Long runs of homozygosity are associated with Alzheimer's disease

Sonia Moreno-Grau
Maria Victoria Fernández
Fabiana Farias
John P. Budde
Carlos Cruchaga

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs
Authors
Sonia Moreno-Grau, Maria Victoria Fernández, Fabiana Farias, John P. Budde, Carlos Cruchaga, O. Maroñas, and et al.
Long runs of homozygosity are associated with Alzheimer’s disease

Sonia Moreno-Grau1,2, Maria Victoria Fernández2,3,4, Itziar de Rojas3,1,2, Pablo García-González3,1, Isabel Hernández1, Fabiana Farias3,4, John P. Budde3,4, Inés Quintela3, Laura Madrid3, Antonio González-Pérez6, Laura Montreuil1, Emilio Alarcón-Martín1, Montserrat Alegret1, Olalla Maroñas5, Juan Antonio Pineda7, Juan Macías8, The GR@ACE study groupDEGESCO consortium, Marta Marquié1,2, Sergi Valero1,2, Alba Benaque1, Jordi Clarimón2,8, Maria Jesus Bullido1,2,9,10, Guillermo García-Ribas11, Pau Pástor12, Pascual Sánchez-Juan2,13, Victoria Álvarez14,15, Gerard Piñol-Ripoll2,16, Jose María García-Alberca17, José Luis Royo18, Emilio Franco-Macias19, Pablo Mín1,20, Miguel Calero22,21, Miguel Medina3,21, Alberto Rábano22,21, Jesús Ávila2,24, Carmen Antúnez25, Luis Miguel Real26,27, Adelina Orellana1, Ángel Carracedo28,26, María Eugenia Sáez26, Lluís Tarraga12, Mercè Boada1,2, Carlos Cruchaga3,4 and Agustín Ruiz1,2, for the Alzheimer’s Disease Neuroimaging Initiative

Abstract
Long runs of homozygosity (ROH) are contiguous stretches of homozygous genotypes, which are a footprint of inbreeding and recessive inheritance. The presence of recessive loci is suggested for Alzheimer’s disease (AD); however, their search has been poorly assessed to date. To investigate homozygosity in AD, here we performed a fine-scale ROH analysis using 10 independent cohorts of European ancestry (11,919 AD cases and 9181 controls.) We detected an increase of homozygosity in AD cases compared to controls \( \beta_{AVROH} \) (CI 95%) = 0.070 (0.037–0.104); \( \beta_{FRoH} \) (CI95%) = 0.043 (0.009–0.076); \( P = 3.91 \times 10^{-5} \). ROHs increasing the risk of AD (OR > 1) were significantly overrepresented compared to ROHs increasing protection \( (p < 2.20 \times 10^{-16}) \). A significant ROH association with AD risk was detected upstream the \( HS3ST1 \) locus (chr4:11,189,482–11,305,456), \( \beta \) (CI 95%) = 1.09 (0.48 – 1.48), \( p \) value = 9.03 \times 10^{-4} ), previously related to AD. Next, to search for recessive candidate variants in ROHs, we constructed a homozygosity map of inbred AD cases extracted from an outbred population and explored ROH regions in whole-exome sequencing data \((N = 1449)\). We detected a candidate marker, rs117458494, mapped in the \( SPON1 \) locus, which has been previously associated with amyloid metabolism. Here, we provide a research framework to look for recessive variants in AD using outbred populations. Our results showed that AD cases have enriched homozygosity, suggesting that recessive effects may explain a proportion of AD heritability.

Introduction

Alzheimer’s disease (AD) is a neurodegenerative disorder that is the leading cause of dementia worldwide1.

AD presents a strong genetic component. Autosomal dominant mutations have been linked to familial early onset AD (EOAD) (<65 years): mutations in \( PSEN1 \), \( PSEN2 \), and \( APP \). These findings lead to the role of amyloid metabolism as disease-causing mechanism5. Despite that, dominant causes account for a minority of both familial and sporadic EOAD cases, suggesting that autosomal recessive loci might cause most EOAD cases (~90%)6. However, only two recessive mutations in the \( APP \) gene
The sporadic form of late-onset AD (LOAD) (>65 years) has a polygenic background. Heritability estimation for LOAD is, roughly, 70%6. Although, near to 40 loci has been associated with LOAD risk9–12, genetic variance captured by genome-wide strategies fall in a range of 7–31%9,13, explaining a limited part of disease heritability. Current genetic findings were made using an additive mode of inheritance, which overlooks the relevance of non-additive genetic components, i.e., the recessive model. Despite the fact these components could explain a fraction of disease heritability.

It is well known that inbreeding increases the incidence of recessive diseases. The probability of detecting a recessive locus increases in offspring of consanguineous unions14, because the partners share alleles identical-by-descent. This recent parental relatedness points to genuine regions of autozygosity. Long runs of homozygosity (ROHs)—long stretches of consecutive homozygous genotypes (>1 Mb)—are a recognized signature of recessive inheritance and provide a measure of inbreeding in studied populations. Thus far, they have been used for homozygosity mapping15. Population history, e.g., historical bottlenecks or geographical isolation, also influences homozygosity levels in individual genomes6,16,17.

An excess of homozygosity has been associated with the risk of AD in individuals of Caribbean-Hispanic and African-American ancestries18–20, suggesting the presence of inbreeding and potentially autosomal recessive AD (arAD) cases nested in these populations. Conversely, this association was not replicated for individuals of European ancestry21,22. Several factors might explain these inconsistencies, among them it has been estimated that large sample sizes (12,000–65,000) are required to detect an excess of homozygosity in outbred populations23. Thus, previous studies might be underpowered.

The limited number of deeply characterized consanguineous families, the difficulties in finding familial information for sporadic AD individuals (mainly due to the late onset of the disease) and the reduced size of intragenerational pedigrees in western countries make the search for recessive patterns of inheritance in AD complex. Furthermore, follow-up of candidate ROHs in sequencing data might be a necessary step in the definitive mapping of an arAD locus, but it has been poorly assessed to date.

Assessing the impact of homozygosity in the genetic architecture of AD, and subsequent follow-up of homoygous regions remains a challenge. To the best of our knowledge, this is the largest genomic data set exploring the influence of homozygosity in AD (n = 21,100). First, we investigated whether AD individuals from a European outbred population presented an excess of homozygosity relative to controls. Next, we measured the degree of inbreeding in AD cases. To prioritize regions with potential recessive loci, we constructed a homozygosity map of genomic regions overrepresented in detected inbred AD cases. Finally, we performed further exploration of several promising candidate ROHs using whole-exome sequencing (WES) data.

