Supplementary Information for Lea et al. "**Dominance rank-associated immune gene expression is widespread, sex-specific, and a precursor to high social status in wild male baboons**"

Table of contents:

Supplementary Materials and Methods

Study subjects, sample collection, and sample processing

All study subjects were members of a long-term study population of yellow baboons (*Papio cynocephalus*, with some admixture from a closely related species, the anubis baboon, *P. anubis* [1]) that has been monitored by the Amboseli Baboon Research Project (ABRP) since 1971 [2]. Animals in the study population are individually recognized and observed on a near daily basis from birth onwards. Thus, the ages of individuals born in the study population were known to within a few days' error. For study subjects that immigrated into the study population as adults (n = 10 males in our data set), ages were estimated by trained observers based on morphological features and comparison to known-age animals [3]. Dominance hierarchies were constructed monthly for every social group in the study population based on the outcomes of dyadic aggressive encounters. Ordinal dominance ranks were assigned to every adult based on these hierarchies, such that low numbers signify high rank/social status and high numbers signify low rank/social status [4].

Blood samples were collected from each study subject (n = 61) in May through August of 2012-2016 following well-established procedures [5–8]. Briefly, animals were immobilized by an anesthetic-bearing dart delivered through a hand-held blow gun, and, following immobilization, were quickly transferred to a processing site for blood sample collection. At the processing site, we collected two types of samples for each individual:

(i) 2 – 4 mL whole blood in a CPT vacutainer tube (Becton, Dickinson, and Company) to isolate peripheral blood mononuclear cells (PBMCs). CPT tubes were stored overnight at the field site and shipped the next day to the Institute of Primate Research (IPR) in Nairobi. At IPR, PBMCs were purified, antibody stained for cell surface markers that discriminate monocytes

(CD3$^-$, CD20$^-$, CD14$^+$), Natural Killer cells (CD3$^-$, CD20$^-$, CD16$^+$), B cells (CD3$^-$, CD20$^+$), helper T cells (CD3$^+$, CD8$^-$, CD4$^+$), and cytotoxic T cells (CD3$^+$, CD8$^+$, CD4$^-$), and profiled for PBMC composition using flow cytometry on a BD FacsCalibur machine. To distinguish T and B cells, we stained 0.5 million purified PBMCs with 3 ul anti-CD3-APC-Cy7 (clone SP34-2, BD Biosciences #557757), 5 ul anti-CD4-FITC (clone L200, BD Biosciences #550628), 1 ul anti-CD20-PE-Cy7 (clone L27, BD Biosciences #335793), and 5 ul anti-CD8-PE (clone 3B5, Invitrogen #MHCD0804). To distinguish Natural Killer and monocyte cells, we stained a second aliquot of 0.5 million purified PBMCs with 3 ul anti-CD3-APC-Cy7, 5 ul anti-CD16-PE (clone 3G8, BD Biosciences #560995), and 5 ul anti-CD14-FITC (clone 322A-1 MY4, Beckman Coulter #6603262).

(ii) 1 mL of whole blood in each of two TruCulture tubes (Myriad RBM) to assess the cytokine and gene expression response to lipopolysaccharide (LPS). For each animal, blood was collected into one tube that contained cell culture media alone (the 'NULL' tube) and a second tube that contained culture media plus 0.1 ug/mL lipopolysaccharide (the 'LPS' tube). NULL and LPS tubes were then incubated in parallel at 37 ºC for 10 hours. Following incubation, we collected serum for cytokine profiling, lysed the red blood cell fraction (PureGene Red Cell Lysis Buffer, QIAGEN), and collected white blood cells for gene expression profiling. Serum samples and RNAlater-preserved white blood cells (ThermoFisher Scientific) were stored at -20 ºC until transport to the United States.

Following sample collection, study subjects were allowed to regain consciousness in a covered holding cage until they were fully recovered from the effects of the anesthetic, and then released near their social group.

Generation and processing of cytokine data

For a subset of individuals (n=29; n=18 males and 11 females), we measured circulating levels of 23 cytokines involved in the immune response (Dataset S1). Specifically, we used serum isolated from both the LPS and NULL condition TruCulture tubes to perform cytokine profiling with the MILLIPLEX MAP Non-Human Primate Cytokine Magnetic Bead Panel (EMD Millipore) following the manufacturer's instructions. All samples were assayed in duplicate, and all cytokine work was performed by the Immunology Unit of the Duke University Regional Biocontainment Laboratory.

We excluded a given cytokine from downstream analyses if more than half of our samples did not exceed the lower limit of quantification for that cytokine. Further, we computed the correlation between normalized cytokine values for duplicate samples and excluded measures with $R^2<0.8$ between replicates. We did not exclude any individual samples from analyses. For the remaining 15 cytokines that passed our filters, we tested for differences between LPS and NULL condition samples using linear mixed effects models implemented in the R package 'nlme' [9] . Specifically, we modeled each set of normalized cytokine values as a function of condition (NULL or LPS), age of the donor, sex of the donor, and individual identity (as a random effect). We extracted the p-values associated with the condition effects and corrected for multiple hypothesis testing using an FDR approach [10,11] (Figure S1).

Generation and low level processing of mRNA-seq data

For each TruCulture sample, we extracted RNA from white blood cells stored in RNAlater (ThermoFisher Scientific) using the RNeasy mini kit (QIAGEN) following the manufacturers' instructions. RNA quality was assessed for a random subset of samples (n=36)

using an Agilent RNA 6000 Nano kit and an Agilent 2100 Bioanalyzer (mean ± SD of RIN values = 8.56 ± 0.86).

For each sample, we used 200 ng of total RNA as the input for mRNA isolation using the NEBNext Poly(A) mRNA Isolation Module (New England BioLabs). We generated mRNA-seq libraries for high-throughput sequencing from the isolated mRNA using the NEBNextUltra RNA Library Prep Kit for Illumina (New England BioLabs), following the manufacturers' instructions. We pooled 10-12 samples per lane of sequencing (100 bp paired-end) on an Illumina HiSeq 2500. We recovered a mean of 18.03 ± 9.94 (SD) million reads per individual (Dataset S2).

Following sequencing, we trimmed Illumina adapter sequence and low quality bases from the ends of the reads using the default settings in Trimmomatic [12]. We mapped trimmed reads to the anubis baboon genome (*Panu* 2.0) using the STAR aligner and the recommended two-pass method [13]. For each gene, we collated the number of reads that overlapped any annotated exon using the program HTSeq [14] and NCBI's *Panu* 2.0 RefSeq exon annotations [15]. In downstream analyses, we only included genes with mean RPKM values > 1 in both the NULL or LPS condition. We retained 7576 genes after applying these filters. At this stage, we also removed the LPS condition sample from one individual who appeared not to respond to stimulation (RPKM value for the IL6 gene was <1 in the LPS condition). This filtering left us with n=121 total samples, 61 from the NULL condition and 60 from the LPS condition.

Prior to analysis, we normalized the read count data using the function '*voomWithQualityWeights'* in the R package *limma* [16]. Further, we removed known batch effects (i.e., the year of sample collection) as well as effects of cell type composition using linear models implemented in *limma*. To do so, we first performed PCA on the relative abundance data

for each of 5 cell types described above, and used the loadings from the first two principal components (which together explained 84.44% of the total variance) as covariates in linear models (see Figure S3 for analyses of rank effects on cell type composition). Finally, to obtain the PCA projection shown in Figure 2, we computed the covariance matrix of normalized, batch- and cell type-corrected gene expression values for our set of filtered genes and used this matrix as the input for the '*prcomp'* function in R.

Genotyping

We used genotype data to confirm that paired LPS and NULL samples were matched to the same individual, to estimate pairwise genetic relatedness, and to perform Mendelian randomization. To do so, we called variants across all regions within 200 kb of an annotated gene (i.e., within the gene body or within 200 kb of the transcription start or end site) using HaplotypeCaller from the Genome Analysis Toolkit (GATK v3.3.0). For all steps, we followed the Best Practices for variant calling using RNA-seq data (https://www.broadinstitute.org/gatk/guide/article?id=3891). After genotyping, we retained sites that passed the following filters: variant quality score ≥100; QD < 2.0; MQ < 35.0; FS > 60.0; HaplotypeScore >13.0; MQRankSum < −12.5; and ReadPosRankSum < −8.0. Additionally, we used the program vcftools [17] to remove variant calls with quality scores < 10, as well as sites that had a mean depth of coverage < 5x or that were not in Hardy-Weinberg equilibrium ($p<0.05$). This filtering left us with 99,760 SNPs. We imputed data for missing genotype values (10.93%) using default settings in Beagle [18].

To obtain our final call set, we averaged the filtered, imputed genotype calls from the LPS and NULL conditions for each individual at each locus (resulting in a numeric value

between 0 and 2 for each; note that for all individuals, genotype calls from the two conditions were identical at >99% of genotyped sites). To estimate pairwise relatedness between individuals, we used the '*relatedness2'* option in vcftools [17,19].

