center
University of North Carolina at Chapel Hill
123 W. Franklin Street, Room 403-C
Chapel Hill, NC 27516-2524

Add Health

https://doi.org/10.17615/C6ND4B

Table of Contents

1. Introduction

During Wave IV, Add Health collected biological specimens from a large, nationally representative sample of young adults. Given the size of the Wave IV sample, its geographic distribution, and in-home setting of the respondent interviews, biological specimen collection involved practical, relatively non-invasive, cost-efficient and innovative methods. These methods included collection of saliva by trained and certified field interviewers, salivary buccal cell lysis and DNA stabilization in the field, then shipment to a central lab for DNA extraction, genotyping, and archiving. The collection of saliva followed the interview and collection of cardiovascular and anthropometric measures (Entzel et al. 2009). It preceded the collection of capillary whole blood (Whitsel et al. 2012) and data on respondent use of prescription and select over-the-counter medications (Tabor et al. 2010). Further details on the design of Add Health Waves I-IV are available elsewhere (Harris 2012; Harris et al. in press).

Included in the updated Add Health Wave IV candidate gene data release are the following:

Variable number tandem repeat (VNTR) and single tandem repeat (STR) polymorphisms

- Monoamine Oxidase A Upstream VNTR (MAOA_V)
- Dopamine D4 Receptor Exon 3 VNTR (DRD4)
- Dopamine Transporter *SLC6A3* 3'-Untranslated Region VNTR (DAT1)
- Serotonin Transporter-Linked Polymorphic Region (HTTLPR)
- "Triallelic" HTTLPR alleles (LALGS; TRI)
- Dopamine D5 receptor dinucleotide repeat in the 5' region (DRD5)
- Monoamine Oxidase A dinucleotide repeat (MAOCA1)

Single nucleotide polymorphisms (SNPs)

- Dopamine D2 Receptor TaqIA SNP rs1800497 in the 3' UTR (DRD2)
- Catechol O-Methyltransferase (COMT) val158met SNP rs4680 (S000005)
- Serotonin Transporter (5HTT, Locus Symbol *SLC6A4*) rs12945042 (S000006)

This document summarizes the rationale, equipment, protocol, genotyping, data cleaning, and quality for each measure listed above. Documentation of other (genetic; inflammatory; immune) measures based on additional genotyping of salivary buccal cell DNA and assay of dried capillary whole blood spots will be provided in separate reports.

2. General Overview of Data Collection

A Blaise computer-assisted interview (CAI) program guided trained and certified field interviewers (FIs) through the saliva collection process. Help screens with step-by-step measurement instructions

were accessible within the program. Each FI also carried a Job Aids Booklet that served as a quick hard-copy reference guide to study protocols. Respondents were free to decline any or all measurements while participating in the interview, including saliva collection. Consent information will be provided in a separate data file.

Some measurement protocols were revised in the period between the Wave IV Pretest (conducted in 2007) and the Main Study (conducted in 2008). Where the Pretest and Main Study data collection protocols differed significantly, this report documents the key differences between them. Pretest cases in the Wave IV data set are flagged for identification.

3. Saliva Collection

3.1. Rationale

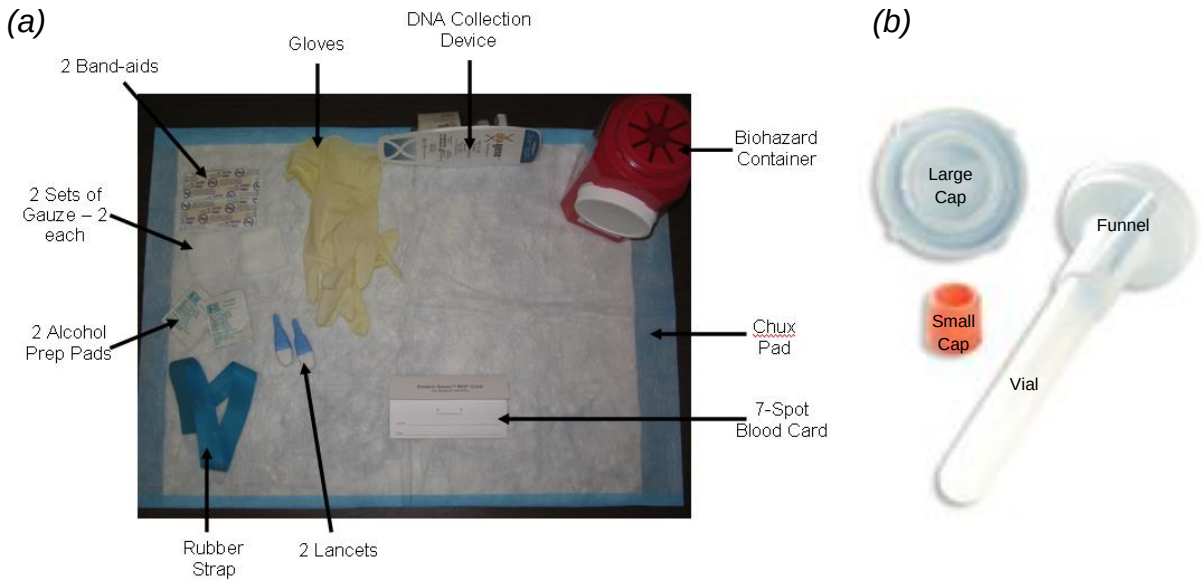
The selection of biological measures for Wave IV was driven by scientific knowledge of the leading causes of health problems at this and future developmental stages of the Add Health cohort (aged 24-32 at Wave IV), the role of specific biological processes in causation, and the ability of specific measures to characterize these processes (Harris 2012). Within these scientific criteria, choice was constrained by the feasibility, validity, and reliability of methods used to obtain biological specimens in a large non-clinical field setting and assay techniques and measurement strategies used to capture biological phenomena of interest in this life stage.