Patients and methods
The overview of the proposed strategy for ROH detection and subsequent prioritization is depicted in Fig. 1.

Genotyping data
This study includes 10 independent genome-wide data sets comprising a total sample of 21,100 unrelated individuals (11,921 AD cases and 9181 individual controls) of European ancestry (Supplementary Table 1). The recruitment and phenotyping, has been described previously12.

Genotype-level data for each cohort was processed by applying identical quality control and imputation procedures, as previously reported12. Next, we generated a merged data set combining imputed genotypes (MAF > 0.05; imputation quality R^2 > 0.90) from available data sets. We calculated identity-by-descent (IBD) with PLINK 1.9 to generate a cohort of unrelated individuals of European ancestry (Supplementary Fig. 1). All possible pairs had Pi-hat < 0.1875, a Z0 ≥ 0.75 and a Z1 ≤ 0.25. Imputed markers with call rates >0.95 and MAF > 0.05 in the merged data set were selected for ROH calling (N_{SNPs} = 2,678,325).

Runs of homozygosity (ROHs) exploration
1-Identification of individual ROHs
Individual ROH calling was conducted using the observational genotype-counting approach implemented in PLINK (v1.09) (https://www.cog-genomics.org/plink/1.9/), as it outperforms additional methods in ROH detection and it is applicable to outbred populations24. ROH detection was performed for each individual study and for the merged data set using imputed genotypes. We used a sliding window of 50 SNPs of 5000 Kb in length to scan the genome. In order to manage genomic regions with a small number of genotyping errors and discrete missingness, one heterozygote and five missing calls per window were tolerated. These parameters were similar to those described previously25. The minimal number of SNPs in a ROH was set to 100 SNPs26,27. We empirically explored two minimal length cut-offs to consider a ROH, 1 Mb and 1.5 Mb. ROHs < 1.5 Mb might reflect LD patterns of ancient origin rather than the consanguineous cultural practices and genetic isolation captured with ROHs > 1.5 Mb28. SNPs were included in a ROH if >5% of the sliding window was homozygous. The maximum
The distance between two consecutive SNPs was set to 1000 Kb apart, and SNP density to at least 1 SNP in 50 Kb.

**2-Exploration of homozygosity parameters**

To assess the data quality and genetic architecture of detected ROHs (>1 Mb and >1.5 Mb) in each individual study and in the whole dataset, we calculated: (a) the mean of the total length of ROH or sum of ROH (SROH); (b) the average ROH length (AVROH); (c) the number of ROHs (NROH); and (d) ROH-based estimates of the inbreeding coefficient, F (FROH) per individual. AVROH is the SROH divided by NROH per subject. FROH represents the proportion of homozygous segments in the autosomal genome per individual (Eq. 1). For individuals, this would be the SROH detected divided by a factor of 3,020,190 Kb, the total autosomal genome length according to the GRCh37.p13 assembly. We further explored whether the effect of homozygosity parameters was similar when: (1) ROH length was set to 1 or 1.5 Mb; and (2) the analysis was performed per data set or in the final merged database. Results emerging from these exploratory analyses are shown in Supplementary Figs. 2–3, Supplementary Tables 2–3, and Supplementary Methods. According to them, we decided to conduct downstream analyses with ROH calling at 1.5 Mb in the merged data.

\[
F_{\text{ROH}} = \frac{\text{SROH}(\text{Kb})}{\text{Autosomal genome (Kb)}} \tag{1}
\]

Copy number variants (CNV), particularly hemizygous deletions, are known to cause spurious ROHs. However, prior studies have demonstrated that the impact of performing ROH calling with or without CNVs is only 0.3% of the total ROH length. To assess the impact of CNVs deletions, we also conducted ROH calling after removing common CNV deletions extracted from the Database of Genomic Variants (DGV) (http://dgv.tcgag.ca). The same exercise was conducted after removing CNVs detected in GR@ACE dataset. Further description of CNV calling is provided in Supplementary material.
3-Identification of consensus ROHs

Consensus ROHs were defined as overlapping segments between individual ROHs observed in different individuals, with DNA segment match of at least 95% for non-missing SNP markers. Consensus ROH calling was performed using PLINK 1.9. To prevent the detection of false-positive ROHs, we extracted those consensus ROHs with a length >100 Kb and >3 SNPs.

4-Analyses

4a-Association analysis between homozygosity parameters and AD risk

To assess the quality of the data in each individual study, we explored sample distribution for each of four homozygosity parameters: NROH, SROH, AVROH, and FROH. An exploratory analysis was depicted with violin plots, which combine a box plot with a kernel density plot, using the ggplot2 package from R (Supplementary Figs. 4 and 5). The inverse rank normal transformation was performed to generalize homozygosity parameters using “rankNorm” option in the RNOmni package in R. Transformed distributions are shown in Supplementary Fig. 6. To test the association of homozygosity parameters with AD status, we developed a generalized linear model for a binominal outcome, using R for individual-level data. We tested three models, adjusting per: (1) cohort and the first four principal components (PCs) resulting from ancestry analysis. See Eq. 2; (2) cohort, PCs and age; (3) cohort, PCs, age and gender. We also conducted a sensitivity analysis excluding control individuals <60 years old (See the “Results” section).

\[ Z = \beta_1 \text{Homozygosity Parameter} + \beta_2 \text{Cohort} + \beta_3 \text{PC1} + \beta_4 \text{PC2} + \beta_5 \text{PC3} + \beta_6 \text{PC4} + e \]  
\[ \]  

4b-Association analysis between consensus ROHs and AD

The association between the phenotype and consensus ROHs was explored using a logistic model, for ROHs in or near to previously identified AD loci extracted from de Rojas et al.30 and non-AD ROHs. The model was adjusted per cohort, and the first four PCs as covariates for downstream analysis. Covariate models adjusted for age and gender, in addition to cohort and PCs, were also calculated. Regression-based results were corrected for multiple testing using a Bonferroni correction.

