Published in final edited form as:

J Med Primatol. 2011 August ; 40(4): 224–232. doi:10.1111/j.1600-0684.2011.00486.x.

High-Throughput SNP Discovery and the Search for Candidate Genes for Long-Term SIVmac Nonprogression in Chinese Rhesus Macaques (*Macaca mulatta*)

J Satkoski Trask^{1,2}, WT Garnica^{1,2}, RS Malhi^{3,4}, S Kanthaswamy^{1,2}, and DG Smith^{1,2}

¹ Department of Anthropology, University of California, Davis, CA, 95616, USA

² California National Primate Research Center, University of California, Davis, CA, 95616, USA

³ Department of Anthropology, University of Illinois, Urbana-Champaign, IL, 61801, USA

⁴ Institute for Genomic Biology, University of Illinois, Urbana-Champaign, IL, 61801, USA

Abstract

Genetic differences between Indian and Chinese rhesus macaques contribute to the phenotypic variance of clinical trials, including experimental infection with SIVmac. The completion of the rhesus genome has facilitated the discovery of several thousand markers. Although the marker density necessary for whole genome association mapping of phenotypes has not yet been achieved, a SNP map will help researchers investigate variation in candidate genes. We developed a genome-wide SNP map for rhesus macaques containing 3,869 validated markers with an average distance of 0.88 megabases. We used the program VarLD to identify genomic areas with significant differences in linkage disequilibrium (LD) between Indian-derived and Chinese rhesus macaques, assuming that these areas provide the greatest potential for differential selection in these regional populations. These genomic areas provide entry to more detailed study of gene function. This method is also applicable to the study of differences in biomarkers between regional populations of other species.

Keywords

Linkage; Disequilibrium; Primate; Biomedicine; Selection

Introduction

The National Institutes of Health currently fund eight National Primate Research Centers in the United States, housing approximately 26,000 animals, mostly Indian origin rhesus macaques (*Macaca mulatta*). Although many of these animals are utilized specifically as breeders, most research incorporating rhesus macaques employ them as a model of human disease, including colon cancer, type-2 diabetes and asthma, but especially HIV [Hadfield, 2001]. Failures of HIV vaccine clinical trials have underscored an increased need for more innovative research tools for non-human primates [Hayden, 2008]. The value of the rhesus model was illustrated by the discovery that pre-exposure prophylaxis with topical drugs can prevent HIV infection in humans, a discovery initially made in rhesus macaques and one of the first successful translations from rhesus research to human clinical studies [Folks, 2010].

Corresponding Author: Jessica Satkoski Trask, Department of Anthropology, 330 Young Hall, One Shields Avenue, University of California, Tel: 530-752-8570, Fax: 530-752-8885, jasatkoski@ucdavis.edu.

Work was performed at the Department of Anthropology, University of California-Davis.

However, this transition from basic to translational research, and from there to preclinical and clinical investigation, requires a more complete understanding of primate models and more effective ways to analyze them.

Linkage analysis has been widely used in studies of the human genome. However, multiple limitations of this method have been identified, specifically that genomic variants influencing susceptibility many diseases are rare, requiring sample sizes in the tens of thousands to support statistical associations, and the need for high SNP density in areas of low linkage disequilibrium (LD) [Wang et al., 2005]. These limitations become much more pronounced in rhesus macaques, where sample sizes can be quite limited (although deep pedigrees, uncommon in human studies, are often available) [Hadfield, 2001] and density of validated SNPs is currently limited to just over a thousand markers, as opposed to the millions available for humans.

Despite the boundaries imposed on genetic investigation of the rhesus model, or perhaps because of them, alternate approaches to genomics are necessary. Development of a single-nucleotide polymorphism (SNP) map for the rhesus macaque will allow researchers to understand the background genetic variation of research subjects and directly investigate variation in genes of interest. In addition, SNPs could be useful markers for colony management, conservation and paternity assignment [Hadfield, 2001], areas of little interest to researchers focusing on the human genome.

To this end, the completion of the rhesus genome map has facilitated the discovery of several thousand markers, but the marker density necessary for whole genome association mapping of phenotypes has not yet been achieved. Currently, the rhesus macaque has not been characterized as a genetic model organism and the search for rhesus candidate genes often begins with the interrogation of the human genome. Many research groups have reported inter-individual variation in progression of SIVmac infection, and the relationship between long-term nonprogression to simian-AIDS and Chinese ancestry [Joag et al., 1994; Trichel et al., 2002; Ling et al., 2002a; Ling et al., 2002b]. However, long-term nonprogression and elite controller status appear to be much more common in rhesus macaques than in humans, suggesting that a more efficient search for candidate genes would begin with the investigation of variation present in the genome of rhesus macaques, rather than that of humans. This study utilizes differences in linkage disequilibrium between genomes of Indian-derived and Chinese rhesus macaques to identify regions potentially under selection. We have identified genes in these areas and used available functional information to identify potential candidates for the primary immunological difference between these two populations – SIVmac nonprogression. The prioritized gene list will provide a point of entry to more detailed studies of gene function.