Genes play an important role in many health and behavioral trajectories. The influence of genetic factors on health is conditioned by the social, cultural, institutional, and physical environments in which individuals live, work, and play. The multi-level, multi-dimensional, and longitudinal design of Add Health, with its wealth of social, behavioral, and biological data across the early life course, makes it well-suited to study gene by environment interactions. We therefore collected saliva from the entire national sample to enable buccal cell DNA extraction, genotyping, and archiving.

3.2. Equipment

Equipment included gloves, Chux-type absorbent underpad, and a salivary DNA collection device (Oragene™, DNAgenotek, Ottawa, Ontario, Canada) including a funnel, 2 ml vial (OG-300), large cap containing a stabilizing lysis buffer, and a small cap used in shipment (*Exhibit 1a-b*).

Exhibit 1. (a) Workspace and (b) salivary DNA collection device.



3.3. Protocol

3.3.1. Main Study

3.3.1.1. Consent

Informed consent to participate in Add Health Wave IV – including consent for saliva collection – was obtained before respondent interview data collection. At that time, respondents agreed to provide either (a) saliva for currently planned genotyping and DNA archival for future genotyping, (b) saliva for currently planned genotyping, but not DNA archival for future genotyping, or (c) no saliva. Consent to saliva collection for currently planned genotyping and DNA archival are tabulated in *Exhibit 2*.

Exhibit 2. Consent by race/ethnicity

	Consent, n (%)*	
Race/Ethnicity	**To Collection**	**To Archive**
Hispanic	2,393 (96%)	1,883 (75%)
Black, Non-Hispanic	3,348 (96%)	2,523 (72%)
White, Non-Hispanic	8,051 (97%)	6,822 (82%)
Other, Non-Hispanic†	1,335 (95%)	997 (71%)
All	15,140 (96%)‡	12,234 (78%)‡

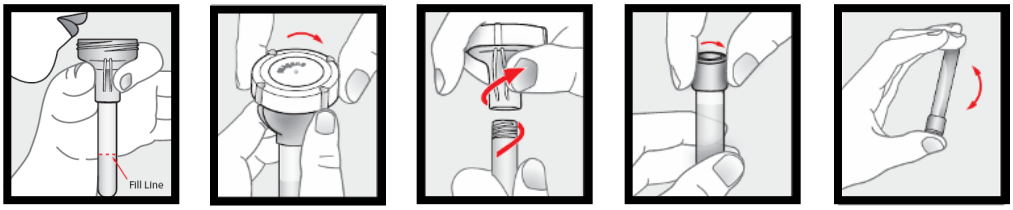
*Unweighted number and percentage, excluding four incarcerated respondents for whom saliva collection was prohibited. †Asian, American Indian, or other. ‡13 respondents with missing Wave I race are included in the Consent to Collection totals; 9 respondents missing Wave I race are included in the Consent to Archive totals.

Respondents initially refusing consent to saliva collection before respondent interview data collection were asked to confirm their refusal before beginning any biological data collection. Respondents subsequently deciding to participate in a previously refused saliva collection changed their response in the check box and initialed the change on the consent form. Likewise, respondents initially consenting to, but subsequently refusing saliva collection changed and initialed the change on the consent form.

3.3.1.2. Saliva Collection

After confirming that respondents consented to saliva collection for currently planned genotyping, FIs identified latex-allergic respondents and used latex-free gloves when working with them. FIs told respondents that they would collect a small amount of saliva for genotyping. To do so, FIs gave respondents the funneled vial (*Exhibit 1b*), then directed them to spit into it until saliva reached the 2 ml fill line (*Exhibit 3*). FIs asked respondents not to expectorate in an attempt to generate saliva more quickly. Once full, respondents returned the funneled, saliva-containing vial to FIs. FIs attached the large cap (*Exhibit 1b*) to the funnel by turning it clockwise. On attachment, ~2 ml of stabilizing lysis buffer was released into the saliva. FIs removed the capped funnel from the vial by twisting it counter-clockwise. FIs tightly attached the small cap (*Exhibit 1b*) to the vial. FIs mixed the buffer and saliva by inverting the recapped vial five times.

Exhibit 3. Saliva collection procedure



3.3.1.3. Saliva Shipment

FIs bar code labeled the recapped vial with the corresponding respondent biospecimen ID. Thereafter, FIs packaged the vial in an approved container (SafTpak #STP-210, Edmonton, Alberta, Canada) and shipped it in a FedEx Priority Overnight envelope to the University of Colorado, Institute for Behavioral Genetics (IBG, Andy Smolen, Ph.D., Director, Genotyping Laboratory, Boulder, CO) for DNA extraction, genotyping, and depending on consent, archiving.

