PNAS
www.pnas.org

Supplementary Information for

Genetically-defined cellular correlates of the baseline brain MRI signal

Jie Wen, Manu S. Goyal, Serguei V. Astafiev, Marcus E. Raichle, and Dmitriy A. Yablonskiy

Corresponding authors:

Marcus E. Raichle, MD,

Email: mraichle@wustl.edu

Dmitriy A. Yablonskiy, PhD,

Email: yablonskiyd@wustl.edu

This PDF file includes:

Figs. S1, S2, S3 and S4

Captions for Datasets S1, S2, S3, S4, S5, S6, S7 and S8

Other supplementary materials for this manuscript include the following:

Datasets S1, S2, S3, S4, S5, S6, S7 and S8

www.pnas.org/cgi/doi/10.1073/pnas.1808121115

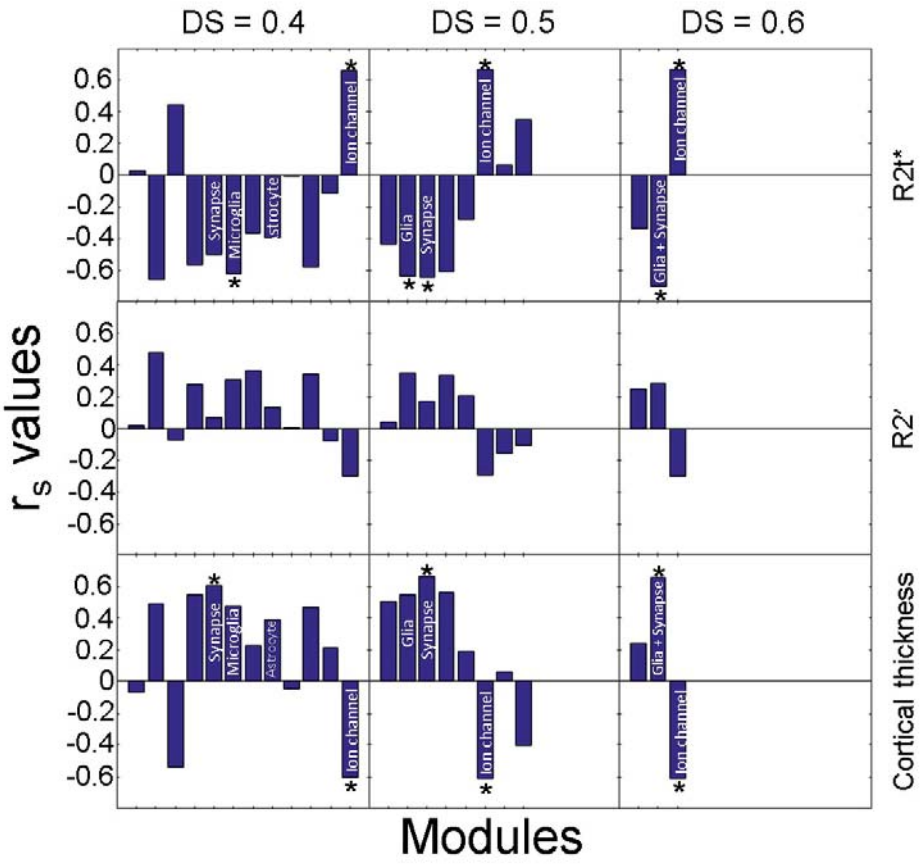

Fig. S1. Correlations between MEs and MRI features (R2t*, R2' and cortical thickness) using different DS (1) thresholds and fixed ME threshold (=0.5). Cellular affiliations of different modules are marked in the bars. Significant correlations (averaged $| r_s | > 0.6$ and $p < 0.05$ for each of the Allen brain) were marked by *. All p-values were adjusted for multiple comparisons using false discovery rate (FDR) method with Benjamini-Hochberg procedure. When using different DS threshold (0.4, 0.5 and 0.6), the number of genes in the WGCNA analysis changed – 3339 genes for DS = 0.4, 1995 genes for DS = 0.5 and 1077 genes for DS = 0.6. As expected, reducing DS threshold increases the number of genes entering the WGCNA analysis and results in modules with more specific cellular affiliation but with lower statistical power. Increased DS threshold, results in less genes and modules with less specific cellular affiliation. The choice of DS threshold of 0.5 presents a reasonable compromise between specificity of cellular affiliation and statistical power of the analysis using all 6 Allen brains.

