

1 Genetic analysis of cognitive preservation in the midwestern
2 Amish reveals a novel locus on chromosome 2.

3

4 Author names and affiliations: Leighanne R Main*_{a,b,c},
5 Yeunjoo E Song_{b,c}, Audrey Lynn_{b,c}, Renee A Laux_b, Kristy
6 L Miskimen_b, Michael D Osterman_b, Michael L Cuccaro_{d,e},
7 Paula K Ogrocki_{f,g}, Alan J Lerner_{f,g}, Jeffery M Vance_{d,e}, M
8 Denise Fuzzell_b, Sarada L Fuzzell_b, Sherri D Hochstetler_b,
9 Daniel A Dorfsman_{d,e}, Laura J Caywood_d, Michael B
10 Prough_d, Larry D Adams_d, Jason E Clouse_d, Sharlene D
11 Herington_d, William K Scott_{d,e}, Margaret A Pericak-Vance
12_{d,e}, Jonathan L Haines*_{a,b,c}

13

14_a Departments of Genetics and Genome Sciences, Case
15 Western Reserve University, 10900 Euclid Ave, Cleveland,
16 OH, USA, 44106

17_b Department of Population and Quantitative Health
18 Sciences, Case Western Reserve University School of
19 Medicine, 10900 Euclid Ave, Cleveland, OH, USA, 44016

20_c Cleveland Institute of Computational Biology, Case
21 Western Reserve University School of Medicine, 10900
22 Euclid Ave, Cleveland, OH, USA, 44106

23 d John P Hussman Institute for Human Genomics, University
24 of Miami Miller School of Medicine, 1501 NW 10th Ave,
25 Miami, FL, USA, 33136
26 e Dr. John T Macdonald Foundation Department of Human
27 Genetics, University of Miami Miller School of Medicine,
28 1501 NW 10th Ave, Miami, FL, USA, 33136
29 f Department of Neurology, University Hospitals Cleveland
30 Medical Center, 11100 Euclid Ave, Cleveland, OH, USA,
31 44106
32 g Department of Neurology, Case Western Reserve
33 University School of Medicine, 10900 Euclid Ave, Cleveland,
34 OH, USA, 44106
35
36 *Corresponding Authors: Leighanne R Main
37 (lrm91@case.edu), and Jonathan L Haines
38 (jlh213@case.edu)

39 ABSTRACT

40 INTRODUCTION: Alzheimer disease (AD) remains a
41 debilitating condition with limited treatments and additional
42 therapeutic targets needed. Identifying AD protective genetic
43 loci may identify new targets and accelerate identification of
44 therapeutic treatments. We examined a founder population
45 to identify loci associated with cognitive preservation into
46 advanced age.

47 METHODS: Genome-wide association and linkage analyses
48 were performed on 946 examined and sampled Amish
49 individuals, aged 76-95, who were either cognitively
50 unimpaired (CU) or impaired (CI).

51 RESULTS: 12 SNPs demonstrated suggestive association
52 ($P \leq 5 \times 10^{-4}$) with cognitive preservation. Genetic linkage
53 analyses identified >100 significant ($\text{LOD} \geq 3.3$) SNPs, some
54 which overlapped with the association results. Only one
55 locus on chromosome 2 retained significance across multiple
56 analyses.

57 DISCUSSION: A novel significant result for cognitive
58 preservation on chromosome 2 includes the genes *LRRTM4*
59 and *CTNNA2*. Additionally, the lead SNP, rs1402906,
60 impacts the POU3F2 transcription factor binding affinity,
61 which regulates *LRRTM4* and *CTNNA2*.

62 ABBREVIATIONS

- 63 AD – Alzheimer Disease
- 64 AGDB – Anabaptist Genealogical Database
- 65 CI – Cognitively impaired
- 66 CU – Cognitively unimpaired
- 67 GWAS – Genome-wide association study
- 68 HLOD – Heterogeneity logarithm of the odds
- 69 LD – Linkage disequilibrium
- 70 LOD – Logarithm of the odds
- 71 MCI – Mild cognitive impairment
- 72 MCMC – Markov chain Monte Carlo
- 73 QC – Quality control

74

75 1 BACKGROUND

76 Dementia, defined broadly as a decline or loss in memory
 77 and other cognitive abilities, is an overarching term that
 78 incorporates many underlying conditions. Alzheimer disease
 79 (AD), the most common form of dementia (~60-80% of
 80 cases), affects more than five million people in the U.S. and
 81 is becoming more prevalent as the world population ages.[1]
 82 To date there have been over 70 genetic loci associated with
 83 *increased* risk for AD-associated dementia,[2–4] some of
 84 which have been targets for therapeutic development.
 85 However, no existing treatments or therapies addressing AD
 86 and other dementias have been sufficient to significantly

87 ameliorate AD pathology or delay onset and decline.[5] An
 88 alternative and less often used approach for generating
 89 therapeutic targets is the identification of genetic variation
 90 that associates with a *decreased* risk for AD-associated
 91 dementia (e.g., protective loci). Focusing on protective loci is
 92 a strategy demonstrating higher success rates than focusing
 93 on loci with increased risk when translated into
 94 therapeutics.[6]
 95
 96 Phenotypically, protective AD genetic loci result in
 97 maintenance of cognitive abilities (e.g., cognitive
 98 preservation). To identify loci promoting cognitive
 99 preservation, we analyzed members of the Ohio and Indiana
 100 Amish, a founder population of European ancestry. For our
 101 analyses, we identified older individuals who were
 102 cognitively unimpaired (CU) but at high risk for developing
 103 AD, defined as having a first-degree relative with dementia.
 104 Analyses included cognitively impaired (CI) individuals as
 105 controls. The relatively uniform lifestyle of the midwestern
 106 Amish reduces the variability of environmental influences on
 107 many phenotypes, while their endogamous cultural norms
 108 increase the frequency of some normally rare genomic
 109 variants.[7,8] Both genetic association analyses and family-

110 based linkage analyses were used to identify loci associated
111 with cognitive preservation.

112 2 METHODS

113 2.1 Data Collection and Cognitive Screening

114 The analyzed dataset consisted of Amish individuals from
115 Ohio and Indiana who were 76-95 years old at their last
116 exam. Individuals underwent a panel of neurocognitive
117 assessments to evaluate their cognitive status. Assessments
118 included the Consortium to Establish a Registry for
119 Alzheimer's Disease (CERAD) neuropsychological
120 assessment[9], the Modified Mini-Mental State (3MS)
121 exam[10], the AD8 Dementia Screening Interview[11], and
122 the Trail Making Test for Dementia Studies.[12] Cognitive
123 data from 2,096 individuals were available for analysis.
124 Individuals were classified as cognitively impaired (CI) based
125 on established testing thresholds, with determinations
126 confirmed by a clinical adjudication board. Categories
127 defined by the adjudication board included cognitively
128 unimpaired for age-normed benchmarks (CU), mild cognitive
129 impairment (MCI), borderline impairment, or cognitively
130 impaired (CI). Additional personal and medical information
131 was assessed to provide context for other factors related to
132 cognition scores, such as self-reported depression.

133 Individuals unable to be categorized as CI or CU were
134 excluded from genetic analyses. All participants provided
135 written informed consent for participation in this study.

136

137 2.2 Genotyping

138 Blood was collected from participants for DNA extraction. All
139 samples were genotyped via the Multi-Ethnic Genotyping
140 Array (MEGA^{ex}) or Global Screening Array (GSA) genotyping
141 chips. The MEGA^{ex} chip recognized over 2 million loci, and
142 the GSA chip recognized 660,000 loci. Custom markers
143 were also included on the MEGA^{ex} chip (6,000 SNPs),
144 including disease-associated loci previously identified in
145 Amish and European populations.[13,14]

146

147 2.3 Genotype Quality Control

148 Initial quality control (QC) was done as previously
149 described.[15] 256,978 SNPs passed QC across both chips.
150 After additional pruning for linkage disequilibrium (LD),
151 167,196 SNPs were used to inform the number of effective
152 independent tests done in the association analyses.

153

154 Genotyped individuals were filtered for sex discrepancies
155 and low call rates (<96%). Familial relationships were

verified across samples. Age was defined as follows for all analyses: for CU individuals, age at the last study visit; for CI individuals, age at CI diagnosis. Participants whose age or familial relationships were missing were not included in the current analyses. Additional samples were dropped due to being under 75 years of age ($n = 748$) or a diagnosis other than CI/CU ($n = 343$). After QC, 946 individuals were available for analysis (66% CU; 59% female, Supplementary Table 1). Mean age for both CU and CI individuals was 82 years (Supplementary Figure 1).

