PNAS
www.pnas.org

Supplementary Information for

Large-scale genomic study reveals robust activation of the immune system following advanced Inner Engineering meditation retreat.

Vijayendran Chandran[1,2,‡], Mei-Ling Bermúdez[1], Mert Koka[1], Brindha Chandran[1], Dhanashri Pawale[3], Ramana Vishnubhotla[3,4], Suresh Alankar[5], Raj Maturi[6], Balachundhar Subramaniam[7], Senthilkumar Sadhasivam[3]

[‡] **Corresponding author:**
Dr. Vijayendran Chandran Ph.D.,
Assistant Professor, Departments of Pediatrics and Neuroscience,
University of Florida, School of Medicine,
1600 SW Archer Road, Box 100296, Gainesville, FL 32610-0296, USA.
E-mail: vijayendran@ufl.edu

This PDF file includes:
Supplementary text
Supplementary Figures Legends: Fig S1 – Fig S5
Supplementary Dataset Legends: Dataset S1 – S7
SI References

SUPPLEMENTARY INFORMATION TEXT

Evaluation of immune cell proportions using cellular deconvolution

To analyze immune cell proportions in the whole blood samples, formatted data with gene symbols were uploaded to the CIBERSORTx web portal (https://cibersortx.stanford.edu/) and LM22 gene signature was utilized (1). LM22 is a signature matrix file consisting of 547 genes that accurately distinguish 22 mature human hematopoietic populations isolated from peripheral blood (1). Bulk RNA-seq data was deconvoluted using the signature matrix with bulk mode batch correction to remove variances between different platforms. Two-way ANOVA test was used to analyze differences in the abundances of different cell types before and after meditation. The p-values were corrected for multiple testing using the Benjamini-Hochberg method. $P < 0.05$ was considered statistically significant.

Differential gene expression analyses

Raw data was log transformed and checked for outliers. Across samples, Pearson correlation and clustering based on variance were used as quality-control measures. One sample from T1 was not included in any analysis due to low read counts. Variance stabilizing transformation normalization method from DESeq2 package was utilized. Normalized data was processed with 'RemoveBatchEffect' function from 'limma' package for batch and cell type composition correction. Next, a linear model was fitted across the dataset, contrasts of interest were extracted, and differentially expressed genes for each contrast were selected using an empirical Bayes test statistic (2). Differential expression analyses on cell type composition corrected or non-corrected data revealed 98% overlap in the number of genes differentially expressed. Hence, we utilized batch and cell type composition corrected data for all downstream analyses.

Construction of co-expression networks

A weighted signed gene co-expression network was constructed using the normalized dataset to identify groups of genes (modules) associated with meditation following a previously described algorithm (3, 4). Briefly, we first computed the Pearson correlation between each pair of selected genes yielding a similarity (correlation) matrix. Next, the adjacency matrix was calculated by raising the absolute values of the correlation matrix to a power (β) as described previously (3). The parameter β was chosen by using the scale-free topology criterion (3), such that the resulting network connectivity distribution best approximated scale-free topology. The adjacency matrix was then used to define a measure of node dissimilarity, based on the topological overlap matrix, a biologically meaningful measure of node similarity (3). Next, the genes were hierarchically clustered using the distance measure and modules were determined by choosing a height cutoff for the resulting dendrogram by using a dynamic tree-cutting algorithm (3). Utilizing this network analysis, we identified modules (groups of genes) differentially expressed across different time points before and after meditation and calculated the first principal component of gene expression in each module (module eigengene). Next, we correlated the module eigengenes with time points before and after meditation to select modules for functional validation.

Transcriptional regulatory network analyses

We applied the network-based integrative NetBID (5) algorithm to identify critical drivers associated with meditation using batch and cell type composition corrected gene expression

profiles. We first reverse-engineered a meditation-specific regulatory network by using the SJARACNe algorithm (6) from 388 RNA-Seq profiles of individuals before and after meditation. We then applied the activity inference algorithm in NetBID to identify drivers whose network activity are significantly different between T3 versus other time points. In short, meditation-specific network was inferred based on the gene expression profiles based on mutual information dependency. Transcription factors and signaling factors with a high number of differentially expressed targets in the network were regarded as potential drivers. Gene set enrichment analysis (GSEA) and curated gene sets from MSigDB (7) were used to identify potential biological functions of the drivers and targets. The activity of the potential driver is calculated from the expression of its targets, and differential activities were inferred using NetBID. The drivers whose activities are significantly correlated with the timepoint T3 were utilized for enrichment analyses. We performed gene set enrichment analysis with default parameters using pathways derived from gene sets from Molecular Signatures Database (7). All network plots were constructed using the Cytoscape software (8).

Transcription factor binding site enrichment analyses

Transcription factor binding site (TFBS) enrichment analysis was performed using the 'TFBSenrich' function from 'RegFacEnc' package (https://tfenrichment.semel.ucla.edu/). TFBS enrichment analysis was performed by scanning the canonical promoter region (1000bp upstream of the transcription start site) for the top 200 genes (based on kME) present in the meditation-associated co-expression modules. Next, we utilized TFBS position weight matrices (PWMs) from JASPAR (746 motifs) and HOCOMOCO (769 motifs) databases (9, 10) to examine the enrichment for corresponding TFBS within each module. For TFBS enrichment all the modules were scanned with each PWMs using Clover algorithm (11). To compute the enrichment analysis, we utilized three different background datasets (1000 bp sequences upstream of all human genes, human CpG islands and human chromosome 20 sequence). When a TFBS is over-represented (based on the P-values obtained relative to all the three corresponding background datasets) we considered it to be significant, which increases our confidence in these predictions.

