# Supplemental Materials

## ADDITIONAL RESULTS

### *Comparisons of rare variant frequencies with humans*

We determined the distribution of alternative allele frequencies (AAF) in the 133 rhesus macaques and compared this with the same statistic for human samples (Supplemental Figures S15A and B). For low-frequency SNPs (AAF<0.04) the proportion of SNVs is higher in the human population than in Indian rhesus, while for AAF>0.05, the proportion of SNVs is lower in human. This indicates that rare variants make up a larger proportion of total SNVs in the human genome than the Indian rhesus genome. This may be due to the recent explosion of human population size (Keinan and Clark 2012). However, the distribution of AAF in Chinese rhesus is quite different from Indian rhesus (especially in the low- and high-frequency range of SNPs, see Supplemental Figure S15B), suggesting that Indian and Chinese rhesus populations likely experienced quite different recent demographic histories.

### *Residual variant intolerance score tests*

We also investigated gene-specific evidence of natural selection in the rhesus lineage, which can reveal aspects of evolutionary process or history, as well as identify candidate loci for disease-related research (Rhesus Macaque Genome Sequencing and Analysis Consortium *et al* 2007; Vitti et al. 2013). We calculated residual variation intolerance scores (RVIS), a statistic originally developed for prioritizing variants and genes according to their likelihoods of producing disease in humans (Petrovski et al. 2013). The RVIS scores for specific genes can be affected by either positive or negative selection. These analyses identified 22 genes with the smallest RVIS (P<0.001) as candidate genes for negative selection and 35 genes with large RVIS (P<0.001) as candidate genes for positive or balancing selection (see Supplemental Figure S16).

We also performed an additional check for a possible effect of mapping errors on the observed signals of positive selection. Cryptic (unrecognized) gene copy number differences among individuals may produce incorrect SNV calls, but will also produce characteristic patterns of inter-animal variation in read coverage for the genes showing copy number variation (CNV). If there is no copy number variation within our sample for a given gene, then average coverage in the region of that gene (x) will be proportional to average coverage across the whole genome (y). We can screen for CNV effects on RVIS using a regression of y onto x, and compare the

regression results for genes under putative positive selection with those showing potential evidence of purifying selection to determine whether any effect of CNV and mapping errors is more apparent for genes identified as under positive selection. We used the average coverage across the whole genome in each Indian rhesus sample, and then calculated the average coverage in each genic region for each Indian rhesus sample. Next, we calculated the regression of read depth across the genome (y) against read depth for genes for which RVIS analysis indicated positive selection (x). The results of the regressions are shown in Supplemental Table S9. We found that the adjusted $R^2$ values are very high for positive genes, and adjusted $R^2$, minimum and maximum residuals are not significantly different in genes flagged as candidates for positive selection versus those scored as experiencing purifying selection ($p>0.05$ in *t*-test and Mann–Whitney *U* test). Therefore, we do not find evidence for any significant effect of CNV-induced mapping errors on the RVIS statistics for genes under positive selection compared to those genes under purifying selection. A further note: because positive selection may decrease the heterozygosity in affected genic regions, heterozygosity is not suitable for measuring the effects of mapping errors caused by gene CNV.

Our findings regarding RVIS are consistent with prior observations that immune response genes frequently undergo adaptive change in mammalian species. This RVIS approach in nonhuman primates may be useful in identifying genetic models of human disease because variants in a gene under strong negative selection in both humans and a nonhuman primate species will be more likely to produce deleterious phenotypic effects in parallel in the two species.

## ADDITIONAL METHODS

### *RVIS analyses*

The coding sequence co-ordinates of genes on rhesus autosomes were downloaded from ENSEMBL database (ftp://ftp.ensembl.org/pub/release-80/gtf/macaca_mulatta/Macaca_mulatta.MMUL_1.80.gtf.gz). A total of 18,466 genes with variants in transcripts were obtained from our SNV dataset. The annotation of these variants in the rhesus genome was downloaded from the ENSEMBL database VEP annotation (ftp://ftp.ensembl.org/pub/release-80/variation/VEP/). Using these annotations, we classified the variants in coding regions into non-functional (synonymous) and functional (including missense, nonsense, and splicing) variants. These variants were mapped onto the baboon (*Papio anubis*)

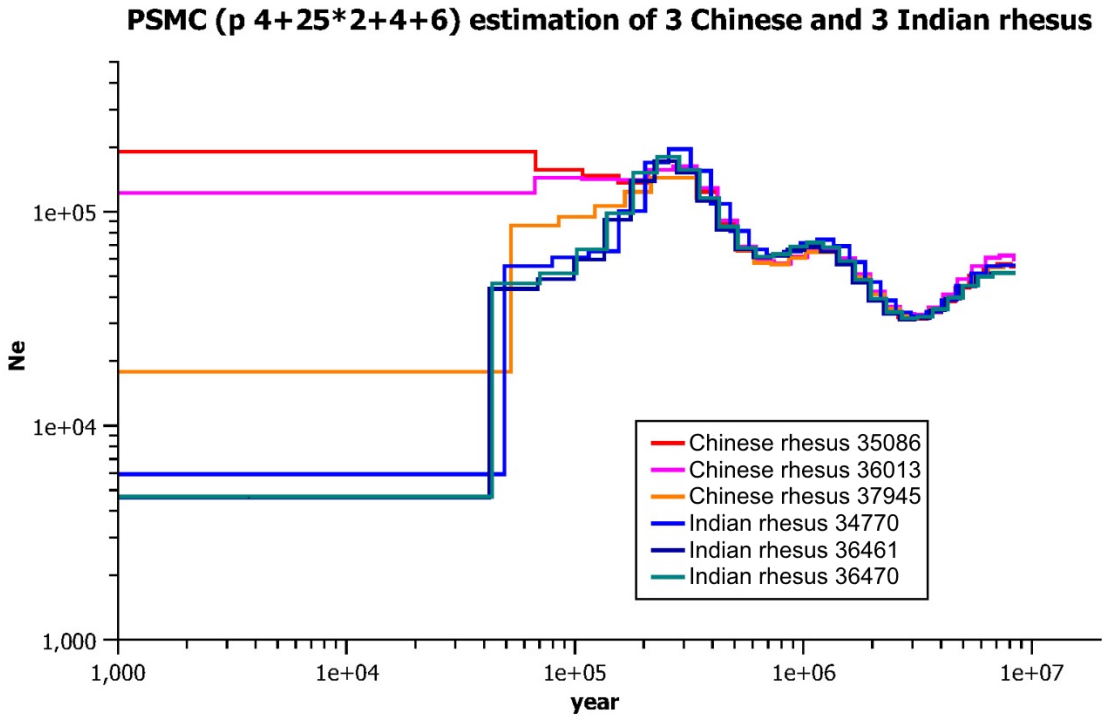
reference genome sequence using liftOver tools, and we treated the reference allele in the baboon reference sequence as the ancestral allele for macaques. (If the ancestral allele was missing, the major allele was treated as the ancestral allele.) We next defined a threshold dividing "common" from "rare" variants with the derived allele frequency (DAF) > 0.01. In this analysis we determined the number of non-functional, common variant sites within the coding region of a gene. We then regressed the number of common functional variants on the number of non-functional coding-region variants in 17,787 IRh genes (CRh sample size is too small) and calculated the standardized residual as the RVIS.

In order to identify the genes not evolving neutrally in IRh population, we simulated 130M genes according to the IRh demographic history described in this paper to get the null distribution of RVIS under neutrality. The genotypes of $1.3 \times 10^6$ ($10^6$ 10kb-, $2 \times 10^5$ 50kb- and $10^5$ 100kb- segment) genes in 123 samples were simulated by *scrm* software, as described under our demographic analyses. We counted the numbers of total SNVs and common (DAF>0.01) SNPs in each gene, calculated the RVIS for each gene and obtained the distribution of RVIS across 123 samples simulated as the null distribution of RVIS under neutral evolution of rhesus genes (see Supplemental Figure S17). As controls, we checked the effects of segmental length and the number of genes considered on the distribution of RVIS, and found that those factors do not affect the RVIS distribution significantly (P>0.05, Mann–Whitney *U* test). Finally we determined the P-value for each IRh gene under this null distribution of RVIS for theoretically neutral genes. Genes with RVIS significantly deviating from the mean of RVIS in the null distribution are candidates for genes affected by selection. Large positive RVIS values located in the positive tail of the null distribution indicate that the gene is highly tolerant to the accumulation of functional variants, which can be due to either positive or balancing selection. Large negative RVIS values indicate an intolerance to functional variants, which is likely due to strong negative selection.

*Ion Torrent re-sequencing*

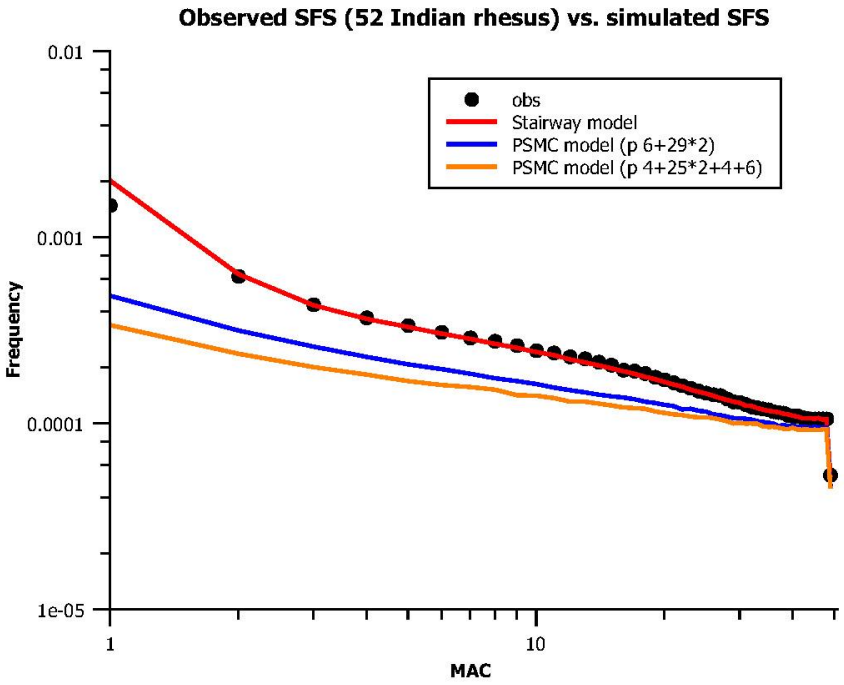
To further assess the reliability of the final SNV call set, we performed another independent validation by sequencing. We used the Ion Torrent PGM platform (Thermo Fisher Scientific Inc.) to sequence custom designed amplicons covering rhesus macaque SNVs with a range of minor allele frequencies. The total number of SNVs re-tested this way was 611. We found that, based on the Ion Torrent data, the inferred false discovery rate (FDR) for doubletons (SNVs with minor allele observed twice in our data) was 5.8%, for tripletons (observed three times) was 5.6% and for common SNV genotypes (MAF 2-3% in our dataset) was 8.6%. These FDR rates

of course include both the false negative errors from the Ion Torrent data and the false positive errors in the Illumina data.

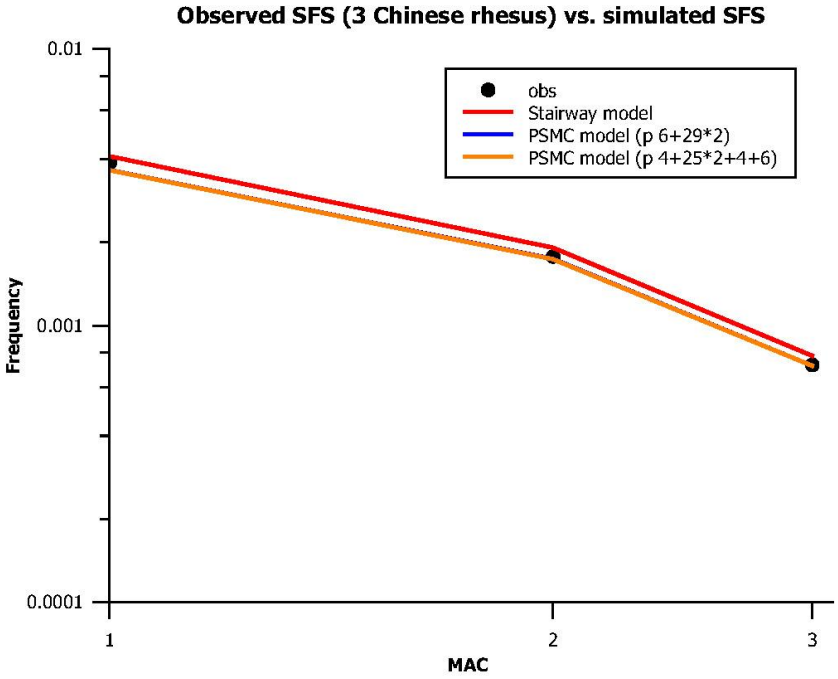



**Supplemental Figure S1**: Demographic histories inferred by PSMC with pattern "4+25*2+4+6". PSMC estimations with pattern "4+25*2+4+6" for 3 high-coverage Chinese rhesus (35086, 36013 and 37945) and 3 high-coverage Indian rhesus (34770, 36461 and 36470). Ne: effective population size.

