Supplementary table 1 – Neuropsychological tests used to adjudicate cognitive diagnosis.

| Cognitive domain                      | Tests                                                                 |
|---------------------------------------|----------------------------------------------------------------------|
| Language                              | Boston Naming Test, Letter Fluency, Animal Fluency                   |
| Visuospatial                          | Modified Rey-Osterreith Figure Copy, Simple Drawings, Block Design   |
| Attention/Information processing speed| Trail Making Test A, Digit Symbol Subtest                            |
| Memory                                | Logical Memory, Modified Rey-Osterreith Figure Recall, California Verbal Learning Test |
| Executive                             | Stroop Neuropsychological Screening Test, Trail Making Test Part B, Clock Drawing |
Supplementary table 2 – Proteins related to White matter hyperintensities in LLD subjects

| Protein                                           | Disease effect | p-value (permuted) | q-value (permuted) |
|---------------------------------------------------|----------------|--------------------|--------------------|
| Apolipoprotein A I (Apo AI)                       | -9.36          | 0.0003             | 0.07               |
| Aldose Reductase                                  | -6.49          | 0.0015             | 0.07               |
| Kidney Injury Molecule 1 (KIM-1)                  | 13.04          | 0.0016             | 0.07               |
| Neutrophil Gelatinase Associated Lipocalin (NGAL)| -19.36         | 0.001              | 0.07               |
| CD40 Ligand (CD40L)                               | -18.12         | 0.007              | 0.15               |
| Insulin like Growth Factor Binding Protein 2 (IGFBP-2) | 7.95         | 0.007              | 0.15               |
| Insulin like Growth Factor Binding Protein 3 (IGFBP-3) | -12.43       | 0.004              | 0.15               |
| Macrophage Colony Stimulating Factor 1 (MCSF)     | 8.48           | 0.008              | 0.15               |
| Tissue Inhibitor of Metalloproteinases 1 (TIMP-1) | 4.14           | 0.005              | 0.15               |
| Vitronectin                                       | -4.46          | 0.007              | 0.15               |
| Cancer Antigen 19 9 (CA 19 9)                     | 13.33          | 0.011              | 0.15               |
| FASLG Receptor (FAS)                              | 7.20           | 0.011              | 0.15               |
| Luteinizing Hormone (LH)                          | 7.59           | 0.011              | 0.15               |
| Neuron Specific Enolase (NSE)                     | -10.14         | 0.010              | 0.15               |
| Neuropilin 1                                      | 3.43           | 0.013              | 0.17               |
| YKL 40                                            | 9.80           | 0.017              | 0.21               |
| CD5 Antigen like (CD5L)                           | -7.07          | 0.019              | 0.21               |
| Epidermal Growth Factor (EGF)                     | -16.57         | 0.021              | 0.21               |
| Interleukin 8 (IL-8)                              | 9.14           | 0.022              | 0.21               |
| Receptor tyrosine protein kinase erbB 3 (ErbB3)   | 13.33          | 0.024              | 0.21               |
| Stem Cell Factor (SCF)                            | -4.17          | 0.023              | 0.21               |
| Insulin like Growth Factor Binding Protein 6 (IGFBP-6) | -9.65        | 0.028              | 0.24               |
| Apolipoprotein a (Lpa)                            | -18.23         | 0.03               | 0.24               |
| Clusterin                                         | -2.79          | 0.033              | 0.25               |
| Osteoprotegerin                                   | 3.38           | 0.032              | 0.25               |
| Tissue type                                      | Value 1 | Value 2 | Value 3 |
|-------------------------------------------------|---------|---------|---------|
| Tissue type Plasminogen activator (tPA)         | -4.32   | 0.036   | 0.25    |
| Tyrosine kinase with Ig and EGF homology domains 2 (TIE-2) | -3.40   | 0.037   | 0.25    |
| Fibrinogen                                      | 6.19    | 0.039   | 0.26    |
| Mesothelin (MSLN)                               | 6.68    | 0.041   | 0.26    |
| C Peptide                                       | -16.84  | 0.042   | 0.26    |
| Fatty Acid Binding Protein liver (FABP liver)   | 11.99   | 0.046   | 0.28    |
| Tumor necrosis factor receptor 2 (TNFR2)        | 3.12    | 0.049   | 0.29    |
Supplemental table 3 – Proteins related to gray matter volume in LLD subjects