Materials and methods

Institutional Animal Care and Use Committee approval for this study was obtained and all Federal and State guidelines for animal care and use were followed. We developed a genome-wide SNP map for rhesus macaques containing 4,140 validated markers with an average inter-marker distance of 0.88 megabases. Our method of SNP discovery is described in detail in [Mahli et al. 2007]; the goal of this method was to maximize the number of SNPs discovered that were variable between (but not necessarily within) populations of Indian and Chinese rhesus macaques. Briefly, DNA from a western Chinese-origin rhesus macaque was submitted to 454 Life Sciences [Roche Diagnostics, Branford, CT] for large-scale parallel pyrosequencing, producing a total of 339,967 reads with an average read length of 104 bp. These reads were aligned against the (Indian-origin) rhesus genome [Gibbs et al., 2007] and approximately 23,000 polymorphisms were identified bioinformatically.

The chromosome and nucleotide position of each pyrofragment, as well as single copy number, was confirmed with the BLAST [Altschul et al., 1990] network service at the website of the National Center for Biotechnology Information [NCBI, www.ncbi.nlm.nih.gov]. Only fragments producing a single, high-quality match were selected for further analysis. Quality-checking and marker validation is described in [Satkoski et al. 2009]. Markers were spaced as equally as possible across chromosomes while avoiding repeat and centromeric regions. We employed Illumina [San Diego, CA] GoldenGate technology to genotype the quality-checked markers in a geographically representative sample of 29 Indian-origin and 21 Chinese rhesus macaques (see Table 1 for sample representation). Markers without a minimum quality (GenTrain) score of 0.4 were excluded from further analysis. A maximum of 5% missing data was allowed for both markers and individuals, resulting in 3869 markers on all 20 autosomes and the X chromosome.

We used the program VarLD [Ong and Teo, 2010] to identify areas of the rhesus macaque genome with the greatest difference in LD between our samples of Indian-origin and Chinese rhesus macaques. Although there are several accepted methods for calculating LD, VarLD expresses the value as r^2 , or the correlation coefficient between allele frequencies of two markers [Hedrick and Kumar, 2001]. VarLD calculates LD separately for each population. LD is calculated within a sliding window of 50 SNPs, which then moves downstream by one SNP and recalculates LD. The r^2 value for each window is subtracted from the value of the corresponding window in the other population to determine the raw VarLD score. The raw scores are pooled and standardized to give them a zero mean and unit variance. Percentiles from this distribution are calculated for each LD difference and assigned the corresponding p value.

To create a list of candidate genes for phenotypic differences between Indian and Chinese rhesus macaques, the identified regions were compared to the NCBI rhesus genome annotation. These regions were classified as outside of a known gene, inside of a known gene, or located within a gene that has received preliminary annotation through gene prediction.

The identified genes were used as search terms in the Gene Ontology Website (AmiGO, <http://www.geneontology.org>) to identify the ontology term associations of each identified product. As term associations are not available for rhesus macaques, human term associations and reference data from the Universal Protein Resource [UniProt, www.uniprot.org] were used. The one exception was gene Eaf6, which has only been fully annotated in *Drosophila* (fruit fly). In this case, fruit fly term associations were used, and reference data was taken from FlyBase [flybase.org]. All term associations were then counted with the GO Classification Counter Tool [Hu et al., 2008] and divided into the three main GO classes, “molecular function”, “cellular component” and “biological processes”, as well as up to 127 child terms. These functional clues allowed us to prioritize genes for their potential to influence SIVmac nonprogression.

Results

For each chromosome, we predicted the optimal distance between markers assuming perfectly even distribution by dividing the number of markers by the length of the chromosome in base pairs. The actual average distance between markers is the mean value of the distances between adjacent markers on each chromosome. The genomic distribution of SNPs is shown in Figure 1 and each average distance is equal to or less than the prediction, with the exception of chromosome 11. Although placing SNP markers preferentially in linkage blocks to act as “tag” markers is more efficient, it requires a

detailed knowledge of the position and extent of haplotypes within the genome [Johnson et al., 2001]. In the absence of this information, positioning SNPs equidistant throughout the genome maximizes the probability of identifying regions of high LD.

Forty-one statistically significant LD differences between Indian-origin and Chinese rhesus macaques were detected on chromosomes one, four, five and eleven. All significant LD differences are summarized in Table 2. The standardized VarLD scores for each of these chromosomes are shown in Figure 2. Twenty-six of the identified positions were not in any known or predicted gene. Thirteen of the identified positions were in predicted genes (including one position in two overlapping predicted genes) and two were in annotated genes. Chromosome one contained the only region with an LD difference in the 99.99th percentile. Although there were no known genes at this position, LD differences at adjacent positions containing two genes, *ELAVL4* and *EPS15* were in the 99th percentile, forming a block of highly significant LD differences between nucleotide positions 41807851 and 56902623.

