

Social environment is associated with gene regulatory variation in the rhesus macaque immune system

Jenny Tung^{a,1,2}, Luis B. Barreiro^{a,3}, Zachary P. Johnson^b, Kasper D. Hansen^c, Vasiliki Michopoulos^b, Donna Toufexis^{b,d}, Katelyn Michelini^a, Mark E. Wilson^b, and Yoav Gilad^{a,1}

^aDepartment of Human Genetics, University of Chicago, Chicago, IL 60637; ^bYerkes National Primate Research Center, Emory University, Atlanta, GA 30322;

^cDepartment of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21202; and ^dDepartment of Psychology, University of Vermont, Burlington, VT 05405

Edited by Gene E. Robinson, University of Illinois at Urbana–Champaign, Urbana, IL, and approved March 6, 2012 (received for review February 15, 2012)

Variation in the social environment is a fundamental component of many vertebrate societies. In humans and other primates, adverse social environments often translate into lasting physiological costs. The biological mechanisms associated with these effects are therefore of great interest, both for understanding the evolutionary impacts of social behavior and in the context of human health. However, large gaps remain in our understanding of the mechanisms that mediate these effects at the molecular level. Here we addressed these questions by leveraging the power of an experimental system that consisted of 10 social groups of female macaques, in which each individual's social status (i.e., dominance rank) could be experimentally controlled. Using this paradigm, we show that dominance rank results in a widespread, yet plastic, imprint on gene regulation, such that peripheral blood mononuclear cell gene expression data alone predict social status with 80% accuracy. We investigated the mechanistic basis of these effects using cell type-specific gene expression profiling and glucocorticoid resistance assays, which together contributed to rank effects on gene expression levels for 694 (70%) of the 987 rank-related genes. We also explored the possible contribution of DNA methylation levels to these effects, and identified global associations between dominance rank and methylation profiles that suggest epigenetic flexibility in response to status-related behavioral cues. Together, these results illuminate the importance of the molecular response to social conditions, particularly in the immune system, and demonstrate a key role for gene regulation in linking the social environment to individual physiology.

inflammation | social gradient | differential gene expression | sociogenomics

The social organization of many group-living mammals, including many primates, is marked by readily recognizable differences in the social environments experienced by individual group members. The causes and consequences of variation in the social environment have been of long-standing interest in behavioral ecology and evolution, and strong evidence indicates the importance of the social environment in human health as well. In particular, a large body of work in humans and nonhuman primates indicates that adverse social conditions can have important consequences, both for disease susceptibility in the short term (1–3) and for evolutionarily important parameters such as fertility and survival over the long term (4–7). Characterizing the mechanisms underlying these effects is therefore a priority not only in the study of evolution and behavior but also for understanding the protective and pathological effects of the social environment in our own lives.

Social status is one of the most important predictors of the quality of an individual's social environment. Consequently, social status has been intensively studied in nonhuman primates, including as a model for the effects of social stress and socioeconomic status (SES) in humans (reviewed in ref. 8). Social status in nonhuman primates is encoded by dominance rank, which defines which individuals yield to other individuals during competitive encounters. In settings in which hierarchies are strongly enforced

or subordinates have little social support, low dominance rank can lead to chronic stress, immune compromise, and reproductive dysregulation (3, 9). In socially housed captive female rhesus macaques (*Macaca mulatta*), for instance, the physiological effects of dominance rank include changes in glucocorticoid (GC) and sex steroid hormone regulation (10), serotonergic and dopaminergic signaling (11), and lymphocyte counts and proliferation rates (12). Interestingly, many of these rank-associated effects are detectable even in the absence of rank-related asymmetries in access to resources, suggesting that the stress of social subordination alone can trigger a physiological response. This relationship echoes the effects of social stress and SES in humans, in which observed “social gradients” in disease risk remain in large part unexplained by resource access alone (8, 13, 14).

Much remains unresolved, however, about the physical intermediates that link the social environment with immunological and physiological changes, especially on the molecular level. In particular, we know little about how social status impacts gene regulation, either in primates or in mammals more generally, although several lines of evidence suggest that these effects may be important. First, the potential for social regulation of gene expression is supported by correlations between social integration, early-life SES, and gene expression variation in humans (15–17). Second, gene–environment interactions identified in humans and nonhuman primates frequently involve allelic variants that act via altering gene expression levels (e.g., 18, 19). Finally, social hierarchies are known to influence gene expression in other organisms. Dominance rank ascendancy in social cichlids, for example, is associated with strong induction of the transcription factor *egr1* and its downstream target, gonadotropin-releasing hormone (20). Similarly, in honey bees, genome-wide gene expression profiles strongly differentiate queen bees from worker castes, sterile workers from reproductive workers, and active foragers from hive workers (21, 22). Although the nature of social hierarchies differs between insects, fish, and primates, these lines of evidence suggest that social status might also directly influence gene regulation in primates. Further, the

Author contributions: M.E.W. and Y.G. designed research; J.T., L.B.B., Z.P.J., V.M., and K.M. performed research; J.T., L.B.B., Z.P.J., K.D.H., D.T., M.E.W., and Y.G. contributed new reagents/analytic tools; J.T., L.B.B., and K.D.H. analyzed data; and J.T., L.B.B., Z.P.J., K.D.H., V.M., D.T., M.E.W., and Y.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequence (bisulfite sequencing for DNA methylation) and gene expression datasets reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (SuperSeries no. GSE34129).

¹To whom correspondence may be addressed. E-mail: jt5@duke.edu or gilad@uchicago.edu.

²Present address: Department of Evolutionary Anthropology and Duke Population Research Institute, Duke University, Durham, NC 27708.

³Present address: Department of Pediatrics, Sainte-Justine Hospital Research Centre, University of Montreal, Montreal, QC, Canada H3T 1C5.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1202734109/-DCSupplemental.

strong ties between social status and disease risk in humans suggest that such effects might be particularly pronounced in the primate immune system.

To test this possibility, we turned to social groups of female rhesus macaques, a well-established model for the physiological effects of social stress and social status. Specifically, we focused on experimentally constructed social hierarchies, in which dominance rank assignments were imposed on each female by manipulating her order of introduction into a new social group [earlier introduction predicted higher rank (23); *Results*]. Using this powerful animal model, we were able to test for an association between rank and gene regulation in social groups in which dominance ranks were experimentally assigned, and in which other known environmental influences were carefully controlled. Together, our findings indicate that dominance rank in female rhesus macaques explains substantial variation in gene expression levels in peripheral blood mononuclear cells, an important component of immune surveillance and defense.

Results

Associations Between Dominance Rank and Gene Expression Levels.

To investigate whether and to what degree dominance rank influences gene regulation, we profiled gene expression levels in peripheral blood mononuclear cells (PBMCs) from 49 female rhesus macaques in 10 replicate social groups, using the Illumina HT-12 BeadChip. Each social group was composed of five females (we sampled only four females in one group), resulting in five dominance rank values for each group. Dominance ranks were strongly correlated with introduction order (Spearman's $\rho = 0.70$, $P = 2.62 \times 10^{-8}$, $\rho = 0.80$ when excluding a group in which multiple individuals recently changed ranks; these individuals were subjects of the plasticity analysis reported below) but not with female age ($P = 0.34$, $\rho = -0.14$), parity ($P = 0.12$, $\rho = -0.23$), time since ovariectomy ($P = 0.16$, $\rho = 0.20$), or time since removal from original breeding colony ($P = 0.52$, $\rho = 0.09$). Differences in rank were reflected in different social experiences, including rates of received aggression (Fig. 1A). After filtering for probes that mapped uniquely to the macaque genome and for genes that were detectably expressed in our sample, we considered the normalized expression levels of 6,097 genes in each female (listed in [Dataset S1](#)). We identified a strong global signature of dominance rank in the gene expression data. In particular, the first principal component of the gene expression data, which accounted for 20.2% of variance in the dataset, was significantly correlated with dominance rank ($R^2 = 0.10$, $P < 0.03$; Fig. 1B).

We then proceeded to analyze the contribution of dominance rank to variation in expression levels for each gene. To do so, we used a linear mixed-effects model in which residual variation in gene expression levels, after taking into account differences across social groups, was treated as the response variable. Dominance rank was incorporated as a fixed effect, and the significance of the dominance rank–gene expression relationship was assessed based on the strength of this effect. We identified an association between interindividual gene expression variation and dominance rank for 987 genes (16.2% of the 6,097 genes we considered; false discovery rate = 10%; Fig. 1C and [Dataset S1](#)). Within the set of 987 rank-associated genes, 535 genes (8.8% of all genes we considered) were more highly expressed in high-ranking individuals, and 452 genes (7.4%) were more highly expressed in low-ranking individuals.

To explore the biological functions associated with these genes, we performed categorical enrichment analysis based on publicly available gene set annotations (24) ([Dataset S2](#)). We identified little evidence for enrichment of specific functional categories in the rank-associated gene set as a whole, although the most significant category, interleukin signaling ($P < 0.02$; proportion of false discoveries at this P value threshold, $q = 0.30$), suggested a role for immune function. Consistent with this possibility, when conditioning on the direction of the rank effect on gene regulation, we observed enrichments for interleukin

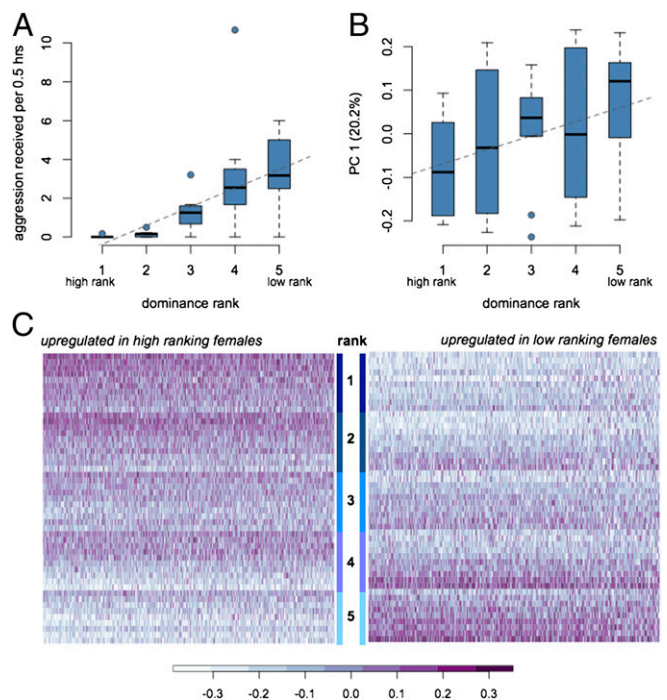


Fig. 1. Parallel effects of dominance rank on social environment and gene expression levels. (A) Low-ranking individuals experience more aggression from group mates than high-ranking individuals ($P < 10^{-6}$, $R^2 = 0.42$, $n = 49$). (B) This social experience is mirrored by gene expression profiles that vary according to rank. Principal component (PC)1 explains 20.2% of overall variance in gene expression and is correlated with rank ($P = 0.03$, $R^2 = 0.10$, $n = 49$). (C) Heatmaps of \log_2 -transformed gene expression levels for rank-associated genes. Values are shown after controlling for differences in means among social groups; 0 roughly corresponds to mean expression levels.

signaling ($P = 0.01$, $q = 0.05$), T-cell activation ($P = 0.01$, $q = 0.05$), and chemokine and cytokine inflammation ($P = 0.02$, $q = 0.07$) among genes more highly expressed in low-ranking individuals. Examples of such genes (Fig. 2) include *PTGS2* (a proinflammatory signaling molecule that is negatively regulated by immunosuppressive glucocorticoids), *IL8RB* (a receptor for the strongly proinflammatory cytokine *IL8*, which is associated with neutrophil migration into injured tissue), and *NFATC1* (which is associated with the transcriptional response to T-cell stimulation).

