**Title**: Transcriptomic Profiles of Sepsis in the Human Brain

**Authors**: Angela C. Bustamante1, Kristopher Opron2, William J. Ehlenbach3, Eric B. Larson4, Paul K. Crane5, C. Dirk Keene6, Theodore J. Standiford1, Benjamin H. Singer1

**Affiliations**:

1 Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Michigan Health System, Ann Arbor, MI; 2 Bioinformatics Core, University of Michigan, Ann Arbor, MI; 3Department of Medicine, University of Wisconsin, Madison, WI 4Kaiser Permanente Washington Health Research Institute, Seattle, WA; 5Department of Medicine, Division of General Internal Medicine, University of Washington School of Medicine, Seattle, WA; 6Department of Pathology, University of Washington School of Medicine, Seattle, WA

**Supplemental Methods:**

1. **Participant Selection**

The Adult Changes in Thought (ACT) study was approved by institutional review boards at Group Health (now Kaiser Permanente Washington) and the University of Washington. An initial cohort of participants who died while hospitalized were selected from the ACT autopsy cohort (N=89) (1). Medical records for each patient were reviewed to determine cause of death and history of sepsis using a validated and structured instrument (2). Patients were excluded if there was insufficient documentation, an indeterminate cause of death, or history of brain injury based on review of the medical records. In total, 24 patients were identified for inclusion as participants in this study, 12 participants who died of sepsis and 12 participants who died of non-infectious critical illness while hospitalized. Additional information regarding patient cause of death can be found in Supplemental Table S1.

1. **Sample Preparation and RNA Sequencing**

Frozen post-mortem parietal cortex gray matter brain tissue samples were supplied by the ACT brain bank. Parietal cortex tissue samples were collected by trained personnel following a standard protocol at the University of Washington. Participant RNA was isolated from sections of parietal cortex tissue using a standard phenol chloroform RNA isolation protocol at the University of Michigan. All samples were isolated by the same individual who was blind to case/control status of the tissue samples. Following isolation, RNA was treated with DNase I to remove any potential contaminating DNA (Qiagen, Valencia, CA). RNA samples were preliminarily quantified using the NanoDrop 1000 (ThermoScientific). Samples were sent to the University of Michigan Advanced Genomics Core facility to assess RNA integrity as well as prepare and sequence cDNA libraries. RNA concentration and RNA integrity number was determined using the Bioanalyzer (Agilent, Wilmington, DE). cDNA libraries were prepared using the pico-input rRNA depletion cDNA kit (Takara Bio, Mountain View, CA), generated libraries were subsequently sequenced on Illumina’s Hi-Seq platform (Illumina, San Diego, CA). In total 24 RNA samples were sequenced. Raw data was exported and sent to the University of Michigan Bioinformatics Core for processing. Individual level gene count data is available at NCBI GEO (GSE135838).

1. **Preprocessing of RNAseq Data**

Raw count data were downloaded from the University of Michigan Advanced Genomics Core facility’s server and processed by the University of Michigan Bioinformatics Core. The Bioinformatics Core concatenated the read files into a single fastq file for each sample. Raw read quality was checked using the program FastQC version v0.11.3 (3) to identify potential quality problems from features of the data (e.g. low quality score, over-expressed sequences, and inappropriate GC content). The Tuxedo Suite software package was used to align reads, perform differential analysis, and perform post analysis diagnostics (4–6). Reads were aligned to the reference genome including mRNAs and lncRNAs (UCSC genome build hg19) using both TopHat version 2.0.13 and Bowtie2 version 2.2.1. Default settings were used for alignment, with the exception of “--b2-very-sensitive”, which allows the software to spend more time searching for valid alignments, and “—no-coverage-search” which is recommended for reads >45bps. FastQC was used for a second round of quality control (post-alignment), to ensure that only high quality data would be used for all analyses. Cufflinks/CuffDiff (version 2.1.1) was used for expression quantitation, normalization, and differential expression analysis, using UCSC hg19 as the reference genome sequence. More specifically, the default library normalization method, ‘classic fpkm’, was employed on this data, ‘classic fkpm’ sets the library size factor to 1, meaning there is no scaling performed on the FPKM values or fragment counts. Within Cufflinks/Cuffdiff the setting “--compatible-hits-norm” was used. Diagnostic plots were generated using the CummeRbund R package.