Testing for associations between rank and gene expression

To identify genes for which gene expression was significantly predicted by dominance rank, we used linear mixed effects models implemented in the R package 'EMMREML' [20]. Specifically, for each gene in our dataset, we ran the following model:

$$y_i = \mu + r_i\beta_{r1} * I(s_i = 0) + r_i\beta_{r2} * I(s_i = 1) + a_i\beta_{a1} * I(s_i = 0) + a_i\beta_{a2} * I(s_i = 1) + c_i\beta_c + g_i + e_i, \quad (1)$$

$$g_i \sim MVN\left(0, \sigma_g^2 \boldsymbol{K}\right),$$

$$e_i \sim MVN(0, \sigma_e^2 \boldsymbol{I})$$

where y_i is the gene expression level estimate for sample i, μ is the intercept, c_i is a binary variable indicating whether sample i is from the control or LPS condition (1=control and 0=LPS), and β_c is the corresponding estimate of the condition effect. I is an indicator variable for sex (s_i; 0=female and 1=male). a_i and r_i represent the age and dominance rank, respectively, of the focal individual at the time of sample collection. e_i is a random effects term to control for environmental noise, and g_i is a random effects term to control for kinship and other sources of genetic structure. $\boldsymbol{K}$ is an n x n matrix that contains estimates of pairwise genetic relatedness derived from genotype data. σ_g^2 and σ_e^2 are the genetic and environmental variance components, respectively. $\boldsymbol{I}$ is the identity matrix, and MVN denotes the multivariate normal distribution. We chose to use a mixed effects model of this type in order to exclude false positive associations between dominance rank and gene expression that could emerge if ranks are more similar

between related individuals (as we know to be true in female baboons) and gene expression patterns are also more similar between related individuals (which is often the case for gene expression because gene expression levels are partially heritable in this and other populations: [5,21]). Mixed models that fit a random effect to account for genetic non-independence therefore test for associations between predictor and response variables of interest (here, dominance rank and gene expression), beyond that explained by genetic covariance between the study subjects [22,23].

We also tested for interactions between dominance rank and condition (NULL or LPS), as previous work has shown that rank effects on gene expression are more pronounced after LPS stimulation [24]. To do so, we ran the following model using data from males only (the sex where additive effects of dominance rank were common; n=70 samples from 31 individuals):

$$y_i = \mu + r_i\beta_r + a_i\beta_a + c_i\beta_c + (r_i * c_i)\beta_{rxc} + g_i + e_i \quad (2)$$

where $(r_i * c_i)$ represents the interaction between dominance rank and condition, and β_{rxc} is the effect size of the interaction term. All other terms are as described above.

As an alternative approach to testing for interactions, we tested for effects of male dominance rank on the magnitude of the gene expression response to LPS, using the fold change in gene expression levels between LPS and NULL conditions as the outcome variable. Specifically, for each individual, we subtracted the *voom* normalized gene expression values estimated for the NULL sample from the normalized values for the LPS sample (*voom* normalized values are already $\log_2$-transformed, so subtraction in this case is equivalent to fold change). Using these values, we ran the following model where all predictor variables are as described above except y_i, which in equation 3 denotes the $\log_2$ fold-change response to LPS:

$$y_i = \mu + r_i\beta_r + a_i\beta_a + g_i + e_i \quad (3)$$

For each gene, we extracted the p-value associated with the rank effect (nested within sex from equation 1, or without nesting from equation 3, for males only) or the rank interaction with condition (from equation 2). We corrected these distributions for multiple hypothesis testing using an FDR approach, and considered genes to be rank-associated if they passed a 5% FDR [10,11]. As described in the main text, we identified few rank x condition interactions or effects of rank on fold-change gene expression. Rather, genes that were more highly expressed in high-ranking (low-ranking) individuals at baseline tended to remain so after LPS stimulation, including those in innate immune defense and inflammation-related pathways (see also Figure S8).

Annotation of rank-associated genes

We performed Gene Ontology (GO) enrichment analyses using the Cytoscape module ClueGO [25], using one-sided Fisher's Exact Tests and a Benjamini-Hochberg FDR approach to correct for multiple hypothesis testing [26]. To reduce our multiple testing burden and to account for the nested nature of GO terms, we focused our analyses on terms that: (i) were within levels 3-8 of the Biological Process GO set; (ii) included at least 10 expressed genes from our data set; and (iii) included > 5% of all genes in the GO term in the test gene set. We report significant terms as those that were enriched in the test gene set at a 5% FDR (full results are provided in Dataset S4-5).

To investigate rank-related polarization of the TLR4 signaling pathway, we used previously compiled lists of genes associated with a MyD88- or TRIF-dependent response (obtained from antigen stimulation experiments in MyD88 or TRIF knock-out mice [27]). 234/542 of the MyD88-dependent genes and 165/400 of the TRIF-dependent genes identified by

[27] had expressed orthologs in our dataset. Using these gene sets, we performed two analyses. First, we asked whether the distribution of dominance rank effect sizes differed between MyD88-versus TRIF-induced genes (Mann-Whitney U test). To do so, we focused on those genes that were significantly associated with rank in males and also upregulated in response to LPS. Second, we asked whether male social status predicted composite expression variation across all genes in the MyD88 or TRIF-dependent sets. To do so, we extracted, for each individual, the median normalized, batch- and cell type composition-corrected gene expression level for all genes measured in the LPS condition that were dependent on MyD88 or TRIF for antigen-stimulated up-regulation. Using these median values, we used Spearman's rank correlations to ask whether dominance rank predicted median gene expression levels for the set of MyD88 versus TRIF-induced genes.

Comparison of rank-associated genes in female macaques and male baboons

Previous work [24] reported strong, causal effects of dominance rank on gene expression in captive rhesus macaques. Specifically, Snyder-Mackler et al. manipulated female social status (n=45) and profiled gene expression in sorted cell populations, as well as in leukocytes at baseline and following immune stimulation with LPS. They found that genes associated with innate immune function and a pro-inflammatory phenotype were upregulated in low-ranking animals, who also mounted a stronger response to LPS. To compare our results with theirs, we compared our estimates of standardized rank effects in males (β_{r2}) to female macaque standardized rank effect estimates from leukocytes unexposed or exposed to LPS (Table S13 from [24]). In the macaque study, social status was measured using Elo scores, such that higher numbers indicated higher social status; in our study, social status was measured using ordinal

ranks, such that higher numbers indicated lower social status. Therefore, for visualization (Figure 3 and Figure S9), we polarized effect sizes from both studies so that a negative beta was equivalent to higher expression of a given gene in high status individuals. We used Spearman's rank correlations to estimate the consistency of effect size estimates between datasets, and a binomial test to understand whether effect size estimates were directionally consistent more often than expected by chance.

Behavioral mediation analyses

To ask whether behaviors associated with high or low social status in males mediate the relationship between dominance rank and gene expression, we first created an index of received and initiated harassment for each individual. To do so, we extracted observations of dyadic agonistic encounters from the Amboseli Baboon Research Project's long-term database, BABASE. Data on these encounters are collected in the context of random-order focal sampling [28], where observers move through the group to locate and follow known individuals according to a predetermined list. Hence, records of agonisms are sampled in an unbiased, representative manner.

We summed the number of initiated or received agonisms involving each individual for the six month period prior to sample collection, and corrected this value for observer effort [29]. Specifically, we regressed the sum of initiated agonisms or sum of received agonisms (separately) against a measure of observer intensity, calculated as the number of focal animal samples performed on adult females in a given social group and month, divided by the total number of adult females in the group (following [29]; focal samples are concentrated periods of observation focused on a single individual and collected in randomized order for target animals

in each social group [28]). Observer intensity estimates were calculated separately for each of the 6 months spanning the period prior to sample collection, and then averaged to obtain a single value for linear regression. Finally, we extracted the residuals from the linear regression of initiated or received agonisms on observer effort and used these values in downstream analyses.

Next, we asked whether our indices of initiated or received harassment could explain the observed rank-gene expression associations, focusing specifically on genes for which this relationship was significant in males. For each gene, we were interested in estimating the indirect effect of male dominance rank on gene expression levels through the mediating variable (initiated or received agonisms). The strength of the indirect effect was estimated as the difference between the effect of rank in two models: the 'unadjusted' model that did not account for the mediator, and the effect of rank in an 'adjusted' model that incorporated the mediator, m_i. The unadjusted model, including only data from males, was as follows:

$$y_i = \mu + r_i\beta_r + a_i\beta_a + c_i\beta_c + g_i + e_i \quad (3)$$

Notations are consistent with equations 1 and 2. The adjusted model was:

$$y_i = \mu + r_i\beta_r + a_i\beta_a + c_i\beta_c + m_i\beta_m + g_i + e_i \quad (4)$$

where m_i was observer effort-corrected rates of initiated or received agonisms, respectively. To assess the significance of each indirect effect, we performed 1000 iterations of bootstrap resampling to calculate 95% confidence intervals for each mediator. We considered an indirect effect to be significant if (i) the lower bound of the 95% confidence interval did not overlap with 0 and (ii) the absolute effect size of the rank effect decreased when the mediating variable was included in the model.