Protein-Protein Interaction (PPI) Network Analyses

We constructed an experimentally validated protein–protein interaction (PPI) network using both turquoise and brown meditation-associated co-expression gene network modules. We created all possible combinations of gene pairs present in these co-expression networks and identified all experimentally verified interaction data (in humans dataset) for their corresponding proteins in the STRING database (integration of the following databases: BIND, DIP, GRID, HPRD, IntAct, MINT, and PID) (12), constructing the protein network by force-directed layout organized by significantly enriched functional pathways (Fig. 3C). Nodes correspond to genes and edges to PPI. The size of each node in the PPI network correspond to differential expression (Z-score). Node color represents upregulation (red) or downregulation (blue). Nodes with highlighted border (green) correspond to enriched or driver TFs.

Gene Ontology, Pathway enrichment and PubMed Analyses

Gene ontology and pathway enrichment analysis was performed using the Fisher's Exact Test with 'funcEnrich.Fisher' function from 'NetBID' package (https://github.com/jyyulab/NetBID). A list of differentially regulated transcripts for a given modules were utilized for enrichment

analyses. We performed gene set enrichment analysis with default parameters using pathways derived from gene sets from Molecular Signatures Database (7). For PubMed analyses we determined the association with the following key-words: “interferon signaling” and “interferon pathway” in the PubMed database for every gene using R (http://cran.r-project.org/).

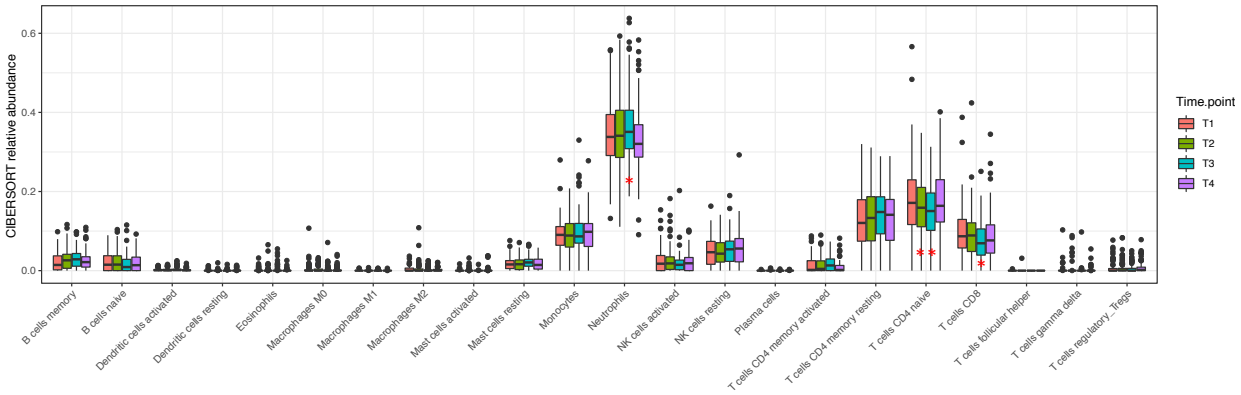
Quantitative real-time PCR

qRT-PCR was performed using a Bio-Rad CFX96 real-time PCR system according to the manufacturer’s instructions. Briefly, RNA was harvested from human whole blood samples and cDNA was produced using SuperScript VILO IV master mix (cat #11756050). iTaq Universal SYBR Green Supermix was used to quantify amplification of cDNA. Primers we designed using Primer-BLAST (13) and was verified to amplify one product by verifying one peak present on the dissociation curves, and standard curves were performed to show that this assay is sensitive to changes in each gene. The following primer sequences were utilized: STAT1_F: CAGCTTGACTCAAAATTCCTGGA, STAT1_R: TGAAGATTACGCTTGCTTTTCCT, STAT2_F: CCAGCTTTACTCGCACAGC, STAT2_R: AGCCTTGGAATCATCACTCCC, TRIM22_F: CTGTCCTGTGTGTCAGACCAG, TRIM22_R: TGTGGGCTCATCTTGACCTCT. Ten biological replicates (individuals) were used for each timepoints, and three technical replicates were performed for each sample. The relative expression of genes was calculated using the $2^{-\Delta\Delta CT}$ method.