1. **Differential Expression Analysis**

The differential expression analysis was performed by the University of Michigan Bioinformatics Core facility using their own internally developed scripts to format and annotate the differential expression data output from CuffDiff. Briefly, differentially expressed genes and transcripts were identified based on three criteria: test status = “OK”, FDR $\leq $ 0.05, and fold difference ≥ ± 0.6. In this analysis FDR (false discovery rate) was estimated using the Benjamini-Hochberg method and is the default setting in CuffDiff (7). Genes and isoforms were annotated using NCBI Entrez GeneIDs and text descriptions.

1. **Weighted Gene Co-Expression Network Analysis**

Weighted gene co-expression networks were generated for using WGCNA version 1.68, an R package available in Bioconductor (<http://bioconductor.org>) (8–10). Weighted gene co-expression network analysis (WGCNA) was performed on RNA sequencing count data that had been filtered to remove low and non-expressing probes. Data were subsequently subjected to variance stabilizing transformation using the R package DESeq2 version 1.22.2. The filtering step removed any features/genes with less than 10 reads total. After filtering 20,967 genes remained and were used as the input for WGCNA. WGCNA was used to generate co-expression networks for sepsis by hierarchical clustering to detect modules, or gene clusters, based on the individual expression or count data for each gene. A correlation matrix was generated during this step and was transformed into an adjacency matrix through the use of a power function (β=7), ultimately creating the weighted gene network. Scale free topology at this power is approximately 0.9, which is the ideal value for scale free topology. Automatic network construction was employed with the function “blockwiseModules”. Modules were created using default parameters as well as the following: signed network type, mergeCutHeight of 0.1, and a minimum module size of 30. Newly created modules were then correlated (biweight midcorrelation) with variables of interest (sepsis, RNA integrity, age, biologic sex, Braak score, and CERAD score) based on the individual module eigengene. The module eigengene describes the first principal component of an individual module and is considered to be representative of the gene expression profile of the module (8). Therefore, a positive correlation indicates an overall pattern of over-expression or upregulation, whereas a negative correlation indicates the overall pattern is under-expressed or down-regulated. Further examination of modules and their individual gene members is necessary to better understand the roles and contributions of each gene within the module. The newly generated modules were provided a unique color identifier. Several iterations of the network construction were performed to optimize the analysis and ensure the best parameters were selected.

1. **Gene Enrichment Analysis**

Gene enrichment analyses were performed on modules significantly (p≤0.05) correlated with sepsis using two different tools, the enrichment analysis built into WGCNA and the bioinformatics tool DAVID version 6.8 (https://david.ncifcrf.gov/). Gene lists comprised of the official gene symbols for each gene were used as input for the enrichment analyses. Both analyses used the default settings. The gene enrichment analysis performed within WGCNA used the full list of genes 20,937 as the input for the background.

**REFERENCES**:

1. Miller JA, Guillozet-Bongaarts A, Gibbons LE, Postupna N, Renz A, Beller AE, Sunkin SM, Ng L, Rose SE, Smith KA, Szafer A, Barber C, Bertagnolli D, Bickley K, Brouner K, Caldejon S, Chapin M, Chua ML, Coleman NM, Cudaback E, Cuhaciyan C, Dalley RA, Dee N, Desta T, Dolbeare TA, Dotson NI, Fisher M, Gaudreault N, Gee G, *et al.* Neuropathological and transcriptomic characteristics of the aged brain. *Elife* 2017;6:L231–L242.

2. Iwashyna TJ, Odden A, Rohde J, Bonham C, Kuhn L, Malani P, Chen L, Flanders S. Identifying patients with severe sepsis using administrative claims: patient-level validation of the Angus implementation of the international consensus conference definition of severe sepsis. *Med Care* 2014;52:e39-43.

3. Andrew S. FastQC: A quality control tool for high throughput sequence data. at <http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>.

4. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* 2013;31:46–53.

5. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;25:1105–1111.

6. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10:R25.

7. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B-Statistical Methodol* 1995;57:289–300.

8. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.

9. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oles AK, Pages H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 2015;12:115–121.

10. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.