2.4 Amish Genealogy

To evaluate familial relationships in this dataset, we consulted the Anabaptist Genealogical Database (AGDB).[16] All individuals were connected into a single, large, 14-generation pedigree ($n=8,222$) generated using pedigree[17] (Figure 1A). These relationships were used to correct for relatedness in our genetic analyses. To conduct chromosome-wide genetic linkage analyses smaller pedigrees were required, so PedCut[18] was used to create sub-pedigrees with at least two cognitively preserved individuals in each (see Figure 1B for an example). The 14-generation pedigree was split based on pedigree size, or bit-

size. Bit sizes evaluated were 21 ($n = 103$ sub-pedigrees) and 22 ($n = 98$) for linkage analysis on the autosomes. An additional 23-bit ($n = 94$) pedigree size was used for linkage analysis on the X chromosome.

183

2.5 Principal Component Analysis and Association Analysis

Principal Component Analysis (PCA) was performed via Principal Component Analysis in Related Samples (PC-AiR), part of the GENetic Estimation and Inference in Structured samples (GENESIS) R package.[19] PC-AiR adjusted for cryptic relatedness to generate principal components (PCs) based on the overall population structure, not just family structure. PC-AiR created kinship estimates for each individual based on the pedigree relationship information obtained from the 14-generation pedigree (Figure 1A). Model-free estimation of recent genetic relatedness (PC-Relate)[20] then used the PCs generated from PC-AiR to adjust for the population substructure and ancestry to provide an accurate estimate of recent genetic relatedness due to family structure, correcting for the extensive and complex relatedness in this dataset by generating a Genetic Relationship Matrix (GRM). GENESIS software was then

202 used for association analysis.[19] Association analysis was
203 completed using covariates of age, sex, and the GRM,
204 adjusting for the first two PCs.

205
206 To accommodate the LD structure specific to the Amish, we
207 used the simpleM method[21–23] to define appropriate
208 significance thresholds for the study population. SimpleM
209 uses PCA and linkage disequilibrium (LD) between SNPs to
210 help estimate the effective number of independent tests for
211 the dataset. This method generated a significance threshold
212 of 6.4×10^{-7} .

213
214 The X chromosome was also investigated for association.
215 9,852 X-linked SNPs were available after QC. XWAS[24]
216 was used to perform sex-stratified association testing. Two
217 tests were performed, one treating males as heterozygous
218 for SNPs on the X chromosome, and the other treating
219 males as homozygous for SNPs on the X chromosome, both
220 yielded to the same results. Age and the first five PCs were
221 included as covariates.

222

2.6 Linkage Analysis

Due to computational complexity, the 14-generation 8,222-person Amish pedigree (Figure 1A) was divided into smaller sub-pedigrees (e.g., Figure 1B) to perform both nonparametric (NPL) and parametric linkage analysis. A total of 295 sub-pedigrees were used in linkage analyses (Supplementary Table 2).

Model-based (parametric) and model-free (nonparametric or NPL) linkage analyses were performed using MERLIN.[25] To test for linkage, the null hypothesis of no linkage was compared to the likelihood of linkage ($0 \leq \alpha < 0.5$) between the phenotype and SNP within each sub-pedigree. A logarithm of the odds (LOD) score was calculated within each sub-pedigree and then summed across sub-pedigrees. In addition, the proportion of sub-pedigrees linked to a particular SNP (α) was calculated as part of a heterogeneity LOD score (HLOD).

For parametric two-point and multipoint linkage tests, both dominant and recessive models were tested on sub-pedigrees. These pedigrees each contained at least three but up to 11 CU individuals, with varying numbers of CI

246 individuals ranging from zero to two. For tests modeling a
247 dominant inheritance model, a trait allele frequency (AF) of
248 0.001 was used with penetrances of 0.0001, 1.0, and 1.0 for
249 zero, one, and two copies of the trait allele, respectively. For
250 the recessive model, the trait AF was 0.05 with penetrance
251 of 0.0001, 0.0001, and 1.0 for zero, one, and two copies of
252 the trait allele, respectively.

253
254 In nonparametric linkage (NPL) and parametric analyses,
255 both two-point and multipoint linkage was examined. Two-
256 point NPL and parametric analyses were performed on the
257 full set of post-QC SNPs ($n = 256,978$), while multipoint NPL
258 and parametric analyses were carried out on two different
259 sets of pruned SNPs to minimize the impact of linkage
260 disequilibrium (LD). An r^2 threshold of 0.16 was used to
261 define a set of 85,857 pruned SNPs for the multipoint
262 analyses.[26] MERLIN's NPL_{all} function was utilized for NPL
263 analyses to construct a linear model of linkage. LOD and
264 HLOD scores surpassing the defined thresholds for
265 suggestive (≥ 1.86) and significant linkage (≥ 3.3)[27] were
266 examined in more depth.

267

MINX software[25], an extension of MERLIN, was used for X chromosome linkage analysis. For two-point NPL and parametric analyses, a set of 9,852 X chromosome SNPs was used, and for multipoint NPL and parametric analyses, a pruned set of 153 SNPs with an $r^2 < 0.02$ and $MAF \geq 0.40$ was used. Multipoint NPL analysis utilized a sex-averaged genetic map[28], and model penetrances, adjusted for males and females, were the same as defined for the autosomal linkage analyses above.

For the subset of the loci where HLOD scores surpassed the significance threshold, 21- and 22-bit pedigrees driving the positive HLOD scores were examined for more distant connections allowing for the construction of larger pedigrees. 21- or 22-bit sub-pedigrees were merged when either identical AGBD IDs or first-degree relative AGBD IDs appeared (see Figure 1C for a larger pedigree example constructed for chromosome 2-linked sub-pedigrees). Six loci, one on each of chromosomes 1, 2, 3, 7, 11 and 17 met significance criteria, and larger pedigrees were possible for five of these loci, the exception being the chromosome 3 locus. MORGAN[29,30] was used to calculate region-specific LOD scores for these larger, more complex

291 pedigrees using both dominant and recessive models. Any
 292 linkage originally appearing under dominant inheritance
 293 patterns was verified by MCMC analysis also utilizing
 294 dominant parameters, recessive results were treated
 295 similarly.
 296
 297 MORGAN allowed for larger pedigrees and provided an
 298 alternative linkage method, Markov chain Monte Carlo
 299 (MCMC), to investigate the same loci. The trait allele
 300 frequency for CU was coded as very rare ($CU = 0.0001$) for
 301 all analyses, similar to, but slightly more restrictive than
 302 previous models run with MERLIN, in order to specify a
 303 single founder for each allele. Additionally, MCMC
 304 penetrance models utilized the same penetrances specified
 305 in MERLIN analyses. MCMC was performed with a
 306 maximum of 8 SNPs, spaced at 1-5 cM intervals. The
 307 genetic maps generated by Halldorsson et al., 2019 were
 308 used.[31] D' and r^2 values were generated specifically for our
 309 Amish population to verify that the resulting LOD scores
 310 were not influenced by LD.

311

312 3 RESULTS

313 3.1 Genome-Wide Association Analysis for 314 Cognitive Preservation

315 No SNPs reached the genome-wide significance threshold
316 defined by simpleM for this population, but eleven loci
317 (chromosomes 1, 3, 5, 6, 7, 11, 13, 14, 15, 16) surpassed
318 the suggestive threshold ($p \leq 10^{-4}$) (Supplementary Table 3
319 and Figure 2). Genomic inflation was 0.98 after the GRM,
320 age, and sex covariates were applied to the model
321 (Supplementary Figure 2A).

322

323 No SNPs on the X chromosome were significant or
324 suggestive under either sex-stratified model (Figure 2). The
325 genomic inflation value was 0.98 for the X chromosome
326 (Supplementary Figure 2B).