Comparison of transcriptional profiles with publicly available datasets

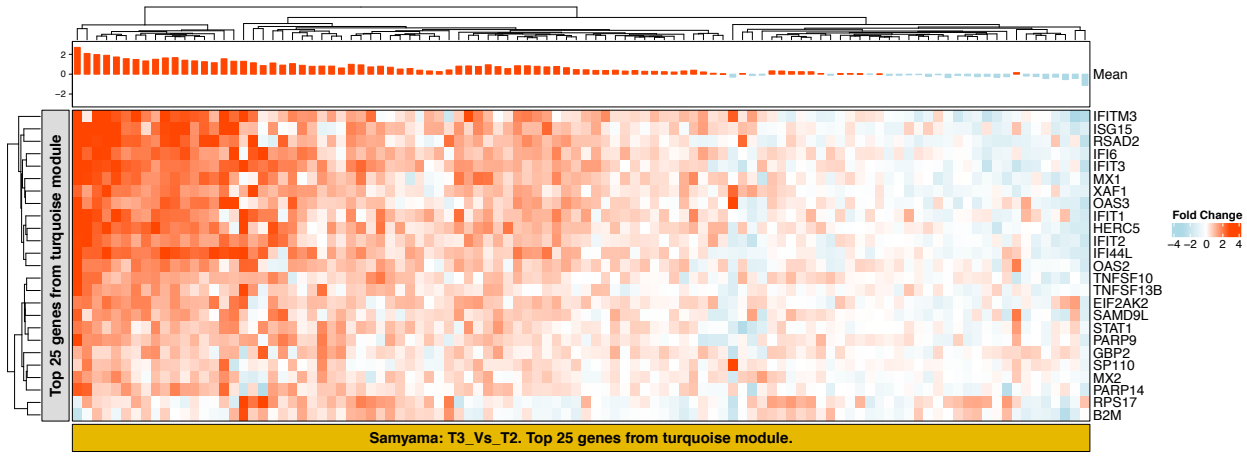
Gene expression data sets were downloaded from the Gene Expression Omnibus (GEO), read into R, preprocessed and normalized using variance stabilizing transformation normalization method from DESeq2 package. We then calculated the correlation of gene expression between samples, and outliers with mean sample correlations more than two to three standard deviations below average were omitted until no outliers remained. Using the ‘limma’ package, a linear model was fitted across the dataset, contrasts of interest were extracted, and differentially expressed genes for each contrast were analyzed using an empirical Bayes test statistic (2). We analyzed several publicly available different datasets and compared it with meditation dataset generated and analyzed in this study. We analyzed: dataset generated from leukocyte samples from hospitalized patients with or without COVID-19 (GSE157103), dataset from whole blood of MS patients and controls (GSE41850), and exercise training datasets (GSE111554, GSE111553).

SUPPLEMENTARY FIGURES

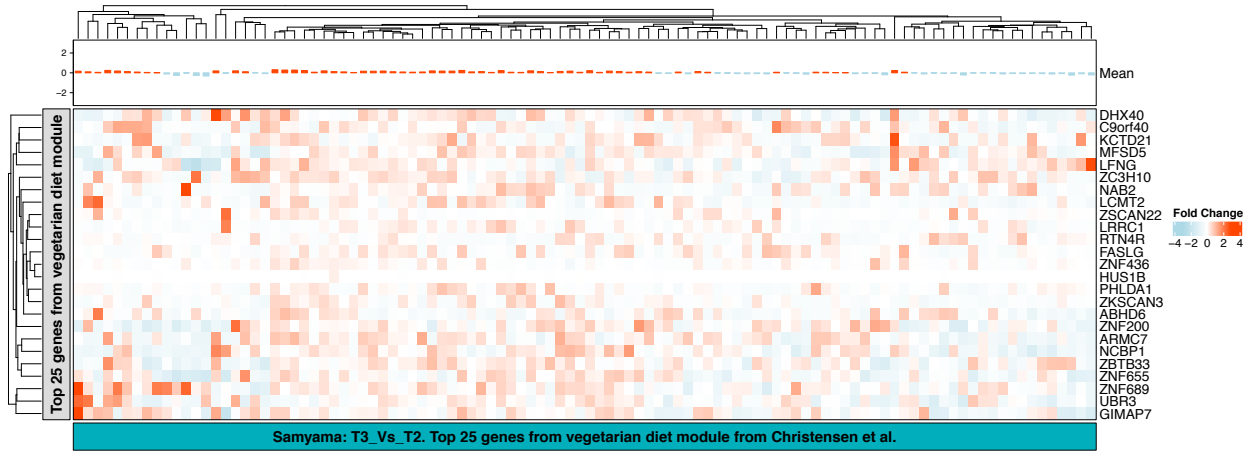


Supplementary Fig. S1. The quantity of 22 different immune cell types estimation in the meditation gene expression data using CIBERSORT-Relative deconvolution. (A) The box plot visualizing the relative abundance of 22 immune cell subsets. CIBERSORT-derived immune cell relative scores were used to determine the abundance of immune cells in all the samples derived from four-time pre-and post-meditation (T1-T4). Significantly higher abundance of neutrophils and significantly lower abundance of CD8+ T-cells and naïve CD4+ T-cell was observed at T3 compared to T1. Comparisons were performed by employing two-way ANOVA test. *P < 0.05.

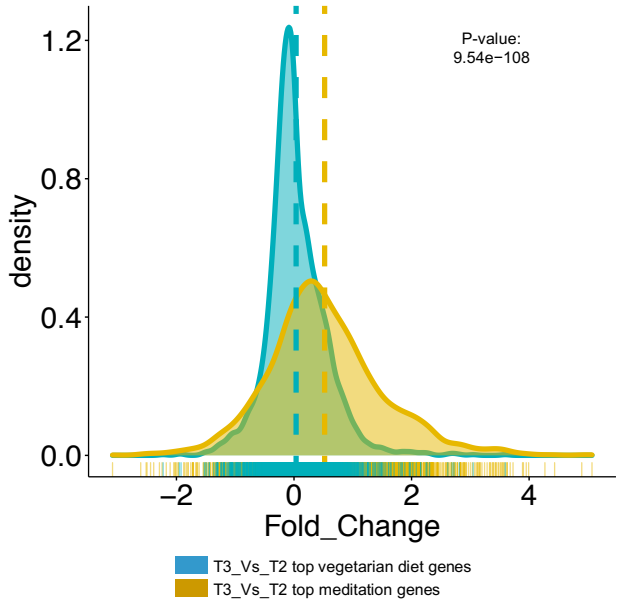
A


Mean
Top 25 genes from turquoise module
IFITM3
ISG15
RSAD2
IFI6
IFIT3
MX1
XAF1
OAS3
IFIT1
HERC5
IFIT2
IFI44L
OAS2
TNFSF10
TNFSF13B
EIF2AK2
SAMD9L
STAT1
PARP9
GBP2
SP110
MX2
PARP14
RPS17
B2M
Fold Change
Samyama: T3_Vs_T2. Top 25 genes from turquoise module.