| Protein                                                                 | Disease effect | p-value (permuted) | q-value (permuted) |
|------------------------------------------------------------------------|----------------|--------------------|--------------------|
| Angiotensin Converting Enzyme (ACE)                                    | -3.96          | 0.002              | 0.18               |
| Apolipoprotein C III (ApoC III)                                        | -4.55          | 0.004              | 0.18               |
| Glutathione S Transferase alpha (GST alpha)                            | -7.60          | 0.007              | 0.18               |
| Macrophage Colony Stimulating Factor 1 (M-CSF)                         | -4.71          | 0.007              | 0.18               |
| Myeloperoxidase (MPO)                                                 | 7.40           | 0.007              | 0.18               |
| N terminal prohormone of brain natriuretic peptide (NT pro-BNP)        | -1.12          | 0.005              | 0.18               |
| Sortilin                                                               | -3.52          | 0.005              | 0.18               |
| YKL-40                                                                 | -7.59          | 0.005              | 0.18               |
| Beta-2 Microglobulin (B2M)                                             | -2.91          | 0.009              | 0.19               |
| Alpha-2 Macroglobulin (A2Macro)                                        | -5.39          | 0.01               | 0.21               |
| Receptor tyrosine protein kinase erbB 3 (ErbB3)                       | -8.49          | 0.01               | 0.23               |
| Thyroid Stimulating Hormone (TSH)                                      | -7.46          | 0.01               | 0.23               |
| Fatty Acid Binding Protein heart (FABP heart)                          | -3.26          | 0.01               | 0.24               |
| Alpha 1 Microglobulin (A1Micro)                                       | -1.95          | 0.03               | 0.25               |
| Apolipoprotein A I (ApoA I)                                           | 3.12           | 0.03               | 0.25               |
| AXL Receptor Tyrosine Kinase (AXL)                                     | -2.12          | 0.03               | 0.25               |
| B cell activating factor (BAFF)                                        | -2.10          | 0.04               | 0.25               |
| CD40 antigen (CD40)                                                   | -2.58          | 0.03               | 0.25               |
| Chromogranin A (CgA)                                                  | 8.76           | 0.03               | 0.25               |
| Collagen IV                                                            | -2.58          | 0.04               | 0.25               |
| Complement Factor H Related Protein 1 (CFHR1)                         | -2.76          | 0.03               | 0.25               |
| Cortisol                                                               | 4.05           | 0.03               | 0.25               |
| Hepatocyte Growth Factor (HGF)                                         | 3.46           | 0.04               | 0.25               |
| Interferon-gamma Induced Protein 10 (IP10)                             | -4.64          | 0.02               | 0.25               |
| IL6 receptor subunit beta (IL6R beta) | -1.66  | 0.03  | 0.25  |
| IL-12 Subunit p40 (IL12 p40)       | 2.30   | 0.02  | 0.25  |
| Latency Associated Peptide of TGF-beta 1 (LAP TGF b1) | 2.48   | 0.03  | 0.25  |
| Thrombomodulin™                     | -1.88  | 0.02  | 0.25  |
| Vascular Cell Adhesion Molecule 1 (VCAM1) | -1.89  | 0.02  | 0.25  |
Supplementary Table 4 - Model performance (uncorrected) with different number of proteins on predicting LLC-MCI

| Model protein number | accuracy | sensitivity | specificity |
|----------------------|----------|-------------|-------------|
| 2                    | 0.700    | 0.795       | 0.583       |
| 3                    | 0.813    | 0.864       | 0.750       |
| 4                    | 0.763    | 0.841       | 0.667       |
| 5                    | 0.763    | 0.841       | 0.667       |
| 6                    | 0.738    | 0.818       | 0.639       |
| 7                    | 0.700    | 0.750       | 0.639       |
| 8                    | 0.738    | 0.773       | 0.694       |
| 9                    | 0.763    | 0.795       | 0.722       |
| 10                   | 0.738    | 0.795       | 0.667       |
| 11                   | 0.775    | 0.795       | 0.750       |
| 12                   | 0.750    | 0.795       | 0.694       |
| 13                   | 0.700    | 0.795       | 0.583       |
| 14                   | 0.750    | 0.841       | 0.639       |
| 15                   | 0.713    | 0.818       | 0.583       |
| 16                   | 0.675    | 0.727       | 0.611       |
| 17                   | 0.688    | 0.773       | 0.583       |
| 18                   | 0.688    | 0.750       | 0.611       |
| 19                   | 0.688    | 0.750       | 0.611       |
| 20                   | 0.700    | 0.750       | 0.639       |
### Supplementary Table 5 - Pathway analysis of differentially expressed proteins related to white mater hyperintensities

