Both rare and common genetic variants contribute to autism in the Faroe Islands

Short title: The genetic architecture of autism in the Faroe Islands

Claire S Leblond[1,2,3,4¶*], Freddy Cliquet[1,2,3,4¶], Coralie Carton[1,2,3,4¶], Guillaume Huguet[1,2,3,4], Alexandre Mathieu[1,2,3,4], Thomas Kergrohen[1,2,3,4], Julien Buratti[1,2,3,4], Nathalie Lemière[1,2,3,4], Laurence Cuisset[5], Thierry Bienvenu[5], Anne Boland[6], Jean-François Deleuze[6], GenMed consortium, Tormodur Stora[7], Rannva Biskupstoe[8], Jónrit Halling[9], Guðrið Andorsdóttir[8], Eva Billstedt[10], Christopher Gillberg[10,11&], and Thomas Bourgeron[1,2,3,4&]

[1] Institut Pasteur, Human Genetics and Cognitive Functions Unit, Paris, France

[2] CNRS UMR 3571 Genes, Synapses and Cognition, Institut Pasteur, Paris, France

[3] University Paris Diderot, Sorbonne Paris Cité, Paris, France

[4] Centre de Bioinformatique, Biostatistique et Biologie Intégrative, Paris, France

[5] Laboratoire de Génétique et Biologie Moléculaires, Hôpital Cochin, HUPC, Paris, France

[6] Centre National de Recherche en Génomique Humaine (CNRGH), Institut de Biologie François Jacob, CEA, Université Paris-Saclay, Evry, France

[7] Department of Psychiatry, National Hospital Faroe Islands, Tórshavn, Faroe Islands

[8] Ministry of Health Genetic Biobank of the Faroes Tórshavn Faeroe Islands

[9] Faculty of Science and Technology, The University of the Faroe Islands, Tórshavn, Faroe Islands

[10] Gillberg Neuropsychiatry Centre, Institute of Neuroscience and Physiology, Gothenburg University, Gothenburg, Sweden

[11] University of Glasgow, Glasgow, Scotland

* Corresponding author

E-mail: claire.leblond@pasteur.fr

¶ These authors contributed equally to this work.

& These authors also contributed equally to this work.

Abstract

The number of genes associated with autism is increasing, but few studies have been performed on epidemiological cohorts and in isolated populations. Here, we investigated 357 individuals from the Faroe Islands including 36 individuals with autism, 136 of their relatives and 185 non-autism controls. Data from SNP array and whole exome sequencing revealed that individuals with autism compared to controls had a higher burden of copy-number variants ($p < 0.05$), higher inbreeding status ($p < 0.005$) and higher load of homozygous deleterious variants ($p < 0.01$). Our analysis supports the role of several genes/loci associated with autism (e.g. *NRXN1*, *ADNP*, 22q11 deletion) and identified new truncating (*e.g. GRIK2*, *ROBO1, NINL* and *IMMP2L*) or recessive deleterious variants (*e.g. KIRELL3* and *CNTNAP2*) affecting autism-risk genes. It also revealed three genes involved in synaptic plasticity, *RIMS4*, *KALRN* and *PLA2G4A,* carrying *de novo* deleterious variants in individuals with autism without intellectual disability. In summary, our analysis provides a better understanding of the genetic architecture of autism in isolated populations by highlighting the role of both common and rare gene variants and pointing at new autism-risk genes. It also indicates that more knowledge about how multiple genetic hits affect neuronal function will be necessary to fully understand the genetic architecture of autism.

Author summary

Autism is characterized by problems in social communication and stereotyped behavior. To improve our understanding of the genetic architecture of autism, we studied a group of children and adolescents from the Faroe Islands, an isolated population living on archipelago located in the North Atlantic Ocean, halfway between Norway, Iceland and Scotland since the 9th century. The genetic profile of this small but genetically homogenous epidelmiological cohort of the Faroe individuals revealed that both rare and common genetic variants contribute to the susceptibility to autism. Our analysis supports the role of several genes previously associated with autism and points at new candidate genes involved in the neuronal connectivity of the brain. In summary, our analysis provides a comprehensive framework to analyze the complex heterogeneity of autism in order to improve the diagnostic, the care and the integration of individuals with autism.

Introduction

Autism spectrum conditions (ASCs; henceforth 'autism') are diagnosed in 1-2% of the population worldwide and are characterized by atypical social communication and the presence of restricted interests, stereotyped and repetitive behaviors. Individuals with autism can also suffer from psychiatric and medical conditions including intellectual disability (ID), epilepsy, motor control difficulties, attention-deficit hyperactivity disorder (ADHD), tics, anxiety, sleep disorders, depression or gastrointestinal problems (1). The genetic susceptibility to autism can vary from one individual to another. In some cases, a single *de novo* variant can be detected. On the contrary, in some cases, the genetic architecture is more complex and involved thousands of common genetic variants, each one with low impact but collectively increasing the susceptibility to autism (2). Most of our knowledge on the genetics of autism comes from studies on unrelated individuals with autism who do not share a recent common ancestor. Several studies investigated families with autism from countries where consanguinity is high (3,4), but the genetic architecture of autism in isolated populations remains largely unknown.

The Faroe Islands is an archipelago located in the North Atlantic Ocean, halfway between Norway, Iceland and Scotland (Fig 1A). The population (approximately 49,000 inhabitants) was founded in the 9th century by a small number of emigrants from Norway. The population remained at a small size for centuries until they experienced a rapid expansion in the 1800s. Previous genetic studies indicated that individuals from Scotland, Norway, Sweden, Ireland, Iceland and British Isles have significantly contributed to the current gene pool of the Faroese population (5,6).

We previously showed that the prevalence of autism in the Faroe Islands (0.94% of the population (7–9)) was similar to many other western countries. In this study, we ascertained the genetic profile of 357 individuals including an epidemiological cohort of 36 individuals

94 with autism born between 1985 and 1994 (Fig 1 and S1 Fig), their relatives (n=136) and a
95 group of 185 controls. We first investigated the known causes of autism and then identified
96 new candidate genes. We also ascertained the impact of inbreeding and the load of deleterious
97 homozygous variants on the risk of autism. Finally, both rare and common genetic variants
98 were used to stratify individuals with autism and to compare their genetic and clinical
99 profiles.
100

Results

The genetic diversity in the Faroe Islands

A total of 67 children and adolescents with autism were detected in a total population study of individuals aged 8-17 years living in the Faroe Islands and born between 1985 and 1994 (8,10). Thirty-six of the 67 individuals (54% of the total group) signed (or had parents who signed) informed consent forms and were included in the genetic study. Participants and non-participants in the genetic study were similar in terms of gender and cognitive abilities (Fig 1B). In addition, we collected DNA from 136 of their relatives and from 185 "non-autism controls. The genetic profile included a high-density Illumina SNP array interrogating > 4.3 millions of single nucleotide polymorphisms (SNPs) and a whole exome sequencing (WES) to discover new variants. Using identical-by-state (IBS) genomic distance (see materials and methods), we first compared the population structure of individuals from the Faroe Islands with worldwide populations (Fig 1C). All individuals were clustered in the Faroese population (with the exception of seven controls, but still with European genetic background). Using admixture, we showed that individuals from the Faroe Islands have their genome in majority constituted from “European component” (S2 Fig). As expected from the demographic history, individuals from the Faroese population displayed a higher degree of inbreeding compared with others world populations (Fig 1D).

Contribution of *de novo* variants

We ascertained the burden of *de novo* variants since they are key players in the genetics of autism (11,12). The *de novo* variants were identified for 31 independent families including 28 individuals with autism and 45 siblings for whom DNA of both parents was available (see Clinical notes in S1 Appendix for the pedigrees). The combined analysis of genotyping and WES data revealed the presence of *de novo* chromosomal abnormalities and exonic copy-

number variants (CNVs) in 3 out of 26 individuals with autism (11.5%) and 1 out of 43 siblings (2.3%). One female PN400129 had a trisomy of chromosome 21 and was diagnosed with autism, ID and Down syndrome (S3 Fig and S1 Table). One female PN400533 with atypical autism without ID carried a *de novo* 2.9 Mb deletion on chromosome 22q11 causing DiGeorge/velocardial syndrome. A male PN400115 with atypical autism without ID carried a *de novo* 425.5 kb deletion removing the six first exons of the *NRXN1α*.We also found a 91.4 kb deletion removing all exons of *ADNP* in a male with autism and ID (PN400125). The deletion was not found in the mother and was most likely *de novo*, but father's DNA was not available and none of the SNPs within the deletion were informative to confirm the *de novo* status of the deletion. The *de novo* CNV observed in the sibling (PN400170) was a duplication of 782 kb affecting 5 genes (*CNPY1*, *DPP6*, *EN2*, *HTR5A*, *INSIG1* and *PAXIP1*).