IBG received the FedEx Priority Overnight envelopes containing single vials. They scanned the FedEx tracking number and bar code-labeled vial into a database to confirm receipt, then unpackaged and examined the vials for damage / leakage. They also keyed the receipt date, combined buffer / saliva volume (ml), and condition of the vial alongside the biospecimen ID. They sorted the vials by archive/no archive status and stored them at 15-30°C, a temperature associated with DNA stability in

the lysis buffer for at least five years (Iwasiow et al. 2011; Rylander-Rudqvist et al. 2006). Nine (0.06%) vials were empty and 24 (0.15%) were damaged or leaking on arrival. The mean (standard deviation) and range of the combined buffer / saliva volume was 3.72 (0.63) and 0 to 6.0 ml.

3.3.2. Pretest Variations

During the Pretest, urine collection preceded saliva collection. No urine was collected in the Main Study. Further, consent to saliva collection and, separately, DNA archiving, was obtained immediately before saliva collection in the Pretest.

4. DNA Assay Procedures

4.1. DNA Extraction and Yield

Genomic DNA was isolated from buffer / saliva samples using ZymoResearch (Irvine, CA) Silicon-A™ plates according to the manufacturer's protocol. 500 µl of Oragene™ solution was used, and the final elution volume was 150 µl. Samples were diluted 1:20 in Tris-EDTA, pH 8.0, and the DNA was quantified using Picogreen® (Invitrogen, Carlesbad, CA) fluorescence. The mean (standard deviation) and range of the isolated DNA yield was 33 (25) and 0 to 400 µg.

4.2. VNTR and STR Genotyping

The methods for genotyping the five VNTR and STR loci are summarized below. All primer sequences are written 5' to 3'.

4.2.1. Monoamine Oxidase A Upstream VNTR

Variable name: **MAOA_V**

The MAOA gene, which maps to Xp11.3-11.4, contains a 30 bp VNTR in the 5' regulatory region of the gene (Samochowiec et al. 1999) which may consist of two to five repeats. This VNTR has been shown to affect the expression, and likely the activity of MAOA (Sabol et al. 1998). In vitro experimental evidence indicates that MAOA alleles can be pooled into two groups based on their effects on transcriptional efficiency (Denney et al. 1999; Sabol et al. 1998). The first group (low expressing) consists of the 2R and 3R alleles. The second group (high expressing) combines the 3.5R, 4R and 5R alleles, although some investigators consider the 5R allele "low expressing" and combine it with the 2R and 3R alleles. The MAOA_V polymorphism was assayed by a modification (Haberstick et al. 2005) of a published method (Sabol et al. 1998). The primer sequences were forward: 6FAM-ACA GCC TGA CCG TGG AGA AG; and reverse: GAA CGG ACG CTC CAT TCG GA and yield products of 291 (2R), 306 (2.5R), 321 (3R), 336 (3.5R), 351 (4R), and 381 (5R) bp.

4.2.2. Dopamine D4 Receptor Exon 3 VNTR

Variable name: **DRD4**

The dopamine D4 receptor gene, which maps to 11p15.5, contains a 48 bp VNTR polymorphism in the third exon (Van Tol et al. 1992), which results in ten allelic products comprised of 2-11 repeat units. The most common variants consist of two, four, and seven repeats (Van Tol et al. 1992). This VNTR, which codes for the proline-rich third cytoplasmic loop of the receptor protein has been shown to affect the function of the D4 receptor in vivo: the large variants (7R and larger) appear to blunt the intracellular response to dopamine in vitro as compared with the D4.2 and D4.4 variants (Asghari et al. 1995). The assay (Anchordoquy et al. 2003) was a modification of an extant method (Lerman et al. 1998). The primer sequences were forward: VIC-GCT CAT GCT GCT GCT CTA CTG GGC; and reverse: CTG CGG GTC TGC GGT GGA GTC TGG; and yield products of 279 (2R), 327 (3R), 375 (4R), 423 (5R), 471 (6R), 519 (7R), 567 (8R), 615 (9R), 663 (10R) and 711 (11R) bp. A rare allele of 366 bp was also found. We report this as 3.39R to indicate that it is composed of a 3R allele plus 39 bp (9 short of a usual 48 bp repeat). We have seen this allele, which to our knowledge has not been reported previously, in other data sets. This allele was not reported previously in the Wave III genotyping data release.

4.2.3. Dopamine Transporter (*SLC6A3*) 3'-Untranslated Region VNTR

Variable name: **DAT1**

The dopamine transporter locus, which maps to 5p15.3, contains a 40 bp VNTR in the 3' untranslated region (UTR) of the gene. This element was originally reported as containing 3 to 11 copies and has been shown to be associated with expression of the DAT protein in lymphocytes (Vandenbergh et al. 1992) and in human striatum (Heinz et al. 2000). The assay (Anchordoquy et al. 2003) was a modification of the method of Vandenbergh et al. (1992). The primer sequences were forward: 6FAM-TGT GGT GTA GGG AAC GGC CTG AG; and reverse: CTT CCT GGA GGT CAC GGC TCA and yield products of 200 (3R), 360 (7R), 400 (8R), 440 (9R), 480 (10R), 520 (11R) and 600 (13R) bp. The 4R, 5R and 6R alleles were not seen. The 3R allele was not reported previously in the Wave III genotyping data release.

