Genetic variation and gene expression across multiple tissues and developmental stages in a non-human primate

Anna J. Jasinska[1,2], Ivette Zelaya[3], Susan K. Service[1], Christine B. Peterson[4], Rita M. Cantor[1,5], Oi-Wa Choi[1], Joseph DeYoung[1], Eleazar Eskin[5,6], Lynn A. Fairbanks[1], Scott Fears[1], Allison E. Furterer[7], Yu S. Huang[1,8], Vasily Ramensky[1], Christopher A. Schmitt[1,9], Hannes Svardal[10], Matthew J. Jorgensen[11], Jay R. Kaplan[11], Diego Villar[12], Bronwen L. Aken[13], Paul Flicek[13], Rishi Nag[13], Emily S. Wong[13], John Blangero[14], Thomas D. Dyer[14], Marina Bogomolov[15], Yoav Benjamini[16], George M. Weinstock[17], Ken Dewar[18], Chiara Sabatti[19], Richard K. Wilson[20,21], J. David Jentsch[22,23,24], Wesley Warren[20], Giovanni Coppola[1,25], Roger P. Woods[23,25], Nelson B. Freimer[1,5*]

[1]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA;
[2]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland
[3]Interdepartmental Program in Bioinformatics, University of California Los Angeles, Los Angeles CA, USA
[4]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston TX, USA
[5]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA,USA
[6]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA
[7]Interdepartmental Graduate Program in Neuroscience, University of California Los Angeles, Los Angeles CA, USA
[8]Current address: State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China
[9]Current address: Department of Anthropology, Boston University, Boston, MA, USA
[10]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK
[11]Department of Pathology, Wake Forest School of Medicine, Winston-Salem, NC, USA
[12]University of Cambridge, Cancer Research UK Cambridge Institute, Cambridge, UK
[13]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK
[14]South Texas Diabetes and Obesity Institute, UTHSCSA/UTRGV, Brownsville, TX, USA
[15]Faculty of Industrial Engineering and Management, Technion, Haifa, Israel
[16]Department of Statistics and Operation Research, Tel Aviv University, Tel Aviv, Israel

41 [17]The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA
42 [18]Department of Human Genetics, McGill University, Montreal, Quebec, Canada
43 [19]Departments of Biomedical Data Science and Statistics, Stanford University,
44 Stanford, California, USA
45 [20]The McDonnell Genome Institute, Washington University School of Medicine, St.
46 Louis, MO, USA
47 [21]Current Address: Institute for Genomic Medicine, Nationwide Children's Hospital,
48 Columbus, OH, USA
49 [22]Department of Psychology, University of California, Los Angeles, Los Angeles,
50 California, USA
51 [23]Department of Psychiatry and Biobehavioral Sciences, David Geffen School of
52 Medicine, University of California, Los Angeles, Los Angeles, CA, USA
53 [24]Current Address: Department of Psychology, Binghamton University, Binghamton,
54 NY, USA
55 [25]Department of Neurology, David Geffen School of Medicine, University of
56 California, Los Angeles, Los Angeles CA, USA
57 * to whom correspondence should be addressed at: nfreimer@mednet.ucla.edu
58
59

By analyzing multi-tissue gene expression and genome-wide genetic variation data in samples from a vervet monkey pedigree, we generated a transcriptome resource and produced the first catalogue of expression quantitative trait loci (eQTLs) in a non-human primate model. This catalogue contains more genome-wide significant eQTLs, per sample, than comparable human resources, and reveals sex and age-related expression patterns. Findings include a master regulatory locus that likely plays a role in immune function, and a locus regulating hippocampal long non-coding RNAs (lncRNAs), whose expression correlates with hippocampal volume. This resource will facilitate genetic investigation of quantitative traits, including brain and behavioral phenotypes relevant to neuropsychiatric disorders.

Efforts to understand how genetic variation contributes to common diseases and quantitative traits increasingly focus on the regulation of gene expression. Most loci identified through genome-wide association studies (GWAS) lie in non-coding portions of the genome[1], and are enriched for eQTLs; SNPs that regulate transcript levels, primarily those of nearby genes[2]. This observation suggests that eQTL catalogs may signpost specific variants responsible for GWAS signals[3].

The majority of known human eQTLs have been identified in lymphocytes or lymphoblastoid cell lines obtained from adults[4]. As normal development and function in complex organisms depends on tightly regulated gene expression at specific developmental stages in specific cell types, most existing datasets describing human transcriptome characterization likely miss data relevant to understanding disease[5]. This lack is particularly striking for brain and behavior disorders, given the inaccessibility of the most relevant tissues in living individuals and the enormous modifications that occur in these tissues across development[6].

The Genotype Tissue Expression (GTEx) project, using samples obtained from several hundred post-mortem donors[7], has begun to remedy the lack of human data connecting genotypic variation and multi-tissue transcriptome variation. GTEx provides an eQTL catalog, from multiple tissues, that is the most extensive such resource available[7]. However limitations of GTEx, inherent to human research, motivate the generation and investigation of equivalent resources from model organisms. Advantages of model systems include: (1) the feasibility of controlling for inter-individual heterogeneity in environmental exposures and of minimizing the interval between death and tissue preservation; (2) the practicability of obtaining sizable numbers of multi-tissue samples across a full range of developmental stages; and (3) the opportunity to systematically assess phenotypes of interest in individuals carrying particular eQTL variants. Because of the similarities between humans and non-human primate (NHP) species in behavior, neuroanatomy, and brain circuitry[8,9,10], NHP eQTLs may be particularly valuable for our understanding of neuropsychiatric disorders.

We report here, in 58 Caribbean vervets (*Chlorocebus aethiops sabaeus*) from the Vervet Research Colony (VRC) extended pedigree, the first NHP resource combining genome-wide genotypes[11], multi-tissue expression data across post-natal development, and quantitative phenotypes relevant to human brain and behavior, in a setting in which key environmental exposures have been carefully controlled (Online Methods). The Caribbean vervets are an Old World monkey population that has expanded dramatically from a founding bottleneck occurring with the introduction of West African vervets to the Caribbean in the 17th Century[10]; it has experienced a drastic reduction in genetic variation and, like recently expanded human population isolates, displays enrichment for numerous potentially deleterious alleles (Ramensky, unpublished data).

Through necropsies performed under uniform conditions, we obtained both brain and peripheral tissue samples from the 58 vervets included in this study, whose

genomes were also sequenced[13]. Using these resources we have delineated cross-tissue expression profiles for seven of these tissues, across multiple developmental stages from birth to adulthood. We identified numerous local and distant eQTLs in each tissue, including a master regulatory locus that, via *IFIT1B*, a gene with a hypothesized role in immune function, modulates expression in blood cells of multiple genes on several chromosomes. Additionally, we demonstrated the relevance of vervet tissue-specific eQTLs to higher-order traits, using hippocampus-specific local eQTLs to identify a set of lncRNAs associated with hippocampal volume, a phenotype related to neuropsychiatric disorders[12].

Results

We investigated two datasets. Dataset 1, described previously[13], consists of gene expression levels obtained by hybridizing all available whole blood-derived RNA samples from the VRC pedigree (N=347) to Illumina HumanRef-8 v2 microarrays, which we used because no vervet arrays are available. After filtering out probe sequences not represented in the vervet genome[14] or containing common vervet SNPs[11], we estimated expression levels at 6,018 probes, corresponding to 5,586 unique genes (Supplementary Data 1, Supplementary Table 1). Dataset 2 consists of RNA sequencing (RNA-Seq) reads from seven tissues collected under identical conditions from each of 58 sequenced VRC monkeys (representing 10 developmental stages, from birth through adulthood, Online Methods). Five of these tissues play prominent roles in cognitive and behavioral phenotypes [15-17]: Brodmann area 46 [BA46], a cytoarchitectonically defined region which encompasses most of the dorsolateral prefrontal cortex (DLPFC); hippocampus; caudate nucleus, a component of the dorsal striatum; pituitary gland; and adrenal gland. The other two tissues (cultured skin fibroblasts and whole blood) are relatively accessible, and thus widely used in studies aimed at identifying biomarkers. We assessed expression of an initial set of 33,994 annotated genes. Before analyzing Dataset 2, we minimized spurious signals by excluding genes expressed in fewer than 10% of individuals or at a level lower than one read per tissue. The gene numbers after this exclusion step are listed by tissue and biotype in Supplementary Table 2. A principal components analysis (PCA) of Dataset 2 showed that, overall, expression levels clustered more by tissue than by individual (Supplementary Fig. 1). Most genes were expressed in multiple tissues; 137 genes demonstrated strong expression in only a single tissue (Supplementary Table 3).

Multi-tissue Expression Data: Variation By Age, Sex, Cellular Composition, and Technical Factors

The availability, in Dataset 2, of multiple samples from both sexes at each age point enabled us to examine developmental trajectories and sex differences in gene expression for each tissue. To maximize our ability to observe patterns, we conducted PCA on the expression of the 1,000 most variable genes, separately by tissue (Fig. 1). Comparison of the ranks of expression of the orthologs of these genes in matched tissues in humans and rhesus macaques yielded Spearman correlations

155 of between ~0.5-0.8 and ~0.3-0.4, respectively (Supplementary Material and
156 Supplementary Tables 4-6).

157 Among the seven vervet tissues, the patterns in BA46 and caudate display the
158 clearest association with development; PC1 (20.1% of BA46 variability and 18.5% of
159 caudate variability) distinguishes the vervets in a nearly linear manner, with
160 increasing age. All tissues except fibroblast show a sharp demarcation in expression
161 pattern between males and females; this differentiation is observed on PC1 for
162 hippocampus and pituitary (19.3% and 16.2% of variability, respectively), on PC2
163 for BA46, caudate and blood (15.5%, 17.4%, and 3.2% of variability, respectively),
164 and on PC3 for adrenal (8.2% of variability).

165 As an initial, descriptive exploration of the biology underlying these tissue-related
166 expression patterns, we identified, in the brain and endocrine tissue, the genes in
167 the top and bottom 10% of the distribution of PC loadings on PCs 1, 2, and 3 (200
168 genes total per tissue, per PC). We evaluated the known functions of these genes,
169 which contribute most to the variance explained by the PCs in relation to sex (BA46,
170 caudate, hippocampus, pituitary, and adrenal, see Supplementary Table 7,
171 Supplementary Material) or age (BA46 and caudate, Supplementary Table 8).

172 Age-related expression patterns in BA46 and caudate highlight numerous genes that
173 are essential for nervous system development or that are implicated in human
174 diseases. For example, three thrombospondin genes controlling synaptogenesis
175 show a clear developmental pattern in BA46; *THBS1* and *THBS2* are upregulated in
176 neonates, while *THBS4*, a gene upregulated during human brain evolution[18], shows
177 increasing expression across development (Fig. 2). Supplementary Fig. 2 illustrates
178 striking age-related expression patterns in BA46 and caudate observed for other
179 notable genes (see Supplementary Material). Supplementary Fig. 3 displays
180 developmental expression profiles for the orthologs of these genes in human and
181 rhesus macaque brain tissues that are most equivalent to vervet BA46 and caudate
182 (Online Methods); the overall patterns are roughly similar to, but less pronounced
183 than those we observed in vervet. Given the PCA results showing an age-related
184 component to gene expression variation that differs by tissues, we conducted a
185 differential expression analysis, using age as both a continuous and a categorical
186 predictor in two different linear models. Nearly 8,000 genes across all seven tissues
187 show significant differential expression by age for either analysis, mostly with very
188 small effects (Supplementary Table 9)

189 We considered that cell-type heterogeneity could influence the interpretation of our
190 expression and eQTL results, particularly for blood and the three brain tissues. To
191 evaluate such heterogeneity we conducted a transcriptional deconvolution analysis
192 of these tissues, using published data[19,20] (Supplementary Fig. 4-7). We estimated
193 the diversity of cell types per sample in each tissue by calculating entropy and
194 observed that blood has substantially higher diversity of cell types than do the three
195 brain tissues (Supplementary Fig. 8).