Finally, we used clustering analysis [modulated modularity clustering (25)] to investigate whether modules of rank-associated genes (>9 genes) with highly correlated expression patterns reflected coherent biological functions. Consistent with our previous results, we found that the largest module (112 genes; mean $r = 0.58$) was enriched for immune-related processes ($P = 0.002$, $q = 0.08$). Additionally, the most highly coregulated module (13 genes; mean $r = 0.63$) was enriched for specific immune-related gene subsets, including response to IFN- γ ($P = 7.50 \times 10^{-6}$, $q = 9.0 \times 10^{-5}$) and macrophage activation ($P = 6.31 \times 10^{-4}$, $q = 3.80 \times 10^{-3}$). Thus, rank is associated with functionally coordinated changes in immune gene expression.

Predictive Relationships Between Gene Expression and Rank. The strong effect of dominance rank on gene expression levels suggested that gene expression data alone might be sufficient to predict rank attributes. To test this possibility, we performed 1,000 iterations of a leave-k-out procedure. In each iteration, we fit a support vector machine (SVM)-based model (26) using a training set composed of data from 39 randomly chosen individuals (79.6% of the dataset). We used this model to learn how gene expression data from these females were associated with

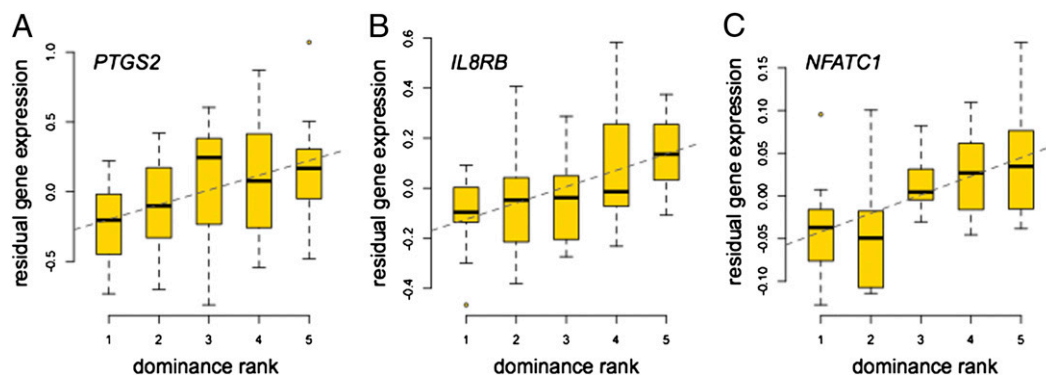


Fig. 2. Rank-gene expression associations among inflammation-related immune genes. Low-ranking females tend to overexpress inflammation-related genes: (A) *PTGS2* ($P = 0.004$); (B) *IL8RB* ($P = 0.003$); and (C) *NFATC1* ($P < 10^{-3}$).

their relative position in the social hierarchy, specifically to an individual's classification as high-ranking (rank class A, including ranks 1 and 2), middle-ranking (rank class B, including rank 3), or low-ranking (rank class C, including ranks 4 and 5) (*SI Appendix*). We then used the model to predict rank class in the test set of 10 individuals (20.4% of the dataset) initially removed from the data.

Using this approach, we correctly predicted rank class for a median of 8 of the 10 individuals in a test set (80% predictive accuracy; Fig. 3). We also calculated absolute error in model prediction by summing the absolute value of the true rank class minus the predicted rank class across all test set individuals. The median error rate, three errors per run, was significantly smaller than the same values calculated based on random prediction ($P = 0.002$; Fig. 3).

Although rank hierarchies in female macaques tend to be stable, female dominance ranks do sometimes change, particularly upon replacement of individuals within a group. As a result of such changes, we were able to opportunistically sample RNA from the same individual, while at different ranks, for seven females, five of whom experienced changes in ordinal rank (1–5) that switched them into a different relative rank class (A–C) as well. These samples allowed us to test the plasticity of the gene expression response to dominance rank. Specifically, we asked whether gene expression data could predict relative rank position for the same individual after her rank had changed. In this case, we used the 49 samples in the original dataset as our training set, and the 7 samples obtained at a different time as our test set (all seven individual females were also represented in the training set, while occupying different ranks). We found that the gene expression data were sufficient to classify six of the seven females (85.7%) in the test set into the correct rank class (Fig. 3).

Regulatory Mechanisms Underlying Rank Effects. Our data indicated that dominance rank has a strong relationship with gene expression levels. We next evaluated the contribution of two mechanisms that could account for these effects, focusing on the 987 rank-associated genes. Specifically, we investigated whether

rank-dependent differences in tissue composition and GC signaling might help explain these associations. In addition, we focused on a smaller dataset of six females (the set of rank 1 and rank 5 individuals from three social groups; it was not feasible to generate whole-genome bisulfite sequencing data for our full set of 49 individuals, so we focused on the extremes of the rank distribution) to explore the possible relationship between dominance rank and another mechanism suggested to mediate social environmental effects on the genome, DNA methylation.

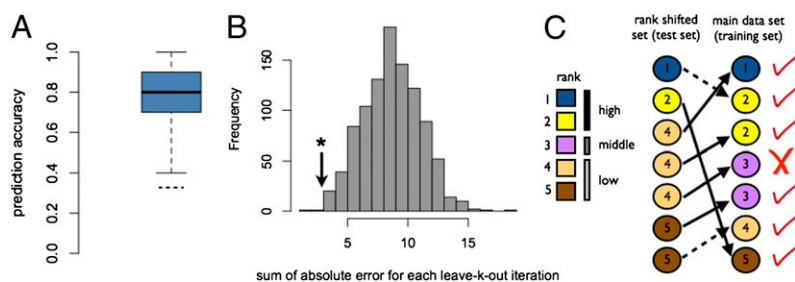
Relationship Between Tissue Composition and Rank-Associated Gene Regulation.

We used fluorescence-activated cell sorting (FACS) analysis to estimate the proportion of the four main PBMC types (CD4⁺ T cells, CD8⁺ T cells, B cells, and monocytes) in blood samples from 39 of the individuals included in our study. Consistent with prior reports, we identified a relationship between PBMC composition and dominance rank: low-ranking females had a reduced proportion of CD8⁺ T cells (a marker associated with cytotoxic T cells) relative to high-ranking individuals ($P < 0.05$, $n = 39$; Fig. 4A). Such changes in cellular composition could, in principle, lead to differences in PBMC gene expression levels by rank, even in the absence of other gene regulatory changes.

To test this possibility, we assessed whether variation in tissue composition could help explain the relationship between dominance rank and gene expression levels for the 987 rank-associated genes. To do so, we first characterized gene expression levels in pure populations of each of the four main PBMC types, again using Illumina HT-12 arrays. We then asked whether the strength of evidence supporting a direct relationship between rank and gene expression levels (measured by a partial correlation) was significantly reduced after taking into account (*i*) differences in cell-type composition among individuals and (*ii*) differences in gene expression levels across cell types.

Specifically, we used our data to calculate the expected gene expression level for each gene in each individual, given tissue composition effects alone. Our estimate was based on the average of the expression levels for the gene across cell types (based on the

Fig. 3. Gene expression data are sufficient to predict relative position in the social hierarchy. (A) Boxplot of predictive accuracy for 10 training set individuals, across 1,000 cross-validation iterations (dashed line shows expected accuracy under random prediction), and (B) histogram of the sum of absolute error between predicted rank class and true rank class, under random assignment. Black arrow, median sum of absolute error across the 1,000 true leave-k-out iterations ($P = 0.002$). (C) Shifts in dominance rank experienced by seven females in the dataset. Solid arrows, cases in which females changed rank classes; dashed arrows, cases in which females changed rank within rank classes; check marks, correct predictions; x, incorrect prediction.



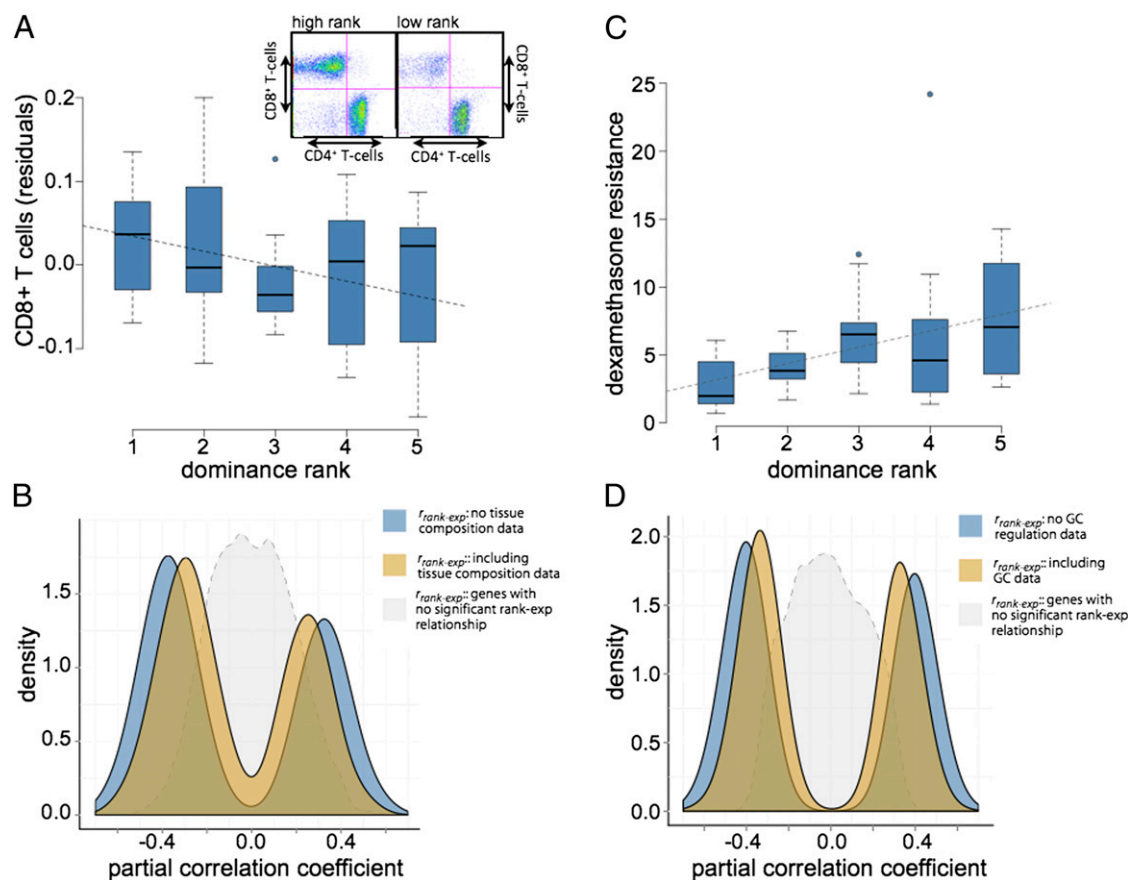


Fig. 4. Effects of tissue composition and glucocorticoid-mediated regulation on rank-associated gene expression levels. (A) Low-ranking individuals exhibit lower proportions of CD8⁺ T cells in PBMCs ($P = 0.047$, $n = 39$; y axis shows the residuals of T-cell proportions after controlling for social group). (Inset) Example data for a rank 1 female and a rank 5 female; x axis shows staining for CD4⁺ (helper) T cells and y axis shows staining for CD8⁺ (cytotoxic) T cells. (C) Low-ranking individuals exhibit decreased GC negative feedback ($P = 0.005$, $n = 49$). The y axis represents levels of circulating cortisol 17 h post-dexamethasone administration. (B and D) Distribution of rank–gene expression partial correlation coefficients for genes in which we inferred a contribution of tissue composition (B, $n = 301$ genes) or an effect of GC regulation (D, $n = 596$ genes). Blue density plots show the distribution of partial correlation coefficients without taking into account tissue composition or GC data; tan density plots show the same distribution after considering these data. The magnitude of the rank–gene expression relationship decreases but remains distinct from the background set of all genes (in gray; $n = 6,097$ genes).

cell type-specific data), weighted by the proportional representation of that cell type in each sample (based on the FACS analysis). After incorporating these estimates into our analysis, we found that tissue composition effects made a modest but significant contribution to the rank–gene expression relationship for 301 of the 987 rank-associated genes (30.5% at a nominal P value ≤ 0.05). As expected, genes that exhibited stronger evidence for a tissue composition effect were also more asymmetrically expressed in CD8⁺ (e.g., cytotoxic) T cells relative to other PBMC types ($P < 10^{-45}$). Overall, however, the distribution of correlations between rank and expression levels for the 301 genes was still markedly nonzero even after taking tissue composition into account (Fig. 4B).

