Published in final edited form as:

Methods. 2009 September ; 49(1): 56–62. doi:10.1016/j.ymeth.2009.05.014.

Strategies for Performing Genotype-Phenotype Association Studies in Nonhuman Primates

Christina S. Barr

Laboratories of Neurogenetics and Clinical and Translational Studies, NIH/NIAAA, Rockville, MD, 20852

Abstract

Anthropoid primate models offer opportunities to study genetic influence on alcohol consumption and alcohol-related intermediate phenotypes in socially and behaviorally complex animal models that are closely related to humans, and in which functionally equivalent or orthologous genetic variants exist. This review will discuss the methods commonly used for performing candidate-gene based studies in rhesus macaques in order to model how functional genetic variation moderates risk for human psychiatric disorders. Various *in silico* and *in vitro* approaches to identifying functional genetic variants for performance of these studies will be discussed. Next, I will provide examples of how this approach can be used for performing candidate gene based studies and for examining gene by environment interactions. Finally, these approaches will then be placed in the context of how function-guided studies can inform us of genetic variants that may be under selection across species, demonstrating how functional genetic variants that may have conferred selective advantage at some point in the evolutionary history of humans could increase risk for addictive disorders in modern society.

Keywords

Macaque; Genetic; Comparative; Primate; EMSA; Reporter Assay; GxE Interactions

1. Introduction

The neurobiological systems that influence addiction vulnerability may do so by acting on reward pathways, behavioral dyscontrol, and vulnerability to stress and anxiety. Anthropoid primate models offer opportunities to study genetic influence on alcohol consumption and alcohol-related intermediate phenotypes in socially and behaviorally complex animal models that are closely related to humans, and in which functionally equivalent or orthologous genetic variants exist [1]. Genomic studies performed in nonhuman primates have translational value for investigating effects of genetic variation on stress reactivity, temperament, and reward sensitivity in alcohol-naïve subjects, and for understanding how genetic variation modifies stress- and alcohol-induced neuroadaptation, neuropathology, and treatment response.

Address correspondence to: Christina S. Barr, VMD, PhD, NIH/NIAAA, 5625 Fishers Lane, Rm 3S-32, Rockville, MD 20852, Phone: 301.496.8127, Fax: 301.435.9278, cbarr@mail.nih.gov.

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.

There are a number of research groups that have been investigating genetic variations in the rhesus macaque that contribute to the expression of traits that have been linked with human alcohol problems and other psychiatric disorders (i.e., stress reactivity, behavioral dyscontrol, aggression and reward seeking/sensitivity). What has emerged from this body of work is the fact that, in many cases, the variants that are identified and studied in the macaque are functionally similar to those present in human populations, and some findings suggest that some of these variants may have been maintained by selection in both species [2,3]. Such data reinforce the utility of the macaque model for studying how relatively common genetic variants, which are associated with traits that may be adaptive in certain environmental contexts, can increase vulnerability to stress-related or alcohol problems (see Section 5). This review will discuss the methods commonly used for performing candidate-gene based studies in outbred populations of rhesus macaques to model how functional genetic variation moderates risk for human psychiatric disorders.

2. Approaches and Challenges

Various approaches for performing genetic studies may be employed using nonhuman primates, and the population composition and structure are important factors to consider in determining which may be most appropriate or powerful. Many of the primate centers and breeding programs have large, pedigreed populations of nonhuman primates. These populations provide excellent tools for performing unbiased, whole-genome linkage studies to determine genotype-phenotype correlations in laboratory primates for which, unlike humans, environmental contributions can be minimized or controlled [5,6]. In recent years, the popularity of candidate-based studies has also risen [1]. Allele-based association studies can be used to examine variation for continuous traits and are much more powerful than linkage studies at detecting loci that account for only a small percentage of the variance for a complex trait, such as temperament [6,7]. However, performance of these studies in inbred populations can be challenging because of the fact that spurious genotype-phenotype associations may arise. The simultaneous genotyping of a high number of markers may help investigators to correct for this potential confound, providing a mechanism by which the specificity of effects of certain genetic markers can be determined and the confound of overall relatedness minimized [2,8]. This type of approach may also be quite useful for determining ancestry in admixed populations as well [9].

In free-ranging populations of macaques, inbreeding avoidance is achieved by male dispersal. Because of concerns relating to interrelatedness among the study subjects, many nonhuman primate colony managers have been employing breeding practices that mimic those observed in nature, with females born into the colony being maintained as breeders for several years, and potential sires being obtained each year from outside sources or distantly related groups of animals. These practices, in combination with the opportunity to use data collected across a large number of birth cohorts, can result in a low degree of overall relatedness among animals included in any given dataset [8]. While these types of populations may not be ideally suited for performing genome-wide linkage studies, they do provide opportunity to perform candidate gene-based studies in populations that have degrees of relatedness that approximate those observed in some human populations of study [10].