Mendelian randomization analysis: selection of intermediate phenotype and instrumental variables

Mendelian randomization (MR) is a form of instrumental variable analysis that uses a genetic variant (the instrument) to test whether an intermediate phenotype (in our case, PC2 of gene expression variation) is causal to a hypothesized outcome (in our case, dominance rank) [30]. Intuitively, MR can be thought of as analogous to a randomized controlled trial, where study participants are randomly allocated to a treatment or control group. This design avoids confounding between the treatment and outcome of interest, such that causal inference is unambiguous. In MR, genotypes are assumed to be randomly distributed with respect to potential confounding variables, and also are assumed to "randomize" each study subject into higher or lower values of the intermediate phenotype under genetic control. MR has been widely used in biomedical analyses [31], for example to test for a causal relationship between HDL cholesterol and myocardial infarction [32]. More recently, genetic effects on molecular phenotypes (e.g., expression or methylation quantitative trait loci) have also been leveraged in an MR framework [33], for example to test the causal relationship between DNA methylation levels and traits related to cardiovascular disease [34].

Valid MR instruments must meet three criteria:

First, they must be robustly associated with the intermediate phenotype. In our analysis, we used projections onto PC2 of the overall gene expression data for males alone as the intermediate phenotype (n=36 unique individuals, n=70 samples). PC2 was strongly associated with male rank (rho=0.44, $p=1.26 \times 10^{-4}$), and explained 6.7% of the overall variance in male gene expression levels. Gene Ontology categories that contributed strongly to PC2 (primarily gene sets involved in the innate/TLR4-mediated immune response) are shown in Figure 5 and Dataset

S6, based on mean loading across constituent genes for each category (excluding GO categories with < 10 genes; significance was assessed by comparison to an empirical null distribution calculated from permuting PC2 loadings across all genes). To identify potential instruments associated with PC2, we refiltered our initial genotype dataset (n = 99,760 SNPs) to only include variants with a MAF>5% in the dataset of male baboons, and, in cases where a SNP was in linkage disequilibrium with one or more nearby (<10 kb) candidate SNPs, we randomly retained one of the linked SNPs. This filtering left us with 39,461 SNPs. We then used a linear mixed effects model [20] to test for an association between SNP genotype and PC2 (controlling for genetic relatedness in the sample), and retained only those that passed a 5% FDR [10] (Figure 5) (n = 51 SNPs). To avoid redundancy among our instruments, we associated each of these 51 SNPs with its closest gene and retained the SNP with the lowest p-value for each gene (n=47 SNPs). Finally, we retained only SNPs close to genes that loaded highly on PC2 (i.e., that had loading scores in the highest or lowest decile). This filtering left us with 20 candidate SNP instruments.

Second, valid MR instruments must be related to the outcome variable only through an association with the intermediate phenotype, and not through any direct effect of the instrument on the outcome. In other words, in our analysis, genotype cannot be directly associated with dominance rank. To test for this requirement, we used linear models to estimate the relationship between SNP genotype for each of the 20 candidate SNP instruments and dominance rank, controlling for PC2. We removed SNPs that showed any evidence of a relationship with dominance rank after controlling for PC2 ($p<0.05$), leaving us with 16 strong instruments (mean PVE for the correlation between a given SNP and PC2 (± SD) = 27.28 ± 6.64%). The

distribution of candidate instruments in gene bodies, coding sequences, exons, and 5' and 3' UTRs is shown in Figure S12.

Third, valid MR instruments should be unrelated to confounding factors that could bias the relationship between the intermediate phenotype and the outcome. This requirement is the most difficult to formally prove. However, we are unable to propose any plausible third variable that both predicts genotype at the 16 variants we analyzed *and* affects the relationship between gene expression and dominance rank. Genetic background/population structure is a candidate, as this population is affected by admixture between anubis and yellow baboons, and ancestry could potentially affect dominance rank. Body size is a second candidate, as larger size does predict rank, and it could conceivably influence immune cell gene expression captured by PC2. However, when we tested for associations between each of the 16 instruments and hybrid score (a measure of anubis baboon ancestry [35]) or body mass index at the time of sampling, we found no evidence for either relationship (linear model: all $p>0.05$ after FDR correction). We further tested for bias in our instruments as a result of population structure by including the following components in our linear mixed models to identify SNP-PC2 associations: (i) the top 5 PCs from a principal components analysis of the genotype data, incorporated as fixed effects, or (ii) the covariance matrix derived from the genotype data (using the 'cov' function in R) as the *K* matrix. In both cases, we saw minimal effects on the estimate of the SNP-PC2 relationship for our 16 instruments, suggesting that population structure does not impact our results (correlation between SNP-PC2 effect sizes estimated from the model in the main text versus a model that included PCs as fixed effects: $p=1.42\text{x}10^{-12}$, $r^2=0.973$, or a model that substitutes the kinship matrix with the genetic covariance matrix: $p=1.16\text{x}10^{-10}$, $r^2=0.949$).

Finally, we note that because our MR analysis specifically tests whether genotype *effects* on immune gene expression (PC2) are positively correlated with genotype *effects* on dominance rank (for cases in which genotype does not independently predict rank), it does not require dominance rank to be a stable individual characteristic. Positive correlations indicate that males who are "genetically randomized" into lower values of PC2 are more likely to be higher ranking than otherwise expected. This interpretation allows MR analysis to be applied to dynamic phenotypes (e.g., HDL and LDL cholesterol levels [32,36]).

Implementation of Mendelian randomization analysis

Using the 16 instrumental variables (SNP genotypes) that passed our filters above and were robust to potential confounding variables, we compared effect sizes estimated from the following models: (i) a linear model testing for an effect of genotype on dominance rank and (ii) a linear mixed model testing for an effect of genotype on PC2. Intuitively, if gene expression is causal to dominance rank, individuals with genotypes that predispose them toward low PC2 gene expression values should tend to also be high rank (low PC2 values are associated with high social status; Figure 1). Consequently, the effect sizes from the two sets of linear models should be positively correlated. To test this prediction, we used the MR Egger method [37] implemented in the R package 'MendelianRandomization' [38]. MR Egger accounts for horizontal pleiotropy, in which a genetic variant affects the outcome via a biological pathway other than the intermediate phenotype. However, we obtain very similar results using more traditional approaches such as the weighted median (beta=1.26; $p=4.04x10^{-16}$) and inverse-variance weighted methods (beta=1.513; $p=8.53x10^{-5}$) [38]. Further, we obtain very similar results when rerunning the MR Egger analysis after iteratively removing each one of the 16 instruments,

suggesting that outlier instruments do not impact our conclusions (beta>0 when 16/16 instruments were iteratively removed and $p<0.05$ when 15/16 instruments were iteratively removed; for the last instrument, $p=0.105$). An overview of our MR pipeline is provided in Figure S10.

We also implemented MR analyses at the single gene level, where gene expression levels for the focal gene are the intermediate variable rather than the composite measure of gene expression captured by PC2. Specifically, for each gene that was significantly associated with male rank in our data set and for which we also detected a significant *cis*-eQTL (FDR < 5%), we tested for a relationship between effect sizes estimated from the following models: (i) a linear mixed model testing for an effect of *cis* genetic variation on gene expression and (ii) a linear model testing for an effect of genotype on dominance rank. To compare the two effect sizes, we used the ratio of coefficients method, also known as the Wald method, as described in [39]). In this analysis, our instruments are consequently eQTL, rather than QTL for a composite measure of rank-associated gene expression (i.e., PC2). We were interested in implementing this single gene approach both to understand the robustness of our conclusions to different methodologies, and also to compare against a "control" data set in which the study design precluded gene expression effects on dominance rank. Specifically, we implemented the same MR pipeline using genotype and gene expression data from female rhesus macaques [24], where dominance rank was experimentally manipulated and must therefore be causal to gene expression (we initially implemented the MR pipeline described in the main text for this data set, but found few strong instruments for PC2 of gene expression variation). As expected, we found no evidence for a relationship between the effect sizes estimated from models (i) and (ii) for the female macaques, where dominance rank was experimentally imposed, but we do observe a significant relationship

between the two effect sizes for many rank-associated genes in male baboons. An overview of the single gene pipeline, as well as results for both the baboon and macaque data sets, are presented in Figure S11.

Figure S1. *Ex vivo* stimulation with lipopolysaccharide (LPS) induces changes in the abundance of immune signaling molecules. Comparison of levels of serum cytokines and immune defense molecules in NULL and LPS samples, for all cytokines that met our filtering criteria (see methods). P-values (uncorrected) represent the effect of treatment controlling for age, sex, and batch effects in a linear model framework.

