MESA-MIND-Longitudinal-Proteomics-cSVD

LekkiWood (LekkiWood@Gmail.com)

2026-01-16

Table of contents

Important Notes - read me first

Version control

- Always check that you have the most recent version of this document, which - unless I am sending you unfinalized work - is available here.
  - An easy check for version control is to make sure this date: 2026-01-16. is the same as on the GitHub file here.
- The code for this is analysis available in the same repository (targets master file here and individual functions here)

Regarding the use of site & race/ethnicity

- Some MESA field centers did not collect data on individuals from select race / ethnicity groups (see: Table 1). Therefore, only race/ethnicity was included in analyses, with the exception of sample descriptives.

Table 1: Field Center by Race / Ethnicity Frequencies

Characteristic	Field Center					
	WFU N = 713[1]	**COL** N = 607[1]	**JHU** N = 491[1]	**MN** N = 710[1]	**MW** N = 683[1]	**UC**
Race or ethnicity						
Non-Hispanic White	331 / 713 (46%)	150 / 607 (25%)	256 / 491 (52%)	438 / 710 (62%)	321 / 683 (47%)	12
Chinese American	0 / 713 (0%)	6 / 607 (1.0%)	0 / 491 (0%)	0 / 710 (0%)	219 / 683 (32%)	24
Black/African-American	379 / 713 (53%)	194 / 607 (32%)	235 / 491 (48%)	0 / 710 (0%)	143 / 683 (21%)	7
Hispanic	3 / 713 (0.4%)	257 / 607 (42%)	0 / 491 (0%)	272 / 710 (38%)	0 / 683 (0%)	18

[1] n / N (%)

Step 1: Cleaning and Formatting Proteins

Input file names

- A table of protein abundances: SMP_IntensityNormalized_20251005.csv
- Sample information to link TOPMed IDs to unique MESA SHARe ID and exam combinations: Mapping_SMP_Plate_20251005.csv
- Keys to link Olink IDs to names compounds: MESAOlink3k_proteinKeys_03292023.csv
- A file to bridge SHARe ids (sidno) with MESA IDs (idno) MESA-SHARE_IDList_Labeled.csv

Raw file info

- The raw protein abundance file contained information on N=3040 protein assays, including those used for QC.
- When removing assays for QC, the raw protein abundance file contained information on N=2941 proteins.
- The protein abundance file contained information on N=14051 sample IDs (i.e., unique participant/exam combinations), including bridging samples.
- After removing QC samples (including bridging, controls, and one duplicate) the protein abundance file contained information on N=12739 sample IDs (i.e., unique participant/exam combinations).

Formatting

- Bridging (and other QC) samples were removed.
- Protein assays used for QC were removed.
- Proteins that should be excluded due to QC warnings (variable "QC_warning" set to "EXCLUDED") were removed, even though these do not have NPX values.
- Data were put into wide format, with "SampleID" as the unique ID, "OlinkID" forming the variable names (protein identifiers), and values taken from the "NPX" column.
  - In wide format, the file contained information on N=12739 unique sample IDs.
  - In wide format, the file contained information on N=0 duplicated sample IDs. [1]
- SHARe IDs, and subsequently MESA IDs, were merged into the file with exam information.

Table 2: Final N by exam

Exam	N_Pps
1	5949
5	3917
6	2873

- At this point, the range of unique SHARe ID by exam combinations was N=0 - 1. This indicates no sample ID were duplicated in the assays.
- The formatted protein file was used to calculate the coefficient of variation (CV) using the formula: CV = sqrt(2^(sigma^2)-1).
- A variable called "Retain" was created to indicate whether each protein was (1) unique (i.e., included on only one panel); (2) duplicated, and across all panels had the lowest CV; or (3) duplicated, and across all panels did not have the lowest CV.
- A final table of protein abundances, with additional variables for SHARe ID, MESA ID, Exam, TOPMed ID and Batch, was created after the steps above, with proteins duplicated across more than one panel cleaned such that only the one with the lowest CV is retained. This file was used in the analysis
- The number of participants, stratified by exam, in the final file is available in Table 2:

Step 2: Format Phenotypes

Input files

- Covariates from E1: MESAe1FinalLabel02092016.dta
- Covariates from E5: MESAe5_FinalLabel_20140613.dta
- Covariates from E6: MESAe6_FinalLabel_20220513.dta
- Afib info: SHARe_MesaEventsThruYear2020_AF_DS.txt
- ApoE info: MESA_ApoE_03102014.sas7bdat
- Incident CVD: MESAEvThru2020AllCohort_20241120.dta
- Microbleeds: MESAe6as253as301_BMRICMB_08052025.csv
- Perivascular spaces: MESAe6as253as301_BMRIPVS_20250310.csv
- White matter hyperintensities: MESAe6anyFIRST_BMRIWMHVol_20240422.csv
- Intracranial volumes: SHARe_AncilMesaAF_BMRIROIVol_DS.txt
- Fractional anisotropy: mesae6anyfirst_bmriTotalFAMUSE_20250828.csv
- White matter hyperintensities: MESAe6anyFIRST_BMRIWMHVol_20240422.csv

Formatting

Outcomes:

- Microbleeds were coded as 0/1, where 0= no microbleeds (value: 0) and 1 = presence of microbleeds (all non-zero values except missing). Then, those images with a low image quality (value = 4; N=0) were recoded to missing.
- Perivascular spaces (variable: epvs_wholebrain_vol) were recoded to missing where the variable 'pvs_exclude' was coded as 1 (N=0).
- White matter hyperintensities (variable: wm_wmh) were divided by 1000 to convert to ml (following Rizwan's code), and those where the variable wmh_exclude had a code of 1 were set to missing.
- Fractional anisotropy (variable: wmfa) was coded to missing where the variable fa_exclude had a value of 1.

Covariates

True time invariant covariates

- Race/ethnicity, gender, and highest education level were all taken from exam 1 data. ApoE information was taken from its own dataset (above).
- ApoE was coded 0/1/2 where 0= no e4 isoform (codes 22, 23, 33), 1 = e4 isoform (24, 34, 44), and 2 = no isoform data. The ApoE variable was formatted as a factor.
- Gender was coded such that female = 0 and male = 1.
- Education was recoded 0/1, such that 0 = less than high school (codes: 0: NO SCHOOLING / 1: GRADES 1-8 / 2: GRADES: 9-11) and 1= high school or more (all other codes, excluding missing).
- Race/ethnicity was recoded retaining the original MESA coded whereby 1=White American; 2= Chinese American, 3=Black, African-American, and 4 = Hispanic. Race/ethnicity was coded as a factor variable.