The macaque model offers a unique opportunity for studying how genetic and environmental factors relate to intra- and inter-specific variation in traits that, in humans, are known risk factors for alcohol use disorders. However, the use of this model offers multiple challenges. When studying nonhuman primates, there is tension between the ethics-driven necessity for reducing the number of research subjects and the science-driven need for sufficient statistical power. Even across the nation's primate centers, the number of animals for which any given phenotype is available is typically limited. Therefore, we and others have adopted various

approaches for increasing power in performing genetic studies in outbred populations of rhesus macaques. One method is in performing ancestral or cladistic clustering of haplotypes [2] in order to reduce the number of groups for comparison, approaches that have been used in human genetics studies for which datasets are particularly small [9]. The major limitation of using these methods is that if the functional marker is on a derived haplotype, then the effects of allelic variation may not be appreciated, particularly if the functional allele is rare. Another, more commonly used, approach is to perform function-guided linkage. This approach can be very powerful, but functional characterization of variants can be costly and time consuming. For these reasons, it can advantageous to make an *a priori* determination of which variants are likely to be functional.

2.1 Identification of Functional Variants

In some instances, a functional variant may be easily identified, for example, a frameshift mutation or a single nucleotide polymorphism (SNP) that is nonsynonymous. However, these types of variants are relatively rare. Not only are coding sequences more likely to be under strong purifying selection (Text Box), but some of these variants can be particularly deleterious. Studies in animals and in humans have demonstrated that variation within promoter or other regulatory regions can contribute to both strain-dependent and within-species inter-individual variation in behavior [11-13]. In fact, in psychiatric genetic studies, the functional variants that are implicated to increase risk for a given disorder are often located in regulatory, rather than coding, regions [14]. In rhesus macaques, many of the genetic studies performed have focused on variants present in the non-coding, regulatory regions as well [2,3,8,15-17].

There are a number of approaches for homing in on putatively functional noncoding genetic variants, some of which rely predominantly what is already known. In other words, these rely largely on literature searches and *in silico* analyses using publically available databases. The ideal case is that a regulatory region for a gene of interest has been characterized in terms of what specific regions are important, the treatments and/or transcription factors that influence expression, thus providing empirical support that a given sequence is functionally important. However, when this information is not available, there are means by which functional roles can be inferred. The wealth of information about the degree and patterns of genetic variation in humans, rodents and domestic dogs, in addition to the fact that the genomes of a large number of vertebrate species have been sequenced, provides opportunity for determining genomic regions that are under selection both within and across species. Though there are certainly exceptions, patterns of variation can often inform us of functional importance (Text Box).

In the laboratory, *cis*-acting functional SNPs can be identified or inferred using a number of different approaches. High-throughput strategies have been used to determine whether genotype or haplotype predict various molecular outcome measures, for example alternative splicing or gene expression levels [18,19]. Although these approaches are generally lacking in specificity, they have the advantage of offering the ability to simultaneously analyze thousands of genes. Another, more focused approach, which is both sensitive and specific, involves assessment of differential allele expression in a tissue of interest [19]. RT-coupled 5'nuclease assays, which use coding SNPs as endogenous reporters and, therefore, provide an allele-based measure of mRNA expression, can be particularly informative because they can be used to examine effects of a specific allele/haplotype on the level of expression in a tissue of interest. New strategies involving performance of genotyping arrays in order to simultaneously detect functional variation and examine levels of expression may provide a high-throughput alternative for this type of approach [20]. Parallel DNA- and mRNA-sequencing using next-generation sequencing technologies [21,22] also promises to permit genome-wide identification of functional SNPs within *cis*-acting regulatory elements. Variants that alter

exon-intron splicing or that drive alternative transcription start site usage may be similarly identified using these new strategies [23,24].

2.2 Empirical Determination of Function

Once functionally important regions have been identified, there are various algorithms for predicting features, such as exon-intron splicing, mRNA stability, or transcription factor binding sites (TFBS), many of which are freely available online (Table 1) [25-30]. In the cases in which a SNP appears to create or disrupt a putative TFBS, functionality can be demonstrated empirically using a number of different approaches. Though there are a number of methods for assessing functionality, the most common and simplistic approaches are Electrophoretic Mobility Shift Assay (EMSA, or gel shift assay) and *in vitro* reporter assays. Gel shift assays can be used to determine whether transcription factors bind to a region of interest and whether the variant in question alters the degree of binding. An alteration in the pattern of DNA-protein interactions is suggestive of the fact that a SNP alters TF binding and, potentially, transcriptional control. Whether altered DNA-protein interactions also translates into differences in transcriptional control can then be assessed with the use of an *in vitro* reporter system, in which reporter constructs containing either the nonvariant or variant alleles are tested for their abilities to drive promoter activity.

