UCLA
UCLA Previously Published Works

Title

Identification of brain transcriptional variation reproduced in peripheral blood: an approach for mapping brain expression traits

Permalink

https://escholarship.org/uc/item/36b331gd

Journal

Human Molecular Genetics, 18(22)

ISSN

0964-6906

Authors

Jasinska, Anna J
Service, Susan
Choi, Oi-wa
et al.

Publication Date

2009-11-15

DOI

10.1093/hmg/ddp397

Peer reviewed

Identification of brain transcriptional variation reproduced in peripheral blood: an approach for mapping brain expression traits

Anna J. Jasinska[1], Susan Service[1], Oi-wa Choi[1], Joseph DeYoung[1], Olivera Grujic[1], Sit-yee Kong[1], Matthew J. Jorgensen[3], Julia Bailey[1], Sherry Breidenthal[1], Lynn A. Fairbanks[1], Roger P. Woods[4], J. David Jentsch[2] and Nelson B. Freimer[1,*]

[1]Center for Neurobehavioral Genetics and [2]Department of Psychology, University of California, Los Angeles, CA, USA, [3]Department of Pathology, Section on Comparative Medicine, Wake Forest University Health Sciences, Winston-Salem, NC, USA and [4]Departments of Neurology and Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA

Genome-wide gene expression studies may provide substantial insight into gene activities and biological pathways differing between tissues and individuals. We investigated such gene expression variation by analyzing expression profiles in brain tissues derived from eight different brain regions and from blood in 12 monkeys from a biomedically important non-human primate model, the vervet (*Chlorocebus aethiops sabaeus*). We characterized brain regional differences in gene expression, focusing on transcripts for which inter-individual variation of expression in brain correlates well with variation in blood from the same individuals. Using stringent criteria, we identified 29 transcripts whose expression is measurable, stable, replicable, variable between individuals, relevant to brain function and heritable. Polymorphisms identified in probe regions could, in a minority of transcripts, confound the interpretation of the observed inter-individual variation. The high heritability of levels of these transcripts in a large vervet pedigree validated our approach of focusing on transcripts that showed higher inter-individual compared with intra-individual variation. These selected transcripts are candidate expression Quantitative Trait Loci, differentially regulating transcript levels in the brain among individuals. Given the high degree of conservation of tissue expression profiles between vervets and humans, our findings may facilitate the understanding of regional and individual transcriptional variation and its genetic mechanisms in humans. The approach employed here—utilizing higher quality tissue and more precise dissection of brain regions than is usually possible in humans—may therefore provide a powerful means to investigate variation in gene expression relevant to complex brain related traits, including human neuropsychiatric diseases.

INTRODUCTION

Gene expression measures provide an intermediate phenotype linking underlying genetic variation with higher order phenotypes observed at the cellular, organ or organism level. Such intermediate phenotypes are particularly important for efforts to understand the genetic basis of heritable human neurobiological traits; the etiological complexity of such traits has continued to frustrate genetic mapping efforts.

The use of gene expression profiles as intermediate phenotypes has been motivated by the demonstration that transcript levels from healthy human subjects, as measured in peripheral blood mononuclear cells and derivative lymphoblastic cell lines, are highly heritable traits which are usually stable over

*To whom correspondence should be addressed at: UCLA Center for Neurobehavioral Genetics, Gonda Center, Room 3506, 695 Charles E. Young Drive South, PO Box 951761, Los Angeles, CA 90095, USA. Tel: +1 3107949571; Fax: +1 3107949613; Email: nfreimer@mednet.ucla.edu

time (1–3). Numerous expression Quantitative Trait Loci (eQTL) involved in natural inter-individual differences in regulation of gene activity have been detected in blood cells (4). Identification of *trans*-regulatory elements controlling gene expression at different genomic locations and *cis*-regulatory elements involved in local expression regulation have provided the basis for better understanding transcriptional regulatory processes.

Inter-individual variation in cortical gene expression has been correlated with genome-wide genotypic variation (5). It is, however, extremely difficult to assemble sufficiently large samples of human brains to enable well-powered eQTL mapping studies relevant to brain and behavioral phenotypes. Such mapping studies, therefore, depend on a systematic strategy for identifying brain-related transcripts that demonstrate high inter-individual variability and that can be assessed in readily accessible surrogate tissues. The accessibility of blood makes it an attractive substitute for brain tissues, but there is so far little direct evidence for its suitability for this purpose.

Studies of different types of acute brain injury in rodents have revealed overlapping gene expression profiles between brain and blood (6–8). In humans, the utility of blood as a surrogate for brain tissues has been supported mainly through indirect investigations of expression profiles in relation to specific neurological and psychiatric diseases (9–13). Such studies have shown that, in both brain and blood samples, expression patterns distinguish between samples drawn from patients or control individuals.

Much less data are available from direct comparisons of brain and blood expression patterns. One such study (14) observed a correlation of about 0.5 in transcript levels among all genome-wide transcripts expressed both in brain and whole blood, including substantial correlations in transcripts representing putative candidate genes for schizophrenia. Further studies are needed to extend our database of transcripts whose levels of expression are reliably concordant between brain and blood. Such information is necessary for the development of large scale studies aiming to map eQTL that are relevant to our understanding of brain and behavioral traits.

We describe here the results of direct comparisons of transcript levels between blood and brain samples in a set of related, apparently healthy vervet monkeys, members of an extended pedigree [the Vervet Research Colony (VRC)] in a species (*Chlorocebus aethiops*) widely employed as a model system in biomedical research. The recent divergence time (23–25 million years) between hominoids and Old World monkeys (15) is reflected in extensive similarities in neurophysiology and neuroanatomy, and in conservation of genomic sequences (~94% sequence identity between humans and vervets) that enables the use of human-specific tools to assess gene expression in Old World monkeys.

The key factors in performing reliable comparisons between blood and brain tissue gene expression are the collection of samples under uniform, well-controlled conditions, the use of a stringent protocol of tissue dissection and sample preservation, and the use of high-quality RNA for gene expression studies. Each of these goals is more readily achievable in a non-human primate model than in human studies. Large-scale gene expression profiling of human brain samples is limited by the frequently poor quality of preserved material, and inconsistencies between specimens in term of tissue dissection and preservation, factors which may have contributed to inconsistent or discrepant results in efforts to replicate gene expression profiling studies of neuropsychiatric phenotypes (9,16–18). Additionally, environmental variables, such as medication use, that may substantially influence gene expression—are much more difficult to control for in investigations of humans compared with non-human primates.

Our study has taken advantage of several features of the VRC, to maximize the reliability of expression comparisons between brain and blood, and to use the information from such comparisons to design well-powered eQTL mapping studies. We were able to obtain replicate measurements of blood expression from several monkeys, and then to collect from the same monkeys, multiple tissues, including brain samples, under highly controlled conditions. The colony management practices minimize environmental variability between the monkeys. These features enabled us to (i) characterize the extent to which inter-individual variation in brain gene expression—of transcripts selected using stringent criteria—is reflected in peripheral blood and (ii) to hypothesize that such variation largely reflects genetic variation. Using such a stringent approach, we identify an initial set of transcripts whose expression levels show low intra-individual variation and large inter-individual variation and are sufficiently heritable to make them candidate expression phenotypes for eQTL mapping.

RESULTS

We determined gene expression profiles across nine tissues—eight brain regions [cerebellar vermis, pulvinar, head of caudate, hippocampus, occipital pole, orbital frontal cortex, frontal pole, dorsolateral prefrontal cortex (DLPFC)] and peripheral blood—in 12 male vervets. Having transcript measures from different tissues from the same individuals allowed us to evaluate, for each transcript, sources of transcript level variation within and between individuals (Fig. 1). We further focused on two classes of transcripts characterized by high variation of expression across brain regions or high variation between individuals. High spatial and temporal inter-individual variation determined, respectively, between brain tissues and blood and between independent blood samplings, allowed us to investigate heritable brain gene expression traits in peripheral blood. The brain and blood gene expression data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (19) and are accessible through GEO Series accession number GSE15301 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15301).

