

# Functional Genomics of *Serotonin Receptor 2A (HTR2A)*: Interaction of Polymorphism, Methylation, Expression and Disease Association

Virginia R. Falkenberg · Brian M. Gurbaxani · Elizabeth R. Unger · Mangalathu S. Rajeevan

Received: 7 June 2010 / Accepted: 25 September 2010 / Published online: 13 October 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** Serotonergic neurotransmission plays a key role in the pathophysiology of neuropsychiatric illnesses. The functional significance of a promoter polymorphism, –1438G/A (rs6311), in one of the major genes of this system (*serotonin receptor 2A, HTR2A*) remains poorly understood in the context of epigenetic factors, transcription factors and endocrine influences. We used functional and structural equation modeling (SEM) approaches to assess the contributions of the polymorphism (rs6311), DNA methylation and clinical variables to *HTR2A* expression in chronic fatigue syndrome (CFS) subjects from a population-based study. *HTR2A* was up-regulated in CFS through allele-specific expression modulated by transcription factors at critical sites in its promoter: an E47 binding site at position –1,438, (created by the A-allele of rs6311 polymorphism), a glucocorticoid receptor (GR) binding site encompassing a CpG at position –1,420, and Sp1 binding at CpG methylation site –1,224. Methylation at –1,420 was strongly correlated with methylation at –1,439, a CpG site that is dependent upon the G-allele of rs6311 at position –1,438. SEM revealed a strong negative interaction between E47 and GR binding (in conjunction with cortisol level) on *HTR2A* expression. This study suggests that the promoter polymorphism (rs6311) can affect both transcription factor binding and promoter methylation, and this along with an individual's stress response can impact the rate of *HTR2A* transcription in a genotype and

methylation-dependent manner. This study can serve as an example for deciphering the molecular determinants of transcriptional regulation of major genes of medical importance by integrating functional genomics and SEM approaches. Confirmation in an independent study population is required.

**Keywords** Chronic fatigue syndrome · Serotonergic neurotransmission · *Serotonin receptor 2A (HTR2A)* · Polymorphism · Gene expression · DNA methylation · Cortisol · Structural equation modeling

## Introduction

Regulatory genetic and epigenetic variations constitute major determinants of differential gene expression via the loss or gain of transcription factor binding or DNA methylation sites. Case-control studies have identified a number of functional polymorphisms, such as single-nucleotide polymorphisms (SNPs) and variable number tandem repeats (VNTRs), in the promoter regions of medically important genes including a group of genes involved in serotonergic signaling, the serotonin transporter (*SLC6A4*), *monoamine oxidase A (MAOA)* and *serotonin receptor 2A (HTR2A)*. Functional polymorphisms in all three of these genes have been associated with major depression, chronic fatigue syndrome (CFS) and post-traumatic stress disorder (PTSD), as well as a variety of other mood disorders (Li and He 2008; Smith et al. 2008; Uher and McGuffin 2008).

Current studies to evaluate functional polymorphisms have generally been restricted to either in vitro reporter gene assays, gel-shift assays to assess DNA-protein interactions, or allele-specific analysis of mRNA (Knight 2005). DNA methylation analyses are usually carried out independently

**Electronic supplementary material** The online version of this article (doi:10.1007/s12017-010-8138-2) contains supplementary material, which is available to authorized users.

V. R. Falkenberg · B. M. Gurbaxani · E. R. Unger · M. S. Rajeevan (✉)  
Division of Viral and Rickettsial Diseases, Centers for Disease Control & Prevention, Atlanta, GA 30333, USA  
e-mail: mor4@cdc.gov

from analyses of genetic polymorphisms and gene-environmental interactions (Gopalakrishnan et al. 2008). Because gene expression is a multi-factorial process, isolated genomic studies rarely result in a defined molecular mechanism or integrated biological model to explain these genetic associations with disease. Simultaneous examination of genetic and epigenetic variations in the context of environmental and developmental responsive signaling molecules (for example, cortisol secretion in response to stress) may provide a better assessment of their influence on gene expression and clarify their pathogenic role. We employed an integrated experimental and computational approach to assess the influence of genetic polymorphisms and DNA methylation on *HTR2A* expression in peripheral blood mononuclear cells (PBMC). *HTR2A* is implicated in the regulation of serotonergic neurotransmission and the hypothalamic-pituitary-adrenal (HPA) axis (Lanfumeey et al. 2008; Lister-Williams et al. 1998; Porter et al. 2004). Subjects were derived from a population-based study of CFS, a complex disease of unknown etiology but with major etiologic roles hypothesized for both HPA-axis-mediated cortisol signaling (hypocortisolism) and serotonergic neurotransmission (Cleare 2003; Cleare et al. 1995; Heim et al. 2000; Smith et al. 2008).

To integrate and evaluate the interactions between polymorphisms, methylation, stress hormones (e.g. cortisol) and gene expression, we applied structural equation modeling (SEM), a statistical approach that has been widely used in the psychosocial sciences (Schreiber 2008). SEM allows the user to design latent variables that can be connected to various observed data points. The algorithm then does model fitting to find the path coefficients and error variances that best describe the observed data based on the latent variables and pathways in the model. This allows the user to assess the contribution of various observed data points to the outcome in the best-fit model.

Our results reveal that changes in *HTR2A* expression are associated with allele-specific expression differences modulated by both genetic polymorphisms and DNA methylation at critical transcription factor binding sites (TFBS). This computational approach to integrate regulatory genetic polymorphisms, epigenetic variations and gene expression changes is a novel approach to decipher the molecular determinants of disease and will have applications to other complex illnesses.

## Methods

### Ethics Statement

This study adhered to human experimental guidelines of US Department of Health and Human Services and the

Helsinki Declaration. The CDC Human Subjects committee approved the study protocol, and all subjects gave written informed consent.

### Subjects and Illness Classification

Subjects from a population-based CFS surveillance study in Wichita, KS, USA were recruited for a previously described 2-day in-hospital case-control study (Vernon and Reeves 2006). Subjects were classified according to the 1994 research case definition as evaluated by standardized questionnaires (Reeves et al. 2005). Of the initial 227 subjects recruited, 13 did not have genotype information for rs6311 (−1438G/A) and were dropped from this study leaving 214 enrollees (rs6311 genotypes: 83 GG, 95 GA, and 36 AA). Methylation analysis was done on 185 subjects with sufficient residual DNA (71 GG, 84 GA, 30 AA). Mean age and body mass index (BMI) of these 185 subjects were  $50 \pm 8.8$  years and  $29 \pm 4.8$ , respectively. Ninety-four percent of subjects were Caucasians, and 82% of subjects were female in this analysis.