Relationship Between GC Signaling and Rank-Associated Gene Regulation. As expected based on previous work, lower-ranking females in our sample exhibited diminished GC negative feedback following experimental administration of dexamethasone (a synthetic GC; $P < 0.01$, $n = 49$; Fig. 4C), as well as a more sluggish GC regulatory response to an experimentally induced acute stressor ($P < 0.07$; SI Appendix, Fig. S4). Such patterns are characteristic of chronic stress exposure and dysregulated hypothalamic-pituitary-adrenal axis (HPA) activity.

We therefore asked whether variation in GC-mediated signaling explained, at least in part, the relationship between dominance rank and gene expression for the 987 rank-associated genes. To do so, we used an analysis similar to that for tissue

composition effects. Specifically, we asked whether the strength of evidence for a direct relationship between rank and gene expression levels was significantly reduced after taking into account data on GC regulation. We found that the rank–gene expression relationship could be partially explained by dexamethasone suppression data for 596 of the 987 target genes (60.5%, at a nominal P value ≤ 0.05), including known targets of GC-mediated signaling such as the GC receptor (*NR3C1*) and the proinflammatory gene *PTGS2* ($P = 0.015$ and $P = 0.002$, respectively). Similar to our findings for tissue composition, the effects of GC regulation were modest (Fig. 4D and SI Appendix, Fig. S5). However, the contribution of GC-mediated regulation may be twofold, as GC physiology can also impact tissue composition (27). Indeed, for 203 genes, both GC-mediated regulation and tissue composition contributed to the rank–gene expression relationship, a larger overlap than expected by chance (hypergeometric test; $P = 0.001$). This overlap is also suggested by joint analysis of GC resistance and tissue composition data, which tends to, but does not always, account for more of the rank–gene expression relationship than either dataset alone (SI Appendix, Fig. S5).

Dominance Rank and DNA Methylation Patterns. Finally, we tested whether DNA methylation, a regulatory mechanism that has been linked to social environmental effects (28), might also play a role in the association between dominance rank and gene expression. In contrast to our analyses of tissue composition and

GC resistance, dominance rank effects on DNA methylation have not previously been described. To explore this possibility, we generated whole-genome bisulfite sequencing data (11- to 14-fold CpG coverage) from PBMC DNA from three rank 1 females and three rank 5 females. We then investigated the relationships between DNA methylation, dominance rank, and rank-associated gene expression.

We first attempted to identify individual genomic regions that exhibited a detectable association with rank (rank differentially methylated regions; rankDMRs). We used an empirical approach similar to the one previously used for the analysis of bisulfite sequencing data from cancer and healthy tissues (29). We identified 7,089 putative rankDMRs in the genome, and found that these regions were more likely to be located close to transcription start sites (TSS) than randomly distributed sets of similarly long CpG stretches ($P < 0.01$). Additionally, rankDMRs that fell close to genes (within 20 kb) were more likely to be associated with TSS for genes whose expression levels were associated with dominance rank (two-tailed Fisher's exact test, $P < 0.033$).

To assess whether DNA methylation levels might therefore contribute to the rank–gene expression associations, we then collated DNA methylation data in and around the 987 rank-associated genes. We found that methylation data from these regions (within 20 kb of the TSS) clearly discriminated between high- and low-ranking females (Fig. 5). DNA methylation data distinguished rank-associated genes and non-rank-associated genes as well. Specifically, features related to rank-associated differences in methylation (*SI Appendix, Table S3*) differentiated between the 987 rank-associated genes and a set of 1,000 rank-independent genes (those with the least evidence for rank-associated gene expression) at a prediction accuracy of 58.1% ($P < 0.01$ based on comparison with permuted data). Thus, information about rank-associated gene expression patterns is embedded within differential DNA methylation data, and epigenetic changes might account, at least in part, for some of the rank–gene expression associations we observed.

Discussion

Taken together, our findings reveal a strong and widespread association between dominance rank and gene regulation in PBMCs. Our results reinforce the idea that sensitivity to the social environment is reflected in changes in gene expression in the immune system, supporting an increasingly widely recognized link between neural, endocrine, and immune function (30). Moreover, our results demonstrate that these associations also appear to be highly plastic. Not only were gene expression data sufficient to robustly predict relative dominance rank but gene expression profiles also tracked dominance rank shifts closely enough to allow us to predict different rank positions for the same individuals across time. These observations indicate that any causal relationship between dominance rank and gene regulation likely begins with rank, rather than vice versa. Because these study subjects did not experience a completely uniform social history and because unknown variables related to our

method of imposing rank (by order of social introduction) might still play a role in the observed rank dynamics, further experiments—for example, imposition of a second controlled rank manipulation on the same individuals—will be necessary to robustly test this hypothesis. However, the current results support the idea that changes in gene regulation help to explain links between the social environment and physiology, potentially supplying an important piece to the puzzle of how social effects “get under the skin” (3, 31).

Changes in gene expression themselves require a mechanistic explanation. A strength of the experimental approach we used here is that it allowed us to move beyond a simple description of gene expression patterns to also investigate the regulatory mechanisms associated with these effects. Our observations suggest that variation in the expression levels of up to 694 (70.3%) of the 987 rank-associated genes may be partially explained by variation in tissue composition or GC regulation. These findings draw a clear connection between gene regulation and prior work on the physiological effects of rank: GCs, for example, have long been implicated in rank effects in the endocrinological literature (9, 32, 33), and also have important downstream effects on immune function.

In addition, our exploratory analysis identified a signal of both rank and rank-associated gene expression in DNA methylation data. The DNA methylation differences that we observed between ranks were modest relative to studies on different tissues or species, or in the context of cancer (e.g., 29). However, changes in epigenetic marks in healthy adults are likely to be muted in comparison, although evidence for such changes is rapidly accumulating (34, 35). Our findings suggest that the timescale for social effects on epigenetic variation also extends to adulthood, and indicate that such effects may include natural components of social structure, such as dominance rank.

Further work will be necessary to more finely evaluate the relative effects of GC regulation, tissue composition, and DNA methylation—among other regulatory mechanisms—on rank-associated gene expression. For instance, *NR3C1*, the GC receptor gene, plays a key role in linking behavior, HPA axis-mediated stress responses, and gene expression (30). The gene regulatory role of *NR3C1* has previously been linked to other social environmental responses via transcription factor motif analysis (16, 17), and could be further investigated using chromatin immunoprecipitation approaches. Tissue composition effects should also be further dissected. Although we identified an important contribution of CD8⁺ T cells among the four cell types we investigated, additional PBMC subtypes could be relevant (and our estimates of the contribution of tissue composition may therefore be an underestimate). Natural killer cells, for example, are of known importance in the stress response (36), and although they constitute a relatively small proportion of the PBMC pool, their contribution to social stress effects may be disproportionate. Finally, variation across sexes, environmental conditions, species, and genetic backgrounds will be important to take into account, especially as comparative data already indicate that the effects of social status can be highly context-dependent (9), and

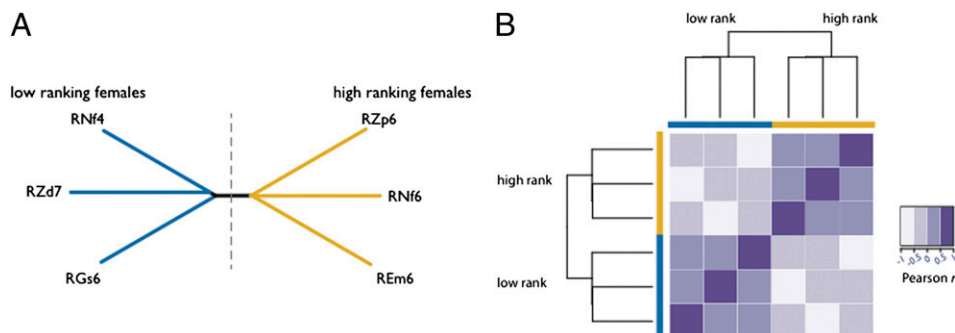


Fig. 5. DNA methylation levels distinguish high-ranking and low-ranking females. (A) A neighbor-joining tree based on DNA methylation levels within 20 kb around the transcription start sites of rank-associated genes (445,059 CpG sites). (B) A heatmap of the same data reveals positive pairwise correlations between individuals of the same rank but not between individuals of different ranks.

that high rank can, in some cases, also induce stress (37). Studying the effects of social status in captivity, as we did here, could exacerbate dominance rank effects if females are less able to distance themselves from their group mates. Alternatively, decreased resource competition in captive animals relative to wild populations could produce the opposite effect.

In conclusion, our results emphasize important connections between behavior, social status, and change at the molecular level. Although in one sense these relationships emphasize the potential costs of adverse social environments, the plasticity of these effects is also encouraging, suggesting that mitigating social stress may confer rapidly realized physiological benefits. Our results motivate efforts to develop a nuanced understanding of social effects on gene regulation, with the aim of both exploring its evolutionary and ecological consequences and addressing its effects on human health.

Materials and Methods

We obtained blood samples from 49 adult female rhesus macaques (members of 10 experimentally constructed social groups), and purified PBMCs from each sample (including replicate samples from a separate time period for seven females while they occupied different dominance ranks). In rhesus macaques, dominance ranks are determined by interactions within a social group among the animals themselves. Thus, weighting the directionality of these interactions by experimentally manipulating the order of introduction into groups is the most direct method of experimentally altering rank. Gene expression profiling was performed using the Illumina HT-12 Expression

BeadChip. To assess the relative proportions of CD4⁺ T cells, CD8⁺ T cells, monocytes, and B cells, we conducted flow cytometry on antibody-stained PBMCs. For cell type-specific gene expression analyses, we physically separated each cell type for five individuals using a BD FACSAria. To identify genes for which gene expression was rank-associated, we used a linear mixed-effects model (38). To test the predictive value of gene expression for relative rank position, we used the program *svm-multiclass* (26). To investigate the contributions of tissue composition and GC signaling, we tested for a reduction in the strength of the rank–gene expression partial correlation when conditioned on tissue composition-related or GC data. DNA methylation was assessed via bisulfite sequencing in six individuals (three rank 1 and three rank 5, from three different social groups). Reads were mapped and methylation levels were estimated using *bismark* (39). The predictive value of differential methylation data for differential gene expression was assessed using *svm-perf* (40). Further details for all experimental and statistical procedures can be found in *SI Appendix*. Data included in this study were obtained in accordance with Institutional Animal Care and Use Committee protocols approved by Emory University (IACUC 2862008Y).