Of critical importance in performing these types of studies is that the appropriate experimental conditions be selected. For example, in selecting a nuclear extract, one should ideally use one from a tissue that expresses high levels of the gene of interest. In terms of *in vitro* reporter design, it is important that promoter constructs include critical regions for driving transcription of that particular gene. In addition, the selection of an appropriate cell line and treatment conditions are critical for generating data relevant to the genotypic and phenotypic variables of interest. An example of this process is illustrated in Figure 1.

3. Linking functional genetic variation with phenotype

As mentioned above, one limitation of performing behavioral genetic studies in primates is the typically small number of phenotyped individuals available for study, making replication in independent populations challenging, if not impossible. For this reason, we have used a multi-tiered approach in order to establish the functional significance of newly-discovered genetic variants in the rhesus macaque. With the *in vitro* functional effects as a foundation, effects of genetic variation on various traits of interest can be thoughtfully tested so as to capture the maximum amount of information while avoiding repeated testing. As an example of this, we previously reported that a SNP on a major *CRH* haplotype predicted increased sensitivity to low doses of corticosteroids. Since low doses of corticosteroids influence *CRH* expression in the non-stressed state, we predicted that this variant would contribute to phenotypic variation observed in the absence of stress [2]. However, there is an additional *CRH* promoter SNP that both increases sensitivity to forskolin stimulation and disrupts feedback inhibition by dexamethasone [31], which would be predicted to drive increased levels of *CRH* transcription during periods of stress. We have found that, consistent with our *in vitro* findings, whereas the previously published variant predicts baseline levels of *CRH* and *ACTH*, a bold temperament, and high risk alcohol drinking, the latter influences behavioral and endocrine stress reactivity and stress-induced alcohol consumption – with no effects in the absence of stress. These findings may indicate that both types of variants (those that affect transcription at baseline and during periods of stress) could increase risk for alcohol dependence, but via distinct, or even opposing, pathogenic pathways. This may be consistent with the fact that, in humans, the two major subtypes of alcohol dependence are represented by individuals who are motivated to consume alcohol either for reward or for its anxiolytic effects. From a purely pragmatic perspective, a multi-tiered approach may be particularly valuable for primate genetics studies,

since the effects of genotype on different phenotypic variables potentially permits there to be “replications” across datasets and/or centers.

4. Gene by Environment (GxE) Interactions- Lessons from the Rhesus Macaque

Stress is a universal condition of life but, if chronic, severe or occurring during critical developmental windows, it can contribute to a variety of disease vulnerabilities [32], particularly disorders of the brain [33]. Individuals exposed to early or persistent life stress are at increased risk for developing nearly all of the psychiatric disorders. Among these are post-traumatic stress disorder, alcoholism, other addictions to both licit and illicit substances, depression, anxiety disorders, and antisocial personality disorder. Stress also leads to structural losses and premature aging of the brain [34]. A hallmark of early stress exposure is increased stress reactivity throughout life, setting the stage for a variety of pathologies, all of which might be better understood by determining how early life stress epigenetically programs gene expression. Studies of heritability of psychiatric disease and studies of a small number of functional loci modifying stress response have clearly demonstrated that there are gene $\times$ stress interactions, such that the effect of inherited genotype is modified by stress [14]. In some instances, functional alleles confer risk for behavioral problems and diseases only in the context of stress exposure. Because of their complex social structures, behaviors, and genetic similarities to humans, primates are useful for modeling the heritability of traits related to human psychiatric and stress-related disorders. This is especially relevant to examination of genes that increase anxiety or interact with environmental stressors, as the key mediators of stress response in brain differ between rodents and catarrhine primates (Old World monkeys and hominoids, < 25 mya divergence from humans) [35,36]. Rhesus macaques show individual differences in temperament and stress reactivity, and we have repeatedly demonstrated the use of this primate model for identification of genetic effects and GxE interactions that translate to the human condition.