Using the WES data, we detected the presence of *de novo* single nucleotide variants (SNVs) and small insertions/deletions (indels)(S2 Table). Overall, the rate of *de novo* exonic SNV/indels was similar to other studies (13) and was not different in individuals with autism (0.93) and their siblings (0.96). The variants were considered as probably deleterious when they were likely gene disruptive (LGD, for example stop gain or frame shift variant) or missense events with a combined annotation dependent depletion (CADD) score > 30 (MIS30) (14). There was also no significant increase in the rate of *de novo* deleterious variants in individuals with autism compared to their siblings and no significant enrichment in genes associated with autism (SFARI genes (15)) or expressed in the brain (Brain genes, see subjects and methods for gene selection). Nevertheless, several deleterious variants were identified in known genes for autism (*MECP2*) or compelling candidate genes (*RIMS4, KALRN, PLA2G4A)* (S4 Fig). Clinical details on the individuals with autism carrying those variants are available in the S1 Appendix.

Contribution of rare CNVs and SNVs/indels variants

The overall burden of exonic CNVs was higher in individuals with autism compared to controls for both deletions ($p = 0.02$) or duplication ($p = 0.006$) (Fig 2A and S3 Table). The burden of deletions was also higher for autism-associated genes listed in the SFARI database ($p = 0.01$), for genes intolerant to loss-of-function variant (pLI > 0.9) (16) ($p = 0.01$) or genes expressed in the brain ($p = 0.02$). For duplications, only genes expressed in the brain were more frequently duplicated in individuals with autism compared to controls ($p = 0.005$). These differences however do not survive corrections for multiple tests and we had no significant difference between individuals with autism and their siblings. Among the SFARI genes affected by the CNVs, we identified a 58 kb maternal inherited deletion including the *IMMP2L*, a 2 Mb paternal inherited duplication on the pseudo-autosomal region 1 including *SHOX* and *ASMT*, and a 39 kb maternal inherited duplication of *TBL1XR1* (S3 Table).

Our analysis was restricted to CNVs affecting exons, but a large 357 kb duplication within intron 5 of the *NLGN1* gene and covering a long *NLGN1* antisense noncoding RNA was paternally inherited in a female (PN400102) with autism and no ID. There was no such rare intronic CNVs in the SFARI genes in siblings and controls.

We then run gene-wise association tests from rare SNV/indels (MAF < 5%). None of associations were gene wide significant, but *KIF1B*, *FOXD3*, *RNF181* and *SDPR* were among the top genes ($p < 0.001$) detected by both the CMC-EMMAX and SKAT-O analyses (Fig 2B, S5 Fig and S4-S5 Tables). For *FOXD3*, *RNF181*, and *SDPR* the association was mainly driven by one missense variant, whereas for *KIF1B*, several variants contributed to the association. Single variant tests were performed (S6 Table) and among the hits with $p < 0.009$, two variants affect genes previously associated with neurodevelopmental disorders (NDD). A variant p.R276W in *NINL*, a gene previously associated with autism (17) was more frequent in individuals with autism (P400121 was homozygous) compared to controls

($p = 0.0005$). This Ninein-like protein is known to associate with motor complex and interact with the ciliopathy-associated proteins lebercilin, USH2A and CC2D2A. A *CHAMP1* variant (p.A586S) was more frequent in individuals with autism compared with controls ($p < 0.003$) and it was previously shown that *de novo* variants in *CHAMP1* cause ID with severe speech impairment (18).

Finally, we found truncating variants affecting several SFARI genes (*e.g. PRODH*, *ERBB4*, *GRIK2*, *ROBO1*, *RBMS3*, and *IMMP2L*; Fig 2B and S7 Table), but the number of individuals with autism carrying these variants was too low to detect a significant association.

Contribution of recessive variants

Since inbreeding increases the risk for individuals to carry homozygous deleterious variants, we first compared the inbreeding status of the individuals with autism, their relatives and controls. Patients ($p < 0.005$) as well as their siblings ($p < 0.01$) had a higher inbreeding coefficient compared with controls (Fig 3A). More interestingly, we found that individuals with autism were carrying more deleterious homozygous variants (LGD, MIS30, gnomAD MAF < 1%) than controls ($p < 0.0005$; Fig 3B and S8 Table). Genes carrying deleterious homozygous variants in affected individuals were significantly enriched in the combined gene-set lists (SFARI + pLI > 0.9 + Brain genes) compared to controls ($p < 0.05$; Fig 3C).

In one consanguineous family, we found a *KIRREL3* homozygous damaging missense variant (p.R562L) affecting a conserved residue in the cytoplasmic domain of this synaptic adhesion molecule (19) listed in SFARI and associated with NDD (20,21) (22) (Fig 3D). Interestingly, the female PN400528 with autism and a normal IQ was also homozygous for another deleterious variant (p.N687K) affecting *TECTA*, a SFARI gene associated with autism (21) and deafness (23).

In another family, the male PN400579 was homozygous for two variants affecting *CNTNAP2* and *PEX6* (Fig 3E). Recessive *CNTNAP2* variants are associated with Pitt-Hopkins like syndrome 1 and cortical dysplasia-focal epilepsy syndrome (MIM #610042). The *CNTNAP2* p.E680K variant affects a highly conserved amino acid within the fibrinogen domain of the protein. Recessive *PEX6* variants are associated with Heimler syndrome 2 a recessive peroxisome disorder characterized by sensorineural hearing loss, amelogenesis imperfecta and nail abnormalities, with or without visual defects (MIM #616617). The homozygous variant p.R601Q carried by the male with autism can be considered pathogenic since it was previously detected in several independent patients diagnosed with Heimler syndrome 2 (24). Details on the clinical profiles of the families are available in S1 Appendix.

Contribution of common variants

In order to complete the genetic profile of all the individuals, we investigated the contribution of the common variants (MAF > 5%) using three complementary approaches: (i) genome wide association studies (GWAS) in cases and controls using three models (allelic, recessive and dominant), (ii) a burden/collapsing test that aggregates the variants located in a gene, (iii) a calculation of the autism polygenic risk score.

The results of the GWAS are presented in S6-S7 figs and a list of the top 45 loci that display p-values under 10^{-5} are shown in Fig 2B and S9 Table. Two of these loci were located within or nearby genes previously associated with NDD. *TNIK* on chromosome 3q26 (rs1492859; $p = 5.59\text{x}10^{-7}$) encodes a key synaptic partner of *DISC1*, a gene associated with psychiatric disorders (25). *TMEFF2* on chromosome 2q32.3 (rs6737056; $p = 3.47\text{x}10^{-7}$) is highly expressed in the brain and is known to modulate the AKT/ERK signaling pathways that are supposedly perturbed in autism (26). *TMEFF2* is also a target of *CHD8*, a chromatin remodeling gene associated with autism (27,28). Using the gene-based analysis implemented

in MAGMA (29), only one gene *WHAMM* on chromosome 15q25.1 had a $p < 10^{-5}$ (S10-S11 Tables). *WHAMM* is expressed in the brain and acts as a regulator of membrane dynamics that functions at the interface of the microtubule and actin cytoskeletons (30).

Finally, we ascertained the autism polygenic risk score (PRS-autism) for each individual (Fig 4). The PRS-autism was calculated using PRSice-2 from a previous GWAS using over 16,000 individuals with autism (31) who do not overlap with this sample. We found a significantly higher PRS-autism in individuals with autism compared to controls ($p < 0.01$; Fig 4A). Remarkably, in the autism group, the PRS-autism was significantly higher in individuals without ID compared to those with ID ($p < 0.05$, Fig 4B).