Disease-specific analyses were restricted to the 39 CFS and 49 non-fatigued (NF) control subjects who met the criteria for diagnosis and had material for methylation analysis. The CFS and NF subjects did not differ in age (NF,  $50.53 \pm 8.3$ ; CFS,  $50.59 \pm 9.2$ ,  $P = 0.9748$ ), sex (NF females, 79.6%; CFS females, 87.2%,  $P = 0.4036$ ), BMI (NF,  $28.4 \pm 4.7$ ; CFS,  $29.8 \pm 4.5$ ,  $P = 0.1767$ ) or race (NF Whites, 94%; CFS Whites, 93%,  $P = 0.9999$ ; Table 1).

### DNA/RNA Isolation

Blood collection and separation of PBMCs from this study population were described previously. Both DNA and RNA were extracted from the same PBMC sample by Trizol DNA/RNA extraction protocol (Invitrogen, Carlsbad, CA). Methods to assess the quality and quantity of DNA and RNA preparations were described earlier (Sorensen et al. 2009).

### Genotyping and Methylation Analyses

Markers −1438G/A (rs6311) and C102T (rs6313) were genotyped as part of a previous study (Smith et al. 2008).

**Table 1** Demographic analysis of CFS and NF subjects in the study

Characteristics	NF ( $n = 49$ )	CFS ( $n = 39$ )	$P$ values
Age (mean $\pm$ SD)	$50.53 \pm 8.3$	$50.59 \pm 9.2$	0.9748
Female (%)	79.6	87.2	0.4036
Caucasians (%)	94	93	0.9999
BMI (mean $\pm$ SD)	$28.4 \pm 4.7$	$29.8 \pm 4.5$	0.1767

**Table 2** Primer sets for quantitation of site-specific *HTR2A* CpG methylation by pyrosequencing

Amplicon	Amplification primers <sup>a</sup>	Sequencing primers <sup>b</sup>	CpG position <sup>c</sup>
1	FW: GGAAAGGGTTTTTTTTATAAGTAAGGTAAG RV: TCATTTTCATCAAACCTCCCTCTAT	Sq1: <b>GGTTTTGGATGGAAGTG</b> <i>TYRGATATTATAGTATTTYG</i> Sq2: <b>AAGAGGGAGAAGAAAAAGT</b> <i>TTGTTGGTTYGTTTTTTG</i>	–1,439, –1,420 –1,224
2	FW: AGGGAGGTTTGATGAAATGAA RV: ACTTTCCTCCCTAAAAATTCTCAT	Sq3: <b>TTTTTTTATGTAGTTTTTTTGTG</b> <i>YGATTTTGAGGGGTTYGTGAATGATTTTA</i> <i>AATGTGTGTTTGTGAGGYGAGTYGTATA</i> Sq4: <b>ATAGGGAGGGAGGAAT</b> <i>TTAGTYGAGTYGTGTAGAGGAAGTTAATA</i> <i>GGATTTAGTAGTYGGGA</i> Sq5: <b>TAGTTTAATTATTGTGAGAGATG</b> <i>TAGYGAGTTATAGAATAATAAATGTATTTTA</i> <i>TGTGTGRATTTTG</i> Sq6: <b>GGATTTTGAAGATAAAATGTAAG</b> <i>TTTTATGTYGTTATATTTATTGTTGTGTA</i> <i>ATTTTTTTTTYGGTTG</i>	–1,065, –1,052, –1,017, –1,012 –988, –983, –949 –817 –753, –721
3	FW: GGGAGGAAAGTTGTTTGTTAAT RV: ACCCACACTCTATAACACTAAAAC	Sq7: <b>GTTTAGGATATTTATTTTTT</b> <i>YGAGYGTATAAAAAAATTTTGTAATTTTA</i> <i>TGTTAAAAGTTTATTTTGTTTTTTTGTTTTY</i> <i>GG</i> Sq8: <b>GGAAGATAAAAAATTTAGTT</b> <i>TYGGGAGAATAGTATGTATATTAGTTT</i>	–314, –310, –253 –125

<sup>a</sup> FW, forward primer; RV, reverse primer (biotinylated)

<sup>b</sup> Sequencing primers are indicated in bold. Sequences to analyze are italicized with the CpG sites underlined. There are two sequencing primers for amplicon 1 (Sq 1 and 2), 4 for amplicon 2 (Sq3–Sq6) and 2 for amplicon 3 (Sq 7 and 8)

<sup>c</sup> CpG nucleotide positions are listed in the order (→) they are read from the corresponding pyrograms

MatInspector and MatBase software programs (Cartharius et al. 2005) were used to identify TFBS in the *HTR2A* promoter prior to designing primers for functionally important cytosine guanine dinucleotide (CpG) sites (Table 2) using Assay Design Software (Biotage, Sweden). Pyrosequencing was used to determine CpG methylation levels as reported earlier (Rajeevan et al. 2006) except that PCR was done with 2.0 mM MgCl<sub>2</sub>.

#### Gene Expression Analysis

The –1438G/A (rs6311) and C102T (rs6313) polymorphisms are in complete linkage disequilibrium. This information was used to design real-time reverse-transcription PCR (RT-PCR) and pyrosequencing assays encompassing the C102T exonic marker. Real-time RT-PCR was used to determine total *HTR2A* expression in 117 subjects of whom 31 were CFS and 15 were NF subjects. Real-time RT-PCR was conducted as described previously using Light Cycler (Sorensen et al. 2009) except that it generated a 197 bp *HTR2A* product using the biotinylated

forward (AGCTCAACTACGAACTCCCTAATG) and reverse (TCCAGTTAAATGCATCAGAAGTGT) primers. *HTR2A* primers were annealed at 62°C resulting in PCR efficiency of 1.96. The primers and annealing temperature for *peptidylprolylase B (PPIB)*, internal reference gene for normalization, were reported previously (Sorensen et al. 2009). The biotinylated LightCycler product was used directly for pyrosequencing with the sequencing primer ATCAGAAGTGTTAGCTTCTC to analyze the sequence CRGAGTTAAAGTCATTACTGTAGAGCC where *R* corresponds to the allelic ratio in mRNA for the *HTR2A* C102T marker. Pyrosequencing was used to determine allele-specific expression in all 22 subjects from the disease-specific expression analysis that was heterozygous for the –1438G/A marker (14 CFS and 8 NF).

#### Cortisol Measurement

Urinary 24-h free cortisol was measured as reported previously (Smith et al. 2009).