327

328 3.2 Linkage Analyses for Cognitive Preservation

329 Pedigrees were grouped via bit-size, chosen based on
330 computational tractability for our analyses. For NPL
331 analyses, families of bit size 21 and 22 were examined ($n =$
332 201). In the two-point NPL analyses, 17 loci on 13

333 chromosomes surpassed the threshold for suggestive
334 linkage (≥ 1.86) (Supplementary Figure 3A). Multipoint NPL
335 analyses resulted in 10 suggestive loci on 10 different
336 chromosomes (Supplementary Figure 3B). No SNPs
337 reached statistical significance in any NPL analysis.
338
339 Two-point dominant linkage analyses generated 72
340 significant (≥ 3.3) HLOD results spread across most
341 chromosomes for both the 21- and 22-bit pedigrees (Figure
342 3 and Supplementary Figure 4A). In multipoint analyses,
343 only chromosomes 12 and 17 had significant results for the
344 dominant model (Supplementary Figure 4B). Two-point
345 recessive linkage analyses generated significant results on
346 11 different chromosomes (Supplementary Figure 4C). No
347 loci reached significance with a multipoint recessive model
348 (Supplementary Figure 4D). Two-point parametric linkage
349 analysis on the X chromosome led to 14 loci surpassing the
350 threshold for suggestive linkage, but no loci were significant.
351 No SNPs were seen as suggestive or significant in multipoint
352 analyses on the X chromosome. See Supplemental Tables 3
353 and 4 for a full list of significant and suggestive loci,
354 respectively, from linkage analyses.
355

356 **3.3 Assessment of Significant and Suggestive**
357 **Genomic Regions**

358 The need to use sub-pedigrees for linkage analysis
359 introduced an unknown amount of variability into the linkage
360 analysis calculations.[32] To address this issue, we
361 implemented an approach to filter the large number of loci
362 surpassing thresholds for suggestive or significant results in
363 GWAS and/or linkage. Some differences were seen across
364 the various bit size pedigrees, therefore, significant HLOD
365 scores in multiple linkage analyses in the same region were
366 a requirement for the locus to undergo further investigation.

367 Regions selected for follow-up were required to have 1) two
368 significant HLOD scores (≥ 3.3) in any two linkage analyses
369 and a significant or suggestive LOD score (≥ 1.86) in at least
370 one other linkage analysis; or 2) one suggestive result in the
371 GWAS (within 5 Mb), a significant HLOD score in any
372 linkage analysis, and a suggestive HLOD score in any other
373 linkage analysis. After applying these criteria, 6 regions of
374 interest were identified for further investigation, located on
375 chromosomes 1, 2, 3, 7, 11, and 17 (Table 1). When probing
376 within a 10 Mb region around each of these loci, we
377 observed two known AD-related genes (*adsp.niagads.org*),
378 *PICALM* (chromosome 11) and *EPHA1* (chromosome 7).

379

380 By examining which sub-pedigrees displayed the highest

381 chromosome-specific LOD scores across multiple analyses,

382 it was noted that some of these sub-pedigrees could be

383 connected either by a specific individual or through a first

384 degree relative. These related sub-pedigrees were combined

385 into larger family pedigrees for five of the loci (chromosomes

386 1, 2, 7, 11, and 17), but were unable to be merged for the

387 chromosome 3 locus. These five larger pedigrees were then

388 used for the MCMC linkage analyses.

389

390 Only the chromosome 2 locus, centered around SNP

391 rs1402906, remained significant ($H_{LOD} = 4.87$) after

392 connection of the sub-pedigrees. MCMC analyses with

393 identical parameters were performed to further verify these

394 results, this time adding additional SNPs upstream and

395 downstream of rs1402906 to widen the genomic window

396 considering potential linkage. These additional analyses

397 each considered 5 SNPs, and again rs1402906 retained the

398 highest significance with a dominant model ($LOD=4.14$) (see

399 Table 2 and Figure 4). The rs1402906 T allele frequency is

400 12% in the Amish, compared to 19% in the ALFA European

401 population (dbSNP).[33] Of the 24 CU individuals who had

402 genotype data available, the genotype counts for rs1402906
403 were 7 TT, 12 CT, and 3 CC (Supplementary Table 6).
404 These CU individuals in the chromosome 2-linked pedigree
405 were all under 88 years of age. There were two CI
406 individuals in the chromosome 2-linked pedigree, both
407 displayed the CC genotype at rs1402906 and were >89
408 years. Despite the co-segregation of the T allele with most
409 CU individuals in this pedigree, the genotype frequencies are
410 essentially the same between CU and CI individuals in the
411 entire Amish population in the study (Supplementary Table
412 6).

413

414 3.4 Chromosome 2p *in Silico* Annotation

415 The UCSC genome browser was used to identify protein
416 coding genes on chromosome 2p11.2-13.1 spanning 10 cM
417 around the peak LOD score at rs1402906.[34] The protein
418 coding genes in this region include *REG1A*, *REG1B*,
419 *CTNNA2*, and *LRRTM1* and *LRRTM4* (Figure 5). In
420 addition, the Open Targets Platform eQTL and PheWAS
421 databases specified an eQTL associated with decreased
422 *CTNNA2* expression with the “T” allele at rs1402906, as well
423 as an increased incidence within a study population of
424 individuals with autism.[35,36] The PheWAS database also

425 indicated a miniscule, but still significant, diminishment within
426 a study population of Bipolar individuals.
427
428 Sequences around the SNP demonstrating the peak LOD
429 score, rs1402906, were investigated for potential promoter
430 regions in the Eukaryotic Promoter Database (EPD)[37], as
431 well as analyzed through PROMO, a transcription factor
432 binding site (TFBS) identification tool developed by
433 ALGGEN.[38,39] Cap Analysis of Gene Expression (CAGE)
434 reads from the FANTOM5 project were also examined to
435 investigate transcription start sites (TSS).[40]
436
437 Although the 2p11.2-13.1 genomic region is relatively devoid
438 of known genes, examination of transcription factor binding
439 sites in this region via ALGGEN shows the transcription
440 factor POU3F2 as having a perfect binding site consensus
441 beginning 4bp upstream of rs1402906 when the genomic
442 sequence contained the T allele at rs1402906.[41]
443 Substitution of the alternative allele C nearly abolished the
444 binding potential for POU3F2, with a low predicted binding
445 score due to two base-pair mismatches in the recognition
446 sequence. Another transcription factor, POU1F1a,
447 demonstrates a perfect binding consensus 10bp upstream of

rs1402906. However, the alleles at rs1402906 do not disrupt the binding. Utilizing the EPD as well as CAGE reads uncovered HK3me1, HK3me3, and H3K27ac sites within 10bp of rs1402906, indicating a potential TSS as well as predicted enhancer activity.[42,43]

4 DISCUSSION

Identifying protective genetic loci as a strategy for finding potential therapeutic targets is a methodology that yields higher success rates for clinical trials than other approaches.[6] For dementia, protective loci are associated with an outcome of cognitive preservation, which can result from either cognitive resilience or resistance. Cognitive resilience is observed when patients display typical dementia pathology, specifically a buildup of amyloid plaques and tangles in the brain in AD, without demonstrating any cognitive impairment or other dementia symptoms.[44] Cognitive resistance is defined by a lack of dementia-associated symptoms without presenting dementia pathologies.[44] As the Amish undergo only pre-mortem diagnosis, their classification falls under cognitive preservation, which is either cognitive resilient or resistant. Previous studies for AD have identified multiple genes that

471 possess alleles that appear to confer some protection
 472 against developing AD including, *APP*[45], *APOE* ϵ 2[46],
 473 *PLCG2*[47], *MS4A*[48], *ABCA1*[49], *RAB10*[50], *SORL1*[51],
 474 *PICALM*[52], *EPHA1*[4], and *CASS4*[4]. Although our *ad hoc*
 475 filtering suggested possible effects near *PICALM* and
 476 *EPHA1*, no SNPs in either gene demonstrated a protective
 477 effect.
 478
 479 GWAS and linkage analyses identified 106 suggestive or
 480 significant loci for cognitive preservation in the Amish
 481 (Supplemental Table 3). To limit false positives, we focused
 482 only on areas yielding suggestive or significant associations
 483 in multiple analyses. This limited our results to six areas of
 484 interest across chromosomes 1, 2, 3, 7, 11, and 17 (Table 1
 485 and Table 2). Notably, an effect of *APOE* ϵ 2 was not
 486 detected in any analysis, however this is likely because of
 487 *APOE* ϵ 3's high allele frequency within the Amish and its
 488 neutral phenotypic outcome, thereby preventing its
 489 enrichment in either the case or control groups. *APOE* ϵ 2
 490 was infrequent in this Amish population, with an allele
 491 frequency of just 4.2%, compared to 19.5% in non-Hispanic
 492 white populations,[53] making it likely too rare for analyses to
 493 detect. A region on chromosome 2p11.2-13.1, centered

494 around rs1402906, demonstrated statistical significance
495 consistently across multiple linkage analyses. By connecting
496 several of the sub-pedigrees, we confirmed the transmission,
497 providing additional evidence of linkage.