Genetic stratification of autism in the Faroe Islands

In order to stratify individuals with autism, we used the number of rare deleterious variants in SFARI genes (including CNVs) and the PRS-autism estimated from the common variants (Fig 5). Hierarchical clustering found four clusters. The first one comprised seven individuals with high PRS-autism and high burden of deleterious variants in SFARI genes. In this cluster, none of the individuals had ID. The second "cluster" included only one individual who was diagnosed with autism and Down syndrome with a low PRS-autism and high burden of SFARI deleterious variants. In the third cluster, fourteen individuals had low PRS-autism and low burden of SFARI genes deleterious variants. In this cluster 50% of the individuals had ID. In the fourth cluster, fourteen individuals had higher PRS-autism, but low burden of SFARI deleterious variants. In this cluster, 29% of the individuals had ID and it includes the three individuals with autism with epilepsy from the cohort.

Discussion

In this study, we investigated a group of individuals with autism that has two distinctive features. First, the group is representative of a general population cohort of all young people living in the Faroe Islands at one point in time, meaning that it was not biased for inclusion/exclusion criteria used for research studies. Secondly, the Faroese population has a more homogeneous genetic background compared to most other populations.

Autism-risk genes in the Faroe Islands

We found a subset of individuals carrying strongly deleterious variants (some of which appeared *de novo*) affecting single gene or chromosomal regions. The chromosomal abnormalities included on case each of trisomy 21 and 22q11 deletion (causing Down and DiGeorge/velocardial syndromes, respectively). This was not surprising to find such known genetic disorders in an epidemiologic cohort since the prevalence of autism in individuals diagnosed with these syndromes is higher than in the general population (16-37% for Down syndrome (32), and 23-50% for DiGeorge/velocardial syndrome deletion (33,34)). We also revealed new variants in known autism-risk genes (*ADPNP*, *NRXN1*, *NINL*, *MECP2*) and points at new compelling candidate genes such as *KALRN, PLA2G4A* and *RIMS4*.

The *KALRN* gene encodes for a guanine nucleotide exchange factor (GEF) with strong homology to *TRIO*, a gene previously associated with autism (35). *KALRN* is expressed in neuronal tissue during embryonic development (36) and has been associated with schizophrenia risk through association analysis and postmortem analyses of individuals with autism' cortical KALRN mRNA and protein levels (37),(38). It is also a binding partner of the Huntingtin and a regulator of structural and functional plasticity at dendritic spines. The *de novo* variant (p.N2024D; CADD=26.7) was never observed in the general population affects a key amino acid of the GEF domain conserved through evolution and present in *Drosophila*

melanogaster and *Caenorhabditis elegans*. The male individual (PN400117) carrying this *de novo KALRN* variant has no ID.

PLA2G4A gene encodes the cytosolic phospholipase A2α that catalyzes the hydrolysis of membrane phospholipids to produce arachidonic acid. Mice lacking *Pla2g4a* display abnormalities in neuronal maturation (narrow synaptic cleft) (39) and long-term potentiation (LTP) (40). The *de novo* variant (p.R485C; CADD=35) has never been observed in the general population and is predicted as a deleterious variant falling in the catalytic domain of the protein. The female PN400102 carrying this *de novo* variant has no ID.

RIMS4 codes for a presynaptic proteins that plays a key role for dendritic and axonal morphogenesis (41). *RIMS1* and *RIMS3* have already been associated with autism (42–44) . The individual (PN400137) carrying the *de novo RIMS4* stop variant (p.Y204*; CADD=38) has a normal IQ (Performance IQ=108, Verbal IQ=116). The variant is predicted to truncate the last quarter of the protein and was never observed in the general population. Interestingly, RIM proteins interact with voltage-dependent Ca(2+) channels (VDCCs) and suppress their activity at the presynaptic active zone to regulate neurotransmitter release. Knockdown of gamma-RIMs (*RIMS3* and *RIMS4*) attenuated glutamate release to a lesser extent than that of alpha-RIMs (*RIMS1* and *RIMS2*). As a consequence, competition between alpha- and gamma-RIMs seems to be essential for modulating the release of glutamate at the synapse (45). We can therefore hypothesize that the *de novo RIMS4* truncating stop variant perturbs the fine-tuning of glutamatergic release at the synapse and contributes to autism.

High inbreeding slightly increases the risk of autism, but no evidence for a founder effect for autism in the Faroe Islands.

In genetic isolates, it is frequent to observe an increase frequency of diseases due to the presence of deleterious variants that were present in the genomes of the small group of

298 migrants who settled the population. In the Faroe Islands, this "founder effect" was
299 documented for several genetic diseases such as Bardet-Biedl syndrome (46), cystic fibrosis
300 (47), 3-Methylcrotonyl-CoA carboxylase deficiency (48), glycogen storage disease type IIIA
301 (49) and retinitis pigmentosa (50). We confirmed that inbreeding in the Faroe Island is higher
302 than expected compared with other populations. The median of inbreeding is
303 $F = 0.015 \pm 0.001$ in the control sample and is similar to the one reported by Binzer et al.
304 (2014) in their study on multiple sclerosis in the Faroe Islands ($F = 0.018$) (51). This level of
305 inbreeding corresponds approximately to children from parents with a second-cousin
306 relationship ($F = 0.016$). We also observed that individuals with autism from the Faroe
307 Islands have significantly higher level of inbreeding and burden of recessive deleterious
308 variants compared to their geographically matched controls. The homozygous deleterious
309 variants carried by individuals with autism were enriched in genes included in our list of
310 genes of interest (*e. g.* high intolerance for loss of function variants, expressed in the brain
311 and previously associated with autism). However, one should note that the increased
312 probability for having autism due to inbreeding in the Faroe Islands is relatively small
313 ($F_{autism} = 0.0189$; $F_{controls} = 0.0148$; OR = 1.28; $p = 0.0047$).
314 In contrast to other genetic conditions, we could not detect a founder effect for autism
315 in the Faroe Islands. Moreover, the loci identified in our study do not overlap with those
316 detected in a previous genetic microsatellite association study in the Faroese population
317 pointing at regions on 2q, 3p, 6q, 15q, 16p, and 18q (52). We also found no overlap between
318 the variants identified in our study and those found in Faroese individuals with autism
319 diagnosed with panic (53) or bipolar (54) disorders. This absence of a founder effect is also in
320 agreement with the epidemiological observation that the prevalence of autism in the Faroese
321 population is not higher compared to more outbred populations.
322

Perspectives

Our study confirms that both rare and common genetic variants contribute to the susceptibility to autism. Indeed, although, we identified previously known genetic causes for autism and pointed at new compelling candidate genes, we also showed a contribution of the common variants illustrated by the higher level of PRS-autism in individuals with autism (especially those with no ID) compared to controls. To date, in the literature, very few genes are identified in individuals diagnosed with autism, but with intact general intelligence. Based on the genes previously reported (*NLGN3*, *NLGN4X*, duplication of *SHANK3*) (55,56) and the genes found in this study (*RIMS4*, *KALRN, PLA2G4A*), it seems that the proteins involved in autism without ID converge to different part of the post- and pre-synapse rather than to pathways such as gene regulation and chromatin remodeling, but this has to be confirmed on larger cohorts. Indeed, the main limitation of our study is the small number of individuals with autism. Several LGD variants affecting autism-risk genes such as *GRIK2* (57) or *ASMT* (58) were found exclusively or more frequently in individual with autism compared to controls, but we lack a replication samples to confirm the contribution of these variants in the susceptibility to autism in the Faroe Islands.

In summary, this study improves our knowledge on the genetic architecture of autism in epidemiological cohorts and in genetic isolates by showing that the contribution of both rare and common gene variants to autism can be detected in small, but genetically homogeneous populations. It also provides new compelling candidate genes and reveals that high inbreeding and high load of homozygous deleterious variants can be a risk factor for autism. Such combined analysis investigating both rare and common gene variants might represent a useful framework to investigate, from groups to individuals, the complex genetic architecture of autism.

Materials and Methods

Ethics statement

This study was approved by the IRB of the "Institut Pasteur" of Paris (IRB00006966 Institut Pasteur, approval 2010-003).

Patients

All individuals with autism in this study were recruited from an epidemiological cohort through a two-phase screening and diagnostic process targeting all children born in the 10-year period from 1985 through 1994 and living in the Faroe Islands in 2002 (7-16 years, n=7,689 children) and 2009 (15-24 years, n= 7,128 children) (8)·(10). The total number of children diagnosed with autism was 67 which corresponds to an autism prevalence of 0.94%. Among the individuals with autism, 23% were diagnosed with childhood autism, 56% with Asperger syndrome and 21% with atypical autism. There were 49 males (73.1%) and 18 females (26.9%). DNA was available for 36 individuals with autism including 11 diagnosed with childhood autism (31%), 17 with Asperger syndrome (47%), and 8 with atypical autism (22%). There were 28 males (78%) and 8 females (22%). The non-autism controls were recruited by issuing an invitation with information on the study to all pupils at the high school level in winter 2008-2009. The schools invited are in Eysturoy, Suduroy and Torshavn. The age of the invited was from 14-24 years. For those under 18 years a letter was sent to the parents that could sign the consent for their children.