4.2.4. Serotonin Transporter-Linked Polymorphic Region

Variable names: **HTTLPR**; **LALGS**; **TRI**

The serotonin Transporter (*SLC6A4*), which maps to 17q11.1-17q12 (Ramamoorthy et al. 1993), contains a 43 bp insertion / deletion (in/del, HTTLPR) polymorphism (not 44 as originally reported) in the 5' regulatory region of the gene (Heils et al. 1996). The in/del in the promoter appears to be associated with variations in transcriptional activity: the long variant (L) has approximately three times the expression of the short promoter (S) with the deletion (Lesch et al. 1996), although this is not a universal finding (Kaiser et al. 2002; Willeit et al. 2001). Several investigators have reported that the

HTTLPR polymorphism affects serotonergic functions in vivo. Individuals with the L/L genotype were found to have significantly higher maximal uptake of serotonin into platelets compared to those with L/S or S/S genotypes (Nobile et al. 1999). The assay for HTTLPR was a modification (Anchordoquy et al. 2003) of the method of Lesch et al. (1996) using the primer sequences from Gelernter et al. (1999) forward: NED-ATG CCA GCA CCT AAC CCC TAA TGT; and reverse: GGA CCG CAA GGT GGG CGG GA, which yield products of 376 (S) or 419 (L) for the two most common alleles. Additional extra-long alleles are found rarely as detailed by Nakamura et al. (2000). According to their nomenclature, the most common S and L alleles contain 14 or 16 repeat units, respectively. Extra-long alleles contain 18, 19, 20 and 22 repeat units. In the Add Health database we report them as such: 14R (Short) 16R (most common Long allele), and extra-long 17R (440 bp), 18R (461 bp), 19R (483 bp), 20R (505 bp) and 22R (549 bp) alleles. The 18R allele was not reported by Nakamura et al. (2000). In absence of any additional information the extra-long alleles should probably be considered "long-like" in function, although there are no direct experimental data to support this. These data (14R to 22R) are reported in the released Wave IV genetic data as the "Biallelic" HTTLPR under the variable name **HTTLPR**. The extra-long alleles were not reported previously in the Wave III genotyping data release. Hu et al. (2005) reported that a SNP (rs25531, A/G) in the Long form of HTTLPR may have functional significance: the more common LA allele is associated with the reported higher basal activity, whereas the less common LG allele has transcriptional activity no greater than the S. These investigators suggest that in tests of association the LG alleles should be analyzed along with the S alleles (Hu et al. 2006).

The SNP rs25531, which allows determination of the LA and LG alleles, was assayed (Wendland et al. 2006) by incubating the polymerase chain reaction (PCR) products with 5 units of MspI (New England Biolabs, Ipswitch, MA) for 90 min at 37ºC. A 152 bp restriction digest fragment is indicative of the LG allele. A second, constant restriction site near the 5' end of the amplicons provides a positive control for each restriction reaction by cleaving all S and L alleles to yield fragments of 283 and 326 bp, respectively. An independent verification of the genotypes, was assayed in the same system with the substitution of the primer sequences from Hu et al. (2005) forward: 6FAM-GCA ACC TCC CAG CAA CTC CCT GTA; and reverse: GAG GTG CAG GGG GAT GCT GGA A; which yield PCR products of 138 (S) and 181 (L) bp. A 97 bp MspI restriction digest fragment is indicative of the LG allele. None of the extra-long alleles were found to have the G form of rs25531, which is in agreement with the sequences reported by Nakamura et al. (2000).

In this Add Health data release, in addition to the "biallelic" HTTLPR alleles (reported as repeat units: 14R, 16R, etc.), we report the "Triallelic" HTTLPR alleles, which take into account the SNP rs25531, in two ways related to two additional variables: (1) the three alleles L-A, L-G, S, and S-G are listed under variable name LALGS; and (2) combined "alleles" S' (consisting of S, S-G, and L-G) and L' (consisting of L-A and extra-long alleles) are listed under variable name TRI. These latter two categories were not reported previously in the Wave III genotyping data release.

4.2.5. Dopamine D5 Receptor Dinucleotide Repeat in the 5' Region

Variable name: **DRD5**

The DRD5 gene has been mapped to Chromosome 4 (4p16.1) and contains a dinucleotide repeat in the 5' untranslated region of the gene (Sherrington et al. 1993) which contains numerous alleles (24 in this data set). We used a minor modification of the Sherrington primers which results in amplicons that are 2 bp smaller than the reference sizes. The Add Health results have been adjusted so that the most common allele (allele 9 of Sherrington et al. 1993) is reported at 148 bp to conform to previous literature (e.g., Vanyukov et al. 2000). The primer sequences were: Forward, 6FAM-TGTATGATCCCTGCA GCATATT, and Reverse, GCTCATGAGAAGAATGGAGTG and produced amplicons of 124-174 bp.