Two genes on chromosome 11, *CCDC65* and *CG32133-PA*, had no association terms and were not analyzed further, leaving 13 genes of interest. The molecular function class, although it contained fewer genes than the other classes (84.6%), had the greatest representation within child terms (37.9%). In contrast the biological process class represented 92.3% of genes and 28.2% of total child terms. The corresponding percentages for the cellular component class were 92.3% and 34.0%, respectively.

Forty-eight unique child terms were associated with the 13 genes of interest. These child terms, the percentage of genes associated with each and the percentage of total child term usage is shown in Table 3. The most commonly occurring child term (15.5% of term occurrences) was “cell”, referring to proteins forming the cell, including in the plasma membrane, cell wall or cell envelope. “Binding”, referring to any non-covalent molecular interaction, was the second most common, with 10.6% of term occurrences. Although binding between gene products and molecules is a major function of the innate immune system’s identification of pathogens, this term is too general. By focusing on DNA, RNA and nucleotide binding specifically, interactions between the host and viral proteins may be inadvertently eliminated, but protein-protein binding not related to the immune system is eliminated as well, reducing the probability of spurious identification. Three genes had nucleotide binding functions (DNA or RNA): *ELAVL4*, *MAST2* and *HIVEP3*. All three of these are located at the proximal end of chromosome 1 in the area of LD difference within or exceeding the 99th percentile.

Discussion

Linkage disequilibrium is employed in this study as a statistical association between nucleotides, identifying alleles at different loci that co-occur at a rate which exceeds that predicted by chance. In the absence of other forces, the co-occurrence of alleles can be predicted by the distance between the loci, with the probability of recombination during meiosis increasing with greater distance [Lynch and Walsh, 1998]. Hernandez et al. [2007] found that LD decayed much faster in Chinese rhesus than in Indian rhesus, with r^2 values at 10 kilobases of 0.15 and 0.52, respectively. However, with an average distance between markers of 880 kilobases, the present SNP map greatly exceeds the distance at which recombination differences in LD between Chinese and Indian individuals would be expected, and alternative explanations for LD differences may be invoked, such as admixture, genetic bottleneck/drift and selection.

There are several well-characterized phenotypic differences between Indian and Chinese *Macaca mulatta*. Chinese-origin males tend to be heavier, longer and taller, and exhibit more sexual dimorphism than their Indian counterparts [Clarke and O'Neil, 1999; Hamada et al., 2005]. Animals with partial Chinese ancestry were found to have greater total protein, higher erythrocyte counts and greater corpuscular hemoglobin than Indian animals [Champoux et al., 1996]. There is wide belief that Chinese rhesus macaques are more aggressive, and individuals of partial Chinese ancestry exhibit more aggression towards both humans and conspecifics, and are more impulsive [Champoux et al., 1997] than those of solely of Indian ancestry.

The phenotypic differences between Indian and Chinese *Macaca mulatta* are of biomedical importance when responses to experimental factors are under even partial genetic control. Conversely, genetic differences between Indian and Chinese rhesus macaques may contribute to the variance of phenotypes studied in clinical trials such as experimental infection by some strains of SIVmac. Candidate genes for the latter phenotype, including those at the CCL3-like [Degenhardt et al., 2009] and major histocompatibility complex (MHC) class I loci have been reported. Of special interest are the class I alleles *Mamu-A*01* [Mothe et al., 2003], *Mamu-B*08* [Loffredo et al., 2007] and *Mamu-B*17* [Mothe et al., 2002], all of which have demonstrated a relationship with control of SIVmac239 viral replication. The *Mamu-B*17* allele is known to bind to at least 16 SIVmac239 peptides to restrict SIV-specific cellular immune response [Mothe et al., 2002], although the specific actions of the other Mamu class I alleles is less well understood. Unfortunately, the rhesus macaque has undergone a major expansion of the class I genes, from six in humans to as many as 22 [Daza-Vamenta et al., 2004]. This duplication has precluded high-throughput interrogation of MHC sequence, due to the difficulty of alignment. Thus, the rhesus MHC, found on chromosome 6, is unrepresented in the SNPs described above. The lack of MHC coverage necessitates a focus on candidate genes that, while still interacting with viral nucleotide or peptide motifs, are located elsewhere in the genome.

Potential explanations for the block of LD difference on chromosome 1 include a severe genetic bottleneck during a glacial event or admixture during re-dispersal from glacial refugia following such an event. However, we have attempted to mitigate the influence of such population-specific effects on LD by genotyping geographically diverse samples of Indian and Chinese rhesus macaques [sampling locations described in Satkoski et al., 2010]. Moreover, such effects should be detectable throughout the entire range of the genome rather than be restricted to a single, or a very few, region(s). A large chromosomal inversion or translocation that reached high frequency in either the Indian or Chinese population could create instant, high LD in the corresponding region. A test of this hypothesis must await the availability of a scaffolded genome from a Chinese rhesus macaque.

