



Published in final edited form as:

Neuroimage. 2016 January 15; 125: 813–824. doi:10.1016/j.neuroimage.2015.10.033.

Integrative Bayesian Analysis of Neuroimaging-Genetic Data with Application to Cocaine Dependence

Shabnam Azadeh^{1,2}, Brian P. Hobbs¹, Liangsuo Ma^{3,8}, David A. Nielsen^{5,6,7}, F. Gerard Moeller^{4,8}, and Veerabhadran Baladandayuthapani^{1,*}

¹Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX

²School of Public Health, The University of Texas Health Science Center, Houston, TX

³Department of Radiology, Virginia Commonwealth University, Richmond, VA

⁴Department of Psychiatry, Pharmacology and Toxicology, Virginia Commonwealth University, Richmond, VA

⁵Menninger Department of Psychiatry and Behavioral Sciences, Houston, TX

⁶Baylor College of Medicine, Houston, TX

⁷Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX

⁸The Institute for Drug and Alcohol Studies, Virginia Commonwealth University, Richmond, VA

Abstract

Neuroimaging and genetic studies provide distinct and complementary information about the structural and biological aspects of a disease. Integrating the two sources of data facilitates the investigation of the links between genetic variability and brain mechanisms among different individuals for various medical disorders. This article presents a general statistical framework for integrative Bayesian analysis of neuroimaging-genetic (iBANG) data, which is motivated by a neuroimaging-genetic study in cocaine dependence. Statistical inference necessitated the integration of spatially dependent voxel-level measurements with various patient-level genetic and demographic characteristics under an appropriate probability model to account for the multiple inherent sources of variation. Our framework uses Bayesian model averaging to integrate genetic information into the analysis of voxel-wise neuroimaging data, accounting for spatial correlations in the voxels. Using multiplicity controls based on the false discovery rate, we delineate voxels associated with genetic and demographic features that may impact diffusion as measured by fractional anisotropy (FA) obtained from DTI images. We demonstrate the benefits of accounting for model uncertainties in both model fit and prediction. Our results suggest that cocaine consumption is associated with FA reduction in most white matter regions of interest in the brain. Additionally, gene polymorphisms associated with GABAergic, serotonergic and dopaminergic neurotransmitters and receptors were associated with FA.

*Corresponding Author, veera@mdanderson.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Bayesian analysis; diffusion tensor imaging; model averaging; voxel-level inference

1 Introduction

Cocaine use is associated with several acute effects on a variety of intracellular pathways such as inhibition of the serotonin, dopamine, and norepinephrine transporters (Han and Gu, 2006). Chronic use of cocaine leads to addiction, which inflicts high costs to individuals and society. Long-term drug addiction is known to cause many health problems, both psychological (depression, paranoia, anxiety) and physical (cardiac disease). Although genetics contribute substantially to addiction vulnerability, the identification of genes associated with this susceptibility has been slow (Li and Burmeister, 2009). Several studies have suggested that certain DNA polymorphisms may impact the extent to which a person is vulnerable to drug abuse and addiction (Nielsen and Kreek, 2012; Kreek et al., 2005b), and may have significant effects on an individual's response to treatment (Berrettini and Lerman, 2005; Kreek et al., 2005a; Nielsen et al., 2014; Bauer et al., 2014).

Recent advances in imaging technologies such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and diffusion tensor imaging (DTI) have enabled non-invasive investigation of neurochemical, functional and structural alterations in the brains of drug-addicted subjects, revealing new insights into the mechanism of addiction (Volkow et al., 2003). Chronic cocaine users have been shown to exhibit subtle abnormalities in different areas of the brain, as measured by DTI (Lim et al., 2008; Moeller et al., 2007; Lane et al., 2010; Moeller et al., 2005; Ma et al., 2009; Romero et al., 2010; Liu et al., 2008). These areas include the anterior and posterior corpus callosum and tracts in the frontal and parietal regions of the brain. In addition, cocaine has been shown to alter the methylation of myelin genes in rodents (Nielsen et al., 2012b). However, the precise mechanisms that cause alterations in white matter remain unknown.

The search for genetic factors associated with a disease is complicated due to the intrinsic confluence of genetic, epigenetic (Nielsen et al., 2012c), environmental and psychological components as well as additional factors that may not have been defined (Craig et al., 2008). Understanding the links between genetics, brain structure and function is an inherently multidisciplinary endeavor that has broad implications for a number of medical disorders, including neurological conditions such as Alzheimer's disease (Kohannim et al., 2011; Braskie et al., 2011), autism (Dennis et al., 2011), bipolar disorder (Frazier et al., 2014), obesity (Ho et al., 2010), and schizophrenia (Braskie et al., 2012).

Statistical integration of imaging and genetic data for investigating brain-wide, candidate-gene associations is challenging because the method of inference must account for (i) spatial dependence among the high-dimensional measurements of imaging phenotypes; (ii) multiplicity adjustments among the numerous tests; and (iii) model uncertainty in the presence of a large set of genetic factors. For example, even with a relatively modest number of genetic variables, such as 23, our study yields a prohibitively large number of possible models (more than 8 million) for conducting inference at a single voxel. Moreover, we

expect associations to be spatially heterogeneous across brain regions due to varied functional and physical features, which thereby precludes inference using any single model applied uniformly to all voxels in the brain (which is typically in the hundreds of thousands).

Various statistical methods have been used to examine relationships between genetics and brain imaging features. Multivariate generalized linear modeling is commonly used to identify associations among multivariate phenotypes and candidate genotypes (Chung et al., 2010; Taylor and Worsley, 2008; Worsley et al., 2004). Component-wise methods offer alternatives that use univariate analysis for each component, with multiplicity control through *post hoc* application of Bonferroni correction, false discovery rates, or random field theory to adjust for multiple comparisons (Heller et al., 2007; Lazar et al., 2002). Principal component analysis is another approach that is used to reduce the dimension of the multivariate phenotype (Formisano et al., 2008; Teipel et al., 2007; Rowe and Hoffmann, 2006; Kherif et al., 2002). In addition, sparse reduced rank regression is often implemented in brain-wide, genome-wide association studies (Vounou et al., 2010). Partial least squares regression can be used to identify linear projections of the multivariate phenotype and genotype on lower dimensional spaces (Chun and Kele, 2010; Krishnan et al., 2011). The effects of multiple genes can be modeled nonparametrically using least squares kernel-based learning methods (Liu et al., 2007; Ge et al., 2012).

The aforementioned statistical methods currently available for voxel-wise analysis fail to account for uncertainty in model selection, and thus are limited for detecting associations between brain features and a large set of genetic covariates that may vary over the volume of interest. In addition, these methods do not allow for the explicit quantification of model uncertainties. To overcome these limitations, we developed a statistical framework that enables probabilistic inference for voxel-wise analysis that leverages Bayesian model averaging procedures. This allows us to not only quantify the genetic associations, but also to explicitly calculate uncertainties in terms of posterior probabilities, which admits precise multiplicity controls for coherent inference and variable selection. Briefly, our framework for integrative Bayesian analysis of neuroimaging-genetic (iBANG) features circumvents the high-dimensional problem by decoupling model fitting and inference. We first conduct a voxel-based Bayesian model averaging procedure, in parallel, to generate posterior probability maps for each genetic variable spanning the entire brain. Subsequently, we incorporate spatial information into our inferential techniques by locally smoothing the probability maps. Finally, we demarcate specific regions of association using multiplicity-corrected false discovery rates. This allows our methods to scale to large datasets as well as provide principled statistical inference while accounting for various sources of variability induced by model selection and spatial heterogeneity. Furthermore, we demonstrate the benefits of accounting for model uncertainties in both model fit and prediction. Figure 1 is a schematic representation of our modeling strategy.

Our methods are applied to data from a recent cocaine addiction study (see Section 2 for details) that sought to identify brain regions that exhibit strong evidence of differential diffusion patterns among candidate genetic variants, cocaine users, and demographic covariates using FA in the white matter space as defined by the Montreal Neurological

Institute (MNI). Our framework is general enough to be applied to any imaging modality to conduct voxel-wise analyses of specific imaging parameters and outcomes of interest.

The structure of our article is as follows: Section 2 summarizes the main aspects of the cocaine addiction study, including the target population and acquisition of brain images and genetic data. Section 3 details the iBANG methodology and Section 4 presents the neurobiological findings from the integrative analysis. Section 5 provides a discussion of the assumptions, implications, and possible extensions of our work. Additional technical, computational and analytic details and results are provided in the supplementary materials.

2 Motivating Data

2.1 Study population

Study participants were recruited to the Center for Neurobehavioral Research on Addictions at the University of Texas Health Science Center at Houston through advertisements for research volunteers. All individuals with cocaine dependence and all non-drug users (controls) were screened for psychiatric and non-psychiatric medical disorders using the Structured Clinical Interview for DSM-IV (First et al., 1996). The study population included 39 cocaine users (29 male and 10 female users) and 19 control participants (12 male and 7 female controls). Table 1 lists the characteristics of the study population. This study was approved by the Institutional Review Boards of the University of Texas Health Science Center and Baylor College of Medicine, as well as the Research and Development committee of the Michael E. DeBakey Veteran Affairs Medical Center.

2.2 Brain image acquisition and processing

Whole brain diffusion-weighted images (DWI) were acquired on a Philips 3.0 T Intera system with a six-channel phased array receiver head coil (Philips Medical Systems, Best, Netherlands). Images were acquired in the transverse plane using a single-shot spin-echo diffusion sensitized echo-planar imaging (EPI) sequence (bfactor=1000 s/mm^2 , repetition time=6100 ms, echo time=84 ms, 44 contiguous axial slices, field-of-view= 240 mm \times 240 mm, 112 \times 112 acquisition matrix, 256 \times 256 reconstructed matrix, 0.9375 mm \times 0.9375 mm reconstructed in-plane resolution, slice thickness=3 mm). The diffusion tensor encoding scheme is based on the uniformly distributed and balanced rotationally invariant Icosa21 tensor-encoding set (21 gradient directions) (Hasan et al., 2001). A SENSE acceleration factor or kspace under sampling was set to be R=2 to reduce EPI image distortions. The diffusion-encoded volumes were acquired with fat suppression. The acquisition time was approximately 7 minutes and resulted in signal-to-noise ratio independent DTI-measure estimation (Hasan, 2007). All the DTI images were processed using the FMRIB Software Library (www.fmrib.ox.ac.uk/fsl, version 5.04) (Jenkinson et al., 2012). For each scan, the DWI images were corrected for eddy current distortions and head motion (Jenkinson and Smith, 2001) after converting the Philips DICOM files into NIFTI format using dcm2nii. Then, brain was extracted from the images using FSL's Brain Extraction Tool (BET) (Smith, 2002). Next, the FMRIBs Diffusion Toolbox (FDT/FSL) (Behrens et al., 2003) was used to fit the data to extract FA and other DTI parameters for each voxel. All the FA images were

aligned to the standard MNI space using the FSL's nonlinear registration (Andersson et al., 2007a,b) with the FMRIB58-FA template image.