The first genetic variant to be studied in our colony of rhesus macaques was (*5-HTT-LPR*), an insertion-deletion polymorphism that was shown to be functionally equivalent to the *HTT-LPR s* allele in humans [37]. This lab has shown that the *5-HTTLPR s* allele interacts with early adversity in the form of peer rearing to influence CSF levels of the serotonin metabolite, 5-HIAA [16], and HPA axis output during social separation stress [38]. The latter was especially true among females [8] and, as adults, peer-reared females with the *l/s* genotype consumed more alcohol [39]. We have extended these analyses, examining not only the effects of *HTT-LPR* genotype on reactivity clusters generated by factor analysis of behaviors collected during stress exposure, but also looking at whether genotype predicts behavioral responses following repeated exposures to a homotypic stressor. We have found that infants with no prior history of adversity that are *s* allele carriers rapidly sensitize to stress, exhibiting high levels of anxiety-like behaviors during later stress exposures. They also are more likely to develop behavioral pathology with repeated stress. Consistent with these findings, infant macaques carrying the *HTT-LPR s* allele that have a *prior* history of stress (in the form of “peer”, or nursery, rearing) are more likely to exhibit anxiety-like and pathological behaviors during stress exposure [40]. These studies have provided an illustration of how genetic variants that are functionally equivalent to those present in humans can be studied using a nonhuman primate model and, furthermore, how primate models can be used for examining how these genetic factors interact with environmental variables, such as exposure to early adversity.

5. Using primate genetics studies to translate across species and situation

In certain instances, genetic variants that are functionally similar or orthologous to those that moderate risk for human psychiatric disorders are maintained across primate species [1]. We

have identified several examples of this phenomenon in rhesus macaques and have studied them in order to model how genetic variation moderates risk for developing psychopathology. Some of these studies have suggested the potential for convergent evolution or allelic variants being maintained by selection in both species [2,41]. These findings reinforce the potential for comparative behavioral genomics studies to demonstrate how relatively common human genetic variants, which are linked to traits that may be adaptive in certain environmental contexts, can increase vulnerability to stress-related disorders or alcohol problems

While the field of behavioral genetics is growing rapidly, most of its research is concerned with the identification of “disease alleles” or gene variation underlying what is considered pathological behavior. Its methods and findings, however, can be applied to a long-standing goal of evolutionary anthropology, to understand how changes in allele frequency can affect divergences in primate behavior. Several studies have identified associations between specific alleles and natural features of behavior and life history strategies. For example, the loss-of-function short (*s*) allele of the serotonin transporter gene promoter length polymorphism (5-*HTT*-LPR), which increases risk for developing depression in the face of adversity, has a functional equivalent in the rhesus macaque (see above). In macaques, this allele is associated with increased endocrine and behavioral stress reactivity as a function of stress exposure [38, 40]. Therefore, this variant appears to increase risk for developing psychopathology, particularly in the context of stress. Despite this, these variants have been maintained in both humans and in rhesus (in addition to some other nonhuman primate species). Moreover, in human populations in which the *s* allele is rare, another loss-of-function variant on the *L* allele background (LA>LG) is present at a higher frequency [42]. In humans, there is also a functional VNTR in the second intron, which appears to be functional [43]. This VNTR is present in a number of primate and nonprimate species and is polymorphic in a number of hominoid species [44].

Although SNPs generally are not conserved across species, there are instances in which functionally similar SNPs occur in the human and rhesus macaque [2,3,45]. Recently, it was demonstrated that gain-of-function 5-*HTT* SNPs have arisen and been maintained in both rhesus and in humans, suggesting that both gain- and loss-of-function variants may be under selection in primates [3]. It is of interest that 5-*HTT* variation not only predicts individual differences in impulse control and stress reactivity [8,16,46], but that it is also associated with adaptive traits in free-ranging macaques, such as earlier male dispersal [47] and male reproductive timing [48]. Whether allelic variation at 5-*HTT* predicts “adaptive” traits in humans has not been elucidated.

As another example, in both rhesus and in humans, there are nonsynonymous SNPs in the portion of the *OPRM1* gene that encodes the N-terminal domain of the receptor (C77G in rhesus macaque and A118G in human), and we and others have observed there to be some similar functions *in vivo* [49-51]. In humans, the 118G allele is believed to increase the likelihood that an individual will abuse alcohol because it increases alcohol-induced euphoria [50]. We have shown that rhesus carrying the 77G allele exhibit increased alcohol-induced stimulation (a marker for the euphorogenic effects of alcohol) and that G allele carriers also consume more alcohol in the laboratory [51]. It would stand to reason that *OPRM1* variation might predict sensitivity to natural rewards as well. Based on the fact that these two variants confer similar functional effects, that both are observed at relatively high frequencies, and, further, that there is an extended region of LD with the A118G allele in humans [52,53], it might be hypothesized that they have evolved as result of similar selective pressures in the two species. Data to directly address this hypothesis are presently not available. However, studies performed in the macaque demonstrate this variant to predict behaviors that could theoretically be under selection. The 77G allele predicts aggressive behavior [45], and we have shown that the G allele carriers form stronger attachment bonds with their mothers during infancy [41], especially as a function of

repeated maternal separation. It is of interest that the effects that we observed during repeated exposures to maternal separation and reunion are similar to those that you might observe during periods of alcohol intake and withdrawal. These types of studies highlight how traits that could have conferred selective advantage at some point in the evolutionary history of humans can increase risk for addictive disorders in modern society.