Next, we sought to estimate whether there was an overrepresentation of risk (\( \beta > 0 \)) or protective (\( \beta < 0 \)) consensus ROHs in our association results at different levels of length and SNP number per consensus ROH. We applied a binominal test using R.

5-The homozygosity map of inbred AD individuals

5a-Identification of inbred individuals

We used F\(_{ROH}\) to detect the subset of inbred individuals within our dataset. F\(_{ROH}\) has been previously shown to better correlate with the unobserved pedigree inbreeding22,31. The cut-off between inbred and non-inbred individuals was set to F\(_{ROH} \geq 0.0156\), which corresponds to a second-degree relation. It was assumed that there are no different biological effects below 0.0156 than in the general population33. The efficient capture of inbred individuals is shown in Supplementary Fig. 7. Next, to explore whether the frequency of consanguinity was higher in AD cases than in controls, we calculated the odds ratio and chi square p values using the epitools package in R.

5b-ROHs prioritization based on inbred AD cases

ROH detection was conducted in the subset of inbred AD cases, applying similar criteria to those previously described. Briefly, considering the long size of homozygous tracts for inbred individuals, there is a higher probability of finding a consensus ROH by chance within consanguineous AD cases than in the general population. Hence, we applied stringent criteria to define consensus ROHs. Consensus ROHs from inbred AD cases with ROH lengths >100 Kb and ROH >100 SNPs were given priority for further analysis. Shared overlapping regions between inbred AD cases and the whole data set were also identified (See bash code in Supplementary Code Material) and selected based on their overrepresentation in AD cases relative to controls (\( \beta > 0.03 \)). Prioritized regions were then explored in sequencing data. We also explored the overlapping of these regions with previously identified AD loci30.

WES data

To meet the objective of exploring most promising ROH candidates in the sequencing data, we used the Knight-ADRC-NIA-LOAD (KANL) cohort34. We excluded autosomal dominant familial cases and sporadic AD cases harboring well-known disease-causing mutations, as they could explain disease status. Thus, this study comprised 986 AD cases and 463 control individuals of European ancestry (See Supplementary Table 1 and Supplementary Fig. 1). Of these, 488 subjects presented at least a triad of three first-degree relatives with AD (See Supplementary Table 2). The candidate gene prioritization strategies using WES

6a-Gene-based analysis

To prioritize genes in consensus ROH regions, we performed gene-based analysis (986 cases vs 463 controls) (Fig. 1). To generate variant sets, variants were filtered out according to minor allele frequency (MAF < 0.01) and functional impact. The allele
frequency cut-off was established according to the Exome Aggregation Consortium (ExAC), non-Finnish European Exome Sequencing project (ESP), and 1000 G. Only those variants predicted to have a high or moderate effect according to SnpEff were included. To compute p-values per gene set, SKAT-O model was applied using R. The models were adjusted to consider the impact of the first two PCs and sex. Genes were filtered out from results if the number of variants included in the model was ≤3.

6b-Variant filtering strategy for inbred AD cases with WES and GWAS data available ROH segments emerging from inbred AD cases are the most promising candidates to harbor autosomal recessive variants. Therefore, we deeply explored ROHs by applying a variant filtering strategy. We explored 488 AD cases with complementary GWAS and WES data. Because there is a low likelihood to identify any novel or causative mutation in available databases, variants with MAF > 0.01 were excluded. All heterozygous variants were removed. Finally, only the variants mapped in individual ROHs were selected.

To map genes within ROHs, we first extracted all the variants located in ROH regions. Next, we individually annotated each one.

Results
ROH parameters are associated with AD risk
We examined the typical characteristics of the four ROH parameters (SROH, NROH, AVROH, FROH) in 21,100 unrelated European individuals from 10 independent cohorts (Supplementary Tables 1–2 and Supplementary Fig. 4). Relationships between the mean NROH and SROH are shown in Fig. 2. The mean NROH was 14.6 ± 4.6, the AVROH was 2.11 ± 0.61 Mb, and the SROH was 31.9 ± 22.2 Mb. These estimations are in accordance with those observed in European individuals, except for the NROH parameter, which was higher than in the previous studies.

Next, we tested the association of the four homozygosity parameters with AD risk. We found that (i) higher inbreeding coefficient ($F_{ROH}$) increased the risk of suffering AD [$\beta_{FROH}$ (CI95%) = 0.043 (0.009–0.076); p value = 0.013] (Table 1); (ii) AD patients presented higher average lengths of ROHs compared to controls [$\beta_{AVROH}$ (CI95%) = 0.07 (0.037–0.104); p value = 3.91 × 10$^{-5}$]; (iii) ROH number was not associated with AD risk after adjusting for age [$\beta_{NROH}$ (CI 95%) = 0.010 (–0.024–0.044); p value = 0.571] (Table 1). Results per cohort are shown in Supplementary Table 4. Notably, a sensitivity analysis conducted excluding: (1) known deletions, i.e., hemizygous segments; and, (2) deletions identified in GR@ACE CNV study; provided comparable results (Supplementary Table 5). After excluding control individuals <60yo, a stable and significant effect remains for AVROH [$\beta_{AVROH}$ (CI 95%) = 0.07 (0.031–0.103); p value = 3.51 × 10$^{-5}$] (Supplementary Table 5).

ROH analysis of AD risk using the whole data set
We identified 21,190 consensus ROHs in the merged data set (N = 21,100). We observed a significant over-
Table 1  Effect of genome-wide homozygosity measures in Alzheimer’s disease for the joint analysis.

| Dataset     | Model 1       | Model 2       | Model 3       |
|-------------|---------------|---------------|---------------|
|              | Beta (CI 95%) | P value       | Beta (CI 95%) | P value       | Beta (CI 95%) | P value       |
| FROH        | 0.051 (0.023–0.078) | 3.25 × 10⁻⁴ | 0.044 (0.010–0.077) | 0.011 | 0.043 (0.009–0.076) | 0.013 |
| AVROH       | 0.027 (0.000–0.055) | 0.051 | 0.074 (0.040–0.106) | 2.16 × 10⁻⁵ | 0.070 (0.037–0.104) | 3.91 × 10⁻⁵ |
| NROH        | 0.043 (0.015–0.071) | 2.48 × 10⁻³ | 0.010 (0.004–0.044) | 0.559 | 0.010 (0.003–0.044) | 0.571 |

Model 1: adjusted per Cohort and PCs; Model 2: Adjusted by cohort, PCs, and age; Model 3: Adjusted by cohort, PCs, and age and gender.