2.3 Genetic data acquisition

Twenty-one candidate genetic variants in seventeen genes that we had hypothesized may play a role in the vulnerability to develop addiction, have previously been associated with addiction vulnerability, known to be involved in psychiatric morbidities, or in neurotransmitter pathways were examined. These have been assessed in prior studies (Nielsen et al., 2012a; Kosten et al., 2013b,a; Spellicy et al., 2013; Kosten et al., 2013b, 2014; Anastasio et al., 2014; Spellicy et al., 2014; Frazier et al., 2014; Liu et al., 2014; Brewer III et al., 2015). This includes several polymorphisms in the dopamine and serotonin transporters, and in the norepinephrine postsynaptic receptor. Genotypes were coded as 0 or 1 following the dominant model. The list of genetic variants, genetic descriptions and genotype coding are given in Table 2.

3 Methods

In this section, we introduce iBANG, a method for integrative Bayesian analysis of neuroimaging-genetic data. The statistical method is motivated by the use of FA values acquired in voxels in the MNI white matter space that spans the entire brain. However, the framework is quite general and can be used for integrative analysis of any continuous image-derived parameters.

3.1 iBANG model formulation

Suppose that the imaging feature (e.g., FA values) at voxel $\nu = 1, \dots, V$ for subject $i = 1, \dots, n$ is represented as $y_i(\nu)$ and the corresponding (k -dimensional) genetic and demographic variables are represented as x_{i1}, \dots, x_{ik} . Assume $\mathbf{Y}(\nu)$ is a column-wise matrix of imaging features $\{y_1(\nu), \dots, y_n(\nu)\}$, \mathbf{X} is a $n \times k$ matrix, and $\boldsymbol{\beta}(\nu)$ is the full k -dimensional vector of regression coefficients. The linear regression model for the ν th voxel is

$$\mathbf{Y}(\nu) = \alpha + \mathbf{X}\boldsymbol{\beta}(\nu) + \sigma\boldsymbol{\varepsilon}, \quad (3.1.1)$$

where α is an intercept, $\sigma \in \mathbb{R}_+$ is a scale parameter, and $\boldsymbol{\varepsilon}$ is random noise (accounting for unknown sources of variation), which follows an n -dimensional normal distribution with zero mean and the identity covariance matrix. Our primary construct for inference is the effect surfaces $\boldsymbol{\beta}(\nu)$'s that capture the associations between the imaging features and each of the k covariates across the brain. However, this requires estimation of $V \times k$ number of parameters, which in our case is $3 \times 10^6 \times 24 \approx 72$ million parameters without accounting for model uncertainty, which represents considerable analytical and computational challenges. Subsequently, accounting for model uncertainty over all possible models, this number increases exponentially to $3 \times 10^6 \times 2^{24}$. While Bayesian model averaging (BMA) involves more effective parameters, by explicitly accounting for model uncertainty, BMA both shrinks the influence of certain variables to zero through the model weights and provides an

unified method of inference for all voxels. Therefore, the BMA approach when applied in this context eschews the dilemma posed by implementation of voxel-wise model selection procedures, which can yield a joint inference that utilizes discordant models at adjacent voxels, and thereby neglects spatial dependence in the model space. To address this, we decouple the model fitting and inference using a three-step component-wise analysis pipeline:

- Step I: Estimate the association between each genetic and demographic variable via voxel-based Bayesian model averaging and obtain posterior probability maps (PPMs) of the associated genetic and demographic features for each covariate.
- Step II: Incorporate spatial information pertaining to voxel locations to smooth the PPMs using prefiltered rotationally invariant nonlocal means to generate smoothed posterior probability maps (sPPMs) of the associated genetic and demographic variables.
- Step III: Conduct rigorous inference on the sPPMs of the genetic and demographic variables using Bayesian false discovery rates to delineate regions of brain activation while controlling for multiplicities.

We detail each of these steps below.

3.1.1 Bayesian model averaging—BMA is a tool used to enumerate model uncertainty by averaging across the best models using each model's posterior probability as weights. BMA considers all 2^k subsets of possible explanatory variables in the analysis but shrinks the influence of certain variables to zero through the model weights (Raftery, 1995). Denoting the space of all 2^k possible models by $\mathcal{M} = \{M_j : j = 1, \dots, 2^k\}$, a specific model M_j has a subset of X_j clinical, demographic and genetic variables, leading to the following reduced equation:

$$Y(\nu) = \alpha + X_j \beta_j(\nu) + \sigma \epsilon \quad (3.1.2)$$

where $\beta_j \in \mathbb{R}^{k_j}$ ($0 < k_j < k$) is the reduced subset of covariates from (3.1.1) with the exclusion of a regressor indicating that the corresponding element of $\beta_{k-k_j}(\nu)$ is zero.

Priors: To conduct Bayesian inference, prior distributions need to be defined for the model space \mathcal{M} and the corresponding parameters β , α and σ . The choice of prior distributions could significantly impact the resulting posterior model probabilities, thereby necessitating careful consideration (Kass and Raftery, 1995; Hoeting et al., 1999; Raftery, 1995; Friston and Penny, 2003). We consider improper noninformative priors for the parameters common to all models, namely α and σ , such that $p(\alpha, \sigma) \propto \sigma^{-1}$, allowing for maximal learning from the data. σ is the residual error variance, and thereby accounts for unknown sources of variability that are not attributable to the covariates. We assume a common prior for σ across the different models, which is the usual approach (Mitchell and Beauchamp, 1988; Raftery et al., 1997). Also, we consider a g -prior structure for $\beta_j(\nu)$ whereby $p\{\beta_j(\nu) | \alpha, \sigma, M_j\}$ is modeled as a k_j -dimensional normal distribution with mean zero and covariance matrix of

$\sigma^2(g\mathbf{X}_j'\mathbf{X}_j)^{-1}$, where $g = 1/\max\{n, k^2\}$. The choice of g is based on theoretical considerations to guarantee asymptotic consistency for selecting the correct model, i.e., ensuring that the posterior probability of the correct model approaches 1 for increasing sample size (Fernandez et al., 2001a). Finally for model selection, we consider a prior distribution over all 2^k possible models. We denote the model probability for the j th model by $p(M_j) = p_j, j = 1, 2, \dots, 2^k$. We assume a uniform distribution on the model space to avoid *a priori* model preference in the absence of prior knowledge.

Posterior computations: The posterior distribution of $\beta_j(\nu)$ can be written as

$$\mathcal{P}\{\beta_j(\nu) | \mathbf{Y}(\nu)\} = \sum_{j=1}^{2^k} \mathcal{P}\{\beta_j(\nu) | \mathbf{Y}(\nu), M_j\} \mathcal{P}\{M_j | \mathbf{Y}(\nu)\}. \quad (3.1.3)$$

The second term $\mathcal{P}\{M_j | \mathbf{Y}(\nu)\}$, the posterior model-specific probability is calculated as follows

$$\mathcal{P}\{M_j | \mathbf{Y}(\nu)\} = \frac{\{\mathcal{L}_{(\nu)}(M_j)\} P(M_j)}{\sum_{j=1}^{2^k} \{\mathcal{L}_{(\nu)}(M_h)\} P(M_h)}, \quad (3.1.4)$$

where the marginal likelihood of model M_j , which we denote by $\mathcal{L}_{(\nu)}(M_j)$, is

$$\mathcal{L}_{(\nu)}(M_j) = \int p\{\mathbf{Y}(\nu) | \alpha, \beta_j(\nu), \sigma, M_j\} p(\alpha, \sigma) p\{\beta_j(\nu) | \alpha, \sigma, M_j\} d\alpha d\beta_j(\nu) d\sigma, \quad (3.1.5)$$

where $p\{\mathbf{Y}(\nu) | \alpha, \beta_j(\nu), \sigma, M_j\}$ represents the sampling model, (3.1.2), $p(\alpha, \sigma)$ and $p\{\beta_j(\nu) | \alpha, \sigma, M_j\}$ are the priors distributions for the intercept, scale, and regression coefficients, respectively. The marginal likelihood can be obtained analytically (Fernandez et al., 2001a) as follows:

$$\mathcal{L}_{(\nu)}(M_j) \sim \left(\frac{g}{g+1}\right)^{k_j/2} \left(\frac{1}{g+1} \mathbf{Y}(\nu)' D_{X_j} \mathbf{Y}(\nu) + \frac{g}{g+1} (\mathbf{Y}(\nu) - \overline{\mathbf{Y}(\nu)})' (\mathbf{Y}(\nu) - \overline{\mathbf{Y}(\nu)})\right)^{-(n-1)/2}, \quad (3.1.6)$$

where $D_{X_j} = I_n - \mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'$ and $\overline{\mathbf{Y}(\nu)}$ is the sample mean of response variable. We use Markov chain Monte Carlo (MCMC) based methods to estimate all model parameters and posterior probabilities. Details of the computational algorithms are briefly described

below. Note that the BMA procedure is performed voxel-wise and hence can run simultaneously for each voxel, which allows the algorithm to be parallelized.