196 We also examined the relationship between the proportion of specific cell types and
197 developmental stage. For BA46 and hippocampus, the proportion of
198 Oligodendrocyte Precursor cells decreases as age increases, which is consistent with
199 data from a prior study in human[21], while the proportion of this cell type in caudate
200 increases with increasing age. Similarly, the proportion of neurons increases as age
201 increases in BA46 and hippocampus, and decreases with increasing age in caudate.
202 (Supplementary Fig. 4-6). We found no correlation between estimated cell
203 proportions and major PC axes in any tissue.

204 We evaluated the potential impact of technical variables on transcriptomic profiles
205 and PC patterns (Supplementary Material). RNA-Seq sample batch demonstrated an
206 association with expression profiles in pituitary and adrenal (PC2) and caudate and
207 pituitary (PC3); we therefore included batch as a covariate in eQTL analysis.

208 **Identification of eQTLs**

209 Whole genome sequencing (WGS) of 721 VRC monkeys has previously provided the
210 first NHP genome-wide, high-resolution genetic variant set[11]: 497,163 WGS-based
211 SNPs that tag common variation genome-wide. Using these SNPs we conducted
212 separate GWAS of Datasets 1 and 2 to identify local (probes/genes < 1 Mb from an
213 associated SNP) and distant (all other probe/gene-SNP associations) eQTLs in each
214 dataset. Covariates in all eQTL analyses included age, sex, and batch.

215 We used SOLAR[22] to estimate heritability of probe expression in Dataset 1,
216 identifying significant heritability for 3,417 probes (out of the 6,018 filtered probes
217 that we evaluated, corresponding to 5,586 unique genes) at a false discovery rate
218 (FDR) threshold < 0.01 (Supplementary Data 1, 2). In a GWAS of each heritable
219 probe, we identified 461 local and 215 distant probes to have one or more eQTLs
220 (significant at Bonferroni-corrected thresholds of 4.8×10^{-8} for local and 1.5×10^{-11}
221 for distant eQTLs, Table 1, Supplementary Data 3). Approximately 35% of probes
222 with a significant eQTL (173/498) displayed at least one local *and* one distant
223 significant association.

224 In Dataset 2 we observed, for each of the five solid tissues, between 361-596 genes
225 with local eQTLs and 30-80 genes with distant eQTLs, and for blood and fibroblasts,
226 60 and 239 genes with local eQTLs and 4 and 43 genes with distant eQTLs,
227 respectively, all at Bonferroni corrected significance thresholds (6.5×10^{-10} [local]
228 and 5.3×10^{-13} [distant]) (Table 1, Supplementary Data 4). The smaller number of
229 eQTLs observed in blood likely reflects heteroegentiy in the proportions of different
230 cell types in this tissue as identified in deconvolution analyses (Supplementary Fig.
231 1, 8); we have no obvious explanation for the relative paucity of eQTLs in
232 fibroblasts, aside from the observation that fewer genes were analyzed in
233 fibroblasts than in tissues with cellular heterogeneity. At Bonferroni significance
234 levels, we had 80% power to detect a significant local eQTL accounting for 11% of
235 variability in expression in Dataset 1, and accounting for 55% of variability in
236 expression in Dataset 2. For about 70% of Bonferroni-significant eQTLs (local and

237 distant and in all tissues), the SNPs demonstrating association had minor allele
238 frequency > 30% (Supplementary Table 10).

239 We considered the possibility that genotypic variation within the vervet pedigree
240 could confound the effects of age in generating the strong loadings on PCs
241 associated with age in BA46 and caudate. Among the 200 genes with such strong
242 loadings, 26 of 200 genes in BA46 showed evidence of an eQTL, and for only one
243 gene (*LOC103219658*) could genotype partially account for the association with age.
244 Similarly, 37 genes showed evidence of eQTLs in caudate, even when using the more
245 liberal FDR controlling procedure. For these 37 genes, we modeled expression as a
246 function of both age and genotype, using the most significant eQTLs, and found that
247 genotype could not account for the association with age (data not shown).

248 We evaluated the enrichment/depletion of cis-eQTLs in genes with age effects, using
249 genes without age effects as reference (Supplementary Table 9). We observe that
250 the genes with age related pattern are actually depleted for eQTLs (Supplementary
251 Table 11), in accord with prior studies predicting that purifying selection results in
252 such depletion in genes that play important roles at particular developmental
253 timepoints[23].

254 **Comparison to Human eQTLs**

255 While the eQTLs summarized in Table 1 exceeded Bonferroni thresholds, we also
256 applied FDR-controlling procedures, to expand the list of local eQTLs for more
257 exploratory investigations, and to make our results comparable to those of GTEx
258 (Table 2). We controlled the FDR for eGenes at 0.05 (Online Methods), accounting
259 for multiple testing using a hierarchical error controlling procedure developed for
260 multi-tissue eQTL analysis[24]. We applied this same procedure to GTEx eQTLs to
261 facilitate the comparison between the datasets.

262 In comparison with GTEx V6, despite having a smaller sample size we identify more
263 local eQTLs (at FDR thresholds applied to both datasets, see Online Methods) for the
264 five solid tissues that were evaluated in both resources (Table 2). We attribute the
265 larger number of local eQTLs identified in the vervet sample, relative to GTEx, to the
266 more homogenous environment of colonied NHPs compared to humans, and to the
267 more uniform process of collecting tissues in this study. We also evaluated the
268 degree to which specific vervet and GTEx eQTLs overlap. All genes with a genome-
269 wide significant vervet eQTL (at FDR <0.05) also display a human eQTL in the same
270 tissue (at p< 0.05), given that the gene has a known human ortholog and was tested
271 in GTEx. Using instead, GTEx's defined significance threshold for orthologous genes
272 (FDR < 0.05), an average of 19% of vervet eQTLs display such a human eQTL (Table
273 2). Restricting the comparison to Bonferroni-significant local eQTLs, an average of
274 23% of vervet eQTLs also have such an eQTL in the same tissue in GTEx
275 (Supplementary Table 12).

276 We also compared our local eQTL results for brain tissues to the Open Access
277 version of human eQTLs from DFPLC, available from CommonMind Consortium
278 (CMC)[25]. More than 87% of vervet brain local eQTL genes with human orthologs in
279 the CMC dataset have a local eQTL at FDR<0.05 in that dataset (Supplementary
280 Material and Supplementary Table 13).

281 **eGene Sharing Among Tissues**

282 We assessed sharing of locally regulated eGenes (genes with a significant local eQTL,
283 see Online Methods) across tissues (Supplementary Fig. 9). We differentiated
284 between tissue-specific and shared eGenes. We observed that the tissue-specific
285 eGenes in all tested tissues except blood are more common than eGenes shared
286 among tissues. The largest number of shared local eGenes was observed between
287 adrenal and pituitary (300), organs inter-regulated in the same neuroendocrine
288 pathway, and then among the three brain regions (239); 229 eGenes are shared
289 across all tissues but blood and 82 eGenes are shared across all seven tissues.

290 **Genomic Distribution of eQTLs**

291 Regulatory variants occur most frequently in functional genomic regions[26], and we
292 find that vervet gene regions encompassing exons, introns and adjacent flanks show
293 a clear enrichment for local eQTLs (Supplementary Fig. 10, Supplementary Table
294 14). Conversely, intergenic regions show a significant deficit of local eQTLs
295 (Supplementary Fig. 10, Supplementary Table 14). As in other primates[27], vervet
296 eQTLs are enriched around gene boundaries (transcription start site [TSS] and
297 transcription end site [TES]) (Supplementary Fig. 11).

298 We used previously published chromatin immunoprecipitation with DNA
299 sequencing (ChIP-Seq) data[28,33] to evaluate eQTL distribution in H3K4me3 enriched
300 regions (promoters) and H3K27ac enriched regions (which include acetylated
301 promoters and enhancers). As H3K4me3 marks are typically conserved across
302 tissues we analyzed them using vervet liver data. As enhancer marks are more
303 tissue specific[29-31] we analyzed H3K27ac marks in both vervet liver and available
304 brain data (caudate and prefrontal cortex) from rhesus macaque[28]. The promoter
305 regions show stronger enrichment for vervet local eQTLs than either genic or
306 H3K27ac-enriched regions (Supplementary Fig. 10, Supplementary Table 14).

307 **Validation of Distant eQTLs: a Master Regulatory Locus on Vervet**
308 **Chromosome 9**

309 Our Dataset 1 is well-powered for discovery of distant eQTLs. Among the 215 genes
310 for which we observed association at genome-wide significance thresholds to one or
311 more distant eQTLs, a locus on CAE9 in which 76 SNPs across a ~500 Kb region
312 displayed genome-wide significant local eQTL signals, stood out for showing
313 association to multiple unlinked genes. For each of these 76 SNPs we identified
314 genome-wide significant distant eQTLs at between five and 14 genes, on different

315 vervet chromosomes, for a total of 2,127 distant SNP-gene associations (Fig. 3,
316 Supplementary Table 15).

317 Because we obtained Dataset 2 using a different platform from Dataset 1, and from a
318 mostly non-overlapping sample (only 6 vervets were in both datasets), we
319 evaluated it for replication of the CAE 9 distant eQTLs, recognizing the limited
320 power of this much smaller dataset. Considering the percent of variance accounted
321 for by the distant eQTLs in Dataset 1 (Supplementary Table 15), we have 82%
322 power to identify eQTLs in Dataset 2, with 58 animals, when the SNP accounts for
323 35% or more of expression variance, using a significance threshold ($p<2.35 \times 10^{-5}$)
324 that accounts for multiple testing of the 76 SNPs to multiple genes (2,127 tests).
325 Two genes, *ST7* (31 SNPs) and *YPEL4* (22 SNPs) replicate association at this
326 threshold, with estimated regression coefficients for these 53 SNP-gene associations
327 being similar in magnitude and direction in the two datasets (Supplementary Table
328 16). We confirmed eight distant associations (*RANBP10, LCMT1, ST7, TMEM57,*
329 *YPEL4, NARF, STXBP1, DEDD2*) across the two datasets, with at least one SNP
330 demonstrating association at a marginal $p<0.05$ (Supplementary Table 15).

331 These results suggest that the CAE 9 eQTL represents a master regulatory locus
332 (MRL). This genomic segment contains a cluster of acid lipase genes and interferon-
333 inducible genes, including *IFIT1B* (Interferon-Induced Protein With
334 Tetratricopeptide Repeats 1B), a gene recently implicated in viral resistance in
335 vervets, but not humans[32]. The same SNPs contributing to the MRL are also local
336 eQTLs for *IFIT1B*, at genome-wide significant levels, however GTEx reports no
337 significant local eQTLs for *IFIT1B* in human blood.