4.2.6. Monoamine Oxidase A Dinucleotide Repeat

Variable name**: MAOCA1**

The dinucleotide repeat in the 3'UTR region (sWXD805) of the MAOA gene was genotyped using modifications of the standard primer sequences (Black et al. 1991). In the course of our studies with non-Add Health, family based samples we found a high frequency of Mendelian errors with the standard MAOCA-1 primers, which we traced to the presence of a SNP, rs5906974 located 3 bp from the 3'-end of the reverse primer. The SNP rs5906974 has a minor allele frequency (MAF) of 0.227 in individuals of European ancestry and of 0.5 in African Americans. Individuals carrying the minor allele, when genotyped with the standard MAOCA-1 primer set often appeared to be falsely homozyogous due to failure of the reverse primer to bind to its target site. The primer was modified so that the site of the SNP was avoided. A second modification was the addition of a GTTTCTT-sequence to the 5'-end of the reverse primer (Brownstein et al. 1996) which yields a more uniform product by promoting more complete addition of a non-templated A-residue by Taq polymerase. The primer sequences were: Forward, 5'-VIC-AGAGACTAGACAAGTTGCACTG-3', and Reverse: 5'-GTTTCTTGATTCTTCACTATC TTGTTA-3' and produced amplicons of 101 to 131 bp.

4.3. PCR Amplification and Analysis

All PCR reactions contained two µl of DNA (20 ng or less), 1.8 mM MgCl2, 180 µM deoxynucleotides, with 7'-deaza-2'-deoxyGTP (Roche Applied Science, Indianapolis, IN) substituted for one-half of the dGTP, forward (fluorescently labeled) and reverse primers (concentrations in *Exhibit 4)* and 1 unit of AmpliTaq Gold® polymerase (ABI, Foster City, CA), in a total volume of 20 µl. Amplifications were performed using a modified (Anchordoquy et al. 2003) touchdown PCR method (Don et al. 1991). A 95°C incubation for 10 min was followed by two cycles of 95°C for 30 s, 65°C for 30 s, and 72°C for 60 s. 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 s, 55°C for 30 s, and 72°C for 60 s and a final 30-min incubation at 72°C.

After amplification, an aliquot of PCR product was combined with loading buffer containing size standard (Rox1000, Gel Company, San Francisco, CA or LIZ1200, ABI Foster City, CA) and analyzed with an ABI PRISM® 3130xl Genetic Analyzer using protocols supplied by the company. Data were analyzed with Genemapper software and the resulting allele sizes were reviewed by two investigators independently. All PCR analyses were repeated twice from scratch, generally on different days and generally by a different investigator. For the Archive samples the reported genotypes are the result of one run with genomic DNA and one with Whole-genome amplified (WGA) DNA. For the non-archive samples two runs with genomic DNA were used. The results of the two runs were compared by a third investigator. In the event of missing or discrepant results, a third run was used to determine the reported genotype.

Exhibit 4. Primers used in PCR reactions

Primer*	Sequence (5'→3')	Concentration (nM)	Size range (bp)
MAOA_V-F	6FAM-ACA GCC TGA CCG TGG AGA AG	200	291-381
MAOA_V-R	GAA CGG ACG CTC CAT TCG GA	200	(2R-5R)
DRD4-F	VIC-GCT CAT GCT GCT GCT CTA CTG GGC	600	279-711
DRD4-R	CTG CGG GTC TGC GGT GGA GTC TGG	600	(2R-11R)
DAT1-F	6FAM-TGT GGT GTA GGG AAC GGC CTG AG	300	200-600
DAT1-R	CTT CCT GGA GGT CAC GGC TCA AGG	300	(3R-13R)
HTTLPR-F	NED-ATG CCA GCA CCT AAC CCC TAA TGT	600	376, 419-549
HTTLPR-R	GGA CCG CAA GGT GGG CGG GA	600	(S, L-xL)
HTTLPR-Hu-F	6FAM-GCA ACC TCC CAG CAA CTC CCT GTA	500	138, 181
HTTLPR-Hu-R	GAG GTG CAG GGG GAT GCT GGA A	500	(S, L)
MAOCA_1_F	VIC-AGAGACTAGACAAGTTGCACTG	370	101-131
MAOCA_1_R	GTTTCTTGATTCTTCACTATC TTGTTA	370	
DRD5_F	6FAM-TGTATGATCCCTGCA GCATATT	250	124-174
DRD5_R	GCTCATGAGAAGAATGGAGTG	250	

*F = forward. R = reverse.

4.4. SNP Genotyping

4.4.1. Dopamine D2 Receptor TaqIA SNP rs1800497 in the 3' UTR

Variable name: **DRD2**

SNP assays were done on either an Applied Biosystems TaqMan® OpenArray® (archived samples) or Illumina BeadXpress® GoldenGate® (non-archived samples) platform. SNP assays were ordered by rs number and genotyping was performed according to company-supplied protocols (Applied Biosystems, Inc.; Illumina, Inc.).

Samples without results from the OpenArray or GoldenGate platforms were re-assayed using TaqMan® assays. The DRD2 TaqIA assay (rs1800497) was done using a fluorogenic 5'nuclease (Taqman®, ABI, Foster City, CA) method (Haberstick and Smolen 2004) on an ABI Prism® 7000 Sequence Detection System using the allelic discrimination mode (Livak 1999). Primer and probe sequences were:

Forward Primer: 5'-GTGCAGCTCACTCCATCCT-3';
Reverse Primer: 5'-GCAACACAGCCATCCTCAAAG-3';
A1 Probe: 5'- VIC-CCTGCCTTGACCAGC-NFQMGB-3';
A2 Probe: 5'- FAM-CTGCCTCGACCAGC-NFQMGB-3'.