Pseudo-time invariant covariates

- Although some variables are technically time invariant, where they were included due to their effects on MRI data, since MRI data are only measured at one exam for this analysis (exam 6), these covariates were always taken from exam 6.
- These 'pseudo time invariant covariates' were: atrial fibrillation, myocardial infarction, congestive heart failure, LDL, systolic blood pressure, hypertension medication, and site (since site seems to affect MRI more than proteins??).
- Afib, MI, and CHF were coded 0/1, such that 0= no diagnosis and 1= afib diagnosis. Missing data was left as missing (this is different to Rizwan who coded missing data as no diagnosis).
- Site was arbitrarily coded as 0=Wake forest, 1= Columbia, 2=Johns Hopkins, 3=University of Minnesota, 4=Northwestern, 5=UCLA
- For the longitudinal analysis only, age at baseline (BL_age) was also included as a covariate, with age excluded from the model - this is because age is conflated with time (days since baseline) introducing a multicollinearity problem.

Time varying covariates

- The following covariates were taken from the exam when the proteins were used, as these were seen to affect proteins more in the short term than they affect MRI (?): kidney function (egfr), BMI , cigarette smoking (never/former/current; coded as ordinal), diabetes status.
- Smoking was harmonized and coded such that 0 = never smoker, 1= past smoker, 2=current smoker.
- Diabetes has harmonized and coded such that 0= no diabetes (including impaired fasting glucose), and 1 = diabetes (treated and untreated).
- The following covariates were taken from exam 6: age (age6c), kidney function (egfr; cepgfr6c), BMI (bmi6c), systolic blood pressure (sbp6c), LDL (ldl6), site(site6c), the use of hypertension medication (htnmed6c; coded 0= no, and 1= yes), cigarette smoking (cig6c), diabetes status (dm036t)

Missingness

The missingness for each variable those who have at least one MRI measure can be found in Table 3

Sample descriptives

- There were N=1429 MESA participants with at least one MRI outcome after the exclusions above.
- Of those with MRI data, N=1307 participants had protein data, equating to N=1307 at exam 1, N=1281 at exam 5, and N=1284 at exam 6
- Sample descriptives are available in Table 4

Table 3: Missingness for those who have at least one MRI measure

Variable	Frequency (N)	Proportion (%)
icv	307	23.9
mb_present	252	19.6
epvs	197	15.3
fa	52	4.05
ldl	34	2.65
egfr	26	2.02
diabetes	25	1.95
wmh	19	1.48
htnmeds	11	0.857
E4	10	0.779
AFprevalent	4	0.312
edu	2	0.156
smoking	2	0.156
sbp	2	0.156
BMI	1	0.0779
MIprevalent	1	0.0779
CHFprevalent	1	0.0779
idno	0	0
sidno	0	0
Exam	0	0
time	0	0
BL_age	0	0
age	0	0
gender	0	0
site	0	0
race	0	0

Table 4: Sample Descriptives

Characteristic	Exam		
	1 N = 1,307[1]	**5** N = 1,281[1]	**6** N = 1,284[1]
Age (y)	56.40 (7.96)	65.84 (7.83)	72.20 (7.79)
Gender			
Female	689 / 1,307 (53%)	676 / 1,281 (53%)	676 / 1,284 (53%)
Male	618 / 1,307 (47%)	605 / 1,281 (47%)	608 / 1,284 (47%)
Field Center			
Wake Forest	239 / 1,284 (19%)	235 / 1,259 (19%)	239 / 1,284 (19%)
Columbia	204 / 1,284 (16%)	199 / 1,259 (16%)	204 / 1,284 (16%)
Johns Hopkins	165 / 1,284 (13%)	161 / 1,259 (13%)	165 / 1,284 (13%)
Minnesota	239 / 1,284 (19%)	232 / 1,259 (18%)	239 / 1,284 (19%)
Northwestern	228 / 1,284 (18%)	227 / 1,259 (18%)	228 / 1,284 (18%)
UCLA	209 / 1,284 (16%)	205 / 1,259 (16%)	209 / 1,284 (16%)
Highest education level			
Up to and including high school	133 / 1,305 (10%)	129 / 1,279 (10%)	130 / 1,282 (10%)
More than high school	1,172 / 1,305 (90%)	1,150 / 1,279 (90%)	1,152 / 1,282 (90%)
Race or ethnicity			
Non-Hispanic White	553 / 1,307 (42%)	545 / 1,281 (43%)	542 / 1,284 (42%)
Chinese American	157 / 1,307 (12%)	156 / 1,281 (12%)	157 / 1,284 (12%)
Black/African-American	350 / 1,307 (27%)	340 / 1,281 (27%)	345 / 1,284 (27%)
Hispanic	247 / 1,307 (19%)	240 / 1,281 (19%)	240 / 1,284 (19%)
BMI (kg/m^2^)	27.95 (5.21)	28.41 (5.36)	28.35 (5.44)
Smoking status			
Never	699 / 1,305 (54%)	609 / 1,277 (48%)	615 / 1,282 (48%)
Former	463 / 1,305 (35%)	574 / 1,277 (45%)	592 / 1,282 (46%)
Current	143 / 1,305 (11%)	94 / 1,277 (7.4%)	75 / 1,282 (5.9%)
LDL levels	107.96 (35.28)	107.96 (34.98)	107.96 (35.28)
systolic blood pressure	126.34 (20.35)	126.37 (20.37)	126.34 (20.35)
Diabetes status			