To evaluate the impact of probe-target sequence incompatibility in our data set, we compared the number of probes widely detected in vervet brain tissues (at least 80% of tissues) and in 193 human cortical samples analyzed by Myers *et al*. (5), using 6791 probes that are in common between the two data sets. At this threshold, 2410 probes were detected in vervet brain and 4622 probes were detected in human cortex. There was considerable overlap in the
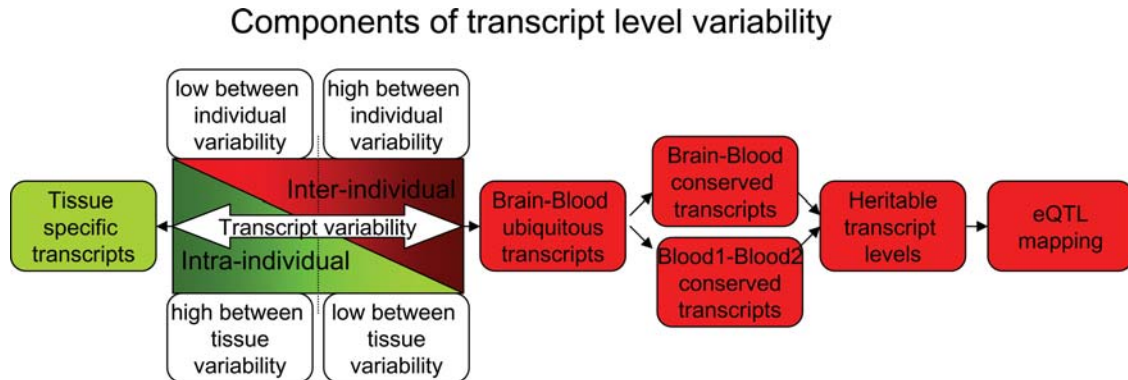
Components of transcript level variability

Figure 1. Components of transcript level variation. Observation of transcript levels across different tissues and individuals allows analysis of variation of transcript levels from the perspective of its two components: inter-individual variation (green) and intra-individual or between tissue (red) variation. Transcripts characterized by much higher intra-individual than inter-individual variation provide insight into the functional relationships between different tissues and may highlight functions attributable to specific tissues. Transcripts characterized by much higher inter-individual than intra-individual variation are of interest as candidates for mapping brain eQTL using blood as a surrogate tissue. Red boxes on the right represent the probe filtering steps leading to selecting transcripts for eQTL mapping characterized by high spatial and temporal inter-individual variation.

probes detected by both studies. In spite of human–vervet sequence differences, 46% (2128/4622) of the probes detected in human cortex were also detected in vervet brain, whereas 88% (2128/2410) of the probes detected in vervet brain were detected also in human cortex, showing that human–vervet sequence differences do not prevent reliable detection of a substantial fraction of the brain transcripts.

Transcripts differentially detected between brain tissues

We used the detection status of all 22 184 probes represented on the Illumina HumanRef-8 version 2 chip across the tissues sampled in tissue set one to determine relationships in gene expression between eight brain regions. Distances between tissues were estimated on the basis of the probes that showed the most striking differences in terms of number of shared detections. Most such probes were either detected in only one tissue or detected in all or most brain tissues. Similarities between all possible pairs of tested tissues are illustrated by a heat map (Fig. 2A). Cortical tissues generally show the greatest similarity within this sample set, but three genes (*KREMEN1, MED13L* and *ZMYM6*) differentiate orbital frontal cortex from frontal pole and one gene (*POLE*) differentiates DLPFC and frontal pole. As reflected in a corresponding dendrogram showing relations between all brain tissues (Fig. 2B), neocortical regions cluster close to hippocampus and are more distantly linked to caudate and pulvinar tissues, with respect to the number of detected samples. Cerebellar vermis and pulvinar tissues are more distant from other brain tissues in terms of the number of differentially detected transcripts on the dendrogram.

When tissues are ranked according to the number of region specific detections (Table 1), cerebellar vermis is first, followed by head of the caudate. Cerebellar vermis shows a tissue-specific presence of 38 transcripts and absence of 54 transcripts. Transcripts with decreased detection in this tissue are significantly enriched for genes associated with developmental processes and coding calcium binding proteins. In the head of the caudate, 22 transcripts are preferentially

detected. Of these transcripts, 18 have a decreased number of detections and are enriched for genes associated with neuronal activities (2.97E−04). Additionally, we grouped together three frontal regions clustering on the bottom of the tissue dendrogram—frontal pole, DLPFC and orbital frontal cortex—and compared them against all other brain tissues, identifying two transcripts (*GPR120* and *RASGRF2*) as specifically expressed in these frontal regions.

Ubiquitously detected transcripts differentially expressed between blood and brain

We focused on transcripts that are ubiquitously detected across tissues and individuals, as these transcripts provide a means to investigate brain-related biology using peripheral blood samples. Of the 22 184 probes on the array, we identified 2481 probes—representing 2430 genes—for which expression was detected in all 12 individuals in all eight brain regions and in blood (Supplementary Material, Table S1). The regional mean expression levels of these ubiquitously expressed 2430 genes were examined in all pairwise comparisons of brain tissues, and the number of differentially expressed probes used as a distance metric in a hierarchical clustering analysis (Supplementary Material, Fig. S2). The relations between tested tissues based on the number of ubiquitously detected transcripts with differential expression levels are mostly concordant with relationships determined based on differentially detected transcripts (previous section) and with relationships known from human studies (20). The only exception to this observation is the clustering together of occipital pole and pulvinar patterns, while hippocampus unexpectedly localizes closer to frontal cortex than does occipital cortex (Supplementary Material, Fig. S2B). Unlike the tissue grouping based on extreme differences in detection count, this tissue clustering is based on a group of ubiquitously detected transcripts that is clearly depleted of transcripts showing the most extreme differences in expression between tissues, such as those whose expression is restricted to a specific tissue.
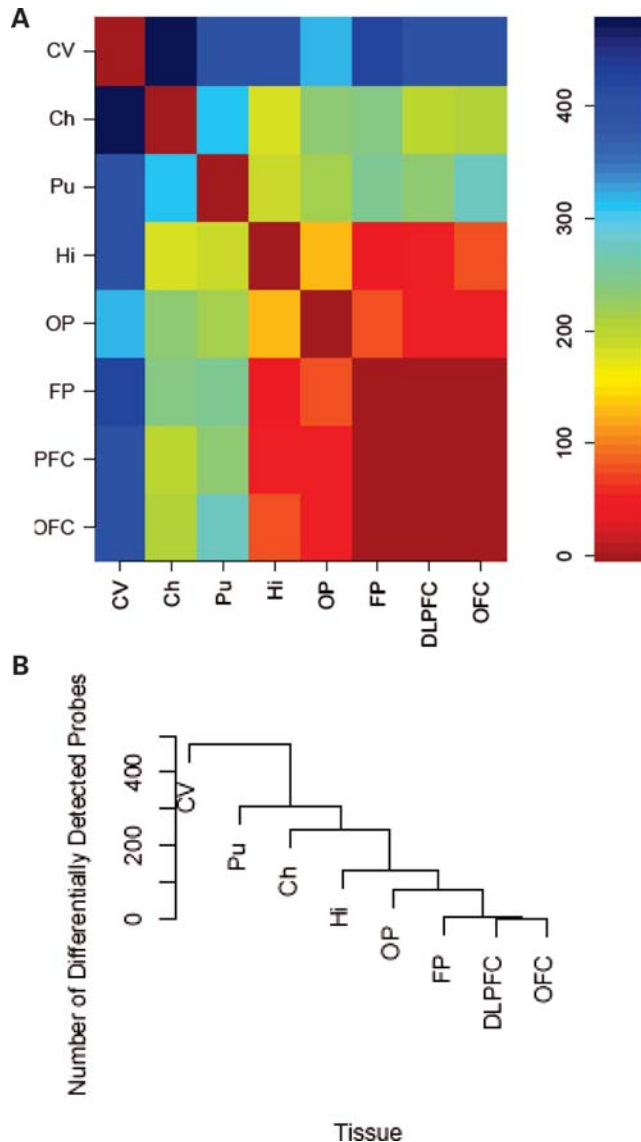
Figure 2. Gene expression differences between brain tissues. Pairwise comparison between all eight brain regions is presented on a heat map (**A**) constructed based on differentially detected transcript. Corresponding hierarchical clustering of tissues is presented on a dendrogram (**B**). Tissue labeling: CV—cerebellar vermis, Pu—pulvinar, Hi—hippocampus, OP—orbital pole, FP—frontal pole, DLPFC—dorsolateral prefrontal cortex, OFC—orbital frontal cortex.

Exclusion of the tissue-specific transcripts generates unexpected topography in the tissue dendrogram, possibly indicating that these transcripts determine important regional features to a greater extent than ubiquitously expressed transcripts. Nevertheless, the ubiquitously expressed transcripts still may suggest important aspects of the biology of different brain regions; as many as 1474 ubiquitously expressed genes show increased or decreased expression levels in specific brain regions, based on pairwise comparisons of mean expression values between tissues (Supplementary Material, Fig. S2A). Consistent with the suggestion from the detection based distance measure, the frontal cortical region shows smaller inter-tissue differences than other tissues tested. Additionally,

a differential mean probe signal was observed for 90 probes between DLPFC and frontal pole, 92 probes between DLPFC and orbital frontal cortex, and 215 probes between frontal pole and orbital frontal cortex.