**A**




**B**




**Supplemental Figure S2**: Comparison of observed SFS with simulated SFS. (**A**) Observed SFS was from 52 low-coverage Indian rhesus. Simulated SFS was obtained by simulating 1 Gb DNA sequence using scrm program assuming the Stairway plot model, the PSMC model with "6+29*2" pattern and the PSMC model with "4+25*2+4+6"

pattern. The Stairway plot model was the inferred demographic history of 75 high-coverage Indian rhesus. The two PSMC model were obtained by averaging the inferred demographic histories of 3 high-coverage Indian rhesus (34770, 36461 and 36470). (**B**) Observed SFS was from 3 low-coverage Chinese rhesus. Simulated SFS was obtained by simulating 1 Gb DNA sequence using scrm program assuming the Stairway plot model, the PSMC model with “6+29*2” pattern and the PSMC model with “4+25*2+4+6” pattern. The Stairway plot model was the inferred demographic history of 6 high-coverage Indian rhesus. The two PSMC model was obtained by averaging the inferred demographic histories of 3 high-coverage Chinese rhesus (35086, 36013 and 37945). The blue line and the orange line in the plot mostly overlapped.

**Figure S3A**

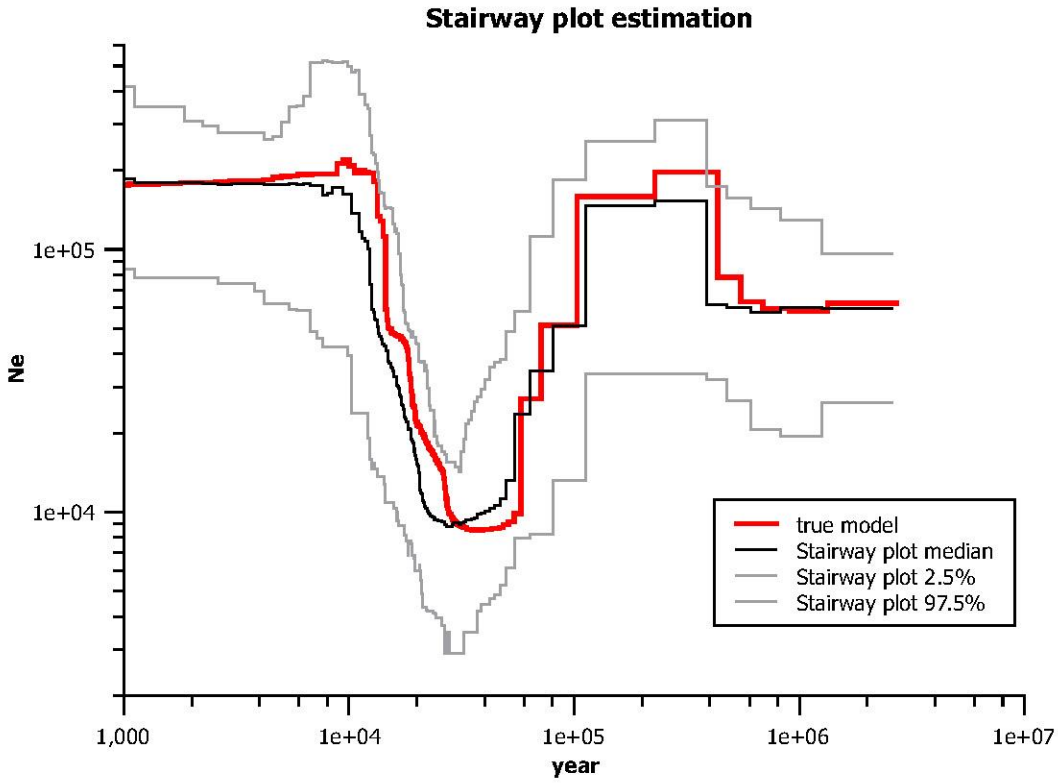

Stairway plot estimation
Ne
1e+05
1e+04
1,000
1e+04
1e+05
1e+06
1e+07
year
true model
Stairway plot median
Stairway plot 2.5%
Stairway plot 97.5%

**Figure S3B**

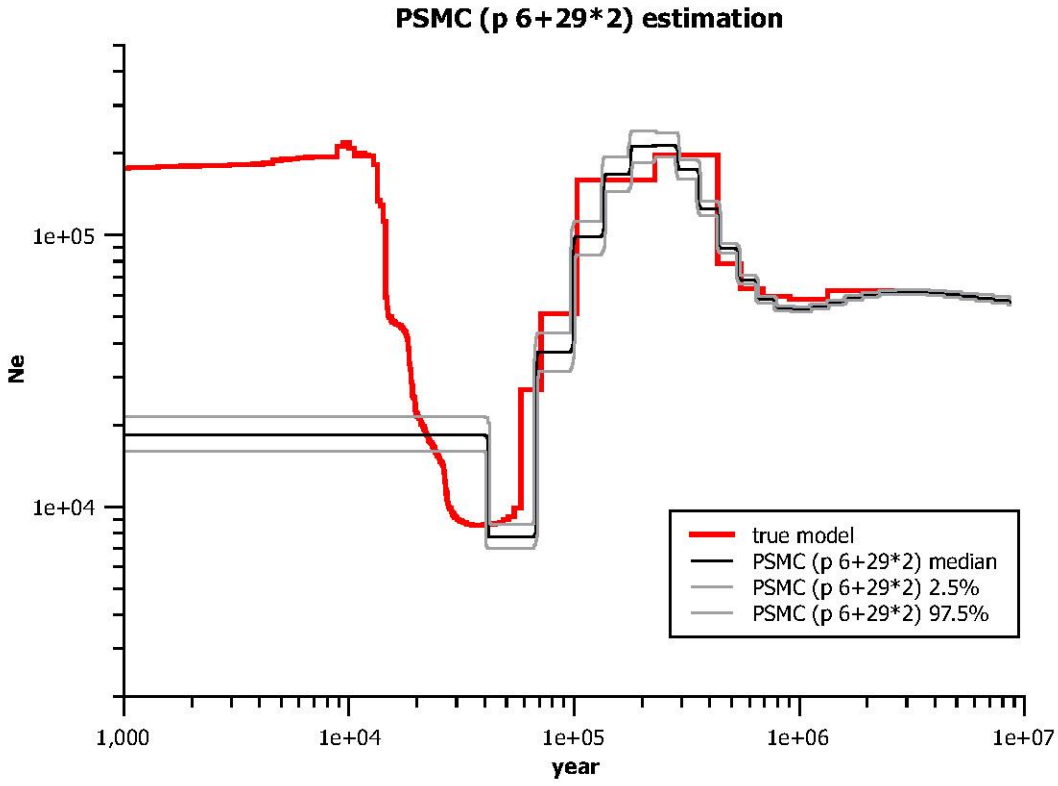
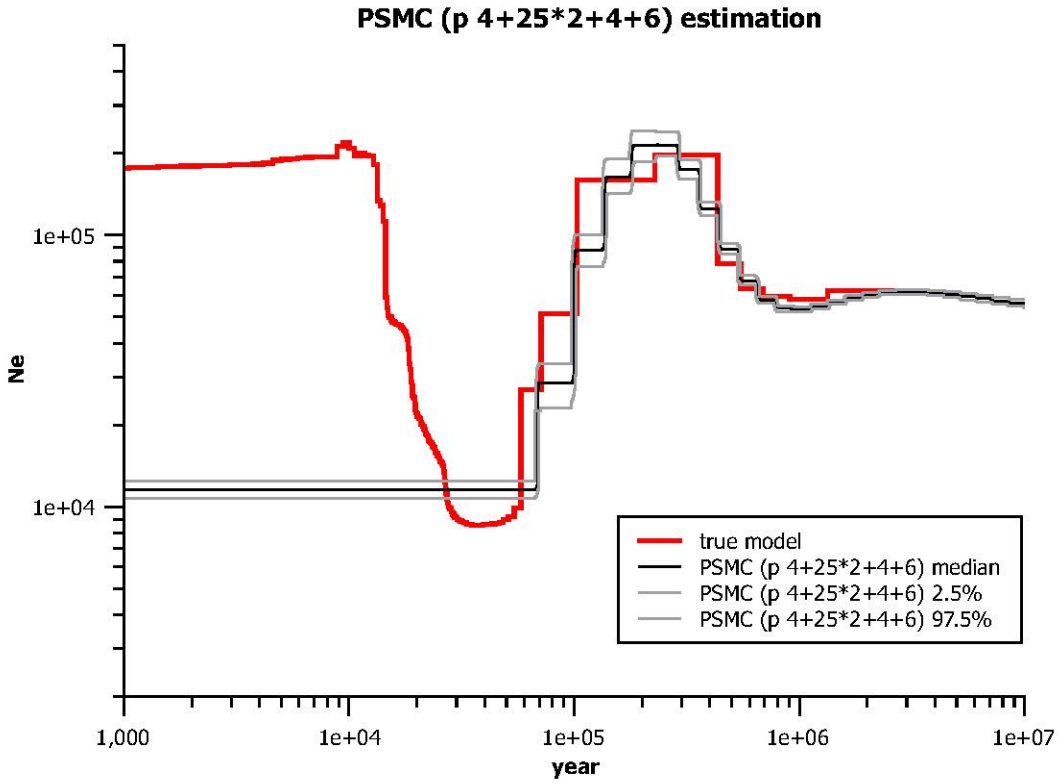
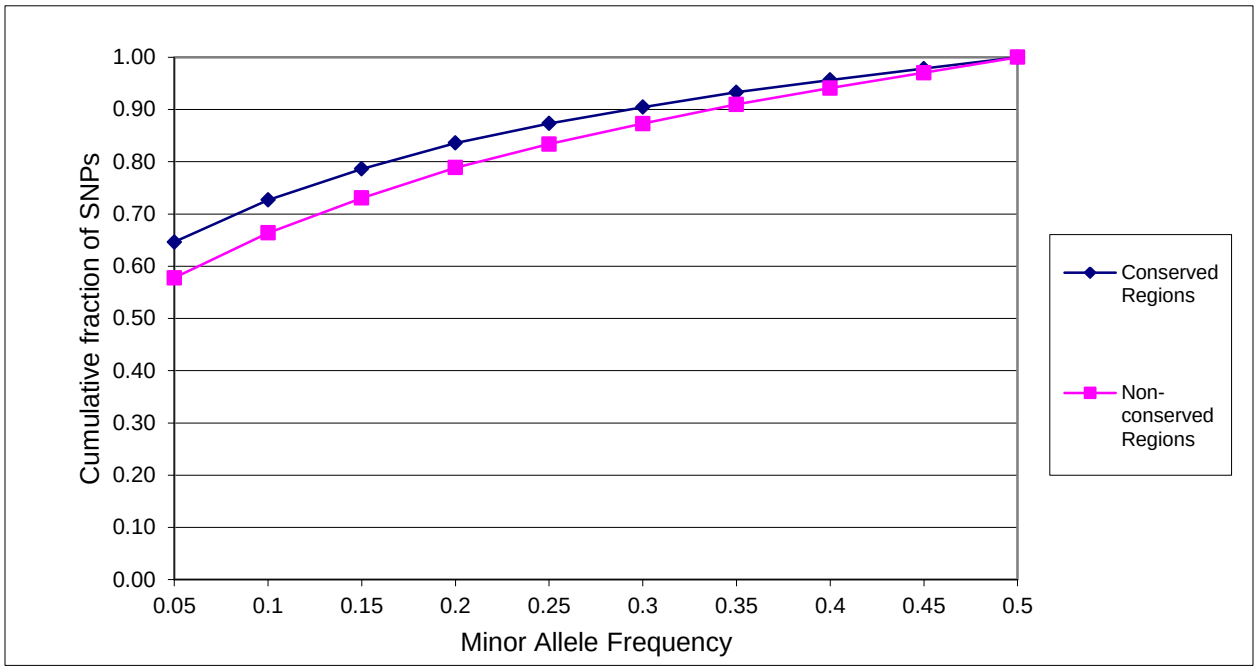

PSMC (p 6+29*2) estimation
Ne
1e+05
1e+04
1,000
1e+04
1e+05
1e+06
1e+07
year
true model
PSMC (p 6+29*2) median
PSMC (p 6+29*2) 2.5%
PSMC (p 6+29*2) 97.5%