6. Emerging strategies

With the emergence of new DNA sequencing technologies (i.e., Solexa, 454 and SOLiD), whole genome sequencing, focused resequencing, transcriptome analysis, and epigenetics studies can be performed very efficiently and on a single platform [54,55]. We have been using new high throughput sequencing methodologies for performance of ChIP-Sequencing (Chromatin Immunoprecipitation followed by sequencing, or ChIP-SEQ) of DNA isolated from macaque hippocampus and immunoprecipitated with an antibody toward H3K4me3, which is a form of the H3 histone protein that is associated with active promoters. This method, therefore, may obtain sequence surrounding “active” genes. As such, this approach may be particularly powerful for identifying genetic variants that are relevant to behavioral differences in the rhesus macaque. While only 25,000 putative (and fewer than 1000 annotated) SNPs have been identified in the rhesus macaque [56], more than 700,000 SNPs (99% novel-identified) were identified with this method [57]. These appear to be valid, given the transition to transversion ratio, and we have validated some of these against SNPs that we previously identified with the use of labor-intensive, PCR-based sequencing. By using this approach, we can efficiently detect sequence variation genome-wide, permitting the creation of custom genotyping arrays for performing large-scale, genome-wide studies in the rhesus macaque.

Summary

Because of their complex social structures, behaviors, and genetic similarities to humans, nonhuman primates are useful for studying how genetic factors influence alcohol consumption. The neurobiological systems that influence addiction vulnerability may do so by acting on reward pathways, behavioral dyscontrol, and vulnerability to stress and anxiety. Rhesus macaques show individual differences in alcohol response and temperament, and such differences are influenced by genetic variants that are orthologous to those present in humans. We anticipate that genetic studies performed in rhesus macaques may have translational value for investigating not only effects of genotype on stress reactivity, temperament, and alcohol response and consumption, but may also allow us to investigate how these variants influence alcohol-induced neuroadaptation, neuropathology, and treatment response.

Text Box

There are a number of analytical methods that rely on examination of patterns of intra- and inter-specific genetic variation that can potentially inform us of regions under genetic selection and, by extension, the existence of functional alleles. The existence of long range, alternative (yin-hang) haplotypes is proposed to be indicative of an allele that is being (or has been) maintained by balancing selection. The idea is that the “favorable allele was going to fixation at a rate greater than that of the breakdown of the haplotype, creating an area of reduced haplotype diversity and increased LD (Smith and Haigh, 74). Extended regions of complete LD (D') can, therefore, be an indication of the presence of a functional locus. Conversely, there have also been suggestions that recombination hotspots may occur in proximity to allelic variants that are under selection. The hypothesis is that, if recombination in that region is reduced, then the selective pressure on functional alleles is increased. Using data available through HapMap (<http://www.hapmap.org/>), we can examine recombination rates in genomic areas of interest. Here, we show that there is a recombination hotspot in a

region 3' to the human *NPY* gene. This is true across Asian (shown), African, and European and populations, potentially indicating there to be an ancient functional allele within this region that has been under selection in humans. Loci in this region may, therefore, be good candidates for performing genetic association studies.

Haplotype map demonstrating areas of increased recombination in a region 3' to the human *NPY* gene (Recombination Hot Spots, indicated in black).

In the last several years, the genomes of a wide variety of vertebrate species have been sequenced, and these data are publically available through various websites, such as the UCSC genome browser (<http://genome.ucsc.edu/cgi-bin/hgGateway>). This provides genome information for species for which genetic variation has not yet been determined, permitting efficient primer design for sequencing. But in addition to this, there are features to the UCSC genome browser that permit extraction of information about sequence conservation. Here, we show an example of the output from UCSC gateway, in which the human *NPY* gene is depicted. Shown are the exons, mRNAs, and ESTs (expressed sequence tags) in addition to CpG islands and conserved TFBS. In addition, measures of both inter- and intraspecific conservation are shown (Conservation and Cons Indels MmCf, respectively). High degrees of conservation are suggestive of purifying selection, indicating functionally important regions. As is typically the case, the coding regions for *NPY* (particularly, Exon 2, which encodes the *NPY* peptide) are those that show the highest degrees of conservation in this region. However, some non-coding segments also appear to be under selection, which may indicate them to be important to regulation of *NPY* transcription or other processes. Of interest, there is also a CpG-rich region at the *NPY* locus, suggesting that epigenetic processes may impact *NPY* function. As such, *NPY* may be an excellent candidate for examining GxE interactions.