Normoglycemia/IFG	1,219 / 1,303 (94%)	1,076 / 1,280 (84%)	994 / 1,259 (79%)
Diabetes (treated or untreated)	84 / 1,303 (6.4%)	204 / 1,280 (16%)	265 / 1,259 (21%)
Takes hypertentsion medicine			
No	537 / 1,273 (42%)	527 / 1,248 (42%)	537 / 1,273 (42%)
Yes	736 / 1,273 (58%)	721 / 1,248 (58%)	736 / 1,273 (58%)
Atrial fibrillation			
No	1,095 / 1,280 (86%)	1,073 / 1,255 (85%)	1,095 / 1,280 (86%)
Yes	185 / 1,280 (14%)	182 / 1,255 (15%)	185 / 1,280 (14%)
Myocardial Infarction			
No	1,245 / 1,283 (97%)	1,220 / 1,258 (97%)	1,245 / 1,283 (97%)
Yes	38 / 1,283 (3.0%)	38 / 1,258 (3.0%)	38 / 1,283 (3.0%)
Coronary Heart Failure			
No	1,261 / 1,283 (98%)	1,236 / 1,258 (98%)	1,261 / 1,283 (98%)
Yes	22 / 1,283 (1.7%)	22 / 1,258 (1.7%)	22 / 1,283 (1.7%)
ApoeE information			
No E4 isoform	916 / 1,274 (72%)	899 / 1,249 (72%)	916 / 1,274 (72%)
E4 isoform	348 / 1,274 (27%)	340 / 1,249 (27%)	348 / 1,274 (27%)
No ApoE data	10 / 1,274 (0.8%)	10 / 1,249 (0.8%)	10 / 1,274 (0.8%)
Kidney function (egfr)	82.72 (15.47)	82.46 (19.44)	76.86 (19.53)
Intracranial volume	1,360,304.37 (145,625.45)	1,361,045.01 (145,456.35)	1,360,304.37 (145,625.45)
Fractional anisotropy	0.40 (0.03)	0.40 (0.03)	0.40 (0.03)
White matter hyperintensities	6.74 (10.15)	6.62 (9.89)	6.71 (10.16)
Enlarged perivascular spaces	3,701.67 (2,374.34)	3,702.34 (2,383.48)	3,701.67 (2,374.34)
Presence of microbleeds?			
No	661 / 1,032 (64%)	652 / 1,017 (64%)	661 / 1,032 (64%)
Yes	371 / 1,032 (36%)	365 / 1,017 (36%)	371 / 1,032 (36%)

[1]Mean (SD); n / N (%)

Step 3: Cross sectional Protein-Wide Association Studies (PWAS)

This section conducts a cross-sectional PWAS, using the set of covariates chosen by Tim & Lekki, and provides background to the longitudinal analysis (main analysis of interest) only.

- This section includes PWAS where proteins were only used at one time point (exam 1 or exam 6), even if this was not the same time point as when the MRI data were used.
- All estimates are standardized.

White matter hyperintensity (WMH)

Model Specification

- The associations between WMH and proteins were analyzed via linear regression
- For WMH the numeric covariates included icv, age, egfr, BMI. [2]
- For WMH the factor covariates included gender, race, edu, smoking, E4. [2]
- WMH was transformed using an inverse normal transformation (with blom constant).

WMH: Exam 6 proteins

- The E6 proteins `->` E6 MRI for WMH included data from N=951 participants.
- Full results can be found here
- A total of N=0 exam 6 proteins were significantly associated with WMH at an FDR corrected `P<.05`.

WMH: Exam 1 proteins

- The E1 proteins `->` E6 MRI for WMH included data from N=939 participants.
- Full results can be found here
- A total of N=0 exam 1 proteins were significantly associated with WMH at an FDR corrected `P<.05`.

Perivascular Spaces (PVS)

Model Specification

- The associations between PVS and proteins were analyzed via linear regression
- For PVS the numeric covariates included icv, age, egfr, BMI. [2]
- For PVS the factor covariates included gender, race, edu, smoking, E4. [2]
- PVS was transformed using an inverse normal transformation (with blom constant).

PVS: Exam 6 proteins

- The E6 proteins -> E6 MRI for PVS included data from N=944 participants.
- Full results can be found here
- A total of N=0 exam 1 proteins were significantly associated with PVS at an FDR corrected P<.05

PVS: Exam 1 proteins

- The E1 proteins -> E6 MRI for PVS included data from N=932 participants.
- Full results can be found here
- A total of N=0 exam 1 proteins were significantly associated with PVS at an FDR corrected P<.05.

Fractional Anisotropy (FA)

Model Specification

- The associations between FA and proteins were analyzed via linear regression
- For FA the numeric covariates included age, egfr. [2]
- For FA the factor covariates included gender, race, edu, smoking, E4. [2]
- FA was transformed using an inverse normal transformation (with blom constant).

FA: Exam 6 proteins

- The E6 proteins -> E6 MRI for FA included data from N=1190 participants.
- Full results can be found here
- A total of N=33 exam 6 proteins were significantly associated with FA at an FDR corrected P<.05 (see Table 5 and Figure 1).

Table 5: Associations Between Exam 6 Proteins and Fractional Anisotropy Reaching an FDR-corrected P<.05 Equivalent