While gene flow and the stochastic effects of genetic drift influence the genome homogeneously, selection can target only one or a few genomic regions, as found in the present study. An evaluation of selection as a potential cause of LD differences requires an examination of the functions of the identified candidate genes. Of the genes exhibiting statistically significant LD between Indian and Chinese rhesus macaques, *HIVEP3* is the most obvious candidate for influencing SIVmac nonprogression. The product of *HIVEP3* is a zinc finger protein with a domain known to bind to the κ B motif of the HIV-1 long terminal repeat. It has no other known function in the regulation of viral or cellular promoters [Seeler et al., 1994]. However, the human sequence of this gene has a sixth exon not found in mice, but present in rhesus macaques, potentially differentiating the function of this gene in primates relative to other mammals [Hicar et al., 2001]. *HIVEP3* has also undergone positive selection in the human lineage relative to chimpanzees [Vamathevan et al., 2008]. A study of over 500 thousand SNPs in human populations found that *HIVEP3*

had at least two intronic SNPs with significantly high genetic distance between African and European populations [Hughes et al., 2008], suggesting genetic differentiation due to selective pressure exerted by HIV-1 infection. The specific functional relationship between *HIVEP3* polymorphism and resistance to HIV in human populations or other organisms is currently poorly understood.

In contrast, *ELAVL4* and *MAST2* appear to be only transcribed in the brain although the former is also expressed in small cell lung tumors [Szabo et al., 1981; Prehaud et al., 2010]. *ELAVL4* appears to have little relationship to SIV infection. Initially, this would also seem to be the case for *MAST2*. It has both a kinase domain and a PDZ scaffolding domain [Walden and Cowan, 1993]. However, Prehaud et al. [2010] remarked that one early feature of viral infection is viral recruitment of large cellular sub-membrane protein scaffolds with the catalytic activity of these scaffolds is regulated through signals transmitted by PDZ domains. The authors infected mice with both virulent and attenuated rabies and found that in uninfected cells, *MAST2* is distributed uniformly throughout the neuron; after infection it redistributes to areas where the viral G protein accumulates. Infection with the virulent strain had the effect of silencing *MAST2*, due to a competing PDZ domain. This prevents neuronal apoptosis and facilitates the spread of the virus. Thus, rabies infection in wild populations is a potential source of selection on *MAST2*.

Given that SIVmac does not exist in wild populations of rhesus macaque, it is unlikely that it has operated as a significant selective force in *M. mulatta* generally. It was not known to exist in captive populations prior to the early 1970's and may have emerged at the California National Primate Research Center (CNPRC) [Mansfield et al., 1995], SIVmac is even more unlikely to have exerted differential selection on Indian-derived versus Chinese individuals, who are both housed together and known to hybridize at the CNPRC [Kanthaswamy et al., 2009]. Thus, a parsimonious but very preliminary hypothesis is that the high LD between Indian and Chinese rhesus macaques for *HIVEP3* results from that gene hitchhiking on another gene under selection due to SIVmac elsewhere in the chromosome one linkage block, possibly (but certainly not limited to) one of the other nine candidate genes in the region. For example, rabies infection has been observed in wild Indian rhesus macaques [Ciani, 1984] and in one year (1999), the Infectious Diseases Hospital, Municipal Corporation, Delhi, diagnosed at least one case of human rabies due to monkey bite [Chhabra et al., 2004]; rabies infection from monkey bites was also reported in Delhi in previous years [Singh et al., 2001]. Researchers hypothesize that endemic rabies in India is at least as old as human habitation [Sudarshan et al., 2007], suggesting that there is also a long history of rabies in primate populations as well. Although rabies is also endemic in China, it is far less common than in India [Hu et al., 2009] and there are no corresponding observations of non-human primate infection, providing a potential but inconclusive source of differential selection on *MAST2* in the two populations. In this scenario, elite control of SIV infection would not be the result of selection on Chinese rhesus, but the wild type rhesus macaque condition. Positive selection on neighboring genes in Indian-derived animals (such as *MAST2*) would then also result in the deleterious susceptibility to simian-AIDS, a condition that did not become a selective force in its own right until introduction of animals into U.S. breeding facilities and infection with SIVmac. However, re-sequencing, gene expression and functional studies of both *MAST2* and *HIVEP3* are necessary to rigorously test this hypothesis.

The genomic regions identified in this study also provide a focus for more detailed studies of causes of other major differences between Indian and Chinese rhesus macaques. Moreover, the same methodology is applicable to the study of regional differences in other species, such as susceptibility to *Plasmodium knowlesi* (Collins et al., 1992, Migot-Nabias et al., 1999; Schmidt et al., 1977) and myocardial degeneration (Vidal et al., 2010) in Mauritian,

but not Philippine and Indonesian, cynomolgus macaques (*M. fascicularis*), from which Mauritian cynomolgus macaques derive, respectively.

Conclusion

The increased reliance of biomedicine on a nonhuman primate model and a trend towards translational science necessitates more complex genetic tools for study of the rhesus macaque. Development of a single-nucleotide polymorphism (SNP) map for the rhesus macaque will allow researchers to understand the background genetic variation of research subjects and directly investigate variation in genes of interest. In this study, we interrogated differences in linkage disequilibrium for an initial foray into genome-wide candidate gene identification for phenotypes of biomedical interest. Although intensive functional validation of these candidate genes is necessary, as well as additional SNP validation, these technologies will be useful in addressing a wide variety of questions regarding genotype-phenotype relationships in rhesus macaques.