Screening and diagnosis

In 2002, screening included the use of the Autism Spectrum Screening Questionnaire (ASSQ) (59). Screen-positive children were thoroughly examined via Diagnostic Interview for Social and Communication Disorder (DISCO-10 in 2002 and DISCO-11 in 2009) (60) of one or both

parents and the Wechsler Intelligence Scale for Children – 3ʳ edition (WISC) or Wechsler Adult Intelligence Scale – Revised (WAIS). Whenever overall and verbal abilities allowed it possible, children were also interviewed in an unstructured/semistructured manner about interests and skills patterns, peer relations, family relationships and about formal general information knowledge. The following diagnostic criteria used when making clinical diagnoses were (a) ICD-10 criteria for childhood autism/autistic disorder; (b) Gillberg criteria for Asperger syndrome; (c) ICD-10 criteria for atypical autism with the added requirement that a case thus diagnosed could not meet full criteria for childhood autism or Asperger syndrome; and (d) ICD-10 criteria for disintegrative disorder.

The majority of children in the atypical autism and Asperger syndrome groups had been tested with the WISC-R. Those with childhood autism had usually been tested on other tests. In those intellectually low-functioning individuals for whom no test was available, IQ was estimated on the basis of the Vineland developmental portion that is part of the DISCO-interview.

Genotyping

The cohort available for the genotyping is shown in Fig 1 and S1 Fig. It includes 36 individuals with autism, 208 controls, 132 close relatives of the individuals with autism (61 siblings and 68 parents) and 10 close relatives from the controls. DNA was extracted from blood leukocytes. The genotyping was performed at the "Centre National de Recherche en Génomique Humaine (CNRGH)" using the Infinium IlluminaOmni5-4 BeadChip (> 4.3 millions of markers) from Illumina. Sample quality controls such as Sex check (based on the X chromosome homozygosity rate), Mendel errors (transmission errors within full trios) and Identity By State (IBS, see section below) were performed using PLINK 1.90 (61).

398 **Population genetic structure**

399 Genome-wide pairwise IBS calculations and Multidimensional scaling (mds) analysis on
400 genome-wide IBS pairwise distances matrix was calculated using PLINK 1.90. IBS values
401 have been calculated for 376 individuals from Faroe Islands and 1,184 individuals from
402 HapMap phase 3 project with the following calculation: 1- (0.5 * IBS1 + IBS2)/N; N is the
403 number of tested markers; IBS1 and IBS2 are the number of markers for which one pair of
404 individuals share either 1 or 2 identical allele(s), respectively. Out of the 376 individuals, 32
405 individuals were removed from further analyses, including 7 ancestry outliers (all controls), 9
406 siblings of controls, one swap and 15 control individuals involved in pairs with IBS score
407 superior to 0.9.

408 For the estimation of the inbreeding status, SNPs with genotyping call rate < 95%,
409 minor allele frequency < 0.05, strong linkage disequilibrium r > 0.5 or failing Hardy
410 Weinberg equilibrium test ($p < 10^{-6}$) were filtered out of the Faroe SNP genotyping dataset.
411 All homozygosity analyses were performed with Plink 1.09 on autosomes including
412 identification of Runs Of Homozygosity (ROH) and Inbreeding coefficients calculation. For
413 ROH detection, a threshold of 50 consecutive homozygous SNPs with a minimum density of
414 1 SNP / 5,000 kb and no minimum length 50 SNPs was used following Howrigan et al.'s
415 guidelines (62). No heterozygous markers were allowed in the 50 SNPs-window. In this analysis,
416 the maximum gap between two consecutive SNPs within a run was set to 5,000 kb. Inbreeding
417 coefficients were calculated by estimating the proportion of the autosomal genome that is in
418 ROH. This method was proposed by McQuillan and al (2008)(63) and has been showed to be
419 the most reliable, especially with small sample size(64). Faroe inbreeding coefficients were
420 compared to inbreeding coefficient of HapMap phase 3 project populations.
421

Genome-wide association study (GWAS)

Prior association analyses, SNPs with genotyping call rate < 90%, minor allele frequency < 0.05 or failing Hardy Weinberg equilibrium test ($p < 10^{-6}$) were filtered out of the Faroe SNP genotyping dataset. The global genome wide genotyping call rate of all the individuals was superior to 90%. A total of 1,690,491 variants and 212 independent individuals (including 36 cases and 176 controls) passed filters and QC. Allelic, recessive and dominant GWAS were performed with Plink 1.09 using Chi-squared statistics. Manhattan and Quantile-Quantile (Q-Q) plots were generated using R. Gene and gene-set (including SFARI, pLI > 0.9 and Brain gene lists) analyses were performed with MAGMA v1.06 (29) using principal components regression and linear regression model, respectively.

Polygenic risk score (PRS) for autism

The computation of the PRS was performed with the tool PRSice2 (65) on the Faroes SNP array data using as a reference the PGC GWAS summary statistics (31). SNPs were not imputed since we used high density arrays (over 4 millions SNPs). For our dataset, PRSice2 with default parameters defined a p-value threshold of 0.197 which gives us a R^2 (squared correlation coefficient) of 0.04.

Whole-Exome Sequencing (WES*)*

Blood leukocytes DNA from 286 individuals was enriched for exonic sequences through hybridization SureSelect Human All Exon V5 (Agilent) by the CNRGH. For 67 individuals for whom the available quantity of DNA was low, they used a low-input protocol using only 200 ng of DNA compared to 3 µg for the normal protocol. The captured DNA was sequenced using a HiSeq 2000 instrument (Illumina). Coverage/depth statistics have been accessed as quality control criteria. We required that more than 90% of each exome have 10X coverage

447 and more than 80% have 20X coverage. Short read sequences were then aligned to hg19 with
448 BWA v0.7.8, duplicate reads were removed with PicardTools MarkDuplicates. Reads with a
449 global quality under 30 and a mapping quality under 20 were excluded from the analysis.
450 Variants were predicted using FreeBayes (66) and GATK (67) with a minimum of 10 reads
451 covering the position. VEP (using RefSeq and Ensembl 91) was used to annotate the variants.
452 We used the GEMINI (68) framework that automatically integrates the VCF file into a
453 database for exploring genetic variant for disease and population genetics. Genetic variants
454 were analyzed using GRAVITY, a Cytoscape (69) plugin designed in the lab specifically for
455 interpreting WES results using Protein-Protein Interaction networks
456 (http://gravity.pasteur.fr/). Gravity uses a user-friendly interface and makes easier prioritizing
457 variants according to damage prediction, mode of inheritance, gene categories and variant
458 frequency in databases. It allows filtering variants with many parameters, such as quality
459 parameters (DQ, MQ, GQ), allelic fraction, frequency of the variant in the cohort or in
460 databases, damage prediction scores (CADD, SIFT, Polyphen2) and many more. Since WES
461 does not detect the *FMR1* amplification, 33 individuals with autism were tested for Fragile-X
462 syndrome using the AmplideX™ *FMR1* PCR kit from Theradiag. There were no individuals
463 with "pre-mutation" or "full-mutation" of CGG repeats in the 5' UTR region of the fragile X
464 mental retardation-1 (*FMR1*) gene.
465
466 **Copy-number variants (CNVs*)***
467 CNVs were identified from both SNP genotyping and WES data. Quality controls were the
468 following: call rate > 0.99, standard deviation of the Log R ratio < 0.35, standard deviation of
469 the B allele frequency < 0.08 and absolute value of the wave factor < 0.05. CNVs were
470 detected by both PennCNV(70) and QuantiSNP(71) algorithm using the following
471 filters: >= 3 consecutive probes, CNV size > 1kb and CNV detection confidence score >= 15.