Reactions containing 20 ng or less of DNA were performed in 10 µl reactions with TaqMan® Universal PCR Master Mix using the standard cycling conditions. Final primer and probe concentrations were 900 mM and 200 mM, respectively. Each 96 well plate included at least one non-template and three known genotype controls. Genotypes were scored independently by two individuals.

4.4.2. Catechol O-Methyltransferase (COMT) val158met SNP rs4680 (RS4680)

Variable name: **S000005**

SNP assays were done on either an Applied Biosystems TaqMan® OpenArray® (archived samples) or Illumina BeadXpress® GoldenGate® (non-archived samples) platform. SNP assays were ordered by rs number and genotyping was performed according to company-supplied protocols (Applied Biosystems, Inc.; Illumina, Inc.).

Samples without results from the OpenArray or GoldenGate platforms were re-assayed using TaqMan® assays. The COMT val^{158}met assay (rs4680) was performed using a fluorogenic 5'nuclease (Taqman®, Applied Biosystems, Foster City, CA) method (Haberstick and Smolen 2004). Forward and reverse primers, and allele specific probes were kindly provided by Dr. Daniel Weinberger (Mattay et al. 2003; and personal communication):

Forward Primer: 5'-TCGAGATCAACCCCGACTGT-3';
Reverse Primer: 5'-AACGGG-TCAGGCATGCA-3';
Val Probe: 5'-FAM-CCTTGTCCTTCA*C*GCCAGCGA- NFQMGB-3';
Met Probe: 5'-VIC-ACCTTGTCCTTCA*T*GCCAGCGAAAT- NFQMGB-3'.

Reactions were performed in an ABI Prism® 7000 Sequence Detection System using the allelic discrimination mode (Livak 1999). Reactions containing 20 ng of DNA were performed in 15 µl reactions with TaqMan® Universal PCR Master Mix using the standard cycling conditions. Final primer and probe concentrations were 900 mM and 200 mM, respectively. Each 96 well plate included at least one non-template and two known genotype controls. Genotypes were scored independently by two individuals.

4.4.3. Serotonin Transporter (5HTT, Locus Symbol *SLC6A4*) rs12945042 (RS12945042)

Variable name: **S000006**

All SNP assays were done on either an Applied Biosystems TaqMan® OpenArray® (archived samples) or Illumina BeadXpress® GoldenGate® (non-archived samples) platform. SNP assays were ordered by

rs number and genotyping was performed according to company-supplied protocols (Applied Biosystems, Inc.; Illumina, Inc.).

5. Data Cleaning

5.1. Wave III – Wave IV Comparison

Candidate gene and DRD2 SNP data from Wave III and Wave IV were compared in respondents with data from both waves. When discrepant, both the Wave III and IV candidate gene samples were re-genotyped and DRD2 SNP data were reviewed to help determine reported genotype. If a Wave IV result was missing, but a Wave III result was available, the Wave III result was included in the Wave IV data file and identified by an allele-specific flag variable. When Wave III and Wave IV values do not agree, the Wave IV value is correct. This was confirmed by re-assaying or reviewing both the Wave III and Wave IV samples. We therefore recommend that the Wave IV results be used for analysis when the same genetic measure is available at both Wave III and Wave IV

5.2. Recoding of Variable HTTLPR

The variable LALGS was used to assign variable HTTLPR values for respondents who did not have HTTLPR results. The following logic was used to create the HTTLPR values: the LALGS value of S corresponds to HTTLPR value of 14R; LALGS values of L-A and L-G correspond to HTTLPR values of 16R; and extra-long values are the same for both LALGS and HTTLPR alleles. Consistency between the variables HTTLPR and LALGS was checked using the same logic previously described. For LALGS and TRI to be consistent the following applies: LALGS values of L-A and extra-long correspond to TRI values of L'; LALGS values of L-G, S, and S-G correspond to TRI values of S'. When LALGS values were not consistent with HTTLPR values both were genotyped again.

5.3. Coding of allele B for MAOA

MAOA is a sex-linked allele and males have values for allele A but are coded to zero for allele B. Consistency checks on MAOA were conducted using both genotypic and self-reported sex. Discrepancies were genotyped again. If a discrepancy still remained then self-reported sex was used to code MAOA. This applies to values for both MAOA_V and MAOCA1.

6. Quality

A test of Hardy Weinberg equilibrium (HWE) was performed for each allele genotype in race/ethnicity-specific strata. Deviations from HWE (at α~0.05) were only identified among blacks for DRD4 and HTTLPR, and among whites for COMT. A race/ethnicity- and sex-stratified random sample of 58 Add Health Wave IV (2008) participants was examined twice in the Main Study, one to two weeks apart. At each examination, saliva was collected according to study protocol, typically by the same FI and at

approximately the same time of day. On both occasions, saliva was processed, shipped, and genotyped by masked IBG personnel as detailed above. Short-term reliability of IBG-assigned genotypes from the two examinations was estimated as a kappa (*K*) statistic, 95% confidence interval (Cicchetti et al. 1971; Cohen 1960; Fleiss et al. 1969) and interpreted using conventional criteria (Landis and Koch, 1977). Reliability was uniformly good to excellent (*Exhibit 5*).

