Methods 49 (2009) 56–62

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Review Article

Strategies for performing genotype–phenotype association studies in nonhuman primates

Christina S. Barr *

Laboratories of Neurogenetics and Clinical and Translational Studies, NIH/NIAAA, 5625 Fishers Lane, Rm. 3S-32, Rockville, MD 20852, USA

ARTICLE INFO

Article history:
Accepted 22 May 2009
Available online 6 June 2009

Keywords:
Macaque
Genetic
Comparative
Primate
EMSA
Reporter assay
GxE interactions

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.

Published by Elsevier Inc.

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.

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 [4,5]. 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 com-

* Fax: +1 301 435 9278.
E-mail address: cbarr@mail.nih.gov.

1046-2023/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.ymeth.2009.05.014

plex 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 interspecific 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.

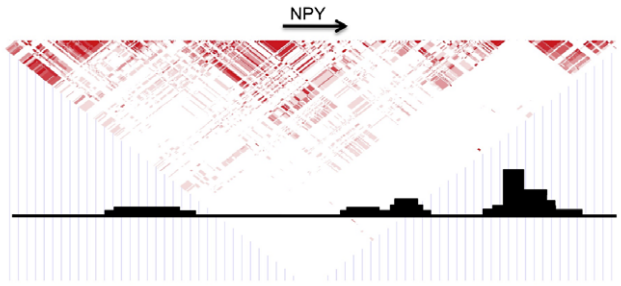
2.1. Identification of functional variants

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 1), 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 noncoding, regulatory regions as well [2,3,8,15–17].

Text Box 1

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-yang) 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).

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 2).

In the laboratory, *cis*-acting functional SNPs can be identified or inferred using a number of different approaches. High-throughput