Estimation of model parameters and posterior probabilities using MCMC: In practice, computing the posterior probabilities is a challenge due to the large amount of summations involved in calculating the posterior probability maps (PPMs). In our application, we have $k=24$ possible clinic/demographic and genetic variables. Hence, we need to calculate the posterior probabilities for each $2^{24} = 16777216$ models and average the required distributions over all models. To ease the computational effort, we adopt the approach of (Madigan et al., 1995) and approximate the posterior distribution on the model space \mathcal{M} by applying MCMC frequencies obtained from a Metropolis-based sampler. Assuming that the chain is currently at model M_s which contains X_s regressors, and X_j denotes a (single) regressor which could include or exclude one of X_s regressors. A new model M_j was proposed randomly by sampling from an uniform distribution on the space containing M_s and all models with one variable more or less than M_s . So the number of regressors in M_j is $X_s \pm X_j$. Two options are effectuated at each iteration of the sampling algorithm. The chain could move to M_j with probability $pr = \min \left\{ 1, \frac{\mathcal{L}(M_j)P(M_j)}{\mathcal{L}(M_s)P(M_s)} \right\}$ or remain at model M_s with probability $1 - pr$ (see e.g. Fernandez et al., 2001b; Raftery et al., 1997, for implementations in the context of linear models). Additional details describing the MCMC software implementation and performance metrics such as mixing and computational times are provided in Section S1 of the supplementary materials.

3.1.2 Smoothing posterior probability maps by prefiltered rotationally invariant non-local means—The BMA procedure above yields voxel-wise PPMs of the clinical, demographic and genetic variables. To account for spatial correlation among voxels that is inherent to the PPMs, a post-hoc smoothing procedure is conducted to generate smoothed posterior probability maps (sPPMs) by borrowing strength from neighboring voxels. Standard (linear) smoothing approaches such as Gaussian kernels could be adopted here. While approaches based on Gaussian filters are capable of reducing image noise, especially in homogeneous areas, they tend to remove some high-frequency signal components as well, which produces images with blurred edges (Ashburner and Friston, 2000). In our context, this could have implications for identification of the resulting significant regions, as a simple Gaussian smoother may yield posterior probability maps with blurred edges.

We follow the method proposed by (Manjón et al., 2012) via prefiltered rotationally invariant nonlocal means (PRI-NLM3D). This method maintains *domain and scale symmetry* i.e since the raw (non-smoothed) posterior probabilities lie in the interval (0, 1) this method maps the (smoothed) probabilities on same interval (0, 1). This is especially helpful since all of our downstream analyses are based on the smoothed probabilities maps. Moreover, from the scale perspective, this technique is similar to the processing methods, since the original FA values are also in (0, 1).

This technique is a combination of the Oracel-based 3D discrete cosine transform (3D-DCT) and rotationally invariant nonlocal means (RI-NLM). The method offers several advantages,

namely that it provides better smoothing performance around the edges and benefits from faster smoothing time in comparison to Oracel-based 3D-DCT and block-wise NLM. Let $\mathcal{P}(\nu)$ denote the PPM of any variable on voxel ν . The PPM can be written as a sum of sPPMs, $\mathcal{P}_s(\nu)$, plus white noise corresponding to voxel-based BMA analysis, which leads to:

$$\mathcal{P}(\nu) = \mathcal{P}^s(\nu) + \eta \quad (3.1.7)$$

The smoothing process includes two steps:

- Step I: Pre-smoothing based on a three-dimensional moving-window DCT hard thresholding.
- Step II: Smoothing the pre-smoothed PPMs using a three-dimensional rotationally invariant version of the nonlocal means.

The main concepts of each step are explained in sequence below.

3D discrete cosine transform smoothing: The DCT is a similarity transformation. The advantage of this technique is robustness and absence of any assumptions concerning PPM statistics beyond sparsity (Guleryuz, 2007). Let $\mathcal{P}_1^s(\nu)$ denote a pre-smoothed posterior. A 3D block DCT (4×4×4 block size, which indicates data cubes) is used. We apply a hard threshold to obtain the local smoothed estimation at block b , $\hat{\mathcal{P}}_b^s$, which is presented as

$$\hat{\mathcal{P}}_b^s = H^{-1}(\hat{c}_b), \quad \text{where } c_b = H(\mathcal{P}_b), \quad \text{and } \hat{c}_b = T(c_b, \tau), \quad (3.1.8)$$

where H represents a 3DDCT, c_b shows the transform coefficients for block b , and T is the hard thresholding operator with threshold τ . By combining all overlapping B blocks at the position of ν and following the weighted rule, we can derive the local estimator of $\hat{\mathcal{P}}_1^s(\nu)$:

$$\hat{\mathcal{P}}_1^s(\nu) = \frac{\sum_{b=1}^B \theta_b \hat{\mathcal{P}}_b^s(\nu)}{\sum_{b=1}^B \theta_b}, \quad \text{where } \theta_b = \frac{1}{1 + \|\hat{c}_b\|_0}, \quad (3.1.9)$$

and B represents the number of overlapping blocks used to calculate $\hat{\mathcal{P}}_1^s(\nu)$, and θ_b denotes a weight for block b , which is a propositional inverse of the $\hat{c}_b L_0$ norm. L_0 norm of \hat{c}_b corresponds to the number of nonzero coefficients of block b after the hard thresholding.

Nonlocal means smoothing: Nonlocal means (NLM) smoothing is defined on the basis of voxels with similar neighborhoods (small 3D volumetric patches) that tend to have similar posterior probability values (Buades et al., 2005). The NLM estimator of the smoothed posterior probability at voxel ν is defined as

$$\hat{\mathcal{P}}_2^s(\nu) = \frac{\sum_{\nu' \in \Omega} w(\nu, \nu') \mathcal{P}(\nu')}{\sum_{\nu' \in \Omega} w(\nu, \nu')}, \quad \text{where} \quad w(\nu, \nu') = e^{-\frac{\|N_\nu - N_{\nu'}\|}{h^2}}, \quad (3.1.10)$$

and Ω represents the search volume, $w(\nu, \nu')$ represent the similarity weights for any two 3D patches N_ν and $N_{\nu'}$ centered around voxels ν and ν' , and h^2 denotes the bandwidth, which determines the extent of the smoothing.

We use the PRI-NLM3D method to obtain $\hat{\mathcal{P}}^s(\nu)$, providing an sPPM for each clinical, genetic and demographic variable. The PRI-NLM3D method combines and extends the above steps. We implemented smoothing using an Oracle-based 3D-DCT to find the pre-smoothed PPM of each of the clinical and genetic variables and then applied the rotational invariant NLM to derive the sPPM, $\hat{\mathcal{P}}^s(\nu)$, for each variable (Manjón et al., 2012). We used the extended version DCT filter which is an Oracle-based DCT filter. It only has one parameter τ , and as recommended by Manjon et al (2012), we use $\tau = 2.7 * \sigma$ where σ is the estimated from the data as the standard deviation of random noise. The approach is similar to Discrete cosine transforms and wavelet thresholding methods of (Mallat, 1999). Additionally, we assume noise follows a Rician distribution (Nowak, 1999). The procedure is computationally inexpensive, with computation times of less than one minute when implemented using the PPMs that resulted from the cocaine study, for each of the 24 maps consisting of close to 300,000 voxels.

3.1.3 Bayesian false discovery rate controls for identification of significant voxels—

The primary objective of this type of integrative analysis is to identify brain regions that exhibit strong evidence of differential activation patterns in the presence or absence of genetic and clinical and demographic factors. After implementing the Bayesian model averaging and smoothing steps, we obtain a smoothed posterior probability map for each variable spanning the entire brain region. Owing to the high-dimensional nature of these neuroimaging phenotypes, the process for delineating significant voxels requires hundreds of thousands of tests, thereby running the risk of detecting thousands of false positive findings by random chance alone, in the absence of adequate multiplicity corrections. The false discovery rate (FDR), or the expected proportion of false positives voxels among all voxels deemed significant, can be controlled in this context through the specification of a threshold for the Bayesian posterior p-value (sometimes referred to as a q-value), (Storey, 2003). The generation of sPPMs from the aforementioned modeling steps admits a straightforward and computationally efficient strategy to implement FDR controls. Specifically, we assume that all voxels for which the resulting sPPM exceeds a given threshold ϕ , $\hat{\mathcal{P}}^s(\nu) > \phi$, characterize locations for which the imaging phenotype is significantly affected by changes in a given genetic variant. Considering the cocaine study, let $\chi_\phi = \{\nu : \hat{\mathcal{P}}^s(\nu) > \phi\}$ represents the set of all significant voxels for a given genetic variant. The significant threshold ϕ can be set to control the Bayesian FDR (Morris et al., 2008; Baladandayuthapani et al., 2010). Suppose that ϕ_δ denotes the threshold value that controls the overall average FDR at level δ . Posterior inference using ϕ_δ implies that given the available evidence, we expect only 100 δ % of the total number of voxels that are declared

significant to represent false-positive locations for which the imaging phenotype is truly unaffected by a given covariate. For all sPPMs, we sort $\hat{\rho}_{\kappa} = \hat{\rho}^s(\nu_{\kappa})$ in ascending order to yield $\hat{\rho}_{(\kappa)}$, $\kappa = 1, \dots, \nu$. Assume κ^* is a count of the number of voxels, with sPPM

exceeding the threshold. Then, $\phi_{\delta} = \hat{\rho}(\xi)$, where $\xi = \max \{ \kappa^* : \kappa^{*-1} \sum_{\kappa=1}^{\kappa^*} 1 - \hat{\rho}_{(\kappa)} \leq \delta \}$. We use $\chi_{\phi_{\delta}}$ to denote the final set of voxel locations that yield evidence of an effect that significantly impacts the FA based on the average Bayesian FDR control at level δ . Note that ϕ may also be specified in consideration of the utility/loss functions such as relative cost of a false-positive versus a false-negative error (Müller et al., 2004).

4 Application to neuroimaging-genetic data in cocaine addiction study

In this section, we present the results of our integrative analysis of the neuroimaging-genetic data obtained from the cocaine addiction study. The iBANG method described in Section 3 was used to identify regions of brain that exhibit strong evidence for differential diffusion patterns among the candidate genetic variants and demographic variables. Before model fitting, each patient's scan was registered to a common brain. Voxels contributing FA values that exceed 0.2 were used in the analysis to capture white matter regions of brain. This resulted in a total of 293,318 voxels for downstream analyses. The analysis involved the three sequential components as described in the methods section. Implementation of step 1 used the BMS package in statistical software R (see Section S1 of the supplementary materials for more details). Step 2 was implemented using the MRI Denoising Matlab toolbox (Manjón et al., 2012). Multiplicity correction was used thereafter to control the rate of false positives at 10%. The elapsed time for analysis of one voxel was 3.894 seconds. For ease of exposition, our results are summarized in the following order. In section 4.1, we present the number of voxels for which a significant association with FA was identified for each genetic variant and demographic feature. In Section 4.2, we summarize the results for the brain regions of interest (ROIs), defined by a well-established white matter atlas. We also provide results obtained from two-sided hierarchical clustering of the ROIs and genetic variants. In Section 4.3, we report the dominant magnitude and directions of the effect on FA alteration for significant genetic variants and demographic feature within each the ROI.