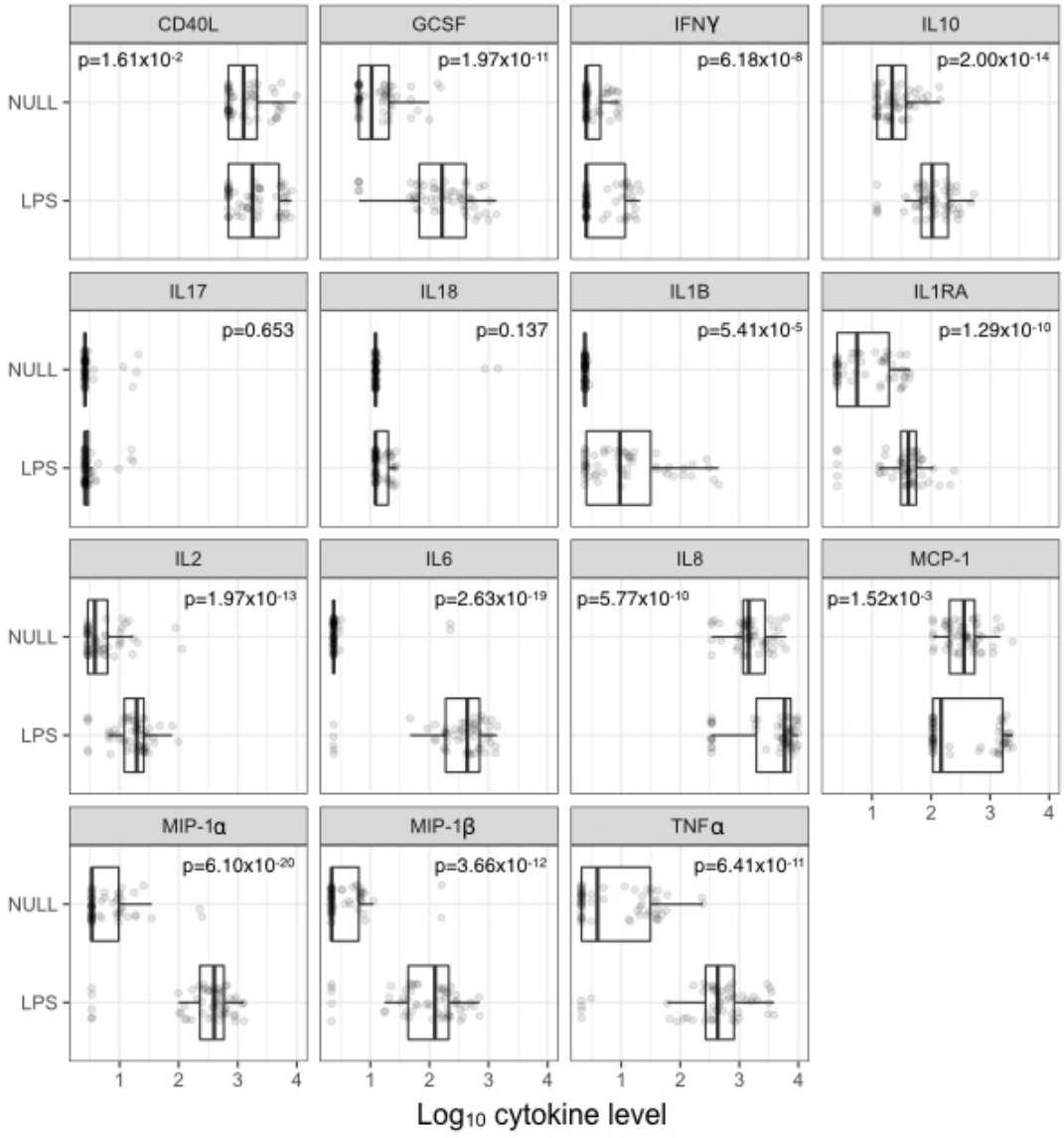


Figure S2. Overview of cell phenotyping strategy. Strategy for identifying populations of five different cell types within each PBMC sample. We gated on live cells and phenotyped these cell populations using the cell surface markers detailed in the SI Materials and Methods. All analyses were performed using FlowJo (FlowJo, LLC, Ashland, OR).

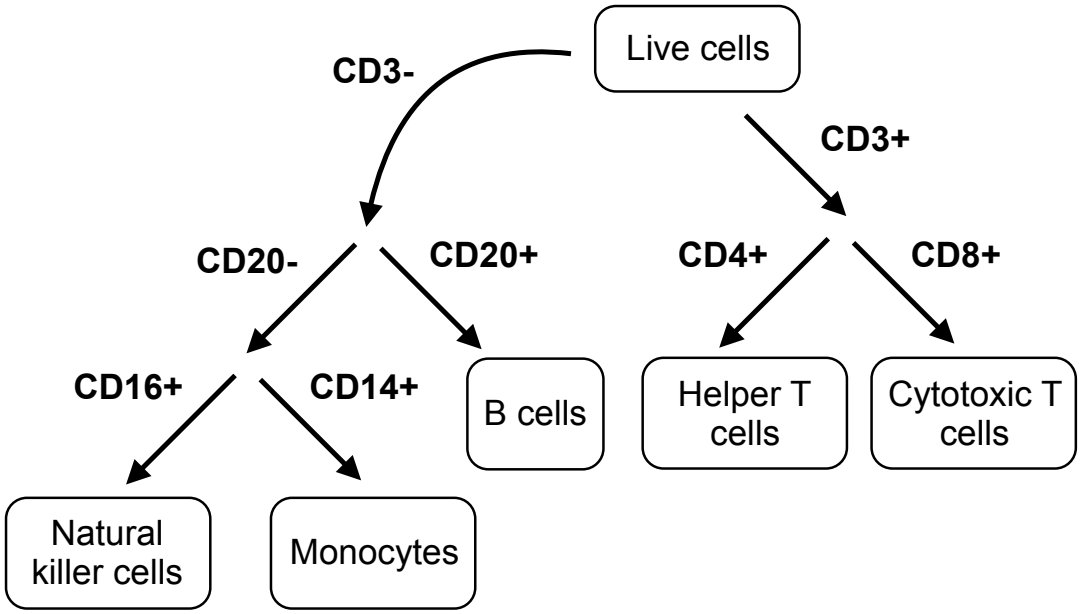


Figure S3. Association between dominance rank and the proportions of five white blood cell populations. Each plot shows the relationship between dominance rank (stratified by sex) and the proportion of a given cell population. P-values represent the effect of dominance rank (nested within sex) controlling for age (also nested within sex) in a linear model framework.

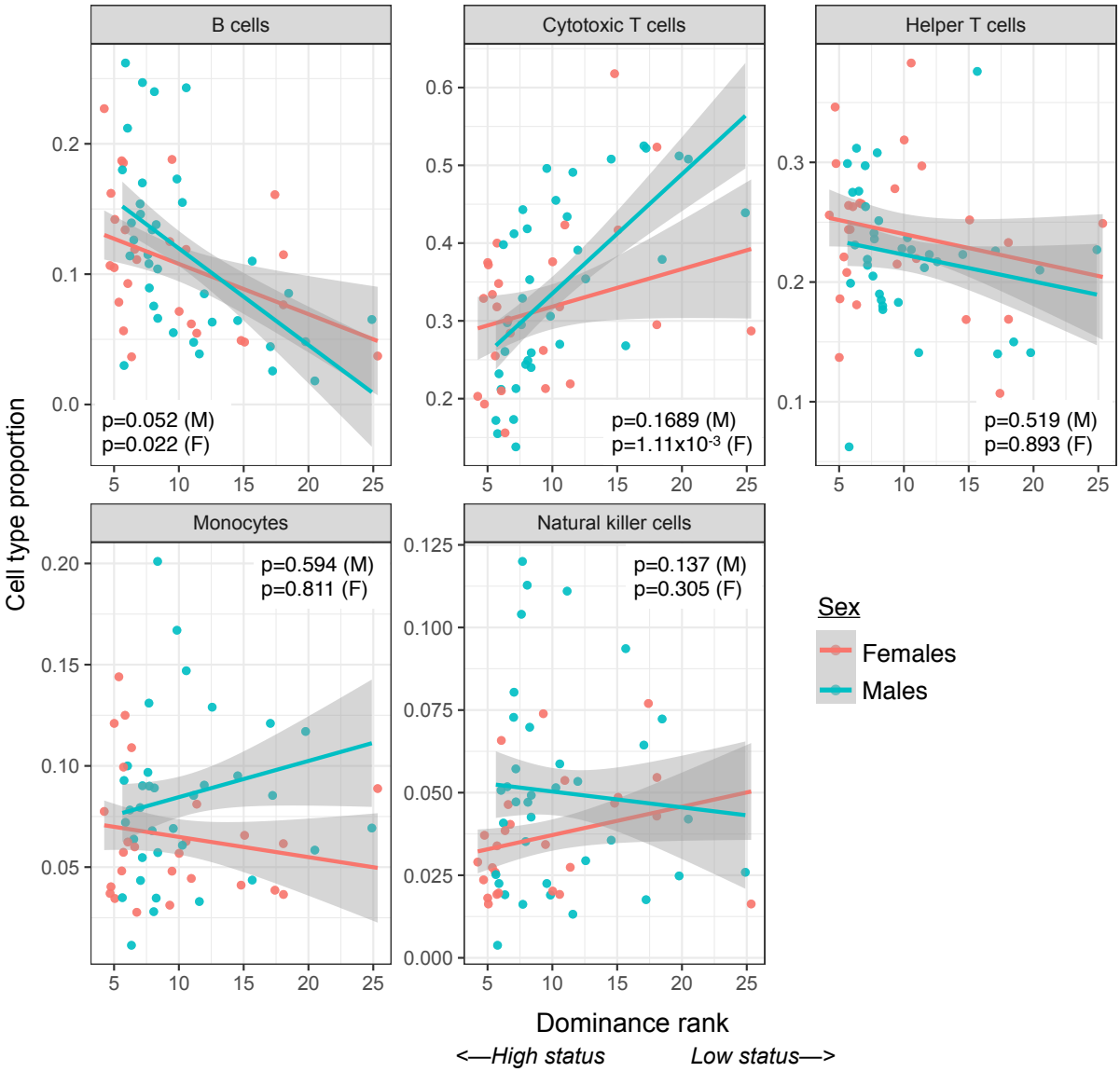


Figure S4. Gene ontology (GO) term enrichment for genes that are significantly (A) up-regulated or (B) down-regulated in the LPS condition in male and female baboons (FDR<1%). Each significant GO term is represented by a node, and related GO terms are colored similarly and connected by edges.