Association between homozygosity parameters and AD status, adjusted per Cohort, PCs, Age and Sex, was conducted in individuals with all available data; N = 19,253.

Table 2  Frequency of consensus ROHs with a potential risk or protective effect in Alzheimer’s disease.

| Dataset Model | N ROH | Risk associations | Protective associations | P value | Probability of success |
|---------------|-------|------------------|-------------------------|---------|------------------------|
| Whole dataset | 21190 | 11974            | 9216                    | < 2.2 × 10⁻¹⁶ | 0.56 |
| Category A    | 1017  | 593              | 424                     | < 2.2 × 10⁻¹⁶ | 0.58 |
| Category B    | 926   | 537              | 389                     | 1.30 × 10⁻⁶  | 0.57 |
| Category C    | 858   | 499              | 359                     | 1.98 × 10⁻⁶  | 0.58 |
| Category D    | 42    | 33               | 9                       | 2.7 × 10⁻⁸   | 0.79 |
| Whole dataset/map of inbreed AD cases | 6636  | 3969             | 2667                    | < 2.2 × 10⁻¹⁶ | 0.60 |

Strategy A, ROHs > 100 kb; >3 SNPs.
Strategy B, ROHs > 100 kb; >25 SNPs.
Strategy C, ROHs > 100 kb; >50 SNPs.
Strategy D, ROHs > 100 kb; >3 SNPs, P < 0.05.

representation of ROHs increasing the risk of suffering AD (p value < 2.20 × 10⁻¹⁶) (Table 2). The same over-representation of risk associations was detected after filtering at several levels based on the length and number of SNPs per consensus ROH (Table 2). When the test was conducted with results adjusted for cohort, PCs, age, and gender, the over-representation of risk associations remained highly significant (p value < 2.20 × 10⁻¹⁶).

We then tested the association of 11 consensus ROH (≥100 Kb and ≥3 SNPs) located in or near to previously identified AD loci (N = 38)³⁰, with AD status (Supplementary Table 6). For these analyses, Bonferroni corrected significance threshold of p = 1.32 × 10⁻³ was prespecified. We detected a strong association near to HS3ST1 locus (consensus ROH length = 115.9 Kb; chr4:11,189,482–11,305,456), (45 AD cases vs 12 controls, β (CI 95%) = 1.09 (0.48–1.48), p value = 9.03 × 10⁻⁴). This region survived age and gender adjustments (Supplementary Table 6), and was detected across 12 out of 16 datasets (Supplementary Table 7). The replication of this specific locus with AD, using ROH methodology, provides new insights of a potential recessive mechanism for this dementia locus. Among other ROH regions in or near to known AD loci (Supplementary Tables 6 and 10), we highlighted a 237 Kb ROH upstream the APP gene (chr21: 26,903,551–27,141,292), by its known role in AD⁷,8,3⁶, detected in 38 AD cases vs 26 controls (26 vs 12 inbred individuals, respectively). For non-previously associated AD regions, none ROH (N = 1006) reached the significance threshold (Bonferroni correction of p = 4.97 × 10⁻⁶). Previous significant consensus ROH (chr8: 37835460–38143780) associated with AD in Europeans²¹ was not detected in this study, which is in line of results from Sims et al.²², failing replication.

We then explored the genes located in significant risk consensus ROHs (p value < 0.05) in gene-based analysis from WES data as well (Fig. 1). A total of 33 ROHs comprising 32 genes were analyzed (included > 3 SNPs in the model; Bonferroni correction p value = 0.0015). The NECAB1 locus (chr8:91,803,921–91,971,630) presented the most significant signal (p = 0.01) (Supplementary Table 8), but none loci reached the multiple test correction threshold.

Homozgyosity mapping of AD using DNA segments identified in inbred cases

We detected 1621 individuals (958 Cases and 663 Controls) presenting a FROH ≥ 0.0156 among the total sample (N = 21,100) (Fig. 2) (Supplementary Table 9). Interestingly, inbreeding over the second degree of
consanguinity was associated with a higher risk of suffering AD [OR (95%, CI) = 1.12 (1.01–1.25); \(p\) value = 0.027], which is in line with our previous results. This supports the idea that an excess of consanguineous individuals is present in the AD population. Accordingly, the search for recessive loci that play a role in AD can first be assessed in consanguineous cases.

After ROH calling in inbred AD cases, we detected 5087 ROHs, and extracted those with \(\geq 100\) Kb and \(\geq 100\) SNPs. We then selected only over-represented regions in AD cases relative to controls in the general analysis (Fig. 1). We prioritized 807 consensus homozygous segments from inbred cases (8.6% of the total autosomal genome) (Fig. 3 and Supplementary Table 10). Among them, 12 ROHs were in or near to a previously identified AD loci (Supplementary Table 10).

After exploring genes in identified ROHs by gene-based analysis from WES data, none of them remained associated after multiple corrections (\(N_{\text{genes tested}} = 1136; \ p\) value = \(3.47 \times 10^{-5}\)) (Supplementary Table 11). Our top signal was detected in the \(FRY\) locus (\(p\) value = 0.001) (Supplementary Table 11).

Considering that recessive variants are expected at very low frequencies, even gene-based analysis would be underpowered to detect significant associations. Therefore, we decided to further prioritize loci by searching homozygous mutations within selected consensus ROHs from inbred AD subjects (Fig. 1). We identified seven AD cases that had eight new (or extremely rare) homozygous variants within long ROH segments (Table 3). All ROH segments with homozygote variants were detected in more than 6 cohorts. Two of these individuals were consanguineous (\(F_{\text{ROH}} > 0.156\)). One had a missense variant (rs140790046, c.926A > G) that encodes p.Asn309Ser change within the \(MKX\) locus. Another carried a rare variant (rs116644203) in the \(ZNF282\) locus, which was in an extremely large region of homozygosity (14.9 Mb) (Table 3). Furthermore, three additional homozygous variants were detected: (i) a variant (rs117458494) in the \(SPON1\) locus, previously related with amyloid metabolism, and (ii) two potential causative variants, carried only by this individual, within a previously identified AD region (\(TP53INP/NDUFAF6\) region) (Table 3).

Further notes and functional effect predictions for these variants are provided in Supplementary Table 12.