B

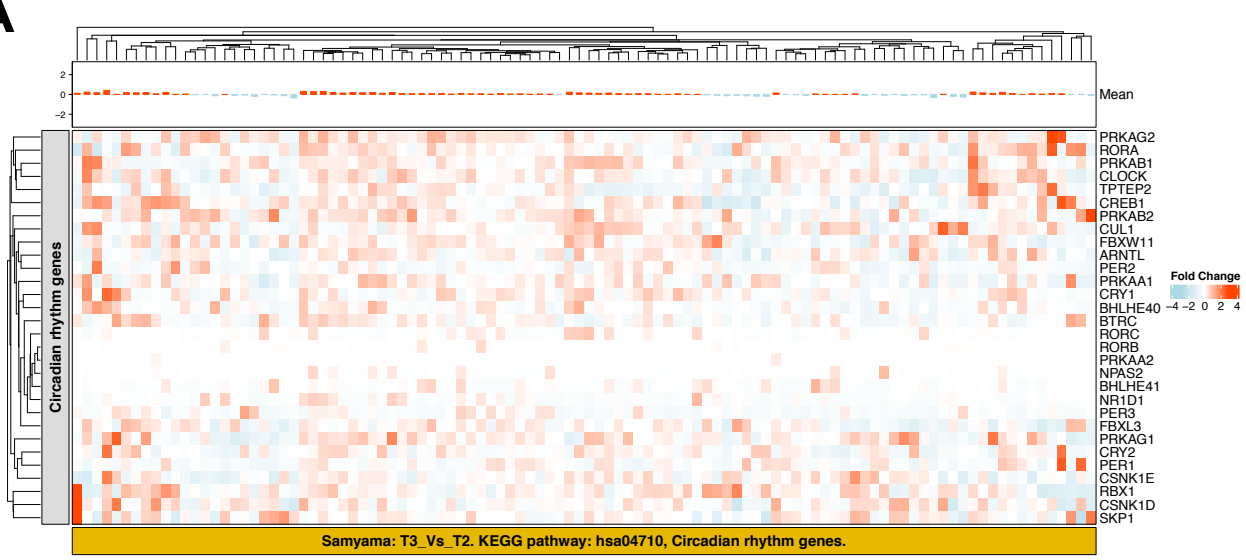

Mean
Top 25 genes from vegetarian diet module
DHX40
C9orf40
KCTD21
MFSD5
LFNG
ZC3H10
NAB2
LCMT2
ZSCAN22
LRRC1
RTN4R
FASLG
ZNF436
HUS1B
PHLDA1
ZKSCAN3
ABHD6
ZNF200
ARMC7
NCBP1
ZBTB33
ZNF655
ZNF689
UBR3
GIMAP7
Fold Change
Samyama: T3_Vs_T2. Top 25 genes from vegetarian diet module from Christensen et al.

C

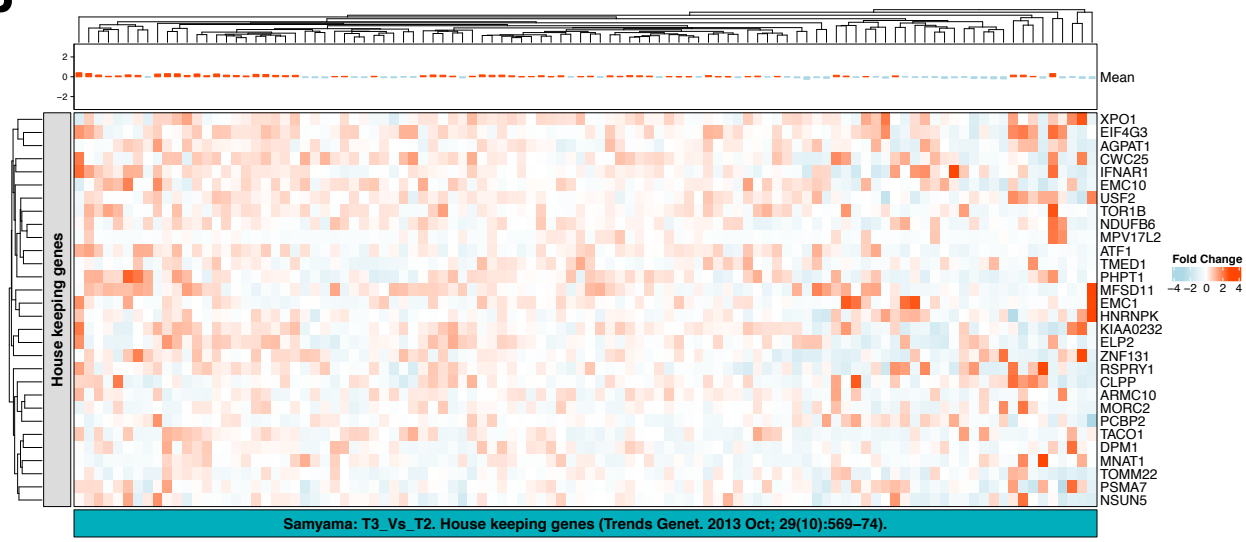

P-value:
9.54e−108
density
Fold_Change
T3_Vs_T2 top vegetarian diet genes
T3_Vs_T2 top meditation genes