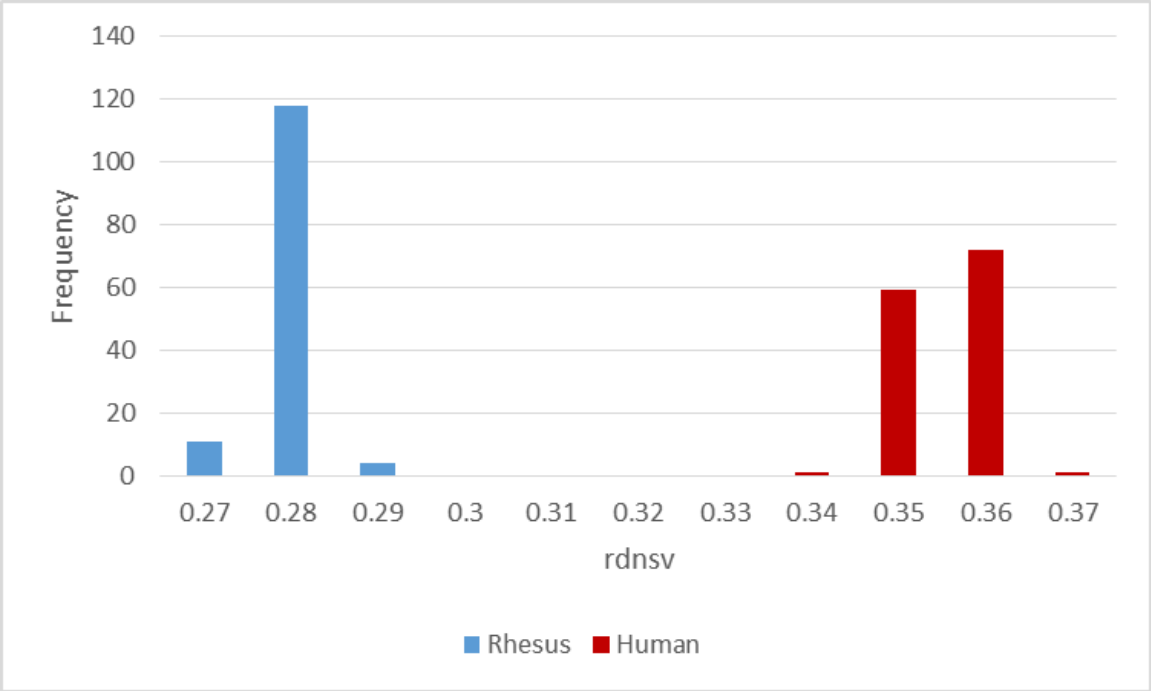
**Figure S3C**



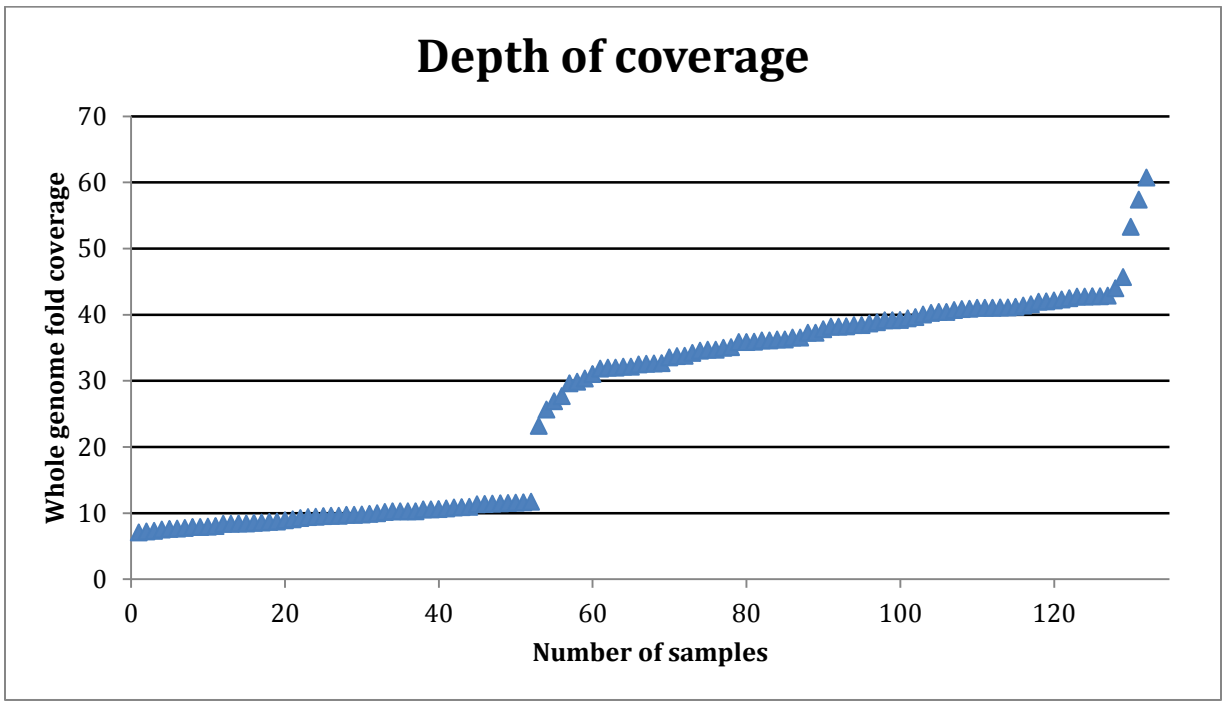

**Supplemental Figure S3**: Stairway plot and PSMC estimations with simulated data assuming a recent population size recovery. A total of 200 simulated DNA sequence samples were simulated using the scrm software. For each simulation, 75 individuals were simulated each with 500 Mb DNA sequences. We assumed the Stairway plot estimation based on 75 high-coverage Indian rhesus is the true demographic model and the ratio of recombination rate and mutation rate is 0.4. Then Stairway plot and PSMC with pattern parameter “6+29*2” and “4+25*2+4+6” were used to infer demographic histories based on the simulated DNA sequences. Only the first individual (of the 75 individuals) were used for PSMC estimations while all 75 individuals were used for the Stairway plot estimation. Median, 2.5% and 97.5% of the 200 estimations from the Stairway plot and PSMC were plotted.




**Supplemental Figure S4.** The distribution of minor allele frequency for rhesus macaque polymorphisms in regions orthologous to the 4.2% of the human genome found to be conserved across 29 mammals (Lindblad-Toh et al. 2011). The distribution of MAF for polymorphisms in the conserved regions is shifted to the left compared with polymorphisms in the remainder of the rhesus macaque genome.

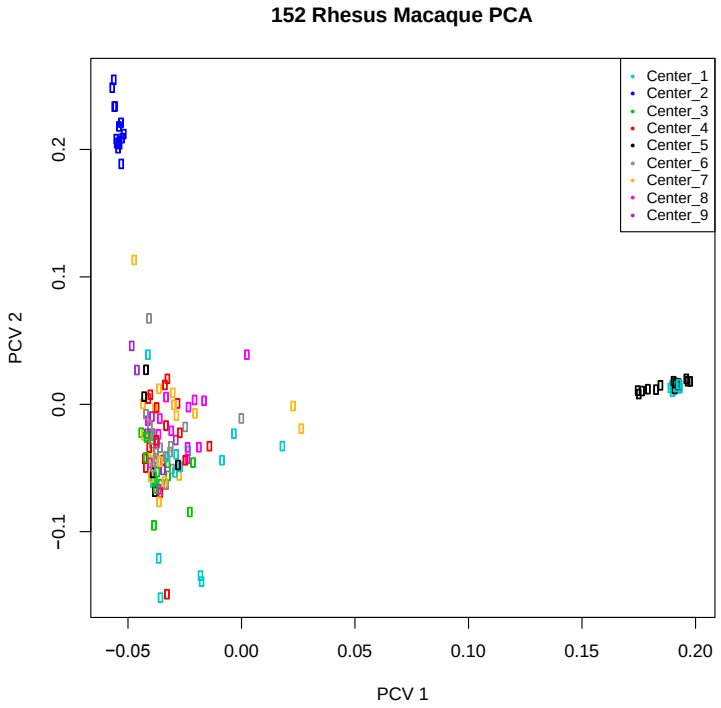



**Supplemental Figure S5:** rdnsv results for rhesus and humans. Blue bars indicate the distribution of rdnsv values for 133 rhesus macaques. The red bars indicate the equivalent distribution of rdnsv for 133 humans from the 1000 Genomes dataset, selected to match whole genome sequence coverage for the rhesus as closely as possible.




**Supplemental Figure S6:** Sequence coverage for each of the 133 rhesus macaque samples, ranked and plotted from lowest coverage to highest. Mean coverage across the full sample set is 26.7x.

## Figure S7A




## Figure S7B

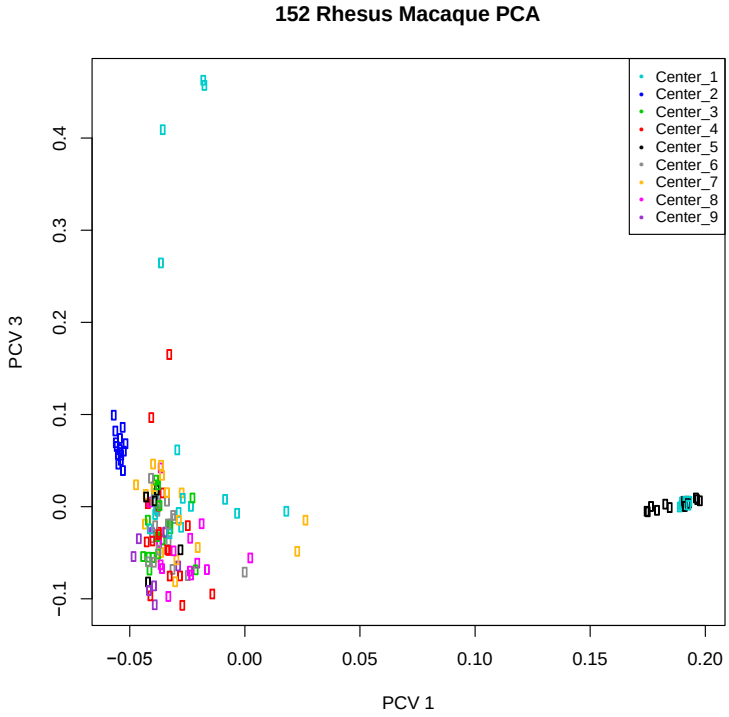
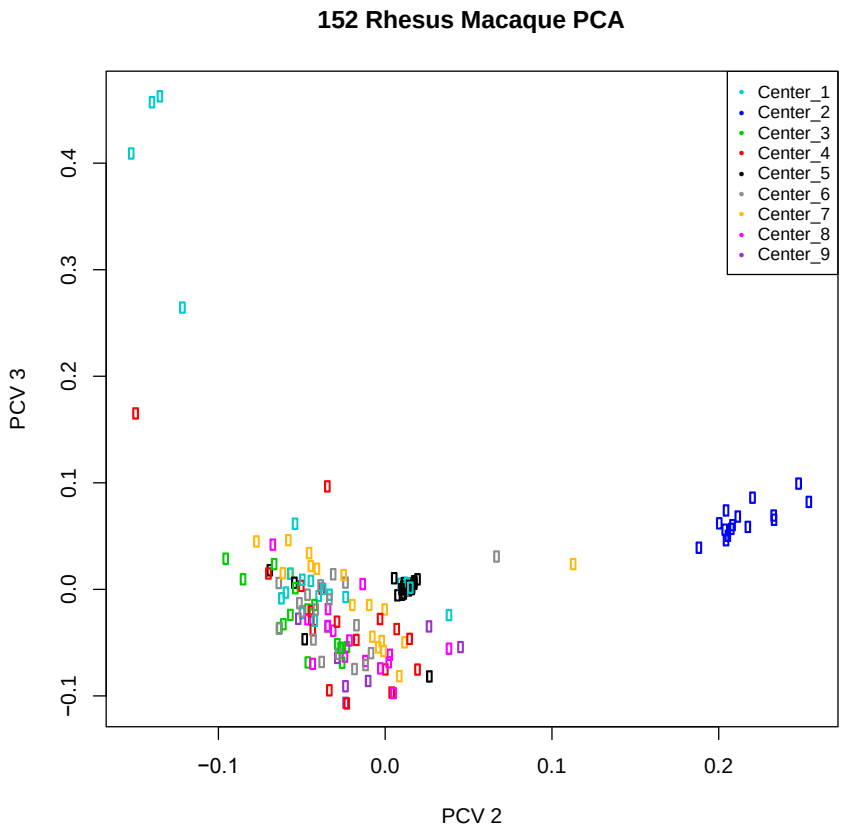
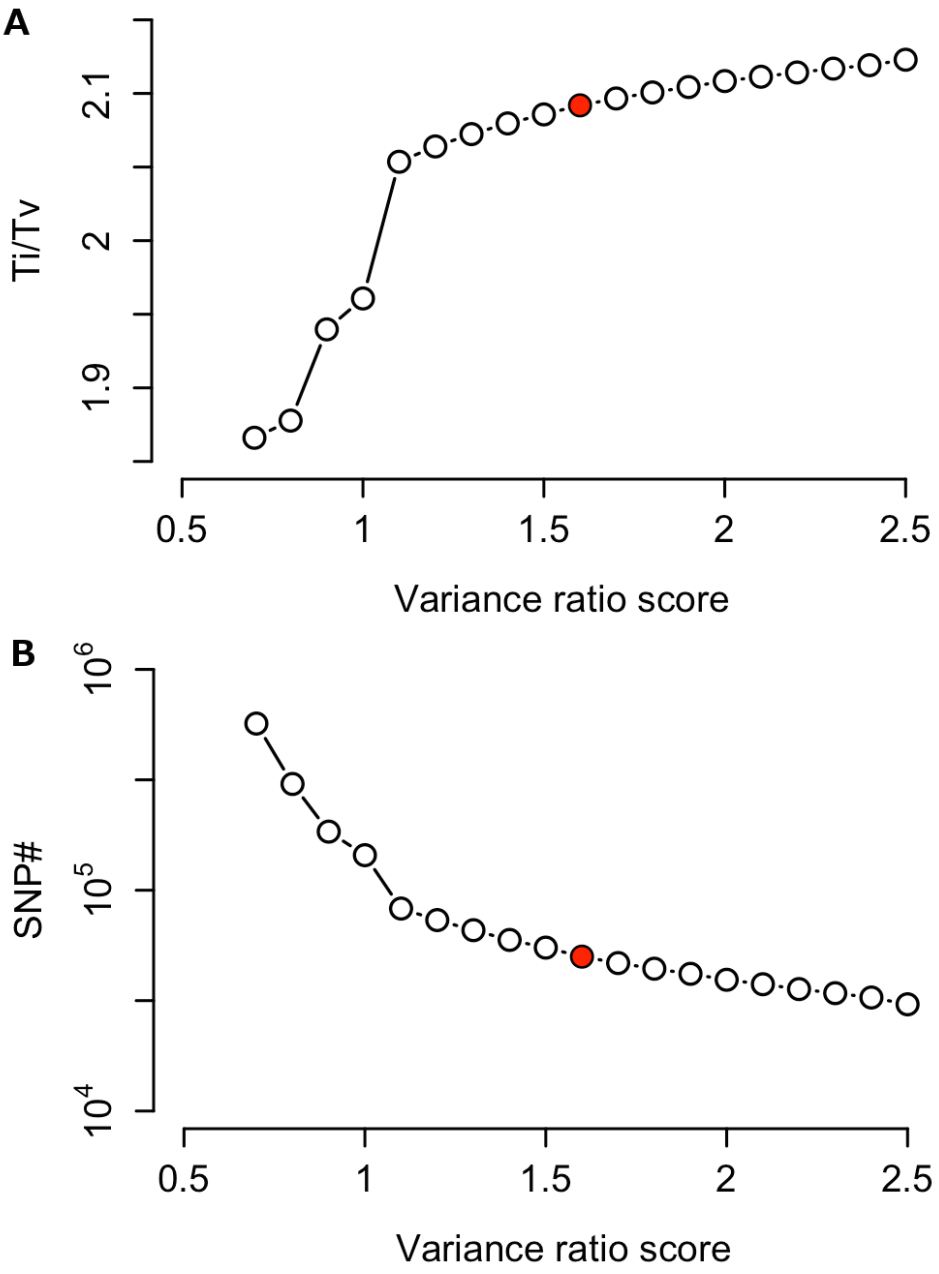