ACKNOWLEDGMENTS. We thank S. Mukherjee and R. Irizarry for advice on data analysis; V. Abadie and members of the Y.G. laboratory for helpful discussions; and S. C. Alberts, E. A. Archie, Z. Gauhar, D. Maestriperi, J. Marioni, J. Pritchard, and three anonymous referees for comments on the manuscript. This work was supported by National Institutes of Health (NIH) Grants GM077959 (to Y.G.) and HD046501, MH081816, and RR00165 (to the Yerkes National Primate Research Center). J.T. was supported by a Chicago Fellows fellowship and K.D.H. was supported by NIH Grant R01HG005220.

- Cohen S, et al. (1997) Chronic social stress, social status, and susceptibility to upper respiratory infections in nonhuman primates. *Psychosom Med* 59:213–221.
- Cole SW, Mendoza SP, Capitanio JP (2009) Social stress desensitizes lymphocytes to regulation by endogenous glucocorticoids: Insights from in vivo cell trafficking dynamics in rhesus macaques. *Psychosom Med* 71:591–597.
- Sapolsky RM (2005) The influence of social hierarchy on primate health. *Science* 308:648–652.
- Altmann J, Alberts SC (2003) Variability in reproductive success viewed from a life-history perspective in baboons. *Am J Hum Biol* 15:401–409.
- Cowlshaw G, Dunbar RIM (1991) Dominance rank and mating success in male primates. *Anim Behav* 41:1045–1056.
- Pusey A, Williams J, Goodall J (1997) The influence of dominance rank on the reproductive success of female chimpanzees. *Science* 277:828–831.
- van Noordwijk MA, van Schaik CP (1999) The effects of dominance rank and group size on female lifetime reproductive success in wild long-tailed macaques, *Macaca fascicularis*. *Primates* 40(1):105–130.
- Sapolsky RM (2004) Social status and health in humans and other animals. *Annu Rev Anthropol* 33:393–418.
- Abbott DH, et al. (2003) Are subordinates always stressed? A comparative analysis of rank differences in cortisol levels among primates. *Horm Behav* 43(1):67–82.
- Michopoulos V, Checchi M, Sharpe D, Wilson ME (2011) Estradiol effects on behavior and serum oxytocin are modified by social status and polymorphisms in the serotonin transporter gene in female rhesus monkeys. *Horm Behav* 59:528–535.
- Morgan D, et al. (2002) Social dominance in monkeys: Dopamine D2 receptors and cocaine self-administration. *Nat Neurosci* 5(2):169–174.
- Pairardini M, et al. (2009) T-cell phenotypic and functional changes associated with social subordination and gene polymorphisms in the serotonin reuptake transporter in female rhesus monkeys. *Brain Behav Immun* 23:286–293.
- House JS, Landis KR, Umberson D (1988) Social relationships and health. *Science* 241:540–545.
- Marmot MG (2006) Status syndrome: A challenge to medicine. *JAMA* 295:1304–1307.
- Chen E, et al. (2009) Genome-wide transcriptional profiling linked to social class in asthma. *Thorax* 64(1):38–43.
- Cole SW, et al. (2007) Social regulation of gene expression in human leukocytes. *Genome Biol* 8:R189.
- Miller GE, et al. (2009) Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. *Proc Natl Acad Sci USA* 106:14716–14721.
- Barr CS, et al. (2004) Sexual dichotomy of an interaction between early adversity and the serotonin transporter gene promoter variant in rhesus macaques. *Proc Natl Acad Sci USA* 101:12358–12363.
- Newman TK, et al. (2005) Monoamine oxidase A gene promoter variation and rearing experience influences aggressive behavior in rhesus monkeys. *Biol Psychiatry* 57(2):167–172.
- Burmeister SS, Jarvis ED, Fernald RD (2005) Rapid behavioral and genomic responses to social opportunity. *PLoS Biol* 3:e363.
- Grozinger CM, Fan Y, Hoover SE, Winston ML (2007) Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Mol Ecol* 16:4837–4848.
- Whitfield CW, Cziko AM, Robinson GE (2003) Gene expression profiles in the brain predict behavior in individual honey bees. *Science* 302:296–299.
- Jarrell H, et al. (2008) Polymorphisms in the serotonin reuptake transporter gene modify the consequences of social status on metabolic health in female rhesus monkeys. *Physiol Behav* 93:807–819.
- Thomas PD, et al. (2003) PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res* 31:334–341.
- Stone EA, Ayroles JF (2009) Modulated modularity clustering as an exploratory tool for functional genomic inference. *PLoS Genet* 5:e1000479.
- Joachims T (1999) Making large-scale SVM learning practical. *Advances in Kernel Methods—Support Vector Learning*, eds Scholkopf B, Burges C, Smola A (MIT Press, Cambridge, MA), pp 169–184.
- Fauci AS, Dale DC (1974) The effect of in vivo hydrocortisone on subpopulations of human lymphocytes. *J Clin Invest* 53:240–246.
- Szyf M, McGowan P, Meaney MJ (2008) The social environment and the epigenome. *Environ Mol Mutagen* 49(1):46–60.
- Hansen KD, et al. (2011) Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 43:768–775.
- Irwin MR, Cole SW (2011) Reciprocal regulation of the neural and innate immune systems. *Nat Rev Immunol* 11:625–632.
- Taylor SE, Repetti RL, Seeman T (1997) Health psychology: What is an unhealthy environment and how does it get under the skin? *Annu Rev Psychol* 48:411–447.
- Creel S (2001) Social dominance and stress hormones. *Trends Ecol Evol* 16:491–497.
- Sapolsky RM, Alberts SC, Altmann J (1997) Hypercortisolism associated with social subordination or social isolation among wild baboons. *Arch Gen Psychiatry* 54:1137–1143.
- Guo JU, et al. (2011) Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat Neurosci* 14:1345–1351.
- Miller CA, Sweatt JD (2007) Covalent modification of DNA regulates memory formation. *Neuron* 53:857–869.
- Esterling BA, Kiecolt-Glaser JK, Bodnar JC, Glaser R (1994) Chronic stress, social support, and persistent alterations in the natural killer cell response to cytokines in older adults. *Health Psychol* 13:291–298.
- Gesquiere LR, et al. (2011) Life at the top: Rank and stress in wild male baboons. *Science* 333:357–360.
- Kang HM, et al. (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178:1709–1723.
- Krueger F, Andrews SR (2011) Bismark: A flexible aligner and methylation caller for bisulfite-Seq applications. *Bioinformatics* 27:1571–1572.
- Joachims T (2006) Training linear SVMs in linear time. *Association for Computing Machinery Conference on Knowledge Discovery and Data Mining* (Philadelphia).

Supplementary Information for Tung et al, “Social environment is associated with gene regulatory variation in the rhesus macaque immune system”

1. Supplementary Information Materials and Methods

2. Supplementary Tables

Table S1: Study subjects

Table S2: Summary of raw bisulfite sequencing data

Table S3: Features used for SVM classification of rank-associated or rank-independent gene expression using DNA methylation data

3. Supplementary Figures

Figure S1: Quantile-quantile plot of p values for dominance rank and potential confounding variables.

Figure S2: Leave-k-out prediction results, under random assignment of test set/training set membership.

Figure S3: Hierarchical clustering relationships for cell type-specific gene expression profiles.

Figure S4: The acute stress response to social isolation by rank.

Figure S5: Effect size estimates for putative regulatory mechanisms

Figure S6: Proportion of CpG sites classified as methylated by base pair position and distance from TSS

1. Supplementary Information Materials and Methods

Study population and samples

Individuals included in this study were members of ten social groups of female rhesus macaques housed at the Yerkes NPRC. Most groups had been formed 5 years previously as a part of studies on the relationships between psychosocial stress, reproduction, metabolism, and behavior (e.g., 1). Two groups were formed more recently following the same protocols (Table S1). However, removing these two groups from the data set produced qualitatively similar results as including them ($r = 0.950$ for the correlation between rank effects on gene expression, by gene, estimated with and without including these new groups), and prediction accuracy for relative rank class for these females did not differ from that for other groups ($p = 0.755$). As described previously (2), groups were formed by removing females from the large breeding groups at the Yerkes NPRC Field Station and placing them in separate housing. Prior to new group formation, all females were ovariectomized. Females in the middle part of the dominance hierarchy were selected to ensure that all had a similar social history. Unfamiliar females were randomly introduced sequentially to indoor-outdoor run housing (25 m by 25 m for each area) over the course of one week, until all groups included five adult females. Dominance hierarchies formed quickly with minimal contact aggression. A single male was also housed in each social group at the time these individuals were sampled.

Dominance rank at time of sampling was strongly correlated with order of introduction (Spearman's $\rho = 0.61$, $p = 3.25 \times 10^{-6}$; excluding one group with multiple rank shifts—used in our plasticity analysis— $\rho = 0.72$, $p = 2.94 \times 10^{-8}$; note that because >1 individual is necessary to constitute a “group,” the first two individuals added were tied in terms of order of introduction, such that the maximum value of $\rho < 1$). In contrast, we identified no significant correlation between dominance rank and age ($p = 0.34$), parity ($p = 0.12$), time since ovariectomy ($p = 0.16$), or time since removal from the large breeding colonies ($p = 0.52$). Furthermore, we also identified no signal of any of these variables on gene expression after controlling for social group (beyond that expected by chance: see Figure S1).

We chose to assess rank-dependent gene expression in PBMCs both because of the accessibility of this tissue, and because of the relevance of PBMCs to immune function, which has close ties to social stress. To sample PBMCs, we obtained blood samples for 49 of the 50 individuals in these social groups in heparinized Vacutainer tubes, as well as replicate samples from a different sampling effort for 7 of these 49 females (changes in dominance rank in these samples occurred within the course of 1 year). For one group, we sampled only four individuals because the lowest ranking female had recently been removed from the group and had not yet been replaced. All study subjects had previously been habituated to conscious venipuncture, and group members were sampled within 10 minutes of entrance into the housing area. Peripheral blood mononuclear cells (PBMCs) were extracted from these samples using a Ficoll gradient, and RNA (for gene expression analysis) or DNA (for methylation analysis) was purified from the total PBMC fraction using the Qiagen RNeasy kit or the Qiagen Gentra Puregene kit respectively. For cell-type specific expression analyses, we physically separated $CD3^+/CD4^+$ (e.g., helper T cell), $CD3^+/CD8^+$ (e.g., cytotoxic T cell), monocyte ($CD14^+$), and B cell ($CD20^+$) populations prior to RNA extraction from each cell type.