338 Expression of *IFIT1B* correlates strongly with expression of the distant genes
339 regulated by this eQTL (Supplementary Material, Supplementary Table 17). We
340 conducted mediation analyses in Dataset 1 for a SNP (CAE9_82694171) that, at
341 Bonferroni corrected significance thresholds, is both a distant eQTL for all 14 genes
342 and a local eQTL for *IFIT1B* (Supplementary Table 18). This SNP accounts for 19-
343 37% of the variance in expression level of the 14 genes not located on CAE 9. When
344 we conditioned these analyses on expression of *IFIT1B*, the magnitude of these
345 distant associations diminished substantially, the variance accounted for by this SNP
346 dropping to 10% or less for all 14 genes. These results indicate that *IFIT1B*, under
347 direct control of a local eQTL on CAE 9, likely influences expression of 14 other
348 genes spread across the genome. As suggested by studies in human populations,
349 such phenomenon of mediation by local eQTLs of distant eQTLs provides a further
350 validation of the latter loci[33].

351 **Identification of Hippocampus-Specific eQTLs in a Region Linked to**
352 **Hippocampal Volume**

353 In an initial investigation of the impact of vervet tissue-specific eQTLs on higher
354 order traits we focused on MRI-based hippocampal volume, a highly heritable trait
355 in the VRC ($h^2 = 0.95$)[34], for which the strongest QTL signal genome wide (peak LOD

356 score 3.42) lies in an ~8.3 Mb segment of CAE 18. Power simulations in SOLAR
357 indicate that, in the VRC pedigree, quantitative trait data for 347 vervets (the
358 number with hippocampal volume data) provide 80% power to detect a locus with
359 LOD=2 when locus-specific heritability is > 45%. In the center of the broad region
360 around this linkage peak, two hippocampus-specific local eQTLs were Bonferroni-
361 significant at a genome-wide threshold (Fig. 4).

362 The genome-wide significant eQTLs SNPs reside in, and regulate expression of, two
363 lncRNAs located 168 Kb apart: *LOC103222765* (nine associated local eQTL SNPs)
364 and *LOC103222769* (three associated local eQTL SNPs). An additional lncRNA gene,
365 *LOC103222771*, situated two bp from *LOC103222769*, shows hippocampal specific
366 association to six SNPs at a significance level ($p < 10^{-9}$) just above the genome-wide
367 Bonferroni-corrected threshold. While all three genes display hippocampus-specific
368 eQTLs, the genes themselves are expressed across all seven tissues that we
369 analyzed, and show no significant sex or age specific differences in expression
370 patterns (data not shown). The incomplete database annotation of lncRNAs[35] limits
371 comparative analyses of such genes among primates; a BLAST search found a
372 homolog for *LOC103222765* in the white-tufted-ear marmoset and one for
373 *LOC103222771*, in the crab-eating macaque. While *LOC103222765* overlaps a coding
374 gene (*RAB31*), *LOC103222769* and *LOC103222771* do not overlap exons of any
375 coding genes and therefore are more specifically classified as long intergenic non-
376 coding RNA (lncRNA) genes[36].

377 Given the physical proximity of these lncRNAs, we used multivariate conditional
378 analyses to evaluate whether the regulation of these genes depends on a single or
379 multiple independent eQTLs. For each lncRNA we designated a "lead SNP" (the SNP
380 most significantly associated to its expression, Supplementary Table 19). For both
381 *LOC103222769* and *LOC103222771*, modeling expression as a function of both lead
382 SNPs results in diminished significance levels for both SNPs (Supplementary Table
383 19), suggesting that one eQTL regulates both genes. Modeling *LOC103222765*
384 expression as a function of its lead SNP and the lead SNP of the other two genes, the
385 lead SNP for *LOC103222765* remains significant, while the other two lead SNPs are
386 non-significant, confirming the "distinctness" of this signal (Supplementary Table
387 19). This analysis suggests two eQTLs in this region; one associated with
388 *LOC103222765*, and the second associated with *LOC103222769* and *LOC103222771*.

389 We observed a positive correlation between hippocampal expression of
390 *LOC103222765*, *LOC103222769* and *LOC103222771*, and hippocampal volume as
391 assessed by MRI, in six vervets for which both MRI and RNA-Seq data were
392 available. To extend this observation, we assessed, using an independent platform,
393 quantitative real-time PCR (qRT-PCR), *LOC103222765*, *LOC103222769* and
394 *LOC103222771* hippocampal expression in these six vervets and 10 additional
395 vervets for which both hippocampal RNA and MRI data were available. In this
396 expanded sample set, we identified significant positive correlations (Fig. 5) between
397 *LOC103222765, LOC103222769* and *LOC103222771* expression and hippocampal
398 volume. While the above data suggest that genetic variation in this region regulates

399 these lncRNAs and also has a strong impact on the MRI phenotype, colocalization
400 analysis[37] does not support the hypothesis that a single variant accounts for both
401 the genome-wide linkage (MRI) and GWAS (eQTL) findings (8.2% posterior
402 probability).

403 **Discussion**

404 The data presented here provide the first NHP resource for investigating the genetic
405 contribution to inter-individual variation in gene expression across multiple tissues
406 and development. This resource, in a species closely related to humans,
407 complements GTEx, which has become an essential tool for pinpointing genes, and
408 even variants, underlying human GWAS findings[38,39].

409 Several features differentiate this vervet resource from GTEx, reflecting aspects of
410 the study design that are infeasible in human research. Notably, the age-based
411 sampling design enabled us to delineate tissue-specific expression profiles in
412 relation to developmental trajectories. Delineating these trajectories provides
413 insights into biological processes that may be associated with the expression
414 profiles of particular genes. For example, several genes that contribute to synapse
415 formation and postnatal myelination of the central nervous system[40-43] contribute to
416 the near linear age-related pattern observed in BA46 and caudate and, and suggest
417 that the observed expression pattern reflects this process. Conversely, the lack of
418 such a developmentally specific pattern in the hippocampus may relate to the
419 generation of functional neurons in this tissue that occurs throughout the lifespan,
420 underpinning its functions in learning and memory[44,45].

421 Three factors increased the signal-to-noise ratio of vervet eQTL analyses, relative to
422 human studies: (i) the homogeneity of the vervet sample with respect to
423 environmental exposures; (ii) the greater control over necropsy conditions; and (iii)
424 the restricted genetic background of the recently bottlenecked Caribbean vervet
425 population. These factors enabled us to identify 385 genes with one or more
426 genome wide significant distant eQTLs, including the MRL at *IFIT1B*.

427 The function of *IFIT1B*, one of a cluster of five IFIT genes, is poorly understood. It is
428 a paralog of *IFIT1*, which is involved in innate antiviral immunity in mammals,
429 broadly[46], and in regulation of gut microbiota in mouse[47]. In some mammalian
430 species *IFIT1B* contributes to discrimination between "self versus non-self"
431 transcripts based on the lack of 2' O-methylation on mRNA 5' caps in viruses, a so-
432 called cap0 structure[32]. Vervet *IFIT1B* recognizes and inhibits replication of viruses
433 with cap0-mRNAs, while human *IFIT1B* lacks this function[32]. This functional
434 divergence of *IFIT1B* antiviral activity may reflect the divergence of the human
435 lineage from that of other primates, in exposures and adaptations to particular
436 pathogens, including the arboviruses which are responsible for diseases such as
437 encephalitis, dengue, and yellow fever.

438 Our results suggest that investigation of genes regulated by *IFIT1B* in vervet might
439 reveal mechanisms for its role in defense against viral pathogens. While these genes
440 do not act together in any annotated pathway, recent evidence points to immune
441 functions for the products of several of them. For example, *RANBP10*, a
442 transcriptional coactivator, promotes viral gene expression and replication in HSV-1
443 infected cells[48]. *SUGT1*, a cell cycle regulator, is the homolog of *SGT1*, which plays an
444 essential role in innate immunity in plants as well as mammals[49,50], while *TMEM57*
445 shows genome-wide significant association in human to blood markers of
446 inflammation[51].

447 Just as GTEx data are helping refine signals from human GWAS of complex traits[5],
448 we used vervet hippocampal eQTLs to identify a set of lncRNAs as candidate genes
449 for a higher order phenotype, hippocampal volume. The genetic and environmental
450 homogeneity of the relatively small vervet study sample likely facilitated these
451 findings, and supports the extension of multi-tissue vervet eQTL studies as a
452 strategy for identifying loci with a large impact on higher-order phenotypes,
453 generally. As the tissues examined to date are only a fraction of those available from
454 the same set of vervets, it will be possible to extend the investigations reported here
455 to an additional 60 brain regions and 20 peripheral tissues.

456 While expanding expression resources in other NHP species will create additional
457 opportunities to identify eQTLs that are informative for various biomedical
458 investigations[9,52], the Caribbean vervet is unique among NHPs in having abundant
459 natural populations available for such investigations, with an essentially identical
460 genetic background to the samples studied here[10,13]. For example, the lead SNPs for
461 the eQTLs contributing to hippocampal volume in the VRC each occur at a relatively
462 high frequency in these island populations (Supplementary Material). We therefore
463 anticipate that most findings presented here can be followed up through well-
464 powered association studies.

465 **Online Methods**

466 **<u>Study Sample</u>**

467 The vervet monkeys used in this study are part of the Vervet Research Colony
468 (VRC), established by UCLA during the 1970's and 1980's from 57 founder animals
469 captured from wild populations in St. Kitts and Nevis[10]. In 2008 the VRC was moved
470 to Wake Forest School of Medicine; the MRI phenotypes included in this study were
471 collected when the colony was in California (see Supplementary Material for more
472 details). All of the animals in this study were captive-born, mother-reared and
473 socially-housed in large, indoor-outdoor enclosures, in matrilineal groups that
474 approximated the social structure of wild vervet populations. They had a uniform
475 exposure to light and darkness and were fed a standardized diet.

476 **<u>Gene Expression Phenotypes</u>**

477 Two data sets of gene expression measurements were collected. Dataset 1 consisted
478 of microarray (Illumina HumanRef-8 v2) assays of RNA obtained from whole blood
479 in 347 vervets, while Dataset 2 consisted of RNA-Seq data from 60 animals, with
480 seven tissues assayed in each animal. Six vervets were in both Datasets; no
481 randomization was applied in allocating animals to Datasets and investigators were
482 not blinded to the allocation of animals to Datasets.

483 *Dataset 1: Microarrays From Whole Blood*

484 The microarray data set has been described in Jasinska et al.[13] and is available at
485 NCBI at the BioProject PRJNA115831. Details on RNA extraction, cDNA synthesis,
486 and initial data processing are presented in Supplementary Material. To obtain a set
487 of probes usable in vervet from the Illumina HumanRef-8 v2 microarray (originally
488 developed for assaying gene expression in humans), we used the vervet reference
489 sequence to select probes that contain no vervet indels and demonstrate $\leq$ five
490 mismatches, with a maximum of one mismatch in the 16 nt central portion of the
491 probe. To prevent bias in the measurement of expression due to SNP interference
492 with hybridization, we excluded probes targeting sequences with common SNPs
493 identified in the VRC pedigree. A total of 11,001 probes passed these filters
494 (Supplementary Table 1). Illumina provides a "detection p-value" for each subject
495 and probe; $p<0.05$ indicates significant detection of a given probe in a specific
496 individual. We retained for analysis 6,018 probes that were detected with detection
497 p-values of $p<0.05$ in at least 5% of vervets, and tested for association 3,417 probes
498 that were significantly heritable. Expression data were inverse-normal transformed
499 prior to analysis.