Exhibit 5. Short-term reliability of genotypes in the Main Study, sorted by percent agreement

Candidate Genotype	n*	% Agree	*K*, 95% CI†
rs12945042, Allele B	56	84	71,55-87
rs4680, Allele B	56	84	61,40-83
MAOCA-1, Allele B‡	26	85	81,64-97
MAOCA-1, Allele A	58	86	79, 66-92
rs12945042, Allele A	56	88	58, 30-85
rs4680, Allele A	56	88	68, 50-87
DRD5, Allele B	58	88	85,75-95
DRD2, Allele A	57	88	77, 62-92
DRD4, Allele A	58	88	73, 54-92
MAOA_V, Allele A	58	88	79, 65-93
DRD5, Allele A	58	90	88,78-97
DAT1, Allele A	58	90	80, 67-94
DRD4, Allele B	58	90	83, 71-95
LALGS, Allele B	58	90	81, 67-95
DAT1, Allele B	58	91	65, 39-91
TRI, Allele B	58	91	82, 67-97
MAOA_V, Allele B‡	26	92	84, 63-100
DRD2, Allele B	57	93	57,20-94
HTTLPR, Allele B	58	95	85, 68-100
LALGS, Allele A	58	95	89, 78-100
TRI, Allele A	58	95	89, 78-100
HTTLPR, Allele A	58	97	93, 83-100

*Includes missing genotypes as a category at the first or second visit in ≤7 (13%) of respondents. †Unweighted kappa statistic, 95% confidence interval. For genotypes with ≥ 2 categories, weighting was associated with only slight changes in *K*. ‡Restricted to women.

7. References

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(1):73-78.

Applied Biosystems, Inc. TaqMan® OpenArray™ Genotyping System. Part Number 4400402 Rev. B. Available at http://tools.invitrogen.com/content/sfs/manuals/cms_058178.pdf.

Asghari V, Sanyal S, Buchwaldt S, Paterson A, Jovanovic V, Van Tol HH. Modulation of intracellular cyclic AMP levels by different human dopamine D4 receptor variants. J Neurochem 1995;65(3):1157–1165.

Black GCM, Chen Z-Y, Craig IW, Powell JF. Dinucleotide repeat polymorphism at the MAOA locus. Nucleic Acids Res 1991;19: 689.

Brownstein MJ, Carpten JD, Smith JR. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. BioTechniques 1996; 20:1004-1006, 1008-1010.

Cicchetti DV, Allison T. A new procedure for assessing reliability of scoring EEG sleep recordings. Am J EEG Technol 1971;11:101-109.

Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960;20(1):37-46.

Denney RM, Koch H, Craig IW. Association between monoamine oxidase A activity in human male skin fibroblasts and genotype of the MAOA promoter-associated variable number tandem repeat. Hum Genet 1999;105(6): 542-551.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. "Touchdown" PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 1991;19(14):4008.

Entzel P, Whitsel EA, Richardson A, Tabor J, Hallquist S, Hussey J, Halpern CT, Harris KM. Add Health Wave IV documentation: Cardiovascular and anthropometric measures. UNC Chapel Hill: Carolina Population Center, 2009. Available at http://www.cpc.unc.edu/projects/addhealth/data/guides/Wave%20IV%20cardiovascular%20and%20anthropometric%20documentation%20110209.pdf.

Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. Psychol Bull 1969;72(5):323-327.

Gelernter J, Cubells JF, Kidd JR, Pakstis AJ, Kidd KK. Population studies of polymorphisms of the serotonin transporter protein gene. Am J Med Genet 1999;88(1): 61–66.

Haberstick BC, Lessem JM, Hopfer CJ, Smolen A, Ehringer MA, Timberlake D, Hewitt JK. Monoamine oxidase A (MAOA) and antisocial behaviors in the presence of childhood and adolescent maltreatment. Am J Med Genet B Neuropsychiatr Genet 2005;135B(1): 59-64.

Haberstick BC, Smolen A. Genotyping of three single nucleotide polymorphisms following whole genome preamplification of DNA collected from buccal cells. Behav Genet 2004;34:541-547.

Harris KM. Design features of Add Health. UNC Chapel Hill: Carolina Population Center, 2012. Available at http://www.cpc.unc.edu/projects/addhealth/data/guides/design%20paper%20WI-IV.pdf.

Harris KM, Halpern CT, Whitsel EA, Hussey JM, Killeya-Jones LA, Tabor J, Dean SC. Cohort Profile: The National Longitudinal Study of Adolescent Health (Add Health). Int J Epidemiol, In press.

Heils A, Teufel A, Petri S, Stober G, Riederer P, Bengel D, Lesch KP. Allelic variation of human serotonin transporter gene expression. J Neurochem 1996;66(6):2621-2624.

Heinz A, Goldman D, Jones DW, Palmour R, Hommer D, Gorey JG, Lee KS, Linnoila M, Weinberger DR. Genotype influences in vivo dopamine transporter availability in human striatum. Neuropsychopharmacology 2000;22(2):133–139.