Blood samples for PBMC purification were collected over a period of three weeks (Table S1). During each sampling period, individual ranks were confirmed using focal sampling (3) to record the outcomes of dyadic agonistic interactions, with subordinate status defined when an

animal regularly emitted unequivocal submissive gestures to another animal. Agonistic behaviors were identified using a predefined ethogram. Dominance status was defined by which female submitted to which other females and not by which group member was the most aggressive: the 5th ranking female within a group therefore submitted to all other group members, and the highest ranking (alpha) female submitted to none. However, as shown in Fig. 1a, aggression (largely non-contact threats) received by group mates occurred proportionately more often with lower status. In agreement with patterns of rank stability in large mixed-sex groups, these rank assignments have remained largely stable over the course of observation on these animals, with the exception of the 7 cases we used to test for plasticity of the rank-gene expression relationship

FACS analysis

To assess the relative proportions of the major cell types found in PBMCs, we stained a subsample of the purified PBMCs for 39 individuals with five dye-labeled monoclonal antibodies: anti-CD3⁺ (T cells: CD3-PE-Cy7, BD 557749), anti-CD4⁺ (associated with helper T cells: CD4-PerCP-Cy5.5, BD Pharmingen 552838), anti-CD8⁺ (associated with cytotoxic T cells: CD8-APC, Beckman Coulter IM2469U), anti-CD14⁺ (monocytes: CD14-FITC, Beckman Coulter IM0645U), and anti-CD20⁺ (B-cells: CD20-PE, BD Pharmingen 555623). For the other ten individuals in our data set, a FACS machine was unavailable at the time of sampling. To account for differences in antibody staining efficiency between batches of samples, we used as our measurement, for each sample, the proportion of stained live cells of each cell type among the live cells of any type. To control for differences among social groups, we regressed out social group effects and used the residuals in our subsequent analyses.

To establish expectations for cell type-specific gene expression levels, we antibody stained purified PBMCs as described above for five individuals. These samples were collected separately from the samples used in our main analysis of dominance rank and gene expression levels, and were chosen to represent five different social groups and all five possible rank positions. We submitted the total PBMC samples to the University of Chicago Flow Cytometry Facility for physical separation of cell populations. A CD3⁺/CD4⁺ cell fraction, a CD3⁺/CD8⁺ fraction, a CD14⁺ fraction, and a CD20⁺ fraction were sorted on a BD FACS Aria. RNA extractions were conducted separately for each individual-cell type combination (n = 20) for downstream gene expression profiling (see below; Fig S3). Two of these samples (one of B cells and one of CD3⁺/CD8⁺ T cells, both from the same individual) were clear outliers from all other samples and were therefore removed from subsequent analyses.

Illumina HT-12 array cross-hybridization, probe quality control, and general gene expression patterns

To measure gene expression levels, we utilized a human microarray platform, the Illumina HT-12 Expression BeadChip. The genome sequences of rhesus macaques and humans are largely similar, especially in coding regions (~96.5%). The set of probes we utilized contained only few mismatches (median sequence similarity = 48 of 50 base pairs), such that the attenuation of hybridization intensity was minimal. Since our study only involved comparisons of gene expression levels within a species, the sequence mismatches with respect to the array probes are not expected to result in biased estimates of gene expression levels (4). Indeed, we have previously demonstrated that cross-species hybridization to an array designed for a closely related species (and specifically cross-array hybridization involving rhesus macaques and probes based on human DNA sequence) does not produce measurable bias in analyses of differences in

gene expression levels, or greatly reduce power, for *within-species* comparisons (5). We acknowledge that a high-throughput sequencing strategy would have allowed us to capture gene expression variation in greater detail and avoided the need for cross-species hybridization. However, such an approach was cost prohibitive at the time we collected these data.

All RNA samples were hybridized to Illumina HT-12 Expression BeadChips at the University of California Los Angeles Southern California Genotyping Consortium core facility. For the main analysis of rank-gene expression relationships, we conducted technical replicate hybridizations in duplicate (n = 47 individuals) or triplicate (n = 2 individuals). We performed one hybridization per sample for the cell-type specific analyses, based on the very high levels of concordance between technical replicates for our earlier analyses (mean r between technical replicates = 0.987, range = 0.978 – 0.995). Our study design resulted in 4 – 5 biological replicates of gene expression measurements per cell type.

For probe quality control, we mapped the 50 base pair sequence for each Illumina HT-12 probe to the rhesus macaque genome (rhmac 2.0) using *blat*. We removed any probe that failed to map to the macaque genome (n = 19,384 of 47,232 initial probes) or that mapped to multiple places (n = 3,728) in the macaque genome at 80% identity or higher (40 of 50 base pairs). Note that, although a large percentage of Illumina HT-12 probes were culled as a result of this procedure, this phenomenon also occurs when mapping Illumina HT-12 probes to the human genome (using the same methods, 10,981 probes fail to map uniquely to the hg18 version of the human genome). We also removed any probes that overlapped with the location of known segmental duplicates in rhesus macaque greater than 1 kb in length (n = 753 probes). These steps resulted in a set of 23,367 probes that uniquely mapped to the rhesus macaque genome outside of duplicated regions. We also eliminated probes that were not significantly detected in any sample at $p < 0.001$ (compared to Illumina HT-12 control probes), resulting in a set of 7,303 probes (representing 6,097 genes) in the final data set. For downstream analysis, the gene expression data were \log_2 transformed and quantile normalized between arrays using the R package *lumi* (6).

Principal components analysis on mean-centered gene expression data (after regressing out differences in means across social groups) was performed in R using *prcomp* (*stats* package), with the data scaled to unit variance. In addition to investigating the relationship between dominance rank and the resulting PCs, we also explicitly checked for possible effects of age, parity, time since ovariectomy, and time since removal from the large breeding colonies from which these individuals originated. Among the top 10 PCs, which together account for 60% of the variance in the data, we found that dominance rank also correlates with PC4 (which explains 5.1% of variance in the data; rank-PC4 correlation $p = 0.001$). Time since ovariectomy and time since removal from the breeding colony both correlated with PC8 ($p = 0.016$ and $p = 0.008$, respectively), which explained 3.1% of overall variance. These weak or absent effects of non-rank variables are consistent with the absence of evidence for such effects in gene-by-gene analyses (Figure S1) and the absence of detectable correlations between these variables and dominance rank reported above.

Linear effects of rank on gene expression

To assess linear relationships between inter-individual variation in dominance rank and expression levels for each gene, we used the following linear mixed-effects model, which accounts for relatedness in the sample (e.g., (7, 8)):

$$y_{ij} = \beta_j r_i + u_{ij} + b_j + \varepsilon_{ij}$$

Here, y represents the residuals of the normalized, log transformed gene expression levels after controlling for the effect of social group by regressing out mean differences across groups. We utilized these residuals as our measure of gene expression in all subsequent analyses in order to take account of possible biological differences in means across social groups, and to take account of batch effects: with few exceptions, all individuals in a group were sampled at the same time and processed together. Thus, controlling for social group also effectively controlled for sample batch effects, because social groups were subsumed within batches (multiple social groups were sometimes sampled on the same day). Individuals are indexed by i and genes are indexed by j . For gene j , β_j is therefore the fixed effect of rank r_i , b_j is the intercept, and ε_{ij} is the residual error, assumed to be normally distributed with mean zero and variance s^2 . The term u_{ij} refers to the random effect component of the model, where $\text{Var}(u_{ij})$ equals the estimated genetic variance in y multiplied by the pairwise kinship matrix, K . Models were fitted using the R package *emma* (7), with minor modifications to the source code to accommodate gene expression data (this code is available on the Gilad lab website: http://giladlab.uchicago.edu/data/Tung_Rcode/). We evaluated the significance of β_j , the rank effect, as evidence for a linear relationship between dominance rank and gene expression (Figure S1). False discovery rates were evaluated using the method of Storey and Tibshirani (9), implemented in the R package *qvalue*.

Because the social groups in this study were artificially constructed, the species-typical pattern of matrilineal rank inheritance in rhesus macaques did not pertain to our sample. However, some individuals in the data set as a whole were related. As genetic effects can also have an impact on gene expression levels, we evaluated pairwise relatedness between all individuals in the sample by genotyping 51 highly polymorphic microsatellite loci (2.32% missing data) and estimating relatedness using the program COANCESTRY (10, 11). Mean pairwise relatedness in the sample was 0.028 (+/- 0.051 s.d.), and estimated relatedness was highly correlated with pedigree-based estimates of relatedness available for 7 known dyads (Pearson's $r = 0.899$, $p = 0.00588$).

Prediction of rank class using gene expression data

To investigate the predictive value of gene expression data for relative position in the social rank hierarchy, we defined relative ranks (i.e., rank classes) based on prior work, which treated ranks 1 and 2 in 5-female hierarchies as high ranking and ranks 3, 4, and 5 in these hierarchies as low ranking (e.g., 1). We also included an intermediate class based on the agonism data on these study subjects, which suggested additional separation between rank 3 and ranks 4 and 5 (Figure 1; Tukey's HSD test, $p < 0.05$ for the rank 3 – rank 5 contrast and $p < 0.09$ for the rank 3- rank 4 contrast; in contrast, $p = 0.995$ for the rank 4 – rank 5 contrast). We refer to these classes as class A (high), class B (middle), and class C (low).

For each iteration of our prediction analysis, we divided the data set into a training set of 39 individuals (~80% of the data) and a test set of 10 individuals (~20% of the data). Each test set contained 2 randomly chosen individuals of each rank (results were highly similar under completely random assignment that was not balanced across ranks: Figure S2). We used the support vector machine (SVM) approach implemented in the program *svm-multiclass* (12) to develop a model that related the combined gene expression data across all measured genes to rank class. The SVM model was fit using the training set data only, with gene expression values rescaled from 0 to 1 for stability. We then assessed model predictive ability as the percentage of the test set individuals for whom rank class was correctly assigned by the fitted model. We also performed a complementary assessment of model predictive ability by calculating the absolute

error in model prediction (the sum of the absolute value of the true rank class minus the predicted rank class, across all test set individuals: each classification mistake of individual of true rank class A into rank class C, or vice versa, added a value of 2 errors, while each classification mistake of individual of true rank class A or C into rank class B, or vice-versa, added a value of 1 error). We compared the median sum of absolute errors in our cross-validation iterations to the distribution of sum of absolute errors under random assignment of rank class. Note that although we report the results of a three-class classification scheme in the main text, retaining a two-class division (high versus low) yielded as good or better prediction accuracy (average leave-k-out accuracy within the main sample set = 94.7%; prediction accuracy across temporally separated samples = 85%).

We also tested whether gene expression data predicted rank class across time, in individuals who changed their ordinal rank positions. For prediction across temporal replicates, we restricted our analysis to probes represented on both the HT-12 v. 3.0 array (used for the seven temporal replicates) and the HT-12 v. 4.0 array (used to run all other arrays in the study). We normalized, log-transformed, and removed the effects of social group for this common set of probes only (6,735 probes with detectable signal in our data set versus 7,303 such probes in the main analysis). We then treated the main data set of 49 individuals as a training set and the expression data set for the 7 individuals when at an earlier rank as a test set. Model accuracy was calculated as the percentage of the unlabeled test set individuals for whom relative rank position was correctly assigned by the fitted model. Note that in the main text, we report the results of prediction across temporal replicates after normalizing and scaling (from 0 to 1 for each gene expression measurement, across samples) the training set and test set gene expression data together. An alternative approach is to normalize and scale each data set independently. This approach also resulted in excellent prediction accuracy (100% of test set samples were classified into the correct rank class, versus 85.7% accuracy achieved through normalizing all samples together; we report the more conservative estimate in the text).