500 *Dataset 2: RNA-Seq Data from Seven Tissues*

501 Tissues harvested during experimental necropsies were obtained from 60 vervets
502 representing 10 developmental stages, ranging from neonates (7 days), through
503 infants (90 days and one year), young juveniles (1.25, 1.5, 1.75, 2 years old),
504 subadults (2.5, 3 years old) to adults (4+ years old), with six vervets (3 male and 3
505 female) from each developmental time point. The IACUC protocol number covering
506 the necropsies was A09-512. This necropsy protocol was approved by the IACUC at
507 Wake Forest School of Medicine. Two vervets (a 1.75 year old female and a 7 day old
508 male) for which we did not have WGS data were excluded from the eQTL study.
509 Altogether, in the eQTL study we included 11 vervets below one year old, 23 vervets
510 between one to two years old, and 24 vervets between two and four years old, 29
511 males and 29 females. Details regarding tissue collection and RNA collection
512 procedures are in Supplementary Material.

513 We conducted RNA-Seq for all vervets in seven tissues: three brain tissues (BA46,
514 caudate and hippocampus), two neuroendocrine tissues (adrenal and pituitary) and
515 two peripheral tissues serving as a source of biomarkers (blood and fibroblasts).
516 From purified RNA, we created two types of cDNA libraries, poly-A RNA (fibroblasts,
517 adrenal and pituitary) and total RNA (blood, caudate, hippocampus, BA46) cDNA

518 libraries (Supplementary Table 20, Supplementary Material). For one vervet the
519 RNA-Seq data indicated that the caudate and BA46 samples had been mixed-up, and
520 for this vervet we therefore did not include the data from these two tissues in any
521 analyses. Details on library preparation are in Supplementary Material. The RNA-
522 Seq read data were made available through NCBI as BioProject PRJNA219198.

523 RNA-Seq reads were aligned to the vervet genomic assembly Chlorocebus_sabeus
524 1.1 http://www.ncbi.nlm.nih.gov/assembly/GCF_000409795.2 by the ultrafast
525 STAR aligner[53] using our standardized pipeline. STAR was run using default
526 parameters, which allow a maximum of ten mismatches. Gene expression was
527 measured as total read counts per gene. For paired end experiments, total fragments
528 are considered. Fragment counts that aligned to known exonic regions based on the
529 NCBI *Chlorocebus sabaeus* Annotation Release 100 were quantified using the HTSeq
530 package[54]. The counts for all 33,994 genes were then combined, and lowly
531 expressed genes, defined as genes with a mean in raw counts of < 1 across all
532 samples, as well as genes detected in fewer than 10% of individuals were filtered
533 out. The calcNormFactors function in the edgeR package[55] was applied to normalize
534 counts. Finally, an inverse-normal transform was applied to counts per million prior
535 to analysis.

536 Deconvolution analysis was perfomed in vervet brain and blood tissue using
537 available reference for these tissues. For the brain tissues, gene signatures were
538 obtained from Zhang et al.[20], for blood, cell type specific markers were taken from
539 datasets built into the CellMix package[19]. Cell type composition for each tissue was
540 evaluated using the CellMix R package.

541 *Datasets for comparative expression analysis between species*

542 We performed comparative analysis of gene expression between vervet brain
543 regions and age-matched human and rhesus macaque samples. We compared
544 overall expression profiles between these species and inspected developmental
545 expression patterns of selected genes.

546 We paired age categories between vervet and two primate species with
547 developmental gene expression data available from the Allen Brain Atlas (ABA).
548 Gene expression in human from BrainSpan dataset was assessed using RNA-Seq,
549 and gene expression in rhesus macaque from the NIH Blueprint Non-Human
550 Primate (NHP) Atlas was assessed using microarray[6,52] (Supplementary Tables 21,
551 22). We matched the three vervet brain tissues to the most closely corresponding
552 available tissues in the two other species (Supplementary Table 23).

553 Overall mean levels of expression were compared between species using a rank
554 correlation. For the comparison with human, two independent analyses were
555 performed using two different datasets: GTEx data and ABA developmental data.
556 The rhesus macaque comparison was limited to a single developmental dataset of
557 male animals, also obtained from the ABA. Analyses involving the ABA

558 developmental datasets were limited to the three brain regions most closely related
559 to the brain tissues analyzed in vervets (Supplementary Table 23). For the GTEx
560 comparison, vervet tissues were matched to the five corresponding tissues
561 available: adrenal, blood, caudate, hippocampus and pituitary. As the ABA rhesus
562 macaque dataset included only males, we limited comparisons to male vervets.

563 For each of the three dataset comparisons, vervet raw counts were first converted to
564 RPKM values using the edgeR R package[55]. GTEx and human ABA counts obtained
565 were already normalized to RPKM values; rhesus macaque counts had been
566 normalized using an RMA approach[52]. Mean expression was then calculated by
567 tissue for vervet and comparison datasets. For comparisons to ABA developmental
568 datasests, mean expression was calculated by tissue type and time point, according
569 to matched age groups (Supplementary Tables 21, 22). Vervet gene names were
570 converted to their corresponding human orthologs to ensure gene names matched
571 between vervet and comparison datasets; Genes with no human ortholog were
572 excluded. In addition, genes not present in both vervet and the comparison species
573 dataset were also removed. Variances were then calculated for each gene across the
574 five or three different vervet tissues, for GTEx and ABA comparisons, respectively.
575 The top 1,000 genes with the highest variances were then selected for rank-rank
576 correlation testing. The base R function cor.test was used to perform correlation
577 testing.

578 *Real-time quantitative PCR (qPCR)*

579 Real-time quantitative PCR was performed in two steps. First, reverse transcription
580 (RT) was performed using the SuperScript® III First-Strand Synthesis System (Life
581 Technologies) following the manufacturer's protocol for priming with random
582 hexamers. Custom primers and hydrolysis probes were designed for each lncRNA
583 and three candidate reference genes (Hypoxanthine phosphoribosyltransferase 1,
584 *HPRT1*; Glyceraldehyde 3-phosphate dehydrogenase, *GAPDH*; and Beta-2-
585 Microglobulin, *B2M*) using the Custom TaqMan® Assays Design Tool (Applied
586 Biosystems, Supplementary Table 24). Expression analyses were conducted on the
587 LightCycler™ 480 platform (Roche) using the iTaq® Universal Probes Supermix
588 (Bio-Rad). All qPCR reactions were carried out in triplicate and reactions containing
589 water instead of cDNA were included as negative controls. cDNA samples were
590 diluted 1:5 with water, and a five-point standard curve of four-fold dilutions was
591 prepared for each gene using pooled cDNA as the template. Stability of each
592 candidate reference gene was evaluated using the NormFinder software (v5) in R[56].
593 Quantification was performed using the relative standard curve method, with the
594 geometric mean of the most stably expressed reference genes (*GAPDH* and *HPRT1*)
595 used as an endogenous control for normalization of the interpolated lncRNA
596 quantities. Finally, relative expression levels were generated by dividing the
597 normalized lncRNA quantities by the corresponding quantity in one experimental
598 sample which served as a calibrator. Refer to Supplementary Material for additional
599 experimental details and complete primer and probe sequences information.

600 **Hippocampal Volume Phenotype**

601 Estimates of hippocampal volume were measured in 347 vervets >2 years of age
602 using MRI. Details of the image acquisition and processing protocol were described
603 previously[34] and are outlined in Supplementary Material. Prior to genetic analysis,
604 hippocampal volume was log transformed, regressed on sex and age using SOLAR[22],
605 and residuals used as the final phenotype.

606 **Genotype Data**

607 Genotype data were generated through whole genome sequencing of 725 members
608 of the VRC[11]. Genotypes from 721 VRC vervets that passed all QC procedures can be
609 directly queried via the EVA at EBI (www.ebi.ac.uk/eva) using the PRJEB7923
610 accession number. Two genotype data sets were used in the current study[11]: (1) The
611 Association Mapping SNP Set consists of 497,163 SNPs on the 29 vervet autosomes.
612 In this set of ~500K SNPs, there were an average of 198 SNPs per Mb of vervet
613 sequence, and the largest gap size between adjacent SNPs was 5 Kb. (2) The Linkage
614 Mapping SNP Set consists of 147,967 markers on the 29 vervet autosomes. In this
615 set of ~148K SNPs, there were an average of 58.2 SNPs per Mb of vervet sequence,
616 and the average gap size between adjacent SNPs was 17.5 Kb.

617 The software package Loki[57], which implements Markov Chain Monte Carlo
618 methods, was used to estimate the multipoint identical by decent (MIBD) allele-
619 sharing among all vervet family members from the genotype data. As long stretches
620 of IBD were evident among these very closely related animals, a reduced marker
621 density was sufficient to evaluate MIBD at 1cM intervals; we used a 9,752 subset of
622 the 148K SNP data set. The correspondence between physical and genetic positions
623 in the vervet was facilitated by a vervet linkage map[58], constructed using a set of
624 360 STR markers. Both the physical and genetic position of these markers was
625 known, and genetic locations of SNPs were found by interpolation.

626 **Statistical Analysis**

627 *Principal Components Analysis (PCA)*

628 In Dataset 2, the top 1,000 most variable genes were selected for each tissue, and
629 PCA applied to log2-transformed counts per million, using the singular value
630 decomposition and the prcomp function in R (https://www.R-project.org, version
631 3.2.3). Expression was mean-centered prior to analysis. We examined the genes in
632 the top and bottom 10% of the distribution of PC loadings on PCs 1, 2, or 3 (200
633 genes total per tissue, per PC) where these loadings are taken from the eigen-
634 decomposition of the expression matrix. The gene loadings represent the amount
635 that gene contributes to the PC value for that sample on the axis in question.

636 *Mapping of Gene Expression and Hippocampal Volume Phenotypes*

637 We expected greater power for association analyses of gene expression traits
638 compared to more complex phenotypes. Therefore we applied genome wide
639 association analyses to these traits. For the higher-order phenotype examined
640 (hippocampal volume) we anticipated having power only to detect loci with a much
641 stronger effect, and therefore utilized linkage analysis for this trait.

642 Heritability and Multipoint Linkage Analysis We estimated familial aggregation
643 (heritability) of traits using SOLAR, which implements a variance components
644 method to estimate the proportion of phenotypic variance due to additive genetic
645 factors (narrow sense heritability). This model partitions total variability into
646 polygenic and environmental components. The environmental component is unique
647 to individuals while the polygenic component is shared between individuals as a
648 function of their pedigree kinship. If the variance in phenotype Y due to the
649 polygenic component is designated as σ_g^2 and the environmental component as σ_e^2,
650 then in this model $Var(Y) = \sigma_g^2 + \sigma_e^2$, and the covariance between phenotype values
651 of individuals *i* and *j* is $Cov(Y_i,Y_j)=2\ \varphi_{ij}\ \sigma_g^2$, where φ_{ij} is the kinship between
652 individuals *i* and *j*.