Acknowledgments

The authors acknowledge funding from the following sources: NIH/NCRR R24RR025871 and NIH/NCRR R24RR005090

The authors would like to acknowledge the staff of the University of California (UC), Davis Molecular Anthropology Laboratory for assistance with DNA extraction and preparation. We would also like to acknowledge the staff and faculty of the UC Davis Genome Center, especially Dr. Dawei Lin and the Bioinformatics core, for insight on data analysis, and Dr. Charles Nicolet and the Genome Technologies Core, for valuable discussion of SNP genotyping.

References

- ALTSCHUL SF, GISH W, MILLER W, MYERS EW, LIPMAN DJ. Basic local alignment search tool. *Journal of Molecular Biology*. 1990; 215:403–410. [PubMed: 2231712]
- CAMERON J. An initial genetic linkage map of the rhesus macaque (*Macaca mulatta*) genome using human microsatellite loci. *Genomics*. 2006; 87:30–38. [PubMed: 16321502]
- CHHABRA M, ICHHPUJANI RL, TEWARI KN, LAL S. Human rabies in Delhi. *Indian Journal of Pediatrics*. 2004; 71:217–220. [PubMed: 15080408]
- CHAMPOUX M, KRIETE MF, HIGLEY JD, SUOMI SJ. CBC and serum chemistry differences between Indian-derived and Chinese-Indian hybrid rhesus monkey infants. *American Journal of Primatology*. 39:79–84.
- CHAMPOUX M, HIGLEY JD, SUOMI SJ. Behavioral and physiological characteristics of Indian and Chinese-Indian hybrid rhesus macaque infants. *Developmental Psychobiology*. 1997; 31:49–63. [PubMed: 9222116]
- CIANI AC. A case of infanticide in a free-ranging group of rhesus monkeys (*Macaca mulatta*) in the Jackoo Forest, Simla, India. *Primates*. 1984; 25:372–377.
- CLARKE MR, O'NEIL JAS. Morphometric comparison of Chinese-origin and Indian-derived rhesus monkeys (*Macaca mulatta*). *American Journal of Primatology*. 1999; 47:335–346. [PubMed: 10206210]
- COLLINS WE, SKINNER JC, BROADRSON JR, FILIPSKI VK, MORRIS CM, STANFILL PS, WARREN M. Susceptibility of *Macaca fascicularis* monkeys from Mauritius to different species of *Plasmodium*. *Journal of Parasitology*. 1992; 78:505–511.
- DAZA-VAMENTA R, GLUSMAN G, ROWEN L, GUTHRIE B, GERAGHTY DE. Genetic divergence of the rhesus macaque major histocompatibility complex. *Genome Research*. 2004; 14:1501–1515. [PubMed: 15289473]
- DEGENHARDT JD, DE CANDIA P, CHABOT A, SCHWARTZ S, HENDERSON L, LING B, HUNTER M, JIANG Z, PALERMO R, KATZE M, EICHLER EE, VENTURA M, ROGERS J, MARX P, GILAD Y, BUSTAMANTE CD. Copy number variation of CCL3-like genes affects

