Supplemental information

Al-enabled in silico immunohistochemical characterization for Alzheimer's disease

Bryan He, Syed Bukhari, Edward Fox, Abubakar Abid, Jeanne Shen, Claudia Kawas, Maria Corrada, Thomas Montine, and James Zou
Figure S1: Examples of 2048 x 2048 IHC patches annotated as positive for Tau tangles, positive for amyloid plaques, and negative for both. Related to STAR Methods. (a) Positive NFT examples can occur on a high or low background. (b) Negative patches for NFTs can include lipofuscin or folds in the tissue. (c) Positive amyloid plaque examples can occur on high or low backgrounds. (d) Negative patches for amyloid plaques can include lipofuscin and neuromelanin. All these heterogeneities make the task of in silico IHC challenging.
Figure S2: Performance on additional evaluation dataset of 30 samples collected from consecutive patients. Related to Figures 3 and 4. No exclusion criteria were applied to ensure that the patients are representative. (a-c) AUROCs for individual slides for (a) NFTs, (b) neuritic plaques, and (c) amyloid plaques. (d-f) Receiver operating characteristics curve for (d) NFTs, (e) neuritic plaques, and (f) amyloid plaques.
Figure S3: Registration examples. Related to Figure 1. (a) Example of H&E-LFB and IHC stained slides from the same sample. Matches between key points are shown by lines. (b) Example of a H&E-LFB and an IHC patch at full resolution before and after registration. Three matching blood vessels are circled in different colors. The registration process results in the blood vessels being much more closely aligned.
Figure S4: The receiver operating characteristic curves for identifying patches with each hallmark change from IHC. Related to STAR Methods. Results for and (a) amyloid plaques, (b) NFTs, and (c) neuritic plaques.
Figure S5: Calibration plots for each hallmark change. Related to STAR Methods. Calibration plots for (a) tangles, (b) neuritic plaques, and (c) plaques. The patches are grouped by in silico-IHC’s predicted probability into bins of 10%, and the empirical fraction of true positive patches in each bin are computed. The predictions are reasonably calibrated, even without directly attempting to calibrate the model’s predictions. For all three hallmark changes, the model becomes slightly overconfident for high probability predictions, which may be because the reference IHC image is from a different section of tissue, which may result in the lesions only appearing in the H&E section.
Table S1: Performance of in silico-IHC compared to other choices of neural network architectures. Related to STAR Methods. All these models are trained on the same data.

| Model        | AUROC          |
|--------------|----------------|
|              | AP       | NFT       | NP       |
| in silico-IHC| 0.91 (0.88 - 0.95) | 0.92 (0.87 - 0.94) | 0.88 (0.82 - 0.93) |
| AlexNet      | 0.86 (0.81 - 0.92) | 0.84 (0.79 - 0.89) | 0.80 (0.75 - 0.85) |
| VGG11        | 0.88 (0.83 - 0.93) | 0.90 (0.86 - 0.93) | 0.87 (0.82 - 0.91) |
| ResNet-18    | 0.83 (0.77 - 0.89) | 0.87 (0.81 - 0.91) | 0.83 (0.76 - 0.88) |