Effects of cell type proportion

To estimate the expected expression level of each gene based on the cell type composition of each sample, we first measured the expression level for each gene in pure populations of each of four cell types (helper T cells, cytotoxic T cells, monocytes, and B cells, obtained via FACS sorting described above). Pure populations were collected from five individuals across all ranks and five different social groups, and we considered the median expression value across individuals as the estimated expression level for each gene in a given cell type. For the 39 samples for which we were able to obtain cell type composition data, we then weighted the estimated cell type-specific expression levels for each gene by the proportion of the appropriate cell type in the PBMC pool for each sample. This value corresponded to the expected gene expression level in the PBMC sample from that individual, if cell type composition represented the sole mechanism underlying variation in gene expression.

We used these values to investigate the contribution of tissue composition to the rank-gene expression relationship identified for the 987 rank responsive genes, using a partial correlations approach. We reasoned that, if tissue composition effects significantly contributed to the rank-gene expression relationship, then the magnitude of the rank-gene expression partial correlation should decrease when conditioned on the expected expression level of that gene based on the cell type composition data. The magnitude of this decrease should also be larger than that calculated when the rank-gene expression partial correlation was conditioned on

permuted values of these expected gene expression levels. Hence, cases in which the rank-gene expression relation was likely explained by tissue composition exhibited:

$$\|Cor(r, e | c)\| < \|Cor(r, e | c_{permuted})\|$$

where r represents dominance rank; e represents the residuals of the normalized, log transformed gene expression levels after controlling for the effect of social group; and c represents the expected expression level of a gene based on cell type composition alone.

We conducted 1000 permutations of $c_{permuted}$ in order to establish a null distribution. In cases in which the true rank-gene expression partial correlation fell within this null distribution, we failed to reject the null hypothesis (that tissue composition did not significantly contribute to the rank-gene expression relationship). In cases in which the true partial correlation was smaller than expected by chance based on the null distribution, we interpreted our results as supportive of a rank-gene expression relationship mediated, at least in part, by a tissue composition effect. Because we viewed such an effect as mechanistically conservative (i.e., it does not require additional changes in gene regulation), we set our p-value threshold at a nominal value of 0.05.

Effects of glucocorticoid regulation

To investigate the relationship between dominance rank and glucocorticoid regulation, we used data from a dexamethasone (Dex; a synthetic glucocorticoid) suppression test, which assesses the state of GC negative feedback. Prior to Dex administration, serum samples were obtained from each female at 1100 hours on the day of the assay. At 1730 hours, females received an injection of Dex (0.25 mg/kg) and samples were obtained at 1100 hours the following morning to assay cortisol. The degree of GC negative feedback was assessed by the change in cortisol levels between the post Dex-treatment sample and the control sample. Cortisol assays were performed in the Yerkes NPRC Biomarkers Core Lab using established procedures. Serum levels of cortisol were determined by radioimmunoassay (RIA) with a commercially available kit (Beckman-Coulter/DSL, Webster TX) previously validated for rhesus macaques. Using 25 μ l of serum, the assay has a range from 0.5 to 60 μ g/dl with an inter- and intra-assay CV of 4.9% and 8.7%, respectively. To test the relationship between rank and degree of dexamethasone suppression (Dex resistance), we assessed the significance of the rank term in a linear model relating rank to the change in cortisol levels after Dex administration.

To evaluate the contribution of GC regulatory state to the variation in the expression levels of the 987 rank-responsive genes, we again employed a partial correlations approach. This analysis paralleled the analysis we conducted for tissue composition effects, except that the relationship we evaluated was:

$$\|Cor(r, e | g)\| < \|Cor(r, e | g_{permuted})\|$$

where r represents dominance rank and e represents gene expression levels, as before; and g represents the residuals of data from the Dex clearance trials after controlling for social group effects. As before, we established a null distribution using 1000 permutations of $g_{permuted}$. In cases in which the true rank-gene expression partial correlation fell within this null distribution, we failed to reject the null hypothesis (that glucocorticoid resistance did not significantly contribute to the rank-gene expression relationship). Conversely, in cases in which the true partial correlation was smaller than expected by chance from the null distribution, we interpreted our results as supportive of a rank-gene expression relationship mediated, at least in part, by a GC regulatory effect. Because such an effect is conservative with respect to prior knowledge about the physiology of dominance rank effects, and does not require additional changes in gene regulatory mechanisms, we again set our p-value threshold to a relatively relaxed nominal value

of 0.05. Note however that, unlike the case of for tissue composition, we were not able to estimate gene expression levels under a scenario in which GC-mediated effects are the sole contributor. Thus, it is possible that for some genes identified via this method, the rank-gene expression relationship and the rank-GC signaling relationship are statistically correlated but mechanistically independent.

Joint analysis of tissue composition effects and GC effects were pursued in a parallel manner, except that we evaluated whether:

$$\|Cor(r, e | c, g)\| < \|Cor(r, e | c_{permuted}, g_{permuted})\|$$

Bisulfite sequencing and low level data processing for DNA methylation

To measure DNA methylation levels via whole genome bisulfite sequencing, we prepared four sequencing libraries for each individual from DNA obtained from purified PBMCs, following the method of Lister et al (13). Unmethylated lambda phage DNA was incorporated into each library in order to assess the efficiency of bisulfite conversion. We then sequenced 8 Illumina HiSeq flow cell lanes for each individual, with each library represented on 2 lanes to minimize PCR duplicates (see Figure S6; with the exception of 2 individuals, for whom low coverage and a truncated sequencing run motivated us to conduct four additional lanes of sequencing). All sequencing runs were single-ended and 75 base pairs in length. We truncated the resulting reads to 70 base pairs to remove lower quality bases near the end of the reads, and used the tool *cutadapt* (14) to trim off any remaining adapter sequence incorporated in the read.

The resulting reads were mapped to a combined rhesus macaque genome (rheMac 2.0) and lambda phage genome using *bismark* (15), which uses a fully bisulfite converted reference genome sequence as the basis for read mapping (Table S2). Total coverage for each CpG site and the number of reads for each site that were methylated were evaluated using the *methylation_extractor* tool in *bismark*. In doing so, we masked the first three base pairs of each aligned read based on evidence that the methylation estimates for these positions were systematically biased (Fig S5, see also (16) for description of a similar phenomenon).

Analysis of rank-related DNA methylation levels

For the neighbor-joining and hierarchical clustering analyses of DNA methylation data, we identified the locations of Ensembl-annotated transcription start sites for the set of 987 rank-responsive genes (based on the gene expression data). Annotated TSSs were available for 811 of these genes, and we randomly chose one TSS per gene if multiple TSS were associated with the same gene. We then calculated methylation levels for each CpG site within 20 kb of each of these TSS, based on the raw percentage of reads for that site with evidence for a methylated CpG. We excluded sites with 0 coverage in any individual. Over all sites, this procedure produced a data matrix of 445,059 CpG methylation levels for each of the six individuals in the sample. We used these data for the neighbor-joining (conducted using the R package *ape*) and hierarchical clustering analyses depicted in Figure 5, after regressing out social group from the data for each site.

To assess whether features based on differential methylation data could predict rank-associated differential expression, we used leave-one-out cross-validation on a data set consisting of transcripts from the 987 rank-responsive genes and transcripts from the 1,000 genes with the least evidence for rank-associated variation in expression levels (i.e., largest p-values in the initial analysis of the rank-gene expression relationship). For each leave-one-out iteration, we fit a model relating features based on differential methylation (Table S3) to a binary differential

expression variable (rank-associated genes were assigned a value of 1 and rank-independent genes were assigned a value of -1) using *svm-perf* (12, 17). Overall prediction accuracy was measured as the percent of “left-out” genes (of 1,987 total genes) for which the gene expression state was correctly predicted. We evaluate significance of this prediction accuracy by comparison to prediction accuracies calculated in the same way, but on our data set with the labels (rank-associated or rank-independent) permuted, across 100 permutations.

To identify rankDMRs, we used the *bsmooth* method implemented in the R package *bsseq*. Briefly, this approach uses windows of 70 CpG sites or 1,000 base pairs (whichever is larger) to estimate the binomial probability that a CpG site is methylated, smoothing these estimates across nearby CpG sites. This method allowed us to take account of spatial correlations between nearby CpG sites and weight CpG sites with greater coverage more heavily. We limited our analysis to CpG sites with at least 2x coverage in each of the six individuals. We then analyzed the set of CpG t-statistics produced by comparing the smoothed probabilities of methylation for low ranking individuals to the same values for high ranking individuals. We empirically classified differentially methylated regions as regions in which (i) contiguous CpGs were no further than 300 base pairs away from each other, (ii) all t-statistics were in the extreme end (0.5%) of the genome wide distribution of t-statistic magnitudes, (iii) all t-statistics were in the same direction (consistently more methylated in high ranking individuals, or consistently more methylated in low ranking individuals), and (iv) at least 3 CpG sites were covered and exhibited at least a 10% difference in mean methylation levels between the two classes. DMRs were merged if they were called within 1 kilobase of each other, with no intervening analyzed CpGs.

To test for overrepresentation of rankDMRs near rank-associated differentially expressed genes, we used *bedtools* (18) to identify the TSS closest to each DMR, within 20 kb or less intervening distance (our results are consistent if we use a smaller cutoff of 15 kb or a larger cutoff of 30 kb: $p < 0.03$ for 30 kb and $p < 0.05$ for 15 kb; smaller distances greatly reduce the sample size of genes). This procedure assigned one transcript (gene) to each DMR, provided that the DMR was close to a genic region. Within this set, we counted the number of appearances of rank-associated differentially expressed genes ($n = 164$ TSS, representing 162 discrete genes) and the number of appearances of non-differentially expressed genes ($n = 654$ TSS). We used a two-tailed Fisher’s exact test to compare these numbers to the numbers of cases of differentially expressed (1,351 TSS) and non-differentially expressed (6,577 TSS) genes in general. In all cases in which multiple transcripts for a gene utilized the same TSS, we counted that TSS only once.

2. Supplementary Tables

Table S1. Study subjects¹.

Animal ID	Social group	Date of introduction (month/year)	Rank (at time of sampling)	Sample date (month/year)	Time in group (months)
RBm4	1	07/05	1	07/10	60
ROh4	1	07/05	2	07/10	60
RHn6	1	07/05	3	07/10	60
RMg5	1	10/08	4	07/10	21
RVi4	1	07/05	5	07/10	60
RBe5	2	07/05	1	08/10	61
RHe4	2	07/05	2	07/10	60
RMg3	2	07/05	3	07/10	60
RRb7	2	07/05	4	07/10	60
RZt5	2	07/05	5	07/10	60
RZr2	3	03/10	1	07/10	4
RVh5	3	03/10	2	07/10	4
RNu7	3	04/10	3	07/10	3
RBk7	3	08/10	4	08/10	<1
RGv6	4	07/05	1	07/10	60
RTr4	4	07/05	2	07/10	60
RCK4	4	07/05	3	07/10	60
RWe7	4	07/05	4	07/10	60
RCt4	4	07/10	5	08/10	1
RZp6	5	07/05	1	07/10	60
RIz6	5	07/05	2	07/10	60
RYn5	5	07/05	3	07/10	60
RRu6	5	07/05	4	07/10	60
RZd7	5	07/05	5	07/10	60
RWu4	6	07/05	1	08/10	61
RCv6	6	07/05	3	08/10	61
RDv6	6	07/05	4	08/10	61
RIId7	6	07/05	2	08/10	61
RJc6	6	07/10	5	08/10	61
RNf6	7	07/05	1	08/10	61
RZk6	7	07/05	2	08/10	61
RQq4	7	07/05	3	08/10	61
RFc6	7	07/05	4	08/10	61
RNf4	7	12/08	5	08/10	20
REm6	8	07/05	1	08/10	61
RTv6	8	07/05	2	08/10	61
ROb6	8	07/05	4	08/10	61
RRa7	8	07/05	3	08/10	61
RGs6	8	07/05	5	08/10	61
ROy4	9	07/05	1	08/10	61
RYh4	9	07/05	2	08/10	61
RWb7	9	07/05	3	08/10	61
RFp8	9	02/10	4	08/10	6
RIp7	9	07/09	5	08/10	13
RMu3	10	03/10	1	08/10	5
RVg5	10	03/10	2	08/10	5
RDe3	10	03/10	3	08/10	5
RUo4	10	03/10	5	08/10	5
RMf4	10	03/10	4	08/10	5

¹Individuals highlighted in red were those sampled at two different times while occupying 2 different ranks. At the alternative time point, RRa7 was ranked 4, RWb7 was ranked 5, RVi4 was ranked 2, RBm4 was ranked 4, ROh4 was ranked 1, RIId7 was ranked 4, and RMg5 was ranked 5.