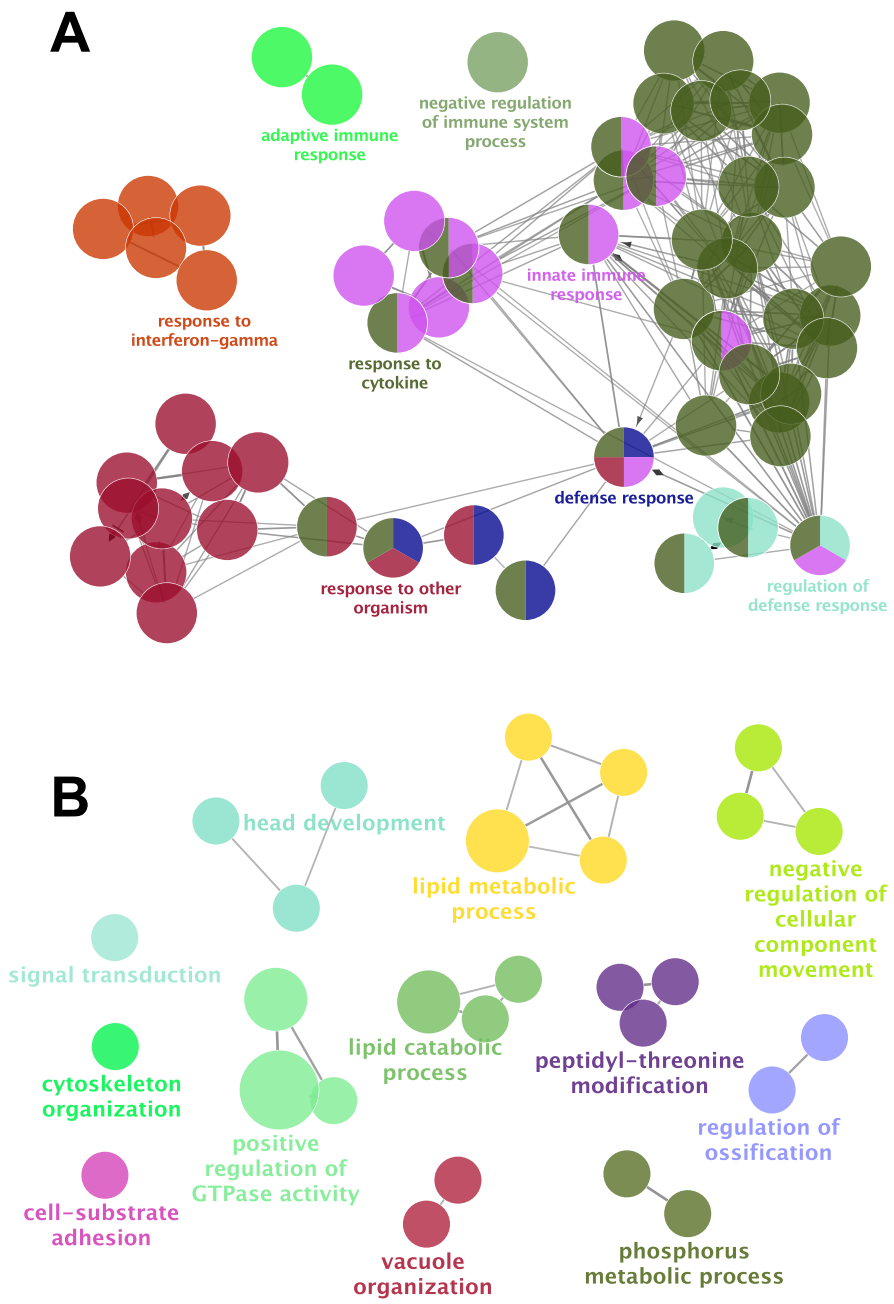


Figure S5. Rank-gene expression relationships in males and females are largely distinct. (A) Comparison of effect sizes for rank effects estimated in males versus females. Points are colored by whether the focal gene was significantly rank-associated in neither sex, one sex, or both sexes (5% FDR). (B) QQ-plot comparing the distribution of p-values associated with the rank effect estimated in females versus males. Comparison is against the expected null distribution (a uniform distribution). In both A and B, p-values were derived from a linear mixed effects model in which rank was nested within sex.

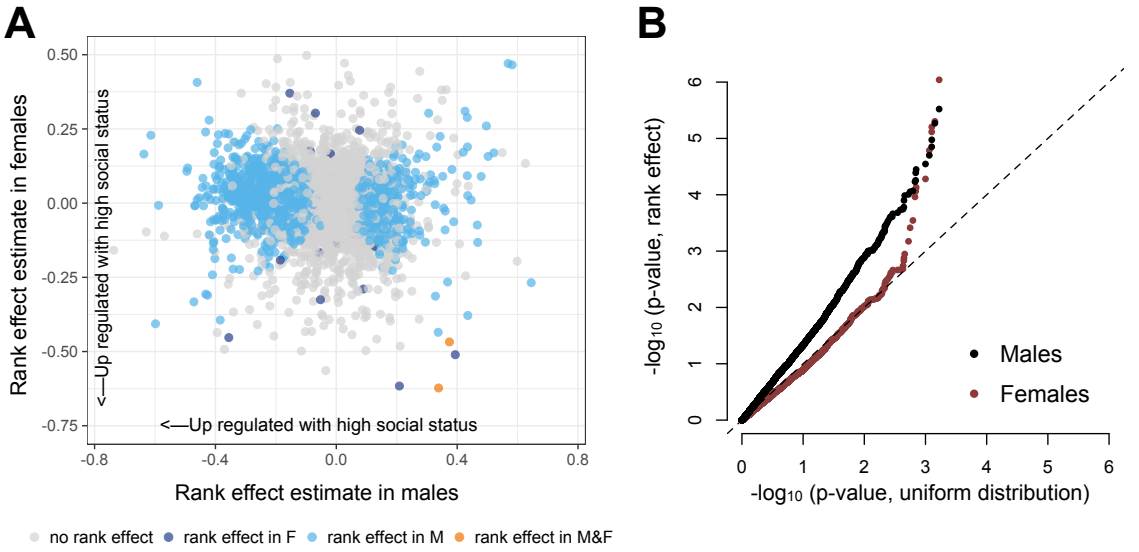


Figure S6. Sample size does not completely explain the difference in the number of significant rank-associated genes detected in males and females. Distribution of the number of rank-associated genes (FDR<5%) detected in (A) males and (B) females, as well as (C) the difference in the number of rank-associated genes found in each sex (number of rank-associated genes in males - number of rank-associated genes in females), after randomly subsampling our dataset 100 times so that the number of samples derived from each sex were matched. Red lines indicate values for the data set described in the main text. Across all subsamples, we consistently found far more rank-associated genes in males than in females (an average of 1387 ± 819.09 s.d. more genes were associated with rank in males compared to females).