Statistical Analysis

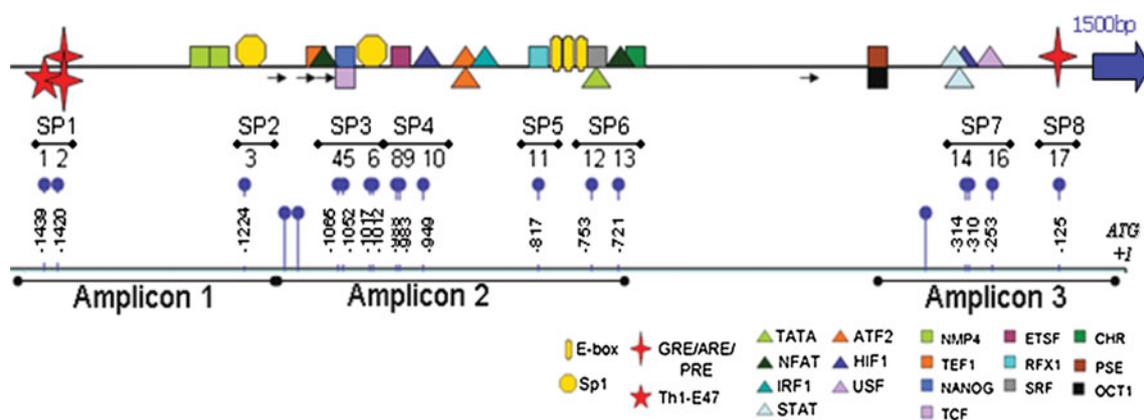
All statistical analyses were done with the Graphpad Prism program (Graphpad Software, CA). Age, BMI and expression data were normally distributed and were analyzed by student's *t* tests, and sex and race were analyzed by Chi-square tests. Mann–Whitney *U* test was used to compare methylation levels between groups. Spearman correlations were used to assess correlation between methylation levels of different CpG sites. Primarily, we used the more conservative Bonferroni correction for multiple hypothesis testing, and associations that did not withstand the Bonferroni correction were tested again by the less conservative false discovery rate (FDR; Reiner et al. 2003). SEM was performed using SSI's LISREL 8.8 software program. Analysis used 46 carriers of the A-allele (13 AA, 33 GA) for rs6311, who have data on *HTR2A* expression and methylation (at least two of −1,439, −1,420, and −1,224 CpG sites) as well as urinary 24-h free cortisol values. Total expression of *HTR2A* was used for the outcome variable "*HTR2A* expression". SEM was performed on z-score transformed data with missing data points imputed using the PRELIS software supplied in the LISREL software suite (15 missing values/330 data points). Path diagrams were designed and run using SIMPLIS syntax, and to increase the degrees of freedom, the error variances of the methylation data were set equal to the calculated technical variance of the methylation assay after z-score normalization of the data (0.1). The *P* value of the presented model, 0.98 is close to the best-fit value of 1.0 indicating that this model fits the observed data very well.

Results

DNA Methylation Varies Along *HTR2A* Promoter

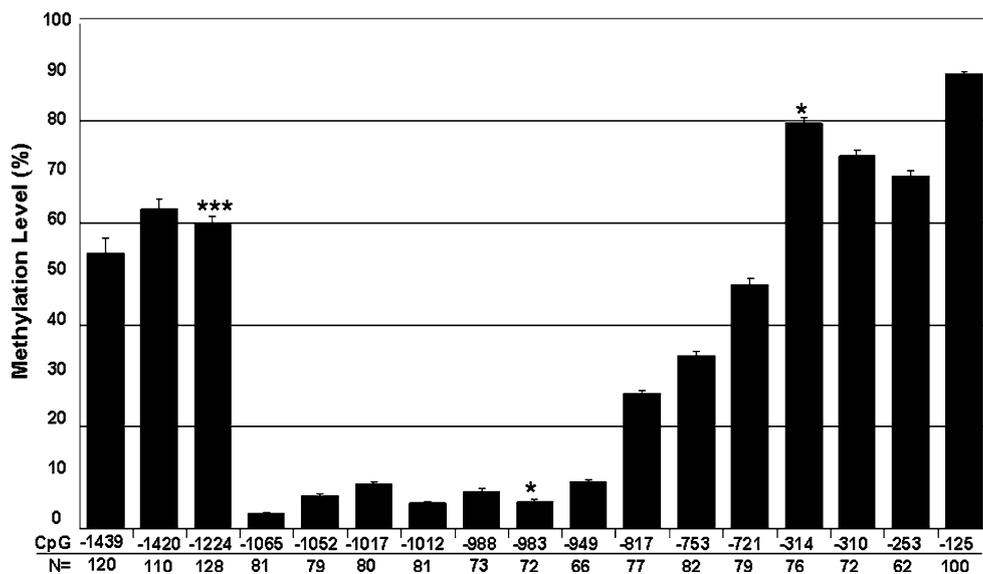
Site-specific methylation levels at 17 out of 20 CpG sites in the 1,500 bp promoter region of *HTR2A* were quantified by pyrosequencing (Fig. 1). Three CpG sites excluded from the analysis neither coincided with any TFBS nor met PCR primer design criteria. There were a total of 3,145 CpG sites (17 unique sites in 185 subjects) to be assayed in this study. Since our goal was to quantify methylation levels at each of the CpG site in every subject, we enforced strict quality control at various stages of the assay that includes bisulfite treatment, PCR product specificity and pyrogram quality for allele quantification. This means that while we have 185 subjects with DNA for methylation assays, final data from all 185 subjects were not equally good to quantify methylation at each of the CpG sites. We removed suboptimal data corresponding to each CpG site, and this process resulted in different numbers of subjects for each CpG site as shown in Figs. 2 and 3. We also carried out separate demographic analyses for each association and verified that the missing subjects did not impact the case–control design of the study.

Methylation levels varied along the *HTR2A* promoter (Fig. 2) with extremely low methylation (mean 6.5%; range 2.9–9.4%) at seven CpG sites (−1,065 to −949). Methylation levels increased in both directions from this region. CpG sites −1,439, −1,420 and −1,224 in the distal region had average methylation levels of 53, 63 and 60%, respectively. CpG sites at −817, −753 and −721 had average methylation levels of 27, 34 and 47%, respectively.



**Fig. 1** A schematic representation of transcription factor binding sites (*top strand*) and the pyrosequencing strategy for the analysis of methylation in *HTR2A* (*bottom strand*). *Top strand*: Location of transcription factor binding sites (TFBS) indicated along with potential transcription initiation sites (*small arrows below line*) and open reading frame (*large arrow at end of line*). Shapes are defined at the *bottom* of the figure. If the TFBS is aligned on the forward strand,

it is placed on *top* of the *line*; if it is aligned to the reverse strand, it is shown below the line. *Bottom strand*: Location of CpG sites *bullet head* numbered 1–17 is listed below the sequencing primers (SP1–SP8) used for their analysis. Position relative to the ATG (+1) is listed for each CpG site. Three different PCR amplicons, shown below CpG sites they encompass, are used with eight different sequencing primers to determine % methylation



**Fig. 2** CpG site-specific methylation of the *HTR2A* promoter. Methylation levels (%)  $\pm$ SE for 17 CpG sites. While there were 185 subjects in the study, not all subjects yielded data for every CpG sites after enforcing the quality control criteria. *N* represents the number of subjects at the indicated CpG sites for analysis after meeting the quality control criteria for methylation assay. In a subset analysis, three CpG sites were found with significant differences in

methylation levels between CFS and NF subjects. Number of subjects in this subset analysis: 23 CFS and 32 NF subjects for site  $-1,224$ , 13 CFS and 16 NF subjects for site  $-983$  and 19 CFS and 16 NF subjects for site  $-314$ . Statistical significance of disease association with methylation indicated by stars over the corresponding site. \*  $P = 0.04$ ; \*\*\* $P = 0.0036$  (Bonferroni corrected alpha = 0.003)

The four most proximally located CpG sites, spanning a 190 bp region, were the most highly methylated sites in the *HTR2A* promoter with average methylation levels of 80% at  $-314$ , 73% at  $-310$ , 69% at  $-253$  and 89% at  $-125$  (Fig. 2).