CNV detections from PennCNV and QuantiSNP were merged using CNVision(17). CNVs with CNVision confidence score < 30, CNV size < 50 kb, overlap > 50% with segmental duplication or known large assembly gaps (greater than 150 kb) or copy number = 2 in pseudo autosomal regions (PARS) in males were filtered out. CNV annotations were performed using ANNOVAR (72) and CNV frequencies in Faroese and in database of genomic variant cohorts (DGV, http://dgv.tcag.ca/dgv/app/home) were assessed using in house python scripts based on reciprocal overlap >= 80%. We also detected CNVs from the WES sequencing data using the XHMM software(73). CNVs with QSOME score < 90, number of targets < 5, or overlap > 50% with segmental duplication or known large assembly gaps (greater than 150 kb) were filtered out. CNV annotations were performed using ANNOVAR (72) and CNV frequencies in Faroese were assessed using in house python scripts based on reciprocal overlap ≥ 80%. *De novo* and inherited CNVs were validated by visual inspection using SnipPeep (http://snippeep.sourceforge.net/).

Gene-set lists and prioritization of variants

Three gene-set lists were used : (i) “SFARI genes” (n=990) that includes genes implicated in autism (15) (Simons Foundation Autism Research Initiative gene database - https://gene.sfari.org/); (ii) “pLI > 0.9 genes” that includes genes with strong probability of being loss-of function intolerant (n=3,230) (74); (iii) “Brain genes” that includes genes specifically or strongly expressed (above 1 Standard Deviation) in fetal or adult human brain using data from Su et al (n=3,591)(75).

A combination of approaches was used to prioritize the genes and to estimate the deleterious effect of a variant. We prioritized genes using gene sets (SFARI genes, pLI >= 0.9 and Brain genes, previously defined). We prioritized Likely Gene Disruptive (LGD) variants (stopgains, splice site variants, frameshift indels) over missense variants or synonymous

497 variants. Additionally, we used the CADD score (14) (a CADD >= 30 means that the variants
498 belong to the 0.1% most deleterious variants) to assess the deleterious effect missense
499 variants. Minor allele frequency (MAF) was estimated in the general population from the
500 gnomAD database(74). In order to filter out common variants that was not listed in gnomAD,
501 we also excluded variants that were present in more than 15% of our Faroese control cohort.
502 For the detection of deleterious homozygous variants, we kept only LGD and MIS30 with
503 MAF <1%.
504
505 **Burden analysis**
506 Rare variant association studies (MAF<5%) were performed using EPACTS v3.2.6
507 (https://genome.sph.umich.edu/wiki/EPACTS). Prior association analysis, variants identified
508 by WES were filtered using VCFtools (http://vcftools.sourceforge.net/man_latest.html) with
509 the following metrics: minimum genotyping quality ≥ 30, min depth of coverage ≥ 10,
510 maximum of missing data ≤ 10, no InDel (small insertion or deletion), only bi-allelic sites and
511 no site failing Hardy Weinberg equilibrium test ($p < 10^{-6}$). The annotation of the variants was
512 done using EPACTS and the variants included in the Gene-wise association analyses were
513 non-synonymous, essential splice site, normal splice site, start loss, stop loss and stop gain
514 variants. Logistic Score Test ("b.score" in S1 Table) was used to test single variant
515 association (n_{Cases}=36 ; $n_{Controls}$=107 ; $n_{Variants}$=155,284). For Gene-wise tests, we used two
516 approaches (including n_{Cases}=36 ; $n_{Controls}$=107 and n_{groups}= 15,005): (i) collapsing burden test
517 using EMMAX (Efficient Mixed Model Association eXpedited (76), "CMC-EMMAX" in S1
518 Table) and (ii) Optimal SNP-set sequence Kernel Association Test ("SKAT-O" in S1 Table).
519 The advantage of the CMC-EMMAX is that this test is accounting for population structure
520 and high relatedness between individual (based on kinship matrix). The advantage of SKAT is

521 that this test is particularly powerful in the presence of protective and deleterious variants and

522 null variants. For both Gene-wise tests, a $10^{-6} \leq MAF \leq 0.05$ was used.

523

Acknowledgements

We would like to thank all the participants of the epidemiological and genetic cohorts, the "progeny" database, Julien Fumey for his R tutorial (https://bioinfo-fr.net/creer-sa-carte-geographique-avec-r) and Fabrice de Chaumont for his help with python.

References

1. Gillberg C. The ESSENCE in child psychiatry: Early Symptomatic Syndromes Eliciting Neurodevelopmental Clinical Examinations. Res Dev Disabil. 2010 Dec;31:1543–51.

2. Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat Rev Neurosci. 2015 Sep;16(9):551–63.

3. Mamidala MP, Kalikiri MK, Praveen Kumar PTV, Rajesh N, Vallamkonda OR, Rajesh V. Consanguinity in India and its association with autism spectrum disorder. Autism Res Off J Int Soc Autism Res. 2015 Apr;8(2):224–8.

4. Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS, et al. Identifying autism loci and genes by tracing recent shared ancestry. Science. 2008 Jul 11;321:218–23.

5. Jorgensen TH, Degn B, Wang AG, Vang M, Gurling H, Kalsi G, et al. Linkage disequilibrium and demographic history of the isolated population of the Faroe Islands. Eur J Hum Genet. 2002 Jun;10:381–7.

6. Jorgensen TH, Buttenschon HN, Wang AG, Als TD, Borglum AD, Ewald H. The origin of the isolated population of the Faroe Islands investigated using Y chromosomal markers. Hum Genet. 2004 Jun;115:19–28.

7. Kocovska E, Biskupsto R, Carina Gillberg I, Ellefsen A, Kampmann H, Stora T, et al. The rising prevalence of autism: a prospective longitudinal study in the Faroe Islands. J Autism Dev Disord. 2012 Sep;42:1959–66.

8. Ellefsen A, Kampmann H, Billstedt E, Gillberg IC, Gillberg C. Autism in the Faroe Islands: an epidemiological study. J Autism Dev Disord. 2007 Mar;37:437–44.

9. Kocovska E, Billstedt E, Ellefsen A, Kampmann H, Gillberg IC, Biskupsto R, et al. Autism in the Faroe Islands: diagnostic stability from childhood to early adult life. ScientificWorldJournal. 2013;2013:592371.

10. Kočovská E, Biskupstø R, Carina Gillberg I, Ellefsen A, Kampmann H, Stórá T, et al. The rising prevalence of autism: a prospective longitudinal study in the Faroe Islands. J Autism Dev Disord. 2012 Sep;42(9):1959–66.

11. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014 Nov 13;515:216–21.

12. Turner TN, Yi Q, Krumm N, Huddleston J, Hoekzema K, F. Stessman HA, et al. denovo-db: a compendium of human de novo variants. Nucleic Acids Res. 2017 Jan 4;45(Database issue):D804–11.

13. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012 May 10;485:242–5.

14. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014 Mar;46:310–5.

15. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism. 2013;4:36.

16. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 18;536(7616):285–91.

17. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015 Sep 23;87(6):1215–33.

18. Hempel M, Cremer K, Ockeloen CW, Lichtenbelt KD, Herkert JC, Denecke J, et al. De Novo Mutations in CHAMP1 Cause Intellectual Disability with Severe Speech Impairment. Am J Hum Genet. 2015 Sep 3;97(3):493–500.

19. Martin EA, Muralidhar S, Wang Z, Cervantes DC, Basu R, Taylor MR, et al. The intellectual disability gene Kirrel3 regulates target-specific mossy fiber synapse development in the hippocampus. eLife. 2015 Nov 17;4:e09395.

20. Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, et al. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell. 2012 Apr 27;149(3):525–37.

21. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012 Apr 4;485:237–41.

22. Kalsner L, Twachtman-Bassett J, Tokarski K, Stanley C, Dumont-Mathieu T, Cotney J, et al. Genetic testing including targeted gene panel in a diverse clinical population of children with autism spectrum disorder: Findings and implications. Mol Genet Genomic Med. 2017 Dec 21;

23. Verhoeven K, Van Laer L, Kirschhofer K, Legan PK, Hughes DC, Schatteman I, et al. Mutations in the human alpha-tectorin gene cause autosomal dominant non-syndromic hearing impairment. Nat Genet. 1998 May;19(1):60–2.

24. Smith CEL, Poulter JA, Levin AV, Capasso JE, Price S, Ben-Yosef T, et al. Spectrum of PEX1 and PEX6 variants in Heimler syndrome. Eur J Hum Genet EJHG. 2016;24(11):1565–71.

25. Wang Q, Charych EI, Pulito VL, Lee JB, Graziane NM, Crozier RA, et al. The psychiatric disease risk factors DISC1 and TNIK interact to regulate synapse composition and function. Mol Psychiatry. 2011 Oct;16(10):1006–23.

