**Supplemental Text**

**Methods**

**Participant selection**

The following exclusion criteria were applied for participant enrollment: 1) head injury with loss of consciousness greater than 1 hour or cognitive sequela; 2) current/past illicit drug use or positive drug screen for methamphetamines, amphetamines, or cocaine at screening or entry; 3) any of the following laboratory abnormalities: PT/PTT > the upper limit of normal (ULN) or INR > 1.1, hemoglobin < 9.0 mg/dL, ALT > 5x ULN, serum creatinine > 2x ULN or creatinine clearance < 30 cc per min by Cockroft-Gault formula; 4) acute illness within 30 days prior, persistent and active AIDS-defining OI of any organ system or autoimmune disease; 5) current or recent fevers or meningeal signs suggestive of CNS opportunistic infection; 6) CNS opportunistic infection, past or present; 7) history of pre-existing neurologic disease to include stroke, multiple sclerosis or psychiatric illness including schizophrenia, bipolar disorder, anxiety disorder, panic attacks, major depression, or posttraumatic stress disorder. Patients with past depression that is controlled and patients with or minor depressive symptoms will be allowed to enroll; 8) known learning disability including dyslexia or unable to read or write basic Thai; 9) positive Hepatitis C serology (Hepatitis C Ab); 10) confusion or other signs and symptoms of metabolic encephalopathy or delirium; 11) other conditions that could explain neurocognitive decline in the opinion of the investigator such as hypothyroidism, vitamin B12 deficiency or neurosyphilis; 12) pregnancy or metal objects that would preclude MRI**.**

**Flow cytometry**

Monoclonal antibodies used for surface staining and flow cytometry were as follows (BD Biosciences unless otherwise noted): CD14-PB (M5E2), CD16-PE-Cy7 (3G8), CD3-AL700 (SP342), CD4-BV605 (OKT4, Biolegend), and CD8a-BV785 (RPA-T8); and markers of activation: CD169-PE (7-239, Bio-Rad), CD163-PE Cy5.5 (GHI/61, Biolegend), CD86-PE Cy5 (IT2.2), and HLA-DR-APC-H7 (L243). Flow cytometry data analysis was performed using FlowJo (v. 9.7.5, Tree Star).

**cDNA synthesis and quantitative gene expression**

RNA from FACS-sorted cells lysed in 10 l of Superscript III Platinum One-Step qRT-PCR buffer (Cells Direct 1X reaction mix, Superscript III RT Platinum Taq and 0.2x gene-specific Taqman assay mix) was reverse transcribed and pre-amplified on a Gene Amp PCR System 9700 (Thermo Fisher Scientific) as follows: 50oC for 15 min, 95oC for 2 min and 18 rounds of 95oC for 15 seconds and 60oC for 4 min [34]. cDNA was diluted 1:5 in DNA Suspension Buffer (Teknova). 3.6 l of diluted cDNA was mixed with 4.4 l of Real-time Reaction mix comprising of 50 l of Fluidigm Sample Loading Reagent and 500 l TaqMan Universal PCR Master mix. All TaqMan assays (Life Technologies) consisted of FAM-MGB probes that span exon-exon junctions and passed qualification tests to establish both efficient and linear amplification as well as multiplexing capability [39-41]. 4 l of each assay was mixed with 4 l of Fluidigm GE Assay Loading Reagent. Samples and assays were loaded onto the Fluidigm 96.96 Dynamic Array for multiplexed qPCR using program M96 Standard Protocol.

BioMark qPCR gene expression data (Ct values) was processed in JMP where Ct values were converted to expression threshold (Et) values using the formula 40-Ct, where 40 represents the number of qPCR cycles performed. Linear expression values were calculated using the formula 2(Et-13), given Et=13 corresponds to a single copy of RNA. The median of the six cDNA replicates was used as a single expression value for each gene and study participant.

Finally, the Maximum Likelihood Estimate (MLE) model used for viral gene estimation in cell populations allows for estimation of the concentration of positive cells in a sample based on the number of replicate wells at varying dilutions determined to be positive for a gene of interest. To estimate the precision of the MLEs, bootstrapping was used to construct 95% confidence intervals: each positive/negative well was assumed to be an independent observation, and 10,000 different datasets of the same size as the original were generated to determine the plausible values for the concentration estimate.

**Statistical Analysis**

 For median host gene analysis, 10,000 datasets were generated by randomly selecting one replicate for each participant and performing the same testing within each of those datasets. The ratio of group medians (similar to fold-change) was computed for each gene, and volcano plots were used to assess any global patterns in expression across all genes. Boxplots were produced for select genes that showed evidence of differences between clinical diagnosis groups.

DAVID functional annotation analysis was utilized to test for independence between upregulation or downregulation and the annotated pathways or functions. Up- vs. downregulation was determined by comparing group medians, and raw p<0.05 was the threshold used to determine each gene list. The default categories for comparison were analyzed, which included KEGG pathways, Biocarta pathways, and biological processes ascribed by the Gene Ontology Consortium (<http://www.geneontology.org/>). Modified Fisher’s exact tests were used to assess the association between annotated categories and membership in a given gene list. Then p-values were adjusted to control the false discovery rate across the multiple comparisons involving numerous categories.

Principal component analysis (PCA) was applied to the 96 gene expression data after centering the data and scaling to unit variance. The scores related to each of the components (as defined by linear combinations of median replicate gene expression values) that explained at least 10% of the variability in the data were compared between diagnosis groups.