Methylation levels at three out of 17 CpG sites ( $-1,224$ ,  $-983$ , and  $-314$ ) differed significantly between CFS and NF control subjects (Fig. 2). At  $-1,224$ , NF subjects had significantly higher methylation than CFS subjects ( $P = 0.0036$ ), whereas at CpG sites  $-983$  and  $-314$ , NF subjects showed lower methylation than CFS subjects ( $P = 0.04$  for both sites). None of these associations remained significant after the conservative Bonferroni correction for multiple testing (Bonferroni corrected alpha = 0.003). However, multiple hypothesis testing correction by false discovery rate (FDR 33%, alpha = 0.058) indicated that while one of these three associations could have arisen by chance the other two are likely to be true positives.

#### Genotype-Dependent Methylation

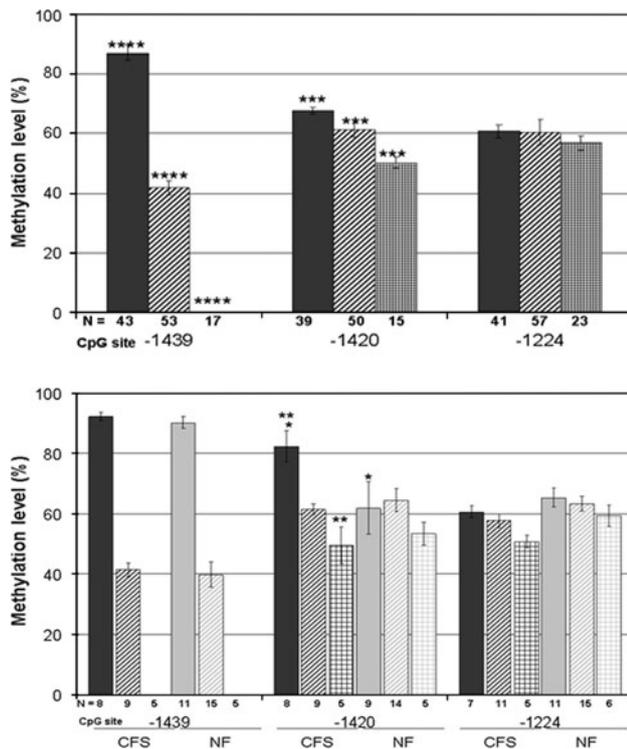
Since CpG site  $-1,439$  is lost as a result of G to A sequence variation (SNP rs6311) at  $-1,438$ ,  $-1438G/A$  genotype-dependent methylation at CpG sites  $-1,439$ ,  $-1,420$  and  $-1,224$  was examined in all subjects for whom data were available (Fig. 3 top panel). Methylation at  $-1,439$  was very strongly dependent upon the  $-1438G/A$  genotypes (GG 87.5%, GA 43%, AA 0%;  $r^2 = 0.999$ ), as

expected. Methylation of the adjacent CpG site  $-1,420$  was also slightly genotype-dependent, but the association was significant only in comparison with the AA genotype ( $P < 0.00085$ , Bonferroni corrected alpha = 0.0006). No genotype-dependent effect on methylation was observed on the remaining 15 CpG sites.

Genotype-dependent methylation also differed between CFS and NF subjects (Fig. 3 bottom panel). CFS subjects with GG genotype had significantly higher methylation at  $-1,420$  (a CpG site 19 bp downstream from the SNP site) than NF subjects with GG genotype ( $P = -0.038$ ). The association between methylation at  $-1,420$  between GG and AA genotypes was significant only for CFS subjects ( $P = 0.0007$ , Bonferroni corrected alpha = 0.003). To further investigate CFS-associated methylation differences, pairwise Spearman correlation matrices were created for all 17 CpG sites for CFS and NF subjects separately. Methylation levels at positions  $-1,439$  and  $-1,420$  were correlated significantly in CFS subjects ( $P = 0.000000076$ , Bonferroni corrected alpha = 0.00003) but not in the NF subjects ( $P = 0.37$ ) reflecting the differential genotype-dependent methylation that is associated with CFS.

#### Allele-Specific Expression of *HTR2A*

*HTR2A* expression was determined in PBMCs from 117 subjects for whom RNA was available (42 GG, 54 GA, 21



**Fig. 3** Methylation at selected CpG sites in *HTR2A* stratified by rs6311 polymorphism. *Top panel:* Methylation stratified by rs6311 genotypes (GG black bar, GA striped bar, AA checked bar) at positions  $-1,439$ ,  $-1,420$  and  $-1,224$  for all subjects having methylation data. *N* represents the number of subjects within each genotype at the indicated CpG sites for analysis after meeting the quality control criteria for methylation assay. Mean methylation levels (%)  $\pm$ SE. \*\*\*\* = all associations significant ( $P < 1E-28$ ). \*\*\* = only AA versus GA and AA versus GG significant ( $P < 0.00085$ , Bonferroni corrected alpha, 0.006). *Bottom panel:* Methylation stratified by rs6311 genotypes and disease status. CFS (GG black bar, GA dark striped bar, AA dark checked bar) and NF (GG gray shaded bar, GA light striped bar, AA light checked bar) at positions  $-1,439$ ,  $-1,420$ , and  $-1,224$ . *N* indicates the number of subjects meeting the quality control criteria for methylation assay within each genotype at the indicated CpG site and stratified by CFS versus NF status. \* = GG-CFS versus GG-NF ( $P = 0.038$ ). \*\* = GG-CFS versus AA-CFS ( $P = 0.0007$ ; Bonferroni corrected alpha = 0.003)

AA for SNP rs6311). Correlation between total *HTR2A* expression in each subject and methylation at each CpG site was analyzed using a Spearman correlation matrix. None of the individual CpG site-specific methylation levels was significantly associated with total expression when correction for multiple hypotheses was applied. *HTR2A* expression in subjects with AA genotype tended to be higher than in those with GA or GG genotypes, but this did not exhibit statistical significance for any of the genetic models (additive, dominant or recessive) tested, perhaps due to the rather large standard error for the AA genotype (Fig. 4a).