Among the 2430 ubiquitously expressed genes, we identified genes showing region—specific differential expression, and we determined functional categories that were under—and over-represented among these genes (Supplementary Material, Table S2). Consistent with the results of the differential detection measures, cerebellar vermis, and head of caudate show the largest differences from other tissues in the number of genes showing differential mean expression levels (790 and 383, respectively). Genes differentially expressed between cerebellar vermis and all other tissues are associated with two interrelated functions, metabolic processes of nucleic acids and mRNA transcription. These two biological processes are significantly over-represented ($6.4\text{E}-11$, $7.06\text{E}-03$, respectively) among 533 up-regulated genes and under-represented ($1.00\text{E}-05$, $3.84\text{E}-02$) among 257 down-regulated genes, which may indicate distinctive transcriptional regulation mechanisms acting in this brain region. In head of caudate, a group of 244 up-regulated genes includes structural proteins of small and large units of cytoplasmatic (21) and mitochondrial (5) ribosomes. As a result, among the genes preferentially expressed in this tissue, molecular function of ribosomal proteins is significantly over-represented ($2.16\text{E}-04$), suggesting that specific protein synthesis mechanisms and regulation may be characteristic of this brain region. Among genes down-regulated (139) in head of caudate, there is an over-representation of kinases ($4.05\text{E}-02$).

Genes ubiquitously detected in brain and blood tissues

Among 2481 transcripts that are ubiquitously detected in brain and blood tissues, 2430 also show differential expression between tissues. This group of 2430 transcripts does not include transcripts showing the most extreme differences in cross-tissue expression patterns, such as transcripts expressed exclusively on a single region. A large number of such differentially expressed transcripts (1474) are still widely detected across brain and blood tissues. This set of genes is clearly depleted of genes related to many brain-specific functions including signal transduction or neurogenesis, while it is enriched (with *P*-values less than 0.05) in genes related to the ubiquitin proteasome system, Parkinson disease (e.g. *SNCA*, synuclein-alpha) and Ras pathways as well as genes involved in the various biological processes and molecular functions related to mRNA and protein metabolism (Supplementary Material, Table S3).

Over representation of genes involved in basic biological processes is consistent with a high representation of housekeeping (HK) genes which, by definition, are widely expressed in numerous human tissues and are involved in maintenance of basal cellular functions. More than 44% of recognized HK genes (255/575) are ubiquitously detected in vervet brain and blood tissues. Among these, HK genes are genes widely used in quantitative RT–PCR as control genes whose expression is assumed to be constant among samples (21,22). It is therefore noteworthy that from the genes utilized as endogenous controls, *ARHGDIA*, *POLR2A* and *RPS18*

Table 1. Region-specific transcripts identified based on differential detection

Brain region	Genes with increased regional number of detections	Genes with decreased regional number of detections
Cerebellar vermis	38: *ARHGEF10L, BARHL1, C12orf28, CD2AP, CDH15, CDH23, CERKL, CLDN3, CNNM4, CTRL, DNAJC12, <u>E2F7</u>, EBF3, ECM2, EN2, ERN1, FAT2, <u>GABRA6</u>, ITGA11, KCND1, KIAA0889, KRT72, MRGPRF, <u>MRPL43</u>, NPFF, NRK, NUPR1, ONECUT1, P2RX1, PCP2, PGAM2, PLD5, RAB37, RAB4A, RPS6KA1, <u>TFAP2B</u>, <u>THRSP</u>, ZNF662*	BP[a]: ↑developmental processes 1.63E−02
		MF[a]: ↑select calcium binding protein 3.57E−02 54: *ADAMTS8, ALDH4A1, BAI1, BCL11A, C1orf165, C1QL1, C22orf25, C3orf54, C6orf117, C9orf117, CDH8, CHSY-2, CPNE4, CRABP1, CSNK1G1, CUGBP2, <u>CXCL14</u>, DDN, DUSP15, EFEMP1, EGFL7, ERF, FBLN1, FLJ33590, FZD8, GPR26, KCNB2, <u>KCNQ3</u>, <u>LHX2</u>, LOC730413, <u>LRRC7</u>, LY6H, MAFB, NELF, NELL1, NR2E1, OAF, P11, PCDH10, PDE2A, <u>PNKD</u>, PRG2, PTGS1, RAB11FIP4, RAI14, <u>RASAL2</u>, RASD2, RPE65, <u>SAMD5</u>, SDCBP2, SH3PX3, ST3GAL1, TNNI3, TSHZ2*
Pulvinar	15: *ARHGAP24, <u>C1orf76</u>, CYP26A1, FLJ45455, IGSF1, LHX9, NKD1, PLAC2, PNOC, SEMA4G, SLC25A24, TCF7L2, TMEM41A, TSPAN2, VEGFC*	8: *ADPRHL1, ARHGAP26, BAIAP2, C20orf103, CCKBR, RYR2, SLC16A11, TMOD1*
Head of caudate	22: *AMBN, CKM, DRD1, DRD2, DRD2, ERAF, GPR52, GPR88, HCRT, ISL1, MYH3, NPPB, PTPN7, RGS9, RXRG, SFN, TMC8, TMEM16B, TMEM45A, UPB1, UPP2, VTCN1*	BP[a]: ↑neuronal activities 2.97E−04
		18: *ACVR2A, BDNF, CLIP1, <u>CYP26B1</u>, <u>ELAVL2</u>, FAM80A, <u>GALNT9</u>, GEFT, GK, GRM2, LIN7A, MEX3B, MKX, NTNG1, PTPN3, SLC17A6, SNCB, TBKBP1*
Hippocampus	2: *LXN, NRG1*	
Occipital pole	2: *LAG3, OSTN*	
Frontal regions[b]	2: *GPR120, RASGRF2*	1: *CA12*

The comparisons were done on a pairwise basis between tissues, and a probe was differentially detected in two tissues if in one tissue it was detected in less than two subjects and in the other tissue it was detected in nine or more subjects. This table is based on all 22 184 transcripts. Underlined—transcripts exclusively detected in all samples from a given brain region while was absent in all samples form other brain regions or transcripts absent in all samples from a given tissue while detected in all samples from other brain regions.
[a]Statistically significant functional associations of Panther ontology terms with listed gene transcripts; BP, biological process; MF, molecular function.
[b]Orbital frontal cortex, DLPFC and frontal pole versus all other tissues.

showed differential regional expression and therefore do not suit the purpose of tissue-to-tissue normalization. As expected, this group of HK genes is enriched for genes involved in maintenance of constitutive functions such as protein and mRNA metabolism, ribosomal activity, energy release and cytoskeletal regulation (Supplementary Material, Table S4).

Here among the genes broadly expressed in brain and blood tissues, we detected a considerable number of ubiquitous transcripts that were previously reported as stable across tissues HK genes. Analysis of blood and high-quality brain tissues from precisely dissected regions revealed that levels of almost all these ubiquitous transcripts vary between at least one pair of tested tissues in the current study.

Variation of transcript profiles in brain and blood tissues

To identify probes showing correlated expression profiles in brain and blood tissues, we estimated, for each brain region, the Spearman rank correlation (SRC) between the paired brain and blood expression measures. Additionally, to identify probes with greater inter-individual variation than intra-individual variation, we used a variance components approach to estimate the percent variation (PV) attributable to the inter-individual component and within monkey component (between tissues). Both measures (SRC and PV) for the whole brain and blood data set are available on the Integrated Vervet Monkey Genomics website: http://genomequebec.mcgill.ca/compgen/submit_db/vervet_web, which enables searches for similarities of expression profiles between eight brain regions and blood for specific genes. Examples of brain–blood expression patterns are shown in Supplementary Material, Figure S3, for highly correlated profiles (left column), moderately correlated profiles (middle column) and poorly correlated profiles (right column).