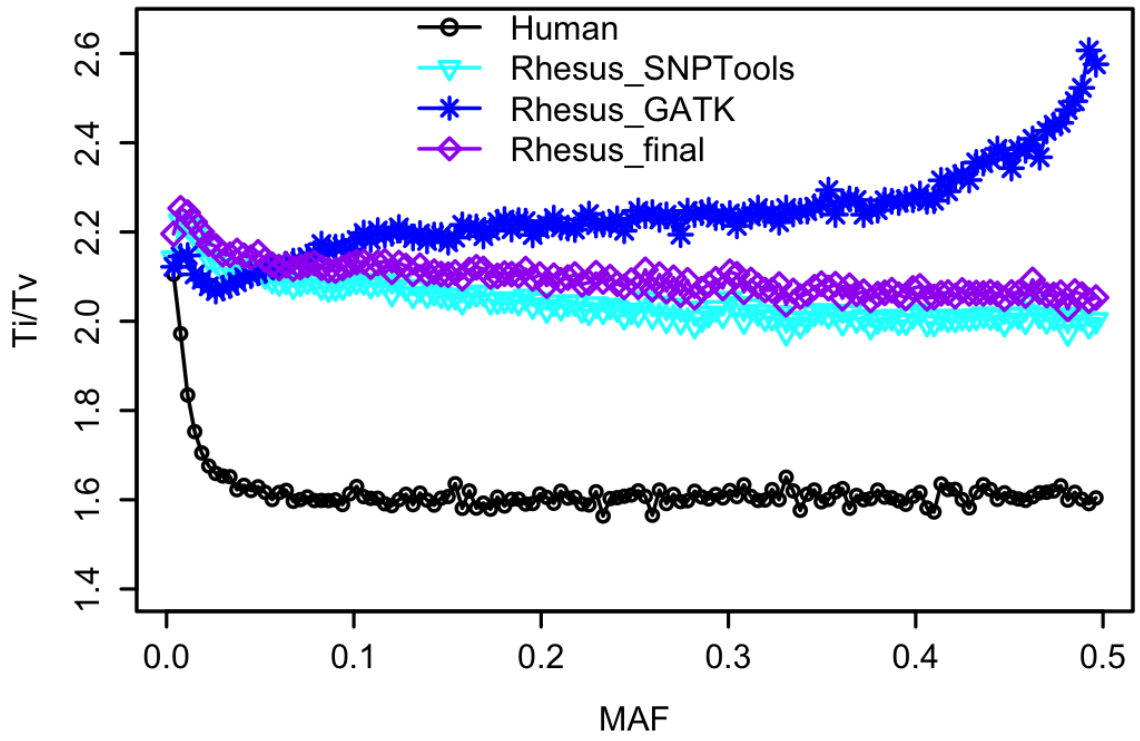
**Figure S7C**



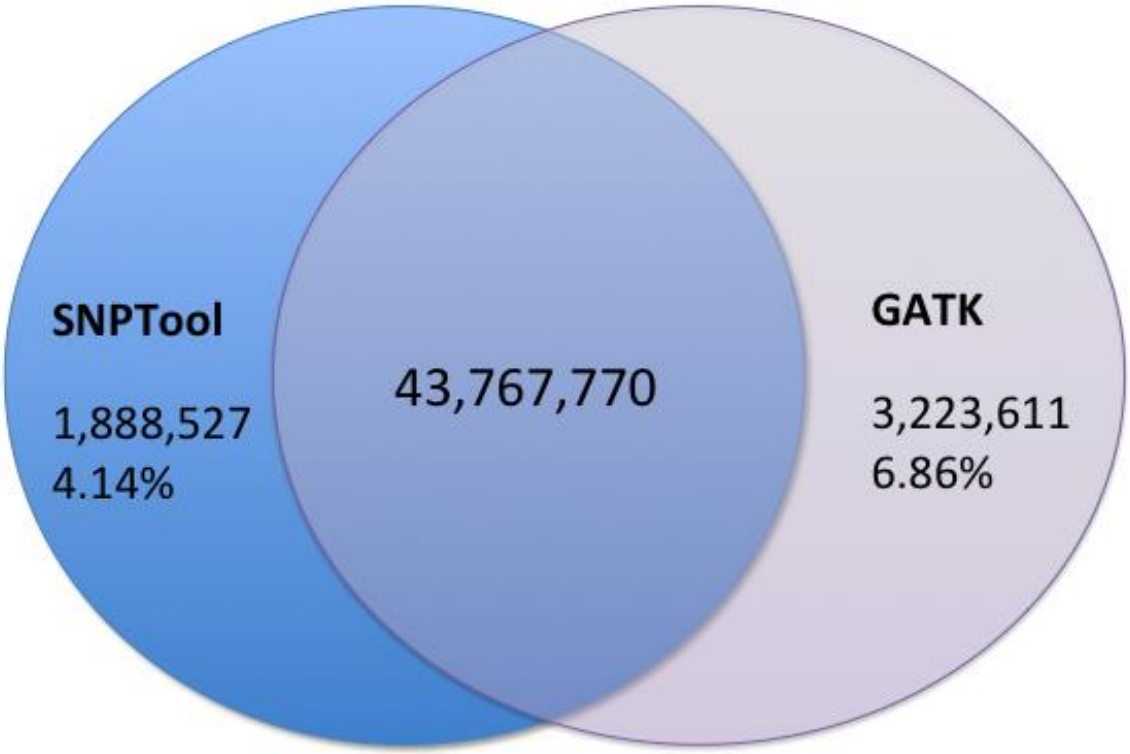

**Supplemental Figure S7**: Principal components plots of SNV genotypes across 152 rhesus macaques. PC1 clearly separates Chinese-origin animals from Indian-origin animals. PC2 separates one colony of Indian-origin rhesus macaques from a larger cluster of other Indian-origin animals. The scatter among the Indian-origin animals along PC1 shows that several likely have some degree of Chinese-origin ancestry.



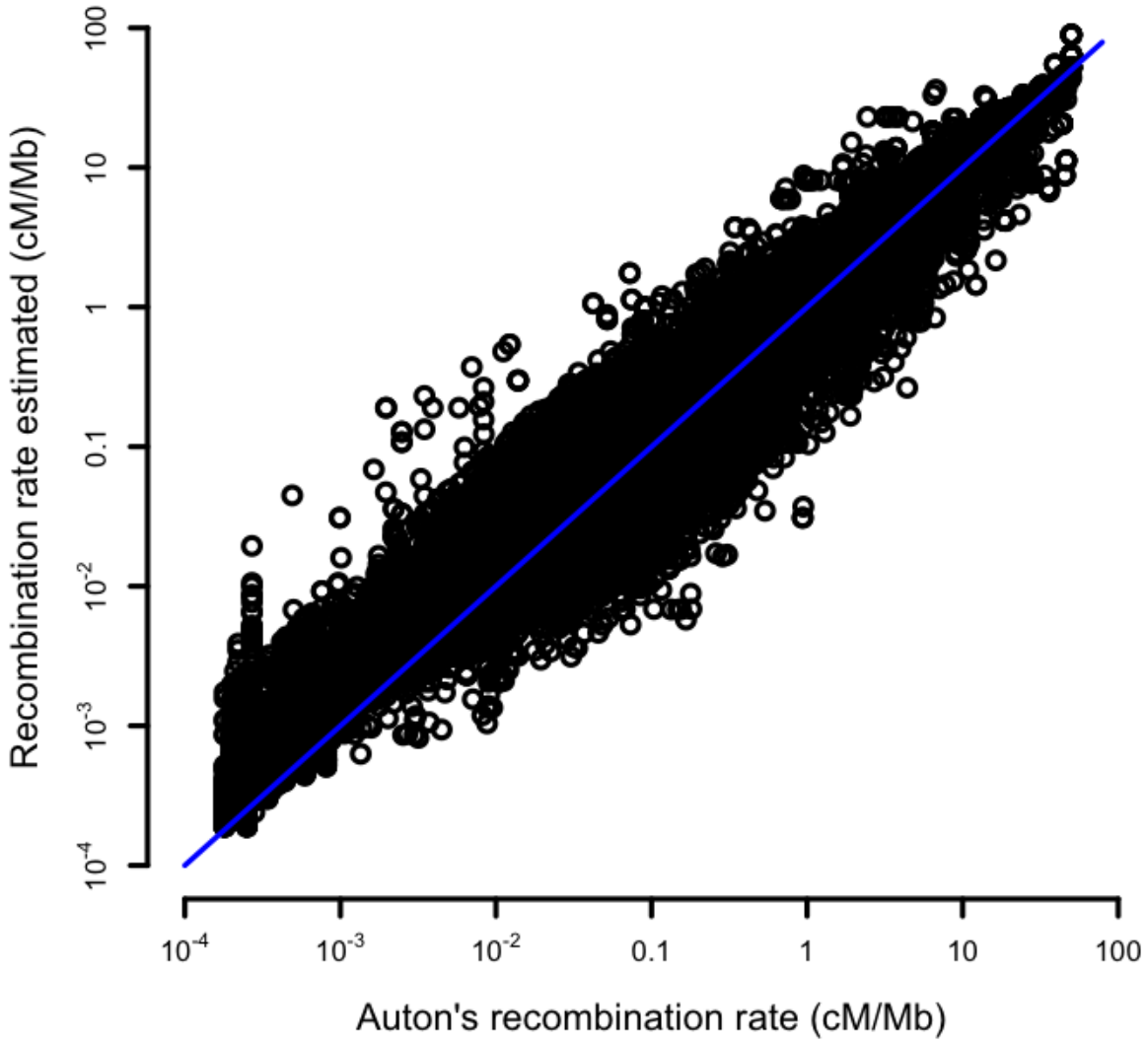

**Supplemental Figure S8.** Ti/TV ratios (panel A) and the numbers of SNPs (panel B) observed in the autosomal SNV datasets across a range of variant ratio scores used in SNPTools SNV discovery analysis. The red point is the final cutoff (variant ratio score = 1.6) used for subsequent data processing (genotype likelihood estimation and imputation, etc.).