Protein	N	Beta	SE	P	P-FDR	Olink ID	Protein CV	UniProt ID
CLSTN2	1189	-0.09	0.03	4.78×10^{-4}	4.51×10^{-2}	OID20664	26.60	Q9H4D0
CRIM1	1189	-0.12	0.03	6.84×10^{-5}	1.33×10^{-2}	OID20701	33.72	Q9NZV1
PLAUR	1189	-0.14	0.03	3.47×10^{-6}	2.03×10^{-3}	OID20764	35.07	Q03405
FSTL3	1189	-0.14	0.03	3.11×10^{-5}	9.09×10^{-3}	OID20782	38.82	O95633
REG4	1189	-0.10	0.03	3.27×10^{-4}	3.67×10^{-2}	OID20784	40.16	Q9BYZ8
PPL	1190	-0.11	0.03	1.79×10^{-4}	2.75×10^{-2}	OID30564	41.64	O60437
CRELD1	1190	-0.10	0.03	4.68×10^{-4}	4.51×10^{-2}	OID30619	41.45	Q96HD1
RNASE1	1190	-0.13	0.03	5.18×10^{-5}	1.16×10^{-2}	OID30672	28.16	P07998
DCC	1190	0.11	0.03	3.43×10^{-5}	9.12×10^{-3}	OID30953	33.38	P43146
LRTM2	1190	0.10	0.03	2.48×10^{-4}	3.30×10^{-2}	OID31020	36.45	Q8N967
PALM2	1190	-0.11	0.03	3.24×10^{-4}	3.67×10^{-2}	OID31096	30.99	Q8IXS6
LMOD1	1189	-0.15	0.03	9.05×10^{-6}	4.41×10^{-3}	OID30212	19.03	P29536
FAM20A	1190	-0.10	0.03	3.51×10^{-4}	3.79×10^{-2}	OID30348	30.10	Q96MK3
FABP3	1190	-0.13	0.03	3.89×10^{-5}	9.48×10^{-3}	OID30351	53.39	P05413
CTHRC1	1190	-0.14	0.03	8.89×10^{-7}	1.16×10^{-3}	OID30393	27.92	Q96CG8
ACTA2	1190	-0.12	0.03	1.60×10^{-4}	2.59×10^{-2}	OID20079	27.45	P62736
IL6	1190	-0.10	0.03	2.04×10^{-4}	2.85×10^{-2}	OID20101	35.49	P05231
IL19	1190	-0.10	0.03	5.00×10^{-4}	4.57×10^{-2}	OID20158	34.07	Q9UHD0
CLEC5A	1190	-0.14	0.03	3.01×10^{-6}	2.03×10^{-3}	OID20165	47.48	Q9NY25
MEP1B	1190	0.10	0.03	3.63×10^{-4}	3.79×10^{-2}	OID20168	36.75	Q16820
SIGLEC7	1190	-0.10	0.03	1.08×10^{-4}	1.85×10^{-2}	OID20190	33.60	Q9Y286
LILRA5	1190	-0.12	0.03	2.37×10^{-5}	7.69×10^{-3}	OID20209	34.07	A6NI73
CTSL	1190	-0.13	0.03	1.19×10^{-6}	1.16×10^{-3}	OID20228	31.03	P07711
UMOD	1190	0.11	0.03	5.99×10^{-5}	1.25×10^{-2}	OID20237	40.83	P07911
GDF15	1190	-0.13	0.03	1.82×10^{-5}	6.65×10^{-3}	OID20251	39.14	Q99988
CA3	1190	-0.10	0.03	4.41×10^{-4}	4.45×10^{-2}	OID20271	57.80	P07451
SELE	1190	-0.10	0.03	1.89×10^{-4}	2.77×10^{-2}	OID20290	64.86	P16581
CCL14	1190	-0.12	0.03	9.74×10^{-5}	1.78×10^{-2}	OID20401	37.76	Q16627
TNFRSF6B	1190	-0.10	0.03	5.41×10^{-4}	4.79×10^{-2}	OID20964	59.96	O95407
TIMP4	1190	-0.10	0.03	2.99×10^{-4}	3.64×10^{-2}	OID21147	37.68	Q99727
OMG	1190	0.12	0.03	1.15×10^{-5}	4.82×10^{-3}	OID21352	71.71	P23515
S100A12	1190	-0.10	0.03	2.83×10^{-4}	3.60×10^{-2}	OID21374	94.18	P80511
HAVCR1	1190	-0.17	0.03	5.28×10^{-10}	1.54×10^{-6}	OID21422	95.89	Q96D42