Among the 2481 widely detected probes, 825 show PV >0.55 and SRC >0.55. For the 2481 ubiquitously detected probes, the number of brain regions with PV >0.55 or with SRC >0.55 was skewed in favor of excluding a majority of probes and varied from zero brain regions (PV: $n = 1329$, SRC: $n = 1574$) to all eight brain regions (PV: $n = 52$; SRC: $n = 28$). The probes exceeding these PV and SRC thresholds for each brain region are shown in Supplementary Material, Table S5. Correlated expression patterns and inter-individual variation (according to the above criteria) in all eight brain regions were attributable to 23 genes: *SPOCK2, HSBP1, CLN3, CIRBP, ANXA11, PNKP, CCT2, RPS20, GGA2, SCO1, CCDC115, DDOST, DUSP11, MRPL51, DMTF1, RPL31, RPL35A, LARP5, SS18L2, TUBA1B, C9orf114, SRF, MRPS17.* Even though these genes displayed correlated patterns across all brain tissues, ten of them showed regionally increased or decreased expression levels, for example, three-ribosomal genes *RPS20, MRPL51* and *RPL35A* are up-regulated in the head of the caudate in comparison to all other brain tissues. Supplementary Material,

Table S6 shows the number of probes that met PV and SRC criteria for comparison between blood and each brain region. Both methods consistently identified hippocampus as the brain region most dissimilar to peripheral blood with regard to transcriptional profiles, but did not show considerable variation among other brain tissues. Probes that have passed the PV and SRC thresholds, for at least one brain region, are probes that show similar profiles in brain and blood expression and have more inter-monkey variation than intra-monkey variation. Therefore, these selected probes can be further studied in blood samples to identify brain gene expression traits.

For investigating specific regional brain functions, genes expressed only in one or a few brain regions only are of interest. Such regional specificity, however, is likely to predict lack of expression of transcripts in peripheral tissues and therefore the impossibility of using correlated brain and blood expression patterns to guide studies of such transcripts in peripheral blood. We verified that only three transcripts in our data set (*LHX1*, *PPP1R1B* and *RGS9*) meet both tissue-specific detection and brain–blood correlation criteria.

Variation of transcript profiles in peripheral blood in replicate samples

To assess the reproducibility of gene expression profiling, which could be affected by either technical variation or transcript level stability over time, we used expression data set two, derived from 18 monkeys each sampled twice for blood. From the initial list of 22 184 probes, we identified 1880 probes that were detected in peripheral blood in all 36 replicate samples from 18 monkeys. We required detection in all replicates in order to identify the most reliable transcripts for future eQTL mapping in blood samples. The group of 1839 genes represented by these 1880 probes is greatly enriched for genes involved in protein and mRNA metabolism and various signaling pathways, as well as pathways characteristic of peripheral blood such as lymphocyte activation and inflammation (Supplementary Material, Table S7). Among the most overrepresented biological processes in this set of genes are those implicated in oxidative phosphorylation, apoptosis, immunity and defense, and cell cycle, structure and motility. Nucleic acid binding, and ribosomal and cytoskeletal functions are the molecular functions most enriched among these genes.

To determine the biological reproducibility over time of the gene expression profiles in these 1880 probes, we assessed the within monkey versus between monkey (between duplicate samplings) variance, including sex as a fixed effect in the model. Examples of the similarity in expression signal between replicate samples is shown in Supplementary Material, Figure S4, for highly correlated replicates (left column), moderately correlated replicates (middle column) and poorly correlated replicates (right column). There were significant differences (at the 0.05 level, uncorrected for multiple testing) between males and females in blood expression data for 238 of the 1880 probes that passed the detection threshold.

For each of the 18 vervets with duplicate blood expression measures, correlation of duplicate measures across all probes was at least 0.9 (range across the 18 vervets was 0.89–0.99). To identify transcripts with stable levels in blood but differentially expressed between monkeys, we focused on

transcripts which had more variation between monkey than within monkey (temporal transcript variation in blood and technical reproducibility) as defined by the percent of total variance attributable to the within monkey component PV > 55%. Among the 1880 probes that passed the detection threshold in peripheral blood replicate samples, there were 134 probes (representing 133 genes) with PV > 0.55, indicating that for these probes the majority of the total variation in probe signal was between monkey rather than within monkey (between replicates over time). Both high inter-individual variation and intra-individual reproducibility make these selected genes suitable candidates for genetic mapping of their eQTL.

Selection of candidate transcripts for eQTL mapping

We merged results from data sets one and two to select expression traits for future eQTL mapping. We examined the 2481 probes that passed detection thresholds in brain and blood expression data set one to identify probes that passed PV and SRC thresholds for both brain–blood similarity (825) and the PV threshold for biological reproducibility (Fig. 3A) from data set two (130). We identified 53 of 2491 probes that met both these criteria, i.e. having correlated expression profiles in brain and blood (for at least one brain region) and showing more inter-individual variation than intra-individual (between tissues) variation. Next, we limited the list of probes to the probes that meet the detection threshold (36 measurements) for the biological reproducibility data set (Fig. 3B). The reduced subset of 32 probes passing all variation, reproducibility and detection thresholds is presented in Table 2.

Significant correlation and PV of the *TUBA1B* transcript and also PV of additional transcripts (*BAT1*, *C19orf62*, *EIF1* and *SUV420H1*) was observed between all brain regions and blood, suggesting a common regulatory mechanism acting across brain tissues. Region-specific correlation of brain expression with blood was observed in caudate (*EIF1*), cerebellar vermis (*SMOX* and *TSPAN14*), DLPFC (*SLC25A23*), frontal pole (*RAB5A*), hippocampus (*STOM*) and orbital frontal cortex (*ERAL1* and *TFE3*); no such correlation was observed specifically for occipital pole and pulvinar.

Using transcript levels measured in the peripheral blood of 347 individuals (tissue set three) from the extended vervet pedigree with known and genetically confirmed structure, we estimated the heritability of the 32 selected transcripts (Fig. 3B, Table 2). Twenty-nine of these expression traits showed heritability at a significance level less than 0.05 and 25 transcripts showed heritability at $P < 0.001$; 62.5% of these traits displayed heritability estimates of ≥ 0.4. The generally high heritability of these transcripts suggests that selecting transcripts whose inter-individual variation in transcript levels is greater than their intra-individual variation, identifies transcripts whose regulation has a strong genetic component. Evidence consistent with this hypothesis is provided by much lower estimates of heritability in a comparison set of 32 transcripts chosen randomly from the data set (data not shown).

Polymorphisms located within a probe sequence may cause differential hybridization which mimics differential expression results and lead to inaccurate estimates of the heritability of levels of particular transcripts. To assess the effects of such
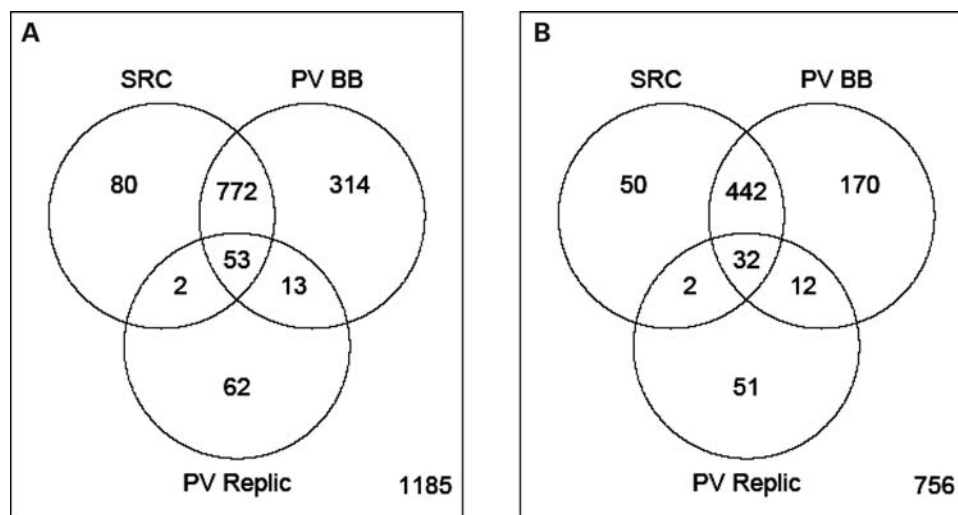
Figure 3. Selection of candidate transcripts for mapping brain eQTL in peripheral blood. The diagram represents the set of probes in the brain–blood gene expression comparison that passed the 55% threshold for PV (PV BB) and the 0.55 threshold for the SRC for any of the brain regions, and the set of probes that passed the 55% threshold for PV in blood replicate samples PV Replic. (**A**) All 2481 probes that passed detection thresholds in brain and blood expression data. Considering just the brain–blood similarity analysis, 825 probes (33%) had both PV > 0.55 and SRC > 0.55. (**B**) The subset of 1515 probes that also passed detection thresholds in the replicate blood sample data set. Although the absolute number of probes is reduced according to detection criteria, the distribution of number of probes to pass PV and SRC thresholds in both brain–blood similarity and the PV threshold for biological reproducibility is very similar; for example 474 probes (31%) had both PV > 0.55 and SRC > 0.55.

polymorphisms on our results, we sequenced probes for 16 transcripts showing significant heritability in the vervet monkey pedigree. Ten of these probes were monomorphic. For the six probes in which we detected SNPs (*TUBA1B*, *TMED3*, *BAT1*, *CDKN1A*, *C19orf62* and *SMOX*), we compared expression level measures between different SNP genotype classes to observe possible correlations between signal intensity and genotype. Four of these six probes showed marked correlation between signal intensity and genotype, raising the possibility that probe hybridization properties rather than differential gene regulation are responsible for observed inter-individual variation in these transcripts. A probe for the *SMOX* transcript did not show such signal intensity–genotype correlation, most likely due to the low frequency of the minor allele. The probe for the *CDKN1A* transcript showed consistent differences between genotypes in seven tested tissues but not in hippocampus and pulvinar. This observation suggests that the *CDKN1A* probe is sensitive to transcript level, but confirmation of this interpretation will require use of an alternative gene expression assay such as quantitative real-time PCR.