Table S2. Summary of raw bisulfite sequencing data.

Individual	Rank	Total number of reads ¹	Total uniquely mapped reads ²	Number of spike-in reads ³	Conversion efficiency	Mean CpG coverage
REm6	1	7.24e8	4.39e8	2.99e5	99.6%	11.30
RGs6	5	7.95e8	4.61e8	3.28e5	99.6%	11.78
RNf4	5	6.57e8	4.41e8	3.27e5	99.5%	13.91
RNf6	1	6.04e8	3.91e8	3.22e5	99.6%	11.80
RZd7	5	6.36e8	4.11e8	3.37e5	99.5%	14.05
RZp6	1	5.76e8	3.58e8	2.91e5	99.6%	10.90

¹For REm6 and RGs6, 2.00e8 and 2.18e8 reads were 50 bp reads instead of 75 (truncated to 70 base pairs for mapping purposes) bp reads

²With apparent PCR duplicates removed

³Reads mapped to cl857 lambda DNA after filtering for PCR duplicates and removing non-uniquely mapped reads

Table S3. Features used for SVM classification of rank-associated or rank-independent gene expression using DNA methylation data.

1	distance to nearest DMR
2	log10 distance (in base pairs) to the nearest DMR
3	t-statistic area (sum of all t-statistics for CpGs in a DMR) for the nearest dmr
4	number of CpG sites covered by the nearest dmr
5	width of the nearest DMR in base pairs
6	t-statistic area x distance (bps) to the nearest DMR
7	t-statistic area x log10 distance (bps) to the nearest DMR
8	mean t-statistic across all CpGs 1kb upstream of the TSS
9	mean t-statistic across all CpGs 5kb upstream of the TSS
10	mean t-statistic across all CpGs 20kb upstream of the TSS
11	mean t-statistic across all CpGs 50kb upstream of the TSS
12	maximum absolute value of the t-statistic 1kb upstream of the TSS
13	maximum absolute value of the t-statistic 5kb upstream of the TSS
14	maximum absolute value of the t-statistic 20kb upstream of the TSS
15	maximum absolute value of the t-statistic 50kb upstream of the TSS
16	mean t-statistic across all CpGs 1kb downstream of the TSS
17	mean t-statistic across all CpGs 5kb downstream of the TSS
18	mean t-statistic across all CpGs 20kb downstream of the TSS
19	mean t-statistic across all CpGs 50kb downstream of the TSS
20	maximum absolute value of the t-statistic 1kb downstream of the TSS
21	maximum absolute value of the t-statistic 5kb downstream of the TSS
22	maximum absolute value of the t-statistic 20kb downstream of the TSS
23	maximum absolute value of the t-statistic 50kb downstream of the TSS
24	number of CpGs with t-stats in the majority direction, 1 kb upstream of the TSS
25	% of CpGs in the majority direction, 1 kb upstream of the TSS
26	number of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$ (0.5% of the genome-wide distribution), 1 kb upstream of the TSS
27	% of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$, 1 kb upstream of the TSS
28	number of CpGs with t-stats in the majority direction, 5 kb upstream of the TSS
29	% of CpGs in the majority direction, 5 kb upstream of the TSS
30	number of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$ (0.5% of the genome-wide distribution), 5 kb upstream of the TSS
31	% of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$, 5 kb upstream of the TSS
32	number of CpGs with t-stats in the majority direction, 20 kb upstream of the TSS
33	% of CpGs in the majority direction, 20 kb upstream of the TSS
34	number of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$ (0.5% of the genome-wide distribution), 20 kb upstream of the TSS
35	% of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$, 20 kb upstream of the TSS
36	number of CpGs with t-stats in the majority direction, 50 kb upstream of the TSS
37	% of CpGs in the majority direction, 50 kb upstream of the TSS
38	number of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$ (0.5% of the genome-wide distribution), 50 kb upstream of the TSS
39	% of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$, 50 kb upstream of the TSS
40	% of CpGs in the majority direction, 1 kb downstream of the TSS
41	1kb num t stats > abs(2.97) downstream

42	number of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$ (0.5% of the genome-wide distribution), 1 kb downstream of the TSS
43	% of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$, 1 kb downstream of the TSS
44	number of CpGs with t-stats in the majority direction, 5 kb downstream of the TSS
45	% of CpGs in the majority direction, 5 kb downstream of the TSS
46	number of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$ (0.5% of the genome-wide distribution), 5 kb downstream of the TSS
47	% of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$, 5 kb downstream of the TSS
48	number of CpGs with t-stats in the majority direction, 20 kb downstream of the TSS
49	% of CpGs in the majority direction, 20 kb downstream of the TSS
50	number of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$ (0.5% of the genome-wide distribution), 20 kb downstream of the TSS
51	% of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$, 20 kb downstream of the TSS
52	number of CpGs with t-stats in the majority direction, 50 kb downstream of the TSS
53	% of CpGs in the majority direction, 50 kb downstream of the TSS
54	number of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$ (0.5% of the genome-wide distribution), 50 kb downstream of the TSS
55	% of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$, 50 kb downstream of the TSS
56	total number of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$, within 1 kb on either side of the TSS
57	total number of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$, within 5 kb on either side of the TSS
58	total number of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$, within 20 kb on either side of the TSS
59	total number of t-stats for which $\text{abs}(\text{t-stat}) > 2.97$, within 50 kb on either side of the TSS

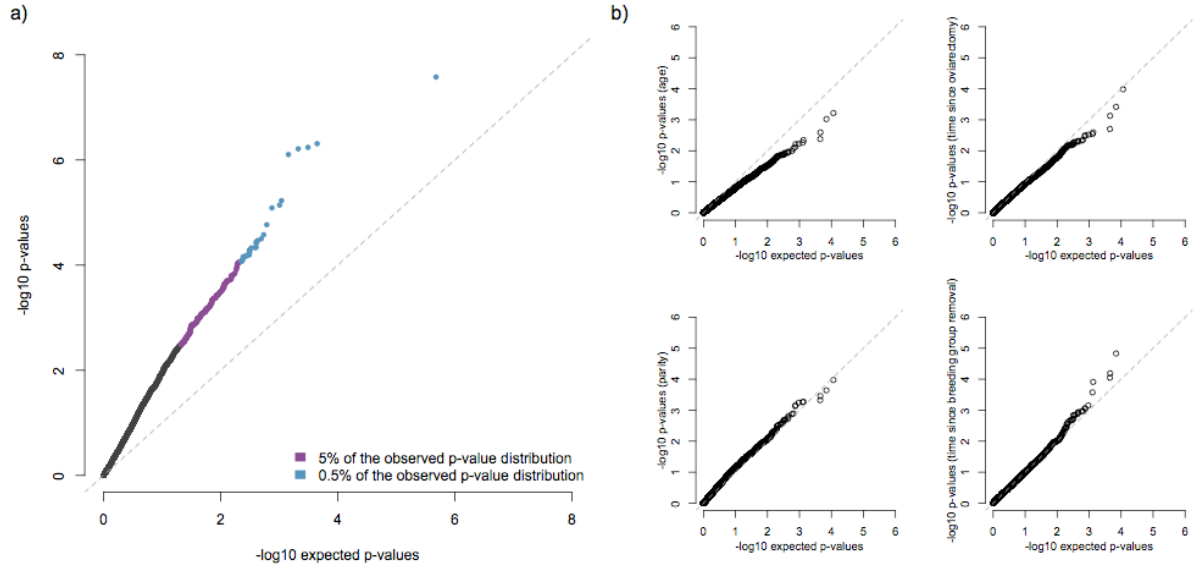


Figure S1. Quantile-quantile plot of p values. a) Quantile distribution of p values for the relationship between dominance rank and gene expression levels in the 6,097 genes we analyzed in our main data (y-axis) compared to quantiles from theoretical p values from a uniform distribution (x-axis); b) Q-Q plots of p values for possible confounding variables (age, parity, time since ovariectomy, time since removal from breeding groups) analyzed using a linear mixed effects model parallel to that run for dominance rank. None of these variables exhibits an enrichment for low p-values, indicating that they do not explain substantial variance in the gene expression data set.

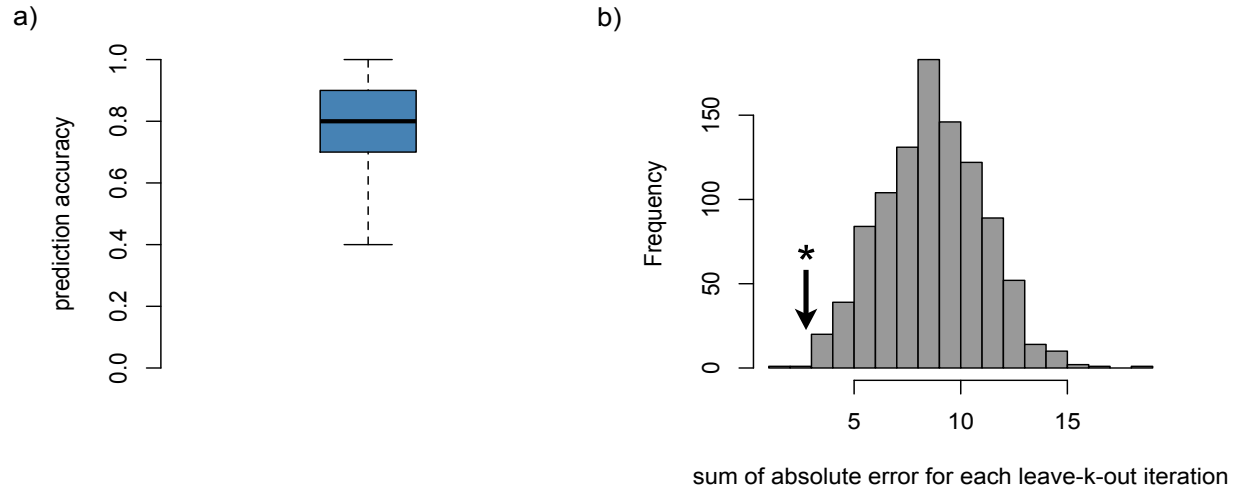


Figure S2. Leave-k-out prediction results, under random assignment of test set/training set membership instead of random assignment with removal of an equal number of individuals of each rank (as reported in the main text). a) boxplot of predictive accuracy for 10 training set individuals obtained across 1000 leave-k-out iterations followed by cross-validation, and b) histogram of the sum of absolute error between predicted rank class and true rank class, if rank classes were randomly assigned. The black arrow and asterisk show the median sum of absolute error across the 1000 true leave-k-out iterations ($p = 0.011$).