Hu X, Oroszi G, Chun J, Smith TL, Goldman D, Schuckit MA. An expanded evaluation of the relationship of four alleles to the level of response to alcohol and the alcoholism risk. Alcohol Clin Exp Res 2005;29(1):8-16.

Hu XZ, 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(5):815-826.

Illumina, Inc. GoldenGate® Assay. Pub. No. 370-2004-006 03Oct06. Available at http://res.illumina.com/documents/products/workflows/workflow_goldengate_assay.pdf

Iwasiow RM, Desbois A, Birnboim HC. Long-term stability of DNA from saliva samples stored in the Oragene self-collection kit. DNAgenotek, 2011. Available at http://www.dnagenotek.com/US/pdf/PD-WP-005.pdf.

Kaiser R, Muller-Oerlinghausen B, Filler D, Tremblay PB, Berghofer A, Roots I, Brockmoller J. Correlation between serotonin uptake in human blood platelets with the 44-bp polymorphism and the

17-bp variable number of tandem repeat of the serotonin transporter. Am J Med Genet 2002;114(3):323-328.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-174.

Lerman C, Caporaso N, Main D, Audrain J, Boyd NR, Bowman ED, Shields PG. Depression and self-medication with nicotine: the modifying influence of the dopamine D4 receptor gene. Health Psychol 1998;17(1):56-62.

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(5292):1527–1531.

Livak KJ. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genetic Analysis: Biomol Eng 1999;14:143-149.

Mattay VS, Goldberg TE, Fera F, Hariri AR , Tessitore A, Egan MF, Kolachana B, Callicott JH, Weinberger DR. Catechol O-methyltransferase val158-met genotype and individual variation in the brain response to amphetamine. Proc Nat Acad Sci 2003;100:6186-6191.

Nakamura M, Ueno S, Sano A, Tanabe H. The human serotonin transporter gene linked polymorphism (5-HTTLPR) shows ten novel allelic variants. Mol Psychiatry 2000;5(1):32–38.

Nobile M, Begni B, Giorda R, Frigerio A, Marino C, Molteni M, Ferrarese C, Battaglia M. Effects of serotonin transporter promoter genotype on platelet serotonin transporter functionality in depressed children and adolescents. J Am Acad Child Adolesc Psychiatry 1999;38(11):1396-1402.

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. Proc Natl Acad Sci USA 1993;90(6):2542–2546.

Rylander-Rudqvist T, Håkansson N, Tybring G, Wolk A. Quality and quantity of saliva DNA obtained from the self-administrated Oragene method—a pilot study on the cohort of Swedish men. Cancer Epidemiol Biomarkers Prev 2006;15(9):1742-1745.

Sabol SZ, Hu S, Hamer D. A functional polymorphism in the monoamine oxidase A gene promoter. Hum Genet 1998;103(3):273-279.

Samochowiec J, Lesch KP, Rottmann M, Smolka M, Syagailo YV, Okladnova O, Rommelspacher H, Winterer G, Schmidt LG, Sander T. Association of a regulatory polymorphism in the promoter region of the monoamine oxidase A gene with antisocial alcoholism. Psychiatry Res 1999;86(1):67-72.

Sherrington R, Mankoo B, Attwood J, Kalsi G, Curtis D, Buetow K, Povey S, Gurling, H. Cloning of the human dopamine D5 receptor gene and identification of a highly polymorphic microsatellite for the DRD5 locus that shows tight linkage to the chromosome 4p reference marker RAF1P1. Genomics 1993;18:423–425, 1993

Tabor J, Whitsel EA. Add Health Wave IV documentation: Prescription Medication Use. UNC Chapel Hill: Carolina Population Center, 2010. Available at http://www.cpc.unc.edu/projects/addhealth/data/guides/medication_documentation.pdf.

Vandenbergh DJ, Perisco AM, Hawkins AL, Griffin CA, Li X, Jabs EW, Uhl GR. Human dopamine transporter gene (DAT1) maps to chromosome 5p15.3 and displays a VNTR. Genomics 1992;14(4):1104–1106.

Van Tol HH, Wu CM, Guan HC, Ohara K, Bunzow JR, Civelli O, Kennedy J, Seeman P, Niznik HB, Jovanovic V. Multiple dopamine D4 receptor variants in the human population. Nature 1992;358(6382):149-152.

Vanyukov MM, Moss HB, Kaplan BB, Kirillova GP, Tarter RE. Antisociality, substance dependence, and the DRD5 Gene: A preliminary study. Amer J Med Genet (Neuropsych Genet) 2000;96:654–658.

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(3):224-226.

Whitsel EA, Tabor JW, Nguyen QC, Cuthbertson CC, Wener MH, Potter AJ, Killeya-Jones LA, Harris KM. Add Health Wave IV documentation: Measures of glucose homeostasis. UNC Chapel Hill: Carolina Population Center, 2012. Available at http://www.cpc.unc.edu/projects/addhealth/data/guides/Glucose_HbA1c.pdf.

Willeit M, Stastny J, Pirker W, Praschak-Rieder N, Neumeister A, Asenbaum S, Tauscher J, Fuchs K, Sieghart W, Hornik K, Aschauer HN, Brucke T, Kasper S. No evidence for in vivo regulation of midbrain serotonin transporter availability by serotonin transporter promoter gene polymorphism. Biol Psychiatry 2001;50(1):8-12.