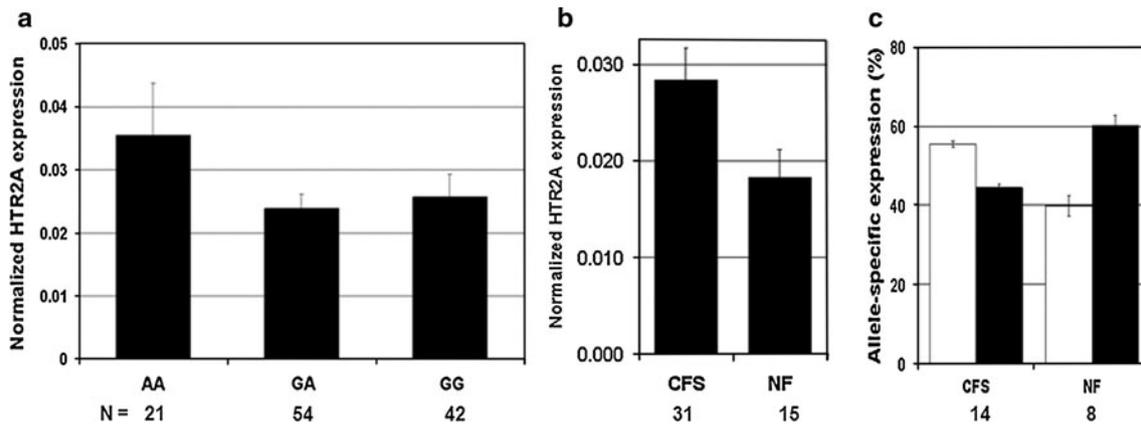
When analysis was restricted to those classified as CFS or NF (31 CFS and 15 NF) and stratified by disease status,

the relative expression of *HTR2A* was 1.56-fold higher in CFS compared to NF subjects ( $P = 0.01$ ; Fig. 4b). For the 22 heterozygous subjects (rs6311 GA) in this group (14 CFS, 8 NF), the allele-specific expression is shown in Fig. 4c. The A-allele contributed a higher percentage of total expression than the G-allele in CFS subjects ( $P = 0.019$ ), whereas the opposite was true (though not significant) for NF subjects ( $P = 0.064$ ). This disease-dependent differential expression of the A-allele was also seen in the overall expression of *HTR2A*, in that CFS subjects with the AA genotype had significantly higher expression of *HTR2A* than NF subjects with AA ( $P = 0.036$ , 6 CFS, 5 NF; data not shown). This disease-specific difference may contribute to the large standard error for the AA genotype in the analysis with all 117 subjects.

#### Modeling Transcriptional Regulation of *HTR2A*

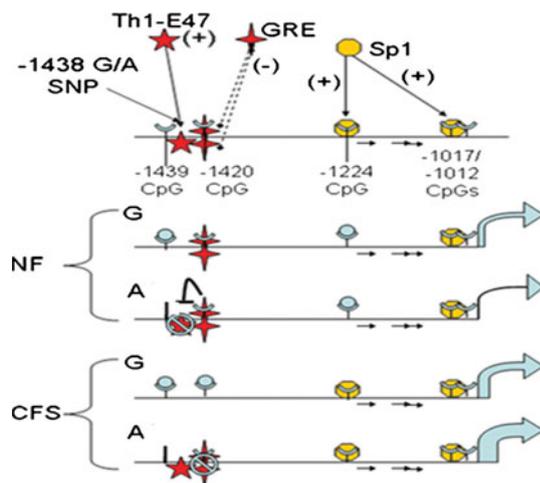
Multiple lines of evidence identified transcription factors that regulate *HTR2A* expression through binding at positions  $-1,438$ ,  $-1,420$  and  $-1,224$ ; E47 binds at  $-1,439$ , glucocorticoid receptor (GR) binds at  $-1,420$  and Sp1 binds at  $-1,224$  (Fig. 1; Supplementary Table 1). Binding of Sp1 at  $-1,224$  was identified as early as 1995 through sequencing, gel retardation and DNase I footprinting assays (Zhu et al. 1995) and also conserved in *Rattus norvegicus* and *Mus musculus*. Recently, we demonstrated multiple lines of evidence for the binding of GR at  $-1,420$  through computational analysis, competitive electrophoretic mobility shift assays (EMSA) using both HeLA and purified GR, and showing that this GRE binding site is highly conserved (79–95%) in *Rattus norvegicus*, *Mus musculus* and *Pan troglodytes* (Falkenberg and Rajeevan 2010). Experimental verification of E47 binding at  $-1,438$  specific to the A-allele of *HTR2A* was reported recently by our group using computational and competitive EMSA (Smith et al. 2008). We also found this E47 binding site to be conserved in *Mus musculus* (76%). Based on these substantial computational and experimental evidences for these different binding sites, we propose a qualitative regulatory model of *HTR2A* expression (Fig. 5). In this model, increased methylation of the CpG at position  $-1,224$  in NF subjects would result in reduced binding of the Sp1 transcriptional activator and an overall lower expression of *HTR2A*. Higher *HTR2A* expression in CFS subjects, particularly higher expression of the A allele, would be accounted for by the creation of a binding site for the transcriptional activator E47 at  $-1,438$ , along with the deficiency of GR binding at  $-1,420$  under a state of hypocortisolism (Heim et al. 2000) and the potential for increased methylation at this site.

We used SEM (Fig. 6) to provide a quantitative analysis of this model. We focused on the A-allele since its



**Fig. 4** Total and allele-specific expression of *HTR2A*. **a** *HTR2A* expression in all subjects stratified by  $-1,438$  G/A genotypes. **b** *HTR2A* expression stratified by disease status. Bars show mean normalized expression  $\pm$  SE for each group. **c** Allele-specific expression of  $-1,438$ G/A polymorphism. Bar graph shows the

percentage  $\pm$  SE of G (filled bar) or A (opened bar) allele in mRNA from heterozygote subjects as determined by pyrosequencing of cDNA. *N* represents the number of subjects under each group stratified genotype status (**a**) or disease status (**b** and **c**)