4.1 Genetic variants and clinical and demographic features associated with FA

Table 3 presents the resultant numbers of voxels for which statistically significant differential expression of mean FA was evident between the factor levels of each candidate genetic variant or clinical/demographic feature.

Genetic variants associated with FA: Our study suggests that the impact of *GAD1*^a on diffusion in the white matter of the brain was extensive in comparison to the other 20 genetic variants. A total of 5217 voxel locations were found to be significantly associated with *GAD1*^a, nearly 4 times more than the second most influential genetic variant, *HTR2A*.

Figure 2 depicts the multi-slice sagittal views of the neuroanatomic locations of *significant regions* (SRs) in the white matter of the brain that were impacted by *GAD1*^a and *GAD1*^b respectively. In addition, 1207 voxels were identified as significant for *GAD1*^b. In Section S5 of supplementary materials, Figure S9 displays more anatomic locations of SRs for

GAD1^a. Both *GAD1^a* and *GAD1^b* code for glutamic acid decarboxylase (*GAD*), a rate-limiting enzyme in synthesis of *GABA* in inhibitory interneurons. *GABA* is the major inhibitory neurotransmitter in the brain, GABAergic neurotransmission plays a critical role in drug-reward and drug-seeking behavior (Hyman and Malenka, 2001). Expression of the *GAD1* gene is highly regulated by neuronal activity in the prefrontal cortex of the brain in cocaine users (Enoch et al., 2012). SRs were also evident in more than 1000 voxels for each of the following genetic variants: *HTR2A*, a 5-*HT* receptor subtype; and *SLC6A4^b*, a mediator of reuptake of 5-*HT*; *TH* involved in the synthesis of dopamine; *SLC6A3^b*, a mediator of reuptake of dopamine; and *ADRA1A*, a norepinephrine postsynaptic receptor. In Section S5 of supplementary materials, each row of Figure 10 displays the anatomic locations of SRs for one of these genetic variants.

Clinical and demographic features associated with FA: Diffusion in the white matter of the brain as measured by FA was significantly impacted by cocaine abuse to the extent of 3100 voxels. Figure 3 depicts the neuroanatomic regions for which significant differences in the mean FA were evident between cocaine users and healthy controls. In Section S5, Figure S11 of supplementary materials more anatomic locations of SRs for cocaine abuse is displayed. In addition, both demographic variables (age and gender) significantly impacted at least 1000 voxels.

We evaluated the partial effects of cocaine consumption in the presence of the of other independent variables, consisting of gene variants and demographic features. Our study confirms previous studies that have found that numerous anatomical brain regions involved in the induction and long-term sensitization to cocaine are impacted by the consumption of the substance (Stein and Fuller, 1993; Hammer and Cooke, 1994; Hadfield, 1995). Several imaging studies have demonstrated that cocaine consumption leads to alterations in neuronal activity in the prefrontal cortex (Beyer and Steketee, 1999; Jentsch and Taylor, 1999; Vanderschuren and Kalivas, 2000).

4.2 Cluster analysis of ROIs, genetic variants, and clinical and demographic features

We further refined our inference using the white matter atlas developed by the Johns Hopkins University (JHU) (Wakana et al., 2004). The atlas enabled us to map the locations of significant voxels into the 48 white matter regions of interest (ROIs) detailed in Table 4.

Cluster analyses: Two-sided hierarchical clustering of ROIs and genetic variants, and clinical/demographic features was implemented using counts of significant voxels obtained within each ROI as defined by the JHU atlas, with the aim of better understanding functional interdependencies among genes and regions of the brain. Ward's method was used to minimize the total within-cluster variance (Szekely and Rizzo, 2005). Figure 4 presents the results of the cluster analysis, which resulted in four clusters among genetic-demographic features and three clusters of JHU ROIs. Table 5 reports cluster assignments among the white matter ROIs (top) and genetic variants (bottom). Genetic variants belonging to the same genes predominately clustered together, such as *SLC6A4^b* and *SLC6A4^a* in cluster one, and *GAD1^a* and *GAD1^b* in cluster three.

4.3 Magnitude and direction of FA alteration in ROIs

Figures 5 and 6 depict the extent and nature of FA alterations among the candidate genetic variants and clinical and demographic features considered within each ROI of the JHU white matter atlas. In Figure 5, we plot the proportion of significant voxels with positive regression coefficients. Green represents FA enhancement; yellow characterizes the absence of FA enhancement. In Figure 6, red depicts the proportion of voxels with negative regression coefficients, suggesting that these features are predominately associated with FA attenuation. In both Figures, white represents the absence of significant voxels in that particular brain region-by-feature combination. Figure 5 suggests that both $GADI^a$ and $GADI^b$ are associated with FA enhancement in most regions defined by the JHU white matter atlas; whereas Figure 6 demonstrates an association with FA diminishment in cocaine users.

4.4 Comparison of Bayesian model averaging versus full model

We also compared our Bayesian model averaging (BMA) based inference to a full Bayesian model (Full) fit with no model averaging – to determine the necessities/gains of accounting the uncertainty inherent in the model selection process. To this end, we undertook several, more formal comparisons to describe the extent of relative differences between BMA and Full model inferences on the basis of several metrics such as goodness-of-fit measures, predictive performance, significant voxels, and point estimation of the regression coefficients.

Specifically, we compared BMA and Full model inference using two goodness-of-fit metrics: (i) approximate Deviance information criterion (aDIC), and (ii) Bayesian information criterion (BIC), which are indicative of the (relative) fit of our models to the observed data. Using both metrics we found that the BMA-based methods consistently outperformed the full model (see Section S2 of the supplementary materials for more details). We further evaluated the predictive performance of our models using log predictive scores (LPS). We evaluated the predictive performance of LPS considering four different test-training scenarios by splitting the sample to test and train sets of {33%, 25%, 20%, 10%}. In all scenarios, the BMA-based LPS scores are lower than the Full model, thus indicating better predictive performance (see Section S3 of the the supplementary materials for more details). Finally, we compared the regression coefficient estimates of BMA and Full model the entire white matter of the brain for significant factors: cocaine abuse, $GADI^a$, and $GADI^b$. Results depicted the disadvantage of using full Bayesian model since the range of coefficients were very narrow with most of the values close to zero. As a result BMA detects much larger number of significant voxels as compared to full model, which we conjecture is due to this over-shrinkage of the estimates (see Section S4 of the the supplementary materials for more details.)

5 Discussion

This article presents iBANG, a general method for the integrative Bayesian analysis of voxel-wise neuroimaging and genetic data. The key features of our modeling strategy are to allow for the quantification of model uncertainty via Bayesian model averaging, account for spatial correlation via local smoothing, and explicitly control for multiplicity via Bayesian

false discovery rate procedures. In addition, by decoupling the model fitting and inference, we are able to scale our methods to conduct voxel-wise analyses of the entire white matter of the brain with multiple genetic variables. Although our methods are motivated by a specific neuroimaging-genetic study in cocaine addiction –our methods are general and applicable to any voxel-wise imaging modality.

We exemplify our methods using neuroimaging-genetic data from a cocaine addiction study. We analyze the integrity of the white matter of the brain and demonstrate the possible effect of candidate genetic variants and demographic factors on white matter impairment in users of cocaine. Results of this analysis showed that gene polymorphisms associated with the synthesis of *GABA*, serotonin, and dopamine and the function of the corresponding receptors were associated with brain FA in cocaine-dependent subjects. At least some previous preclinical and clinical studies have found a relationship between cocaine use and *GAD1*. A study by (Enoch et al., 2012) showed that *GAD1* and *GAD2* expression levels in postmortem brains were related to cocaine use. In addition, our findings of reduced white matter FA in cocaine users are consistent with previous studies (Beyer and Steketee, 1999; Jentsch and Taylor, 1999; Vanderschuren and Kalivas, 2000; Lim et al., 2002; Moeller et al., 2005). However, our methods showed evidence of FA diminishment in most brain regions of interest, which is an interesting finding that is worthy of further investigation.

To identify significant regions of the brain associated with genetic variables, we applied voxel-wise FDR correction to the smoothed posterior probability maps (sPPMs) for each genetic covariate of interest, independently. Another alternate and interesting approach, as pointed out by one of the reviewers, is to use the notion of a “topological FDR” (Chumbley and Friston, 2009) wherein topological boundaries of activation are identified around local inflection points i.e. *a priori* clustering of voxels (e.g via spatial volumes) into discrete sets and treated as the units of inference. As argued by the authors, this topological perspective refines the interpretation as well as inference, and allows for more rigorous control of a smaller multiple comparison problem. In a classical (frequentist) and single covariate (one treatment) settings, the authors propose an approach that combines the FDR procedure on p-values obtained from SPM with Random Field Theory (RFT) to find regional (defined by topological property) activation of the underlying signal. We take a different tack in this article. From a Bayesian viewpoint, the sPPMs inherently account for the spatial structure in the data (as opposed to raw PPMs) and thus define contiguous region of activation. Our main motivation for doing this is to allow our methods to scale to large datasets as well as provide principled statistical inference while accounting for various sources of variability induced by model selection and spatial heterogeneity. Extending our FDR approach to a more global topological level, while no doubt interesting, is a non-trivial one, for the following reasons: (i) the approach must accommodate two hierarchical levels of multiplicity – within and across genetic covariate maps; (ii) requires prior null and alternative hypotheses specifications at each level and (iii) in a DTI context, this would also require a more precise definition of a topological feature. We leave these tasks for future consideration.