- rate of progression to simian-AIDS in rhesus macaques (*Macaca mulatta*). PLoS Genetics. 2009; 5:1–11.
- FOLKS TM. AIDS animal model comes of age. Journal of Medical Primatology. 2010;10.1111/j.1600–0684.2010.00460.x
- GIBBS R. RHESUS MACAQUE GENOME SEQUENCING AND ANALYSIS CONSORTIUM. Evolutionary and biomedical insights from the rhesus macaque genome. Science. 2007; 316:222–234. [PubMed: 17431167]
- HADFIELD RM, PULLEN JG, DAVIES KF, WOLFENSOHN SE, KEMNITZ JW, WEEKS DE, BENNETT ST, KENNEDY SH. Towards developing a genome-wide microsatellite marker set for linkage analysis in the rhesus macaque (*Macaca mulatta*): identification of 76 polymorphic markers. American Journal of Primatology. 2001; 54:223–231. [PubMed: 11468752]
- HAMADA Y, WATANABE T, CHATANI K, HAYAKAWA S, IWAMOTO M. Morphometrical comparison between Indian- and Chinese-derived rhesus macaques (*Macaca mulatta*). Anthropological Science. 2005; 113:183–188.
- HAYDEN EC. US plans more primate research. Nature. 2008; 453:439. [PubMed: 18497784]
- HEDRICK P, KUMAR S. Mutation and linkage disequilibrium in human mtDNA. European Journal of Human Genetics. 2001; 9:969–972. [PubMed: 11840186]
- HERNANDEZ RD, HUBISZ MJ, WHEELER DA, SMITH DG, FERGUSON B, ROGERS J, NAZARETH J, INDAP A, BOURQUIN T, MCPHERSON J, MUZNY D, GIBBS R. Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. Science. 2006; 316:240–243. [PubMed: 17431170]
- HICAR MD, LIU Y, ALLEN CE, WE LC. Structure of the human zinc finger protein HIVEP3: molecular cloning, expression, exon-intron structure and comparison with paralogous genes HIVEP1 and HIVEP2. Genomics. 2001; 71:89–100. [PubMed: 11161801]
- HU R, TANG Q, TANG J, FOOKS AR. Rabies in China: an update. Vector-borne and zoonotic diseases. 2009; 9:1–12.
- HU ZL, BAO J, REECY JM. CateGORizer: a web-based program to batch analyze gene ontology classification categories. Online Journal of Bioinformatics. 2008; 9:108–112.
- HUGHES AL, WELCH R, PURI V, MATTHEWS C, HAQUE K, CHANOCK SJ, YEAGER M. Genome-wide SNP typing reveals signatures of population history. Genomics. 2008; 92:1–8. [PubMed: 18485661]
- JOHNSON GCL, ESPOSITO J, BARRATT BJ, SMITH AN, HEWARD J, DI GENOVA G, UEDA H, CORDELL HJ, EAVES IA, DUDBRIDGE F, TWELLS RCJ, PAYNE F, HUGHES W, NUTLAND S, STEVENS H, CARR P, TUOMILEHTO-WOLF E, TUOMILEHTO J, GOUGH SCL, CLAYTON DG, TODD JA. Haplotype tagging for the identification of common disease genes. Nature Genetics. 2001; 29:233–237. [PubMed: 11586306]
- KANTHASWAMY S, GILL L, SATKOSKI J, GOYAL V, MALLADI V, KOU A, BASUTA K, SARKISYAN L, GEORGE D, SMITH DG. Development of a Chinese-Indian hybrid (Chindian) rhesus macaque colony at the California National Primate Research Center by introgression. Journal of Medical Primatology. 2009; 38:86–96. [PubMed: 18715266]
- LEE AS, GUTIÉRREZ-ARCELUS M, PERRY GH, VALLENDER EJ, JOHNSON WE, MILLER GM, KORBEL JO, LEE C. Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies. Human Molecular Genetics. 2008; 17:1127–1136. [PubMed: 18180252]
- LOFFERDO JT, MAXWELL J, QI Y, GLIDDEN CE, BORCHARDT GJ, SOMA T, BEAN AT, BEAL DR, WILSON NA, REHRAUER WM, LIFSON JD, CARRINGTON M, WATKINS DI. Mamu-B*08-positive macaques control simian immunodeficiency virus replication. Journal of Virology. 2007; 81:8827–8832. [PubMed: 17537848]
- LYNCH, M.; WALSH, B. Genetics and Analysis of Quantitative Traits. Sinauer Associates; Sunderland, Massachusetts: 1998. Sources of genetic variation for multilocus traits; p. 81–103.
- MALHI RS, SICKLER B, LIN D, SATKOSKI J, TITO RY, GEORGE D, KANTHASWAMY S, SMITH DG. MamuSNP: a resource for rhesus macaque (*Macaca mulatta*) genomics. PLoS ONE. 2:e438. [PubMed: 17487284]