Supplementary Fig. S2. Comparison of meditation and diet-associated specific transcriptional gene networks. (A) Heatmap depicting the expression of the top 25 genes (rows) from the meditation specific network (turquoise module) across samples (columns) for the timepoint comparisons, T3 (after meditation) versus T2 (before meditation), showing significant changes across samples. **(B)** Heatmap depicting the expression of the top 25 vegetarian diet-specific genes (rows) in samples before and after meditation (columns) for the timepoint comparisons, T3 (after meditation) versus T2 (before meditation), showing no significant changes across samples. For both heatmaps, the red color corresponds to gene upregulation and blue to downregulation. Mean gene expression levels are shown as a bar-plot on top of each heatmap. (**C**) The density plots showing the distribution of log2 fold change for both gene lists (top meditation and diet-associated genes) and the significance of the variability in the expression levels between the two groups are calculated by a two-sample t-test.

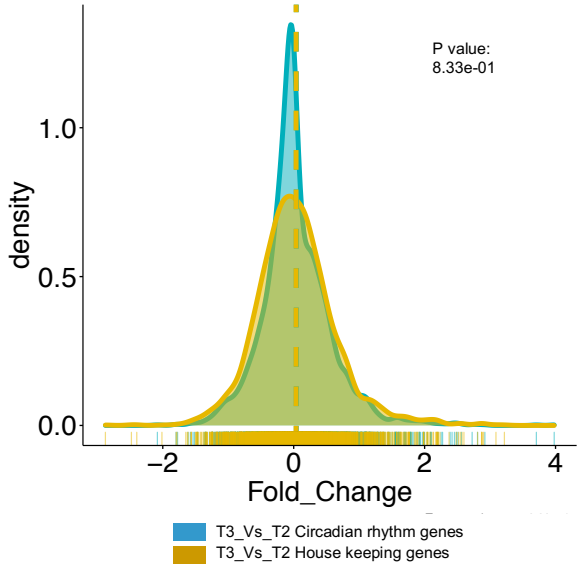
A


Mean
Circadian rhythm genes
PRKAG2
RORA
PRKAB1
CLOCK
TPTEP2
CREB1
PRKAB2
CUL1
FBXW11
ARNTL
PER2
PRKAA1
CRY1
BHLHE40
BTRC
RORC
RORB
PRKAA2
NPAS2
BHLHE41
NR1D1
PER3
FBXL3
PRKAG1
CRY2
PER1
CSNK1E
RBX1
CSNK1D
SKP1
Fold Change
−4 −2 0 2 4
Samyama: T3_Vs_T2. KEGG pathway: hsa04710, Circadian rhythm genes.