498

499 Because significant allelic association or genetic linkage
500 results do not necessarily identify causal variations, a 10 Mb
501 region around the chromosome 2 locus peak was evaluated
502 to develop a broader assessment of possible functional
503 elements. While this region was not accompanied by
504 association in the GWAS, PROMO predicted a POU3F2
505 TFBS with perfect consensus beginning 4bp upstream of
506 rs1402906 when the genomic sequence contained the T
507 allele. Though there remained a possibility of binding in the
508 presence of the C allele at that site, consensus was
509 markedly diminished as any potential binding would require
510 only partial adherence to the site, with two of the seven base
511 pairs in the binding locus being mismatched.[41] A majority
512 (87%) of the individuals demonstrating cognitive
513 preservation in the chromosome 2-linked pedigree (Figure
514 1C and Figure 4) have this T allele. Of the three CC-
515 genotyped CU individuals, two married into the pedigree
516 (i.e., were not otherwise closely related to the other pedigree

members), with the third CC genotyped individual (age 80) being below the average age of onset for CI in the Amish population (age 82) and a child of one of the married-ins. Additionally, the average age of onset for CI was 89.5 in this chromosome 2-linked pedigree used for MCMC analysis, implying that these individuals may still be too young to display CI onset or may have other protective factors located elsewhere in the genome as the average age of onset for this family is older than the general Amish population.

POU3F2 is a known regulator of the nearby gene *Leucine Rich Repeat Transmembrane neuronal 4 (LRRTM4)*.^[54] The leucine-rich repeat transmembrane proteins (LRRTMs) are primarily expressed in the nervous system where they assist in synaptic differentiation and development.^[55,56] Murine *in vivo* studies demonstrated that reduced *LRRTM4* expression led to deficits in excitatory synapse density in dentate gyrus granule cells.^[57] Increased expression of *LRRTM4* in the hippocampus improves memory in aging rats.^[58] Clinically, *LRRTM4* has been associated with Tourette syndrome and autism spectrum disorder, as well as increased risk of suicide attempt in bipolar disorder.^[59,60] It is possible that altered transcription in an upstream

540 regulatory region, possibly due to the potential for POU3F2
541 binding, may play a role in expression differences of
542 *LRRTM4*, which could further contribute to improved
543 cognitive ability in CU patients, as observed in rodent
544 models.
545
546 In eQTL and PheWAS databases searched, the “T” allele at
547 rs1402906 was associated with increased incidence of
548 autism spectrum disorder as well as lower *CTNNA2*
549 expression. Notably, the POU1F1a TF also had perfect
550 consensus binding in the region containing rs1402906 and
551 POU1F1a regulates the *CTNNA2* gene[61]. POU1F1a
552 overexpression induces increased expression of *Leucine*
553 *Rich Repeat Transmembrane neuronal 1 (LRRTM1)*,[61] a
554 gene 312 kb away from rs1402906 and within the region of
555 significant linkage on 2p11.2-13.1. *LRRTM1* is highly
556 expressed in the brain and is associated with handedness
557 and schizophrenia in humans.[62] Deletions of this gene lead
558 to hippocampal shrinking, loss of synaptic density, reduced
559 long-term potentiation in hippocampal neurons, and memory
560 deficiencies in mice.[63,64]
561

562 *LRRTM1* resides within the seventh intron of the *Catenin*
563 *Alpha 2 (CTNNA2)* gene in an anti-sense orientation, 262kb
564 away from rs1402906. *CTNNA2* shares promoters with
565 *LRRTM1*. [65] These shared promoters induce alternative
566 transcripts of *CTNNA2* containing additional 5' exons
567 expressed almost exclusively in the brain. Importantly, these
568 promoters are bi-directional and may regulate both *LRRTM1*
569 and the alternative splicing of *CTNNA2* as a single functional
570 unit. [66] Exon VIIb of *CTNNA2*, located on the sense strand
571 corresponding to the *LRRTM1* location, is included more
572 frequently in transcription than other alternative exons,
573 especially in the brain. The ratio of expression levels of full-
574 length *CTNNA2* and this isoform may be important for
575 regulating cadherin-dependent adhesion. [66] We note that a
576 previous publication from our group using a subset of our
577 Amish dataset found linkage to dementia in the region
578 surrounding *CTNNA2* and suggested it as a potential risk
579 locus for dementia. [32] Here, the "T" allele at rs1402906,
580 linked to cognitive preservation, was an eQTL noted to
581 decrease *CTNNA2* expression.

582

583 *CTNNA2* is highly expressed in the brain in both mice and
584 humans. [32,67] It is well categorized in mice, where it

585 regulates function during murine neuronal development,
586 specifically synapse formation and plasticity.[67]
587 Mechanistically, *CTNNA2* regulates synaptic spine
588 formation, stabilization, and turnover.[68] *CTNNA2* also
589 forms a complex with beta-catenin, which interacts with
590 presenilin, a known contributor to AD pathology.
591 Destabilization of this beta-catenin complex can lead to
592 neuronal apoptosis in mice and humans.[69] The location of
593 potential POU3F2 and POU1F1a binding sites at rs1402906,
594 near HK3me1, HK3me3, and H3K27ac sites indicate
595 predicted enhancer activity nearby a TSS. Thus, POU1F1a
596 binding may initiate transcription of either *LRRTM1*, an
597 alternative splicing variant of *CTNNA2*, or both, as
598 transcription of these genes has been shown to be
599 coregulated, with multiple TSSs for each gene existing within
600 30 bps of each other.[66] Similarly, POU3F2 binding may
601 initiate altered transcription of *LRRTM4*, another gene
602 involved in neurocognition.[58]
603
604 Our analyses highlight the genetic complexity of cognitive
605 preservation in the Amish, with different sub-pedigrees
606 appearing to drive disparate regions of the genome
607 associated with cognitive preservation. Numerous regions

608 yielded significant or suggestive results, but a region on
609 chromosome 2p11.2-13.1 showed the most robust linkage in
610 an area previously associated with dementia in the Amish,
611 suggesting the possibility of both protective and risk
612 regulatory elements for cognition in this region of the
613 genome. These results lay the foundation for additional
614 studies to identify loci conferring protection against cognitive
615 impairment.

616

617 5 REFERENCES

- 618 [1] 2021 Alzheimer's disease facts and figures.
619 Alzheimers Dement 2021;17:327–406.
620 <https://doi.org/10.1002/ALZ.12328>.
- 621 [2] Kunkle BW, Grenier-Boley B, Sims R, Bis JC,
622 Damotte V, Naj AC, et al. Genetic meta-analysis
623 of diagnosed Alzheimer's disease identifies new
624 risk loci and implicates A β , tau, immunity and
625 lipid processing. Nat Genet 2019;51:414–30.
626 <https://doi.org/10.1038/s41588-019-0358-2>.
- 627 [3] Williamson J, Goldman J, Marder KS. Genetic
628 Aspects of Alzheimer Disease. Neurologist
629 2009;15:80.
630 [https://doi.org/10.1097/NRL.0B013E318187E76](https://doi.org/10.1097/NRL.0B013E318187E76B)
631 B.
- 632 [4] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj
633 AC, Sims R, Bellenguez C, et al. Meta-analysis
634 of 74,046 individuals identifies 11 new
635 susceptibility loci for Alzheimer's disease. Nature
636 Genetics 2013 45:12 2013;45:1452–8.
637 <https://doi.org/10.1038/ng.2802>.