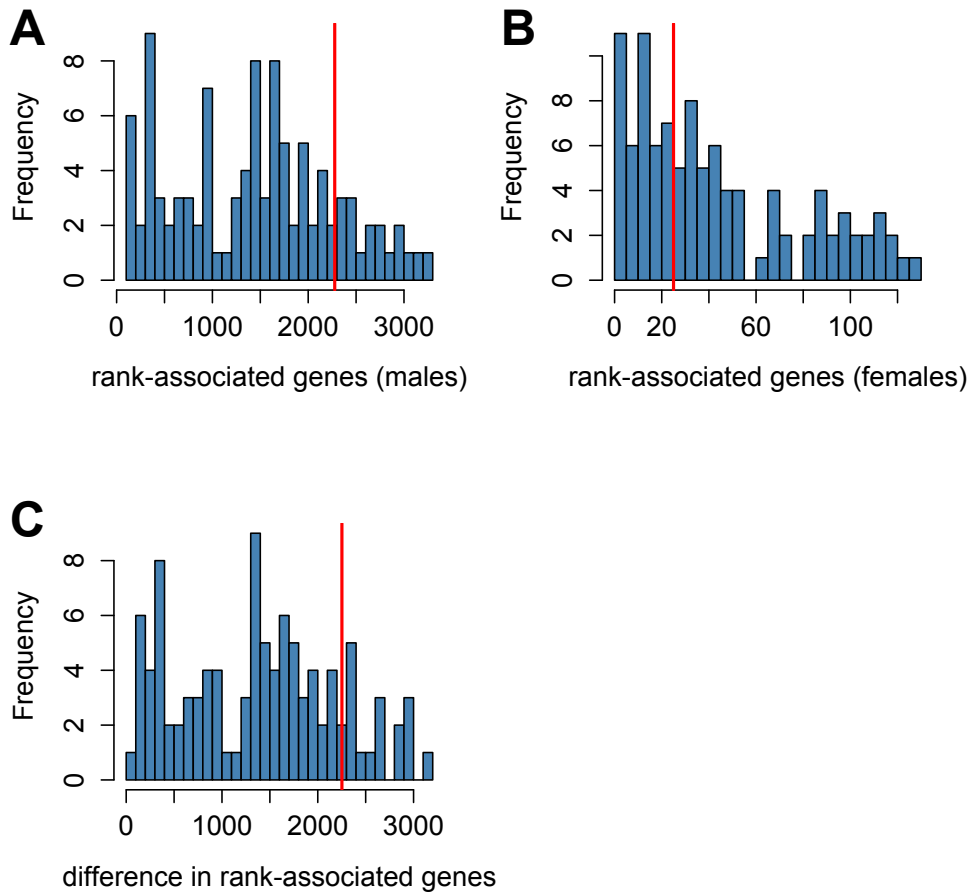


Figure S7. Social status has weak effects on the strength of the response to LPS stimulation. (A) QQ-plot comparing the distribution of p-values for a rank x condition interaction effect estimated from a linear mixed effects model (for males and females separately; all models controlled for age, dominance rank, and condition as fixed effects) against the expected uniform distribution. 5 and 0 genes exhibit a significant (FDR<5%) rank x condition interaction in males and females, respectively, although the QQ-plot for males suggests that detection of interaction effects is constrained by power. (B) Magnitude of the rank effect estimated in males in the LPS and NULL conditions (rho=0.619, $p<10^{-10}$). Effect sizes are derived from a linear mixed effects model using male data only, in which rank effects were nested within condition. Genes with no rank effects in either condition (FDR>20% in both LPS and NULL conditions), rank effects in the LPS or NULL condition only (FDR>20% in one condition and <5% in the other), or rank effects in both conditions (FDR<5% in both conditions) are highlighted as described in the legend.

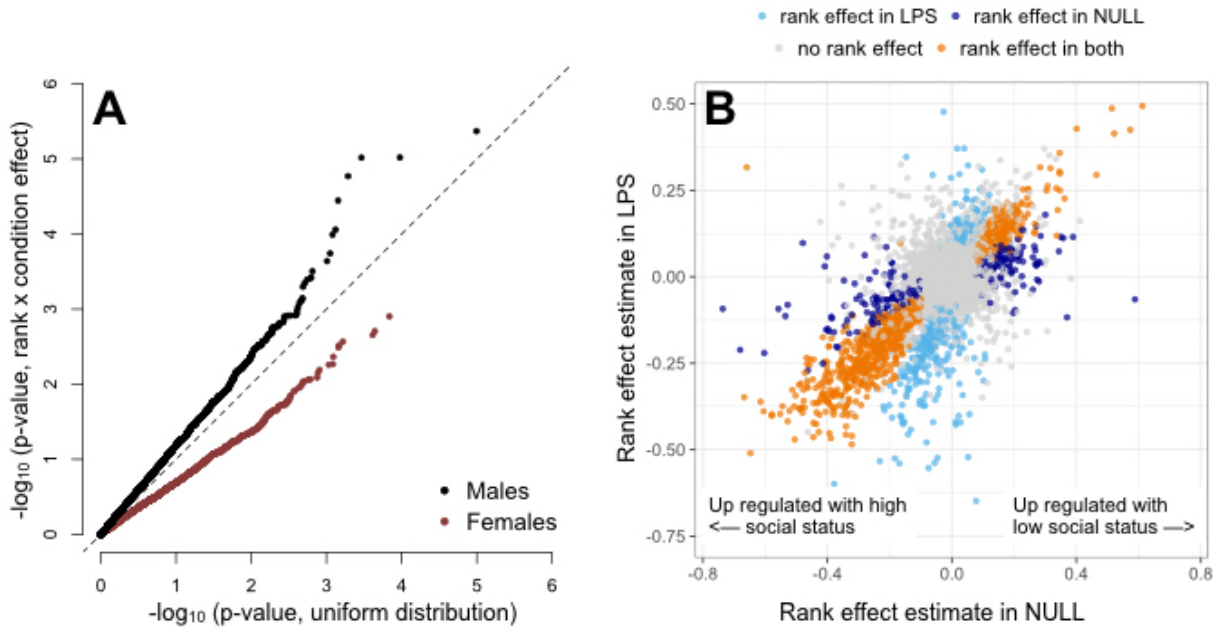


Figure S8. High status males exhibit higher expression of pro-inflammatory genes compared to low status males, in both NULL and LPS condition samples. Each point represents the median expression level for a given sample, across all genes included in the following GO annotations: 'regulation of IL6 production', 'toll-like receptor signaling pathway', and 'regulation of inflammatory response' (all three categories are enriched among genes significantly upregulated in high status males, p<10^-6). Lines connect samples collected from the same male, and are colored by quartiles of dominance rank.

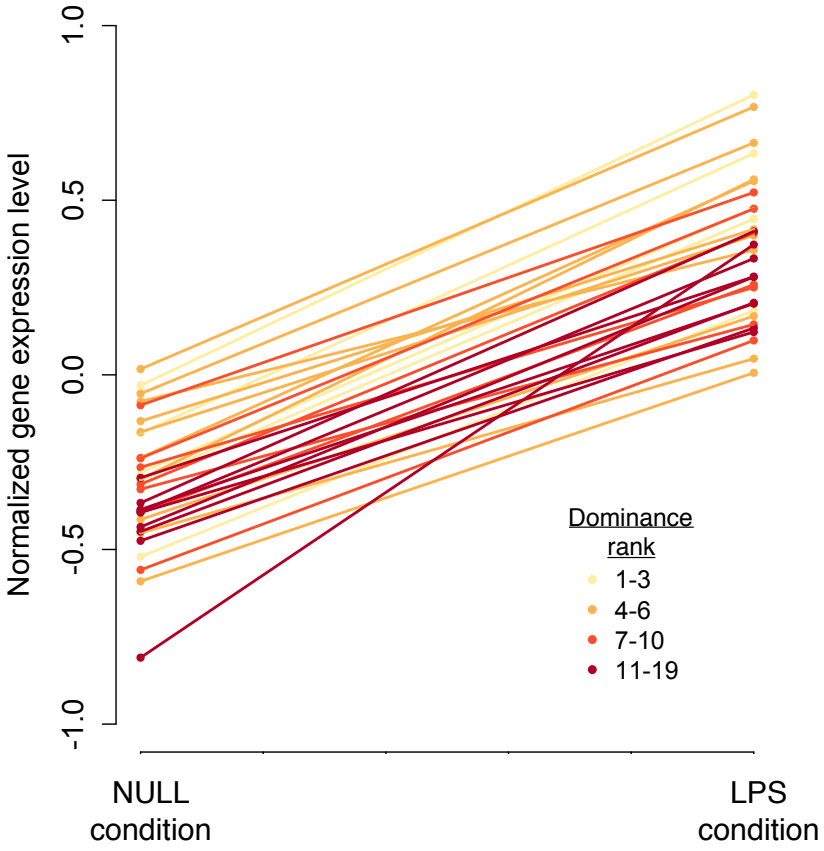


Figure S9. Genes up-regulated in low-ranking captive female macaques are up-regulated in high-ranking wild male baboons. X-axis: effect of rank on gene expression reported in [24], for leukocytes incubated in the presence (A; LPS condition, as shown in Figure 3B and repeated here for comparison to the NULL) or absence (B: NULL condition) of lipopolysaccharide. Effect sizes were estimated from linear mixed effects models, in which dominance rank was nested within condition. Y-axis: parallel results from wild male baboons. Effect sizes and p-values are from Spearman's rank correlations, and sign-reversed for the macaque data set for easier comparison to baboons.

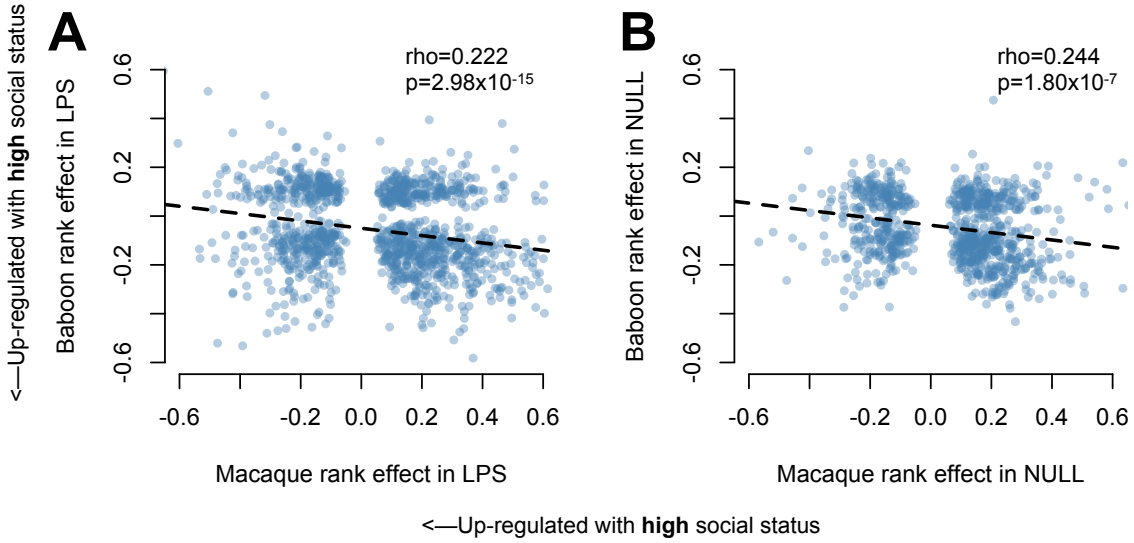


Figure S10. Overview of filtering procedures for Mendelian randomization analyses.