Output from USCS, demonstrating sequence conservation at the human *NPY* gene

Acknowledgments

I thank Qiaoping Yuan for the recombination hotspot figure, and the investigators at NIAAA and NICHD for their contributions to the work discussed in this review. The work described was funded by the NIAAA and NICHD intramural programs.

Literature Cited

1. Barr CS, Goldman D. *Addict Biol* 2006;11:374–385. [PubMed: 16961765]
2. Barr CS, Dvoskin RL, Yuan Q, Lipsky RH, Gupte M, Hu X, Zhou Z, Schwandt ML, Lindell SG, McKee M, Becker ML, Kling MA, Gold PW, Higley JD, Heilig M, Suomi SJ, Goldman D. *Arch Gen Psychiatry* 2008;65:934–944. [PubMed: 18678798]
3. Vallender EJ, Priddy CM, Hakim S, Yang H, Chen GL, Miller GM. *Genes Brain Behav* 2008;7:690–697. [PubMed: 18445138]
4. Fairbanks LA, Newman TK, Bailey JN, Jorgensen MJ, Breidenthal SE, Ophoff RA, Comuzzie AG, Marin LJ, Rogers J. *Biol Psychiatry* 2004;55:642–647. [PubMed: 15013834]
5. Rogers J, Kohcunov P, Lancaster J, Shelledy W, Glahn D, Blangero J, Fox P. *Human Brain Mapp* 2007;28:576–583.
6. Risch N, Merikangas K. *Science* 1996;273:1516–1517. [PubMed: 8801636]
7. Long AD, Langley CH. 1999;9:720–731.
8. Barr CS, Newman TK, Schwandt ML, Shannon C, Dvoskin RL, Lindell SG, Taubman J, Thompson W, Champoux M, Lesch KP, Goldman D, Suomi SJ, Higley JD. *Proc Natl Acad Sci U S A* 2004;101:12358–12363. [PubMed: 15302939]