Fourteen of the heritable transcripts showed differential expression levels across brain regions. This group includes transcripts specifically elevated in cerebellar vermis (5), head of caudate (3) and hippocampus (1). It will be of interest to examine further the genetic determinants of regional gene expression that may be involved in specific tissue functions.

DISCUSSION

We are developing genome-wide gene expression resources for the vervet that permit investigation of variation between different tissues within an individual or between individuals. Genome-wide gene expression studies may provide comprehensive insight into gene activities and biological pathways differing between various tissues and individuals. Brain regional differences in gene expression levels relate to specific functions of brain tissues including disease symptoms distinctively affecting specific brain regions, while variation in brain gene expression profiles among individuals indicates possible genetic factors regulating gene transcript levels. Our research addressed both kinds of gene expression variation—between brain regions and between individuals—by expression profiling in brain tissues derived from eight brain regions and blood from 12 vervet monkeys.

Regional brain expression profiles in vervet

We employed the vervet monkey as a non-human primate model to assure precisely dissected high-quality brain tissues, which are difficult to obtain from human subjects. Gene expression profiles from distinct vervet brain regions are generally comparable to profiles of equivalent human brain regions, as indicated by the similarity of the hierarchical structure of tissue dendrograms (20). On the basis the observed overall conservation of gene expression profiles between vervet and human brain regions, we expect that identification of genes that are preferentially expressed or that show exclusive expression in specific vervet brain regions may provide a means to identify genes with region-specific functions in humans. Several of these genes are already known to be involved in region-specific functions, for example the caudate-specific transcript RGS9, which is inversely correlated with striatal dopamine metabolism (23).

Candidates for mapping eQTL

The primary focus of our study was to identify correlated brain and peripheral blood transcriptional profiles that could be reproducibly measured over time. We selected

Table 2. Thirty-two candidate genes for mapping brain eQTL in peripheral blood

Probe ILMN_	Gene symbol	Entrez Gene name	Brain expression[a]	H2r P-value	H2r	Covariate[b]	Relation to disease
1657857[c]	*TMEM14C*	Transmembrane protein 14C		7.32E−22	0.71	B	
1764813[c]	*B3GALTL*	Beta 1,3-galactosyltransferase-like	CV↓	6.98E−14	0.78	SAB	Peters-plus syndrome (MIM#261540)
1669878	*GUSB*	Glucuronidase, beta		1.01E−13	0.64	B	Mucopolysaccharidosis VII (MIM #253220)
1719316[d]	*TMED3*	Transmembrane emp24 protein transport domain containing 3		2.35E−13	0.64	SB	
1686626[d]	*BAT1*	HLA-B associated transcript 1	CV↑	4.11E−13	0.60	AB	Asthma, myocardial infarction[e]
1812325	*TMEM111*	Transmembrane protein 111	CV↓	9.48E−13	0.75	B	
1800261[d]	*TUBA1B*	Tubulin, alpha 1b		8.23E−11	0.65	SAB	
1718831	*TMEM57*	Transmembrane protein 57	CV↑	1.19E−10	0.58	SAB	
1768031[c]	*DEDD2*	Death effector domain containing 2	CV↑OP↓	7.92E−09	0.56	SAB	
1784602[d]	*CDKN1A*	Cyclin-dependent kinase inhibitor 1A		1.01E−08	0.47	B	Cancer[e]
1692026[c]	*SUV420H1*	Suppressor of variegation 4-20 homolog 1	CV↑	8.71E−08	0.43	None	
1708516[c]	*PRTFDC1*	Phosphoribosyl transferase domain containing 1		1.02E−07	0.44	SAB	
1806937[d]	*C19orf62*	Chromosome 19 open reading frame 62		1.92E−07	0.51	SAB	
1679949	*SLC25A23*	Solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 23	Ch↑	4.31E−07	0.43	SA	
1686099	*ZFYVE1*	Zinc finger, FYVE domain containing 1		6.54E−07	0.43	B	
1718706	*ERAL1*	Era G-protein-like 1		6.93E−07	0.47	B	
1726456	*SLC3A2*	Solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2	CV↓	9.37E−07	0.46	AB	
1766657[c]	*STOM*	Stomatin	CV↓	9.41E−07	0.45	B	
1761176	*GRIPAP1*	GRIP1 associated protein 1		5.65E−06	0.34	B	
1775380[d]	*SMOX*	Spermine oxidase		2.97E−05	0.41	SB	
1785060	*TSPAN14*	Tetraspanin 14		7.95E−05	0.41	SA	
1764826[c]	*TFE3*	Transcription factor binding to IGHM enhancer 3		1.81E−04	0.28	AB	Renal carcinoma (MIM#605074)
1808875	*RAB5A*	RAB5A, member RAS oncogene family	CV↑Ch↓	2.29E−04	0.35	B	
1712312[c]	*RAB11A*	RAB11A, member RAS oncogene family	H↑	2.52E−04	0.34	AB	
1779735[c]	*LOC389541*	Chromosome 7 open reading frame 59		6.56E−04	0.31	SB	
1739641[c]	*MTMR3*	Myotubularin related protein 3	CV↓	3.17E−03	0.28	B	
1737074	*LOC440589*	Hypothetical LOC440589	CV↓ Ch↑	1.24E−02	0.22	B	
1662905	*NME1-NME2*	NME1−NME2 read-through transcript	CV↓Ch↑	2.28E−02	0.20	B	Cancer[e]
1716260	*HSPA8*	Heat shock 70 kDa protein 8		3.86E−02	0.16	SB	Kidney failure[e]
1803846	*EIF1*	Eukaryotic translation initiation factor 1		7.51E−02	0.17	SB	
1682299	*NDUFA11*	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 11		9.05E−02	0.14	SAB	
1695420	*CLTA*	Clathrin, light chain (Lca)	Ch↓	3.68E−01	0.03	B	

[a]Genes ↑up- or ↓down-regulated in a single-brain region compared with all other regions, CV, cerebellar vermis; Ch, head of caudate; OP, occipital pole; H, hippocampus.
[b]Important covariates: A, age; S, sex; B, batch.
[c]Probe region uninterrupted by SNPs in vervet.
[d]Probe region containing SNPs in vervet.
[e]Genetic Association Database (GAD).

stable blood biomarkers of brain gene expression that may warrant further investigation of expression eQTL differentially regulating transcript levels in the brain across individuals. Our approach to select promising candidates for mapping brain eQTL using blood as a surrogate tissue included very restrictive criteria for probe filtering. We focused on transcripts with both ubiquitous tissues expression and with much higher inter-individual than intra-individual variation, which resulted in selecting a relatively small number of candidate transcripts (32). However, both the high proportion of heritable transcripts in this selected set (90%) and their high estimated heritabilities demonstrate the utility of this approach for identifying transcripts whose

variation has a strong genetic component. These heritable expression phenotypes can be genetically mapped using the VRC pedigree, which has proven to be an efficient tool for linkage studies (24,25). It is hypothesized that heritable variations in transcript levels are more directly related to underlying genetic variation than many other nervous system traits, and therefore these molecular phenotypes are attractive candidates for genetic mapping. Indeed several studies using recombinant inbred mouse strains have clearly demonstrated the utility of eQTL studies focused on brain tissues for mapping of complex traits related to these tissue (27–30).