**Discussion**

This study represents the largest analysis of homozygosity conducted for AD. Our estimates of homozygosity provide a robust evidence suggesting that recessive allelic architecture might define a portion of AD heritability.

Previous AD ROH studies in European populations have shown negative results for the association of ROH parameters with AD. First studies had very modest sample sizes (\(N < 3000\), vs \(N_{\text{present study}} = 21,100\)).
likely were underpowered. Then, these studies used a ROH calling lengths set to 1 Mb\textsuperscript{21,22}. This generates substantial inflation in the inbreeding coefficient ($F_{ROH}$) and makes undetectable the enrichment in consanguinity due to unspecified noise (Supplementary material). These reasons might explain initial failures. We encourage other groups to conduct ROH analysis in new unrelated populations, but with large enough sample sizes and redefining the ROH lengths at least to 1.5 Mb, to better capture the recessive component of AD.

At the present study we identified a study-wise significant ROH association close to the $HS3ST1$ gene (~200 Kb). Genetic markers near to this ROH (~300 kb) have been previously associated with AD using additive models\textsuperscript{38,39}, and $HS3ST1$ locus was differentially expressed in the brain of AD cases versus controls\textsuperscript{38}. Our finding reinforces the association of this region with AD, and further suggests the role of recessiveness in explaining underlying associations. High-resolution mapping across this ROH could help to identify the causative mutation.

This study failed replication of previously detected ROH at chr8:37835460–38143780\textsuperscript{21}. Although, both studies include TGEN cohort, overlapping to some extent, the default technical parameters for ROH definition were completely different (ROH calling: 1 Mb vs 1.5 Mb). We assume that technical differences of the present study respect to prior ones, might be critical points impacting replication of ROH findings, in addition to other causes, e.g. population-specific genetic patterns, or, even, random chance.

A strength of the present study comes from our effort to prioritize consensus ROHs according to the homozygosity map of inbred AD individuals, performed by the first time in AD, and our capacity to explore them in sequencing data. This strategy lets us to find interesting candidate recessive variants in: $MKX$ and $ZNF282$ genes, identified in two independent inbred AD cases; $TP53INP1/NDUFAF6$ genomic region, previously associated with AD\textsuperscript{12,40}; and $SPON1$ locus. The $SPON1$ locus deserves a further explanation as it is directly related with APP metabolism, a key player in AD physiopathology. APP cleavage through $\beta$-secretases produces amyloid-beta (A$\beta$), which later accumulates in AD brains\textsuperscript{8}. $SPON1$ has been found to bind to APP, inhibiting its $\alpha/\beta$ cleavage\textsuperscript{37}, and to APOE family of receptors\textsuperscript{41}. Markers in this gene have been related to dementia severity\textsuperscript{42} and with the rate of cognitive decline\textsuperscript{43}. Considering prior findings and the present result, it would be biologically plausible that the presence of recessive variants in $AP\beta$\textsuperscript{45,46}, or its biological partners directly influences the amyloid cascade. Thus, we believe that $SPON1$ could be considered an interesting candidate, which deserves future resequencing efforts.
genotyped using different platforms and shared a small proportion of directly genotyped markers. Given that lower SNP density could impact the accuracy of the study, high quality imputed markers were used ($r^2 > 0.90$, MAF > 0.05). Second, to reduce dataset heterogeneity we use a set of European individuals; applied the same GWAS quality control per study; generated a merged dataset including common variants across datasets; and controlled all our analyses by cohort, to account with potential confounding.

We assumed that differences in the ROH parameters between the cases and controls are modest. In that sense, we are not expecting a very large percentage of recessive AD cases, but we expect a fraction, in the same way, that it occurs for autosomal dominant forms (<1%). Considering that, the reported findings are supporting the hypothesis of this work, a group of recessive mutation may explain a portion of AD cases. However, we suspect that the existence of a large non-allelic heterogeneity is preventing its identification.

Our gene-based analysis strategy did not show significant associations. With a decreasing allele frequency and high locus heterogeneity, the power to detect genes of interest also decreases. Despite our effort to include WES data in the present study, the available sample size could be underpowered.

The potential impact of CNV deletions on ROH analysis must be taken into consideration. Thus, we assessed its effect on our analyses, but no differences were found before and after CNV exclusion (Supplementary Table 5), which is in agreement with the previous studies. Clonal mosaicism, due to aging, could also generate spurious ROHs. At the present study an age-dependent increase in the NROH was detected in the control group (Supplementary Table 13), which partially disappeared after excluding consensus ROHs associated with age ($p < 0.05$) (See Supplementary Material, and Supplementary Table 13 and 14). We assumed that these DNA segments might contain somatic alterations, confounding ROH associations. Among age-related ROH regions, we identified some loci previously associated with AD, e.g., RORA, CD2AP, HS3ST1, and amyloid-beta burden, e.g., GLIS3; suggesting that some known AD regions could be affected by this phenomenon. These findings deserve future investigations. Despite the existence of ROH segments associated to age and somatic mosaicism phenomena, our most significant findings largely supported adjustments by age. Therefore, we feel that the major observations of this study are not affected by age-related instability of the human genome.

In summary, we demonstrated the existence of an inbreeding effect in AD and efficiently captured a fraction of inbred individuals from outbred populations, providing an improved strategy to look for recessive alleles, and to conduct future large-scale homozygosity mapping studies in AD. Furthermore, the exploration of complementary sequencing data gave an added value to this research, providing a subset of potential candidates harboring recessive variants. In any case, the proposed candidates would need confirmation in larger series. Greater efforts and larger collections of individuals with GWAS and sequencing data are needed to confirm the present findings.

Our understanding of the dynamics of population genomics in AD is far from complete, but ROH analyses provide us with a means to go further and might be an alternative strategy to uncover the genetic loci underlying AD.