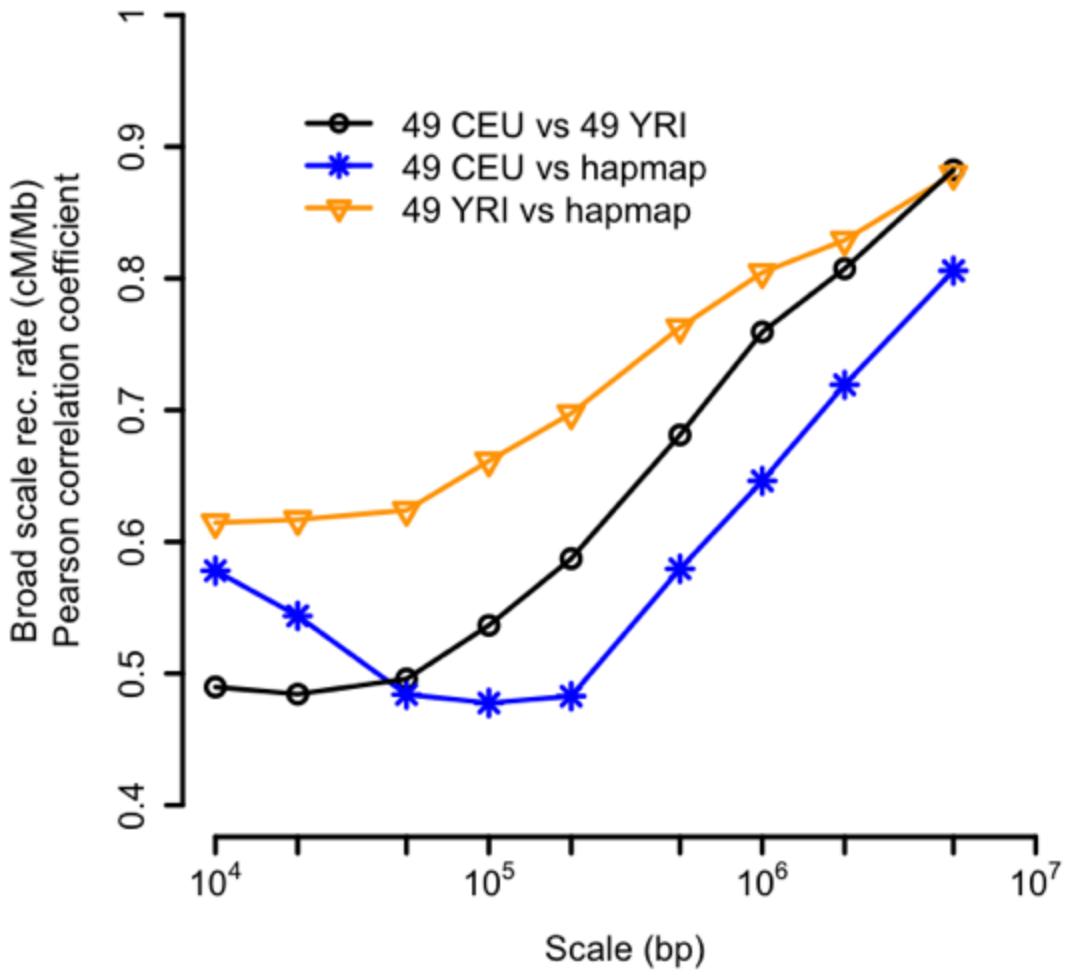

**Supplemental Figure S9**. Changes in Ti/Tv ratios with minor allele frequency (MAF) in rhesus datasets (133 samples) called by SNPTools, GATK toolkits, intersection of SNPTools and GATK called (Rhesus_final), and human dataset (1000 Genome data, phase 1, 133 samples randomly selected from 1092 samples).



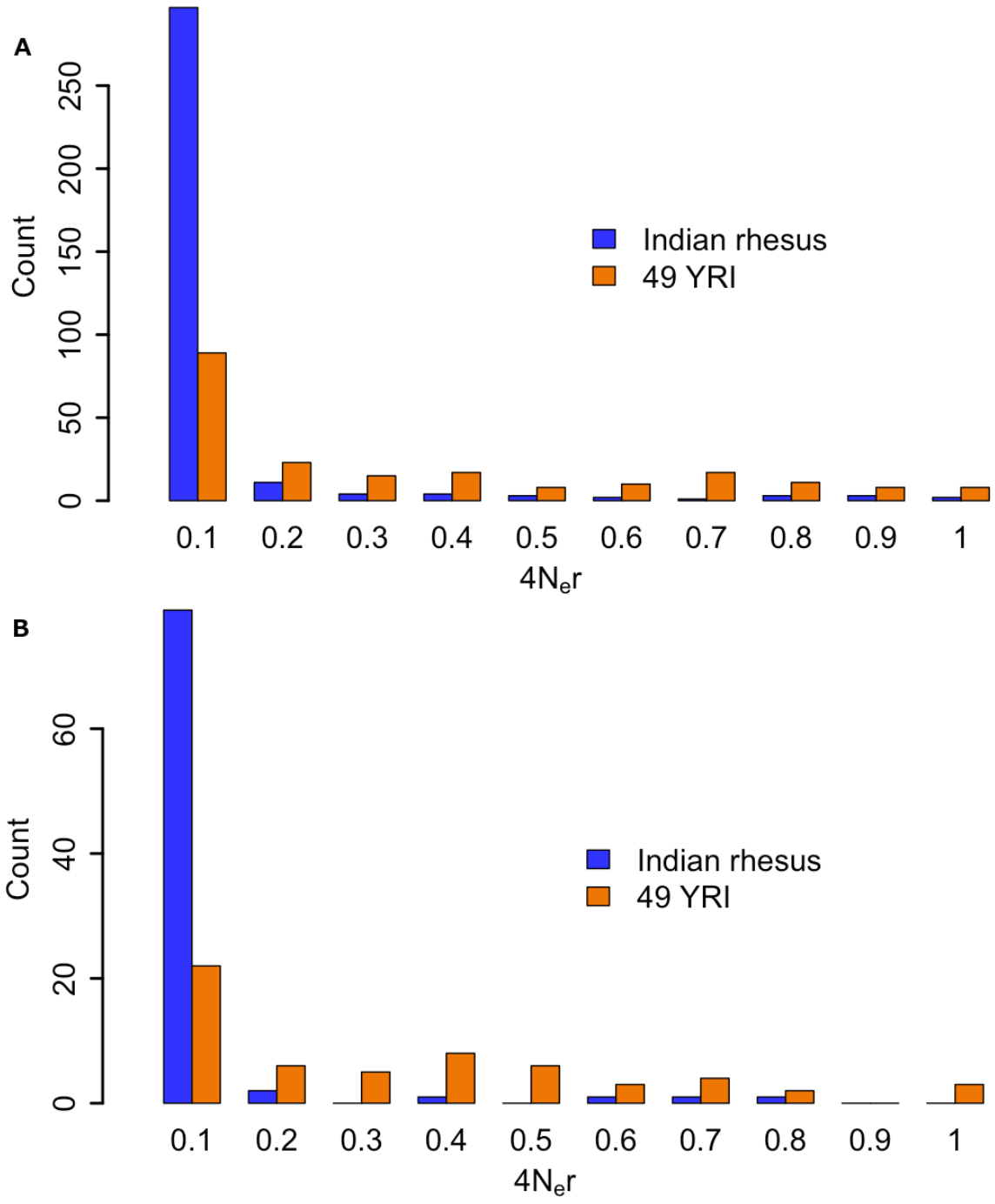

**Supplemental Figure S10.** Venn diagram of intersection of SNV sites from autosomes of *Macaca mulatta* (n = 133 samples) discovered using SNPTools and GATK.



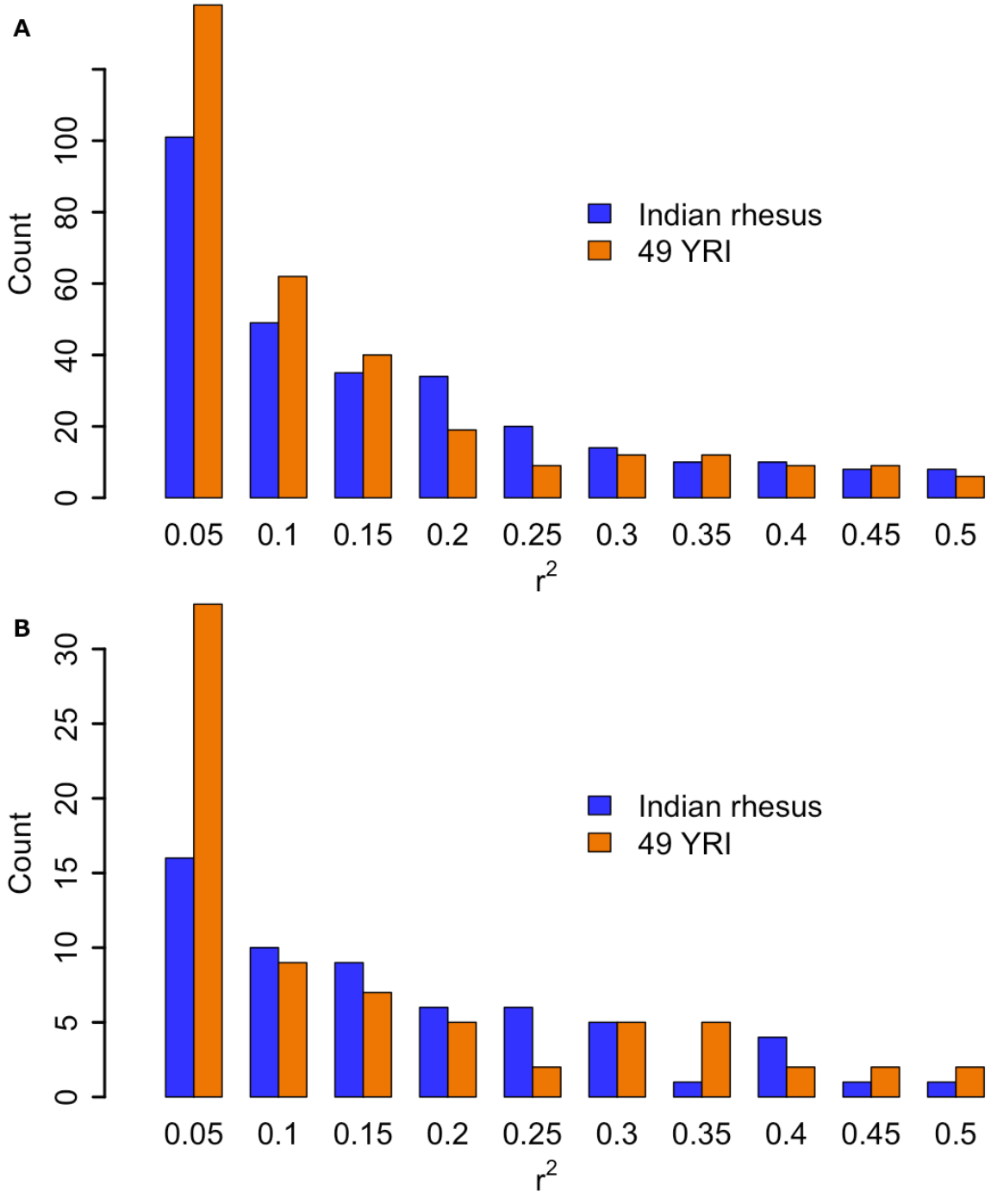

**Supplemental Figure S11.** Comparison of recombination rates on chimpanzee chromosome 19 estimated by software LDHat and the original results reported by Auton et al (2012). The blue line is the expectation of recombination rate being equal in both datasets.



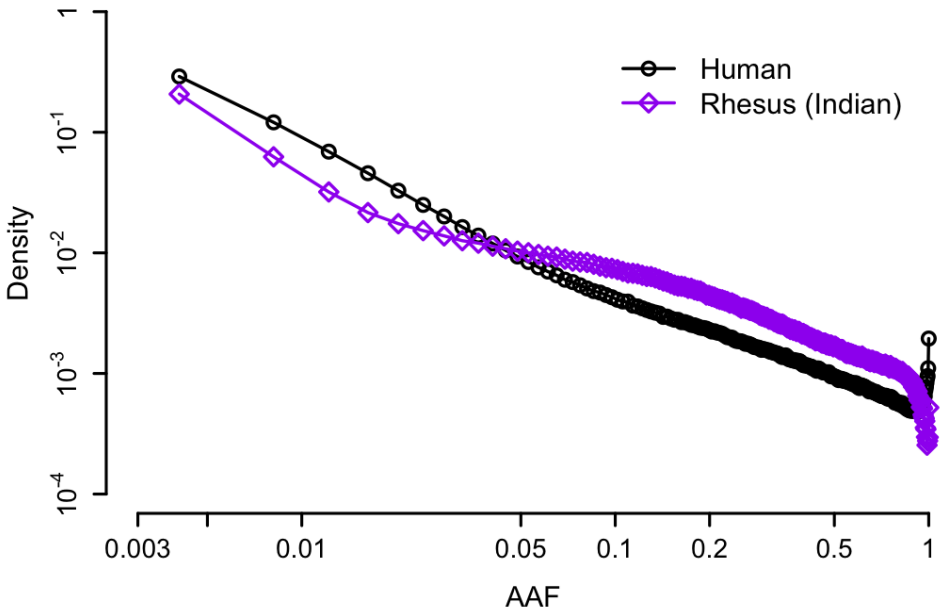

**Supplemental Figure S12**. Pearson correlation coefficients at different scales between recombination rates/genetic distances estimated in 49 CEU vs 49 YRI (black) samples, 49 CEU samples vs Hapmap (blue), and 49 YRI samples vs Hapmap (orange). Hapmap genetic map data is downloaded from (The International HapMap Consortium 2007).