| Biological pathways and moldecular process                                                      | p-value (permuted) | q-value (permuted) | match gene                  |
|-------------------------------------------------------------------------------------------------|--------------------|--------------------|----------------------------|
| KEGG P53 SIGNALING PATHWAY                                                                     | 0.001              | 0.027              | IGFBP3 /FAS /THBS1 /        |
| BIOCARTA PPARA PATHWAY                                                                         | 0.004              | 0.054              | APOA1 /INS /FABP1 /         |
| GO BP TRANSMEMBRANE RECEPTOR PROTEIN TYROSINE KINASE SIGNALING PATHWAY                           | 0.023              | 0.152              | EGF /ERBB3 /TEK /           |
| GO BP REGULATION OF PROTEIN METABOLIC PROCESS                                                  | 0.024              | 0.152              | IGFBP3 /TIMP1 /INS /        |
| BIOCARTA KERATINOCYTE PATHWAY                                                                  | 0.030              | 0.152              | FAS /EGF /TNFRSF1B /        |
| GO BP LEUKOCYTE ACTIVATION                                                                    | 0.034              | 0.152              | CD40LG /IL8 /INS /          |
| GO BP CYTOKINE PRODUCTION                                                                      | 0.044              | 0.152              | APOA1 /CD40LG /INS /        |
| GO BP NEGATIVE REGULATION OF PROTEIN METABOLIC PROCESS                                         | 0.054              | 0.152              | IGFBP3 /TIMP1 /INS /        |
| GO BP CELL ACTIVATION                                                                         | 0.059              | 0.152              | CD40LG /IL8 /INS /          |
| GO MF PROTEIN TYROSINE KINASE ACTIVITY                                                         | 0.068              | 0.152              | NRP1 /ERBB3 /TEK /          |
| REACTOME INTEGRIN CELL SURFACE INTERACTIONS                                                    | 0.069              | 0.152              | VTN /FGA /THBS1 /           |
| GO BP ENZYME LINKED RECEPTOR PROTEIN SIGNALING PATHWAY                                         | 0.076              | 0.152              | EGF /ERBB3 /TEK /           |
| GO BP ORGAN MORPHOGENESIS                                                                      | 0.078              | 0.152              | NRP1 /EGF /IL8 /FABP1 /     |
| REACTOME FORMATION OF PLATELET PLUG                                                            | 0.079              | 0.152              | APOA1 /TIMP1 /EGF /CLU /FGA /THBS1 / |
Supplementary Table 6 - Pathway analysis of differential expressed proteins related to gray matter volume (q value < 0.1)

| Biological pathways and molecular process | p-value (permuted) | q-value (permuted) | match gene     |
|------------------------------------------|--------------------|-------------------|----------------|
| BIOCARTA DENDRITIC CELL PATHWAY          | 0.001              | 0.08              | CD40 / IL12B   |

Supplementary table 7 - Common significant pathways associated with both whole brain gray matter volume and white matter hyperintensities (both with p value <0.1).

| Biological pathways and molecular process | GMV    | WMHs   |
|------------------------------------------|--------|--------|
| REACTOME FORMATION OF PLATELET PLUG      | 0.047  | 0.079  |
| GO BP ORGAN MORPHOGENESIS                | 0.040  | 0.078  |
| BIOCARTA PPARA PATHWAY                   | 0.019  | 0.004  |
| GO BP ENZYME LINKED RECEPTOR PROTEIN SIGNALING PATHWAY | 0.046 | 0.076 |
Supplementary figure 1 – Distributions of missing values.

Histogram of count_missing
Supplementary Figure 2 - **Protein biomarker prediction performance assessment flowchart with corrected model selection bias.** To search for the optimal number of features (2-20 proteins) in the prediction model, we applied nested cross-validation (CV) utilizing two nested CV loops. The dataset is initially divided into one sample as the test set and the remaining as the training set. Then leave-one-out cross validation (LOOCV) is applied on the training set only using all the classifiers (choosing top 2-20 proteins) and the classifier with the smallest error rate is selected and used to build the model on the training set. Finally, the model is evaluated on the left-out test sample and the procedure is repeated until all samples are left-out once. In this manner, the left-out test sample is independent of the model selection stage, including the selection of model with the minimum error rate, and the procedure guarantees an unbiased error estimate.
Supplementary neuroimaging information: Region of Interest (ROI) definitions for regions used in correlation matrices (Ref. 33).