- 638 [5] Asher S, Priefer R. Alzheimer's disease failed
639 clinical trials. Life Sci 2022;306:120861.
640 <https://doi.org/10.1016/J.LFS.2022.120861>.
- 641 [6] Harper AR, Nayee S, Topol EJ. Protective
642 alleles and modifier variants in human health
643 and disease. Nat Rev Genet 2015;16:689–701.
644 <https://doi.org/10.1038/NRG4017>.
- 645 [7] Strauss KA, Puffenberger EG. Genetics,
646 medicine, and the Plain people. Annu Rev
647 Genomics Hum Genet 2009;10:513–36.
648 [https://doi.org/10.1146/ANNUREV-GENOM-](https://doi.org/10.1146/ANNUREV-GENOM-082908-150040)
649 [082908-150040](https://doi.org/10.1146/ANNUREV-GENOM-082908-150040).
- 650 [8] Crowley WK. OLD ORDER AMISH
651 SETTLEMENT: DIFFUSION AND GROWTH.
652 Annals of the Association of American
653 Geographers 1978;68:249–64.
654 [https://doi.org/10.1111/J.1467-](https://doi.org/10.1111/J.1467-8306.1978.TB01194.X)
655 [8306.1978.TB01194.X](https://doi.org/10.1111/J.1467-8306.1978.TB01194.X).
- 656 [9] Welsh KA, Butters N, Mohs RC, Beekly D,
657 Edland S, Fillenbaum G, et al. The Consortium
658 to Establish a Registry for Alzheimer's Disease
659 (CERAD). Part V. A normative study of the
660 neuropsychological battery. Neurology
661 1994;44:609–609.
662 <https://doi.org/10.1212/WNL.44.4.609>.
- 663 [10] Teng E, Chui H. The Modified Mini-Mental State
664 (3MS) examination. J Clin Psychiatry
665 1987;48:314–8.
- 666 [11] Galvin JE, Roe CM, Morris JC. Evaluation of
667 Cognitive Impairment in Older Adults: Combining
668 Brief Informant and Performance Measures.
669 Arch Neurol 2007;64:718–24.
670 <https://doi.org/10.1001/ARCHNEUR.64.5.718>.
- 671 [12] Arnett JA, Labovitz SS. Effect of Physical Layout
672 in Performance of the Trail Making Test. vol. 7.
673 1995.
- 674 [13] Kuzma A, Valladares O, Cweibel R, Greenfest-
675 Allen E, Childress DM, Malamon J, et al.
676 NIAGADS: The NIA Genetics of Alzheimer's

- 677 Disease Data Storage Site. Alzheimer's &
678 Dementia 2016;12:1200–3.
679 <https://doi.org/10.1016/J.JALZ.2016.08.018>.
- 680 [14] Beecham GW, Bis JC, Martin ER, Choi SH,
681 DeStefano AL, Van Duijn CM, et al. The
682 Alzheimer's Disease Sequencing Project: Study
683 design and sample selection. *Neurol Genet*
684 2017;3.
685 <https://doi.org/10.1212/NXG.0000000000000194>
686 .
- 687 [15] Osterman MD, Song YE, Adams LD, Laux RA,
688 Caywood LJ, Prough MB, et al. The genetic
689 architecture of Alzheimer disease risk in the
690 Ohio and Indiana Amish. *Human Genetics and*
691 *Genomics Advances* 2022;3:100114.
692 <https://doi.org/10.1016/J.XHGG.2022.100114>.
- 693 [16] Agarwala R, Biesecker LG, Schäffer AA.
694 Anabaptist genealogy database. *American*
695 *Journal of Medical Genetics - Seminars in*
696 *Medical Genetics* 2003;121 C:32–7.
697 <https://doi.org/10.1002/ajmg.c.20004>.
- 698 [17] Garbe JR, Da Y. A software tool for the
699 graphical visualization of large and complex
700 populations. *Yi Chuan Xue Bao* 2003;30:1193–
701 5.
- 702 [18] Liu F, Kirichenko A, Axenovich TI, Van Duijn
703 CM, Aulchenko YS. An approach for cutting
704 large and complex pedigrees for linkage
705 analysis. *European Journal of Human Genetics*
706 2008;16:854–60.
707 <https://doi.org/10.1038/ejhg.2008.24>.
- 708 [19] Gogarten SM, Sofer T, Chen H, Yu C, Brody JA,
709 Thornton TA, et al. Genetic association testing
710 using the GENESIS R/Bioconductor package
711 n.d.
712 <https://doi.org/10.1093/bioinformatics/btz567>.
- 713 [20] Conomos MP, Reiner AP, Weir BS, Thornton
714 TA. Model-free Estimation of Recent Genetic

- 715 Relatedness. *Am J Hum Genet* 2016;98:127.
716 <https://doi.org/10.1016/J.AJHG.2015.11.022>.
- 717 [21] Gao X, Becker LC, Becker DM, Starmer JD,
718 Province MA. Brief Report Avoiding the High
719 Bonferroni Penalty in Genome-Wide Association
720 Studies. *Genet Epidemiol* 2010;34:100–5.
721 <https://doi.org/10.1002/gepi.20430>.
- 722 [22] Gao X, Starmer J, Martin ER. A Multiple Testing
723 Correction Method for Genetic Association
724 Studies Using Correlated Single Nucleotide
725 Polymorphisms. *Genet Epidemiol* 2008;32:361–
726 9. <https://doi.org/10.1002/gepi.20310>.
- 727 [23] Gao X. Multiple Testing Corrections for Imputed
728 SNPs. *Genet Epidemiol* 2011;35:154–8.
729 <https://doi.org/10.1002/gepi.20563>.
- 730 [24] Gao F, Chang D, Biddanda A, Ma L, Guo Y,
731 Zhou Z, et al. XWAS: A Software Toolset for
732 Genetic Data Analysis and Association Studies
733 of the X Chromosome. *Journal of Heredity*
734 2015;666–71.
735 <https://doi.org/10.1093/jhered/esv059>.
- 736 [25] Abecasis GR, Cherny SS, Cookson WO, Cardon
737 LR. Merlin-rapid analysis of dense genetic maps
738 using sparse gene flow trees 2002.
739 <https://doi.org/10.1038/ng786>.
- 740 [26] Boyles AL, Scott WK, Martin ER, Schmidt S, Li
741 YJ, Ashley-Koch A, et al. Linkage disequilibrium
742 inflates type I error rates in multipoint linkage
743 analysis when parental genotypes are missing.
744 *Hum Hered* 2005;59:220–7.
745 <https://doi.org/10.1159/000087122>.
- 746 [27] Ott J, Wang J, Leal SM. Genetic linkage analysis
747 in the age of whole-genome sequencing. *Nat*
748 *Rev Genet* 2015;16:275.
749 <https://doi.org/10.1038/NRG3908>.
- 750 [28] MERLIN Quick Reference n.d.
751 [https://csg.sph.umich.edu/abecasis/Merlin/refere](https://csg.sph.umich.edu/abecasis/Merlin/reference.html)
752 [nce.html](https://csg.sph.umich.edu/abecasis/Merlin/reference.html) (accessed November 1, 2023).

- 753 [29] Wijsman EM, Rothstein JH, Thompson EA.
754 Multipoint Linkage Analysis with Many
755 Multiallelic or Dense Diallelic Markers: Markov
756 Chain–Monte Carlo Provides Practical
757 Approaches for Genome Scans on General
758 Pedigrees. *Am J Hum Genet* 2006;79:846.
759 <https://doi.org/10.1086/508472>.
- 760 [30] Tong L, Thompson E. Multilocus lod scores in
761 large pedigrees: combination of exact and
762 approximate calculations. *Hum Hered*
763 2008;65:142–53.
764 <https://doi.org/10.1159/000109731>.
- 765 [31] Halldorsson B V., Palsson G, Stefansson OA,
766 Jonsson H, Hardarson MT, Eggertsson HP, et
767 al. Characterizing mutagenic effects of
768 recombination through a sequence-level genetic
769 map. *Science* 2019;363.
770 <https://doi.org/10.1126/SCIENCE.AAU1043>.
- 771 [32] Cummings AC, Jiang L, Velez Edwards DR,
772 Mccauley JL, Laux R, Mcfarland LL, et al.
773 Genome-wide association and linkage study in
774 the Amish detects a novel candidate late-onset
775 Alzheimer disease gene. *Ann Hum Genet*
776 2012;76:342–51. <https://doi.org/10.1111/J.1469-1809.2012.00721.X>.
- 778 [33] Phan L, Jin Y, Zhang H, Qiang W, Shekhtman E,
779 Shao D, et al. ALFA: Allele Frequency
780 Aggregator. National Center for Biotechnology
781 Information, US National Library of Medicine
782 2020. www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/
783 (accessed July 3, 2023).
- 784 [34] Rao SSP, Huntley MH, Durand NC, Stamenova
785 EK, Bochkov ID, Robinson JT, et al. A 3D map
786 of the human genome at kilobase resolution
787 reveals principles of chromatin looping. *Cell*
788 2014;159:1665–80.
789 <https://doi.org/10.1016/j.cell.2014.11.021>.
- 790 [35] Mountjoy E, Schmidt EM, Carmona M,
791 Schwartzentruber J, Peat G, Miranda A, et al. An
792 open approach to systematically prioritize causal