653 Whole genome multipoint linkage analysis of hippocampal volume was also
654 implemented in SOLAR, which uses a variance components approach to partition
655 the genetic covariance between relatives for each trait into locus-specific heritability
656 and residual genetic heritability. Linkage analysis was performed at 1cM intervals
657 using the likelihood ratio statistic.

658 Association Analysis Association between specific SNPs and gene expression
659 phenotypes was evaluated using EMMAX[59]. EMMAX employs a linear mixed model
660 approach, where SNP genotype is a fixed effect, and correlation of phenotype values
661 among individuals is accounted for using an identity by state (IBS) approximation to
662 kinship. Association analyses used the full set of 497,163 SNP markers, and for both
663 Dataset 1 and Dataset 2 included age (where in Dataset 2 age, in days, corresponds
664 to developmental stage), sex, and sample batch as covariates. It is common to try to
665 account for unmeasured factors influencing global gene expression by including
666 probabilistic estimation of expression residuals (PEER) factors as covariates[60]. We
667 considered the controlled nature of the study environment and experimental design
668 to preclude the need for this adjustment.

669 *Colocalization of eQTL and Hippocampal Volume QTL*

670 We evaluated the posterior probability that the hippocampal volume QTL and the
671 hippocampus local eQTLs on CAE 18 share a single, common causal variant using
672 COLOC[37]. The same variants were tested in both analyses, and only six vervets
673 overlapped between the two data sets.

674 *Multiple Testing Considerations in eQTL*

675 We used a Bonferroni correction to account for multiple testing across genes, SNPs,
676 and tissues as our primary error-controlling strategy for the identification of eQTLs.

677 Thresholds for Dataset 2 were more stringent, as more genes were tested than in
678 Dataset 1 (~25K vs. ~3K) and multiple tissues were analyzed in Dataset 2. Dataset 1
679 was analyzed association to 3,417 heritable probes. The local eQTL significance
680 threshold (4.8×10^{-8}) was corrected for the testing of SNPs within 1 Mb of 3,417
681 probes, and the distant eQTL significance threshold (1.5×10^{-11}) accounted for
682 genome-wide testing of 3,417 probes. Dataset 2 significance thresholds were
683 constructed in a similar fashion, but also accounted for testing of 191,263 gene-
684 tissue combinations (the number of genes tested per tissue is in Table 1). The RNA-
685 Seq local eQTL threshold was 6.5×10^{-10}, and the distant eQTL threshold was $5.3 \times$
686 10^{-13}.

687 To identify multi-tissue eGenes, the tissues in which they are active, and the
688 associated SNPs in each of these tissues, we used TreeBH, a hierarchical approach
689 testing proposed in Bogomolov et al.[24] which extends the error-controlling
690 procedure characterized in Peterson et al.[61] to the multi-tissue eQTL setting. To
691 apply this method, the hypotheses are grouped into a tree with three levels: genes in
692 level 1, tissues in level 2, and SNPs in level 3. Testing proceeds sequentially starting
693 from the top of the tree in a manner that accounts for each previous selection step.
694 This method allows control of the FDR of local eGenes (defined as those genes
695 whose expression is regulated in at least one tissue by some genetic variants located
696 within 1 Mb of the gene) and of the expected average false discovery proportion of
697 the tissues in which we claim this regulation is present across the discovered
698 eGenes. P-values are defined by building up from the bottom of the tree. Specifically,
699 to obtain a p-value for the null hypothesis of no local regulation for a given gene in a
700 given tissue (corresponding to a hypothesis in level 2 of the tree), we applied Simes'
701 combination rule[62] to the p-values obtained via EMMAX for the hypotheses of no
702 association between the expression of the gene in the tissue and each of the SNPs in
703 the local neighborhood (corresponding to the hypotheses in level 3 of the tree). To
704 obtain a p-value for the null hypothesis of no local regulation for a given gene in any
705 of the tissues under study (corresponding to a hypothesis in level 1 of the tree), we
706 applied Simes' combination rule to the gene x tissues p-values just described. We
707 then tested the global null hypotheses of no local regulation in any tissue for all the
708 genes in our study, applying the Benjamini Hochberg procedure[63] to control the FDR
709 at the 0.05 level. For those genes for which we were able to reject the null
710 hypotheses of no local regulation, we examined the tissue-specific p-values,
711 applying the Benjamini Bogomolov procedure that allows the identification of
712 significant findings controlling for the initial selection[64]. Finally, the individual SNPs
713 responsible for regulation of the gene in each tissue were identified, again using a
714 selection-adjusted threshold as described in Bogomolov et al.[24] An R package
715 implementing this procedure is available
716 at http://www.bioinformatics.org/treeqtl/.[65]

717 We compared the number of eGenes identified in each tissue using the above
718 procedure with the results of GTEx (Analysis Release V6; dbGaP Accession
719 phs000424.v6.p1). We downloaded all eQTL association results for tissues in

720 common with our study, and applied this same hierarchical procedure to the GTEx
721 results to identify eGenes.

722 *Association between local eQTLs and genomic features*

723 We estimated the possible enrichment of eQTLs in exons, introns, flanking regions,
724 intergenic regions, and regulatory regions using logistic regression in a generalized
725 linear mixed model (GLMM), using the GMMAT software[66]. We categorized each SNP
726 in two binary dimensions (local eQTL and located in or near a given region). A SNP
727 was considered a local eQTL if it was associated (at Bonferroni significance
728 thresholds) to gene expression of a gene within 1 Mb, in any tissue, in either Dataset.
729 Local eQTL status was the outcome variable, and a separate GLMM logistic
730 regression performed for each region. A matrix of r^2 values among all SNPs was
731 included as a random effect to account for lack of independence among SNPs.
732 GLMMs are computationally very demanding and the full set of 497,163 SNPs could
733 not analyzed in one model. We LD pruned the SNP data, agnostic to eQTL status and
734 region, and used 18,464 genome-wide SNPs based on LD pruning the entire set of
735 497,163 SNPs at $r^2<0.6$ in 14 unrelated individuals. This SNP set included 1,202
736 local eQTLs.

737 *Enrichment of local eQTLs in near TSS/TES*

738 Our examination of potential enrichment of local eQTLs near TSS/TES gene regions
739 was purely descriptive and involved no hypothesis testing. We restricted our
740 summary to the 27,196 genes that were <0.5 Mb in size, and the 426,403 SNPs that
741 were within 200kb of the TSS/TES of these genes (or in between the TSS/TES). In
742 this set of 426,403 SNPs, 17,595 were local eQTLs to one (or more) of the 27,196
743 genes (at Bonferroni significance levels), in one (or more) tissues in either Dataset,
744 and were within 200 Kb of the TSS/TES of the gene(s) to which they were
745 associated. For each gene, we created 10 Kb distance bins on either side of the
746 TSS/TES, and tallied the proportion of SNPs in the bin that were local eQTLs for the
747 gene. As the distance between TSS and TES varied by gene, we binned distances in
748 this area by deciles of the total distance.

749 **Data Availability**

750 The RNA-Seq datasets generated in the current study are available in the NCBI Gene
751 Expression Ominibus repository, https://www.ncbi.nlm.nih.gov/gds/?term=PRJNA219198.
752 The other data sets, microarray and genotype, analysed during the current study are
753 available in the NCBI Gene Expression Ominibus
754 repository, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15301 (microarray
755 data) and the EMBL-EBI, https://www.ebi.ac.uk/ena/data/view/ERP008917 (genotype
756 data).

757

758 References

759 1. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-
760 wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*
761 **106**, 9362-7 (2009).
762 2. Nicolae, D.L. *et al.* Trait-associated SNPs are more likely to be eQTLs:
763 annotation to enhance discovery from GWAS. *PLoS Genet* **6**, e1000888
764 (2010).
765 3. Albert, F.W. & Kruglyak, L. The role of regulatory variation in complex traits
766 and disease. *Nat Rev Genet* **16**, 197-212 (2015).
767 4. Gilad, Y., Rifkin, S.A. & Pritchard, J.K. Revealing the architecture of gene
768 regulation: the promise of eQTL studies. *Trends Genet* **24**, 408-15 (2008).
769 5. Gibson, G., Powell, J.E. & Marigorta, U.M. Expression quantitative trait locus
770 analysis for translational medicine. *Genome Med* **7**, 60 (2015).
771 6. Kang, H.J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature*
772 **478**, 483-9 (2011).
773 7. Mele, M. *et al.* Human genomics. The human transcriptome across tissues and
774 individuals. *Science* **348**, 660-5 (2015).
775 8. Jennings, C.G. *et al.* Opportunities and challenges in modeling human brain
776 disorders in transgenic primates. *Nat Neurosci* **19**, 1123-30 (2016).
777 9. Rogers, J. & Gibbs, R.A. Comparative primate genomics: emerging patterns of
778 genome content and dynamics. *Nat Rev Genet* **15**, 347-59 (2014).
779 10. Jasinska, A.J. *et al.* Systems biology of the vervet monkey. *ILAR J* **54**, 122-43
780 (2013).
781 11. Huang, Y.S. *et al.* Sequencing strategies and characterization of 721 vervet
782 monkey genomes for future genetic analyses of medically relevant traits.
783 *BMC Biol* **13**, 41 (2015).
784 12. Stein, J.L. *et al.* Identification of common variants associated with human
785 hippocampal and intracranial volumes. *Nat Genet* **44**, 552-61 (2012).
786 13. Jasinska, A.J. *et al.* Identification of brain transcriptional variation reproduced
787 in peripheral blood: an approach for mapping brain expression traits. *Hum*
788 *Mol Genet* **18**, 4415-27 (2009).
789 14. Warren, W.C. *et al.* The genome of the vervet (Chlorocebus aethiops sabaeus).
790 *Genome Res* **25**, 1921-33 (2015).
791 15. Arnett, M.G., Muglia, L.M., Laryea, G. & Muglia, L.J. Genetic Approaches to
792 Hypothalamic-Pituitary-Adrenal Axis Regulation. *Neuropsychopharmacology*
793 **41**, 245-60 (2016).
794 16. McEwen, B.S., Gray, J.D. & Nasca, C. 60 YEARS OF NEUROENDOCRINOLOGY:
795 Redefining neuroendocrinology: stress, sex and cognitive and emotional
796 regulation. *J Endocrinol* **226**, T67-83 (2015).
797 17. Nestler, E., Hyman, S., Holtzman, D. & Malenka, R. *Molecular*
798 *Neuropharmacology: A Foundation for Clinical Neuroscience*, (McGraw-Hill
799 Education / Medical, 2015).
800 18. Caceres, M., Suwyn, C., Maddox, M., Thomas, J.W. & Preuss, T.M. Increased
801 cortical expression of two synaptogenic thrombospondins in human brain
802 evolution. *Cereb Cortex* **17**, 2312-21 (2007).