9. Zhou Z, Zhu G, Hariri AR, Enoch MA, Scott D, Sinha R, Virkkunen M, Mash DC, Lipsky RH, Hu XZ, Hodgkinson CA, Xu K, Buzas B, Yuan Q, Shen PH, Ferrell RE, Manuck SB, Brown SM, Hauger RL, Stohler CS, Zubieta JK, Goldman D. *Nature* 2008;452:997–1001. [PubMed: 18385673]
10. Robin RW, Chester B, Rasmussen JK, Jaranson JM, Goldman D. *Am J Psychiatry* 1997;154:1582–1588. [PubMed: 9356568]
11. Young LJ, Winslow JT, Nilsen R, Insel TR. *Behav Neurosci* 1997;111:599–605. [PubMed: 9189274]
12. Rosso L, Keller L, Kaessmann H, Hammond RL. *Biol Lett* 2008;23:375–378. [PubMed: 18430667]
13. Meyer-Lindenberg A, Kolachana B, Gold B, Olsh A, Nicodemus KK, Mattay V, Dean M, Weinberger DR. *Mol Psychiatry*. 2008Epub ahead of print
14. Caspi A, Moffitt TE. *Nat Rev Neurosci* 2006;7:583–590. [PubMed: 16791147]
15. Newman TK, Syagailo YV, Barr CS, Wendland JR, Champoux M, Graessle M, Suomi SJ, Higley JD, Lesch KP. *Biol Psychiatry* 2005;57:167–172. [PubMed: 15652876]
16. Bennett AJ, Lesch KP, Heils A, Long JC, Lorenz JG, Shoaf SE, Champoux M, Suomi SJ, Linnoila MV, Higley JD. *Mol Psychiatry* 2002;7:118–122. [PubMed: 11803458]
17. Chen GL, Vallender EJ, Miller GM. *Hum Genet* 2008;122:645–657. [PubMed: 17972101]
18. Nembaware V, Lupindo B, Schouest K, Spillane C, Scheffler K, Seoighe C. *BMC Genomics* 2008;9:265. [PubMed: 18518984]
19. Pastinen T, Hudson TJ. *Science* 2004;306:647–650. [PubMed: 15499010]
20. Serre D, Gurd S, Ge B, Sladek R, Sinnett D, Harmsen E, Bibikova M, Chudlin E, Barker DL, Dickinson T, Fan J, Hudson TJ. *PLoS Genet* 2008:e1000006. [PubMed: 18454203]
21. Wang Z, Gerstein M, Snyder M. *Nat Rev Genet* 2009;10:57–73. [PubMed: 19015660]
22. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niaz F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. *Nature* 2008;452(7189):872–876. [PubMed: 18421352]
23. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. *Nat Genet* 2008;40:1413–1415. [PubMed: 18978789]
24. Sultan M, Schulz MH, Richard H, Magen A, Kligenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, O'Keefe S, Haas S, Vingron M, Lehrach H, Yaspo ML. *Science* 2008;320(5891):956–960. [PubMed: 18599741]
25. Kel-Margoulis OV, Kel AE, Reuter I, Deineko IV, Wingender E. *Nucleic Acids Res* 2002;30:332–334. [PubMed: 11752329]
26. Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, Kel OV, Ignatieva EV, Ananko EA, Podkolodnaya OA, Kolpakov FA, Podkolodny NL, Kolchanov NA. *Nucleic Acids Res* 1998;26:362–367. [PubMed: 9399875]
27. Shahi P, Loukianjouk S, Bohne-Lang A, Kenzelmann M, Kuffer S, Maerens S, Eils R, Grone HJ, Gretz M, Brors B. *Nucleic Acids Res* 2006;34:D115–118. [PubMed: 16381827]
28. Burge C, Karlin S. *J Mol Biol* 1997;268:78–94. [PubMed: 9149143]
29. Sumazin P, Chen G, Hata N, Smith AD, Zhang T, Zhang MQ. *Bioinformatics* 2005;21:31–38. [PubMed: 15333453]
30. Zuker M. *Nucleic Acids Res* 2003;31(13):3406–3415. [PubMed: 12824337]
31. Barr CS, Dvoskin RL, Gupte M, Sommer W, Sun H, Schwandt ML, Lindell SG, Kasckow JW, Suomi SJ, Goldman D, Higley JD, Heilig M. *Proc Nat Acad Sci*. in press
32. McEwen BS. *Ann NY Acad Sci* 2001;933:265–277. [PubMed: 12000027]
33. Sapolsky RM. *Science* 2005;308:648–652. [PubMed: 15860617]
34. Landfield PW, McEwen BS, Sapolsky RM, Meaney MJ. *Science* 1996;272:1249–1251. [PubMed: 8650531]
35. Sanchez MM, Young LJ, Plotzky PM, Insel TR. *J Comp Neurol* 1999;408:365–377. [PubMed: 10340512]
36. Sanchez MM, Young LJ, Plotzky PM, Insel TR. *J Neurosci* 2000;20:4657–4668. [PubMed: 10844035]

37. Lesch KP, Meyer J, Glatz K, Flugge G, Hinney A, Hebebrand J, Klauck S, Poustka A, Poustka F, Benger D, Mossner R, Riederer P, Heils A. *J Neural Transm* 1997;104:1259–1266. [PubMed: 9503271]
38. Barr CS, Newman T, Shannon C, Parker CC, Dvoskin RL, Schwandt ML, Champoux M, Lesch KP, Goldman D, Suomi SJ, Higley JD. *Biol Psychiatry* 2004;55:733–8. [PubMed: 15039002]
39. Barr CS, Newman T, Lindell SG, Shannon C, Champoux M, Lesch KP, Goldman D, Suomi SJ, Higley JD. *Arch Gen Psychiatry* 2004;61:1146–1152. [PubMed: 15520362]
40. Spinelli S, Schwandt ML, Lindell SG, Newman TK, Heilig M, Higley JD, Suomi SJ, Goldman D, Barr CS. *Dev Psychopathol* 2007;19:977–987. [PubMed: 17931429]
41. Barr CS, Schwandt ML, Lindell SG, Higley JD, Maestripieri D, Goldman D, Suomi SJ, Heilig M. *Proc Natl Acad Sci* 2008;105:5277–5281. [PubMed: 18378897]
42. Hu XZ, Lipsky RH, Zhu G, Akhtar LA, Taubman J, Greenberg BD, Xu K, Arnold PD, Richter MA, Kennedy JL, Murphy DL, Goldman D. *Am J Hum Genet* 2006;78:815–826. [PubMed: 16642437]
43. Fiskerstrand CE, Lovejoy EA, Quinn JP. *FEBS Lett* 1999;458:171–174. [PubMed: 10481059]
44. Soeby K, Larsen SA, Olsen L, Rasmussen HB, Werge T. *Am J Hum Genet* 2005;136:53–57.
45. Miller GM, Bendor J, Tiefenbacher S, Yang H, Novak MA, Madras BK. *Mol Psychiatry* 2004;9:99–108. [PubMed: 14699447]
46. Champoux M, Bennett AJ, Shannon C, Higley JD, Lesch KP, Suomi SJ. *Mol Psychiatry* 2002;7:1058–1063. [PubMed: 12476320]
47. Trefilov A, Berard J, Krawczak M, Schmidtke J. *Behav Genet* 2000;30:295–301. [PubMed: 11206084]
48. Krawczak M, Trefilov A, Berard J, Bercovitch F, Kessler M, Sauermann U, Croucher P, Nürnberg P, Widdig A, Schmidtke J. *Biol Reprod* 2005;72:1109–1113. [PubMed: 15635127]
49. Chong RY, Oswald L, Yang X, Uhart M, Lin PI, Wand GS. *Neuropsychopharmacology* 2006;31:204–211. [PubMed: 16123758]
50. Ray LA, Hutchison KE. *Alcohol Clin Exp Res* 2004;28:1789–1795. [PubMed: 15608594]
51. Barr CS, Schwandt ML, Lindell SG, Chen SA, Goldman D, Suomi SJ, Higley JD, Heilig M. *Arch Gen Psychiatry* 2007;64:369–376. [PubMed: 17339526]
52. Luo X, Zuo L, Kranzler H, Zhang H, Wang S, Gelernter J. *Am J Med Genet B Neuropsychiatr Genet* 2008;147B:1028–1039. [PubMed: 18213616]
53. Zhang H, Luo X, Kranzler HR, Lappalainen J, Yang BZ, Krupitsky E, Zvartau E, Gelernter J. *Hum Mol Genet* 2006;15:807–819. [PubMed: 16476706]
54. Mardis ER. *Nat Methods* 2007;4(8):613–614. [PubMed: 17664943]
55. Gilchrist DA, Fargo DC, Adelman K. *Methods*. 2009
56. Rhesus Macaque Genome Sequencing and Analysis Consortium. Gibbs RA, Rogers J, et al. *Science* 2007;316(5822):222–234. [PubMed: 17431167]
57. Barr CS, et al. Genome-Wide Identification of Epigenetic Markers of Early Life Stress in Primate Brain. *Biol Psychiatry*. 2009