**Acknowledgements**

We would like to thank patients and controls who participated in this project. The Genome Research at Fundación ACE project (GR@ACE) is supported by Fundación bancaria “La Caixa”, Grifols SA, Fundación ACE, and ISCIII (Ministry of Health, Spain). We also want to thank the private sponsors who support the basic and clinical projects of our institution (Piramal AG, Laboratorios Echevarne, Aracón Biotech S.A., and Fundación ACE). We are indebted to the Trinitat Port-Carbó legacy and her family for their support of Fundación ACE research programs. Fundación ACE is a participating center in the Dementia Genetics Spanish Consortium (DEGESTC). M.R. and M.B. receive support from the European Union/EFPIA Innovative Medicines Initiative Joint Undertaking ADAPTED and MOPEAD projects (Grants No. 115975 and 115985, respectively). M.B. and A.R. are also supported by national grants P11/02434, P11/01861, P11/01474, and P11/01301. Acción Estratégica en Salud is integrated into the Spanish National R+D+i Plan and funded by ISCIII (Instituto de Salud Carlos III). The present work was performed as part of the Biochemistry, Molecular Biology and Biomedicine doctoral program of S. Moreno-Grau at Universitat Autònoma de Barcelona (Barcelona, Spain). This work was supported by grants from the National Institutes of Health (R01AG044546, P01AG003991, R1FAG053303, R01AG059851, U01AG058922, R1FAG058501, and R1FAG057777) and the Alzheimer’s Association (NIRG-11-200110, BAND-14-338165, AARG-16-441560, and BFG-15-362540). The recruitment and clinical characterization of research participants at Washington University were supported by NIH P50 AG05681, P01 AG023991, and P01 AG026276. This work was supported by access to equipment made possible by the Hope Center for Neurological Disorders, and the Departments of Neurology and Psychiatry at Washington University School of Medicine. For additional information about GR@ACE & DEGESTC consortia, please contact with aruz@fundaciace.com The GR@ACE study group is composed by: Abdelnour C, Aguilera N, Alarcon E, Alegret M, Benaque A, Boada M, Buiza-Rueda M, Cañabate P, Carracedo A, Corbatón A, de Rojas I, Diego S, Espinosa A, Gaifajinet J, García González P, Gil S, Guitart M, González Pérez A, Hernández I, Ibarria, M, Lafuente A, Macias J, Marcólu D, Martín E, Martínez MT, Marqués M, Mauleón Á, Monté-Rubio G, Montreuil L, Moreno-Grau S, Moreno M, Orellana A, Ortega G, Pancho A, Peleja M, Pérez-Cordon A, Pineda JA, Precile R, Quintela I, Real LM, Rodríguez-Gómez O, Rosende-Roca M, Ruiz A, Ruiz S, Sáez ME, Sanabria A, Santos-Santos MA, Serrano-Rios M, Sotolongo-Grau O, Tárraga L, Valero S, Vargas L. The DEGESTC consortium is composed by: Adarves-Gómez A-D, Alarcón-Martín E, Alvarez I, Alvarez V, Amen-Ferrer G, Antequera M, Artalejo M, Antúnez C, Baquero M, Bernal M, Blesa R, Boada M, Buiza-Rueda D, Bulldo MJ, Burguera JA, Calero M, Carrillo F, Carrión-Claro M, Casajeros MJ, Clarimón J, Cruz-Gamero JM, de Pancorbo MM, de Rojas I, del Ser T, Dietz-Fairen M, Fortea J, Franco E, Frank-García A, García-Alberca JM, García Madrona S, García-Ribas G, Gómez-Garre P, Hernández I, Hevilla S, Jesús S, Laboratorios Echevarne S, Lage C, Legaz A, Lleó A, López-de Munáin A, Martínez MT, Medrano-Reséns C, Moreno-Grau E, Morillo A, Morillo S, Núñez E, Ochando R, Orrego J, Peleja M, Peleja L, Peleja A, Peleja X, Peleja N, Portillo R, Pascual M, Pascual A, Pateras I, Pineda JA, Precile R, Quintela I, Real LM, Rodríguez Gómez O, Rosende-Roca M, Ruiz A, Ruiz S, Sáez ME, Sanabria A, Santos Santos MA, Serrano Rios M, Sotolongo Grau O, Tárraga L, Valero S, Vargas L. The DEGESTC consortium is composed by: Adarves-Gómez A-D, Alarcón-Martín E, Alvarez I, Alvarez V, Amen-Ferrer G, Antequera M, Artalejo M, Antúnez C, Baquero M, Bernal M, Blesa R, Boada M, Buiza-Rueda D, Bulldo MJ, Burguera JA, Calero M, Carrillo F, Carrión-Claro M, Casajeros MJ, Clarimón J, Cruz-Gamero JM, de Pancorbo MM, de Rojas I, del Ser T, Dietz-Fairen M, Fortea J, Franco E, Frank-García A, García Alberca JM, García Madrona S, García Ribas G, Gómez-Garre P, Hernández I, Hevilla S, Jesús S, Laboratorios Echevarne S, Lage C, Legaz A, Lleó A, López-de Munáin A, López-García S,
Moore Grau et al. Translational Psychiatry (2021) 11:1142

Page 10 of 12

Macias D, Manzanares S, Marín M, Marín-Muñoz J, Marín T, Marquéz M, Martín Montes A, Martinez B, Martinez C, Martinez V, Martinez-Lage Álvarez P, Medina M, Mendioroz Irazte M, Menéndez-González M, Mir P, Molinuevo JL, Montel-Rubio G, Montreuil L, Moreno-Grau S, Orellana A, Pastor A.B, Pastor P, Pérez Tur J, Pérez-Tocino T, Pichó Ripeló G, Rabano A, Real de Asúa D, Rodrigo S, Rodríguez-Rodriguez E, Royo JL, Ruiz A, Sanchez del Valle Diaz B, Sánchez-Juan P, Sastre I, Sotolongo-Grau O, Tárraga L, Valero S, Vicente M.P, Vivancos L. Data collection and sharing for this project was partially funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging and the National Institute of Biomedical Imaging and Bioengineering, as well as through generous contributions from the following: Abbvie; the Alzheimer’s Association; the Alzheimer’s Drug Discovery Foundation, Arancon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSp, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; Elrohim, F, Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; FujiRebio; GE Healthcare, IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development Corporation; Tecnología para la Automatización de Farmacia, AGIA IDEA (Consejería de Sanidad y Enfermedades de la Infancia y de la Juventud de la Junta de Andalucía); Corporación Tecnológica de Andalucía; CorpoMADE (Consejería de Educación, Deportes, Cultura, y Juventud de la Junta de Extremadura); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Junta de Andalucía); Corporación Tecnológica de Andalucía; Agencia IDEA (Consejería de Sanidad y Servicios Sociales de la Juna...
quality control for GR@ACE dataset. I.O., O.M., and A.C. carried out genotyping procedures for GR@ACE dataset. J.C., M.I.B., G.G.R., P.P., P.S.J., V.A., G.P.R., J.M.A., J. L.R., E.F.M., P.M., M.M., M.M., A.A., J.A., C.A., and A.R. are members of DEGESCO consortium. S.M.G., A.G., L.M., and M.E.S. contributed to data processing. A.B., L.T., M.B., C.C., and A.R. participated in funds procurement. S.M.G. prepared the first draft of the manuscript. S.M.G., M.V.F., C.C., and A.R. were responsible for the critical revision of the manuscript. All authors read and approved the final manuscript.