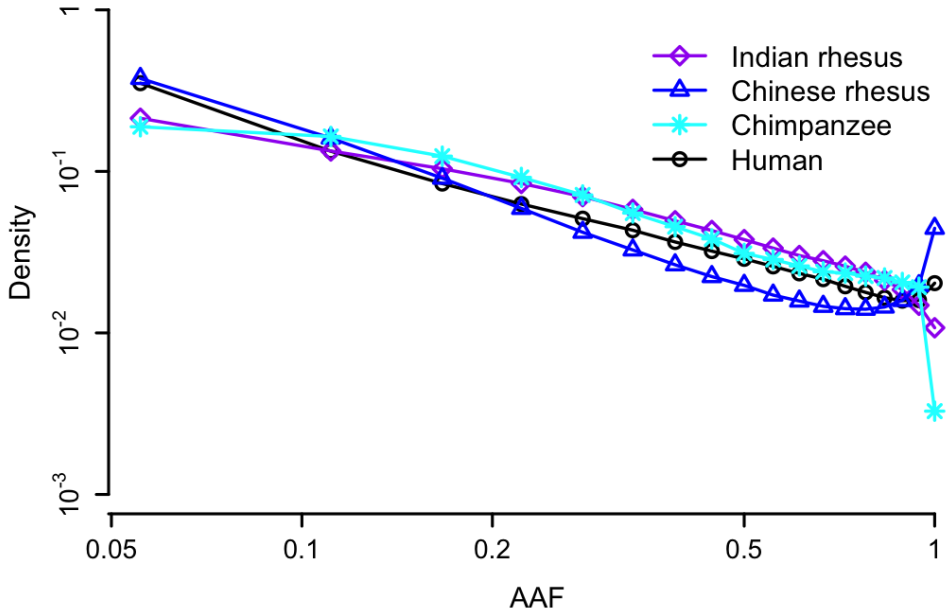

**Supplemental Figure S13.** Distribution of $4N_er$ on 439 non-inverted (panel A) and 95 inverted (panel B) orthologous autosomal syntenic regions estimated directly by software *LDHat* in both Indian rhesus and human (YRI) populations.



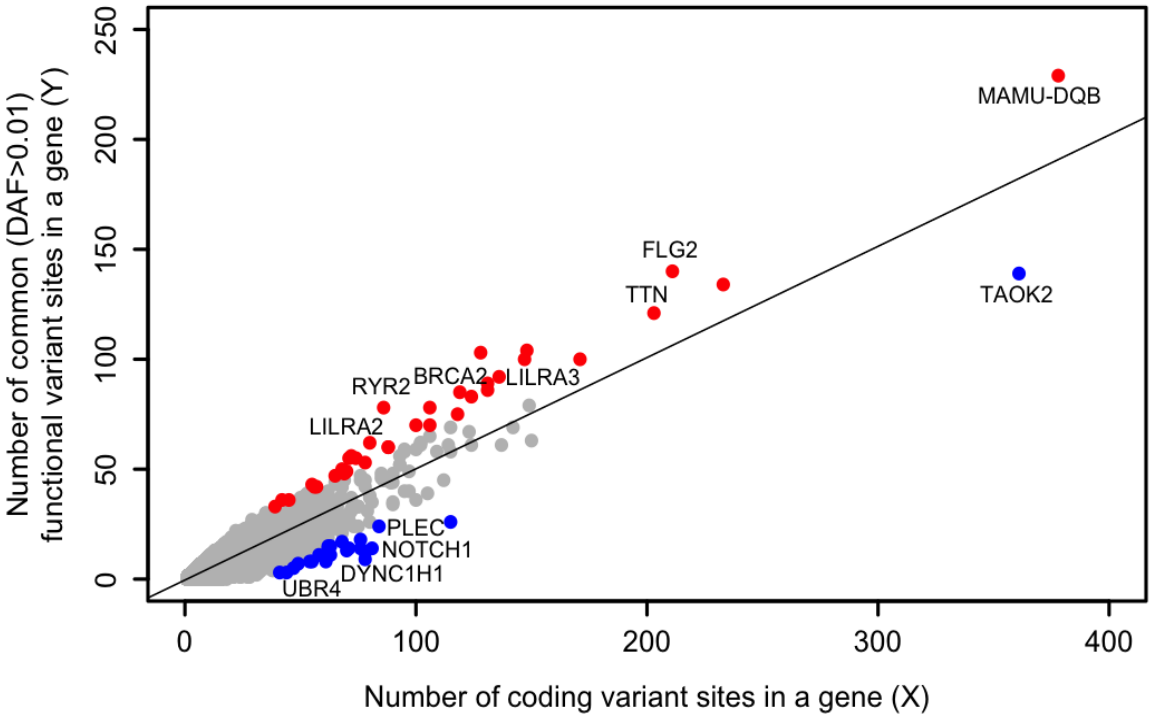

**Supplemental Figure S14.** Distribution of linkage disequilibrium (LD) correlation coefficient ($r^2$) on 439 non-inverted (panel A) and 95 inverted (panel B) orthologous autosomal syntenic regions calculated manually in both Indian rhesus and human (YRI) populations.



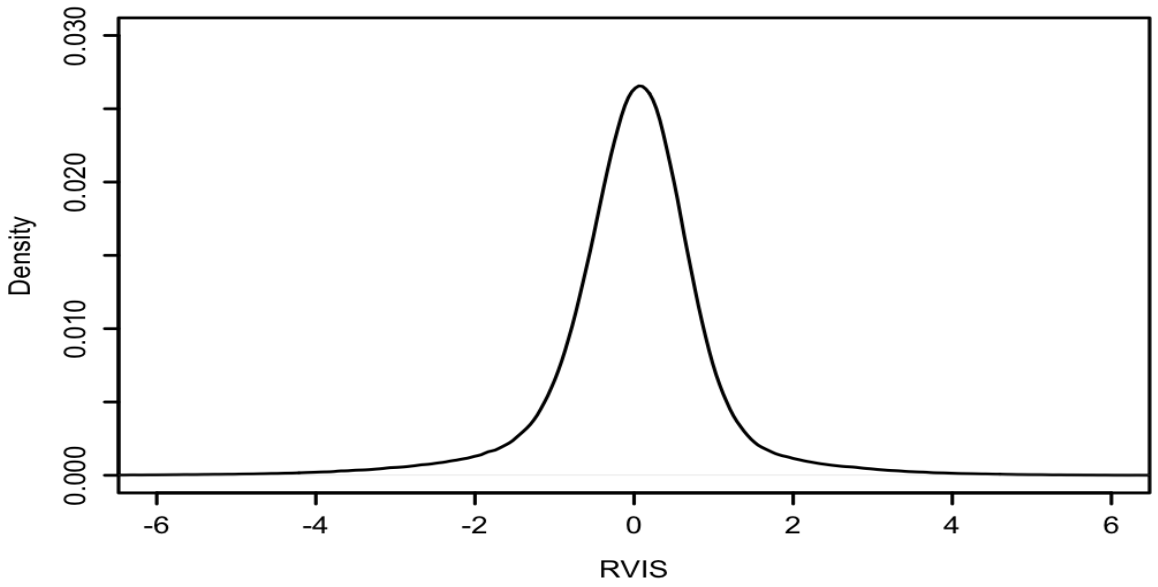

**Supplemental Figure S15A**. Density distribution calculated as the proportion of all SNVs that have alternative allele frequency (AAF) within different frequency bins. This is for autosomal variants in rhesus (only 123 Indian-origin rhesus samples) and human (n=123 samples from 1000 Genomes data, phase 1).




**Supplemental Figure S15B**. Density distribution calculated as the proportion of all SNVs that have alternative allele frequency (AAF) within different frequency bins. This plot compared distributions for AAF in Chinese rhesus (sample size n=9), Indian rhesus (9 samples randomly sampled from 123 samples), chimpanzees (9 samples randomly sampled from 10 samples, downloaded from http://panmap.uchicago.edu) and human (1000 Genomes data, phase 1, 9 samples randomly sampled from 1092 samples).




**Supplemental Figure S16**. Results for RVIS analysis. The grey points represent genes detected under neutral evolution (P>0.001). The red points (n=35) are genes with large positive RVIS that deviates significantly from neutrality (P<0.001). Blue points (n=22) are genes with large negative RVIS deviating significantly from neutrality (P<0.001).




**Supplemental Figure S17.** The null distribution of RVIS for 1.3 x $10^6$ Indian rhesus genes simulated under neutral evolution given the calculated demographics changes from this analysis. These genes include $10^6$ 10kb-, 2 x $10^5$ 50kb- and $10^5$ 100kb-segments.

**Supplemental Table S1: Sources for rhesus macaque DNA samples**

| Institution | Total sample size | Indian-origin | Chinese-origin |
|---|---|---|---|
| California NPRC | 20 | 14 | 6 |
| New England NPRC | 14 | 14 | 0 |
| Oregon NPRC | 17 | 17 | 0 |
| Southwest NPRC | 6 | 6 | 0 |
| Tulane NPRC | 19 | 19 | 0 |
| Wisconsin NPRC | 33 | 33 | 0 |
| Yerkes NPRC | 7 | 7 | 0 |
| Caribbean Primate Res. Center (Cayo Santiago) | 14 | 14 | 0 |
| Anhui University (China) | 3 | 0 | 3 (wild-caught) |
| Total | 133 | 124 | 9 |

**NPRC: National Primate Research Center**

**Supplemental Table S2: Comparison of model fitness**.

| Demographic model | Log-Likelihood | AIC |
|---|---|---|
| Stairway plot model – Indian rhesus | -197249.5 | 394799.0 |
| PSMC model (p 6+29*2) – Indian rhesus | -3041798.2 | 6083766.5 |
| PSMC model (p 4+25*2+4+6) – Indian rhesus | - -5116327.1 | 10232812.2 |
| Stairway plot model – Chinese rhesus | -31920.8 | 63865.6 |
| PSMC model (p 6+29*2) – Chinese rhesus | -17929.6 | 36029.3 |
| PSMC model (p 4+25*2+4+6) – Chinese rhesus | -19084.6 | 38327.2 |

Note: For each demographic model, 1 Gb DNA sequence was simulated with scrm program. Composite likelihood and Akaike information criterion of the observed site frequency spectrum (SFS) of 52 low-coverage Indian rhesus or 3 low-coverage Chinese rhesus was calculated using the SFS as the expected SFS. The Stairway plot model – Indian rhesus was the inferred demographic history of 75 high-coverage Indian rhesus. The PSMC model (p 6+29*2) – Indian rhesus and PSMC model (p 4+25*2+4+6) – Indian rhesus were obtained by averaging the inferred demographic histories of 3 high-coverage Indian rhesus. The Stairway plot model – Chinese rhesus was the inferred demographic history of 6 high-coverage Indian rhesus. The PSMC model (p 6+29*2) – Chinese rhesus and the PSMC model (p 4+25*2+4+6) – Chinese rhesus was obtained by averaging the inferred demographic histories of 3 high-coverage Chinese rhesus.

## Supplemental Table S3: Results from DFE-alpha analyses of rhesus SNVs

| | | | Neutrally Evolving | | | Selectively Evolving | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Percentage of Mutations in $N_e$s Range | | | | alpha | Maximum Likelihood log | | | Population Size | | |
| **Neutral Dataset** | **Epoch** | **Demographic Scenario** | **0-1** | **1-10** | **10-100** | **>100** | | **Neutral** | **Selected** | **AIC** | **N1** | **N2** | **N3** |
| Downstream | 1 | Constant population | 26.96% | 14.35% | 21.55% | 37.14% | -7.83% | -4,580,171.99 | -488,516.80 | 977,039.60 | 100 | | |
| | 2 | Search for the best-fitting population size n2 | 36.37% | 5.47% | 6.30% | 51.86% | -57.96% | -4,561,438.13 | -484,297.23 | 968,604.47 | 100 | 191 | |
| | 3 | Two population size changes | 32.09% | 8.88% | 11.33% | 47.69% | -34.98% | -4,546,359.33 | -483,382.50 | 966,779.00 | 100 | 40 | 170 |
| Intron | 1 | Constant population | 22.10% | 12.23% | 18.83% | 46.84% | 12.53% | -59,096,053.59 | -488,434.80 | 976,875.60 | 100 | | |
| | 2 | Search for the best-fitting population size n2 | 28.94% | 5.78% | 6.94% | 58.34% | -23.68% | -58,889,916.51 | -484,673.02 | 969,356.03 | 100 | 174 | |
| | 3 | Two population size changes | 26.73% | 7.41% | 9.47% | 56.39% | -11.65% | -58,681,661.52 | -483,378.24 | 966,770.48 | 100 | 40 | 170 |
| Synonymous | 1 | Constant population | 15.80% | 9.01% | 14.13% | 61.06% | 38.72% | -592,787.58 | -488,367.17 | 976,740.34 | 100 | | |
| | 2 | Search for the best-fitting population size n2 | 22.51% | 2.75% | 3.08% | 71.67% | 3.12% | -588,824.30 | -483,947.19 | 967,904.38 | 100 | 210 | |
| | 3 | Two population size changes | 19.52% | 5.22% | 6.61% | 68.65% | 19.62% | -587,230.86 | -483,352.60 | 966,719.20 | 100 | 40 | 170 |
| Upstream | 1 | Constant population | 26.91% | 14.33% | 21.52% | 37.24% | -7.65% | -4,591,336.69 | -488,515.70 | 977,037.40 | 100 | | |
| | 2 | Search for the best-fitting population size n2 | 36.30% | 5.46% | 6.29% | 51.95% | -57.66% | -4,573,442.89 | -484,297.17 | 968,604.33 | 100 | 191 | |
| | 3 | Two population size changes | 31.78% | 9.12% | 11.73% | 47.37% | -33.49% | -4,558,290.61 | -483,435.17 | 966,884.34 | 100 | 40 | 170 |

DFE-alpha results for nonsynonymous selected datasets compared to downstream (50kbp), intron, synonymous, and upstream (50kbp) neutral datasets. DFE-alpha was run using three demographic models: 1 epoch (constant population), 2 epoch (one population size change) and 3 epoch (two population changes). Calculated alpha values for each model were used to estimate the proportion of deleterious mutations with effects in four different ranges of fitness effects on a scale $N_e$s. The Akaike information criterion (AIC) was calculated for maximum likelihoods for the selected data set.