- 793 variants and genes at all published human
794 GWAS trait-associated loci. *Nat Genet*
795 2021;53:1527–33.
796 <https://doi.org/10.1038/S41588-021-00945-5>.
- 797 [36] Ghoussaini M, Mountjoy E, Carmona M, Peat G,
798 Schmidt EM, Hercules A, et al. Open Targets
799 Genetics: systematic identification of trait-
800 associated genes using large-scale genetics and
801 functional genomics. *Nucleic Acids Res*
802 2021;49:D1311–20.
803 <https://doi.org/10.1093/NAR/GKAA840>.
- 804 [37] Dreos R, Ambrosini G, Groux R, Perier RC,
805 Bucher P. The eukaryotic promoter database in
806 its 30th year: focus on non-vertebrate
807 organisms. *Nucleic Acids Res* 2017;45:D51–5.
808 <https://doi.org/10.1093/NAR/GKW1069>.
- 809 [38] Messeguer X, Escudero R, Farré D, Núñez O,
810 Martínez J, Albà MM. PROMO: detection of
811 known transcription regulatory elements using
812 species-tailored searches. *Bioinformatics*
813 2002;18:333–4.
814 [https://doi.org/10.1093/BIOINFORMATICS/18.2.](https://doi.org/10.1093/BIOINFORMATICS/18.2.333)
815 333.
- 816 [39] Farré D, Roset R, Huerta M, Adsuar JE,
817 Roselló L, Albà MM, et al. Identification of
818 patterns in biological sequences at the ALGGEN
819 server: PROMO and MALGEN. *Nucleic Acids*
820 *Res* 2003;31:3651.
821 <https://doi.org/10.1093/NAR/GKG605>.
- 822 [40] Noguchi S, Arakawa T, Fukuda S, Furuno M,
823 Hasegawa A, Hori F, et al. FANTOM5 CAGE
824 profiles of human and mouse samples. *Scientific*
825 *Data* 2017 4:1 2017;4:1–10.
826 <https://doi.org/10.1038/sdata.2017.112>.
- 827 [41] Ison J, Rapacki K, Ménager H, Kalaš M, Rydza
828 E, Chmura P, et al. Tools and data services
829 registry: a community effort to document
830 bioinformatics resources. *Nucleic Acids Res*
831 2016;44:D38–47.
832 <https://doi.org/10.1093/NAR/GKV1116>.

- 833 [42] Barski A, Cuddapah S, Cui K, Roh TY, Schones
834 DE, Wang Z, et al. High-resolution profiling of
835 histone methylations in the human genome. *Cell*
836 2007;129:823–37.
837 <https://doi.org/10.1016/J.CELL.2007.05.009>.
- 838 [43] Sayers EW, Bolton EE, Brister JR, Canese K,
839 Chan J, Comeau DC, et al. Database resources
840 of the national center for biotechnology
841 information. *Nucleic Acids Res* 2022;50:D20–6.
842 <https://doi.org/10.1093/NAR/GKAB1112>.
- 843 [44] Arenaza-Urquijo EM, Vemuri P. Resistance vs
844 resilience to Alzheimer disease: Clarifying
845 terminology for preclinical studies. *Neurology*
846 2018;90:695.
847 <https://doi.org/10.1212/WNL.0000000000005303>
848 .
- 849 [45] Kero M, Paetau A, Polvikoski T, Tanskanen M,
850 Sulkava R, Jansson L, et al. Amyloid precursor
851 protein (APP) A673T mutation in the elderly
852 Finnish population. *Neurobiol Aging*
853 2013;34:1518.e1-1518.e3.
854 [https://doi.org/10.1016/J.NEUROBIOLAGING.20](https://doi.org/10.1016/J.NEUROBIOLAGING.2012.09.017)
855 [12.09.017](https://doi.org/10.1016/J.NEUROBIOLAGING.2012.09.017).
- 856 [46] Locke PA, Conneally PM, Tanzi RE, Gusella JF,
857 Haines JL. Apolipoprotein E4 allele and
858 Alzheimer disease: examination of allelic
859 association and effect on age at onset in both
860 early- and late-onset cases. *Genet Epidemiol*
861 1995;12:83–92.
862 <https://doi.org/10.1002/GEPI.1370120108>.
- 863 [47] Sims R, Van Der Lee SJ, Naj AC, Bellenguez C,
864 Badarinarayan N, Jakobsdottir J, et al. Rare
865 coding variants in PLCG2, ABI3, and TREM2
866 implicate microglial-mediated innate immunity in
867 Alzheimer's disease. *Nature Genetics* 2017 49:9
868 2017;49:1373–84.
869 <https://doi.org/10.1038/ng.3916>.
- 870 [48] Jiao B, Liu X, Zhou L, Wang MH, Zhou Y, Xiao
871 T, et al. Polygenic Analysis of Late-Onset
872 Alzheimer's Disease from Mainland China. *PLoS*

- 873 One 2015;10:e0144898.
874 <https://doi.org/10.1371/JOURNAL.PONE.0144898>.
875 8.
- 876 [49] Wang F, Jia J. Polymorphisms of cholesterol
877 metabolism genes CYP46 and ABCA1 and the
878 risk of sporadic Alzheimer's disease in Chinese.
879 Brain Res 2007;1147:34–8.
880 <https://doi.org/10.1016/J.BRAINRES.2007.02.005>.
881 5.
- 882 [50] Ridge PG, Karch CM, Hsu S, Arano I, Teerlink
883 CC, Ebbert MTW, et al. Linkage, whole genome
884 sequence, and biological data implicate variants
885 in RAB10 in Alzheimer's disease resilience.
886 Genome Med 2017;9:1–14.
887 [https://doi.org/10.1186/S13073-017-0486-](https://doi.org/10.1186/S13073-017-0486-1)
888 1/TABLES/4.
- 889 [51] Chou CT, Liao YC, Lee WJ, Wang SJ, Fuh JL.
890 SORL1 gene, plasma biomarkers, and the risk of
891 Alzheimer's disease for the Han Chinese
892 population in Taiwan. Alzheimers Res Ther
893 2016;8:1–9. [https://doi.org/10.1186/S13195-016-](https://doi.org/10.1186/S13195-016-0222-X)
894 0222-X/TABLES/4.
- 895 [52] Santos-Rebouças CB, Gonçalves AP, dos
896 Santos JM, Abdala BB, Motta LB, Laks J, et al.
897 rs3851179 Polymorphism at 5' to the PICALM
898 Gene is Associated with Alzheimer and
899 Parkinson Diseases in Brazilian Population.
900 Neuromolecular Med 2017;19:293–9.
901 [https://doi.org/10.1007/S12017-017-8444-](https://doi.org/10.1007/S12017-017-8444-Z)
902 Z/TABLES/2.
- 903 [53] Beydoun MA, Weiss J, Beydoun HA, Hossain S,
904 Maldonado AI, Shen B, et al. Race, APOE
905 genotypes, and cognitive decline among middle-
906 aged urban adults. Alzheimers Res Ther
907 2021;13:1–16. [https://doi.org/10.1186/S13195-](https://doi.org/10.1186/S13195-021-00855-Y)
908 021-00855-Y/FIGURES/3.
- 909 [54] Rouillard AD, Gundersen GW, Fernandez NF,
910 Wang Z, Monteiro CD, McDermott MG, et al.
911 The harmonizome: a collection of processed
912 datasets gathered to serve and mine

913 knowledge about genes and proteins. Database
914 2016;2016.
915 <https://doi.org/10.1093/DATABASE/BAW100>.

916 [55] Laurén J, Airaksinen MS, Saarma M, Timmusk
917 T. A novel gene family encoding leucine-rich
918 repeat transmembrane proteins differentially
919 expressed in the nervous system. Genomics
920 2003;81:411–21. [https://doi.org/10.1016/S0888-](https://doi.org/10.1016/S0888-7543(03)00030-2)
921 [7543\(03\)00030-2](https://doi.org/10.1016/S0888-7543(03)00030-2).

922 [56] Linhoff MW, Laurén J, Cassidy RM, Dobie FA,
923 Takahashi H, Nygaard HB, et al. An unbiased
924 expression screen for synaptogenic proteins
925 identifies the LRRTM protein family as synaptic
926 organizers. Neuron 2009;61:734–49.
927 <https://doi.org/10.1016/J.NEURON.2009.01.017>.

928 [57] Siddiqui TJ, Tari PK, Connor SA, Zhang P,
929 Dobie FA, She K, et al. An LRRTM4-HSPG
930 complex mediates excitatory synapse
931 development on dentate gyrus granule cells.
932 Neuron 2013;79:680–95.
933 <https://doi.org/10.1016/j.neuron.2013.06.029>.

934 [58] Flores-Páez LA, Pacheco-Rosado J, Alva-
935 Sánchez C, Zamudio SR. Cerebrolysin
936 enhances the expression of the synaptogenic
937 protein LRRTM4 in the hippocampus and
938 improves learning and memory in senescent
939 rats. Behavioural Pharmacology 2020;31:491–9.
940 <https://doi.org/10.1097/FBP.0000000000000530>.