26. Bourgeron T. A synaptic trek to autism. Curr Opin Neurobiol. 2009 Apr;19(2):231–4.

603 27. Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O, et al. Disruptive CHD8
604 Mutations Define a Subtype of Autism Early in Development. Cell. 2014 Jul 3;

605 28. Gallagher L, Becker K, Kearney G, Dunlop A, Stallings R, Green A, et al. Brief report:
606 A case of autism associated with del(2)(q32.1q32.2) or (q32.2q32.3). J Autism Dev
607 Disord. 2003 Feb;33(1):105–8.

608 29. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set
609 analysis of GWAS data. PLoS Comput Biol. 2015 Apr;11(4):e1004219.

610 30. Campellone KG, Webb NJ, Znameroski EA, Welch MD. WHAMM Is an Arp2/3
611 Complex Activator That Binds Microtubules and Functions in ER to Golgi Transport.
612 Cell. 2008 Jul 11;134(1):148–61.

613 31. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium.
614 Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder
615 highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol
616 Autism. 2017;8:21.

617 32. Warner G, Moss J, Smith P, Howlin P. Autism characteristics and behavioural
618 disturbances in ~ 500 children with Down's syndrome in England and Wales. Autism
619 Res Off J Int Soc Autism Res. 2014 Aug;7(4):433–41.

620 33. Vorstman JA, Morcus ME, Duijff SN, Klaassen PW, Heineman-de Boer JA, Beemer
621 FA, et al. The 22q11.2 deletion in children: high rate of autistic disorders and early onset
622 of psychotic symptoms. J Am Acad Child Adolesc Psychiatry. 2006 Sep;45:1104–13.

623 34. Niklasson L, Rasmussen P, Oskarsdottir S, Gillberg C. Autism, ADHD, mental
624 retardation and behavior problems in 100 individuals with 22q11 deletion syndrome. Res
625 Dev Disabil. 2009 Aug;30:763–73.

626 35. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic,
627 transcriptional and chromatin genes disrupted in autism. Nature. 2014 Nov 13;515:209–
628 15.

629 36. Yan Y, Eipper BA, Mains RE. Kalirin-9 and Kalirin-12 Play Essential Roles in
630 Dendritic Outgrowth and Branching. Cereb Cortex N Y N 1991. 2015 Oct;25(10):3487–
631 501.

632 37. Russell TA, Grubisha MJ, Remmers CL, Kang SK, Forrest MP, Smith KR, et al. A
633 Schizophrenia-Linked KALRN Coding Variant Alters Neuron Morphology, Protein
634 Function, and Transcript Stability. Biol Psychiatry. 2018 Mar 15;83(6):499–508.

635 38. Hill JJ, Hashimoto T, Lewis DA. Molecular mechanisms contributing to dendritic spine
636 alterations in the prefrontal cortex of subjects with schizophrenia. Mol Psychiatry.
637 2006;11(6):557–66.

638 39. Qu B-X, Gong Y, Sinclair D, Fu M, Perl D, Diaz-Arrastia R. cPLA2α knockout mice
639 exhibit abnormalities in the architecture and synapses of cortical neurons. Brain Res.
640 2013 Feb 25;1497:101–5.

641 40. Su L-D, Wang D-J, Yang D, Shen Y, Hu Y-H. Retrograde cPLA2α/arachidonic acid/2-
642 AG signaling is essential for cerebellar depolarization-induced suppression of excitation
643 and long-term potentiation. Cerebellum Lond Engl. 2013 Jun;12(3):297–9.

644 41. Alvarez-Baron E, Michel K, Mittelstaedt T, Opitz T, Schmitz F, Beck H, et al. RIM3γ
645 and RIM4γ are key regulators of neuronal arborization. J Neurosci Off J Soc Neurosci.
646 2013 Jan 9;33(2):824–39.

647 42. Dong S, Walker MF, Carriero NJ, DiCola M, Willsey AJ, Ye AY, et al. De novo
648 insertions and deletions of predominantly paternal origin are associated with autism
649 spectrum disorder. Cell Rep. 2014 Oct 9;9(1):16–23.

650 43. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene
651 disruptions in children on the autistic spectrum. Neuron. 2012 Apr 26;74(2):285–99.

652 44. Kumar RA, Sudi J, Babatz TD, Brune CW, Oswald D, Yen M, et al. A de novo 1p34.2
653 microdeletion identifies the synaptic vesicle gene RIMS3 as a novel candidate for
654 autism. J Med Genet. 2010 Feb;47(2):81–90.

655 45. Uriu Y, Kiyonaka S, Miki T, Yagi M, Akiyama S, Mori E, et al. Rab3-interacting
656 Molecule γ Isoforms Lacking the Rab3-binding Domain Induce Long Lasting Currents
657 but Block Neurotransmitter Vesicle Anchoring in Voltage-dependent P/Q-type Ca^{2+}
658 Channels. J Biol Chem. 2010 Jul 9;285(28):21750–67.

659 46. Hjortshøj TD, Grønskov K, Brøndum-Nielsen K, Rosenberg T. A novel founder BBS1
660 mutation explains a unique high prevalence of Bardet-Biedl syndrome in the Faroe
661 Islands. Br J Ophthalmol. 2009 Mar;93(3):409–13.

662 47. Schwartz M, Sørensen N, Brandt NJ, Høgdall E, Holm T. High incidence of cystic
663 fibrosis on the Faroe Islands: a molecular and genealogical study. Hum Genet. 1995
664 Jun;95(6):703–6.

665 48. Thomsen JA, Lund AM, Olesen JH, Mohr M, Rasmussen J. Is L-Carnitine
666 Supplementation Beneficial in 3-Methylcrotonyl-CoA Carboxylase Deficiency? JIMD
667 Rep. 2015;21:79–88.

668 49. Santer R, Kinner M, Steuerwald U, Kjaergaard S, Skovby F, Simonsen H, et al.
669 Molecular genetic basis and prevalence of glycogen storage disease type IIIA in the
670 Faroe Islands. Eur J Hum Genet EJHG. 2001 May;9(5):388–91.

671 50. Ostergaard E, Duno M, Batbayli M, Vilhelmsen K, Rosenberg T. A novel MERTK
672 deletion is a common founder mutation in the Faroe Islands and is responsible for a high
673 proportion of retinitis pigmentosa cases. Mol Vis. 2011;17:1485–92.

674 51. Binzer S, Imrell K, Binzer M, Kyvik KO, Hillert J, Stenager E. High inbreeding in the
675 Faroe Islands does not appear to constitute a risk factor for multiple sclerosis. Mult Scler
676 Houndmills Basingstoke Engl. 2015 Jul;21(8):996–1002.

677 52. Lauritsen MB, Als TD, Dahl HA, Flint TJ, Wang AG, Vang M, et al. A genome-wide
678 search for alleles and haplotypes associated with autism and related pervasive
679 developmental disorders on the Faroe Islands. Mol Psychiatry. 2006 Jan;11:37–46.

680 53. Gregersen NO, Lescai F, Liang J, Li Q, Als T, Buttenschøn HN, et al. Whole-exome
681 sequencing implicates DGKH as a risk gene for panic disorder in the Faroese population.
682 Am J Med Genet B Neuropsychiatr Genet. 2016 Dec 1;171(8):1013–22.

683 54. Ewald H, Flint TJ, Jorgensen TH, Wang AG, Jensen P, Vang M, et al. Search for a
684 shared segment on chromosome 10q26 in patients with bipolar affective disorder or
685 schizophrenia from the Faroe Islands. Am J Med Genet. 2002 Mar 8;114:196–204.

686 55. Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, et al. Mutations
687 of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with
688 autism. Nat Genet. 2003 May;34:27–9.

689 56. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, et al.
690 Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated
691 with autism spectrum disorders. Nat Genet. 2007 Jan;39:25–7.

692 57. Jamain S, Betancur C, Quach H, Philippe A, Fellous M, Giros B, et al. Linkage and
693 association of the glutamate receptor 6 gene with autism. Mol Psychiatry. 2002;7:302–
694 10.

695 58. Melke J, Goubran Botros H, Chaste P, Betancur C, Nygren G, Anckarsater H, et al.
696 Abnormal melatonin synthesis in autism spectrum disorders. Mol Psychiatry. 2008
697 Jan;13:90–8.