Code availability

Script used to identify shared overlapping regions between imbed AD cases and the whole data set is available at Supplementary Code Material. Additional scripts used to conduct ROH calling are available from the authors upon request.

Data availability

GR@ACE dataset analyzed during the current study is available into the European Genome-phenome Archive (EGA) (https://ega-archive.org), which is hosted by the EBI and the CRG, under accession number EGA500001003424 (https://ega-archive.org/studies/EAG500001003424). Additional GWAS datasets used in these studies are available upon request in dbGap. WES dataset used in the study can be downloaded from https://dss.naags.org/studies/sa000004/. Data used in preparing this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher’s note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1038/s41388-020-01145-1.

Received: 13 August 2020 Revised: 27 November 2020 Accepted: 4 December 2020 Published online: 24 February 2021

References

1. 2019 Alzheimer’s disease facts and figures. Alzheimer’s Dement. 15, 321–387 (2019).
2. Sherrington, R. et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer’s disease. Nature 375, 754–760 (1995).
3. Sherrington, R. et al. Alzheimer’s disease associated with mutations in presenilin 2 is rare and variably penetrant. Hum. Mutat. Genet. 5, 985–988 (1996).
4. Goate, A. et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer’s disease. Nature 349, 704–706 (1991).
5. Hardy, J. & Selkoe, D. J. The Amyloid Hypothesis of Alzheimer’s disease: progress and problems on the road to therapeutics. Science 297, 353–356 (2002).
6. Wingo, T. S., Lah, J. J., Levey, A. I. & Cutler, D. J. Autosomal recessive causes likely in early-onset Alzheimer disease. Arch. Neurol. 69, 59–64 (2012).
7. Di Fede, G. et al. A recessive mutation in the APP gene with dominant-negative effect on Amyloidogenesis. Science 323, 1473–1477 (2009).
8. Tomiyama, T. et al. A new amyloid β variant favoring oligomerization in Alzheimer’s-type dementia. Ann. Neurol. 63, 377–387 (2008).
9. Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer’s disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat. Genet. 51, 414–430 (2019).
10. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. Nat. Genet. 51, 404–413 (2019).
11. Sims, R.J. et al. Rare coding variants in PLCL2, AB13, and TRM2 implicate microglial-mediated innate immunity in Alzheimer’s disease. Nat. Genet. 49, 1373–1384 (2017).
12. Moreno-Grau, S. et al. Genome-wide association analysis of dementia and its clinical endophenotypes reveal novel loci associated with Alzheimer’s disease and three causality networks: the GR@ACE project. Alzheimer’s Dement. 15, 1333–1347 (2019).
13. Ridge, P. G. et al. Assessment of the genetic variance of late-onset Alzheimer’s disease. Neurobiol. Aging 41, 200.e13–200.e20 (2016).
14. Garrod, A. E. The incidence of alkaptonuria: a study in chemical individuality. Lancet 1616, 1902–1911 (1902).
15. Lande, E. & Bostein, D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236, 1567–1570 (1987).
16. Jakulka, E. et al. The Genome-wide patterns of variation expose significant substructure in a founder population. Am. J. Hum. Genet. 83, 787–794 (2008).
17. Gross, A. et al. Population-genetic comparison of the Sardinian isolate population in Germany with the German KORA population using genome-wide SNP arrays. BMC Genet. 12, 67 (2011).
18. Vardarajan, B. N. et al. Inbreeding among Caribbean Hispanics from the Dominican Republic and its effects on risk of Alzheimer disease. Genet. Med. 17, 639–643 (2015).
19. Ghani, M. et al. Evidence of recessive Alzheimer disease loci in a Caribbean Hispanic data set: genome-wide survey of runs of homozygosity. JAMA Neurol. 70, 1261–1267 (2013).
20. Ghani, M. et al. Association of long runs of homozygosity with Alzheimer Disease among African American individuals. JAMA Neurol. 72, 1313–1323 (2015).
21. Nalls, M. A. et al. Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer’s disease. Neurogenetics 10, 183–190 (2009).
22. Sims, R. et al. No evidence that extended tracts of homozygosity are associated with Alzheimer’s disease. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. 156, 764–771 (2011).
23. Keller, M. C., Vitscher, P. M. & Goddard, M. E. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetcs 189, 237–249 (2011).
24. Howrigan, D. P., Simonson, M. A. & Keller, M. C. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. BMC Genomics 12, 460 (2011).
25. Joshi, P. K. et al. Directional dominance on stature and cognition in diverse human populations. Nature 523, 459–462 (2015).
26. Lencz, T. et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc. Natl Acad. Sci. USA 104, 19942–19947 (2007).
27. Christofidou, P. et al. Runs of homozygosity: association with coronary artery disease and gene expression in monocytes and macrophages. Am. J. Hum. Genet. 97, 228–237 (2015).
28. McQuillan, R. et al. Runs of homozygosity in European populations. Am. J. Hum. Genet. 83, 359–372 (2008).
29. Sudmant, P. H. et al. An integrated map of structural variation in 2,504 human genomes. Nature 526, 75–81 (2015).
30. Rojas, I. de et al. Common variants in Alzheimer’s disease: novel association of six genetic variants with AD and risk stratification by polygenic risk scores. medRxiv. 19012021 (2020).
31. Gazal, S. et al. Inbreeding coefficient estimation with dense SNP data: comparison of strategies and application to HapMap II. Hum. Hered. 77, 49–64 (2014).
32. Ceballos, F. C., Haznelhurst, S. & Ramsay, M. Runs of Homozygosity in sub-Saharan African populations provide insights into a complex demographic and health history. Hum. Genet. 138, 1123–1142 (2019).
33. Bittles, A. H. & Black, M. L. Consanguinity, human evolution, and complex diseases. Proc. Natl Acad. Sci. USA 107, 1779–1786 (2010).
34. Fernández, M. V. et al. Analysis of neurodegenerative Mendelian genes in clinically diagnosed Alzheimer Disease. PLoS Genet. 13, e1007045 (2017).
35. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6, 80–92 (2012).
36. Shulman, J. M. et al. Genetic susceptibility for Alzheimer disease neurotic plaque pathology. JAMA Neurol. 70, 1150–1157 (2013).
37. Ho, A. & Sudhof, T. C. Binding of F-spondin to amyloid-beta precursor protein: a candidate amyloid-beta precursor protein ligand that modulates amyloid-beta precursor protein cleavage. Proc. Natl Acad. Sci. USA 101, 2548–2553 (2004).
38. Broc, I. J. et al. Dissecting the genetic relationship between cardiovascular risk factors and Alzheimer’s disease. *Acta Neuropathol.* **137**, 209–226 (2018).
39. Wittecoal, A. et al. Meta-analysis of Alzheimer’s disease on 9,751 samples from Norway and IGAP study identifies four risk loci. *Sci. Rep.* **8**, 18088 (2018).
40. Escott-Price, V. et al. Gene-wide analysis detects two new susceptibility genes for Alzheimer’s disease. *PLoS ONE* **9**, e94661 (2014).
41. Hoe, H-S. et al. F-Spondin Interaction with the Apolipoprotein E receptor ApoEr2 affects processing of amyloid precursor protein. *Mol. Cell Biol.* **25**, 9259–9268 (2005).
42. Jahanshad, N. et al. Genome-wide scan of healthy human connectome discovers SPON1 gene variant influencing dementia severity. *Proc. Natl Acad. Sci. USA* **110**, 4768–4773 (2013).
43. Sherva, R. et al. Genome-wide association study of the rate of cognitive decline in Alzheimer’s disease. *Alzheimer’s Dement.* **10**, 45–52 (2014).
44. Laurie, C. C. et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* **44**, 642–650 (2012).
45. Deming, Y. et al. Genome-wide association study identifies four novel loci associated with Alzheimer’s endophenotypes and disease modifiers. *Acta Neuropathol.* **133**, 839–856 (2017).