There are several possible extensions and generalizations of our iBANG models to more general settings. In this work, we considered full MCMC-based posterior sampling approaches to estimate the exact model probabilities for each (genetic) covariate at each

voxel location. Another approximate approach based on Savage-Dickey Bayes factors has been recently proposed by (Rosa et al., 2012). In their approach, first the largest (full) model is computed and subsequently, a post-hoc approach is used to calculate the model evidence for any (reduced) submodel by using a generalization of the Savage-Dickey density ratio (Dickey, 1971). The benefit of this approach is a reduction in computational time, since a potentially huge space of the nested sub-models can be explored using single model fit.

Another natural advancement is to incorporate spatial correlation in the Bayesian model averaging in the first step of the model fitting. We eschew this step because of computational constraints and so avoid evaluating and storing large covariance matrices. However, the use of lower dimensional projections of these matrices, such as principal component analysis (PCA) or independent component analysis (ICA) could potentially circumvent these issues. Our implementation is currently limited to one imaging parameter (FA, in this case). In future extensions, one could simultaneously consider multiple imaging parameters, such as mean diffusivity, axial diffusivity, and radial diffusivity. We leave these tasks for future work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Shabnam Azadeh was supported by the NIDA R25T in Statistical Genetics of Addiction training program at UT MD Anderson Cancer Center. Drs. Hobbs and Baladandayuthapani were partially supported by the NIH through the University of Texas MD Anderson Cancer Center Support Grant (CCSG) (P30 CA016672). Dr. Baladandayuthapani was also partially supported by NIH grant R01 CA160736. Dr. Nielsen was supported by NIH/NIDA DA018197, and through the University of Texas MD Anderson Cancer Center Support Grant DA026120, and the Toomim Family Fund. Dr. Moeller was supported by U54DA038999, and P50DA33935. Dr. Ma was supported by R01 grant (DA034131). Research reported in this publication was supported by the National Institute on Drug Abuse of the National Institutes of Health under Award Number R25DA026120, and P50DA009262. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Special thanks to Ms. LeeAnn Chastain for editing assistance. This material is the result of work supported with resources and the use of facilities at the Michael E. DeBakey VA Medical Center, Houston, TX.

References

- Anastasio N, Liu S, Maili L, Swinford S, Lane S, Fox R, Hamon S, Nielsen D, Cun-ningham K, Moeller F. Variation within the serotonin (5-HT) 5-HT_{2C} receptor system aligns with vulnerability to cocaine cue reactivity. *Translational psychiatry*. 2014; 4:e369. [PubMed: 24618688]
- Andersson JL, Jenkinson M, Smith S. Non-linear optimisation. *FMRIB technical report TR07JA1*. Practice. 2007a Jun.
- Andersson, JL.; Jenkinson, M.; Smith, S., et al. Non-linear registration, aka Spatial normalisation *FMRIB technical report TR07JA2*. FMRIB Analysis Group of the University of Oxford; 2007b.
- Ashburner J, Friston KJ. Voxel-based morphometrythe methods. *Neuroimage*. 2000; 11:805–821. [PubMed: 10860804]
- Baladandayuthapani V, Ji Y, Talluri R, Nieto-Barajas LE, Morris JS. Bayesian random segmentation models to identify shared copy number aberrations for array CGH data. *Journal of the american statistical association*. 2010:105.
- Bauer IE, Soares JC, Nielsen DA. The role of opioidergic genes in the treatment outcome of drug addiction pharmacotherapy: A systematic review. *The American Journal on Addictions*. 2014

- Behrens T, Johansen-Berg H, Woolrich M, Smith S, Wheeler-Kingshott C, Boulby P, Barker G, Sillery E, Sheehan K, Ciccarelli O, et al. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature neuroscience*. 2003; 6:750–757. [PubMed: 12808459]
- Berrettini WH, Lerman CE. Pharmacotherapy and pharmacogenetics of nicotine dependence. *American Journal of Psychiatry*. 2005; 162:1441–1451. [PubMed: 16055765]
- Beyer CE, Steketee JD. Dopamine depletion in the medial prefrontal cortex induces sensitized-like behavioral and neurochemical responses to cocaine. *Brain research*. 1999; 833:133–141. [PubMed: 10375688]
- Braskie MN, Jahanshad N, Stein JL, Barysheva M, Johnson K, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Ringman JM, et al. Relationship of a variant in the NTRK1 gene to white matter microstructure in young adults. *The Journal of Neuroscience*. 2012; 32:5964–5972. [PubMed: 22539856]
- Braskie MN, Jahanshad N, Stein JL, Barysheva M, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Ringman JM, Toga AW, et al. Common Alzheimer's disease risk variant within the CLU gene affects white matter microstructure in young adults. *The Journal of Neuroscience*. 2011; 31:6764–6770. [PubMed: 21543606]
- Brewer AJ III, Nielsen DA, Spellacy CJ, Hamon SC, Gingrich J, Thompson-Lake DG, Nielsen EM, Mahoney JJ III, Kosten TR, Newton TF, et al. Genetic variation of the dopamine transporter (DAT1) influences the acute subjective responses to cocaine in volunteers with cocaine use disorders. *Pharmacogenetics and Genomics*. 2015
- Buades, A.; Coll, B.; Morel, J-M. A non-local algorithm for image denoising. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on; IEEE; 2005.* p. 60-65.
- Chumbley JR, Friston KJ. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *Neuroimage*. 2009; 44:62–70. [PubMed: 18603449]
- Chun H, Kele S. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2010; 72:3–25. [PubMed: 20107611]
- Chung MK, Worsley KJ, Nacewicz BM, Dalton KM, Davidson RJ. General multivariate linear modeling of surface shapes using SurfStat. *NeuroImage*. 2010; 53:491–505. [PubMed: 20620211]
- Craig J, et al. Complex diseases: Research and applications. *Nature Education*. 2008:1.
- Dennis EL, Jahanshad N, Rudie JD, Brown JA, Johnson K, McMahon KL, de Zubicaray GI, Montgomery G, Martin NG, Wright MJ, et al. Altered structural brain connectivity in healthy carriers of the autism risk gene, CNTNAP2. *Brain connectivity*. 2011; 1:447–459. [PubMed: 22500773]
- Dickey JM. The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*. 1971:204–223.
- Enoch M-A, Zhou Z, Kimura M, Mash DC, Yuan Q, Goldman D. GABAergic gene expression in postmortem hippocampus from alcoholics and cocaine addicts; corresponding findings in alcohol-naïve P and NP rats. *PloS one*. 2012; 7:e29369. [PubMed: 22253714]
- Fernandez C, Ley E, Steel MF. Benchmark priors for Bayesian model averaging. *Journal of Econometrics*. 2001a; 100:381–427.
- Fernandez C, Ley E, Steel MF. Model uncertainty in cross-country growth regressions. *Journal of applied Econometrics*. 2001b; 16:563–576.
- First, M.; Spitzer, R.; Gibbon, M.; Williams, J. Structured Clinical Interview for DSM-IV Axis I Disorders-Patient Editions (SCID-I/P, Version 2.0). New York: Biometrics Research Department, New York State Psychiatric Institute; 1996.
- Formisano E, De Martino F, Valente G. Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. *Magnetic resonance imaging*. 2008; 26:921–934. [PubMed: 18508219]
- Frazier TW, Youngstrom EA, Frankel BA, Zunta-Soares GB, Sanches M, Escamilla M, Nielsen DA, Soares JC. Candidate gene associations with mood disorder, cognitive vulnerability, and fronto-limbic volumes. *Brain and behavior*. 2014; 4:418–430. [PubMed: 24944871]

- Friston K, Penny W. Posterior probability maps and SPMs. *Neuroimage*. 2003; 19:1240–1249. [PubMed: 12880849]
- Ge T, Feng J, Hibar DP, Thompson PM, Nichols TE. Increasing power for voxel-wise genome-wide association studies: the random field theory, least square kernel machines and fast permutation procedures. *NeuroImage*. 2012; 63:858–873. [PubMed: 22800732]
- Guleryuz OG. Weighted averaging for denoising with overcomplete dictionaries. *Image Processing, IEEE Transactions on*. 2007; 16:3020–3034.
- Hadfield M. Cocaine. *Molecular neurobiology*. 1995; 11:47–53. [PubMed: 8561967]
- Hammer R, Cooke E. Gradual tolerance of metabolic activity is produced in mesolimbic regions by chronic cocaine treatment, while subsequent cocaine challenge activates extrapyramidal regions of rat brain. *The Journal of neuroscience*. 1994; 14:4289–4298. [PubMed: 8027779]
- Han DD, Gu HH. Comparison of the monoamine transporters from human and mouse in their sensitivities to psychostimulant drugs. *BMC pharmacology*. 2006; 6:6. [PubMed: 16515684]
- Hasan KM. A framework for quality control and parameter optimization in diffusion tensor imaging: theoretical analysis and validation. *Magnetic resonance imaging*. 2007; 25:1196–1202. [PubMed: 17442523]
- Hasan KM, Parker DL, Alexander AL. Comparison of gradient encoding schemes for diffusion-tensor MRI. *Journal of Magnetic Resonance Imaging*. 2001; 13:769–780. [PubMed: 11329200]
- Heller R, Golland Y, Malach R, Benjamini Y. Conjunction group analysis: an alternative to mixed/random effect analysis. *Neuroimage*. 2007; 37:1178–1185. [PubMed: 17689266]
- Ho AJ, Stein JL, Hua X, Lee S, Hibar DP, Leow AD, Dinov ID, Toga AW, Saykin AJ, Shen L, et al. A commonly carried allele of the obesity-related FTO gene is associated with reduced brain volume in the healthy elderly. *Proceedings of the National Academy of Sciences*. 2010; 107:8404–8409.
- Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial. *Statistical science*. 1999:382–401.
- Hyman SE, Malenka RC. Addiction and the brain: the neurobiology of compulsion and its persistence. *Nature reviews neuroscience*. 2001; 2:695–703. [PubMed: 11584307]
- Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. *Neuroimage*. 2012; 62:782–790. [PubMed: 21979382]
- Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Medical image analysis*. 2001; 5:143–156. [PubMed: 11516708]
- Jentsch JD, Taylor JR. Impulsivity resulting from frontostriatal dysfunction in drug abuse: implications for the control of behavior by reward-related stimuli. *Psychopharmacology*. 1999; 146:373–390. [PubMed: 10550488]
- Kass RE, Raftery AE. Bayes factors. *Journal of the american statistical association*. 1995; 90:773–795.
- Kherif F, Poline J-B, Flandin G, Benali H, Simon O, Dehaene S, Worsley KJ. Multivariate model specification for fMRI data. *NeuroImage*. 2002; 16:1068–1083. [PubMed: 12202094]
- Kohannim, O.; Hibar, DP.; Stein, JL.; Jahanshad, N.; Jack, CR.; Weiner, MW.; Toga, AW.; Thompson, PM. Boosting power to detect genetic associations in imaging using multi-locus, genome-wide scans and ridge regression. *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on; IEEE; 2011. p. 1855-1859.*
- Kosten TA, Huang W, Nielsen DA. Sex and litter effects on anxiety and DNA methylation levels of stress and neurotrophin genes in adolescent rats. *Developmental psychobiology*. 2014; 56:392–406. [PubMed: 23460384]
- Kosten TR, Domingo CB, Hamon SC, Nielsen DA. DBH gene as predictor of response in a cocaine vaccine clinical trial. *Neuroscience letters*. 2013a; 541:29–33. [PubMed: 23458673]
- Kosten TR, Wu G, Huang W, Harding MJ, Hamon SC, Lappalainen J, Nielsen DA. Pharmacogenetic randomized trial for cocaine abuse: disulfiram and dopamine β -hydroxylase. *Biological psychiatry*. 2013b; 73:219–224. [PubMed: 22906516]
- Kreek MJ, Bart G, Lilly C, Laforge KS, Nielsen DA. Pharmacogenetics and human molecular genetics of opiate and cocaine addictions and their treatments. *Pharmacological reviews*. 2005a; 57:1–26. [PubMed: 15734726]