**Supplemental Table S4. List of human genes involved in eye or retinal diseases that were searched for putatively functional mutations in the rhesus macaque population survey.**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ABCA4 | CACNA1F | DFNB31 | GRK1 | KIF11 | NPHP4 | PHYH | RGS9 | TIMP3 | WFS1 |
| ABHD12 | CC2D2A | DHDDS | GRM6 | KLHL7 | NR2E3 | PITPNM3 | RGS9BP | TK2 | ZNF408 |
| ADAM9 | CDH23 | DMD | GUCA1B | LCA5 | NRL | PLA2G5 | RHO | TLR3 | ZNF423 |
| AHI1 | CEP164 | DTHD1 | GUCY2D | LRAT | OAT | PROM1 | RIMS1 | TLR4 | ZNF513 |
| AIPL1 | CEP290 | ELOVL4 | GUCY2F | LRIT3 | OFD1 | PRPF3 | RLBP1 | TLR6 | |
| ALMS1 | CERKL | EMC1 | HARS | LRP1B | OPA1 | PRPF31 | ROM1 | TMEM126A | |
| ARL2BP | CFB | ERCC6 | HTRA1 | LZTFL1 | OPA3 | PRPF6 | RP1L1 | TMEM216 | |
| ARL6 | CHM | EYS | IDH3B | MAK | OPN1LW | PRPF8 | RPE65 | TOPORS | |
| BBS10 | CIB2 | FAM161A | IFT140 | MERTK | OPN1SW | PRPH2 | RPGRIP1L | TREX1 | |
| BBS12 | CLRN1 | FBLN5 | IMPDH1 | MFRP | PAX2 | PVRL1 | RS1 | TRIM32 | |
| BBS2 | CNGA3 | FLVCR1 | IMPG1 | MKKS | PCDH15 | RAB28 | SAG | TRPM1 | |
| BBS4 | CNGB1 | FSCN2 | IMPG2 | MKS1 | PDE6A | RAX2 | SDCCAG8 | TTC8 | |
| BBS5 | CNGB3 | FZD4 | INPP5E | MPV17 | PDE6B | RB1 | SEMA4A | TTPA | |
| BBS7 | COL2A1 | GNAT1 | INVS | MTTP | PDE6C | RBP3 | SLC24A1 | TULP1 | |
| BBS9 | COL9A1 | GNAT2 | IQCB1 | MYO7A | PDE6H | RD3 | SNRNP200 | USH2A | |
| C3 | CRB1 | GNPTG | ITM2B | NDP | PDZD7 | RDH12 | SPATA7 | VCAN | |
| CA4 | CRX | GPR125 | KCNJ13 | NEK2 | PEX7 | RDH5 | TEAD1 | VPS13B | |
| CABP4 | CYP4V2 | GPR179 | KCNV2 | NPHP1 | PGK1 | RGR | TIMM8A | WDR19 | |

**Supplemental Table S5. List of rhesus macaque variants scored by HGMD or ClinVar as "disease causing" or "pathogenic." Analysis was performed using WGSA (Liu et al. 2015).**

Please refer to Excel file Supplemental_table _S5.xlsx.

**Supplemental Table S6. Sample names with NCBI SRA accession numbers, colony source of sample, ancestry, sex and sequencing read coverage.**

| NCBI SRA Accession | Sample Name | Colony source of sample | Ancestry | Sex | Sequence Coverage |
|---|---|---|---|---|---|
| SAMN03264739 | MMUL.CH-36390 | California National Primate Research Center | Chinese | F | 11.5 |
| SAMN03264740 | MMUL.CH-36394 | California National Primate Research Center | Chinese | F | 8.6 |
| SAMN03264732 | MMUL.IN-36332 | California National Primate Research Center | Indian | M | 9.7 |
| SAMN03264736 | MMUL.IN-36371 | California National Primate Research Center | Indian | F | 10.7 |
| SAMN03264738 | MMUL.CH-36389 | California National Primate Research Center | Chinese | M | 9.4 |
| SAMN03264734 | MMUL.IN-36357 | California National Primate Research Center | Indian | F | 10.3 |
| SAMN03264735 | MMUL.IN-36359 | California National Primate Research Center | Indian | F | 9.8 |
| SAMN03264737 | MMUL.IN-36374 | California National Primate Research Center | Indian | F | 9.9 |
| SAMN03264733 | MMUL.IN-36355 | California National Primate Research Center | Indian | F | 10.8 |
| SAMN03264762 | MMUL.IN-36460 | Caribbean Primate Research Center | Indian | F | 9.2 |
| SAMN03264763 | MMUL.IN-36467 | Caribbean Primate Research Center | Indian | F | 8.4 |
| SAMN03264767 | MMUL.IN-36477 | Caribbean Primate Research Center | Indian | F | 9.0 |
| SAMN03264768 | MMUL.IN-36476 | Caribbean Primate Research Center | Indian | F | 8.3 |
| SAMN03264773 | MMUL.IN-36475 | Caribbean Primate Research Center | Indian | M | 11.7 |
| SAMN03264775 | MMUL.IN-36474 | Caribbean Primate Research Center | Indian | M | 8.7 |
| SAMN03264779 | MMUL.IN-36466 | Caribbean Primate Research Center | Indian | F | 7.9 |
| SAMN03264676 | MMUL.IN-35250 | New England Primate Research Center | Indian | F | 7.7 |
| SAMN03264677 | MMUL.IN-35252 | New England Primate Research Center | Indian | F | 11.0 |
| SAMN03264678 | MMUL.IN-35253 | New England Primate Research Center | Indian | F | 7.9 |
| SAMN03264679 | MMUL.IN-35254 | New England Primate Research Center | Indian | M | 7.0 |
| SAMN03264681 | MMUL.IN-35256 | New England Primate Research Center | Indian | F | 10.3 |
| SAMN03264683 | MMUL.IN-35259 | New England Primate Research Center | Indian | F | 10.3 |
| SAMN03264695 | MMUL.IN-35717 | Oregon National Primate Research Center | Indian | F | 9.6 |
| SAMN03264696 | MMUL.IN-35718 | Oregon National Primate Research Center | Indian | F | 11.3 |
| SAMN03264697 | MMUL.IN-35722 | Oregon National Primate Research Center | Indian | F | 11.6 |
| SAMN03264699 | MMUL.IN-35724 | Oregon National Primate Research Center | Indian | F | 9.4 |
| SAMN03264700 | MMUL.IN-35728 | Oregon National Primate Research Center | Indian | F | 8.5 |
| SAMN03264702 | MMUL.IN-35730 | Oregon National Primate Research Center | Indian | F | 11.4 |
| SAMN03264703 | MMUL.IN-35732 | Oregon National Primate Research Center | Indian | F | 8.4 |
| SAMN03264725 | MMUL.IN-35969 | Southwest National Primate Research Center | Indian | F | 10.2 |
| SAMN03264726 | MMUL.IN-35972 | Southwest National Primate Research Center | Indian | M | 7.3 |
| SAMN03264727 | MMUL.IN-35975 | Southwest National Primate Research Center | Indian | F | 8.4 |
| SAMN03264716 | MMUL.IN-35895 | Tulane National Primate Research Center | Indian | F | 10.6 |
| SAMN03264718 | MMUL.IN-35907 | Tulane National Primate Research Center | Indian | M | 8.6 |
| SAMN03264719 | MMUL.IN-35916 | Tulane National Primate Research Center | Indian | F | 10.9 |
| SAMN03264721 | MMUL.IN-35921 | Tulane National Primate Research Center | Indian | F | 10.0 |
| SAMN03264722 | MMUL.IN-35923 | Tulane National Primate Research Center | Indian | F | 7.2 |
| SAMN03264724 | MMUL.IN-35957 | Tulane National Primate Research Center | Indian | F | 9.7 |
| SAMN03264715 | MMUL.IN-35883 | Tulane National Primate Research Center | Indian | M | 8.0 |
| SAMN03264717 | MMUL.IN-35902 | Tulane National Primate Research Center | Indian | M | 7.7 |
| SAMN03264720 | MMUL.IN-35919 | Tulane National Primate Research Center | Indian | F | 7.5 |
| SAMN03264605 | MMUL.IN-28499 | Wisconsin National Primate Research Center | Indian | M | 11.3 |
| SAMN03264606 | MMUL.IN-28500 | Wisconsin National Primate Research Center | Indian | F | 9.5 |

| | | | | | |
|---|---|---|---|---|---|
| SAMN03264607 | MMUL.IN-28507 | Wisconsin National Primate Research Center | Indian | M | 10.5 |
| SAMN03264608 | MMUL.IN-28518 | Wisconsin National Primate Research Center | Indian | F | 10.3 |
| SAMN03264609 | MMUL.IN-28535 | Wisconsin National Primate Research Center | Indian | M | 7.9 |
| SAMN03264610 | MMUL.IN-28555 | Wisconsin National Primate Research Center | Indian | M | 7.6 |
| SAMN03264618 | MMUL.IN-30423 | Wisconsin National Primate Research Center | Indian | F | 9.6 |
| SAMN03264619 | MMUL.IN-30424 | Wisconsin National Primate Research Center | Indian | M | 10.6 |
| SAMN03264685 | MMUL.IN-35490 | Yerkes National Primate Research Center | Indian | F | 11.4 |
| SAMN03264689 | MMUL.IN-35496 | Yerkes National Primate Research Center | Indian | F | 11.5 |
| SAMN03264694 | MMUL.IN-35502 | Yerkes National Primate Research Center | Indian | F | 8.9 |
| SAMN03264597 | MMUL.IN-18277 | Yerkes National Primate Research Center | Indian | F | 35.8 |
| SAMN03264598 | MMUL.IN-19466 | Yerkes National Primate Research Center | Indian | M | 36.2 |
| SAMN03264600 | MMUL.IN-24898 | Wisconsin National Primate Research Center | Indian | F | 39.4 |
| SAMN03264613 | MMUL.IN-30119 | Wisconsin National Primate Research Center | Indian | F | 35.9 |
| SAMN03264614 | MMUL.IN-30136 | Wisconsin National Primate Research Center | Indian | M | 29.9 |
| SAMN03264616 | MMUL.IN-30158 | Wisconsin National Primate Research Center | Indian | M | 42.1 |
| SAMN03264620 | MMUL.IN-31505 | Wisconsin National Primate Research Center | Indian | F | 36.5 |
| SAMN03264621 | MMUL.IN-32510 | Wisconsin National Primate Research Center | Indian | F | 25.6 |
| SAMN03264622 | MMUL.IN-32538 | Wisconsin National Primate Research Center | Indian | M | 26.9 |
| SAMN03264623 | MMUL.IN-32754 | Wisconsin National Primate Research Center | Indian | F | 27.7 |
| SAMN03264624 | MMUL.IN-33674 | Wisconsin National Primate Research Center | Indian | M | 32.1 |
| SAMN03264625 | MMUL.IN-33707 | Wisconsin National Primate Research Center | Indian | M | 31.0 |
| SAMN03264627 | MMUL.IN-34597 | New England Primate Research Center | Indian | F | 36.5 |
| SAMN03264629 | MMUL.IN-34600 | New England Primate Research Center | Indian | M | 32.1 |
| SAMN03264630 | MMUL.IN-34602 | New England Primate Research Center | Indian | M | 35.0 |
| SAMN03264635 | MMUL.IN-34762 | Yerkes National Primate Research Center | Indian | F | 38.2 |
| SAMN03264639 | MMUL.IN-34770 | Yerkes National Primate Research Center | Indian | F | 37.3 |
| SAMN03264641 | MMUL.IN-35044 | Oregon National Primate Research Center | Indian | M | 36.2 |
| SAMN03264642 | MMUL.IN-35045 | Oregon National Primate Research Center | Indian | M | 60.7 |
| SAMN03264643 | MMUL.IN-35046 | Oregon National Primate Research Center | Indian | M | 34.2 |
| SAMN03264645 | MMUL.IN-35048 | Oregon National Primate Research Center | Indian | M | 40.0 |
| SAMN03264646 | MMUL.IN-35049 | Oregon National Primate Research Center | Indian | M | 42.7 |
| SAMN03264647 | MMUL.IN-35051 | Oregon National Primate Research Center | Indian | F | 35.1 |
| SAMN03264649 | MMUL.IN-35055 | Oregon National Primate Research Center | Indian | F | 32.6 |
| SAMN03264650 | MMUL.IN-35059 | Oregon National Primate Research Center | Indian | M | 37.2 |
| SAMN03264651 | MMUL.IN-35060 | Oregon National Primate Research Center | Indian | M | 41.2 |
| SAMN03264652 | MMUL.IN-35061 | Oregon National Primate Research Center | Indian | M | 40.8 |
| SAMN03264653 | MMUL.CH-35082 | California National Primate Research Center | Chinese | F | 45.7 |
| SAMN03264657 | MMUL.CH-35086 | California National Primate Research Center | Chinese | F | 32.0 |
| SAMN03264658 | MMUL.IN-35087 | California National Primate Research Center | Indian | F | 30.3 |
| SAMN03264659 | MMUL.IN-35088 | California National Primate Research Center | Indian | M | 32.5 |
| SAMN03264660 | MMUL.IN-35089 | California National Primate Research Center | Indian | F | 33.7 |
| SAMN03264661 | MMUL.IN-35090 | California National Primate Research Center | Indian | M | 31.8 |
| SAMN03264662 | MMUL.IN-35091 | California National Primate Research Center | Indian | F | 57.4 |
| SAMN03264666 | MMUL.IN-35095 | California National Primate Research Center | Indian | M | 36.1 |
| SAMN03264667 | MMUL.IN-35096 | California National Primate Research Center | Indian | F | 35.8 |
| SAMN03264668 | MMUL.IN-35144 | New England Primate Research Center | Indian | M | 32.0 |
| SAMN03264670 | MMUL.IN-35150 | New England Primate Research Center | Indian | M | 33.8 |
| SAMN03264672 | MMUL.IN-35154 | New England Primate Research Center | Indian | M | 33.5 |
| SAMN03264673 | MMUL.IN-35160 | New England Primate Research Center | Indian | M | 36.1 |