- MANSFIELD KG, LERCHE NW, GARDNER MB, LACKER AA. Origins of simian immunodeficiency virus infection in macaques at the New England Regional Primate Research Center. *Journal of Medical Primatology*. 1995; 24:116–122. [PubMed: 8751050]
- MEYLAN E, TSCHOPP J, KARIN M. Intracellular pattern recognition receptors in the host response. *Nature*. 2006; 422:39–44. [PubMed: 16823444]
- MIGOT-NABIAS F, OLLOMO B, DUBREUIL G, MORELLI A, DOMARL O, NABIAS R, GEORGES AJ, MILLET P. *Plasmodium coatneyi*: differential clinical and immune responses of two populations of *Macaca fascicularis* from different origins. *Experimental Parasitology*. 1999; 91:30–39. [PubMed: 9920040]
- MOTHÉ BR, SIDNEY J, DZURIS L, LIEBL ME, FUENGER S, WATKINS DI, SETTE A. Characterization of the peptide-binding specificity of Mamu-B*17 and identification of Mamu-B*17-restricted epitopes derived from simian immunodeficiency virus proteins. *Journal of Immunology*. 2002; 169:210–219.
- MOTHÉ BR, WEINFURTER J, WANG C, REHRAUER W, WILSON N, ALLEN TM, ALLISON DB, WATKINS DI. Expression of the major histocompatibility complex class I molecule *Mamu-A*01* is associated with control of Simian Immunodeficiency Virus SIV_{mac}239 Replication. *Journal of Virology*. 2003; 77:2736–2740. [PubMed: 12552014]
- MURPHY WJ, AGARWALA R, SCHÄFFER AA, STEPHENS R, SMITH JRC, CRUMPLER NJ, DAVID VA, O'BRIEN SJ. A rhesus macaque radiation hybrid map and comparative analysis with the human genome. *Genomics*. 2005; 86:383–395. [PubMed: 16039092]
- ONG RTH, TEO YY. varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. *Bioinformatics*. 2010; 26:1269–1270. [PubMed: 20308177]
- PREHAUD C, WOLFF TERRIENE, LAFAGE M, MEGRET F, BABAUT N, CORDIER F, TAN GS, MAITREPIERRE E, MENAGER P, CHOPY D, HOOS S, ENGLAND P, DELEPIERRE M, SCHNELL MJ, BUC H, LAFON M. Attenuation of rabies virulence: takeover by the cytoplasmic domain of its envelope protein. *Science Signaling*. 2010; 3:ra5. [PubMed: 20086240]
- ROGERS J, GARCIA R, SHELEDY W, KAPLAN J, ARYA A, JOHNSON Z, BERGSTROM M, NOVAKOWSKI L, NAIR P, VINSON A, NEWMAN D, HECKMAN G, SATKOSKI JA, MALHI RS, KANTHASWAMY S, TITO RY, MALLADI VS, SMITH DG. Pyrosequencing as a method for SNP identification in the rhesus macaque (*Macaca mulatta*). *BMC Genomics*. 2008; 9:256. [PubMed: 18510772]
- SATKOSKI TRASKJA, MALHI RS, KANTHASWAMY S, JOHNSON J, GARNICA WT, MALLADI VS, SMITH DG. The effect of SNP discovery method and sample size on estimation of population genetic data for Chinese and Indian rhesus macaques (*Macaca mulatta*). *Primates*. In Press.
- SCHMIDT LH, FRADKIN R, HARRISON J, ROSSAN RN. Differences in the virulence of *Plasmodium knowlesi* for *Macaca irus (fascicularis)* of Philippine and Malayan origins. *The American Journal of Tropical Medicine and Hygiene*. 1977; 26:612–622. [PubMed: 407808]
- SEELER JS, MUCHARDT C, SUESSLE A, GAYNOR RB. Transcription factor PRDII-BF1 activates human immunodeficiency virus type 1 gene expression. *Journal of Virology*. 1994; 68:1002–1009. [PubMed: 8289330]
- SINGH J, JAIN DC, BHATIA R, ICHHPUJANI RL, HARIT AK, PANDA RC, TEWARI KN, SOKHEY J. Epidemiological characteristics of rabies in Delhi and surrounding areas, 1998. *Indian pediatrics*. 2001; 38:1354–1360. [PubMed: 11752732]
- SUDARSHAN MK, MADHUSUDANA SN, MAHENDRA BJ, RAO NSN, ASHWATH NARAYANA DH, ABDUL RAMAN S, MESLIN FX, LOBO D, RAVIKUMAR K, GANGBORAIAH. Assessing the burden of human rabies in India: results of a national multi-center epidemiological survey. *International Journal of Infectious Diseases*. 2007; 11:29–35. [PubMed: 16678463]
- SZABO A, DALMAU J, MANLEY G, ROSENFEID M, WONG E, HENSON J, POSNER JB, FURNEAUX HM. HuD, a paraneoplastic encephalomyelitis antigen, contains RNA-binding domains and is homologous to Elav and Sex-lethal. *Cell*. 1981; 67:325–333. [PubMed: 1655278]
- VIDAL JD, DROBATZ LS, HOLLIDAY DF, GEIGER LE, THOMAS HL. Spontaneous findings in the heart of Mauritian-origin cynomolgus macaques (*Macaca fascicularis*). *Toxicological Pathology*. 2010; 38:293–3023.

- VAMATHEVAN JJ, HASAN S, EMES RD, AMRINE-MADSEN H, RAJAGOPALAN D, TOPP SD, KUMAR V, WORD M, SIMMONS MD, FOORD SM, SANSEAU P, YANG Z, HOLBROOK JD. The role of positive selection in determining the molecular cause of species difference in disease. *BMC Evolutionary Biology*. 2008; 8:273. [PubMed: 18837980]
- WALDEN PD, COWAN NJ. A novel 205-kilodalton testis-specific serine/threonine protein kinase associated with microtubules of the spermatid manchette. *Molecular and Cell Biology*. 1993; 13:7625–7635.
- WANG WYS, BARRATT BJ, CLAYTON DG. Genome-wide association studies: theoretical and practical concerns. *Nature Reviews Genetics*. 2005; 6:109–118.
- YANT LJ, FRIEDRICH TC, JOHNSON RC, MAY GE, MANESS NJ, ENZ AM, LIFSON JD, O'CONNOR DH, CARRINGTON M, WATKINS DI. The high-frequency major histocompatibility complex class I allele *Mamu-B*17* is associated with control of simian immunodeficiency virus SIVmac239 replication. *Journal of Virology*. 80:5074–5077. [PubMed: 16641299]

Figure 1. Predicted and actual average distance between markers on each chromosome. When all chromosomes were combined, the average distance was 0.88 Mb.

Figure 2. Chromosomes with significant standardized VarLD scores. The small dashed line represents the 99.99th percentile, the medium dashed line the 99th percentile and the broad dashed line the 95th percentile.