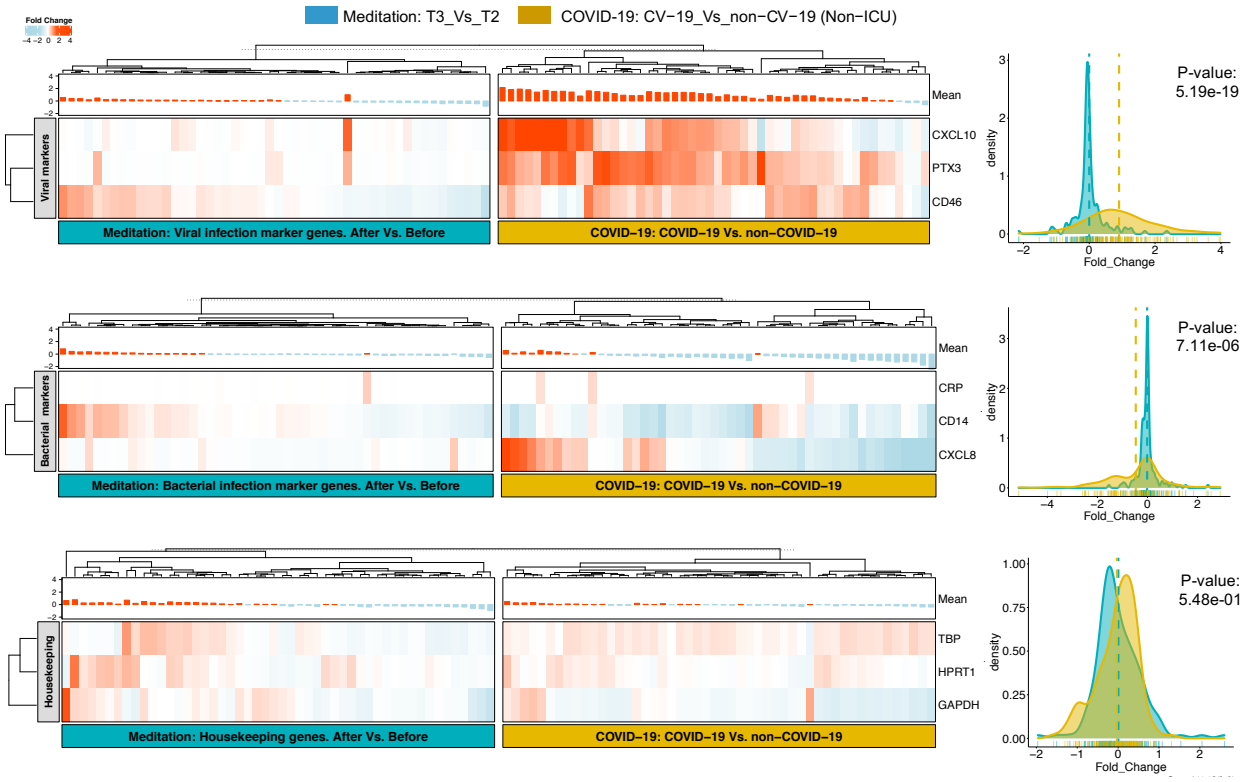
B


Mean
House keeping genes
XPO1
EIF4G3
AGPAT1
CWC25
IFNAR1
EMC10
USF2
TOR1B
NDUFB6
MPV17L2
ATF1
TMED1
PHPT1
MFSD11
EMC1
HNRNPK
KIAA0232
ELP2
ZNF131
RSPRY1
CLPP
ARMC10
MORC2
PCBP2
TACO1
DPM1
MNAT1
TOMM22
PSMA7
NSUN5
Fold Change
−4 −2 0 2 4
Samyama: T3_Vs_T2. House keeping genes (Trends Genet. 2013 Oct; 29(10):569–74).

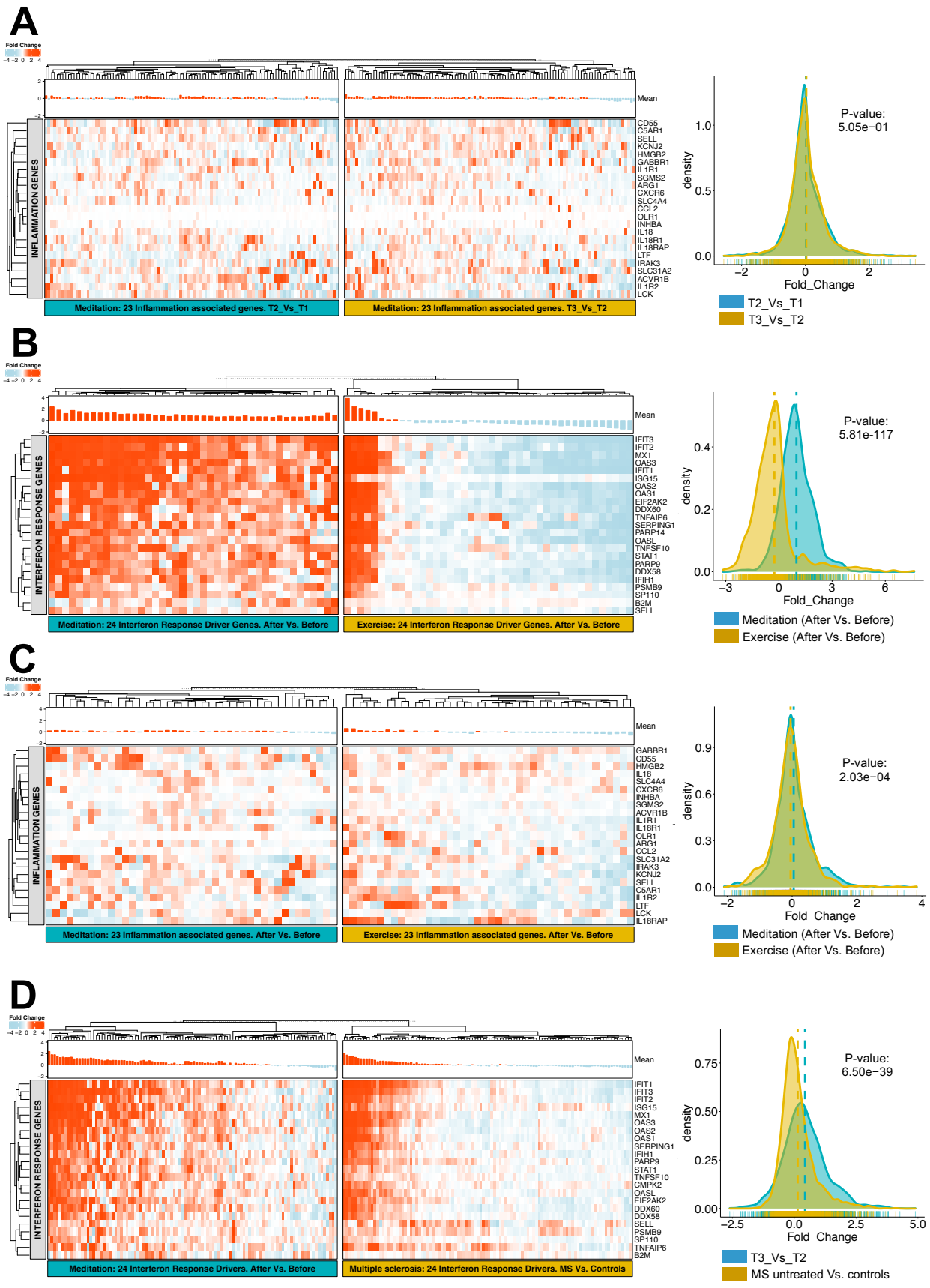
C


P value:
8.33e-01
density
1.0
0.5
0.0
−2
0
2
4
Fold_Change
T3_Vs_T2 Circadian rhythm genes
T3_Vs_T2 House keeping genes