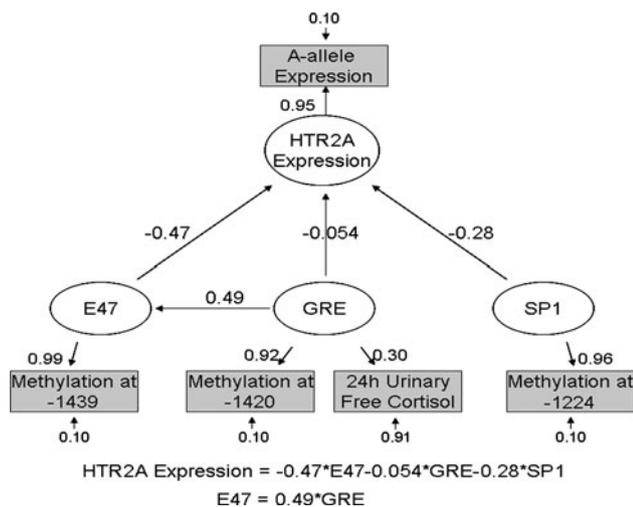


**Fig. 5** Analyses of GR binding and a qualitative regulatory model of *HTR2A* expression integrating genetic, epigenetic and disease associations. The 5' end of the *HTR2A* promoter is shown at the top with relevant experimentally verified TFBS defined. The putative + or – effect on transcription when bound and the location of CpGs in the sequence are also shown. Methylation sites are indicated by  $\nabla$  when largely unmethylated and by  $\ominus$  when largely methylated. Loss of methylation site due to polymorphism is indicated by  $\dashv$ . The effect of genotype and methylation on TF binding, and thus *HTR2A* transcription is indicated by the size of the arrow at the end of each model. The model predicts subjects with AA genotype largely explain the differential expression of *HTR2A* between CFS and NF subjects. Inhibition of E47 binding by cortisol-bound GR in NF subjects is indicated by  $\dashv$  and the inhibited transcription factor is covered by  $\ominus$

differential expression is responsible for a majority of the *HTR2A* expression difference between CFS and NF subjects. The structural equations for this model are:  $E47 = 0.49 * GRE$  and  $HTR2A = -0.47 * E47 - 0.28 * Sp1 - 0.054 * GRE$  where E47, GRE and Sp1 represent the TFBS at positions  $-1,439$ ,  $-1,420$  and  $-1,224$ , respectively. These latent

variables have been defined by observed variables that may influence transcription factor binding. For example, binding of GR at  $-1,420$  should be dependent upon its activation by cortisol as well as methylation of the cognate binding site. Therefore, the GRE latent variable is combined from the observed data on 24-h free cortisol in urine and methylation at  $-1,420$ . The E47 latent variable is defined by the genotype at  $-1,438$  as reflected in the methylation at  $-1,439$ . The Sp1 latent variable is defined only by methylation at  $-1,224$ . This model has a Chi-squared value of 0.47 with a *P* value 0.98 and 4 degrees of freedom. The root mean square error of approximation (RMSEA) is 0.0 (i.e. 0.0 to within two significant figures as determined by LISREL software), and the Akaike information criterion (AIC) is 22.49. The Chi-squared and RMSEA *P* values show that the model fits the data well, and the AIC is an indication of which model is more parsimonious in a comparison of models. According to the SEM results, the GRE doesn't directly contribute much to the transcription of *HTR2A* (the path coefficient is only  $-0.054$ , and is not significant). However, the path coefficient connecting the GRE to the E47 binding site variable is 0.49, (*t* value = 3.19, or *P* value  $< .001$  assuming normally distributed residuals) indicating there is a relatively strong interaction between these variables. This indicates an important indirect contribution of GRE to *HTR2A* transcription through inhibition of E47 binding, supporting the qualitative model shown in Fig. 5. While the path coefficient connecting Sp1 to *HTR2A* expression is marginally not significant (*t* value = 1.81, or *P* value = 0.07), it should be noted that this study is underpowered for detecting effects sizes in this range (i.e. path coefficients  $< 0.30$ ; Miles 2003).

Alternative models were also examined, and fitting the data to a model that includes the link between GRE and E47



**Fig. 6** Structural equation modeling of *HTR2A* transcriptional regulation. A candidate structural equation model (SEM) for A-allele-specific *HTR2A* expression is shown. Latent variables, indicating a computed but not directly observed construct, are shown as ovals, whereas measured values are represented by gray rectangles. Arrows between ovals and boxes show the linear relationship between latent variables and measured variables, while those feeding into the boxes from the periphery of the diagram show how much of the variance of each measurement is due to noise or some other source of unexplained variance (normalized to 1.0, shown as numbers without boxes). Only in the case of cortisol was this unexplained variance estimated by the algorithm. Arrows between latent variables indicate the strength of causal connection between those variables. The statistical significance of these path coefficients is further discussed in the text, but under the assumption of normally distributed residuals in the model then all coefficients  $\geq 0.3$  are significant with a  $P$  value  $\leq 0.05$ . The structural equations for the diagram are also shown in symbolic form. The close fit between this model and the data does not mean that the model shown is the one true model, only that it is statistically consistent with the data

without the direct link between GRE and *HTR2A* expression has slightly better statistics, e.g. the AIC = 20.58, further strengthening the suggested role for GRE. In order to decrease the risk of overfitting of the alternative model discussed earlier and the model shown in Fig. 6, the residual variance (i.e. the measured variance not contributing to the interaction of the latent variables) of four of the five measured variables (not cortisol) was manually set to 0.10, in keeping with a conservative estimate of the performance of the assay. Although estimating the residual variance of 24-h urinary free cortisol resulted in a relatively smaller contribution to the latent variable GRE (path coefficient = 0.30), the contribution is significant with a  $t$  value = 1.96 (Gaussian equivalent  $P$  value = 0.05).

## Discussion

This study suggests that the *HTR2A* gene may be up-regulated in CFS through allele-specific expression. An

integrated analysis of genetics, methylation, gene expression and clinical measurements using SEM reveal that this up-regulation may be dependent on the cis-regulatory transcription factors E47, GR and Sp1. We previously showed that the minor allele A of the promoter polymorphism rs6311 (−1438G/A) in *HTR2A* was more common in CFS than NF subjects (Smith et al. 2008). This SNP, which has also been associated with other complex disorders including depression, PTSD and schizophrenia, results in the creation of a binding site for E47 (Smith et al. 2008) and also results in the loss of CpG methylation site at −1,439. Our current study examined *HTR2A* methylation in subjects from a population-based clinical study of CFS and identified two CpG sites, −1,224 and −1,420 that showed differential methylation between CFS and NF subjects and dependence on sequence variation at position −1,438. We recently demonstrated the first experimental evidence for the binding of GR at CpG site −1,420 (Falkenberg and Rajeevan 2010), whereas binding of Sp1 at CpG site −1,224 and the genotype-dependent binding of E47 at −1,438 were reported earlier (Smith et al. 2008; Zhu et al. 1995). Changes at these cis-regulatory elements, two of which are potentially heritable (methylation at −1,439 and −1,420), may have contributed to increased expression of A-allele and to the overall up-regulation of *HTR2A* in CFS.