- Kreek MJ, Nielsen DA, Butelman ER, LaForge KS. Genetic influences on impulsivity, risk taking, stress responsivity and vulnerability to drug abuse and addiction. *Nature neuroscience*. 2005b; 8:1450–1457. [PubMed: 16251987]
- Krishnan A, Williams LJ, McIntosh AR, Abdi H. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage*. 2011; 56:455–475. [PubMed: 20656037]
- Lane SD, Steinberg JL, Ma L, Hasan KM, Kramer LA, Zuniga EA, Narayana PA, Moeller FG. Diffusion tensor imaging and decision making in cocaine dependence. *PLoS One*. 2010; 5:e11591. [PubMed: 20661285]
- Lazar NA, Luna B, Sweeney JA, Eddy WF. Combining brains: a survey of methods for statistical pooling of information. *Neuroimage*. 2002; 16:538–550. [PubMed: 12030836]
- Li MD, Burmeister M. New insights into the genetics of addiction. *Nature Reviews Genetics*. 2009; 10:225–231.
- Lim KO, Choi SJ, Pomara N, Wolkin A, Rotrosen JP. Reduced frontal white matter integrity in cocaine dependence: a controlled diffusion tensor imaging study. *Biological psychiatry*. 2002; 51:890–895. [PubMed: 12022962]
- Lim KO, Wozniak JR, Mueller BA, Franc DT, Specker SM, Rodriguez CP, Silverman AB, Rotrosen JP. Brain macrostructural and microstructural abnormalities in cocaine dependence. *Drug and alcohol dependence*. 2008; 92:164–172. [PubMed: 17904770]
- Liu D, Lin X, Ghosh D. Semiparametric Regression of Multidimensional Genetic Pathway Data: Least-Squares Kernel Machines and Linear Mixed Models. *Biometrics*. 2007; 63:1079–1088. [PubMed: 18078480]
- Liu H, Li L, Hao Y, Cao D, Xu L, Rohrbaugh R, Xue Z, Hao W, Shan B, Liu Z. Disrupted white matter integrity in heroin dependence: a controlled study utilizing diffusion tensor imaging. *The American journal of drug and alcohol abuse*. 2008; 34:562–575. [PubMed: 18720268]
- Liu S, Green CE, Lane SD, Kosten TR, Moeller FG, Nielsen DA, Schmitz JM. The influence of dopamine β -hydroxylase gene polymorphism rs1611115 on levodopa/carbidopa treatment for cocaine dependence: a preliminary study. *Pharmacogenetics and genomics*. 2014; 24:370–373. [PubMed: 24809448]
- Ma L, Hasan KM, Steinberg JL, Narayana PA, Lane SD, Zuniga EA, Kramer LA, Moeller FG. Diffusion tensor imaging in cocaine dependence: regional effects of cocaine on corpus callosum and effect of cocaine administration route. *Drug and alcohol dependence*. 2009; 104:262–267. [PubMed: 19595517]
- Madigan D, York J, Allard D. Bayesian graphical models for discrete data. *International Statistical Review/Revue Internationale de Statistique*. 1995:215–232.
- Mallat, S. *A wavelet tour of signal processing*. Academic press; 1999.
- Manjón JV, Coupé P, Buades A, Louis Collins D, Robles M. New methods for MRI denoising based on sparseness and self-similarity. *Medical image analysis*. 2012; 16:18–27. [PubMed: 21570894]
- Mitchell TJ, Beauchamp JJ. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*. 1988; 83:1023–1032.
- Moeller FG, Hasan KM, Steinberg JL, Kramer LA, Dougherty DM, Santos RM, Valdes I, Swann AC, Barratt ES, Narayana PA. Reduced anterior corpus callosum white matter integrity is related to increased impulsivity and reduced discriminability in cocaine-dependent subjects: diffusion tensor imaging. *Neuropsychopharmacology*. 2005; 30:610–617. [PubMed: 15637640]
- Moeller FG, Hasan KM, Steinberg JL, Kramer LA, Valdes I, Lai LY, Swann AC, Narayana PA. Diffusion tensor imaging eigenvalues: preliminary evidence for altered myelin in cocaine dependence. *Psychiatry Research: Neuroimaging*. 2007; 154:253–258. [PubMed: 17321725]
- Morris JS, Brown PJ, Herrick RC, Baggerly KA, Coombes KR. Bayesian Analysis of Mass Spectrometry Proteomic Data Using Wavelet-Based Functional Mixed Models. *Biometrics*. 2008; 64:479–489. [PubMed: 17888041]
- Müller P, Quintana F, Rosner G. Hierarchical meta-analysis over related non-parametric Bayesian models. *Journal of Royal Statistical Society, Series B*. 2004; 66:735–749.
- Nielsen D, Harding M, Hamon S, Huang W, Kosten T. Modifying the role of serotonergic 5-HTTLPR and TPH2 variants on disulfiram treatment of cocaine addiction: a preliminary study. *Genes, Brain and Behavior*. 2012a; 11:1001–1008.

- Nielsen DA, Huang W, Hamon SC, Maili L, Witkin BM, Fox RG, Cunningham KA, Moeller FG. Forced abstinence from cocaine self-administration is associated with DNA methylation changes in myelin genes in the corpus callosum: a preliminary study. *Frontiers in psychiatry*. 2012b;3. [PubMed: 22319501]
- Nielsen DA, Kreek MJ. Common and specific liability to addiction: approaches to association studies of opioid addiction. *Drug and alcohol dependence*. 2012; 123:S33–S41. [PubMed: 22542464]
- Nielsen, DA.; Nielsen, EM.; Dasari, T.; Spellicy, CJ. *Pharmacogenomics in Drug Discovery and Development*. Springer; 2014. Pharmacogenetics of addiction therapy; p. 589-624.
- Nielsen DA, Utrankar A, Reyes JA, Simons DD, Kosten TR. Epigenetics of drug abuse: predisposition or response. *Pharmacogenomics*. 2012c; 13:1149–1160. [PubMed: 22909205]
- Nowak RD. Wavelet-based Rician noise removal for magnetic resonance imaging. *Image Processing, IEEE Transactions on*. 1999; 8:1408–1419.
- Raftery AE. Bayesian model selection in social research. *Sociological methodology*. 1995; 25:111–164.
- Raftery AE, Madigan D, Hoeting JA. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*. 1997; 92:179–191.
- Romero MJ, Asensio S, Palau C, Sanchez A, Romero FJ. Cocaine addiction: diffusion tensor imaging study of the inferior frontal and anterior cingulate white matter. *Psychiatry Research: Neuroimaging*. 2010; 181:57–63. [PubMed: 19959341]
- Rosa M, Friston K, Penny W. Post-hoc selection of dynamic causal models. *Journal of neuroscience methods*. 2012; 208:66–78. [PubMed: 22561579]
- Rowe DB, Hoffmann RG. Multivariate statistical analysis in fMRI. *Engineering in Medicine and Biology Magazine, IEEE*. 2006; 25:60–64.
- Smith SM. Fast robust automated brain extraction. *Human brain mapping*. 2002; 17:143–155. [PubMed: 12391568]
- Spellicy C, Harding M, Hamon S, Mahoney J, Reyes J, Kosten T, Newton T, De La Garza R, Nielsen D. A variant in ANKK1 modulates acute subjective effects of cocaine: a preliminary study. *Genes, Brain and Behavior*. 2014; 13:559–564.
- Spellicy CJ, Kosten TR, Hamon SC, Harding MJ, Nielsen DA. ANKK1 and DRD2 pharmacogenetics of disulfiram treatment for cocaine abuse. *Pharmacogenetics and genomics*. 2013; 23:333–340. [PubMed: 23635803]
- Stein EA, Fuller SA. Cocaine's time action profile on regional cerebral blood flow in the rat. *Brain research*. 1993; 626:117–126. [PubMed: 8281422]
- Storey JD. The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics*. 2003:2013–2035.
- Szekely GJ, Rizzo ML. Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method. *Journal of classification*. 2005; 22:151–183.
- Taylor J, Worsley K. Random fields of multivariate test statistics, with applications to shape analysis. *The Annals of Statistics*. 2008:1–27.
- Teipel SJ, Born C, Ewers M, Bokde AL, Reiser MF, Möller H-J, Hampel H. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. *Neuroimage*. 2007; 38:13–24. [PubMed: 17827035]
- Vanderschuren LJ, Kalivas PW. Alterations in dopaminergic and glutamatergic transmission in the induction and expression of behavioral sensitization: a critical review of preclinical studies. *Psychopharmacology*. 2000; 151:99–120. [PubMed: 10972458]
- Volkow ND, Fowler JS, Wang G-J, et al. The addicted human brain: insights from imaging studies. *Journal of Clinical Investigation*. 2003; 111:1444–1451. [PubMed: 12750391]
- Vounou M, Nichols TE, Montana G. Discovering genetic associations with high-dimensional neuroimaging phenotypes: a sparse reduced-rank regression approach. *Neuroimage*. 2010; 53:1147–1159. [PubMed: 20624472]
- Wakana S, Jiang H, Nagae-Poetscher LM, Van Zijl PC, Mori S. Fiber Tract-based Atlas of Human White Matter Anatomy 1. *Radiology*. 2004; 230:77–87. [PubMed: 14645885]

Worsley KJ, Taylor JE, Tomaiuolo F, Lerch J. Unified univariate and multi-variate random field theory. *Neuroimage*. 2004; 23:S189–S195. [PubMed: 15501088]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Highlights

- Statistical framework for integrative Bayesian analysis of neuroimaging-genetic data.
- Delineate significant brain regions associated with genetic variants.
- Fast computational schemes scalable to voxel-wise model fitting and inference.
- Cocaine consumption is associated with fractional anisotropy (FA) reduction.
- Gene polymorphisms associated with GABAergic were associated with FA.