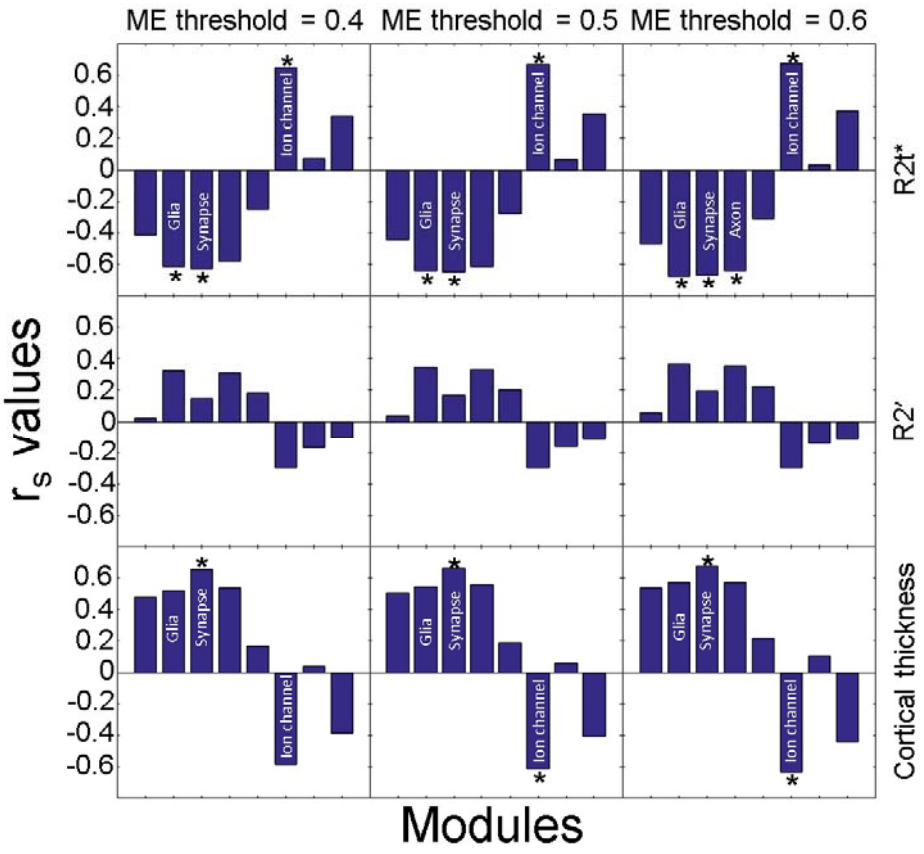

Fig. S2. Correlations between MEs and MRI features (R2t*, R2' and cortical thickness) using different ME correlation thresholds and fixed DS threshold (=0.5). Cellular affiliations of different modules are marked in the bars. Significant correlations (averaged $|r_s| > 0.6$ and $p < 0.05$ for each of the Allen brain) are marked by *. All p-values are adjusted for multiple comparisons using false discovery rate (FDR) method with Benjamini-Hochberg procedure. When using different ME correlation thresholds, the final number of genes that were assigned to non-grey modules changed – 11090 genes for threshold 0.4, 8086 genes for threshold 0.5 and 5191 genes for threshold 0.6. However, the correlation coefficients remained similar for these different thresholds, except for threshold 0.6, the correlation of R2t* with one additional module became significant.

To evaluate the role of this additional module, we analyzed it using ToppGene tool (https://toppgene.cchmc.org) (2) and established that the module contains gene networks related to axon. We also performed a gene enrichment analysis using DAVID Bioinformatics Resources (https://david.ncifcrf.gov/) (3). We found that the enrichment score of genes that are related to axon in this module is relatively low compared to modules that we discussed before (see the details in Fig. S3). The genes in this module that have the highest enrichment scores are related to tubulin and microtubules. This further supports our conclusion on the dominant role that myelinated axons play in forming R2t* signal - the negative correlation of this module with R2t* could be attributed to the neuronal fibers with relatively higher axon to myelin volumetric ratio. By incorporating this

module into our theoretical model can potentially improve understanding of different sub-cellular contributions to R2t* signal.