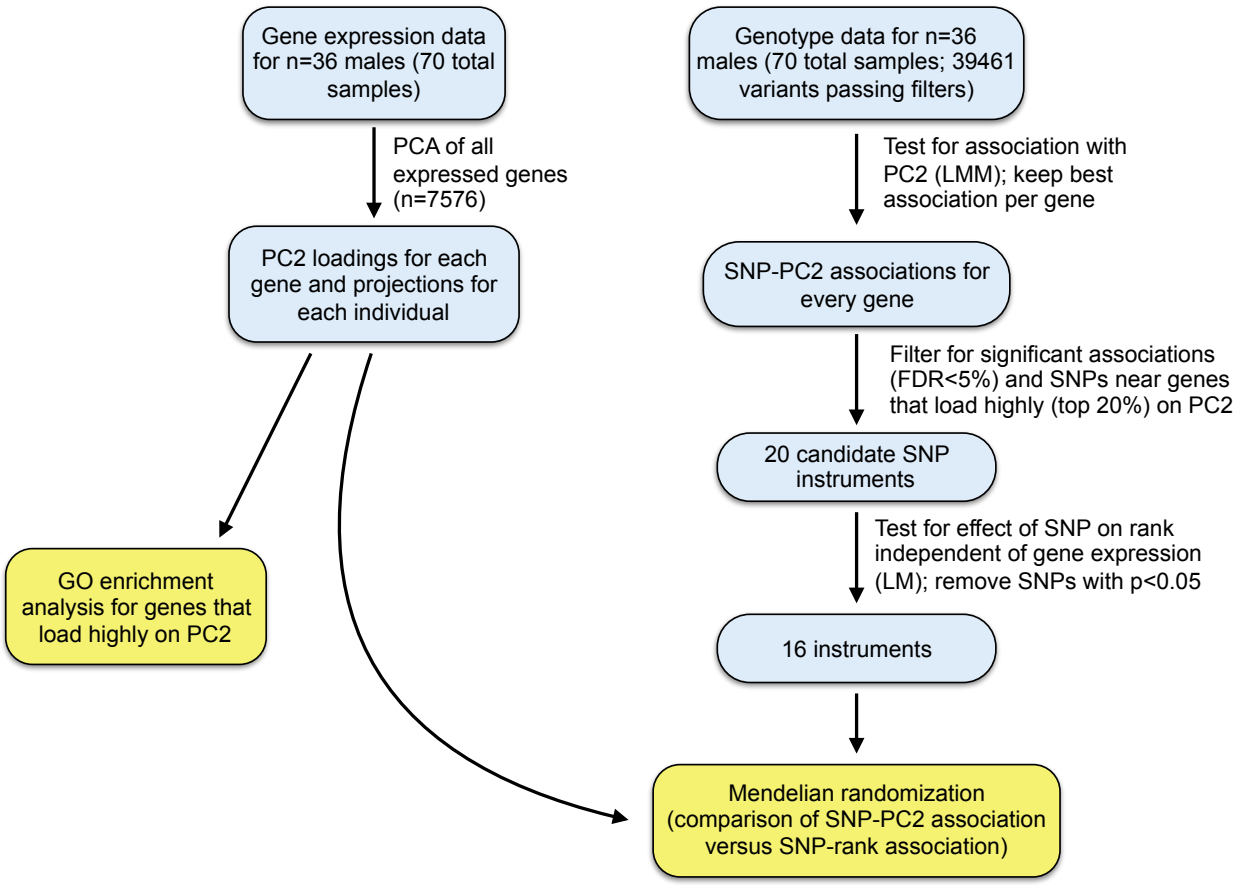


Figure S11. Overview of methods and results for single gene Mendelian randomization analyses. (A) Methodological approach and filtering procedures. The same approach was applied to both the female rhesus macaque and male baboon data sets. Distribution of p-values from a Wald test performed using each of the instruments passing filters in the (B) macaque and (C) baboon data set.

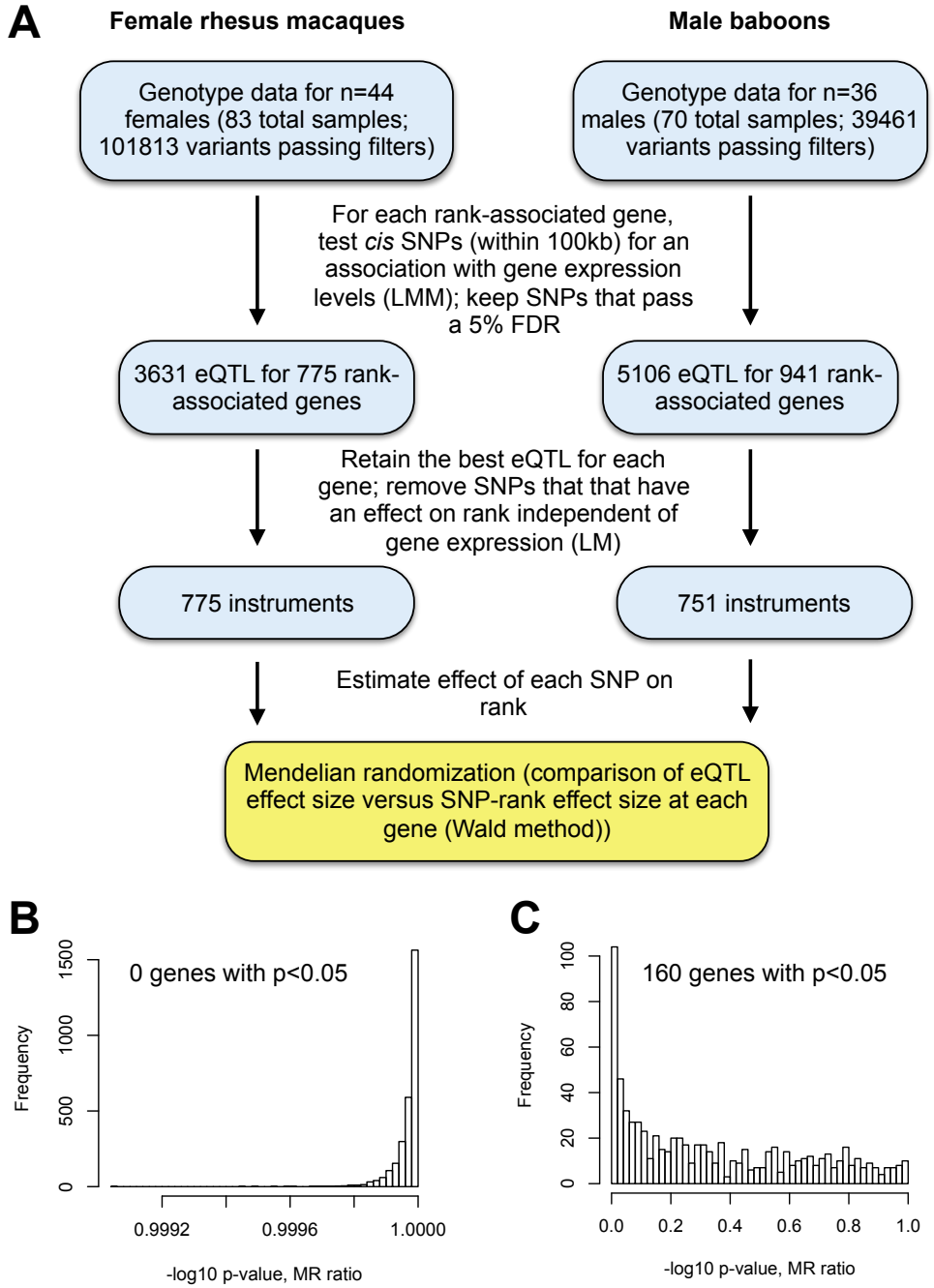
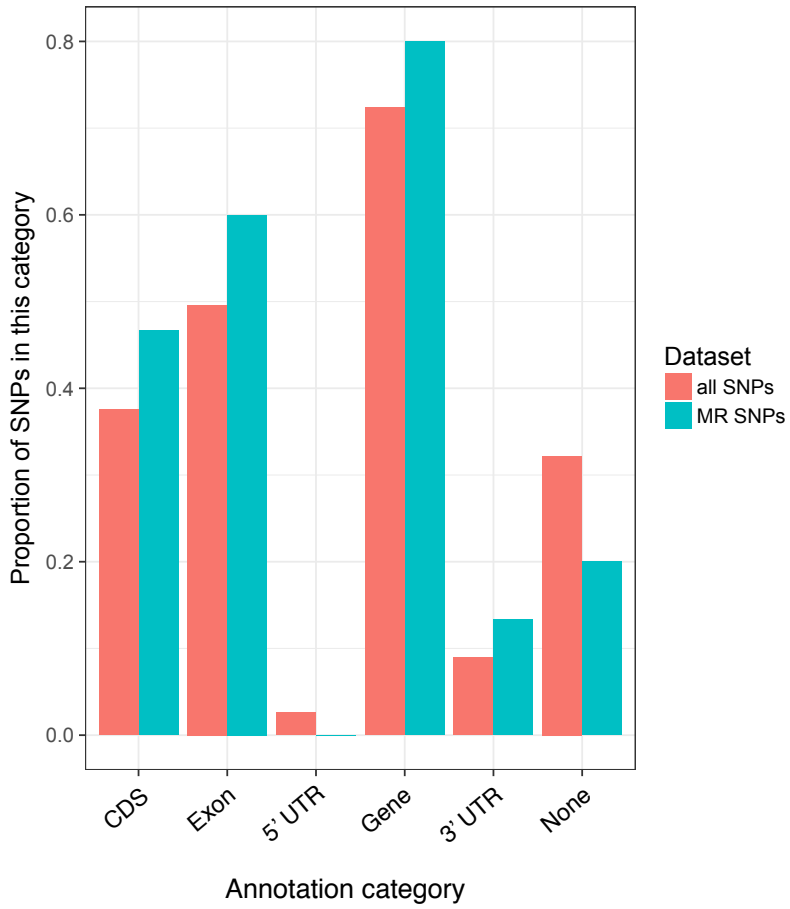