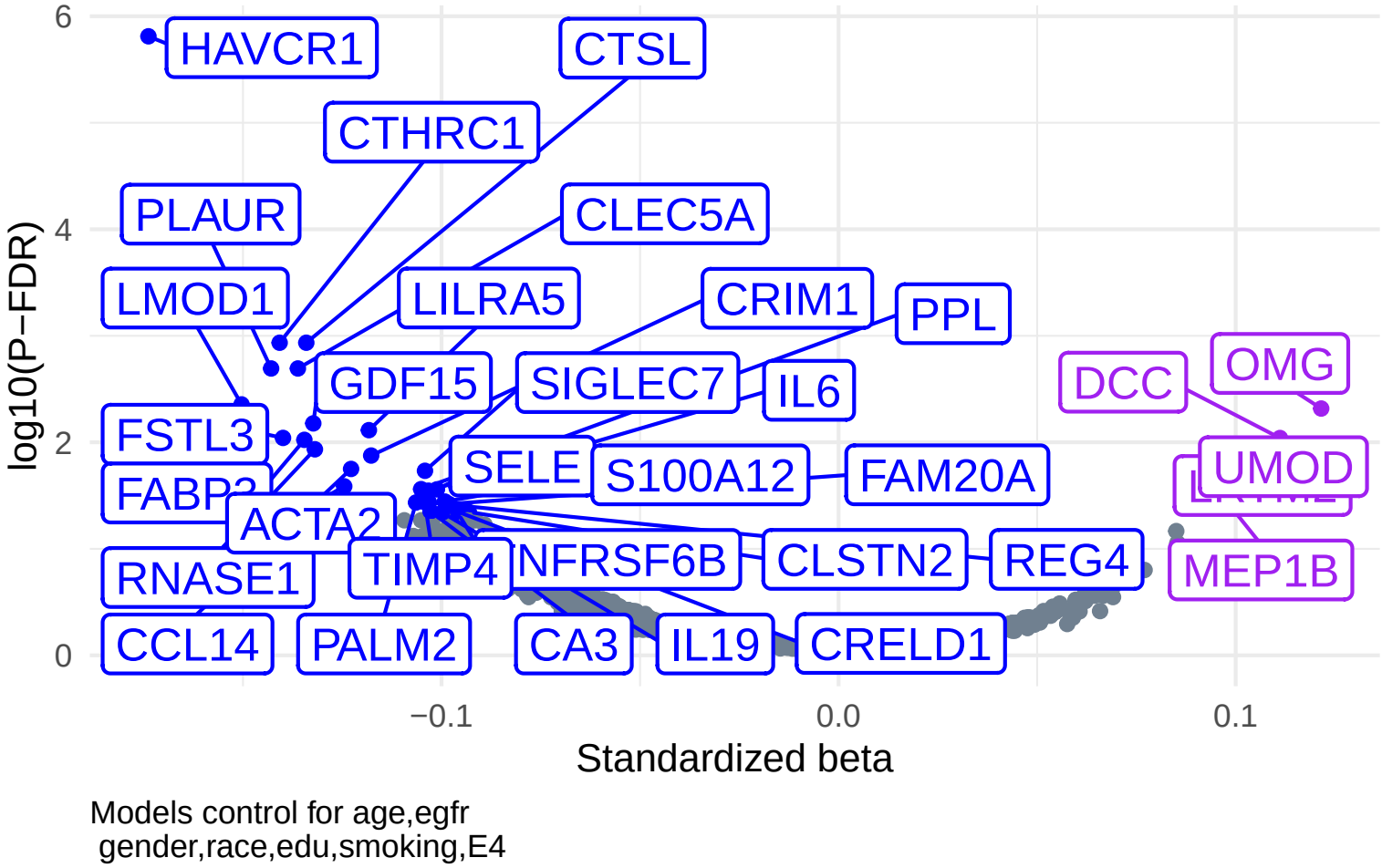

Figure 1: Associations Between Exam 6 Proteins and Fractional Anisotropy

FA: Exam 1 proteins

- The E1 proteins -> E6 MRI for FA included data from N=1187 participants.
- Full results can be found here.
- A total of N=628 exam 1 proteins were significantly associated with FA at an FDR corrected P<.05 (see Figure 4 in Appendix 1).
- Due to the large number of predictors, a penalized model was run (elastic net penalty) on a 50:50 test:training split of the data. The full results can be found here.
- From the N=628 exam 1 proteins significantly associated with FA at an FDR corrected P<.05, the LASSO identified N= proteins as variables of importance. These are described / visualized in **?@tbl-fa-E1-LASSO-reducedcov** and Figure 2 below.

Protein	N	Beta	SE	P	P-FDR	Olink ID	Protein CV	UniProt ID
TNFAIP8	1187	-0.08	0.03	4.36×10^{-3}	3.06×10^{-2}	OID20433	82.37	O95379
IL17A	1187	-0.08	0.03	3.80×10^{-3}	2.79×10^{-2}	OID20469	36.02	Q16552
ITGA6	1187	-0.08	0.03	2.23×10^{-3}	2.11×10^{-2}	OID20528	35.37	P23229
TGFB1	1187	-0.14	0.03	4.68×10^{-7}	5.63×10^{-4}	OID20621	29.64	P01137
TPP1	1187	-0.07	0.03	6.45×10^{-3}	3.79×10^{-2}	OID20750	36.50	O14773
MEPE	1187	0.08	0.03	7.84×10^{-3}	4.20×10^{-2}	OID20753	38.06	Q9NQ76
ATRN	1187	-0.09	0.03	5.44×10^{-4}	1.07×10^{-2}	OID30735	34.41	O75882-2
IGLC2	1187	-0.09	0.03	1.02×10^{-3}	1.39×10^{-2}	OID30740	37.37	P0DOY2
C5	1187	-0.08	0.03	2.18×10^{-3}	2.11×10^{-2}	OID30744	28.65	P01031
F13B	1187	-0.10	0.03	3.11×10^{-4}	9.17×10^{-3}	OID30781	32.26	P05160
FGA	1187	-0.07	0.03	6.85×10^{-3}	3.92×10^{-2}	OID30788	33.22	P02671
CAMLG	1187	-0.07	0.03	9.10×10^{-3}	4.56×10^{-2}	OID30952	59.08	P49069
PTH	1187	-0.08	0.03	2.05×10^{-3}	2.06×10^{-2}	OID30963	113.75	P01270
AZI2	1187	-0.07	0.03	5.66×10^{-3}	3.53×10^{-2}	OID31056	76.07	Q9H6S1
TAX1BP1	1187	-0.10	0.03	1.46×10^{-4}	6.53×10^{-3}	OID31075	58.89	Q86VP1
GAST	1187	-0.10	0.03	2.13×10^{-4}	8.09×10^{-3}	OID31083	40.86	P01350
PTPRR	1187	0.10	0.03	6.43×10^{-4}	1.13×10^{-2}	OID31099	27.29	Q15256
DNAJC21	1187	-0.10	0.03	4.98×10^{-4}	1.02×10^{-2}	OID31104	52.79	Q5F1R6
SMTN	1187	-0.08	0.03	5.12×10^{-3}	3.31×10^{-2}	OID31154	116.70	P53814
SEL1L	1187	-0.09	0.03	1.52×10^{-3}	1.77×10^{-2}	OID30196	17.38	Q9UBV2
LMOD1	1187	-0.14	0.03	6.49×10^{-6}	1.18×10^{-3}	OID30212	19.03	P29536
MYL3	1187	-0.10	0.03	4.94×10^{-4}	1.02×10^{-2}	OID30251	22.36	P08590
NPTX2	1187	-0.09	0.03	3.31×10^{-3}	2.66×10^{-2}	OID30390	27.37	P47972
TSPAN1	1187	-0.08	0.03	5.69×10^{-3}	3.54×10^{-2}	OID20078	21.70	O60635
ADAMTS16	1187	0.08	0.03	4.52×10^{-3}	3.12×10^{-2}	OID20089	29.43	Q8TE57
GP2	1187	-0.10	0.03	5.06×10^{-4}	1.02×10^{-2}	OID20090	32.25	P55259
GDF15	1187	-0.12	0.03	9.78×10^{-5}	5.51×10^{-3}	OID20251	39.14	Q99988
NID1	1187	-0.11	0.03	1.68×10^{-4}	7.10×10^{-3}	OID20362	34.78	P14543
SELP	1187	-0.07	0.03	7.79×10^{-3}	4.20×10^{-2}	OID20379	38.45	P16109
PAK4	1187	-0.09	0.03	9.70×10^{-4}	1.36×10^{-2}	OID20819	33.21	O96013
NXPH1	1187	0.07	0.03	8.23×10^{-3}	4.34×10^{-2}	OID20849	33.12	P58417
CRIP2	1187	-0.09	0.03	9.25×10^{-4}	1.35×10^{-2}	OID20901	39.37	P52943
PTPRN2	1187	0.12	0.03	2.05×10^{-5}	2.50×10^{-3}	OID20912	48.03	Q92932
CDCP1	1187	-0.08	0.03	9.11×10^{-3}	4.56×10^{-2}	OID20940	40.17	Q9H5V8
THBS2	1187	-0.11	0.03	6.88×10^{-5}	5.41×10^{-3}	OID21104	32.08	P35442
IGF2R	1187	-0.09	0.03	1.41×10^{-3}	1.68×10^{-2}	OID21146	33.17	P11717

