Cell culture: The immortalized human cerebral microvascular endothelial cell line (hCMEC/D3) was kindly provided by P-O Couraud, Institut Cochin, France. The cells were cultured in EBM-2 endothelial basal media (Lonza, Walkersville, MD) containing 5% v/v fetal bovine serum (Atlanta biologicals, Flowery Branch, GA), 1% v/v penicillin-streptomycin (Corning, Walkersville, MD), 1.4 µM hydrocortisone (Sigma-Aldrich, St Louis, MO), 5µg/ml ascorbic acid (Sigma-Aldrich, St Louis, MO), 1% v/v chemically defined lipid concentrate (Life Technologies, Grand Island, NY), 10 mM HEPES, 1 ng/ml bFGF (Peprotech, Rocky Hill, NJ). The hCMEC/D3 cells were seeded on collagen-coated Transwell® filters and cultured) at 37 ° C under 5% CO_2 in a humidified chamber.

Illumina TruSeq v2 mRNA and microRNA Protocol: RNA libraries for eight replicates of polarized hCMEC/D3 cell monolayer were prepared according to the manufacturer's instructions using TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego, CA). Briefly, poly-A mRNA was purified from total RNA using oligo dT magnetic beads. The purified mRNA was fragmented at 95°C for 8 min and eluted from the beads. Double stranded cDNA was made using SuperScript III reverse transcriptase, random primers (Invitrogen, Carlsbad, CA), and DNA polymerase I and RNase H. The cDNA ends were repaired and an "A" base is added to the 3' ends. TruSeq paired end index DNA adaptors (Illumina) with a single "T" base overhang at the 3' end were ligated and the resulting constructs were purified using AMPure SPRI beads from Agencourt. The adapter-modified DNA fragments were enriched by 12 cycles of PCR using Illumina TruSeq PCR primers. The concentration and size distribution of the libraries were determined on an Agilent Bioanalyzer DNA 1000 chip and Qubit fluorometry (Invitrogen, Carlsbad, CA). The average RNA integrity number for 8 replicates is 8.3 (+/-0.7). MicroRNA libraries were prepared from 1 ug of total RNA according to manufacturer's instructions for the NEBNext Multiplex Small RNA Kit (New England Biolabs; Ipswich, MA). Libraries (3 samples per lane) were loaded onto paired end flow cells at concentrations of 8-10 pM to generate cluster densities of 700,000/mm^2 following Illumina's standard protocol using the Illumina cBot and cBot Paired end cluster kit version 3. The flow cells were sequenced as 51 X 2 paired end reads on an Illumina HiSeq 2000 using TruSeq SBS sequencing kit version 3 and HCS version 2.0.12.0 data collection software. Base-calling was performed using Illumina's RTA version 1.17.21.3.

RNA-seq Data Analysis: Paired-end RNA-sequencing data alignment and processing was performed using the MAP-RSeq- a comprehensive computational workflow developed at the Mayo Clinic to obtain a variety of genomic features from RNA-seq experiment. The main goal of the MAP-RSeq pipeline is to obtain multiple genomic features, such as gene expression, exon counts, fusion transcripts from RNA-seq data. On an average, 114 million paired-end reads (51 bp) per sample were processed through MAP-RSeq workflow. MAP-RSeq provides quality control reports and summary statistics of sample reads. Total number of reads, mapped reads, number of reads mapped to the genome, and the numbers of reads mapped to junctions were also obtained for each sample. The RNA-seq mapping statistics for all eight BBB replicates are provided in Supplementary Table 1.

Gene Expression Analysis: Gene expression counts were obtained using HT-Seq-count module from MAP-RSeq pipeline for eight hCMEC/D3 replicates. Conditional quantile normalization (CQN) was applied for gene expression counts; normalized data is also available at BBBomics hub.

Identification of Expressed Nucleotide Variants: Expressed single nucleotide variants (eSNVs) from RNA-seq were called using the eSNV-Detect – a computational method developed by our group. The eSNVs observed in the eight replicates were summarized and presented with annotations.