941 [59] Clarke RA, Eapen V. LRRTM4 Terminal Exon
942 Duplicated in Family with Tourette Syndrome,
943 Autism and ADHD. Genes (Basel) 2022;13.
944 <https://doi.org/10.3390/GENES13010066>.

945 [60] Willour VL, Seifuddin F, Mahon PB, Jancic D,
946 Pirooznia M, Steele J, et al. A genome-wide
947 association study of attempted suicide. Mol
948 Psychiatry 2012;17:433.
949 <https://doi.org/10.1038/MP.2011.4>.

950 [61] Herman JP, Jullien N, Guillen S, Enjalbert A,
951 Pellegrini I, Franc JL. Research Resource: A

- 952 Genome-Wide Study Identifies Potential New
953 Target Genes for POU1F1. *Molecular*
954 *Endocrinology* 2012;26:1455–63.
955 <https://doi.org/10.1210/ME.2011-1308>.
- 956 [62] Francks C, Maegawa S, Laurén J, Abrahams
957 BS, Velayos-Baeza A, Medland SE, et al.
958 LRRTM1 on chromosome 2p12 is a maternally
959 suppressed gene that is associated paternally
960 with handedness and schizophrenia. *Mol*
961 *Psychiatry* 2007;12:1129.
962 <https://doi.org/10.1038/SJ.MP.4002053>.
- 963 [63] Bhouri M, Morishita W, Temkin P, Goswami D,
964 Kawabe H, Brose N, et al. Deletion of LRRTM1
965 and LRRTM2 in adult mice impairs basal AMPA
966 receptor transmission and LTP in hippocampal
967 CA1 pyramidal neurons. *Proc Natl Acad Sci U S*
968 *A* 2018;115:E5382–9.
969 <https://doi.org/10.1073/PNAS.1803280115>.
- 970 [64] Dhume SH, Connor SA, Mills F, Tari PK, Au-
971 Yeung SHM, Karimi B, et al. Distinct but
972 overlapping roles of LRRTM1 and LRRTM2 in
973 developing and mature hippocampal circuits.
974 *Elife* 2022;11.
975 <https://doi.org/10.7554/ELIFE.64742>.
- 976 [65] Kask M, Pruunsild P, Timmusk T. Bidirectional
977 transcription from human LRRTM2/CTNNA1 and
978 LRRTM1/CTNNA2 gene loci leads to expression
979 of N-terminally truncated CTNNA1 and CTNNA2
980 isoforms. *Biochem Biophys Res Commun*
981 2011;411:56–61.
982 <https://doi.org/10.1016/J.BBRC.2011.06.085>.
- 983 [66] Trinklein ND, Force Aldred S, Hartman SJ,
984 Schroeder DI, Otilar RP, Myers RM. An
985 abundance of bidirectional promoters in the
986 human genome. *Genome Res* 2004;14:62–6.
987 <https://doi.org/10.1101/GR.1982804>.
- 988 [67] Park C, Falls W, Finger JH, Longo-Guess CM,
989 Ackerman SL. Deletion in *Catna2*, encoding
990 alpha N-catenin, causes cerebellar and
991 hippocampal lamination defects and impaired

- 992 startle modulation. Nat Genet 2002;31:279–84.
 993 <https://doi.org/10.1038/NG908>.
- 994 [68] Abe K, Chisaka O, Van Roy F, Takeichi M.
 995 Stability of dendritic spines and synaptic
 996 contacts is controlled by α N-catenin. Nature
 997 Neuroscience 2004 7:4 2004;7:357–63.
 998 <https://doi.org/10.1038/nn1212>.
- 999 [69] Zhang Z, Hartmann H, Do VM, Abramowski D,
 1000 Sturchler-Pierrat C, Staufienbiel M, et al.
 1001 Destabilization of beta-catenin by mutations in
 1002 presenilin-1 potentiates neuronal apoptosis.
 1003 Nature 1998;395:698–702.
 1004 <https://doi.org/10.1038/27208>.

1005

1006 6 AUTHOR CONTRIBUTIONS

1007 Jonathan L Haines, William K Scott, and Margaret A
 1008 Pericak-Vance helped design the study and obtain
 1009 funding for this research. Ascertainment of data was
 1010 collected and managed by Renee A Laux, Kristy L
 1011 Miskimen, M Denise Fuzzell, Sarada L Fuzzell, Sherri
 1012 D Hochstetler, Laura J Caywood, Jason E Clouse, and
 1013 Sharlene D Herington. Cognitive status was
 1014 determined by Michael L Cuccaro, Paula K Ogrocki,
 1015 Alan J Lerner, and Jeffery M Vance. Yeunjoo E Song,
 1016 Leighanne R Main, Michael D Osterman, Audrey Lynn,
 1017 Daniel A Dorfsman, Larry D Adams, and Michael B
 1018 Prough provided data management, organization, and
 1019 QC. Analyses were performed by Leighanne R Main,

1020 with assistance from Yeunjoo Song and Michael D.
1021 Osterman. Manuscript writing and drafting was
1022 performed by Jonathan L Haines, Audrey Lynn, and
1023 Leighanne R Main. Editing completed by all authors.

1024

1025 7 ACKNOWLEDGMENTS

1026 This research was supported under the National
1027 institutes of Health and National Institute on Aging,
1028 grant AG058066. This research was also supported by
1029 an Alzheimer's Disease Translational Data Science
1030 Training Grant, 5T32AG071474-02.

1031

1032 8 CONFLICT OF INTEREST STATEMENT

1033 The authors declare no conflicts of interest.

1034

1035 9 CONSENT STATEMENT

1036 All individuals ascertained in this study supplied written
1037 informed consent under protocols approved by the
1038 IRBs at Case Western Reserve University and the
1039 University of Miami.

1040

1041 10 Diversity, Equity, and Inclusion Statement

1042 Our study considers an isolated founder population,
1043 not only genetically, but also culturally, as their
1044 lifestyles deviate from modern technocratic habits,
1045 making the Amish an under-represented population in
1046 studies. Additionally, informed consent and community
1047 engagement informed our sampling practices and
1048 subsequent analyses. We also made efforts to ensure
1049 our study cohort consisted of comparable proportions
1050 across each sex.
1051
1052 Keywords: Alzheimer disease, cognitive resilience,
1053 cognitive resistance, GWAS, linkage, isolated founder
1054 population, pedological study

FIGURE 1 (Leighanne Main)

1055 **Figure 1 – Amish pedigrees.** (A) 8,222 individuals
 1056 spanning 14 generations are present in this large pedigree.
 1057 Relationship data was drawn from the AGDB to incorporate
 1058 ascertained individuals into existing family trees. Colored
 1059 lines are used to connect generations between families. (B)
 1060 Example of a 21-bit sub-pedigree generated via PedCut from
 1061 the 14-generation pedigree to create computationally
 1062 tractable units for analyses. An individual's sex was included
 1063 during analyses but was removed here for privacy
 1064 protection. Black shading indicates individuals who were CU
 1065 at their last assessment; clear shading indicates individuals
 1066 with CI. Grey shading indicates individuals for whom no
 1067 information besides relationship data was available. (C)
 1068 Larger pedigree connecting chromosome 2-linked smaller
 1069 pedigrees. To protect privacy, sex information was included
 1070 for analyses but was removed here. Shading indicates the
 1071 same phenotypes as in 1B.

FIGURE 2 (Leighanne Main)

1072 **Figure 2 – Genome-wide association study (GWAS) for**
 1073 **cognitive preservation.** The suggestive association
 1074 threshold ($p=1 \times 10^{-4}$) is depicted with a dashed line; the
 1075 significant association threshold ($p=6.4 \times 10^{-7}$) was defined
 1076 using the simple-M method and is depicted with a solid line.
 1077 SNPs are represented as dots corresponding to genomic
 1078 coordinates within their chromosome (horizontal axis). SNP
 1079 density within each chromosome is shown as a heat map
 1080 above each chromosome number. Chromosome 23 is the X
 1081 chromosome. Eleven loci demonstrated suggestive
 1082 association, located on chromosomes 1, 3, 5, 6, 7, 11, 13,
 1083 14, 15, and 16 (and depicted as pink dots on this figure). No
 1084 SNP surpassed the significance threshold for association.