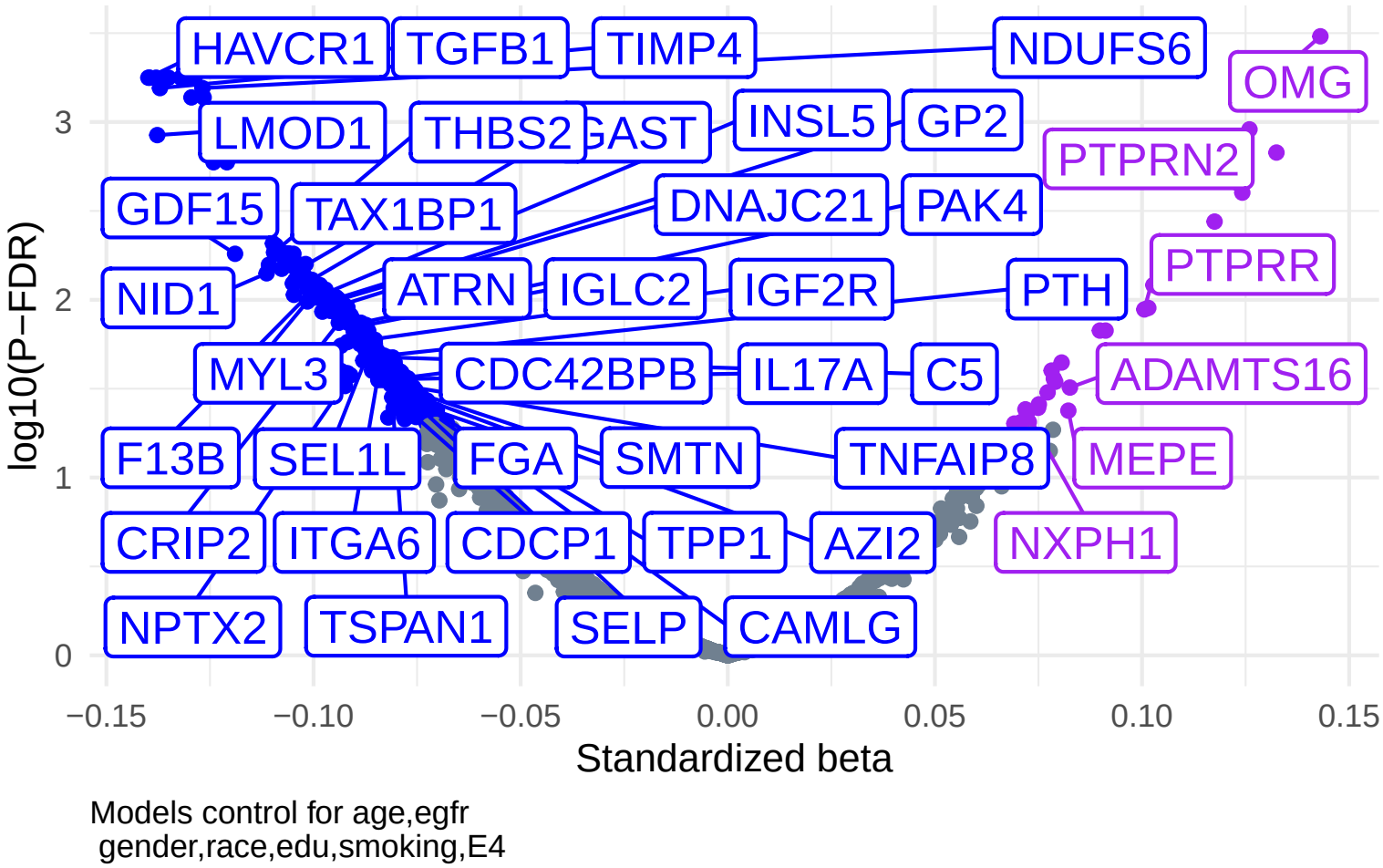

Figure 2: Associations Between Exam 1 Proteins and Fractional Anisotropy, labelling only those selected by LASSO

Microbleeds (MB)

Model Specification

- The associations between MB and proteins were analyzed via logistic regression (MB coded as 'present' vs. 'absent' as above)
- For MB the numeric covariates included age, egfr, BMI. [2]
- For MB the factor covariates included gender, race, edu, smoking, E4. [2]

MB: Exam 6 proteins

- The E6 proteins -> E6 MRI for MB included data from N= 994 participants.
- Full results can be found here
- A total of N=1 exam 6 proteins were significantly associated with MB at an FDR corrected P<.05 (see Table 6 and Figure 3).

MB: Exam 1 proteins

- The E1 proteins -> E6 MRI for FA included data from N=986 participants.
- Full results can be found here
- A total of N=0 exam 1 proteins were significantly associated with MB at an FDR corrected P<.05.

Table 6: Associations Between Exam 6 Proteins and Microbleeds Reaching an FDR-corrected P<.05 Equivalent

Protein	N	OR	Lower CI	Upper CI	P	P-FDR	Olink ID	Protein CV	UniProt ID
ADGRB3	994	1.47	1.23	1.74	1.46×10^{-5}	4.25×10^{-2}	OID20974	53.40	O60242