| ROI (right and left)          | Dorsal limit                                                                 | Ventral limit                  | Anterior limit                          | Posterior limit          | Medial posterior limit                        | Lateral posterior limit                                      |
|-------------------------------|-----------------------------------------------------------------------------|-------------------------------|-----------------------------------------|--------------------------|-----------------------------------------------|-------------------------------------------------------------|
| Ventral frontal cortex        | Middle of the genu of the corpus callosum                                  | -                             | Anterior cingulate sulcus               |                         | Anterior cingulate sulcus                      | Circular Sulcus / Sylvian Fissure                            |
| Dorsal Frontal Cortex         | Middle of the genu of the corpus callosum                                  | Anterior cingulate sulcus     |                                         |                         | Corpus colossum                               | Circular Sulcus & Pre-Central Sulcus (most superiorly)      |
| Subgenual Ant. Cingulate      | Middle of the genu of the corpus callosum                                  | Anterior cingulate sulcus     |                                         |                         | Corpus colossum                               |                                                             |
| Corpus callosum               | Middle of the genu of the corpus callosum                                  | Anterior cingulate sulcus     |                                         |                         | Corpus colossum                               |                                                             |
| Mesial Temporal Cortex        | Best two contiguous planes containing amygdala & hippocampus              |                               |                                         |                         |                                               |                                                             |
| Lower Precuneus               | Lower limit of LPM/RPM                                                      | Splenium of corpus callosum   |                                         |                         | Parieto-occipital sulcus                      |                                                             |
| Middle Precuneus              | Dorsal-most point of occipital lobe                                        | marginal/posterior branch of cingulate sulcus |                         |                         |                                               |                                                             |
| Upper Precuneus               | Dorsal-most point of occipital lobe                                        | marginal/posterior branch of cingulate sulcus | Superior temporal sulcus, inferiority; post. branches of post-central sulcus, superiorty |                         | Parieto-occipital sulcus or back of brain     |                                                             |
| Parietal Cortex               |                                                                             | Calcarine Fissure             | Superior temporal sulcus, Inferiority; post. branches of post-central sulcus, superiorty |                         |                                               | Intraparietal Sulcus                                        |
| Lateral Temporal Cortex       | Dorsal limit of LMT/RMT                                                    | Superior Temporal Sulcus      | Temporo-Occipital Sulcus                |                         |                                               |                                                             |
| Occipital Cortex              | Calcarine Fissure                                                          | Parieto-Occipital Sulcus      | Back of brain                           |                         |                                               |                                                             |
| Occipital Pole                | Dorsal limit of LMT/RMT                                                    | Cerebellum                    | Lateral occipital sulcus                |                         |                                               |                                                             |
| Cerebellar (Reference)        |                                                                             | 2 planes above apex of fourth ventricle |                                         |                         |                                               |                                                             |

Supplementary material:
**Random intercept model (RIM).** To account for the existence of confounders (sex and age), we applied a random intercept model (RIM) (Wang, et al., 2012). For a given protein \(g\), we fit the model:

\[
Y_{gik} = \mu_g + \beta_{g0}X_{0ik} + \sum_{l=1}^{L} \beta_{gl}X_{lik} + \alpha_k + \epsilon_{gik}.
\]

In the model, \(Y_{gik}\) was the protein expression value of protein \(g\) (\(1 \leq g \leq G\)) and main factor \(i\) (e.g., \(i=1\) for LLC-MCI and \(2\) for LLC-NC) in sample pair \(k\) (\(1 \leq k \leq K\)) and \(X_{0ik}\) is the main comparison. \(X_{lik}\) represented values for potential confounding covariate \(l\) (\(1 \leq l \leq L = 2\); e.g., \(X_{lik} = 0\) for male samples and \(X_{lik} = 1\) for female). \(\alpha_k\) was the random intercept from a normal distribution with mean zero and variance \(\tau_{g}^{2}\), which represented the deviation of averaged expression values in the \(k\)th pair from the average of the whole population. Finally, \(\epsilon_{gik}\) were independent random noises that followed a normal distribution with mean zero and variance \(\sigma_{g}^{2}\). Under this model, \(\beta_{g0}\) was the main effect of protein \(g\) and represented the parameter of major interest. To obtain a differential biomarker candidate list associated with this main factor, likelihood ratio test was used to assess the \(p\)-values of testing \(H_0: \beta_{g0} = 0\) (vs \(H_A: \beta_{g0} \neq 0\)). The \(p\)-values were then corrected by Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995; Wolin, et al., 2011) for multiple comparisons. (Wang, et al., 2012) demonstrated using simulation and real data that including the random effects \(\alpha_k\) improved the statistical power.