Supplementary Fig. S3. Comparison of circadian rhythm and housekeeping specific transcriptional gene expression levels. **(A)** Heatmap depicting the expression of the 30 circadian rhythm genes (rows) (KEGG pathway: hsa04710) across samples (columns) for the timepoint comparisons, T3 (after meditation) versus T2 (before meditation), showing no significant changes across samples. **(B)** Heatmap depicting the expression of the random 30 housekeeping genes (rows) in samples before and after meditation (columns) for the timepoint comparisons, T3 (after meditation) versus T2 (before meditation), showing no significant changes across samples. For both heatmaps, the red color corresponds to gene upregulation and blue to downregulation. Mean gene expression levels are shown as a bar-plot on top of each heatmap. (**C**) The density plots showing the distribution of log2 fold change for both gene lists (top circadian rhythm and house keeping genes) and the significance of the variability in the expression levels between the two groups are calculated by a two-sample t-test.



Supplementary Fig. S4. Comparison of meditation and COVID-19 transcriptional profiles showing viral and bacterial marker genes. Heatmaps are depicting the expression of the three bona fide viral markers (top), bacterial markers (middle), and control genes (bottom) across meditation (left - T3 (after meditation) versus T2) and COVID-19 (right - non-ICU COVID-19 patients versus non-ICU non-COVID-19 patients) samples (columns). These marker genes are usually elicited by viral or bacterial infection. For all heatmaps, the red color corresponds to gene upregulation and blue to downregulation. Mean gene expression levels are shown as a bar-plot on top of each heatmap. All the heatmaps are supplemented with density plots showing the distribution of log2 fold change, and the significance of the variability in the expression levels between the two groups are calculated by a two-sample t-test.


A
Fold Change
Mean
INFLAMMATION GENES
CD55
C5AR1
SELL
KCNJ2
HMGB2
GABBR1
IL1R1
SGMS2
ARG1
CXCR6
SLC4A4
CCL2
OLR1
INHBA
IL18
IL18R1
IL18RAP
LTF
IRAK3
SLC31A2
ACVR1B
IL1R2
LCK
Meditation: 23 Inflammation associated genes. T2_Vs_T1
Meditation: 23 Inflammation associated genes. T3_Vs_T2
density
P-value:
5.05e−01
Fold_Change
T2_Vs_T1
T3_Vs_T2
B
INTERFERON RESPONSE GENES
IFIT3
IFIT2
MX1
OAS3
IFIT1
ISG15
OAS2
OAS1
EIF2AK2
DDX60
TNFAIP6
SERPING1
PARP14
OASL
TNFSF10
STAT1
PARP9
DDX58
IFIH1
PSMB9
SP110
B2M
SELL
Meditation: 24 Interferon Response Driver Genes. After Vs. Before
Exercise: 24 Interferon Response Driver Genes. After Vs. Before
P-value:
5.81e-117
Meditation (After Vs. Before)
Exercise (After Vs. Before)
C
GABBR1
CD55
HMGB2
IL18
SLC4A4
CXCR6
INHBA
SGMS2
ACVR1B
IL1R1
IL18R1
OLR1
ARG1
CCL2
SLC31A2
IRAK3
KCNJ2
SELL
C5AR1
IL1R2
LTF
LCK
IL18RAP
Meditation: 23 Inflammation associated genes. After Vs. Before
Exercise: 23 Inflammation associated genes. After Vs. Before
P-value:
2.03e−04
Meditation (After Vs. Before)
Exercise (After Vs. Before)
D
IFIT1
IFIT3
IFIT2
ISG15
MX1
OAS3
OAS2
OAS1
SERPING1
IFIH1
PARP9
STAT1
TNFSF10
CMPK2
OASL
EIF2AK2
DDX60
DDX58
SELL
PSMB9
SP110
TNFAIP6
B2M
Meditation: 24 Interferon Response Drivers. After Vs. Before
Multiple sclerosis: 24 Interferon Response Drivers. MS Vs. Controls
P-value:
6.50e−39
T3_Vs_T2
MS untreated Vs. controls

Supplementary Fig. S5. Comparison of meditation, exercise, and multiple sclerosis specific transcriptional profiles in terms of interferon and inflammatory response. (**A**) Heatmaps are depicting the expression of the 23 inflammatory response genes (rows) across meditation samples (columns) for two different timepoint comparisons, T2 (before meditation) versus T1 (baseline) and T3 (after meditation) versus T2, showing no significant changes. (**B**) Heatmaps are comparing the expression of the 24 meditation-specific IFN driver genes across meditation samples (left) and in samples obtained after exercise training (right) for two different comparisons; after meditation (T3) versus before meditation (T2) and after exercise training versus before exercise training. (**C**) Heatmaps comparing the expression of the 23 inflammatory response genes in samples before and after meditation and exercise training for two different comparisons as described in **B**. (**D**) Heatmaps are comparing the expression of the 24 meditation-specific IFN driver genes across meditation samples (left) and in samples obtained from multiple sclerosis patients (right) for two different comparisons; after meditation versus before meditation and MS patients versus controls. For all heatmaps, the red color corresponds to gene upregulation and blue to downregulation. Mean gene expression levels are shown as a bar-plot on top of each heatmap. All the heatmaps are supplemented with density plots showing the distribution of log2 fold change, and the significance of the variability in the expression levels between the two groups are calculated by a two-sample t-test.