We consider that the 32 transcripts highlighted by this study provide an initial set of candidate transcripts for brain-related

eQTL mapping using peripheral blood measurements. It is likely that further investigations—using larger samples and more powerful technologies for measuring gene expression—will identify a much larger set of transcripts suitable for eQTL analyses. We anticipate, for example, that such investigations may incorporate a substantial number of transcripts for which we found evidence suggestive of brain–blood correlation in expression levels but that we now exclude due to the stringency of our selection criteria. We further emphasize that the use of such rigorous criteria currently precludes us from drawing inferences regarding the overall degree of such brain–blood correlations among the transcripts assessed in this study.

Utility of vervet as a translational model

By applying rigorous inclusion criteria, we identified transcripts whose stable profiles in blood also reflect their profiles in brain. These transcripts therefore provide easily accessible biomarkers of brain gene expression variation. The substantial known similarities between human and vervet in regional brain expression suggest that the transcripts identified here as stably correlated between brain and blood in vervet may be similarly informative in human studies that assay inter-individual variation of expression in peripheral blood. Information from this type of investigation of vervets or other non-human primates could be very valuable, given that reliable brain–blood gene expression comparisons in human subjects are limited by the availability and quality of matched blood and brain tissues.

Studies of expression quantitative traits using brain–blood expression profile correlation in the vervet model also have wider implications for understanding complex human neurobehavioral traits. Several quantitative variables that are relevant to such traits are the focus of active genetic investigation in the vervet, including variation in neuroanatomic features (26), neurochemistry (24) and disease-related behaviors such as impulsivity (27). The opportunity to investigate these phenotypes—in conjunction with gene expression profiling—longitudinally is another important advantage of such a non-human primate model. The availability of such gene expression profiles may greatly facilitate the identification of the genetic variants underlying variation in such traits.

Technical issues

Two types of technical issues influence the interpretation of our results; the first issue concerns the complexity of the tissues analyzed and the second issue concerns the effect of possible incompatibilities between microarray probe sequences and the transcripts under investigation. Tissue samples derived from anatomically different brain regions, even when carefully dissected, show substantial complexity at a cellular level (28). The diversity of cell types in the brain tissues that we have investigated may influence the results presented here regarding gene expression across brain regions and peripheral blood. Different brain regions show different levels of cellular complexity (29). Such cellular heterogeneity may particularly strongly affect low abundance transcripts which are known to be important contributors to the specificity of brain-related functions (29). An alternative approach to studying overall regional brain expression is high resolution analysis of specific cells isolated from tissues using

laser capture microdissection (30), although this approach is not currently practical for large-scale studies.

The sequence divergences between humans and vervets result in imperfect hybridization which may clearly lower transcript detection. However, our observation that most of the probes that were widely detected in vervet samples were also detected in a similar proportion of human brain samples, demonstrates the overall expression similarities between human and vervet brain and indicates that such false-negative results are unlikely to be a substantial issue in interpreting the findings reported here.

False-positive results may result from genetic variants in probe interaction sites, causing differential probe hybridization between individuals. Although commercial microarrays are designed to minimize the inclusion of known common sequence variants in such sites, several studies have shown that a substantial fraction of probes contain at least one SNP (31,32). In our study, it is not currently possible to estimate, on a genome-wide basis, the degree of vervet genetic variation in the sequences corresponding to human probes. Interpretation of gene expression profiling in the vervet will be enhanced in the near future using data generated by the Vervet Genome Sequencing Project (VGSP), which has been initiated recently. The VGSP is not only determining the vervet reference genomic sequence, but is also cataloging common vervet SNPs on a genome-wide basis.

Until such data are available, targeted resequencing is required to determine whether vervet genetic variation may be influencing particular expression results. We have shown that the majority of heritable transcripts that we have suggested as candidates for eQTL mapping, show true inter-individual variation in expression rather than hybridization artifacts. Although we identified genetic variation in more than a third of the probes that we sequenced, it is not evident that this proportion is substantially different than that observed in experiments where the array and the transcript both represent the same species. Rather, the potential for false-positive results reflects the intrinsic drawbacks of array-based approaches for assessing gene expression. This kind of false-positive effect may be enriched among transcripts that show correlation between blood and all brain regions, whereas correlations observed between blood and a single brain region or only few brain regions are less likely to reflect the effect of a SNP on probe hybridization affinity.

As in other species, we anticipate that investigation of gene expression in the vervet will evolve from array-based to more sensitive and specific sequencing-based approaches. It is likely that we will be able to detect a much larger number of transcripts for which the pattern of expression in blood and brain can be correlated, and therefore which will be suitable candidates for genetic mapping of eQTL relevant to brain. The high-quality tissue and RNA resource that we have established will be extremely valuable for mapping brain expression traits using blood as a surrogate tissue.

MATERIAL AND METHODS

Subjects and tissue samples

Tissue specimens were collected from vervet monkeys that were related members of the VRC pedigree. Three tissue

sets were collected within the VRC. Set one consisted of brain and blood samples collected from 12 monkeys. Set two consisted of replicate blood samples collected from 18 monkeys. Set three consisted of blood samples collected from vervets over 2 years old from the entire pedigree, $n = 347$.

Set one sampled for comparison of gene expression between brain and peripheral blood, was collected from 12 male vervets, aged ~3 years at the time of sample collection. The average kinship coefficient among the 12 monkeys represented in tissue set one was 0.025 (min=0.0035, max=0.154). These subjects were euthanized by ketamine (10 mg/kg, intramuscular), followed by overdose with sodium pentobarbital (30–60 mg/kg, intravenous). Circulatory perfusion with cooled isotonic saline was achieved by cardiac cannulation; after ~15 min of perfusion, the calvarium was opened, the brain was removed and coronally blocked and samples were microdissected on a thermostatic platform held at 4–6°C. Tissue was collected from eight brain regions: cerebellar vermis, pulvinar nucleus, head of the caudate, hippocampus, frontal pole, DLPFC, orbital frontal cortex and occipital pole. Localization of sampled regions is shown in Supplementary Material, Figure S1. A sample of 20–180 mg of tissue from each brain region was collected directly into RNAlater reagent (Ambion), immediately stabilizing the RNA profiles.

Set two, enabling gene expression analysis of biological replicate samples, was collected twice, at 20 weeks intervals, from the peripheral whole blood of 18 VRC animals; 12 males (5.5–7.8 years old, mean 6.4 years old) and six females (4.7–7.6 years old, mean 6.1 years old). For blood collection, animals were anesthetized with ketamine HCl (8–10 mg/kg, intramuscular) and blood was collected via femoral venipuncture. The average kinship coefficient among the 18 monkeys sampled twice was 0.026 (min=0, max=0.204).

Set three, enabling estimation of heritability of gene expression data, was obtained from peripheral blood drawn from 347 monkeys aged 2 years and older. For blood samples, 2.5 ml of peripheral blood was drawn from the femoral vein directly to a PaxGene RNA Blood tube (PreAnalyticX) containing a solution to preserve RNA integrity.

RNA samples

Total RNA from whole blood preserved in PaxGene RNA Blood tubes (PreAnalyticX) was extracted using PAXgene Blood RNA Kit (PreAnalyticX). Total RNA from RNAlater-preserved brain tissues was isolated with PefectPure RNA Cell and Tissue Kit (5 Prime) using 20–40 mg of brain tissue homogenized with rotor-stator (Omni International). Since the most important confounding factor affecting expression levels is total RNA quality, we evaluated the total RNA integrity of all of our tissue types (Supplementary Material, Table S8) using the Agilent 2100 Bioanalyzer with the RNA 6000 Nano Assay Kit (Agilent Technologies). Average RNA integrity number (RIN) for eight brain tissue types ranged from 7.5 ± 0.4 for head of caudate to 8.2 ± 0.5 for cerebellar vermis (Supplementary Material, Table S8). All RIN values were greater than 7. Observed RIN values are generally higher and less variable than those usually

obtained from human post-mortem brain tissues. For example, Lipska *et al*. (33) reported RIN for human tissues of 5.7 ± 1.0 (hippocampus) and 6.7 ± 1.3 (DLPFC), whereas we obtained respective RIN values of 7.6 ± 0.4 and 7.5 ± 0.3 for the same tissue types. For blood samples, our average RIN was 9.1 ± 0.6. Total RNA sample concentrations were quantified with RiboGreen RNA (Invitrogen).

Microarray measurements

For assessing transcript levels, we used the Illumina HumanRef-8 v2 chip. This chip provides genome-wide transcriptional coverage of well-characterized genes. Low representation of splice isoforms on this chip do not, however, allow efficient assessment of splicing QTL (sQTL).