While the specific mechanism for the association of allele-specific expression with CFS is not known, we present qualitative (Fig. 5) and quantitative (Fig. 6) models that may account for the potential influence of E47, GR and Sp1 on *HTR2A* expression. As a result of reduced cortisol production in some CFS subjects (Heim et al. 2000), GR binding would also be reduced. This state of hypocortisolism is included in the model as leading to inhibition of GR-mediated transcriptional repression (Sorensen et al. 2009). The qualitative model suggests that in situations of high methylation at −1,420 and low GR activity as in CFS subjects, the A-allele is over-expressed in relation to the G-allele. On the other hand, when methylation at −1,420 is low and GR activity is high as in NF subjects, the G-allele is over-expressed. The joint contributions of transcription factor activation and promoter methylation in the context of sequence variation may explain the lack of consensus between previous investigations that used only isolated functional studies to examine the role of the rs6311 promoter polymorphism in the regulation of *HTR2A* (Myers et al. 2007; Norton and Owen 2005; Parsons et al. 2004; Poleskaya et al. 2006). These joint contributions of several regulatory factors may also explain the lack of simple direct correlations between methylation and expression levels or instances of deviations from simple genetic models as revealed by some results (Fig. 4a) in this study.

Quantitative analysis by SEM supports two significant aspects of the qualitative model of *HTR2A* transcriptional

regulation. First, the most important contribution of GRE to the transcription of *HTR2A* is indirect, mediated through competition with E47 (direct pathway coefficient is only  $-0.054$ , whereas GRE to the E47 pathway coefficient is  $0.49$ ). Second, there is a weaker but suggestive interaction between Sp1 binding and *HTR2A* expression (although the path coefficient of  $0.28$  is not significant, this study is underpowered for determining coefficients in this range).

Although the model posits that GR binding to  $-1,420$  would be important for the inhibition of E47 binding, only a small amount of the variance in the 24-h urinary free cortisol contributes to the model (about 10%, vs. the unaccounted for variance of 90%). However, compared to methylation at  $-1,420$ , variance in cortisol contributes a relatively substantial amount to the latent variable GRE, ( $0.32$  vs.  $0.93$ , or about 25% of the total). In addition, excluding cortisol from the model drives the model (Chi-squared)  $P$  value down to  $0.12$  and the  $P$  value of the RMSEA up to  $0.15$ , both indicating a worse fit. This indirect contribution of GRE to transcriptional regulation also raises the possibility that other GR family members with different kinetics and tissue distributions, such as mineralocorticoid receptor, may be more important for regulating *HTR2A* expression. As illustrated by the use of SEM in this study to provide a quantitative evaluation of the interaction of different TFBS on *HTR2A* allele-specific expression, SEM is an additional computational tool for analyzing complex transcriptional regulation paradigms.

The significances of our observed association of methylation levels between CpG sites at  $-1,439$  and  $-1,420$  in CFS but not in NF subjects are presently unknown. It is not clear how commonly an association of methylation between sites occurs, nor is it clear what accounts for a disease-specific association. Correlation of methylation between adjacent CpG sites in the distal region of the serotonin transporter promoter has been reported (Philibert et al. 2008), so further investigation into this phenomenon is clearly needed. Further investigation is also needed to assess the impact of other SNPs in this region on *HTR2A* expression.

There are several limitations to this study. Since we enforced strict quality control on CpG site-specific methylation analysis, information on methylation levels of all 17 CpG sites on all 185 subjects was not available. This contributed to small sample sizes in certain stratified analyses of genotype-dependent methylation levels. On the other hand, since methylation changes being close to the action of genome and a quantifiable molecular phenotype, it is likely that polymorphic CpG methylation at  $-1,439$  may exhibit higher effect size than complex disease phenotypes that often involve several biological pathways. As expected, the nearly perfect concordance ( $r^2 = 0.99$ ,  $P = 0.0261$ – $0.0392$ ) between methylation levels at  $-1,439$

and  $-1,438$  genotypes (Fig. 3 bottom panel) suggests that our conclusions are unlikely to be influenced by sample size. With respect to the allele-specific expression, the data presented included all 22 CFS and NF subjects heterozygous for the  $-1,438$  G/A promoter polymorphisms. As required by the case definition, this analysis (Fig. 4c) excluded those with exclusionary medical/psychiatric conditions. However, we examined this finding in a larger group of individuals ( $n = 35$ ) by including subjects who reported exclusionary medical and psychiatric conditions but otherwise met criteria for CFS and NF in this study. The results with the larger group of subjects (data not shown) support the original finding (Fig. 4c) of allele-specific expression of *HTR2A* in PBMCs that differ with the fatigue status of the subjects. We recognize that this allele-specific expression finding, although statistically significant, should be considered as suggestive because of the relatively small sample size. Future studies are needed to replicate this finding.

Analytically, the close fit between the SEM model and the data does not mean that the model shown is the one true model, only that it is statistically consistent with the data. The better  $P$  values for this model, given what is typical for SEM in the literature, may be reflecting relative paucity of the data and simplicity of the model instead of overfitting of the data. On the other hand, the small sample size likely contributed to some type II errors in path identification, e.g. the study is underpowered to identify the link between Sp1 and *HTR2A* if indeed the path coefficient is in the range shown. Also, the average 24-h urinary free cortisol used in the SEM is not an ideal measurement for modeling the contribution of cortisol to *HTR2A* gene expression in PBMCs.

Potential influences of E47, GR and Sp1 binding sites were considered in our qualitative and quantitative models to provide a mechanistic understanding of the allele-specific expression. In silico computational analyses and competitive EMSA provide evidence for direct binding of transcription factors E47, GR and Sp1 in the *HTR2A* promoter. While EMSA is an accepted method to validate the possibility of functional binding, in vivo binding assays such as chromatin immunoprecipitation (ChIP) should be considered in future studies. We believe that evidence for direct binding to specific sites is sufficient for modeling and integrating genetic and epigenetic risk in complex traits.

As depression is a recognized comorbid condition in CFS, the specificity of this model for CFS versus depression will require further exploration as serotonin signaling is recognized to be important in depressive disorders. To address the contribution of depression, we tested the association of methylation at each site by stratifying this study population based on Zung depression scores and did

not identify an association (data not shown). However, methylation at one of the CpG sites (−1,224) was negatively correlated ( $r = -0.4315$ ,  $P = 0.001$ ) with increasing Zung scores and thus a potential association between *HTR2A* methylation and depression cannot be totally excluded from this study.

Finally, it should be kept in mind that the expression findings were made in PBMCs. Methylation and gene expression are tissue-dependent, and *HTR2A* expression in PBMCs may only be an indirect reflection of CFS pathogenesis. However, the model presented for *HTR2A* promoter methylation and expression in PBMCs is relevant for understanding mechanisms that may be playing a role in *HTR2A* expression in other tissues like the brain (Le-Niculescu et al. 2009).

This study underscores the importance of the coordinated influence of both genetic and epigenetic variations in determining an individual's susceptibility to disease, and the value of integrated functional and computational genomics for simultaneous identification and quantitative evaluation of multiple cis-regulatory elements in major genes of medical interest. To our knowledge, this is the first study to apply SEM to promoter analysis. Novel findings in this study include disease-specific correlations of methylation at adjacent CpG sites and disease-associated allele-specific expression of *HTR2A* that is influenced by both CpG methylation and genotype-dependent transcription factor binding. These regulatory mechanisms of *HTR2A* expression may play significant roles in the pathology of common complex diseases with central nervous system abnormalities such as CFS, depression and PTSD. Confirmation of these findings in an independent study population is required.