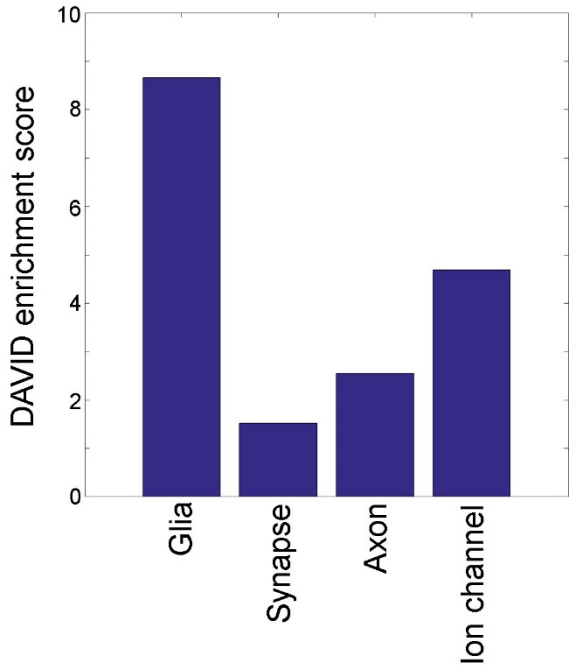

Fig. S3. Gene enrichment analysis of the four significant modules as shown in Fig. S2 when using threshold 0.6. The analysis were performed using DAVID Bioinformatics Resources. The values shown here are the highest enrichment scores of the gene clusters in each module. In the "Glia" module, the gene cluster that has the highest enrichment score is related to Metallothionein. In the "Synapse" module, the gene cluster that has the highest enrichment score is related to synapse and cell junction. In the "Axon" module, the gene cluster that has the highest enrichment score is related to tubulin and microtubules. In the "Ion channel" module, the gene cluster that has the highest enrichment score is related to ion channel.

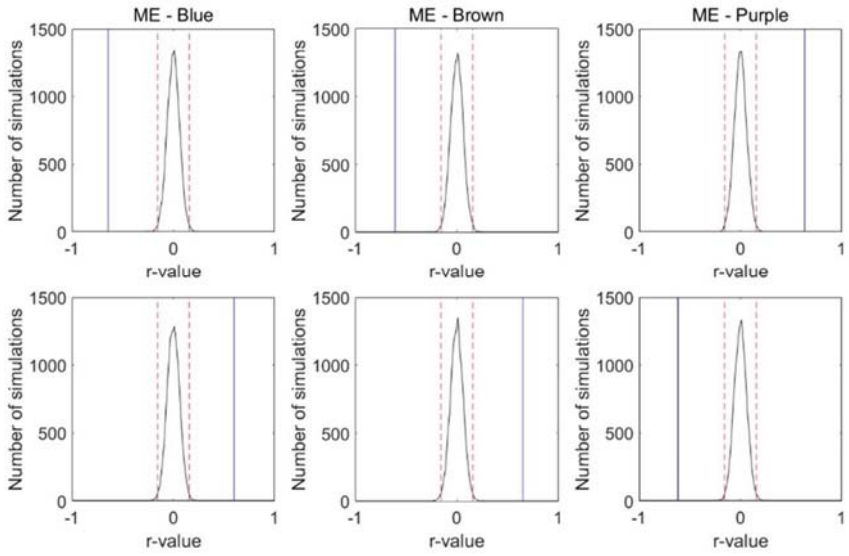

Fig. S4. Stability of the correlations between module eigengenes (ME) and MRI measurements (**top**: R2t*; **bottom**: cortical thickness). Black lines show distributions of r values between original ME and MRI measurements obtained by randomly permuting R2t* and thickness values between FreeSurfer ROIs. Simulations were repeated 10,000 times. The 99% percentiles of the distribution lines are marked by red dashed lines. The r values without permutation are marked as blue lines that are well outside red lines.

Dataset S1 (separate file)

Detailed gene assignment in modules. Gene symbols and module colors that they belong to are provided.

Dataset S2 (separate file)

Results of gene enrichment analysis obtained using “ToppFun” tool in “ToppGene”.

Dataset S3 (separate file)

Results of gene enrichment analysis of Module Blue obtained using DAVID Bioinformatics Resources with a default background setting.

Dataset S4 (separate file)

Results of gene enrichment analysis of Module Blue obtained using DAVID Bioinformatics Resources with a background that contains genes in the AHBA datasets.

Dataset S5 (separate file)

Results of gene enrichment analysis of Module Brown obtained using DAVID Bioinformatics Resources with a default background setting.

Dataset S6 (separate file)

Results of gene enrichment analysis of Module Brown obtained using DAVID Bioinformatics Resources with a background that contains genes in the AHBA datasets.

Dataset S7 (separate file)

Results of gene enrichment analysis of Module Purple obtained using DAVID Bioinformatics Resources with a default background setting.

Dataset S8 (separate file)

Results of gene enrichment analysis of Module Purple obtained using DAVID Bioinformatics Resources with a background that contains genes in the AHBA datasets.

References

1. Hawrylycz M*, et al.* (2015) Canonical genetic signatures of the adult human brain. *Nat Neurosci* 18(12):1832-1844.
2. Chen J, Bardes EE, Aronow BJ, & Jegga AG (2009) ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 37(Web Server issue):W305-311.
3. Jiao XL*, et al.* (2012) DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* 28(13):1805-1806.