| | | | | | |
|---|---|---|---|---|---|
| SAMN03264674 | MMUL.IN-35162 | New England Primate Research Center | Indian | M | 32.5 |
| SAMN03264705 | MMUL.IN-35864 | Tulane National Primate Research Center | Indian | M | 32.7 |
| SAMN03264706 | MMUL.IN-35865 | Tulane National Primate Research Center | Indian | M | 41.0 |
| SAMN03264707 | MMUL.IN-35866 | Tulane National Primate Research Center | Indian | F | 41.0 |
| SAMN03264708 | MMUL.IN-35868 | Tulane National Primate Research Center | Indian | F | 34.7 |
| SAMN03264709 | MMUL.IN-35871 | Tulane National Primate Research Center | Indian | M | 39.1 |
| SAMN03264710 | MMUL.IN-35872 | Tulane National Primate Research Center | Indian | F | 39.2 |
| SAMN03264711 | MMUL.IN-35873 | Tulane National Primate Research Center | Indian | M | 42.7 |
| SAMN03264712 | MMUL.IN-35874 | Tulane National Primate Research Center | Indian | M | 37.8 |
| SAMN03264713 | MMUL.IN-35875 | Tulane National Primate Research Center | Indian | F | 42.9 |
| SAMN03264714 | MMUL.IN-35876 | Tulane National Primate Research Center | Indian | M | 42.7 |
| SAMN03264728 | MMUL.IN-35976 | Southwest National Primate Research Center | Indian | F | 41.1 |
| SAMN03264729 | MMUL.IN-35990 | Southwest National Primate Research Center | Indian | M | 38.2 |
| SAMN03264730 | MMUL.CH-36013 | California National Primate Research Center | Chinese | F | 41.0 |
| SAMN03264761 | MMUL.IN-39345 | California National Primate Research Center | Indian | F | 42.0 |
| SAMN03083651 | MMUL.IN-36468 | Caribbean Primate Research Center | Indian | F | 40.4 |
| SAMN03264769 | MMUL.IN-36461 | Caribbean Primate Research Center | Indian | M | 41.1 |
| SAMN03264770 | MMUL.IN-36462 | Caribbean Primate Research Center | Indian | M | 38.2 |
| SAMN03264771 | MMUL.IN-36463 | Caribbean Primate Research Center | Indian | M | 39.1 |
| SAMN03264774 | MMUL.IN-36473 | Caribbean Primate Research Center | Indian | M | 42.3 |
| SAMN03264780 | MMUL.IN-36471 | Caribbean Primate Research Center | Indian | F | 40.9 |
| SAMN03264781 | MMUL.IN-36470 | Caribbean Primate Research Center | Indian | F | 40.3 |
| SAMN03264764 | MMUL.IN-11414-01b | Wisconsin National Primate Research Center | Indian | F | 34.5 |
| SAMN03264765 | MMUL.IN-11433-07b | Wisconsin National Primate Research Center | Indian | F | 53.3 |
| SAMN03264766 | MMUL.IN-11433-08b | Wisconsin National Primate Research Center | Indian | M | 42.0 |
| SAMN03264741 | MMUL.IN-37730 | Wisconsin National Primate Research Center | Indian | M | 42.7 |
| SAMN03264743 | MMUL.IN-37732 | Wisconsin National Primate Research Center | Indian | F | 44.0 |
| SAMN03264744 | MMUL.IN-37733 | Wisconsin National Primate Research Center | Indian | F | 42.5 |
| SAMN03264745 | MMUL.IN-37734 | Wisconsin National Primate Research Center | Indian | M | 38.4 |
| SAMN03264746 | MMUL.IN-37735 | Wisconsin National Primate Research Center | Indian | M | 41.3 |
| SAMN03264749 | MMUL.IN-37738 | Wisconsin National Primate Research Center | Indian | F | 38.8 |
| SAMN03264750 | MMUL.IN-37739 | Wisconsin National Primate Research Center | Indian | M | 40.7 |
| SAMN03264751 | MMUL.IN-37740 | Wisconsin National Primate Research Center | Indian | M | 41.6 |
| SAMN03264752 | MMUL.IN-37741 | Wisconsin National Primate Research Center | Indian | M | 38.4 |
| SAMN03264753 | MMUL.IN-37742 | Wisconsin National Primate Research Center | Indian | M | 40.4 |
| SAMN03264756 | MMUL.IN-37745 | Wisconsin National Primate Research Center | Indian | F | 39.6 |
| SAMN03264757 | MMUL.IN-37746 | Wisconsin National Primate Research Center | Indian | F | 38.6 |
| SAMN03264758 | MMUL.CH-37854 | Wild Caught | Chinese | M | 34.7 |
| SAMN03264759 | MMUL.CH-37945 | Wild Caught | Chinese | M | 23.2 |
| SAMN03264760 | MMUL.CH-37950 | Wild Caught | Chinese | F | 29.6 |
| SAMN02981228 | MMUL.IN-17573 | Southwest National Primate Research Center | Indian | F | 41.5 |

**Supplemental Table S7. The ethnicity component of 123 human samples (compared to 123 Indian rhesus samples) and 9 samples (compared to 9 Chinese rhesus samples) randomly selected out of 1092 samples of 1000 Genome phase 1 data.**

| Populations | Sample# in 123 samples | Sample# in 9 samples |
|---|---|---|
| ASW | 6 | 1 |
| CEU | 9 | |
| CHB | 12 | |
| CHS | 11 | 1 |
| CLM | 3 | |
| FIN | 12 | |
| GBR | 11 | 1 |
| IBS | 2 | |
| JPT | 8 | 2 |
| MXL | 5 | |
| LWK | 13 | 1 |
| PUR | 4 | |
| TSI | 14 | 2 |
| YRI | 13 | 1 |
| Total | 123 | 9 |

**Supplemental Table S8. Comparison of selected recombination rates on chimpanzee chromosome 19 estimated by software LDHat and the original results reported by Auton et al (Auton et al. 2012).**

| Physical position (BP) | Auton et al's gold standard | Recombination rate estimated |
|---|---|---|
| 32283 | 5.23453 | 5.1763 |
| 160265 | 5.25162 | 5.2019 |
| 171883 | 5.33109 | 5.26139 |
| 171894 | 5.39152 | 5.35215 |
| 171900 | 5.46739 | 5.42645 |
| 173758 | 5.46739 | 5.43385 |
| 175470 | 5.47135 | 5.44415 |
| 176680 | 5.47817 | 5.45525 |
| 180599 | 5.47817 | 5.44439 |
| 183015 | 5.47817 | 5.44247 |
| 183466 | 5.47817 | 5.44535 |
| 189555 | 5.47817 | 5.45053 |
| 189559 | 5.47866 | 5.44922 |
| 190282 | 5.47866 | 5.45372 |
| 193970 | 3.33944 | 3.95515 |
| 194197 | 0.054 | 0.05479 |
| 196522 | 0.05422 | 0.0522 |
| 197075 | 0.05413 | 0.05204 |
| 197783 | 0.05401 | 0.05221 |
| 198230 | 0.05566 | 0.05236 |
| 198580 | 0.1431 | 0.09694 |
| 198823 | 4.80846 | 4.95894 |
| 199762 | 4.81867 | 4.96966 |
| 199816 | 4.82148 | 4.97129 |
| 199821 | 4.82148 | 4.97618 |
| 200214 | 4.84606 | 4.98573 |

**Supplemental Table S9. Regression analysis of read depth coverage for genes scored as under positive or negative selection using RVIS**

| Predicted selection | Gene | Adjusted $R^2$ | F-statistic P-value | Minimum residual | Maximum residual |
|---|---|---|---|---|---|
| Negative selection | TAOK2 | 0.8193 | < 2.2x10^-16 | -2.1803 | 2.0930 |
| | DYNC1H1 | 0.8324 | < 2.2x10^-16 | -2.1468 | 2.1012 |
| | NOTCH1 | 0.6992 | < 2.2x10^-16 | -3.1941 | 2.5455 |
| | DNAH8 | 0.9557 | < 2.2x10^-16 | -1.4054 | 5.8598 |
| | ABCA2 | 0.6458 | < 2.2x10^-16 | -3.8048 | 3.1876 |
| | MYH11 | 0.8135 | < 2.2x10^-16 | -2.1787 | 2.4975 |
| | RBP3 | 0.5778 | < 2.2x10^-16 | -2.1569 | 2.7022 |
| | FASN | 0.5829 | < 2.2x10^-16 | -3.8858 | 2.5988 |
| | MYOM2 | 0.7762 | < 2.2x10^-16 | -2.1843 | 3.4487 |
| | DNAH1 | 0.5508 | < 2.2x10^-16 | -1.8457 | 2.3864 |
| | CSMD1 | 0.942 | < 2.2x10^-16 | -1.5204 | 6.3514 |
| | FLNB | 0.8807 | < 2.2x10^-16 | -2.0161 | 2.8973 |
| | MYH2 | 0.9227 | < 2.2x10^-16 | -1.6807 | 6.6676 |
| | TSHZ1 | 0.8435 | < 2.2x10^-16 | -1.8824 | 5.6681 |
| | SDK1 | 0.8709 | < 2.2x10^-16 | -3.1334 | 3.9598 |
| | RYR3 | 0.9149 | < 2.2x10^-16 | -1.6341 | 4.5853 |
| | MYH3 | 0.7918 | < 2.2x10^-16 | -1.9988 | 2.6697 |
| | ALPK2 | 0.8952 | < 2.2x10^-16 | -1.5189 | 3.8238 |
| | TTN | 0.8084 | < 2.2x10^-16 | -1.2909 | 4.7808 |
| Positive selection | RP1 | 0.9292 | < 2.2x10^-16 | -4.4750 | 4.8282 |
| | LILRA2 | 0.8232 | < 2.2x10^-16 | -1.8336 | 3.2859 |
| | DST | 0.9092 | < 2.2x10^-16 | -1.8448 | 4.6135 |
| | ASPM | 0.8924 | < 2.2x10^-16 | -3.4194 | 4.5248 |
| | BMS1 | 0.9577 | < 2.2x10^-16 | -1.8680 | 5.9789 |
| | SLFN13 | 0.952 | < 2.2x10^-16 | -3.1924 | 5.9652 |
| | LILRA3 | 0.7591 | < 2.2x10^- | -2.3094 | 2.5293 |

| | | | | |
|---|---|---|---|---|
| | | 16 | | |
| BRCA2 | 0.9529 | < 2.2x10^-16 | -2.4408 | 4.8699 |
| FLG2 | 0.3403 | 8.658x10^-13 | -1.3690 | 2.3387 |
| RYR2 | 0.9211 | < 2.2x10^-16 | -5.6916 | 4.3074 |
| MAMU-DQB | 0.775 | < 2.2x10^-16 | -3.3104 | 2.5532 |

* Average coverages on each Indian rhesus sample are calculated in whole genome and genic (listed in the table) regions. Regression is carried out between average coverages in whole genome and each genic region. "$R^2$" is adjusted regression coefficient. "Residual" is the normalized difference between the observed data of y and the fitted values ŷ. And the distribution of residuals is nearly standard normal.

References


Auton A, Fledel-Alon A, Pfeifer S, Venn O, Segurel L, Street T, Leffler EM, Bowden R, Aneas I, Broxholme J et al. 2012. A fine-scale chimpanzee genetic map from population sequencing. *Science* **336**(6078): 193-198.

Keinan A, Clark AG. 2012. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* **336**(6082): 740-743.

Liu X, White S, Peng B, Johnson AD, Brody JA, Li AH, Huang Z, Carroll A, Wei P, Gibbs R et al. 2015. WGSA: an annotation pipeline for human genome sequencing studies. *Journal of medical genetics*.

Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. 2013. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* **9**(8): e1003709.

Rhesus Macaque Genome Sequencing and Analysis Consortium *et al*. 2007. Evolutionary and Biomedical Insights from the Rhesus Macaque Genome. *Science* **316**(5822): 222-234.

The International HapMap Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**(7164): 851-861.

Vitti JJ, Grossman SR, Sabeti PC. 2013. Detecting natural selection in genomic data. *Annual review of genetics* **47**: 97-120.