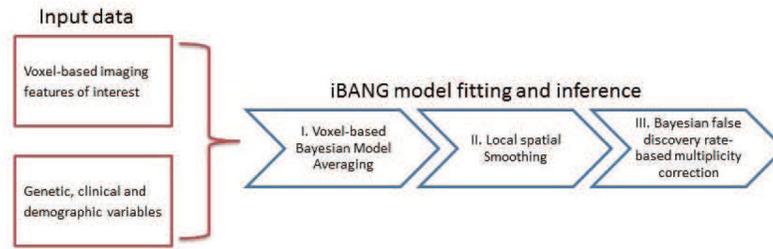


Figure 1.
Schematic representation of iBANG model and analysis

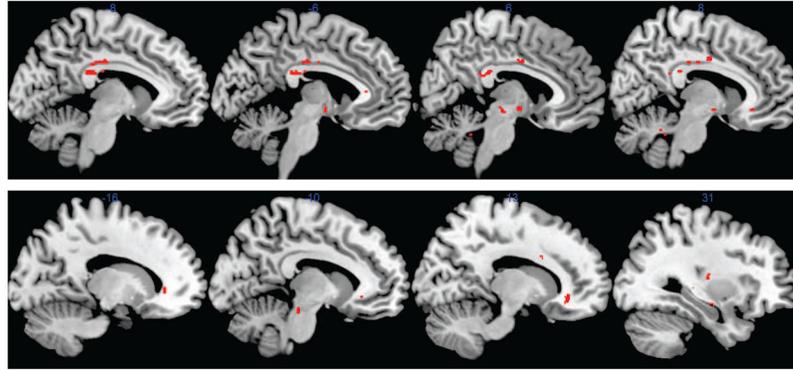


Figure 2. Multi-slice sagittal views displaying significant regions¹ associated with *GAD1^a* and *GAD1^b* respectively from the top to the bottom panel. The SRs, depicted in red, characterize locations for which alteration of mean FA was evident from statistical analysis.

¹In the significant voxels, the adjacent voxels equal to or greater than 20 were grouped and named as significant regions (SRs) to limit noisy images while plotting the sagittal views so that all significant voxels might not be displayed. Significant voxels were grouped in SPM software using MarsBaR tool. Then multi-slice sagittal views were generated using MRICroN software. (Same strategy is used in Figure 3.)

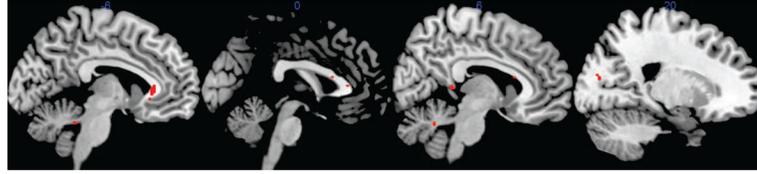


Figure 3. Multi-slice sagittal views displaying significant regions associated with cocaine consumption. The SRs, depicted in red, characterize locations for which alteration of the mean FA was evident from statistical analysis.

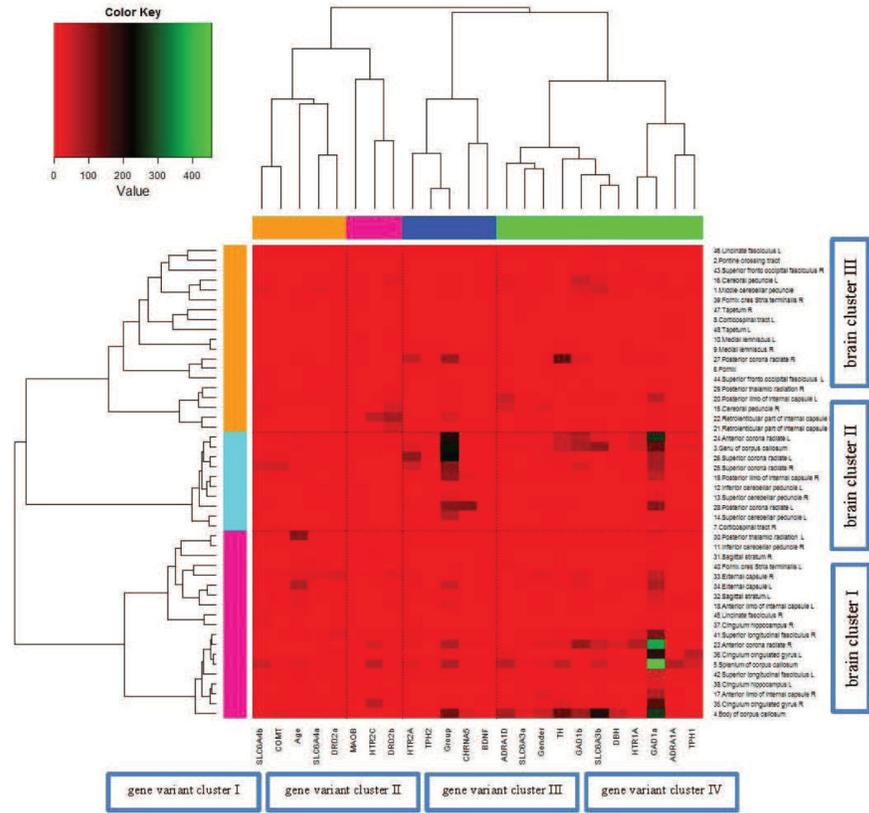


Figure 4. Two-sided hierarchical clustering²of the neuroimaging-genetic findings based on the John Hopkins University white matter atlas. Green indicates an increased number of significant voxels for the ROI of the brain and genetic variant combination. The analysis resulted in four genetic variant clusters, denoted as genetic variant cluster I, genetic variant cluster II, genetic variant cluster III, and genetic variant cluster IV; and three clusters of brain ROIs defined by the JHU atlas, which are denoted as brain cluster I, brain cluster II, and brain cluster III. The individual cluster elements are tabulated in Table 5.

²The bivariate cluster of the neuroimaging-genetic data considering the number of significant voxels was formulated using *gplots* software package in R.

Table 1

Characteristics (mean and range) of study population of cocaine users and controls.

| | Control group(N=19) | Cocaine group(N=39) |
|------------------------------------|---------------------|---------------------|
| Sex (% male) | 63 | 74 |
| Age (years) | 32.69 (21 – 48.8) | 40.38 (22.7 – 54.8) |
| Education (years) | 14.16 (11 – 18) | 12.85 (10 – 16) |
| Lifetime cocaine use (years) | 0 | 13.83 (2 – 30) |
| Cocaine use (days per prior month) | 0 | 14.77 (2 – 30) |
| Ethnicity (% African American) | 84 | 51 |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Genetic Summary: first column presents the list of variant in the study; second column shows SNPs and Insertion/Deletion; third column indicates gene symbol; fourth column presents gene name; and fifth column is genotype coding .

| Genetic variant | SNPs | Gene symbol | Gene name | Code |
|---------------------------|------------------------------------|---------------|--|-------------------------------------|
| <i>HTR2A</i> | rs6311 | <i>HTR2A</i> | 5-hydroxytryptamine (serotonin) receptor 2A, G protein-coupled | (GG=0),(AG,AA=1) |
| <i>HTR2C</i> | rs6318 | <i>HTR2C</i> | 5-hydroxytryptamine (serotonin) receptor 2C, G protein-coupled | (GG=0),(CG,CC=1) |
| <i>TPH1</i> | rs1799913 | <i>TPH1</i> | tryptophan hydroxylase 1 | (GG=0),(GT,TT=1) |
| <i>TPH2</i> | rs4290270 | <i>TPH2</i> | tryptophan hydroxylase 2 | (AA=0),(AT,TT=1) |
| <i>MAOB</i> | rs1799836 | <i>MAOB</i> | monoamine oxidase B | (CC=0),(CT,TT=1) |
| <i>DRD2^a</i> | rs6277 | <i>DRD2</i> | dopamine receptor D2 | (GG=0),(AG,AA=1) |
| <i>DRD2^b</i> | rs2283265 | <i>DRD2</i> | dopamine receptor D2 | (CC=0),(AC,AA=1) |
| <i>DBH</i> | rs1611115 | <i>DBH</i> | dopamine beta-hydroxylase (dopamine beta-monoxygenase) | (CC=0),(CT,TT=1) |
| <i>COMT</i> | rs4680 | <i>COMT</i> | catechol-O-methyltransferase | (GG=0),(AG,AA=1) |
| <i>TH</i> | rs10770141 | <i>TH</i> | tyrosine hydroxylase | (GG=0),(AG,AA=1) |
| <i>HTR1A</i> | rs6295 | <i>HTR1A</i> | 5-hydroxytryptamine (serotonin) receptor 1A, G protein-coupled | (CC=0),(CG,GG=1) |
| <i>GADI^a</i> | rs1978340 | <i>GADI</i> | glutamate decarboxylase 1 (brain, 67kDa) | (GG=0),(AG,AA=1) |
| <i>GADI^b</i> | rs769390 | <i>GADI</i> | glutamate decarboxylase 1 (brain, 67kDa) | (AA=0),(AC,CC=1) |
| <i>ADRA1A</i> | rs1048101 | <i>ADRA1A</i> | adrenoceptor alpha 1A | (GG=0),(AG,AA=1) |
| <i>ADRA1D</i> | rs2236554 | <i>ADRA1D</i> | adrenoceptor alpha 1D | (TT=0),(AT,AA=1) |
| <i>CHRNA5</i> | rs16969968 | <i>CHRNA5</i> | cholinergic receptor, nicotinic, alpha 5 (neuronal) | (GG=0),(AG,AA=1) |
| <i>BDNF</i> | rs6265 | <i>BDNF</i> | brain-derived neurotrophic factor | (CC=0),(CT,TT=1) |
| Insertion/Deletion | | | | |
| <i>SLC6A4^a</i> | 5-HTTLPR in promoter | <i>SLC6A4</i> | solute carrier family 6 (neurotransmitter transporter), member 4 | (LL=0),(LS,SS=1) |
| <i>SLC6A4^b</i> | 17-bp VNTR in intron 2, STin2 | <i>SLC6A4</i> | solute carrier family 6 (neurotransmitter transporter), member 4 | {(12, 12) = 0}, { <i>Other</i> = 0} |
| <i>SLC6A3^a</i> | 40-bp VNTR in the 3' UTR exon 15 | <i>SLC6A3</i> | solute carrier family 6 (neurotransmitter transporter), member 3 | {(10, 10) = 0}, { <i>Other</i> = 0} |
| <i>SLC6A3^b</i> | 30-bp VNTR in intron 8 (Int8 VNTR) | <i>SLC6A3</i> | solute carrier family 6 (neurotransmitter transporter), member 3 | {(6, 6) = 0}, { <i>Other</i> = 0} |