803 19. Gaujoux, R. & Seoighe, C. CellMix: a comprehensive toolbox for gene
804 expression deconvolution. *Bioinformatics* **29**, 2211-2 (2013).
805 20. Zhang, Y. *et al.* An RNA-sequencing transcriptome and splicing database of
806 glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci* **34**, 11929-
807 47 (2014).
808 21. Yu, Q. & He, Z. Comprehensive investigation of temporal and autism-
809 associated cell type composition-dependent and independent gene
810 expression changes in human brains. bioRxiv doi: 10.1101/065292 (2016).
811 22. Almasy, L. & Blangero, J. Multipoint quantitative-trait linkage analysis in
812 general pedigrees. *Am J Hum Genet* **62**, 1198-211 (1998).
813 23. Mahler, N. *et al.* Gene co-expression network connectivity is an important
814 determinant of selective constraint. *PLoS Genet* **13**, e1006402 (2017).
815 24. Bogomolov, M., Peterson, C.B., Benjamini, Y. & Sabatti, C. Testing hypotheses
816 on a tree: new error rates and controlling strategies. arXiv:1705.07529v1
817 [stat.ME] (2017).
818 25. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic
819 risk for schizophrenia. *Nat Neurosci* **19**, 1442-1453 (2016).
820 26. Maurano, M.T. *et al.* Systematic localization of common disease-associated
821 variation in regulatory DNA. *Science* **337**, 1190-5 (2012).
822 27. Tung, J., Zhou, X., Alberts, S.C., Stephens, M. & Gilad, Y. The genetic
823 architecture of gene expression levels in wild baboons. *Elife* **4**(2015).
824 28. Vermunt, M.W. *et al.* Epigenomic annotation of gene regulatory alterations
825 during evolution of the primate brain. *Nat Neurosci* **19**, 494-503 (2016).
826 29. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human
827 epigenomes. *Nature* **518**, 317-30 (2015).
828 30. Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**,
829 554-66 (2015).
830 31. Young, R.S. *et al.* The frequent evolutionary birth and death of functional
831 promoters in mouse and human. *Genome Res* **25**, 1546-57 (2015).
832 32. Daugherty, M.D., Schaller, A.M., Geballe, A.P. & Malik, H.S. Evolution-guided
833 functional analyses reveal diverse antiviral specificities encoded by IFIT1
834 genes in mammals. *Elife* **5**(2016).
835 33. Pierce, B.L. *et al.* Mediation analysis demonstrates that trans-eQTLs are often
836 explained by cis-mediation: a genome-wide analysis among 1,800 South
837 Asians. *PLoS Genet* **10**, e1004818 (2014).
838 34. Fears, S.C. *et al.* Identifying heritable brain phenotypes in an extended
839 pedigree of vervet monkeys. *J Neurosci* **29**, 2867-75 (2009).
840 35. Mattick, J.S. & Rinn, J.L. Discovery and annotation of long noncoding RNAs.
841 *Nat Struct Mol Biol* **22**, 5-7 (2015).
842 36. Ulitsky, I. & Bartel, D.P. lincRNAs: genomics, evolution, and mechanisms. *Cell*
843 **154**, 26-46 (2013).
844 37. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of
845 genetic association studies using summary statistics. *PLoS Genet* **10**,
846 e1004383 (2014).

847 38. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx)
848 pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60
849 (2015).
850 39. Wang, J. *et al.* Imputing Gene Expression in Uncollected Tissues Within and
851 Beyond GTEx. *Am J Hum Genet* **98**, 697-708 (2016).
852 40. Sargiannidou, I. *et al.* Connexin32 mutations cause loss of function in
853 Schwann cells and oligodendrocytes leading to PNS and CNS myelination
854 defects. *J Neurosci* **29**, 4736-49 (2009).
855 41. Bergoffen, J. *et al.* Connexin mutations in X-linked Charcot-Marie-Tooth
856 disease. *Science* **262**, 2039-42 (1993).
857 42. Bond, J. *et al.* ASPM is a major determinant of cerebral cortical size. *Nat Genet*
858 **32**, 316-20 (2002).
859 43. Tang, B.S. *et al.* Small heat-shock protein 22 mutated in autosomal dominant
860 Charcot-Marie-Tooth disease type 2L. *Hum Genet* **116**, 222-4 (2005).
861 44. Eriksson, P.S. *et al.* Neurogenesis in the adult human hippocampus. *Nat Med*
862 **4**, 1313-7 (1998).
863 45. van Praag, H. *et al.* Functional neurogenesis in the adult hippocampus. *Nature*
864 **415**, 1030-4 (2002).
865 46. Pichlmair, A. *et al.* IFIT1 is an antiviral protein that recognizes 5'-
866 triphosphate RNA. *Nat Immunol* **12**, 624-30 (2011).
867 47. Brodziak, F., Meharg, C., Blaut, M. & Loh, G. Differences in mucosal gene
868 expression in the colon of two inbred mouse strains after colonization with
869 commensal gut bacteria. *PLoS One* **8**, e72317 (2013).
870 48. Sato, Y. *et al.* Cellular Transcriptional Coactivator RanBP10 and Herpes
871 Simplex Virus 1 ICP0 Interact and Synergistically Promote Viral Gene
872 Expression and Replication. *J Virol* **90**, 3173-86 (2016).
873 49. Azevedo, C. *et al.* The RAR1 interactor SGT1, an essential component of R
874 gene-triggered disease resistance. *Science* **295**, 2073-6 (2002).
875 50. Mayor, A., Martinon, F., De Smedt, T., Petrilli, V. & Tschopp, J. A crucial
876 function of SGT1 and HSP90 in inflammasome activity links mammalian and
877 plant innate immune responses. *Nat Immunol* **8**, 497-503 (2007).
878 51. Naitza, S. *et al.* A genome-wide association scan on the levels of markers of
879 inflammation in Sardinians reveals associations that underpin its complex
880 regulation. *PLoS Genet* **8**, e1002480 (2012).
881 52. Bakken, T.E. *et al.* A comprehensive transcriptional map of primate brain
882 development. *Nature* **535**, 367-75 (2016).
883 53. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**,
884 15-21 (2013).
885 54. Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with
886 high-throughput sequencing data. *Bioinformatics* **31**, 166-9 (2015).
887 55. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package
888 for differential expression analysis of digital gene expression data.
889 *Bioinformatics* **26**, 139-40 (2010).
890 56. Andersen, C.L., Jensen, J.L. & Orntoft, T.F. Normalization of real-time
891 quantitative reverse transcription-PCR data: a model-based variance

892 estimation approach to identify genes suited for normalization, applied to
893 bladder and colon cancer data sets. *Cancer Res* **64**, 5245-50 (2004).
894 57. Heath, S.C., Snow, G.L., Thompson, E.A., Tseng, C. & Wijsman, E.M. MCMC
895 segregation and linkage analysis. *Genet Epidemiol* **14**, 1011-6 (1997).
896 58. Jasinska, A.J. *et al.* A genetic linkage map of the vervet monkey (Chlorocebus
897 aethiops sabaeus). *Mamm Genome* **18**, 347-60 (2007).
898 59. Kang, H.M. *et al.* Variance component model to account for sample structure
899 in genome-wide association studies. *Nat Genet* **42**, 348-54 (2010).
900 60. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic
901 estimation of expression residuals (PEER) to obtain increased power and
902 interpretability of gene expression analyses. *Nat Protoc* **7**, 500-7 (2012).
903 61. Peterson, C.B., Bogomolov, M., Benjamini, Y. & Sabatti, C. Many Phenotypes
904 Without Many False Discoveries: Error Controlling Strategies for Multitrait
905 Association Studies. *Genet Epidemiol* **40**, 45-56 (2016).
906 62. Simes, R.J. An improved Bonferroni procedure for multiple tests of
907 significance. *Biometrika* **73**, 751-754 (1986).
908 63. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical
909 and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical*
910 *Society. Series B (Methodological)* **57**, 289-300 (1995).
911 64. Benjamini, Y. & Bogomolov, M. Selective inference on multiple families of
912 hypotheses. *J. R. Stat. Soc. B* **76**, 297-318 (2014).
913 65. Peterson, C.B., Bogomolov, M., Benjamini, Y. & Sabatti, C. TreeQTL:
914 hierarchical error control for eQTL findings. *Bioinformatics* **32**, 2556-8
915 (2016).
916 66. Chen, H. *et al.* Control for Population Structure and Relatedness for Binary
917 Traits in Genetic Association Studies via Logistic Mixed Models. *Am J Hum*
918 *Genet* **98**, 653-66 (2016).
919
920

921
922 **Figure Legends**

923 Fig. 1. PCA of 1,000 genes with the most variable expression levels. Analysis was
924 performed separately by tissue; sample size was 60 animals for adrenal, blood,
925 fibroblasts, and pituitary and was 59 for BA46, caudate, and hippocampus. Numbers
926 in the labels for x and y axes indicate the proportion of total variance accounted for
927 by that PC.

928 Fig. 2. Boxplot of log counts per million (CPM) expression in samples of BA46 from
929 58 animals vs. timepoint, for three genes with a strong relationship between
930 expression pattern and age. The inter-quartile range defines the height of the box,
931 and whiskers extend to 1.5x the inter-quartile range. Outliers are indicated as
932 individual points. In each box, the median is represented by the horizontal black bar.

933 Fig. 3. Master regulatory locus on vervet chromosome CAE 9. Upper panel: Ensembl
934 view of the CAE 9 region. Lower panel: The minimum –log10(p-value) for each SNP
935 in association analyses vs. expression in 347 animals of microarray probes on
936 different chromosomes. The symbols are color-coded to represent the number of
937 probes significantly associated to each SNP: 1-2 probes (black), 3-4 probes (yellow),
938 5-6 probes (blue), 7-10 probes (green), 11-14 probes (red). Symbols indicate the p-
939 value from analysis of expression in Dataset 2 (RNA-Seq). Cross: $p<2.35e-05$; X:
940 $p<0.001$; circle: $p>0.001$. The large red X at the top of the plot is CAE9_82694171.

941 Fig. 4. Hippocampal volume QTL and local hippocampal volume eQTLs in RNA-Seq
942 analysis. Top panel: purple dotted line is the multipoint LOD score for hippocampal
943 volume (measured in 347 animals). Circles represent evidence for association of
944 SNPs to hippocampal expression in 58 animals of three genes: *LOC103222765* (red),
945 *LOC103222769* (blue) and *LOC103222771* (green). Solid circles indicate genome-
946 wide significant associations. The region between the black vertical lines is blown
947 up in the middle and bottom panels. The horizontal dotted line represents the
948 genome-wide significant threshold for local eQTLs. Middle panel: SNPs with –
949 log10(p-value)>8 for association to expression in hippocampus, color codes are as
950 in the top panel. Bottom panel: Genes sited between 68.7 and 69 Mb (the eQTL
951 region). Color codes are as in the top panel.

952 Fig. 5. Correlation in 16 animals of hippocampal volume (MRI) with hippocampal
953 expression of *LOC103222765* (left), *LOC103222769* (middle) and *LOC103222771*
954 (right). The expression data are from qRT-PCR. Quantification was performed using
955 the relative standard curve method, with the geometric mean of the reference gene
956 *HPRT1* used as an endogenous control for normalization of the interpolated lncRNA
957 quantities. Hippocampal volume measurements are residuals from a regression on
958 covariates of age and sex. "r" is the Pearson correlation coefficient, and the p-value
959 tests the null hypothesis that r=0.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements. Thanks to Stephanie Groman for assistance with tissue resources. Thanks to Tara Chavanne, Kelsey Finnie, Margaret Long, Jean Gardin and Dianna Swaim for technical assistance with necropsies and tissue collections. This work was supported by the following grants, all from the U.S. National Institutes of Health: U54HG00307907 (to RKW); P40RR019963/OD010965 (to JRK); R01RR016300/OD010980 (to NBF); R37MH060233 (to Daniel Geschwind); UL1DE019580 (to Robert Bilder); PL1NS062410 (to Christopher Evans); RL1MH083270 (to JDJ); P30NS062691 (to NBF and GC); R01MH101782 (to CS and EE). RN, BLA, PF acknowledge support from Wellcome Trust PF, BLA, RN acknowledge support from the Wellcome Trust (grant number WT108749/Z/15/Z) and the European Molecular Biology Laboratory. ESW was supported by an EMBO Advanced Fellowship (aALTF1672-2014).