This chip uses 22 184 probes representing 18 189 unique human genes (or 20 424 unique transcripts) from Reference Sequence database1, Release 17. The Illumina gene expression platform utilizes long 50-mer gene-specific probes that provide both good selectivity and sensitivity (34). Although short oligo-nucleotide probes (32-mers or shorter) are very sensitive to probe-target mismatches such as these resulting from sequence variation between individuals, long cDNA probes are less influenced by such mismatches (35–39). We used intermediate size oligonucleotides expecting that these probes would sufficiently tolerate sequence incompatibilities between human probe sequence and vervet target transcripts and be more robust than shorter probes to possible allelo-specific differences in hybridization efficiency due to vervet-specific SNP variants occurring in probe-interaction sites (40).

cDNA was synthesized and *in vitro* transcribed into biotiny-lated cRNA using the Illumina Totalprep RNA amplification kit, following the manufacturer's instructions (Ambion). Labeled cRNA was hybridized to the HumanRef-8 version 2 (Illumina) gene expression bead-chip. A gene was called detectable by BeadStudio when the detection *P*-value was less than 0.01. The gene expression module of the BeadStudio software version 3.1 (Illumina) was used for initial data processing and background correction. Lumi software was also used to perform a variance-stabilizing transformation that takes advantage of the technical replicates available on every Illumina microarray (usually over 30 randomly distributed beads per probe), and subsequently performs robust spline normalization and quality control of gene expression measures (41). Expression profiles from all collected samples passed sample quality checks in Lumi, except for one pulvinar sample that was excluded from further analysis. Therefore, the total brain and blood sample set used for subsequent analysis consisted of 12 samples from blood and from all brain regions except the pulvinar, which was represented by 11 samples in this study.

Probe filtering

For quality control purposes, gene expression data were filtered based on detection scores defined by Illumina (detection $P < 0.01$). We used two criteria to select probes for further analysis: (i) detection in all brain and blood samples from the 12 animals in the brain and blood data set and (ii) detection in all 36 samples from the biological replicate blood data set

(two samples from each of 18 vervets). Although both steps narrowed down the number of transcripts analyzed to select candidates for brain eQTL mapping, they permit selection of transcripts consistently measured across the samples despite possible sequence incompatibilities in cross species hybridizations. By applying very stringent detection criteria, we excluded also many potentially interesting transcripts that were detected in some but not all brain regions. As such transcripts did not show sufficient correlation between the tissues of interest and blood, we chose not to examine them further through pedigree-wide expression profiling in blood.

Variance component and correlation analysis

For each brain region sampled in tissue set one, we used a mixed model ANOVA and a variance components approach to compare the blood expression levels to those in each of the eight brain regions. Each analysis included 12 paired expression measurements, one from blood and one from a given brain region, from each animal. The total variation in expression for any probe can be divided into the between monkey component and the within monkey (between blood and brain tissues within the same monkey) component. Probes where most of the variation is attributable to the between monkey component show more inter-monkey variation than intra-monkey variation, and therefore have less variation between blood and brain tissue than between monkeys. Prior to performing the variance components analysis, brain and blood measures were standardized to Z-scores separately for each tissue type by subtracting the mean and dividing by the standard deviation (both mean and standard deviation were determined across animals), as overall mean levels of expression may differ in brain tissues and blood. We also estimated the SRC between expression levels in blood and expression levels in each brain region.

Variance components analysis was also used in a similar manner on the biological replicates data set in tissue set two. The total variation in expression for any probe can be divided into the between monkey component and the within monkey (between replicates from the same monkey) component. Probes where most of the variation is attributable to the between monkey component show more inter-monkey variation than intra-monkey variation, and therefore have less variation between replicates than between monkeys. As the biological replicates analysis included both male and female monkeys, sex was included as a fixed effect in the model.

Differential gene expression analysis

Probes that are differentially expressed in brain regions in tissue set one were determined using the limma package (42) of the Bioconductor project, a linear modeling approach that uses an empirical Bayes method in which the t-statistic standard errors have been moderated across probes. This procedure effectively borrows information from the ensemble of probes to aid with inference about each individual probe (43). The correlation of repeated measures from the same vervet was addressed using a version of a mixed model analysis that again borrows information across probes (42). All

pairwise contrasts between the nine tissues (eight brain regions and blood) were assessed. The false-discovery rate procedure was used to control for multiple testing. Only probes that were detected in all twelve individuals in all available brain and blood samples were examined for differential expression.

The number of genes that were differentially expressed between a pair of tissues in tissue set one was used as a distance metric, and hierarchical clustering, using complete linkage (44), was applied to this distance matrix to cluster tissues. This clustering analysis used only probes that were detected in all 12 individuals in all brain and blood samples. Probes that were not detected in some tissues would not be used in the cluster analysis, yet these probes may provide interesting information on differences among tissues. Further comparisons of expression in different brain tissues were performed with all 22 184 probes, using, as a distance metric, the number of differentially detected probes in the two tissues. A probe was considered to be differentially detected in two tissues, if the number of animals in which the probe was detected (based on the above described Illumina detection criteria) in one tissue was less than two and the number of animals in the other tissue was nine or more. Probes detected in three to eight individuals in either of the two tissues being compared were not considered differentially detected.

Heritability estimates

Heritability of gene expression in blood of selected transcripts was estimated with data from 347 subjects from the extended vervet pedigree (tissue set three), using a variance component analysis method as implemented in SOLAR (45). Sex, age and sample grouping during microarray experiment (batch) were included as covariates.

Functional category enrichment analysis

Functional category enrichment analysis was performed using the Panther classification tool (http://www.pantherdb.org/) (46). Briefly, this tool maps the gene list of interest to the Panther ontology and compares to the selected reference gene list to identify over- and under-represented terms. The expected value is determined as the number of genes that are expected in the gene list for a given Panther category, based on the term incidence in the reference list. Total of 17 253 of 18 189 HumanRef-8 v2 chip genes were represented in the Panther ontology and we used them as a reference gene list for comparison with lists of selected genes. *P*-values of binomial statistic were obtained for Panther classification categories for each molecular function, biological process and pathway term. Bonferroni correction was applied to correct for multiple testing. Owing to sequence incompatibilities in cross-species hybridization, the lists of genes selected based on gene expression measures may be biased against rapidly evolving genes and therefore functional categories attributable to such genes may appear as underrepresented in our analysis.

Sequencing of probe interacting region

Genomic DNA from 12 monkeys used for gene expression comparison between brain and blood was used to sequence probe

interacting region in 16 heritable transcripts (Table 2). A standard sequencing procedure with BigDye Terminator v3.1 was used to sequence PCR amplicons.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. None declared.