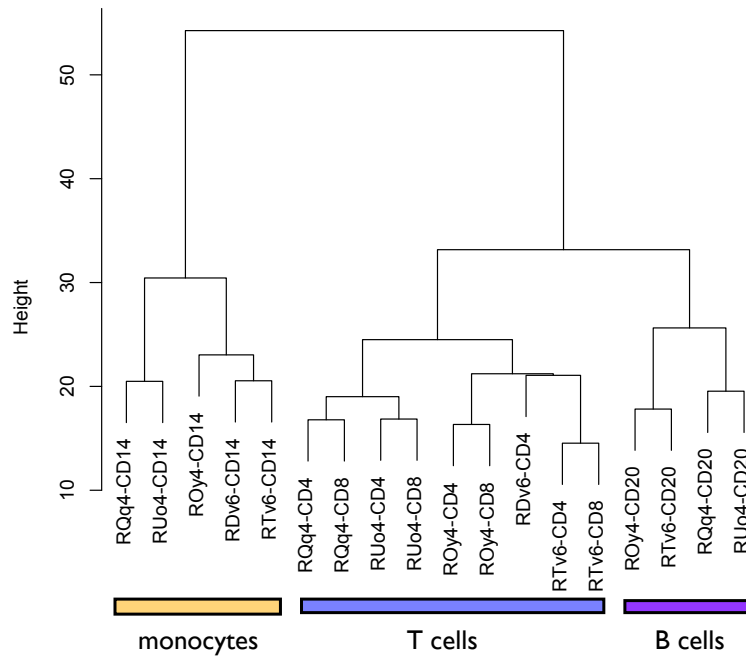


Figure S3. Hierarchical clustering relationships for cell type-specific gene expression profiles. T cells and B cells cluster together, with all monocyte cell fractions outside of this aggregate lymphocyte cluster. Gene expression profiles are more similar within cell types than within individuals, with the exception of CD4⁺ and CD8⁺ T cells, which group together within individuals.

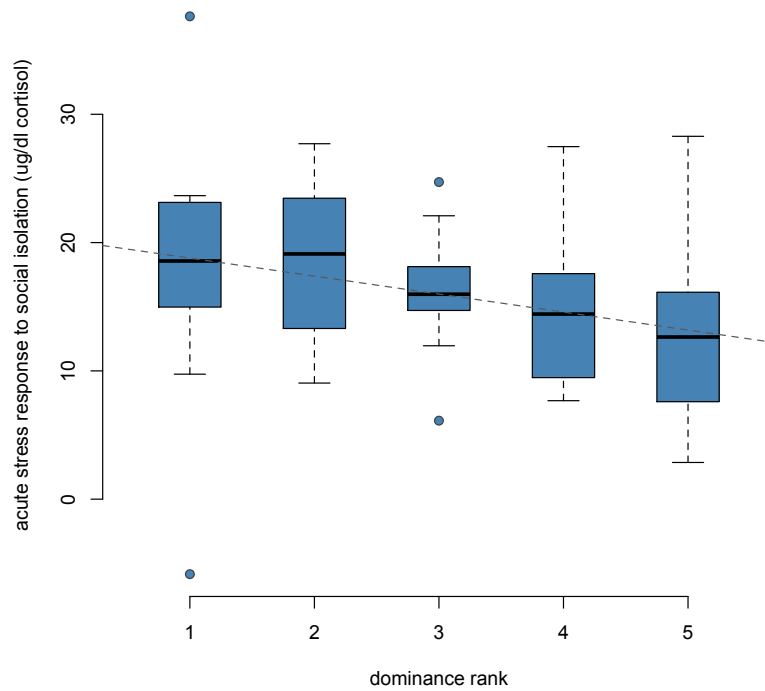


Figure S4. The acute stress response to social isolation tends to be less pronounced in low-ranking individuals. Y-axis depicts the change in cortisol levels between baseline and after social isolation of each individual. Gray line depicts the intercept and slope of a linear regression of this change on dominance rank ($p = 0.067$, $R^2 = 0.070$ $n = 49$).

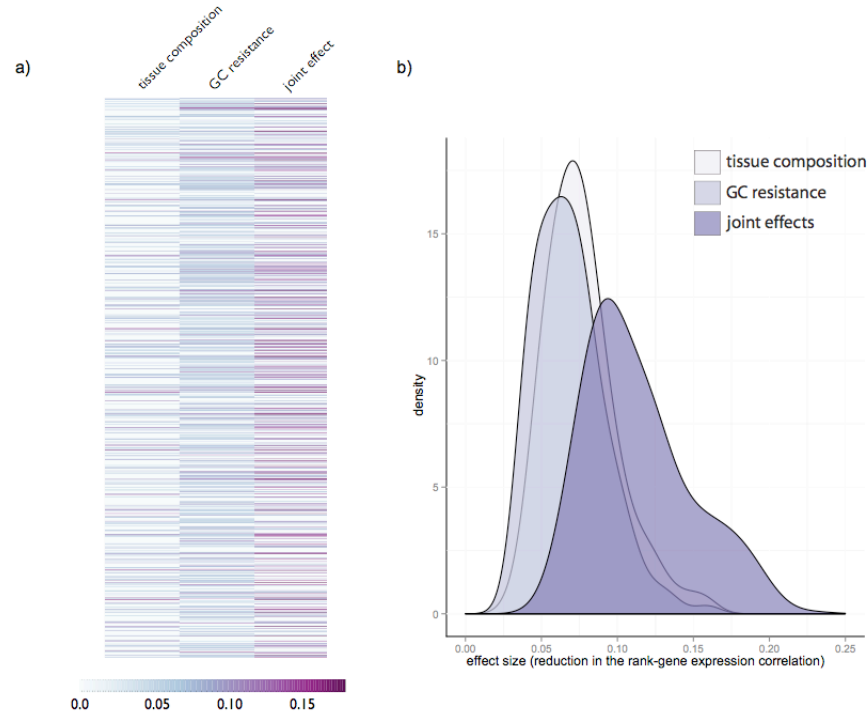


Figure S5. Effect size estimates for putative regulatory mechanisms. a) Reduction in the strength of the dominance rank-gene expression partial correlation when conditioning on tissue composition data (expected gene expression levels based on expression of each gene in each of four individual PBMC cell types and on proportions of these cell types per individual), GC resistance data, or both tissue composition and GC resistance data jointly (e.g., $Cor(rank, gene\ expression \mid tissue\ composition, GC\ resistance)$). Each line represents a gene for which a significant rank association was detected ($n = 987$); genes are ordered alphabetically from top to bottom. Darker colors represent larger effect sizes and values for genes for which no significant effect of the mechanism was detected are zeroed out. All effect sizes were calculated by comparing the median value of the dominance rank-gene expression partial correlation, conditioned on permuted data, to the value of the rank-gene expression partial correlation, conditioned on the candidate mechanism(s). b) Distribution of effect sizes when conditioning the rank-gene expression relationship on tissue composition effects, GC resistance data, or both mechanisms jointly.

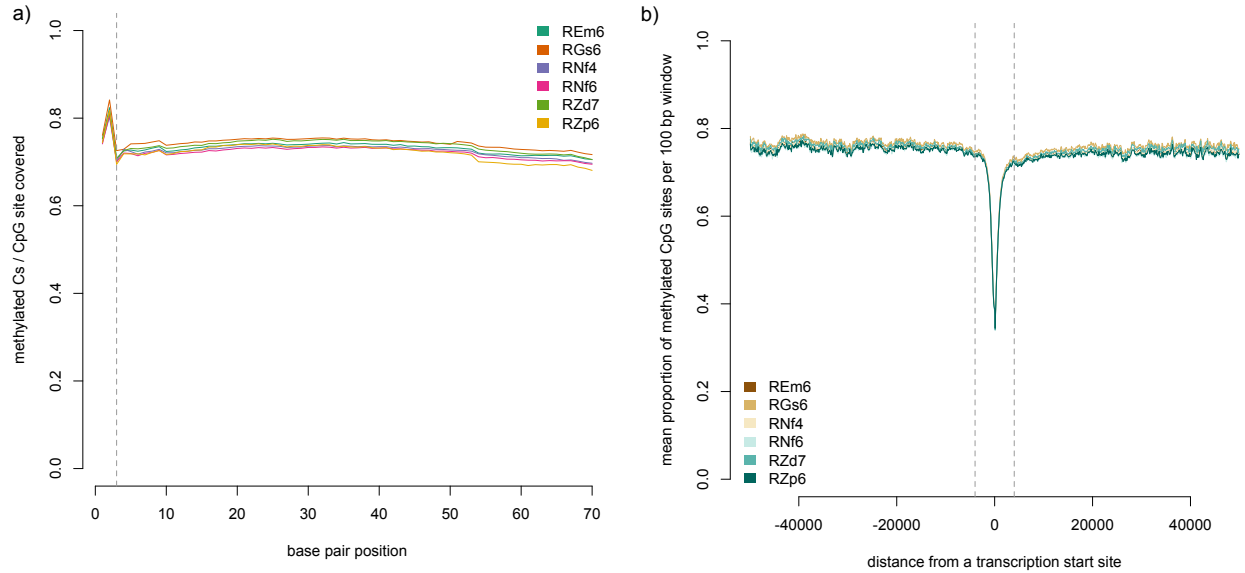


Figure S6. Proportion of CpG sites classified as methylated, by a) base pair position in the sequencing read. The proportion of methylated CpGs observed at each position is expected to be uniform across the length of the read. We observed a systematic bias in DNA methylation estimates for the first three base pair positions in the read, and therefore masked CpG data from these sites when estimating CpG coverage and methylation; and b) distance from annotated transcription start sites. As expected, mean methylation levels drop close to TSS, for all individuals.

References

1. Michopoulos V, Checchi M, Sharpe D, & Wilson ME (2011) Estradiol effects on behavior and serum oxytocin are modified by social status and polymorphisms in the serotonin transporter gene in female rhesus monkeys. *Horm Behav* 59:528-535.
2. Jarrell H, *et al.* (2008) Polymorphisms in the serotonin reuptake transporter gene modify the consequences of social status on metabolic health in female rhesus monkeys. *Physiol Behav* 93:807-819.
3. Altmann J (1974) Observational study of behavior: sampling methods. *Behaviour* 49:227-267.
4. Gilad Y, Rifkin SA, Bertone P, Gerstein M, & White KP (2005) Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. *Genome Res* 15:674-680.
5. Oshlack A, Chabot AE, Smyth GK, & Gilad Y (2007) Using DNA microarrays to study gene expression in closely related species. *Bioinformatics* 23:1235-1242.
6. Du P, Kibbe WA, & Lin SM (2008) lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 24:1547-1548.
7. Kang HM, *et al.* (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178:1709-1723.
8. Yu J, *et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203-208.
9. Storey JD & Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100:9440-9445.
10. Milligan BG (2003) Maximum-likelihood estimation of relatedness. *Genetics* 163:1153-1167.
11. Wang J (2011) COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Mol Ecol Resour* 11:141-145.
12. Joachims T (1999) Making large-scale SVM learning practical. *Advances in Kernel Methods - Support Vector Learning*, eds Scholkopf B, Burges C, & Smola A (MIT Press, Boston).
13. Lister R, *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315-322.
14. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10-12.
15. Krueger F & Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27:1571-1572.
16. Hansen KD, *et al.* (2011) Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 43:768-775.
17. Joachims T (2006) Training linear SVMs in linear time. in *ACM Conference on Knowledge Discovery and Data Mining* (Philadelphia, PA).
18. Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841-842.