FIGURE 3 (Leighanne Main)

1085 **Figure 3 – Dominant two-point parametric linkage**
 1086 **analysis for cognitive preservation.** Dominant linkage
 1087 analysis on 21-bit pedigrees across all autosomes and the X
 1088 chromosome, labeled as chromosome 23. Suggestive (LOD
 1089 ≥ 1.86) and significant (LOD ≥ 3.3) thresholds are denoted
 1090 as the lower and upper dashed horizontal lines, respectively.
 1091 A total of 72 significant results were seen across most
 1092 chromosomes (excluding chromosomes 14, 16, 22, and 23).
 1093 See Supplemental Figure 3 and Supplemental Tables 4 and
 1094 4 for the complete list of all loci surpassing significant and
 1095 suggestive thresholds.

FIGURE 4 (Leighanne Main)

1096 **Figure 4 - Chromosome 2p11.2-13.1 linkage analysis**
1097 **results.** Two-point dominant (green) and recessive (purple)
1098 linkage results, two-point NPL (blue) and multipoint MCMC
1099 (orange) (significance >3.3, upper dashed line; suggestive
1100 >1.86, lower dashed line).

FIGURE 5 (Leighanne Main)


1101 **Figure 5 – UCSC Genome browser for chromosome**
1102 **2p11.2-13.1.** UCSC Genome Browser tracks of protein
1103 coding genes in the significant linkage region on
1104 chromosome 2 p11.2-13.1. The observed chromosomal
1105 region is indicated by a red box on the genomic coordinate
1106 slider. rs1402906 is located at 79Mb, highlighted by a
1107 vertical red line across the viewer, with 5Mb on either side
1108 shown. This region is located upstream of the genes
1109 *CTNNA2* and *LRRTM1*.

TABLE 1 (Leighanne Main)

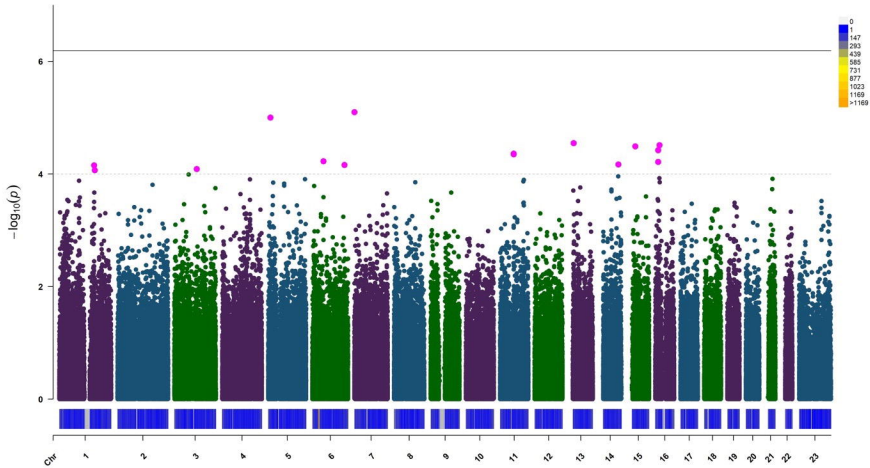
1110 **Table1 – Loci meeting *ad-hoc* filtering conditions for**
 1111 **further investigation.** Chromosome number and
 1112 coordinates are listed for each of the top LOD scores. P-
 1113 value is listed for suggestive GWAS result near the same
 1114 location. For 21- and 22-bit pedigree analyses, significant (\geq
 1115 3.3) and suggestive (≥ 1.86) HLOD scores are listed, with
 1116 significant results in bold. Empty cells indicate that no loci
 1117 surpassed the suggestive or significant threshold for that
 1118 analysis.

TABLE 2 (Leighanne Main)

1119 **Table 2 – MCMC linkage analysis for cognitive**
1120 **preservation.** Chromosome, genomic coordinates, and LOD
1121 score are depicted in the table. LOD scores in bold surpass

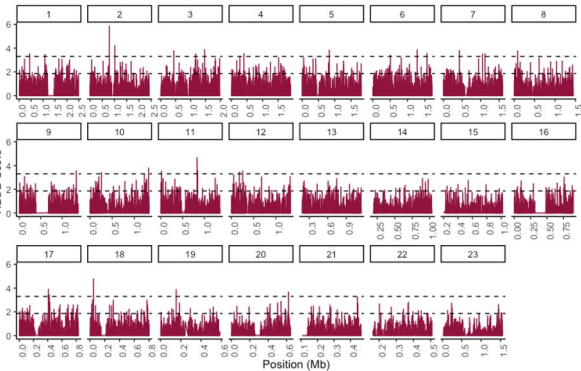


The pedigree chart illustrates a family with four generations. Affected individuals are represented by black diamonds, and unaffected individuals by grey diamonds. The pedigree shows an autosomal recessive inheritance pattern. In the first generation, two unaffected individuals have four children in the second generation. Two of these children are affected, and two are unaffected. The two affected children in the second generation have children in the third generation. One affected child has two children, one affected and one unaffected. The other affected child has one child, who is affected. The two unaffected children in the second generation have children in the third generation. One unaffected child has two children, one affected and one unaffected. The other unaffected child has two children, one affected and one unaffected. The two affected children in the third generation have children in the fourth generation. One affected child has one child, who is affected. The other affected child has one child, who is affected. The two unaffected children in the third generation have children in the fourth generation. One unaffected child has one child, who is affected. The other unaffected child has one child, who is affected.

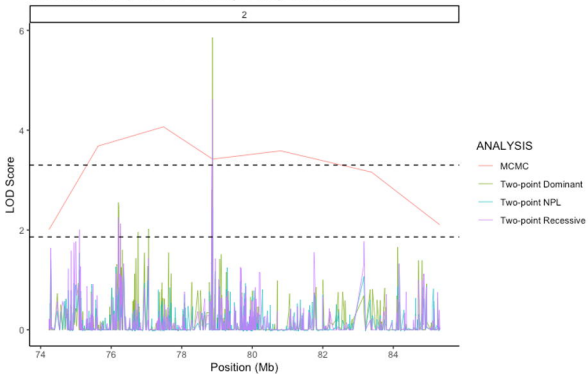


Dominant Parametric Linkage Analysis for Cognitive Preservation

HLOD Score



Chromosome 2 p11.2-13.1 Linkage Analysis Results

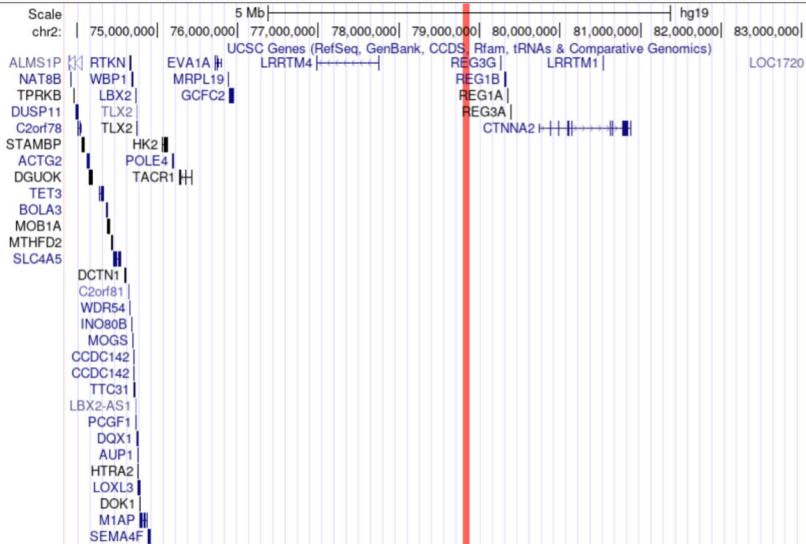


chr2 (p13.1-p11.2)

21

p12

3435



Chr	Location	GWAS	21-bit pedigrees				22-bit pedigrees			
		p-value	Two-point NPL	Two-point Dominant	Two-point Recessive	Multipoint Dominant	Two-point NPL	Two-point Dominant	Two-point Recessive	Multipoint Dominant
1	165,478,920	7.00E-05	1.9		3.4				3.2	
2	78,873,098		2.4	5.9	4.6		1.9	4.9	3	
3	108,251,237	8.20E-05		3.3			2	3.6		
7	139,996,449		2.1	2.4	4.1				3.8	
11	83,081,483		2	4.7	2.9				3.3	
17	4,487,906			2.3		2.3		4.1		4.1

Chromosome	Location (bp) GRCh38	Top LOD Score
1	159,476,580 – 167,672,055	1.97
2	72,972,490 – 80,875,139	4.14
7	138,047,463 – 141,972,804	0.6
11	81,074,649 – 85,224,030	2.64
17	3,492,832 – 5,478,547	0.13