Figure 1. Factors to Consider in Characterizing the Functionality of a Promoter Variant
A. Schematic of a gene regulatory region, with indications of the transcription factor binding sites (TFBS) and agents empirically demonstrated to affect transcription in brain. **B.** Electrophoretic Mobility Shift Assay (EMSA) result, demonstrated with nuclear extract from brain tissue. As different transcription factors might alter expression in different tissues, the selection of nuclear extract for this experiment is critical. The gel shift assay demonstrated DNA-protein interactions when a nuclear extract and ³²P-labeled DNA sequence are co-incubated. Since protein binding to labeled DNA will retard its migration, shifted bands indicate DNA-protein complexes. This result shows there to be several DNA-protein complexes (closed arrows) that are increased as a function of the A allele (vs. allele B),

suggesting increased transcription factor binding for the A allele. Lanes 2, 3, 5, and 6 have cold competitors added to the assay to test for specificity of these interactions. For lane 2, the A competitor is added, and for lane 3, the B competitor is added. The results for the A allele, which show that the cross-competitor (B) is less effective than the A competitor, indicate that the increased DNA-protein interactions for the A allele are not reflective of non-specific DNA binding. **C.** Reporter assay, demonstrating that both the cell line and the nonvariant construct selected are adequate for assessing the effects of relevant agents on promoter activity *in vitro*. Using this system, a stimulatory effect of both TPA and Forskolin (Forsk) are observed, which addition of Dexamethasone (Dex) inhibits Forskolin-stimulated promoter activity. These results are consistent with what has been shown to occur for this regulatory region in brain and, therefore, demonstrate both the construct and cell system to be appropriate. The next step would be to compare the nonvariant and variant alleles in order to determine whether the SNP indicated in panel A influenced promoter activity *in vitro*.

Table 1
Links for *in silico* algorithms useful for screening for functional genetic variation

GeneScan http://genes.mit.edu/GENSCAN.html
Intron/Exon Splice Sites http://bioweb.uwlax.edu/GenWeb/Molecular/Seq_Anal/Splice_Sites/splice_sites.htm
miRNA finder http://www.ma.uni-heidelberg.de/apps/zmf/argonaute/index.html
mRNA Stability http://www.bioinfo.rpi.edu/applications/mfold/
ooTFD http://www.ifti.org/
Tess http://www.cbil.upenn.edu/cgi-bin/tess/tess
TFSearch http://www.cbrc.jp/research/db/TFSEARCH.html
TfSitescan http://www.ifti.org/cgi-bin/ifti/Tfsitescan.pl
Transcription regulatory Element Database http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process=home