Alternate Splicing Analysis: The MISO software was used to evaluate alternative splicing among replicates. Insertion length distributions were pre-calculated from the MAP-RSeq pipeline. Ensembl indexes were constructed from the provided Ensembl hg19, build 37 file. Results were compiled using in-house python scripts. Each table also contains the ENSEMBL gene identifiers, the HUGO gene identifier, start and stop positions, and exon model retention (with the ENSEMBL identifiers).

Circular RNA (CircRNAs) Analysis: The circular RNA workflow, Circ-Seq version 1.0, was used to process eight hCMEC/D3 replicates. The unmapped reads obtained from the MAP-RSeq workflow were used as input. Bowtie version 2.1.0 was used to align reads to the reference genome. Custom python and bash scripts were used to identify and quantitate reads that supported back-splicing events, i.e., RNA transcripts formed from the splicing of 3' tail to 5' head. The BLAT software was used to eliminate false candidates that mapped to multiple locations in the genome. The raw read counts were reported per sample. In order to obtain the RefSeq genes that either overlap or neighbor the circular RNA candidates, intersectBed and closestBed functions were used from the BedTools suite.

Long Intergenic Non-coding RNA Analysis: To identify long intergenic non-coding RNAs (lincRNAs) present in the control samples, the ICQ-lincRNA version 2.0 (lincRNA workflow), was used. The workflow employs the de novo transcriptome assembler StringTie version 1.0.3 to assemble and report all transcripts expressed in the samples. After removal of all knownRNA transcripts in Gencode (version 19), novel RNA candidates are identified through a set of filters for size selection, expression, repeat masker and non-protein coding potential prediction using CPAT and iSeeRNA to arrive at the final list of potential lincRNA candidates. The raw and normalized read counts were reported per sample. The raw values for each lincRNA were normalized to a million and corrected for the lincRNA length to obtain the normalized reads. The closestBed function from BedTools suite was used to identify both the neighboring RefSeq genes and their distance to the lincRNA. If the lincRNA was found upstream of the gene, the distance was reported with a negative sign. Alternatively, a positive distance was reported if the lincRNA was found downstream of the RefSeq gene. A distance of zero implies overlapped lincRNA, which shares exons with the RefSeq gene.

MicroRNA Analysis: Two replicates of hCMEC/D3 cell monolayers were processed through the microRNA workflow CAP-miRSeq, version 1.0. The known microRNAs were called using the miRDeep software (version 2.0.0.5), and were annotated using miRBase (version 19) database. The raw and normalized read counts were reported per sample. Raw reads were normalized to a

million and further computed by dividing each microRNA raw read count by the total number of microRNA reads to arrive at the normalized reads for each sample. TargetScan was used to obtain the computationally predicted gene targets for all microRNAs reported by the CAP-miRSeq workflow.

Data processing: FASTQ files from RNA and microRNA sequencing were aligned to the human genome build NCBI 37.1 (GRCh37), which corresponds to human genome assembly hg19 in UCSC database.

Pathway Analysis: Pathways with at least 5 sequenced genes were rendered using the R pathview pathview, version 1.4.2. Expression gradients for the replicates are depicted with respect to the 25th and 75th quantiles of the pathway expression matrix. Pathway gene features were annotated using KEGGREST, version 1.4.1 and associated with Hugo gene symbols obtained from the Homo.sapiens package, version 1.1.2. RNA expression profiles were overlaid using the R package pathview for 285 of the 291 KEGG pathways, where a minimum of 5 annotated RNA genes were observed. Each gene or node in the pathway diagram is represented by the 8 bands, representing CQN normalized values summarized by their expression means (additional details in supplementary methods).

Data or web portal organization and access: The "BBBomics" site is implemented as a single page web service executed via a Linux/Apache HTTPd/Javascript/JQuery/Boostrap/Perl stack. The query interface allows users to search with gene and microRNA IDs and provides links to relevant GeneCards pages and KEGG pathways. This is the first web portal providing a number of transcriptomic features for any BBB cell line. Its applications are versatile and will be beneficial in identifying coding and noncoding transcripts, mutations (eSNVs), and pathway profiles to perform functional studies. Supplementary methods section of the manuscript consists of instructions of how to query and interpret the data from BBBomics hub. Data used in this study are deposited in the Gene Expression Omnibus web site at GSE76531.