698 59. Ehlers S, Gillberg C, Wing L. A screening questionnaire for Asperger syndrome and
699 other high-functioning autism spectrum disorders in school age children. J Autism Dev
700 Disord. 1999 Apr;29(2):129–41.

701 60. Wing L, Leekam SR, Libby SJ, Gould J, Larcombe M. The Diagnostic Interview for
702 Social and Communication Disorders: background, inter-rater reliability and clinical use.
703 J Child Psychol Psychiatry. 2002 Mar;43(3):307–25.

704 61. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK:
705 A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.
706 Am J Hum Genet. 2007 Sep;81(3):559–75.

707 62. Howrigan DP, Simonson MA, Keller MC. Detecting autozygosity through runs of
708 homozygosity: a comparison of three autozygosity detection algorithms. BMC
709 Genomics. 2011 Sep 23;12:460.

710 63. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc
711 L, et al. Runs of homozygosity in European populations. Am J Hum Genet. 2008
712 Sep;83(3):359–72.

713 64. Gazal S, Sahbatou M, Perdry H, Letort S, Génin E, Leutenegger A-L. Inbreeding
714 coefficient estimation with dense SNP data: comparison of strategies and application to
715 HapMap III. Hum Hered. 2014;77(1–4):49–62.

716 65. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinforma
717 Oxf Engl. 2015 May 1;31(9):1466–8.

718 66. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing.
719 ArXiv12073907 Q-Bio [Internet]. 2012 Jul 17 [cited 2017 Mar 15]; Available from:
720 http://arxiv.org/abs/1207.3907

721 67. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
722 Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA
723 sequencing data. Genome Res. 2010 Sep;20(9):1297–303.

724 68. DeFreitas T, Saddiki H, Flaherty P. GEMINI: a computationally-efficient search engine
725 for large gene expression datasets. BMC Bioinformatics. 2016;17:102.

726 69. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A
727 Software Environment for Integrated Models of Biomolecular Interaction Networks.
728 Genome Res. 2003 Jan 11;13(11):2498–504.

729 70. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. PennCNV: an integrated
730 hidden Markov model designed for high-resolution copy number variation detection in
731 whole-genome SNP genotyping data. Genome Res. 2007 Nov;17:1665–74.

732 71. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, et al. QuantiSNP: an
733 Objective Bayes Hidden-Markov Model to detect and accurately map copy number
734 variation using SNP genotyping data. Nucleic Acids Res. 2007;35:2013–25.

735 72. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants
736 from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep;38:e164.

737 73. Fromer M, Purcell SM. Using XHMM Software to Detect Copy Number Variation in
738 Whole-Exome Sequencing Data. Curr Protoc Hum Genet. 2014 Apr 24;81:7.23.1-21.

739 74. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis
740 of protein-coding genetic variation in 60,706 humans. Nature. 2016 18;536(7616):285–
741 91.

742 75. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the
743 mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A. 2004
744 Apr 20;101(16):6062–7.

745 76. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-Y, Freimer NB, et al. Variance
746 component model to account for sample structure in genome-wide association studies.
747 Nat Genet. 2010 Apr;42(4):348–54.

748

749

Funding

This work was supported by the Institut Pasteur; Centre National de la Recherche Scientifique; the Assistance Publique – Hôpitaux de Paris; the University Paris Diderot; the Simons Foundation; the Fondation pour la Recherche Médicale [DBI20141231310]; the European Commission Horizon 2020 [COSYN]; The human brain project; the European Commission Innovative Medicines Initiative [EU-AIMS no. 115300]; the Cognacq-Jay foundation; the Bettencourt-Schueller foundation; the Orange foundation; the FondaMental foundation; the Conny-Maeva foundation; and the Agence Nationale de la Recherche (ANR) [SynPathy]. This research was supported by the Laboratory of Excellence GENMED (Medical Genomics) grant no. ANR-10-LABX-0013, Bio-Psy and by the INCEPTION program ANR-16-CONV-0005, all managed by the ANR part of the Investment for the Future program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Figure Captions

Fig 1. Genetic background of the Faroese population. A. Geographic localization of the Faroe Islands (R software and "maps" and "mapdata" packages were used to draw the map). B. Demographic comparison of the epidemiological and genetic cohorts of the Faroe Islands. The epidemiological and the genetic cohort are composed of 67 and 36 individuals with autism, respectively. C. Multidimensional scaling plots (MDS) of genome-wide identity by state (IBS) pairwise distances between human populations. Each dot represents an individual and the distance between two dots corresponds to genetic distance based on genome-wide pairwise IBS calculations. D. Degree of inbreeding across HapMap 3 and Faroese populations. The inbreeding coefficients of the Faroe non-autism control individuals (n=176) were compared to the HapMap3 populations. Faroe individuals displayed a higher degree of inbreeding compared with others human populations (Median F: $F_{FAROES} = 0.015$; $F_{ASW} = 0.0014$; $F_{CEU} = 0.0071$; $F_{CHD} = 0.0081$; $F_{CHB} = 0.0083$; $F_{GIH} = 0.011$; $F_{JPT} = 0.0093$; $F_{LWK} = 0.0045$; $F_{MXL} = 0.0095$; $F_{MKK} = 0.0026$; $F_{TSI} = 0.0066$; $F_{YRI} = 0.0028$; Paired samples Wilcoxon test: $W_{ASW} = 12.56$, $p_{ASW} < 0.0001$; $W_{CEU} = 12.27$, $p_{CEU} < 0.0001$; $W_{CHD} = 7.14$, $p_{CHD} < 0.0001$; $W_{CHB} = 7.11$, $p_{CHB} < 0.0001$; $W_{GIH} = -3.06$, $p_{GIH} < 0.002$;$W_{JPT} = -5.83$, $p_{JPT} < 0.0001$; $W_{LWK} = -10.86$, $p_{LWK} < 0.0001$; $W_{MXL} = -5.66$, $p_{MXL} < 0.0001$; $W_{MKK} = -14.52$, $p_{MKK} < 0.0001$; $W_{TSI} = -8.64$, $p_{TSI} < 0.0001$; $W_{YRI} = -14.41$, $p_{YRI} < 0.0001$; p are nominal p-values, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$). N, number; ID, intellectual disability; ASW, African ancestry in Southwest USA (n=83); CEU, Utah residents with Northern and Western European ancestry from the CEPH collection (n=165); CHB, Han Chinese in Beijing, China (n=84); CHD, Chinese in Metropolitan Denver, Colorado (n=85); GIH, Gujarati Indians in Houston, Texas (n=88); JPT, Japanese in Tokyo, Japan (n=86); LWK, Luhya in Webuye, Kenya (n=90); MXL, Mexican ancestry in Los Angeles (n=77), California; MKK, Maasai in

788 Kinyawa, Kenya (n=171); TSI, Toscani in Italia (n=88); YRI, Yoruba in Ibadan, Nigeria
789 (n=167).
790
791 **Fig 2. Rare and common variants in Faroese individuals with autism.** A. Copy-number
792 variant (CNV) analysis among gene-set lists within Faroe individuals. The number of exonic
793 CNV carriers altering any genes or gene-set lists (SFARI genes, pLI > 0.9 genes and Brain
794 genes, see Materials and Methods section) were compared between individuals with autism,
795 siblings and controls (Fisher's exact test: $n_{autism} = 36$, $n_{controls} = 107$, $p_{CNV\_loss\_All\_genes} = 0.02$,
796 $OR_{CNV\_loss\_All\_genes} = 2.79$; $p_{CNV\_loss\_SFARI} = 0.014$, $OR_{CNV\_loss\_All\_SFARI} = 13.25$;
797 $p_{CNV\_loss\_pLI>0.9} = 0.014$, $OR_{CNV\_loss\_pLI>0.9} = 13.25$; $p_{CNV\_loss\_Brain} = 0.024$,
798 $OR_{CNV\_loss\_All\_Brain} = 5.59$; $p_{CNV\_gain\_All\_genes} = 0.0056$, $OR_{CNV\_gain\_All\_genes} = 3.28$;
799 $p_{CNV\_gain\_SFARI} = 0.067$, $OR_{CNV\_gain\_All\_SFARI} = 4.33$; $p_{CNV\_gain\_pLI>0.9} = 0.11$,
800 $OR_{CNV\_gain\_pLI>0.9} = 2.47$; $p_{CNV\_gain\_Brain} = 0.0049$, $OR_{CNV\_gain\_All\_Brain} = 4.76$; p are nominal p-
801 values). B. Heatmap combining signals obtained from rare and common variant association
802 tests and rare deleterious variants altering SFARI genes throughout Faroese individuals with
803 autism. "Burden of non-synonymous" includes results from SKAT-O and CMC-EMMAX
804 obtained from WES data (see Materials and Methods section and S5 fig; $p < 10^{\wedge-3}$). "Common
805 variants" are the top hits of the genome wide association study (GWAS) for both allelic and
806 recessive model obtained from genome-wide genotyping data ($p < 10^{\wedge-6}$). "Rare LGD SFARI
807 variants" are rare likely gene disrupting (LGD) variants altering SFARI genes identified by
808 (MAF < 1% in gnomAD). "The controls freq." column indicates the proportion of non-ASD
809 Faroese controls carrying the corresponding variant. P-Values are nominal. ID, intellectual
810 disability.
811
812 **Fig 3. Genetic recessive mutations in Faroese individuals with autism**. A. Distribution of
813 the inbreeding coefficient in Faroese individuals (Mann Whitney U-test: $n_{autisms} = 36$,