SUPPLEMENTARY DATASET LEGENDS

(Datasets provided as individual .xlsx data files)

Dataset S1. Differentially expressed genes before and after the meditation retreat at four-time points. Genes with a significant differential expression in batch and cell type composition corrected data in all four timepoints (T1, T2, T3, T4) by pair-wise comparison of all six permutation combinations are provided.

Dataset S2. Gene-module membership association based on WGCNA co-expression networks. Nine robust modules identified from datasets generated at four time points before and after meditation are denoted along with genes in these modules, the gene significance (GS) values (correlation of a gene expression profile with a sample trait) with T3, and the module membership (MM) values (Intramodular connectivity).

Dataset S3. Gene ontology enrichment analysis of meditation associated modules. For gene set enrichment analysis for GO terms, we considered GO terms with Fisher's Exact Test P-values less than 0.05. Enriched GO terms are provided for each module in separate tabs. Description of column headers: #Name (Name of the enriched gene set), Total_item (Background size), Num_item (Number of genes in the gene set (filtered by the background list)), Num_list (Number of input genes for testing (filtered by the background list)), Num_list_item (Number of input genes annotated by the gene set (filtered by the background list)), P value (Original P-value from Fisher's Exact Test), Odds_Ratio (Odds ratio from the 2*2 matrix used for Fisher's Exact Test), Intersected_items (A vector of the intersected genes, collapsed by ';'. Number is equal to Num_list_item).

Dataset S4. NetBID analysis identifies meditation-associated drivers. Ninety drivers with significant differential activity in comparing T3 versus other time points are provided in this dataset, along with the differential gene expression values. The dataset provides TF (transcription factor) information, Sig (signaling factor) information, all the DE (differential expression analysis) and DA (differential activity analysis) from multiple comparisons with Z-statistics. It also shows each driver's target gene size and other additional information (e.g. gene biotype, chromosome name, position etc.).

Dataset S5. Gene set enrichment analysis of meditation-associated drivers. Gene set enrichment analysis against the collection of annotated gene sets from MSigDB was utilized to elucidate the functional relevance of ninety meditation-associated drivers. We considered gene sets with Fisher's Exact Test P-values less than 0.05 as enriched. Enriched gene sets terms are provided for up and down drivers in separate tabs. Description of column headers: #Name (Name of the enriched gene set), Total_item (Background size), Num_item (Number of genes in the gene set (filtered by the background list)), Num_list (Number of input genes for testing (filtered by the background list)), Num_list_item (Number of input genes annotated by the gene set (filtered by the background list)), P value (Original P-value from Fisher's Exact Test), Odds_Ratio (Odds ratio from the 2*2 matrix used for Fisher's Exact Test), Intersected_items (A vector of the intersected genes, collapsed by ';'. Number is equal to Num_list_item).

Dataset S6. Analysis of transcription-factor binding-sites (TFBS) enrichment in the meditation specific co-expression modules. For estimation of TFBSs enrichment in the identified corresponding module-genelist (top 200 genes based on connectivity) promoter sequences (1000bp upstream from transcription start site), P-values were obtained relative to three background datasets: 1000-bp of sequence upstream of all human gene, human CpG islands and human chromosome 20 (see methods). Enriched TFBS position weight matrices from both JASPAR and HOCOMOCO databases are provided in this dataset. Enriched TFBSs are provided for each module in separate tabs.

Dataset S7. Literature annotation of enriched transcription factors associated with interferon signaling. Dataset providing enriched transcription factor associated with interferon signaling based on the published literature by testing association with the key-words: 'interferon signaling' and 'interferon pathway' in the PubMed database for every gene. The total number of hits (publications) for each gene is represented.

SI REFERENCES

1. A. M. Newman *et al.*, Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nature biotechnology* **37**, 773-782 (2019).
2. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47 (2015).
3. B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* **4**, Article17 (2005).
4. M. C. Oldham, S. Horvath, D. H. Geschwind, Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 17973-17978 (2006).
5. X. Du *et al.*, Hippo/Mst signalling couples metabolic state and immune function of CD8α(+) dendritic cells. *Nature* **558**, 141-145 (2018).
6. A. Khatamian, E. O. Paull, A. Califano, J. Yu, SJARACNe: a scalable software tool for gene network reverse engineering from big data. *Bioinformatics (Oxford, England)* **35**, 2165-2166 (2019).
7. A. Liberzon *et al.*, The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems* **1**, 417-425 (2015).
8. R. Saito *et al.*, A travel guide to Cytoscape plugins. *Nature methods* **9**, 1069-1076 (2012).
9. I. V. Kulakovskiy *et al.*, HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic acids research* **46**, D252-D259 (2018).
10. O. Fornes *et al.*, JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic acids research* **48**, D87-D92 (2020).
11. M. C. Frith *et al.*, Detection of functional DNA motifs via statistical over-representation. *Nucleic acids research* **32**, 1372-1381 (2004).
12. D. Szklarczyk *et al.*, The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research* **49**, D605-D612 (2021).
13. J. Ye *et al.*, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics* **13**, 134 (2012).