Author contributions: AJJ, JRK, GMW, KD, RKW, JDJ, WW, RPW, and NBF designed the study. AJJ, IZ, OWC, JD, LAF, SF, AEF, YSH, VR, CAS, JDJ, GC, and RPW produced the data. AJJ, IZ, SKS, CP, RMC, EE, LAF, SF, YSH, VR, CAS, HS, DV, BLA, PF, RN, ESW, JB, TDD, MB, YB, CS, and GC analyzed the data. OWC, JD, and MJJ managed data and samples. AJJ, SKS, and NBF wrote the paper. All authors reviewed the final draft.

Author Information:

Reprints and permissions information is available at www.nature.com/reprints

The authors declare that they have no competing financial interests

Correspondence and requests for materials should be addressed to nfreimer@mednet.ucla.edu

Figure 1

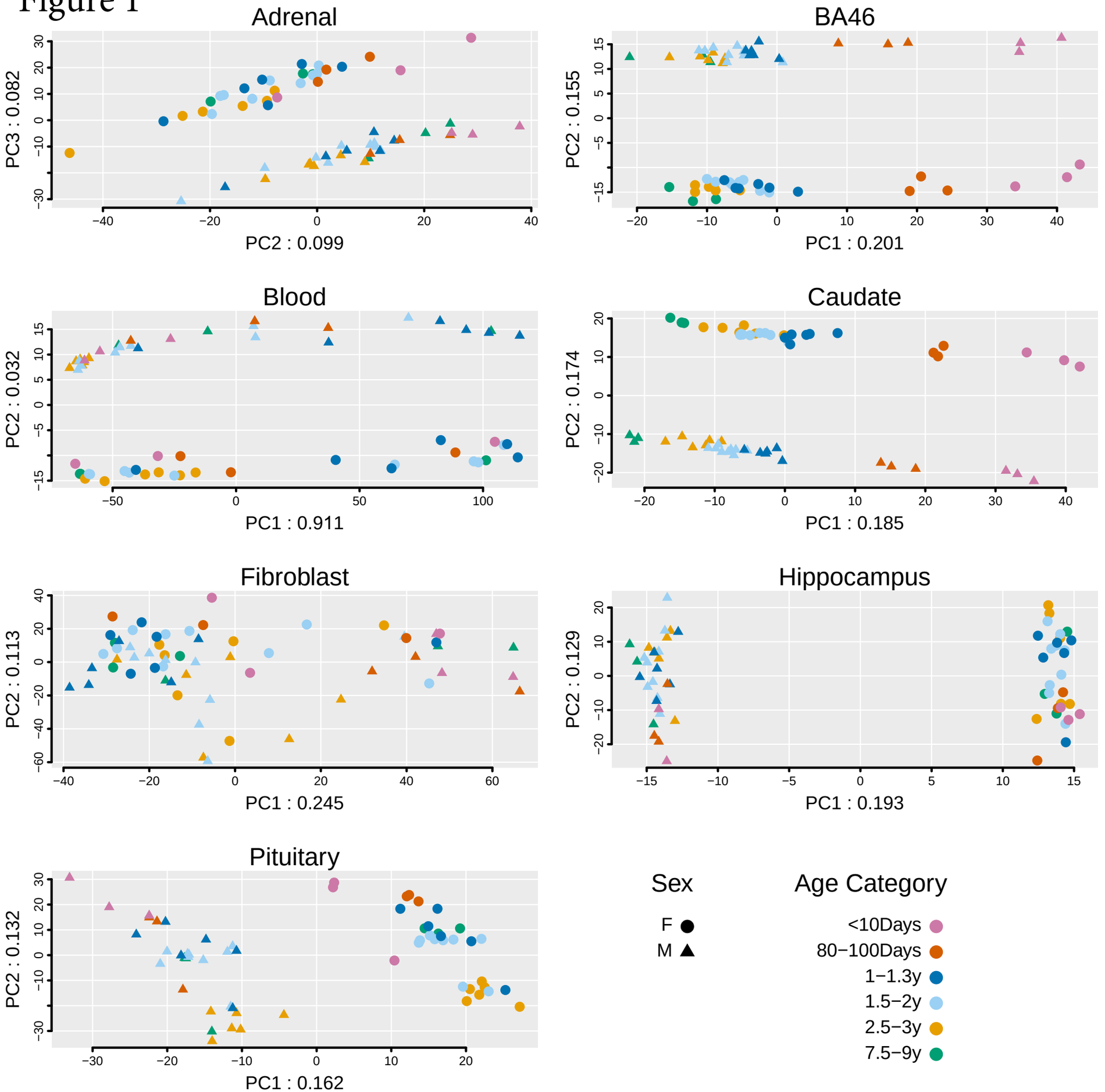

Adrenal
PC3 : 0.082
PC2 : 0.099
BA46
PC2 : 0.155
PC1 : 0.201
Blood
PC2 : 0.032
PC1 : 0.911
Caudate
PC2 : 0.174
PC1 : 0.185
Fibroblast
PC2 : 0.113
PC1 : 0.245
Hippocampus
PC2 : 0.129
PC1 : 0.193
Pituitary
PC2 : 0.132
PC1 : 0.162
Sex
F
M
Age Category
<10Days
80−100Days
1−1.3y
1.5−2y
2.5−3y
7.5−9y

Figure 2

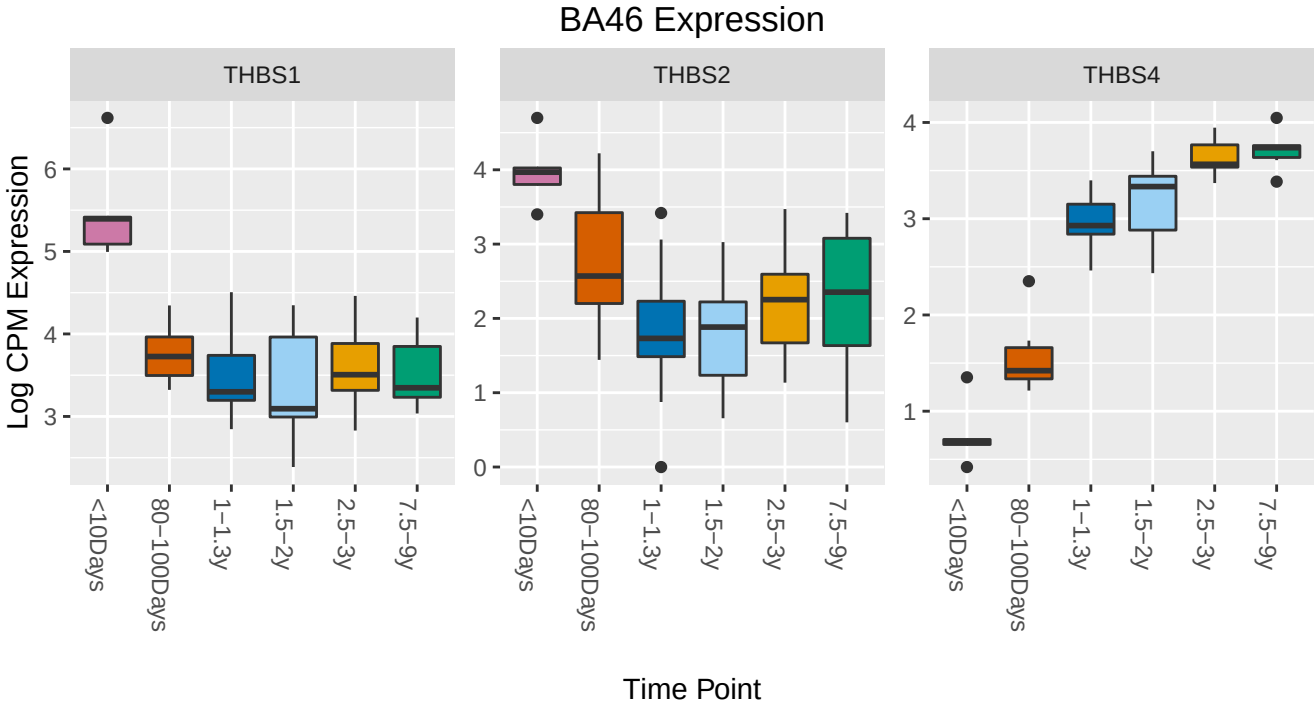
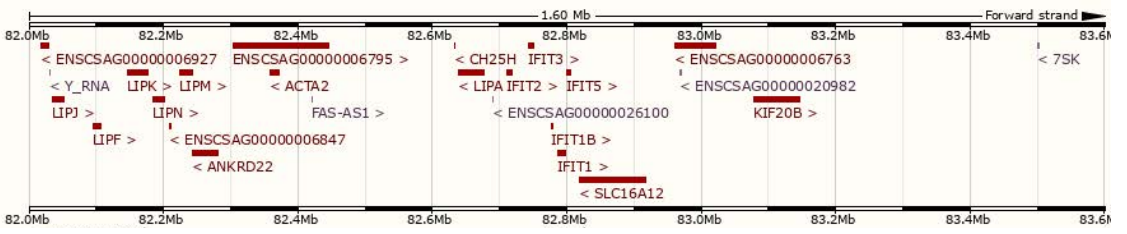

BA46 Expression
THBS1
THBS2
THBS4
Log CPM Expression
Time Point
<10Days
80−100Days
1−1.3y
1.5−2y
2.5−3y
7.5−9y

Figure 3


1.60 Mb
Forward strand
82.0Mb
82.2Mb
82.4Mb
82.6Mb
82.8Mb
83.0Mb
83.2Mb
83.4Mb
< ENSCSAG00000006927
ENSCSAG00000006795 >
< CH25H
IFIT3 >
< ENSCSAG00000006763
< 7SK
< Y_RNA
LIPK >
LIPM >
< ACTA2
< LIPA
IFIT2 >
IFIT5 >
< ENSCSAG00000020982
LIPJ >
LIPN >
FAS-AS1 >
< ENSCSAG00000026100
KIF20B >
LIPF >
< ENSCSAG00000006847
IFIT1B >
< ANKRD22
IFIT1 >
< SLC16A12