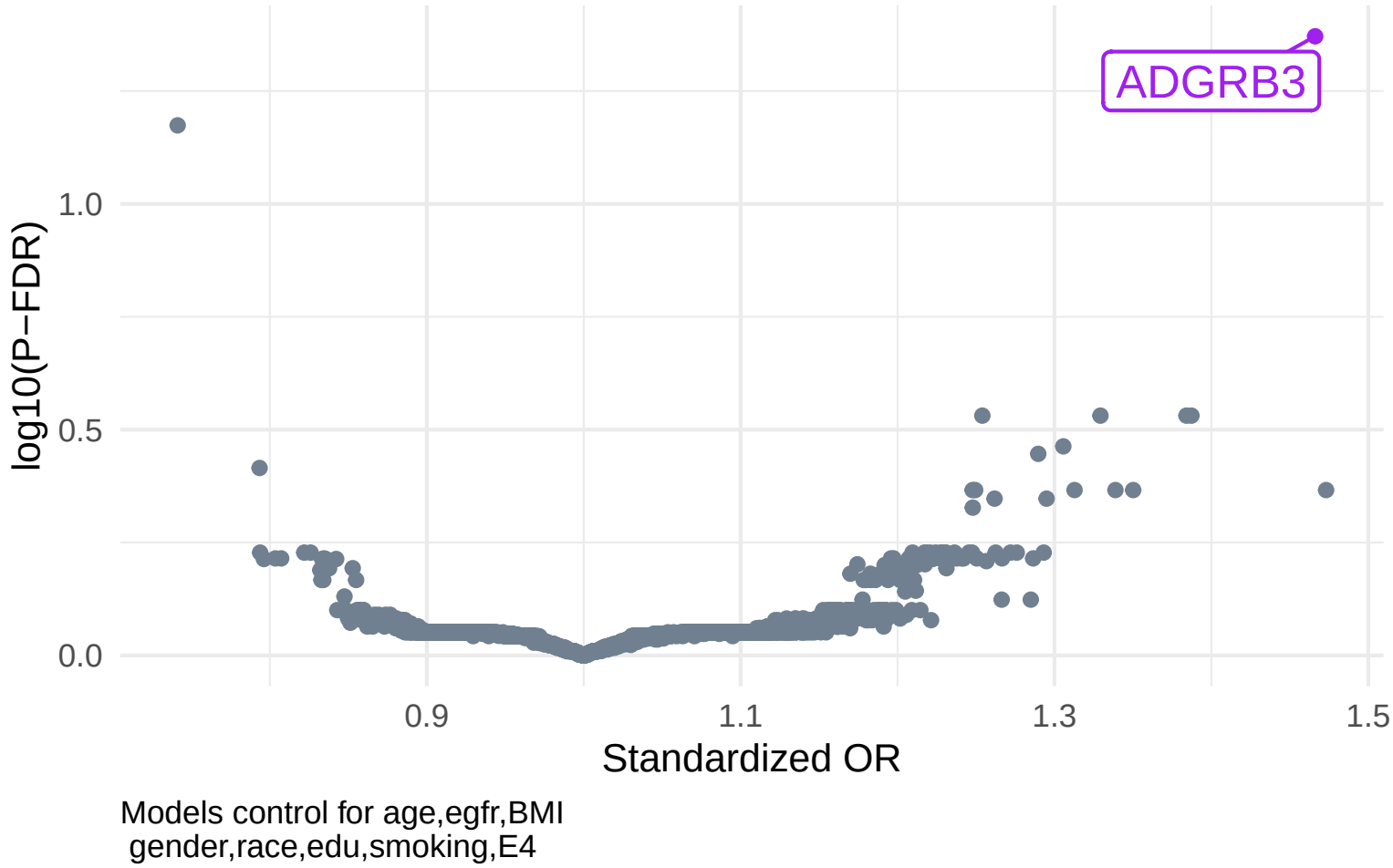

Figure 3: Associations Between Exam 6 Proteins and Fractional Anisotropy

Step 4: Longitudinal models

- This section includes PWAS where proteins were only used at three time points (exam 1, exam 5 and exam 6) in a single model.
- Accordingly, all models were nested within person
- Note: that for the model to have any within-person variance (needed for an MLM) proteins had to be the outcome
- In each model, the MRI was allowed to have an interaction with time, which is the parameter of interest. This parameter is interpreted as showing whether each MRI measure is associated with a different linear change in the protein over time, in the ~20 years prior to the MRI being conducted.
- All estimates are standardized.
- The covariate choice was decided between Tim and Lekki, and mirrors that of the cross-sectional results.

White matter hyperintensity (WMH)

Model Specification

- The associations between WMH and proteins were analyzed via linear regression
- For WMH the numeric covariates included time, BL_age, egfr, icv. [2]
- For WMH the factor covariates included gender, race, edu, smoking, E4. [2]
- WMH was transformed using an inverse normal transformation (with blom constant).

WMH: Longitudinal Associations with Prior Change in Proteins

- The longitudinal WMH model included data from N=2797 participants.
- Full results can be found here
- A total of N=232 proteins were significantly associated with WMH at an FDR corrected P<.05 (see Figure 5 in Appendix 2).

Appendix 1 : Figures displaying full results for assocations between MRI and single time point proteins, where there are many associations