Table 3

Number of significant voxel locations identified for each variable using iBANG analysis, sorted in decreasing order. Bayesian FDRs set to $\delta = 0.1$ for each variable. The top (bottom) panel shows the genetic variants and demographic variables.

| Genetic variants | Description | Number of significant voxels |
|---------------------------|--|------------------------------|
| <i>GAD1^a</i> | glutamate decarboxylase 1 (brain, 67kDa) | 5217 |
| <i>HTR2A</i> | 5-hydroxytryptamine (serotonin) receptor 2A, G protein-coupled | 1413 |
| <i>TH</i> | tyrosine hydroxylase | 1332 |
| <i>GAD1^b</i> | glutamate decarboxylase 1 (brain, 67kDa) | 1207 |
| <i>SLC6A4^b</i> | solute carrier family 6 (neurotransmitter transporter), member 4 | 1125 |
| <i>ADRA1A</i> | adrenoceptor alpha 1A | 1036 |
| <i>SLC6A3^b</i> | solute carrier family 6 (neurotransmitter transporter), member 3 | 1017 |
| <i>ADRA1D</i> | adrenoceptor alpha 1D | 813 |
| <i>HT2CR</i> | 5-hydroxytryptamine (serotonin) receptor 2C, G protein-coupled | 805 |
| <i>COMT</i> | catechol-O-methyltransferase | 721 |
| <i>HT1A</i> | tryptophan hydroxylase 1 | 688 |
| <i>CHRNA5</i> | cholinergic receptor, nicotinic, alpha 5 (neuronal) | 670 |
| <i>TPH1</i> | tryptophan hydroxylase 1 | 666 |
| <i>SLC6A4^a</i> | solute carrier family 6 (neurotransmitter transporter), member 4 | 654 |
| <i>DRD2^b</i> | dopamine receptor D2 | 611 |
| <i>SLC6A3^a</i> | solute carrier family 6 (neurotransmitter transporter), member 3 | 476 |
| <i>DRD2^a</i> | dopamine receptor D2 | 439 |
| <i>DBH</i> | dopamine beta-hydroxylase (dopamine beta-monoxygenase) | 413 |
| <i>BDNF</i> | brain-derived neurotrophic factor | 392 |
| <i>TPH2</i> | tryptophan hydroxylase 2 | 366 |
| <i>MAOB</i> | monoamine oxidase B | 359 |
| Demographics | | |
| Cocaine abuse | Cocaine user=1 vs Control=0 | 3100 |
| Age | | 1127 |
| Gender | | 1086 |

Table 4

Regions of white matter defined by the Johns Hopkins University atlas

| ROI | Description |
|-----|--|
| 1 | Middle cerebellar peduncle |
| 2 | Pontine crossing tract |
| 3 | Genu of corpus callosum |
| 4 | Body of corpus callosum |
| 5 | Splenium of corpus callosum |
| 6 | Fornix |
| 7 | Corticospinal tract R |
| 8 | Corticospinal tract L |
| 9 | Medial lemniscus R |
| 10 | Medial lemniscus L |
| 11 | Inferior cerebellar peduncle R |
| 12 | Inferior cerebellar peduncle L |
| 13 | Superior cerebellar peduncle R |
| 14 | Superior cerebellar peduncle L |
| 15 | Cerebral peduncle R |
| 16 | Cerebral peduncle L |
| 17 | Anterior limb of internal capsule R |
| 18 | Anterior limb of internal capsule L |
| 19 | Posterior limb of internal capsule R |
| 20 | Posterior limb of internal capsule L |
| 21 | Retrolenticular part of internal capsule R |
| 22 | Retrolenticular part of internal capsule L |
| 23 | Anterior corona radiate R |
| 24 | Anterior corona radiate L |
| 25 | Superior corona radiate R |
| 26 | Superior corona radiate L |
| 27 | Posterior corona radiate R |
| 28 | Posterior corona radiate L |
| 29 | Posterior thalamic radiation include optic radiation R |
| 30 | Posterior thalamic radiation include optic radiation L |
| 31 | Sagittal stratum include inferior longitudinal fasciculus and inferior fronto occipital fasciculus R |
| 32 | Sagittal stratum include inferior longitudinal fasciculus and inferior fronto occipital fasciculus L |
| 33 | External capsule R |
| 34 | External capsule L |
| 35 | Cingulum cingulated gyrus R |
| 36 | Cingulum cingulated gyrus L |
| 37 | Cingulum hippocampus R |
| 38 | Cingulum hippocampus L |
| 39 | Fornix cres Stria terminalis cannot be resolved with current resolution R |

| ROI | Description |
|------------|---|
| 40 | Fornix cres Stria terminalis cannot be resolved with current resolution L |
| 41 | Superior longitudinal fasciculus R |
| 42 | Superior longitudinal fasciculus L |
| 43 | Superior fronto occipital fasciculus could be a part of anterior internal capsule R |
| 44 | Superior fronto occipital fasciculus could be a part of anterior internal capsule L |
| 45 | Uncinate fasciculus R |
| 46 | Uncinate fasciculus L |
| 47 | Tapetum R |
| 48 | Tapetum L |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Two-sided hierarchical clustering of the neuroimaging-genetic data: the top panel lists the brain ROIs and the bottom panel lists the genetic variants in each cluster based on two-sided hierarchical clustering of the neuroimaging-genetic data in Figure 4.

| Clusters of brain regions of interest based on two-sided hierarchical clustering of the neuroimaging-genetic data | | | |
|---|---|---|----------------------------|
| brain cluster I | brain cluster II | brain cluster III | |
| 4.Body of corpus callosum | 3.Genu of corpus callosum | 1.Middle cerebellar peduncle | |
| 5.Splenium of corpus callosum | 7.Corticospinal tract R | 2.Pontine crossing tract | |
| 11.Inferior cerebellar peduncle R | 12.Inferior cerebellar peduncle L | 6.Fornix | |
| 17.Anterior limb of internal capsule R | 13.Superior cerebellar peduncle R | 8.Corticospinal tract L | |
| 18.Anterior limb of internal capsule L | 14.Superior cerebellar peduncle L | 9.Medial lemniscus R | |
| 23.Anterior corona radiate R | 19.Posterior limb of internal capsule R | 10.Medial lemniscus L | |
| 30.Posterior thalamic radiation L | 24.Anterior corona radiate L | 15.Cerebral peduncle R | |
| 31.Sagittal stratum R | 25.Superior corona radiate R | 16.Cerebral peduncle L | |
| 32.Sagittal stratum L | 26.Superior corona radiate L | 20.Posterior limb of internal capsule L | |
| 33.External capsule R | 28.Posterior corona radiate L | 21.Retrolenticular part of internal capsule R | |
| 34.External capsule L | | 22.Retrolenticular part of internal capsule L | |
| 35.Cingulum cingulated gyrus R | | 27.Posterior corona radiate R | |
| 36.Cingulum cingulated gyrus L | | 29.Posterior thalamic radiation R | |
| 37.Cingulum hippocampus R | | 39.Fornix cres Stria terminalis R | |
| 38.Cingulum hippocampus L | | 43.Superior fronto occipital fasciculus R | |
| 40.Fornix cres Stria terminalis L | | 44.Superior fronto occipital fasciculus L | |
| 41.Superior longitudinal fasciculus R | | 46.Uncinate fasciculus L | |
| 42.Superior longitudinal fasciculus L | | 47.Tapetum R | |
| 45.Uncinate fasciculus R | | 48.Tapetum L | |
| Genetic variants clusters based on two-sided hierarchical clustering of the neuroimaging-genetic data | | | |
| genetic variant cluster I | genetic variant cluster II | genetic variant cluster III | genetic variant cluster IV |
| <i>DRD2^a</i> | <i>MAOB</i> | <i>HTR2A</i> | <i>ADRA1D</i> |
| <i>COMT</i> | <i>HTR2C</i> | <i>TPH2</i> | <i>SLC6A3^a</i> |
| <i>SLC6A4^b</i> | <i>DRD2^b</i> | <i>BDNF</i> | <i>TPH1</i> |
| <i>SLC6A4^a</i> | | <i>CHRNA5</i> | <i>TH</i> |
| | | | <i>GADI^b</i> |
| | | | <i>SLC6A3^b</i> |
| | | | <i>DBH</i> |
| | | | <i>HTR1A</i> |
| | | | <i>GADI^a</i> |
| | | | <i>ADRA1A</i> |