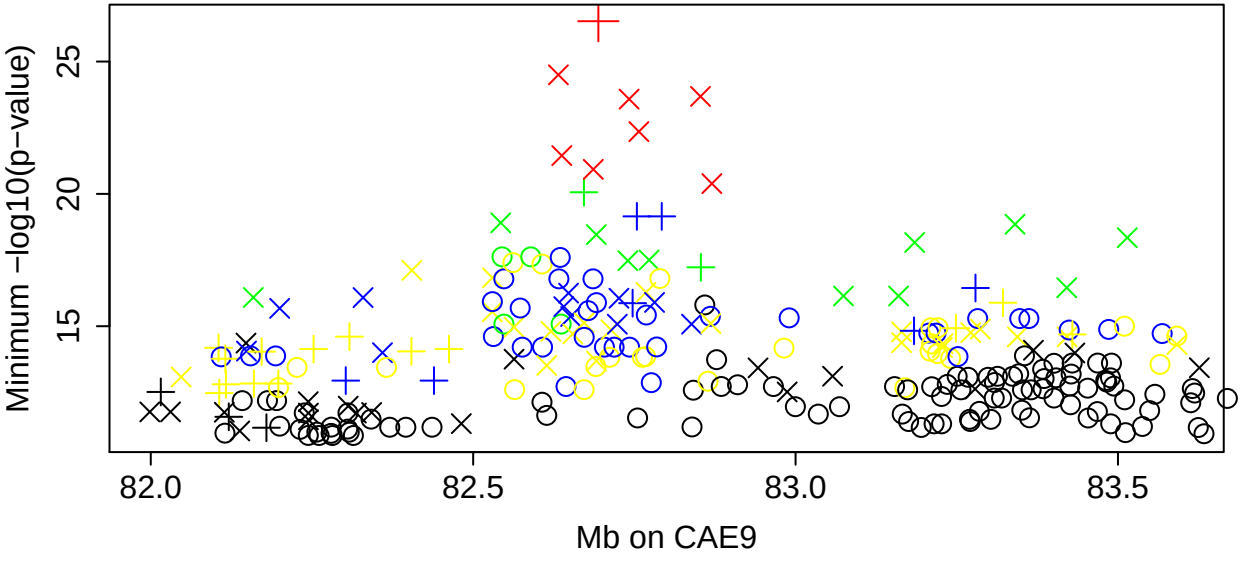

Minimum −log10(p−value)
15
20
25
82.0
82.5
83.0
83.5
Mb on CAE9

Figure 4

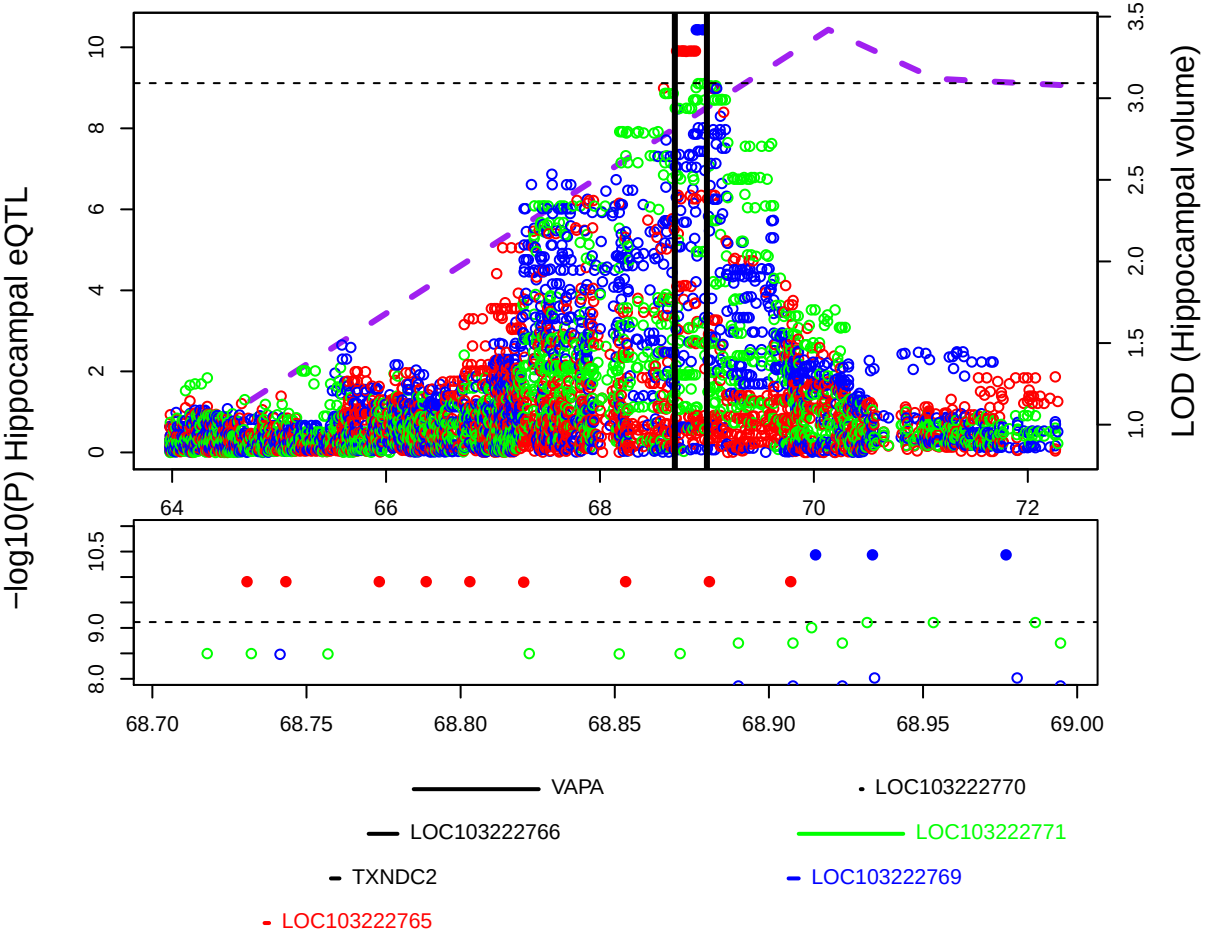

−log10(P) Hippocampal eQTL
LOD (Hippocampal volume)
0
2
4
6
8
10
1.0
1.5
2.0
2.5
3.0
3.5
64
66
68
70
72
8.0
9.0
10.5
68.70
68.75
68.80
68.85
68.90
68.95
69.00
VAPA
LOC103222770
LOC103222766
LOC103222771
TXNDC2
LOC103222769
LOC103222765

Figure 5

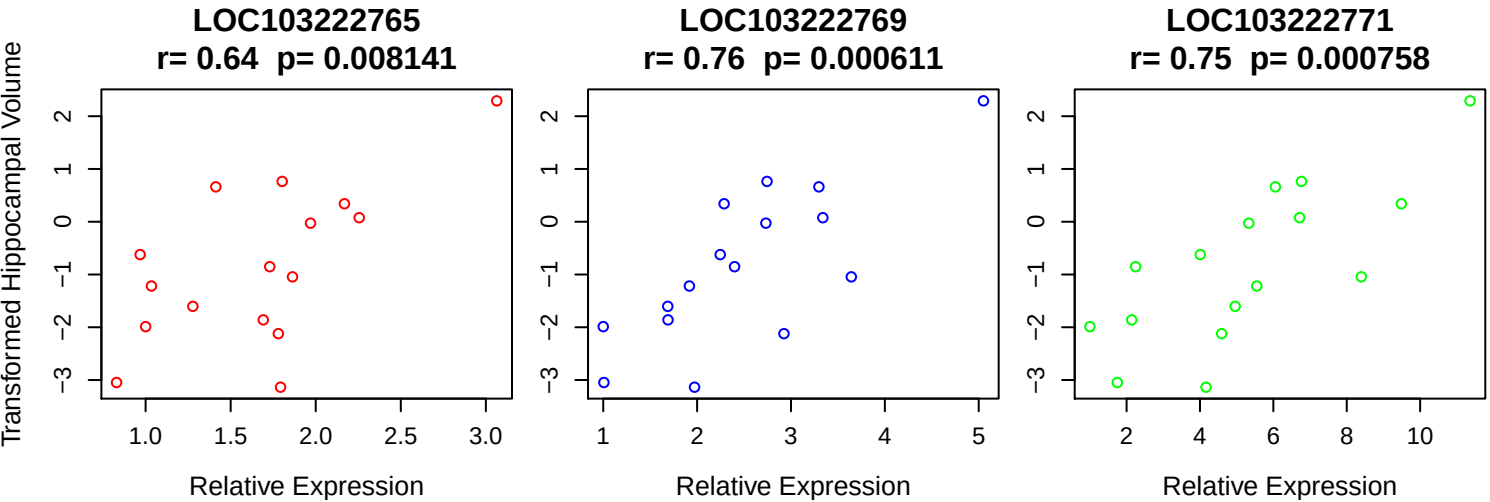

LOC103222765
r= 0.64 p= 0.008141
LOC103222769
r= 0.76 p= 0.000611
LOC103222771
r= 0.75 p= 0.000758
Transformed Hippocampal Volume
Relative Expression
Relative Expression
Relative Expression

Table 1. Gene expression data sets. The number of probes/genes with at least one significant local and distant eQTL (at Bonferroni corrected thresholds) are presented. We have 80% power to detect distant eQTLs accounting for 15% of the variability in expression in Dataset 1 and 66% of the variability in Dataset 2

Tissue	Probes/genes analyzed[a]	Local eQTL[b]	Distant eQTL[c]	%Distant eQTL on same chr
Dataset 1: Microarray				
Blood	3,417	461	215	80.8%
Dataset 2: RNA-seq				
Adrenal	25,187	555	80	54.5%
BA46	27,530	307	30	81.8%
Blood	33,776	60	4	100%
Caudate	28,249	441	47	69.0%
Fibroblast	22,328	239	43	33.2%
Hippocampus	26,957	361	45	70.6%
Pituitary Gland	27,236	596	80	77.5%

[a]microarray dataset (Dataset 1) with an initial set of 22,184 probes on Illumina HumanRef-8 v2 (6,018 probes passed filters described in Supplementary Table 1; 3,417 were heritable); RNA-seq (Dataset 2) with an initial set of 33,994 genes annotated in vervet

[b]Local eQTL are eQTL that are within 1 Mb of the gene. Bonferroni threshold for Dataset 1: 4.8×10^{-8}; Bonferroni threshold for Dataset 2: 6.5×10^{-10}

[c]Distant eQTL are more than 1 Mb away from the gene, and may be on the same or a different chromosome. Bonferroni threshold for Dataset 1: 1.5×10^{-11}; Bonferroni threshold for Dataset 2: 5.3×10^{-13}

Table 2 Comparison of specific genes with local eQTL in Vervet Dataset 2 to GTEx. The number of genes with at least one significant local eQTL in Vervet (at FDR thresholds) are presented.

Tissue	Vervet number of individuals	# Local eQTL Vervet Genes[a]	GTEx number of individuals	GTEx number of eGenes[a]	# Vervet Genes with Human Ortholog	# Orthologous Genes Tested in GTEx[b]	% Tested Genes p<0.05	% Tested Genes p <.05/# tested Genes[c]	% Tested Genes significant genome-wide in GTEx[d]
Adrenal	58	2932	126	2915	1828	1674	100%	28.7%	18.2%
Blood	58	574	338	5438	264	229	100%	70.7%	38.9%
Caudate	57	3140	100	2396	1737	1548	100%	24.6%	14.1%
Hippocampus	58	2437	81	1405	1436	1296	100%	18.4%	9.2%
Pituitary	58	3395	87	2222	1863	1743	100%	20.7%	13.0%

[a]The number of eGenes found in the multi-tissue hierarchical FDR procedure applied to vervet Dataset 2 and to GTEx.
[b]Vervet genes with a human ortholog that were not tested in GTEx were filtered by their QC procedures
[c]The threshold for significance corrected for the number of genes compared between Vervet and GTEx (column 7).
[d]Genes were declared significant by GTEx at an FDR of 0.05.