**Acknowledgments** Support for VR. Falkenberg was provided by the research participation program at the Centers for Disease Control and Prevention (CDC), National Center for Zoonotic, Vector-Borne Enteric Diseases, Division of Viral and Rickettsial Diseases, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the CDC. The authors acknowledge I. Dimulescu and MM. Khin for their laboratory support and WC Reeves for his support and encouragement with this research project.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., et al. (2005). MatInspector and beyond: Promoter analysis based on transcription factor binding sites. *Bioinformatics*, *21*, 2933–2942.
- Cleare, A. J. (2003). The neuroendocrinology of chronic fatigue syndrome. *Endocrine Reviews*, *24*, 236–252.
- Cleare, A. J., Bearn, J., Allain, T., McGregor, A., Wessely, S., Murray, R. M., et al. (1995). Contrasting neuroendocrine responses in depression and chronic fatigue syndrome. *Journal of Affective Disorders*, *34*, 283–289.
- Falkenberg, V. R., & Rajeevan, M. S. (2010). Identification of a potential molecular link between the glucocorticoid and serotonergic signaling systems. *Journal of Molecular Neuroscience*, *41*, 322–327.
- Gopalakrishnan, S., Van Emburgh, B. O., & Robertson, K. D. (2008). DNA methylation in development and human disease. *Mutation Research*, *647*, 30–38.
- Heim, C., Ehler, U., & Hellhammer, D. H. (2000). The potential role of hypocortisolism in the pathophysiology of stress-related bodily disorders. *Psychoneuroendocrinology*, *25*, 1–35.
- Knight, J. C. (2005). Regulatory polymorphisms underlying complex disease traits. *Journal of Molecular Medicine*, *83*, 97–109.
- Lanfume, L., Mongeau, R., Cohen-Salmon, C., & Hamon, M. (2008). Corticosteroid-serotonin interactions in the neurobiological mechanisms of stress-related disorders. *Neuroscience and Biobehavioral Reviews*, *32*, 1174–1184.
- Le-Niculescu, H., Kurian, S. M., Yehyawi, N., Dike, C., Patel, S. D., Edenberg, H. J., et al. (2009). Identifying blood biomarkers for mood disorders using convergent functional genomics. *Molecular Psychiatry*, *14*, 156–174.
- Li, D., & He, L. (2008). Meta-study on association between the monoamine oxidase A gene (MAOA) and schizophrenia. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, *147B*, 174–178.
- Lister-Williams, R. H., Ferrier, I. N., & Young, A. H. (1998). Mood and neuropsychological function in depression: The role of corticosteroids and serotonin. *Psychological Medicine*, *28*, 573–584.
- Miles, J. (2003). A framework for power analysis using a structural equation modelling procedure. *BMC Medical Research Methodology*, *3*, 27.
- Myers, R. L., Airey, D. C., Manier, D. H., Shelton, R. C., & Sanders-Bush, E. (2007). Polymorphisms in the regulatory region of the human serotonin 5-HT<sub>2A</sub> receptor gene (*HTR2A*) influence gene expression. *Biological Psychiatry*, *61*, 167–173.
- Norton, N., & Owen, M. J. (2005). *HTR2A*: association and expression studies in neuropsychiatric genetics. *Annals of Medicine*, *37*, 121–129.
- Parsons, M. J., D'Souza, U. M., Arranz, M. J., Kerwin, R. W., & Makoff, A. J. (2004). The -1438A/G polymorphism in the 5-hydroxytryptamine type 2A receptor gene affects promoter activity. *Biological Psychiatry*, *56*, 406–410.
- Philibert, R. A., Sandhu, H., Hollenbeck, N., Gunter, T., Adams, W., & Madan, A. (2008). The relationship of 5HTT (SLC6A4) methylation and genotype on mRNA expression and liability to major depression and alcohol dependence in subjects from the Iowa Adoption Studies. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, *147B*, 543–549.
- Polesskaya, O. O., Aston, C., & Sokolov, B. P. (2006). Allele C-specific methylation of the 5-HT<sub>2A</sub> receptor gene: Evidence for correlation with its expression and expression of DNA methylase DNMT1. *Journal of Neuroscience Research*, *83*, 362–373.
- Porter, R. J., Gallagher, P., Watson, S., & Young, A. H. (2004). Corticosteroid-serotonin interactions in depression: A review of the human evidence. *Psychopharmacology, (Berl)*, *173*, 1–17.
- Rajeevan, M. S., Swan, D. C., Duncan, K., Lee, D. R., Limor, J. R., & Unger, E. R. (2006). Quantitation of site-specific HPV 16 DNA methylation by pyrosequencing. *Journal of Virological Methods*, *138*, 170–176.

- Reeves, W. C., Wagner, D., Nisenbaum, R., Jones, J. F., Gurbaxani, B., Solomon, L., et al. (2005). Chronic fatigue syndrome—a clinically empirical approach to its definition and study. *BMC Medicine*, *3*, 19.
- Reiner, A., Yekutieli, D., & Benjamini, Y. (2003). Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*, *19*, 368–375.
- Schreiber, J. B. (2008). Core reporting practices in structural equation modeling. *Research in Social and Administrative Pharmacy*, *4*, 83–97.
- Smith, A. K., Dimulescu, I., Falkenberg, V. R., Narasimhan, S., Heim, C., Vernon, S. D., et al. (2008). Genetic evaluation of the serotonergic system in chronic fatigue syndrome. *Psychoneuroendocrinology*, *33*, 188–197.
- Smith, A. K., Maloney, E. M., Falkenberg, V. R., Dimulescu, I., & Rajeevan, M. S. (2009). An angiotensin-1 converting enzyme polymorphism is associated with allostatic load mediated by C-reactive protein, interleukin-6 and cortisol. *Psychoneuroendocrinology*, *34*, 597–606.
- Sorensen, B., Jones, J. F., Vernon, S. D., & Rajeevan, M. S. (2009). Transcriptional control of complement activation in an exercise model of chronic fatigue syndrome. *Molecular Medicine*, *15*, 34–42.
- Uher, R., & McGuffin, P. (2008). The moderation by the serotonin transporter gene of environmental adversity in the aetiology of mental illness: Review and methodological analysis. *Molecular Psychiatry*, *13*, 131–146.
- Vernon, S. D., & Reeves, W. C. (2006). The challenge of integrating disparate high-content data: Epidemiological, clinical and laboratory data collected during an in-hospital study of chronic fatigue syndrome. *Pharmacogenomics*, *7*, 345–354.
- Zhu, Q. S., Chen, K., & Shih, J. C. (1995). Characterization of the human 5-HT<sub>2A</sub> receptor gene promoter. *Journal of Neuroscience*, *15*, 4885–4895.