Table 1

Sources of DNA samples used in this study

India	Geographic Origin	mtDNA haplotype ^a	Sample Size	Source	Sex Ratio
	Uttar Pradesh, Central	Ind1	22	CPRC	
	Uttar Pradesh, Western	Ind1	1	LABS	M=37.9% F=58.6% U=3.4%
	Kashmir	Ind1	3	UM	
	North Central India	Ind2	3	ONPRC	
China					
	Guangdong	ChiE	1	VBS	
	Sichuan	ChiW	10	VBS, TSS	
	Suzhou	ChiW (N=2) ChiE(N=7) Unknown (N=1)	10	CNPRC	M=38.1% F=61.9%

CPRC, Caribbean Primate Research Center; LABS, Laboratory Animal Breeders Services of Virginia; UM, University of Miami; ONPRC, Oregon National Primate Research Center; VBS, Valley Biosystems; TSS, Three Springs Scientific; CNPRC, California National Primate Research Center.

^a GenBank accession numbers available from the authors upon request.

Table 2

Genomic position and *p*-value of significant standardized VarLD scores. Gene symbols with a subscripted (1) were directly annotated in the published rhesus genome, all others are predicted. Entries subscripted (2) had no associations in the AmiGo database, and were excluded from further analysis.

Chromosome	Position (Mb)	Symbol	<i>p</i> value
1	40.38	<i>Eaf6</i>	0.05
1	41.81		0.01
1	43.22		0.05
1	44.58	<i>HIVEP3</i>	0.01
1	45.99		0.01
1	47.43	<i>RNF220</i>	0.01
1	48.83	<i>MAST2</i> ¹	0.05
1	50.25		0.01
1	51.62	<i>AGBL4/BEND5</i>	0.05
1	52.93	<i>ELAVL4</i>	0.01
1	54.26	<i>EPS15</i>	0.01
1	55.59		0.0001
1	56.90		0.01
1	58.19		0.05
1	59.50		0.05
1	60.82		0.05
1	62.13		0.05
1	63.42		0.05
1	64.70	<i>INADL</i>	0.05
1	66.02		0.05
1	67.39	<i>CACHD1</i>	0.05
1	68.74	<i>PDE4B</i>	0.05
1	75.48		0.05
1	76.90		0.05
1	79.71		0.05
1	81.05		0.05
1	82.35		0.05
1	87.52		0.05
4	69.68		0.05
4	128.58		0.05
4	130.05		0.05
5	39.86	<i>GNPDA2</i>	0.05
11	43.39		0.05
11	44.73		0.05
11	46.05	<i>CCDC65</i> ²	0.05

Chromosome	Position (Mb)	Symbol	<i>p</i> value
11	47.46	<i>CG32133-PA</i> ²	0.05
11	66.06		0.05
11	67.58	<i>PTPRB</i> ¹	0.05
11	87.93		0.05
11	89.49		0.05

Table 3

Relative occurrence of GO child terms, by total and genes represented. Although SIVmac nonprogression is the phenotype of interest, these terms potentially include all phenotypic differences between Indian-derived and Chinese rhesus macaques.

GO Child Term	% Genes Represented	% Total Occurrence
cell	92.3	15.5
binding	69.2	10.6
intracellular	76.9	9.7
cytoplasm	69.2	7.1
metabolism	61.5	6.2
catalytic activity	46.2	5.8
hydrolase activity	30.8	3.5
plasma membrane	30.8	3.5
protein binding	46.2	3.5
protein metabolism	30.8	2.2
transport	15.4	2.2
cell organization and biogenesis	15.4	1.8
nucleic acid binding	15.3	1.8
nucleobase, nucleoside, nucleotide and nucleic acid metabolism	23.1	1.8
cell communication	15.4	1.3
cytosol	23.1	1.3
nucleotide binding	15.4	1.3
nucleus	23.1	1.3
peptidase activity	7.7	1.3
protein modification	15.4	1.3
RNA binding	7.7	1.3
signal transduction	15.4	1.3
biosynthesis	15.4	0.9
carbohydrate metabolism	7.7	0.9
endosome	7.7	0.9
ion transport	7.7	0.9
transferase activity	7.7	0.9
organelle organization and biogenesis	7.7	0.8
calcium ion binding	7.7	0.4
catabolism	7.7	0.4
cell differentiation	7.7	0.4
cell proliferation	7.7	0.4
cytoskeletal protein binding	7.7	0.4
cytoskeleton	7.7	0.4
development	7.7	0.4

GO Child Term	% Genes Represented	% Total Occurrence
DNA binding	7.7	0.4
Golgi apparatus	7.7	0.4
kinase activity	7.7	0.4
morphogenesis	7.7	0.4
nucleoplasm	7.7	0.4
phosphoprotein phosphatase activity	7.7	0.4
protein kinase activity	7.7	0.4
protein transport	7.7	0.4
receptor activity	7.7	0.4
reproduction	7.7	0.4
signal transducer activity	7.7	0.4
transcription	7.7	0.4
transcription regulator activity	7.7	0.4