FUNDING

REFERENCES

1. Cheung, V.G., Conlin, L.K., Weber, T.M., Arcaro, M., Jen, K.Y., Morley, M. and Spielman, R.S. (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat. Genet.*, **33**, 422–425.
2. Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M. and Burdick, J.T. (2005) Mapping determinants of human gene expression by regional and genome-wide association. *Nature*, **437**, 1365–1369.
3. Monks, S.A., Leonardson, A., Zhu, H., Cundiff, P., Pietrusiak, P., Edwards, S., Phillips, J.W., Sachs, A. and Schadt, E.E. (2004) Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.*, **75**, 1094–1105.
4. Goring, H.H., Curran, J.E., Johnson, M.P., Dyer, T.D., Charlesworth, J., Cole, S.A., Jowett, J.B., Abraham, L.J., Rainwater, D.L., Comuzzie, A.G. *et al.* (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.*, **39**, 1208–1216.
5. Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P. *et al.* (2007) A survey of genetic human cortical gene expression. *Nat. Genet.*, **39**, 1494–1499.
6. Lu, A., Tang, Y., Ran, R., Ardizzone, T.L., Wagner, K.R. and Sharp, F.R. (2006) Brain genomics of intracerebral hemorrhage. *J. Cereb. Blood Flow Metab.*, **26**, 230–252.
7. Lu, A., Tang, Y., Ran, R., Clark, J.F., Aronow, B.J. and Sharp, F.R. (2003) Genomics of the periinfarction cortex after focal cerebral ischemia. *J. Cereb. Blood Flow Metab.*, **23**, 786–810.
8. Tang, Y., Lu, A., Aronow, B.J. and Sharp, F.R. (2001) Blood genomic responses differ after stroke, seizures, hypoglycemia, and hypoxia: blood genomic fingerprints of disease. *Ann. Neurol.*, **50**, 699–707.
9. Borovecki, F., Lovrecic, L., Zhou, J., Jeong, H., Then, F., Rosas, H.D., Hersch, S.M., Hogarth, P., Bouzou, B., Jensen, R.V. *et al.* (2005) Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease. *Proc. Natl Acad. Sci. USA*, **102**, 11023–11028.
10. Glatt, S.J., Everall, I.P., Kremen, W.S., Corbeil, J., Sasik, R., Khanlou, N., Han, M., Liew, C.C. and Tsuang, M.T. (2005) Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. *Proc. Natl Acad. Sci. USA*, **102**, 15533–15538.
11. Moore, D.F., Li, H., Jeffries, N., Wright, V., Cooper, R.A. Jr, Elkahloun, A., Gelderman, M.P., Zudaire, E., Blevins, G., Yu, H. *et al.* (2005) Using peripheral blood mononuclear cells to determine a gene expression profile of acute ischemic stroke: a pilot investigation. *Circulation*, **111**, 212–221.
12. Tang, Y., Xu, H., Du, X., Lit, L., Walker, W., Lu, A., Ran, R., Gregg, J.P., Reilly, M., Pancioli, A. *et al.* (2006) Gene expression in blood changes rapidly in neutrophils and monocytes after ischemic stroke in humans: a microarray study. *J. Cereb. Blood Flow Metab.*, **26**, 1089–1102.
13. Tsuang, M.T., Nossova, N., Yager, T., Tsuang, M.M., Guo, S.C., Shyu, K.G., Glatt, S.J. and Liew, C.C. (2005) Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **133B**, 1–5.
14. Sullivan, P.F., Fan, C. and Perou, C.M. (2006) Evaluating the comparability of gene expression in blood and brain. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **141B**, 261–268.
15. Steiper, M.E., Young, N.M. and Sukarna, T.Y. (2004) Genomic data support the hominoid slowdown and an Early Oligocene estimate for the hominoid-cercopithecoid divergence. *Proc. Natl Acad. Sci. USA*, **101**, 17021–17026.
16. Matigian, N.A., McCurdy, R.D., Feron, F., Perry, C., Smith, H., Filippich, C., McLean, D., McGrath, J., Mackay-Sim, A., Mowry, B. *et al.* (2008) Fibroblast and lymphoblast gene expression profiles in schizophrenia: are non-neural cells informative? *PLoS ONE*, **3**, e2412.
17. Runne, H., Kuhn, A., Wild, E.J., Pratyaksha, W., Kristiansen, M., Isaacs, J.D., Regulier, E., Delorenzi, M., Tabrizi, S.J. and Luthi-Carter, R. (2007) Analysis of potential transcriptomic biomarkers for Huntington's disease in peripheral blood. *Proc. Natl Acad. Sci. USA*, **104**, 14424–14429.
18. Yao, Y., Schroder, J. and Karlsson, H. (2008) Verification of proposed peripheral biomarkers in mononuclear cells of individuals with schizophrenia. *J. Psychiatr. Res.*, **42**, 639–643.
19. Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
20. Roth, R.B., Hevezi, P., Lee, J., Willhite, D., Lechner, S.M., Foster, A.C. and Zlotnik, A. (2006) Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. *Neurogenetics*, **7**, 67–80.
21. Eisenberg, E. and Levanon, E.Y. (2003) Human housekeeping genes are compact. *Trends Genet.*, **19**, 362–365.
22. Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A. *et al.* (2002) Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl Acad. Sci. USA*, **99**, 4465–4470.
23. Tekumalla, P.K., Calon, F., Rahman, Z., Birdi, S., Rajput, A.H., Hornykiewicz, O., Di Paolo, T., Bedard, P.J. and Nestler, E.J. (2001) Elevated levels of DeltaFosB and RGS9 in striatum in Parkinson's disease. *Biol. Psychiatry*, **50**, 813–816.
24. Freimer, N.B., Susan, K., Ophoff, R., Jasinska, A.J., McKee, K., Villeneuve, A., Belisle, A., Bailey, J.N., Breidenthal, S.E., Jorgensen, M.J. *et al.* (2007) A quantitative trait locus for variation in dopamine metabolism mapped in a primate model using reference sequences from related species. *Proc. Natl Acad. Sci. USA*, **104**, 15811–15816.
25. Jasinska, A.J., Service, S., Levinson, M., Slaten, E., Lee, O., Sobel, E., Fairbanks, L.A., Bailey, J.N., Jorgensen, M.J., Breidenthal, S.E. *et al.* (2007) A genetic linkage map of the vervet monkey (*Chlorocebus aethiops sabaeus*). *Mamm. Genome*, **18**, 347–360.
26. Fears, S.C., Melega, W.P., Service, S.K., Lee, C., Chen, K., Tu, Z., Jorgensen, M.J., Fairbanks, L.A., Cantor, R.M., Freimer, N.B. *et al.* (2009) Identifying heritable brain phenotypes in an extended pedigree of vervet monkeys. *J. Neurosci.*, **29**, 2867–2875.
27. James, A.S., Groman, S.M., Seu, E., Jorgensen, M., Fairbanks, L.A. and Jentsch, J.D. (2007) Dimensions of impulsivity are associated with poor spatial working memory performance in monkeys. *J. Neurosci.*, **27**, 14358–14364.
28. Nelson, S.B., Hempel, C. and Sugino, K. (2006) Probing the transcriptome of neuronal cell types. *Curr. Opin. Neurobiol.*, **16**, 571–576.
29. Geschwind, D.H. (2000) Mice, microarrays, and the genetic diversity of the brain. *Proc. Natl Acad. Sci. USA*, **97**, 10676–10678.
30. Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A. and Liotta, L.A. (1996) Laser capture microdissection. *Science*, **274**, 998–1001.
31. Alberts, R., Terpstra, P., Li, Y., Breitling, R., Nap, J.P. and Jansen, R.C. (2007) Sequence polymorphisms cause many false cis eQTLs. *PLoS ONE*, **2**, e622.
32. Walter, N.A., McWeeney, S.K., Peters, S.T., Belknap, J.K., Hitzemann, R. and Buck, K.J. (2007) SNPs matter: impact on detection of differential expression. *Nat. Methods*, **4**, 679–680.
33. Lipska, B.K., Deep-Soboslay, A., Weickert, C.S., Hyde, T.M., Martin, C.E., Herman, M.M. and Kleinman, J.E. (2006) Critical factors in gene expression in postmortem human brain: Focus on studies in schizophrenia. *Biol. Psychiatry*, **60**, 650–658.
34. Kuhn, K., Baker, S.C., Chudin, E., Lieu, M.H., Oeser, S., Bennett, H., Rigault, P., Barker, D., McDaniel, T.K. and Chee, M.S. (2004) A novel, high-performance random array platform for quantitative gene expression profiling. *Genome Res.*, **14**, 2347–2356.

35. Lipshutz, R.J., Fodor, S.P., Gingeras, T.R. and Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays. *Nat. Genet.*, **21**, 20–24.

36. Naef, F., Lim, D.A., Patil, N. and Magnasco, M. (2002) DNA hybridization to mismatched templates: a chip study. *Phys. Rev. E. Stat. Nonlin. Soft. Matter Phys.*, **65**, 040902.

37. Naef, F. and Magnasco, M.O. (2003) Solving the riddle of the bright mismatches: labeling and effective binding in oligonucleotide arrays. *Phys. Rev. E. Stat. Nonlin. Soft. Matter Phys.*, **68**, 011906.

38. Rennie, C., Noyes, H.A., Kemp, S.J., Hulme, H., Brass, A. and Hoyle, D.C. (2008) Strong position-dependent effects of sequence mismatches on signal ratios measured using long oligonucleotide microarrays. *BMC Genomics*, **9**, 317.

39. Zhang, L., Miles, M.F. and Aldape, K.D. (2003) A model of molecular interactions on short oligonucleotide microarrays. *Nat. Biotechnol.*, **21**, 818–821.

40. Jacquelin, B., Mayau, V., Brysbaert, G., Regnault, B., Diop, O.M., Arenzana-Seisdedos, F., Rogge, L., Coppee, J.Y., Barre-Sinoussi, F., Benecke, A. *et al.* (2007) Long oligonucleotide microarrays for African green monkey gene expression profile analysis. *FASEB J.*, **21**, 3262–3271.

41. Du, P., Kibbe, W.A. and Lin, S.M. (2008) Lumi: a pipeline for processing Illumina microarray. *Bioinformatics*, **24**, 1547–1548.

42. Smyth, G.K., Michaud, J. and Scott, H.S. (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*, **21**, 2067–2075.

43. Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article3.

44. Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) Probability and mathematical statistics. *Multivariate Analysis*. Academic Press Inc.

45. Almasy, L. and Blangero, J. (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.*, **62**, 1198–1211.

46. Cho, R.J. and Campbell, M.J. (2000) Transcription, genomes, function. *Trends Genet.*, **16**, 409–415.