Fractional anisotropy and exam 1 proteins

Appendix 2: Figures for Full Longitudinal Association Results

White Matter Hyperintensities

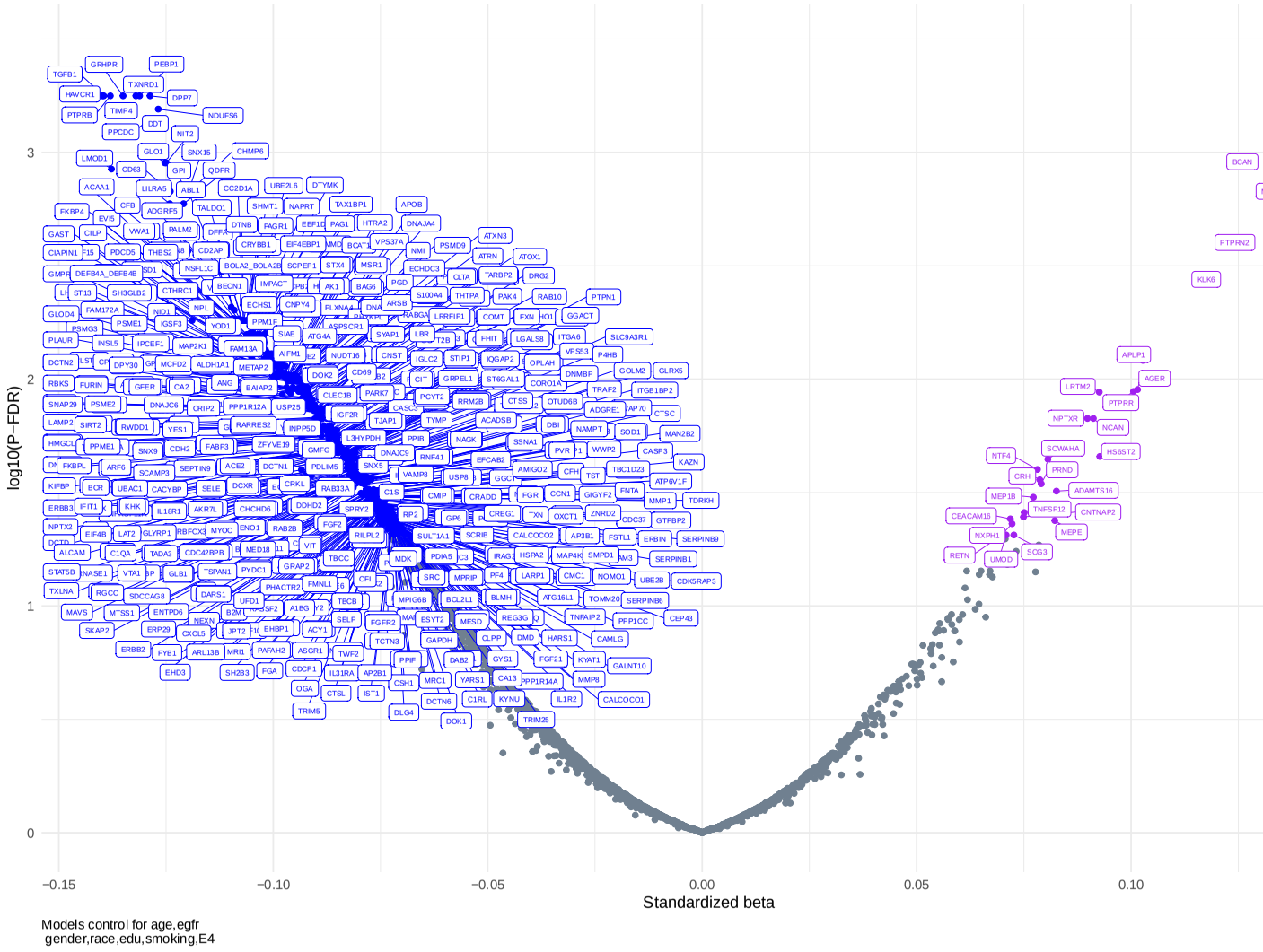

Figure 4: Associations Between Exam 1 Proteins and Fractional Anisotropy

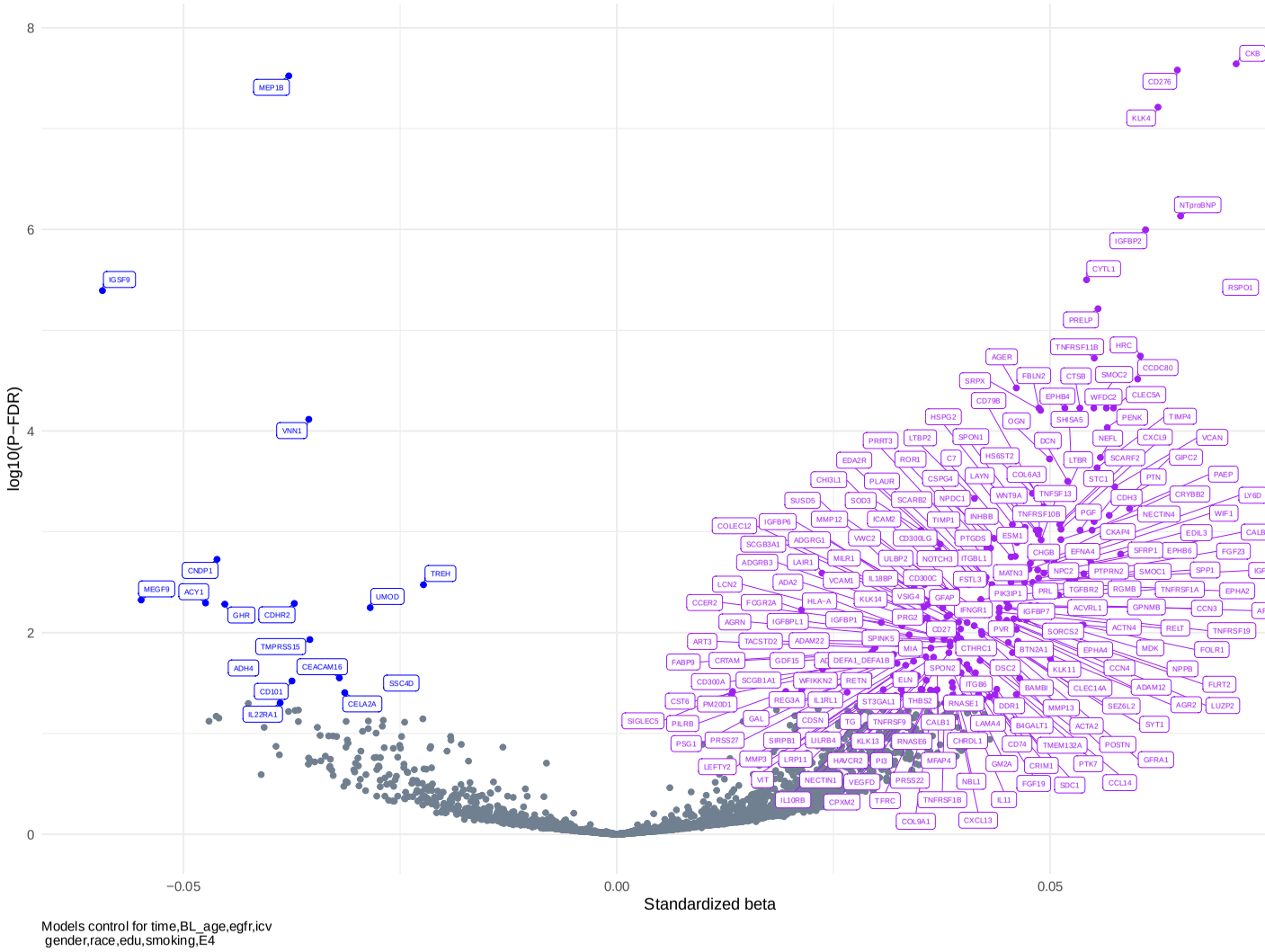

Figure 5: Associations Between Protein Trajectories and Subsequent White Matter Hyperintensities