Figure S12. MR instruments are more likely to occur in genes and regulatory regions. Barplots show the proportion of SNPs falling into each annotation category, for the 16 MR instruments and for all 39,461 SNPs that were considered as candidate instruments. Annotations were taken from the Panu2 GTF file (version 0.90), downloaded from Ensembl. 'CDS' refers to the coding portion of a given gene, and 'gene' is defined as all sequence between the 5' and 3' UTR (and therefore includes all categories except 'none'). SNPs that did not overlap with any annotated regions from the GTF file were assigned to the annotation category 'none'.



Supplementary References

1. Wall JD, Schlebusch SA, Alberts SC, Cox LA, Snyder-Mackler N, Nevonen KA, et al. Genomewide ancestry and divergence patterns from low-coverage sequencing data reveal a complex history of admixture in wild baboons. Mol Ecol. 2016;25: 3469–3483. doi:10.1111/mec.13684
2. Alberts SC, Altmann J. The Amboseli Baboon Research Project: 40 years of continuity and change. In: Kappeler P, Watts DP, editors. Long-Term Field Studies of Primates. New York: Springer; 2012. pp. 261–288.
3. Altmann J, Altmann S, Hausfater G. Physical maturation and age estimates of yellow baboons, Papio cynocephalus, in Amboseli National Park, Kenya. Am J Primatol. 1981;1: 389–399. doi:10.1002/ajp.1350010404
4. Hausfater G. Dominance and reproduction in baboons (Papio cynocephalus): A quantitative analysis. University of Chicago. 1974.
5. Tung J, Zhou X, Alberts SC, Stephens M, Gilad Y. The genetic architecture of gene expression levels in wild baboons. eLife. 2015;4: 1–22. doi:10.7554/eLife.04729
6. Tung J, Primus A, Bouley AJ, Severson TF, Alberts SC, Wray G. Evolution of a malaria resistance gene in wild primates. Nature. 2009;460: 388–91. doi:10.1038/nature08149
7. Tung J, Akinyi MY, Mutura S, Altmann J, Wray G, Alberts SC. Allele-specific gene expression in a wild nonhuman primate population. Mol Ecol. 2011;20: 725–39. doi:10.1111/j.1365-294X.2010.04970.x
8. Altmann J, Alberts S, Haines S, Dubach J, Muruthi PM, Coote T, et al. Behavior predicts genetic structure in a wild primate group. Proc Natl Acad Sci. 1996;93: 5797–5801.
9. Pinheiro J, Bates D, DebRoy S, Sarka D. nlme: Linear and Nonlinear Mixed Effects

Models. R Packag version 31-131. 2017;

10. Dabney A, Storey J. qvalue: Q-value estimation for false discovery rate control. R package version 1.43.0. 2015.

11. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci. 2003;100: 9440–5. doi:10.1073/pnas.1530509100

12. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170

13. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 1–7. doi:doi: 10.1093/bioinformatics/bts635

14. Anders S. HTSeq: Analysing high-throughput sequencing data with Python [Internet]. 2011. Available: http://www-huber.embl.de/users/anders/HTSeq/

15. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44: D733–D745. doi:10.1093/nar/gkv1189

16. Law C, Chen Y, Shi W, Smyth G. Voom! Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. Melbourne, Australia; 2014;15: R29.

17. Danecek P, Auton A, Abecasis G, Albers C a, Banks E, DePristo M a, et al. The variant call format and VCF tools. Bioinformatics. 2011;27: 2156–8. doi:10.1093/bioinformatics/btr330

18. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet. The American Society of Human Genetics; 2008;84: 210–223.

552 doi:10.1016/j.ajhg.2009.01.005
553 19. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust
554 relationship inference in genome-wide association studies. Bioinformatics. 2010;26:
555 2867–2873. doi:10.1093/bioinformatics/btq559
556 20. Akdemir D, Godfrey OU. EMMREML: Fitting mixed models with known covariance
557 structures. 2015.
558 21. Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K, et al. Heritability and
559 genomics of gene expression in peripheral blood. Nat Genet. Nature Publishing Group;
560 2014;46: 430–437. doi:10.1038/ng.2951
561 22. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-
562 model method for association mapping that accounts for multiple levels of relatedness.
563 Nat Genet. 2006;38: 203–8. doi:10.1038/ng1702
564 23. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association
565 studies. Nat Genet. 2012;44: 821–4. doi:10.1038/ng.2310
566 24. Snyder-Mackler N, Sanz J, Kohn JN, Brinkworth JF, Morrow S, Shaver AO, et al. Social
567 status alters immune regulation and response to infection. Science. 2016;354: 1041–1045.
568 doi:10.1126/science.aah3580
569 25. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO:
570 A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway
571 annotation networks. Bioinformatics. 2009;25: 1091–1093.
572 doi:10.1093/bioinformatics/btp101
573 26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful
574 approach to multiple testing. J R Stat Soc. 1995;57: 289–300.

27. Ramsey SA, Klemm SL, Zak DE, Kennedy KA, Thorsson V, Li B, et al. Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. PLoS Comput Biol. 2008;4. doi:10.1371/journal.pcbi.1000021

28. Altmann J. Observational Study of Behavior: Sampling Methods. Behaviour. 1974;49: 227–267.

29. Archie EA, Tung J, Clark M, Altmann J, Alberts SC. Social affiliation matters: both same-sex and opposite-sex relationships predict survival in wild female baboons. Proc R Soc London B Biol Sci. 2014;281: 20141261. doi:10.1098/rspb.2014.1261

30. Smith GD, Ebrahim S. "Mendelian randomization": Can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32: 1–22. doi:10.1093/ije/dyg070

31. Smith GD, Hemani G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23: 89–98. doi:10.1093/hmg/ddu328

32. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, et al. Plasma HDL cholesterol and risk of myocardial infarction: A mendelian randomisation study. Lancet. 2012;380: 572–580. doi:10.1016/S0140-6736(12)60312-2

33. Richmond RC, Hemani G, Tilling K, Davey Smith G, Relton CL. Challenges and novel approaches for investigating molecular mediation. Hum Mol Genet. 2016;25: R149–R156. doi:10.1093/hmg/ddw197

34. Richardson TG, Zheng J, Davey Smith G, Timpson NJ, Gaunt TR, Relton CL, et al. Mendelian Randomization Analysis Identifies CpG Sites as Putative Mediators for Genetic Influences on Cardiovascular Disease Risk. Am J Hum Genet. ElsevierCompany.; 2017;101: 590–602. doi:10.1016/j.ajhg.2017.09.003

35. Tung J, Charpentier MJE, Garfield D a, Altmann J, Alberts SC. Genetic evidence reveals temporal change in hybridization patterns in a wild baboon population. Mol Ecol. 2008;17: 1998–2011. doi:10.1111/j.1365-294X.2008.03723.x

36. Burgess S, Harshfield E. Mendelian randomization to assess causal effects of blood lipids on coronary heart disease: Lessons from the past and applications to the future. Curr Opin Endocrinol Diabetes Obes. 2016;23: 124–130. doi:10.1097/MED.0000000000000230

37. Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44: 512–525. doi:10.1093/ije/dyv080

38. Yavorska OO, Burgess S. MendelianRandomization: An R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol. 2017;46: 1734–1739. doi:10.1093/ije/dyx034

39. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Stat Methods Med Res. 2017;26: 2333–2355. doi:10.1177/0962280215597579