**Variable selection for RIM.** Although the RIM model can effectively adjust for confounding covariates in detecting differential gene expression, considering variable selection could avoid overfitting and increase statistical power. Specifically, all possible RIM models that included at most two (i.e., \(0, 1\) or \(2\)) clinical variables were computed and compared. The model with the smallest Bayesian Information Criterion (BIC) (Schwarz, 1978) value was selected. This additional variable selection allowed the inclusion of different sets of covariates for each gene, which offers more relevant biological conclusions and interpretations (e.g., gene A is confounded with age only while gene B is confounded with both age and sex). Likelihood ratio tests was used to generate \(p\)-values of testing \(H_0: \beta_{g0} = 0\) in each gene for the selected model by BIC.

The obtained \(p\)-values from the best BIC model were, however, not correct \(p\)-values for differential expression since they were biased by the variable selection procedure and the type I error control was voided. The skewed null distribution deviating from uniform distribution between \(0\) and \(1\) showed the need of a permutation analysis for \(p\)-value correction. As a result, we performed a permutation analysis that randomly shuffled the sample labels to generate a null distribution for \(p\)-value assessment (\(B=1000\)). Subsequently, the permutation-corrected and unbiased \(p\)-values were further adjusted by Benjamini-Hochberg procedure for multiple comparisons within each study to control false discovery rate (FDR).

**Prediction analysis and model construction** To investigate the prediction power of protein features on predicting MCI status among LLD patients, support vector machines (SVM) with feature selection was applied to construct the models. SVM has been shown to give overall good performance in comparative studies (Lee, et al., 2005). We use SVM as the classification algorithm with a linear kernel (implemented in R package e1041). Feature selection is applied on the 187 proteins based on 2 criterions: (1) \(p\)-value calculated from t-test between two classes in the training samples, (2) effect size (difference of average of log2 expression between two groups). In selecting the proteins to construct model, proteins were first ranked by \(p\)-values
and then proteins with log2 fold change less than 0.2 were filtered away. The second criterion was necessary because features with significant p-values but small effect sizes were likely ineffective predictors (Spijker, et al., 2010). We tested different SVM models with features from 2 to 20. We evaluated different models with leave-one-out cross validation (LOOCV) and selected the best one that generated the smallest CV error rate. This procedure, however, introduces selection bias from choosing the best model. We applied nested cross validation to estimate an unbiased error rate estimate (Varma and Simon, 2006). Specifically, in every iteration of nested-cross validation, 79 of the 80 samples were treated as training samples and the rest 1 sample is treated as the independent test sample. The models were only evaluated within the 79 training samples with a second round of leave-one-out cross validation. The model with the best CV accuracy was selected from the 79 samples training set and applied to the independent test sample. The whole process was iterated for 80 times and the average performance was reported. The performance evaluation workflow is illustrated in Figure S1. In our analysis, prediction analysis without bias correction generated 81.3% LOOCV accuracy using 3 proteins. With nested cross validation, the estimated accuracy was corrected to 76.3%.

Pathway enrichment analysis: We applied pathway enrichment analysis using Fisher's exact test to identify enriched functional annotation of the identified differentially expressed proteins. Since the original 242 proteins on the assay were pre-selected with relevance to diseases, conventional genome-wide pathway analysis would tend to identify pathways related to the pre-selected proteins with high significance. To correct for the selection bias, we adopted a permutation analysis. Fisher's exact test was first applied on the differentially expressed proteins using the whole-genome as the background. The resulting p-value was treated as the observed statistic. At each permutation iteration, the same number of proteins were randomly subsampled from the 187 proteins and genome-wide Fisher's exact test was similarly applied. The random subsampling was repeated for 1,000 times and the resulting 1,000 p-values formed the null distribution to assess the p-value of the observed statistic. Two-thousand-eleven pathways were downloaded and parsed from the MsigDB database from GO, KEGG and BIOCARTA. Pathways associated with more than 200 genes were excluded to avoid general terms. The pathway enrichment analysis was applied on the differentially expressed proteins associated with the main factors LLC-MCI, whole brain gray matter volume and white matter hyperintensities, respectively.