**Author details**

**The GR@ACE study group**
C. Abdelnour1,2, N. Aguilera1, E. Alarcón-Martín1, M. Alegret1,2, A. Benaque1, M. Boada1,2, M. Buendía1, P. Cañabate1,2, A. Carracedo2,26, A. Corbatón27, I. de Rojas1, S. Diego1, A. Espinosa1,2, A. Gailhajenet1, P. García González1, S. Gil1, M. Guitart1, A. González Pérez6, I. Hernández1,2, M. Ibarria1, A. Lafuente1, J. Macías7, O. Maroñas4, E. Martín1, M. T. Martínez27, M. Marquí1, A. Mauleón1, G. Monté-Rubio1, L. Montreuil1, S. Moreno-Grau1,2, M. Moreno1, A. Orellana1, G. Ortega12, A. Pancho1, E. Pelejà1, A. Pérez-Cordon1, J. A. Pineda8, S. Preckler1, I. Quintela5, L. Tárraga1,2, S. Valero1,2, M. A. Labrador Espinosa2,20, C. Lage2,13, A. Legaz, A. Lleó2,8, A. Frank-García2,10,32, J. M. García-Alberca17, S. García Madrona22, G. Garcia-Ribas11, P. Gómez-Garre2,20, I. Hernández1,2, S. Hevilla17, S. Jesús2,20, M. A. Labrador Espinosa2,20, C. Lage2,13, A. Legaz, A. Lleó8,28, A. López de Munáin33, S. López-García2,13, D. Macias2,20, S. Manzanedo25,34, M. Marín19, J. Marín-Muñoz25, T. Marín17, M. Marquí1,2, A. Martín-Montes29,32, B. Martínez25, C. Martínez15,35, V. Martínez25, P. Martinez-Lage Álvarez26, M. Medina2,21, M. Mendioroz Iriarte37, M. Menéndez-González15,38, P. Mir2,20, J. L. Molinuevo39, L. Montreuil1, S. Moreno-Grau1,2, A. Orellana1, A. B. Pastor21, P. Pastor12, J. Pérez-Tur2,40,41, T. Periñán-Tocino2,20, G. Piñol-Ripoll2,16, A. Rábano2,21,23, D. Real de Asúa42, S. Rodrigo19, E. Rodríguez-Rodríguez2,13, J. L. Royo1,2, A. Ruiz1,2, R. Sanchez del Valle Díaz43, P. Sánchez-Juan2,13, I. Sastre2,19, O. Sotolongo-Grau1, L. Tárraga1,2, S. Valero1,2, M. P. Vicente23 and L. Vivancos25

27Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas, CIBERDEM, Madrid, Spain. 28Hospital Clínico San Carlos, Madrid, Spain. 29Department of Neurology, Hospital Universitario Son Espases, Palma, Spain. 30Servei de Neurologia, Hospital Universitari i Politécnic La Fe, Valencia, Spain. 31BIOMICs, País Vasco; Centro de Investigación Lascaray. Universidad del País Vasco UPV/EHU, Vitoria-Gasteiz, Spain. 32Neurology Service, Hospital Universitario La Princesa, Madrid, Spain. 33Hospital Donostia de San Sebastián, San Sebastián, Spain. 34Fundación para la Formación e Investigación Sanitarias de la Región de Murcia, Murcia, Spain. 35Centro de Investigación y Terapias Avanzadas. Fundación CITA-alzheimer, San Sebastián, Spain. 36NavarraBiomed, Pamplona, Spain. 37Navarrabiomed, Pamplona, Spain. 38Hospital Universitario Central de Asturias, Oviedo, Spain. 39Barcelona Beta Brain Research Center – Fundació Pasqual Maragall, Barcelona, Spain. 40Unitat de Genètica Molecular, Institut de Biomedicina de València-CSIC, Valencia, Spain. 41Unidad Mixta de Neurología Genética. Instituto de Investigación Sanitaria La Fe, Valencia, Spain. 42Hospital Universitario La Princesa, Madrid, Spain. 43Hospital Clinic Barcelona, Barcelona, Spain