814 $n_{controls} = 176$, $n_{siblings} = 30$; $U_{controls.vs.autisms} = 2{,}297$, $p_{controls.vs.autisms} = 0.0047$;
815 $U_{controls.vs.siblings} = 1{,}953$, $p_{controls.vs.\ siblings} = 0.011$; * indicates the one withstanding multiple
816 testing). B. Number of rare LGD+MIS30 homozygous mutations carried per individual (Mann
817 Whitney U-test: $n_{autisms} = 36$, $n_{controls} = 107$, $n_{siblings} = 30$; $U_{controls.vs.autisms} = 1{,}321$,
818 $p_{controls.vs.autisms} = 0.00049$; $U_{controls.vs.siblings} = 1{,}293$, $p_{controls.vs.\ siblings} = 0.025$; * indicates the one
819 withstanding multiple testing). C. Venn diagram of the genes carrying the variants from B.
820 Genes names are in bold and annotated when they are part of our gene-set lists (SFARI genes,
821 pLI > 0.9 genes and Brain genes, see subject and methods section). The plot on the right shows
822 the proportion of individuals in each category carrying at least one mutated gene in our gene-
823 sets lists (Fisher's exact test: $p_{controls.vs.autisms} = 0.03$; $p_{controls.vs.\ siblings} = 0.03$). D. and E. are
824 describing two specific families carrying multiple variants. "0" and "1" refer to wildtype or
825 mutated allele, respectively. The localizations of the variants are indicated along the proteins
826 and alignments throughout species showed the strong conservation of the altered amino acids.
827 LGD, likely gene disruptive; MIS30, missense variants with CADD score ≥ 30; IgD,
828 immunoglobulin domain; TIL, Trypsin Inhibitor-like; FA5/8C, Coagulation factor 5/8 type C
829 domain; LamG, Laminin G domain; EGF, epidermal growth factor like domain; Fibr.,
830 Fibrinogen, alpha/beta/gamma chain, C-terminal globular domain; AAA: ATPases associated
831 domains.
832
833 **Fig 4. Distribution of the polygenic risk score for autism in Faroese individuals**. A.
834 Distribution of the polygenic risk score for autism (PRS-autism) of controls, autisms and
835 siblings (Mann Whitney U-test: $n_{autisms} = 36$, $n_{controls} = 107$, $n_{siblings} = 53$;
836 $U_{controls.vs.autisms} = 2{,}344$, $p_{controls.vs.autisms} = 0.0070$; * indicates the one withstanding multiple
837 testing). B. Distribution of the PRS-autism for the cases without intellectual disability (ID)
838 and the cases with ID (Mann Whitney U-test: $n_{autisms\text{-}with\text{-}ID} = 12$, $n_{autisms\text{-}without\text{-}ID} = 24$; $U_{ID.vs.no\text{-}}$

839 $_{ID}$ = 86, $p_{ID.vs.no-ID}$ = 0.027; * indicates the one withstanding multiple testing). The PRS was
840 calculated using PRSice-2 (see Materials and Methods section).
841
842 **Fig 5. Stratification of autism in Faroese individuals**. On the left, the stratification was built
843 using hierarchical clustering on the number of genes carrying rare deleterious variants altering
844 SFARI genes (MIS30, LGD or CNV) and on the polygenic risk score for autism (PRS-autism).
845 The other columns were not used for the clustering. The genetic profile contains variants with
846 a predicted impact on the condition of the individual with autism, the one in bold are most likely
847 causatives. The clinical profile gives a subset of relevant information for each individual with
848 autism. ID, intellectual disability; M, male; F, female; del, deletion; dup, duplication.
849
850

Supporting information

S1 Fig 1. Pipeline of the study. QC, quality control; SD LRR, standard deviation of the Log R ratio; SD BAF, standard deviation of the B allele frequency; |WF|, absolute value of the wave factor; BWA, Burrows-Wheeler Aligner; GATK, Genome Analysis Toolkit; VEP, Variant Effect Predictor; GQ, genotyping quality; Htz_R, heterozygosity ratio; CNV, copy number variant; SNP/V, single nucleotide polymorphism/variation; WES, whole exome sequencing; Hmz, homozygote; Htz, heterozygote; pLI, probability of being loss-of function intolerant.

S2 Fig. Population stratification of the Faroe Islands. Combination of the ethnic genetic background for each individuals of the Faroe and the individuals from HapMap3 using admixture. Legend of HapMap 3 population: Africa (ASW, African ancestry in Southwest USA (n=83); LWK, Luhya in Webuye, Kenya (n=90); MKK, Maasai in Kinyawa, Kenya (n=171); YRI, Yoruba in Ibadan, Nigeria (n=167)); Asia (CHB, Han Chinese in Beijing, China (n=84); CHD, Chinese in Metropolitan Denver, Colorado (n=85); JPT, Japanese in Tokyo, Japan (n=86)); TSI, Toscani in Italia (n=88); CEU, Utah residents with Northern and Western European ancestry from the CEPH collection (n=165); GIH, Gujarati Indians in Houston, Texas (n=88); MXL, Mexican ancestry in Los Angeles (n=77), California.

S3 Fig. CNVs altering genes involved in neurodevelopmental disorders. A. *De novo* Trisomy 21 in patient with autism, ID and Down syndrome. B. *De novo* deletion of 425.5 kb removing the six first exons of the *NRXN1α* in individual with autism. C. *De novo* 2.9 Mb deletion on chromosome 22q11 in individual with autism and DiGeorge/velocardial syndrome. D. A 91.4 kb deletion removing all exons of *ADNP* in a male with autism and ID. Each dot shows Log R Ratio (LRR; in red) and B allele frequency (BAF; in green). The copy

876 number (CN) is indicated with a blue line. Patients with ID and patients without ID are
877 represented in black and grey, respectively. ID, Intellectual disability.
878
879 **S4 Fig. The *de novo* SNVs in individuals with autism.** Sanger sequencing was performed to
880 validate *de novo* SNVs altering *MECP2*, *KIF17* (A), *PLA2G4A* (B), *RIMS4* (C) and *KALRN*
881 (D). Sanger chromatograms are shown for each trios. Individuals with ID and without ID are
882 represented in black and grey, respectively. The position of the SNVs is indicated on the
883 protein and the amino acid alignment of the region throughout several species (Human,
884 Rhesus, Mouse, Dog, Zebrafish) shows the high conservation of the altered amino acids; ID,
885 Intellectual disability.
886
887 **S5 Fig. Gene-wise association study using the whole exome sequencing data.** To test for
888 Gene-wise association, a collapsing burden test using EMMAX "CMC-EMMAX" (A) and
889 optimal SNP-set sequence Kernel Association Test "SKAT-O" (B) were used. The dashed
890 line indicates p-value < $10^{\wedge-3}$. EMMAX, Efficient Mixed Model Association eXpedited.
891
892 **S6 Fig. Results from the Genome Wide Association Study (GWAS) using different**
893 **models (allelic, Recessive or Dominant).** QQ plots and Manhattan plots for allelic, recessive
894 and dominant GWAS are represented in panel A and B, respectively. QQ, quantile-quantile.
895
896 **S7 Fig. Locus zoom of the top hits detected by the GWAS.** LocusZoom for regional
897 visualization of the top hits isolated from the allelic and recessive GWAS was used.
898
899 **S1 Appendix. Clinical notes**
900

S1 Table: *de novo* CNVs

S2 Table: *de novo* SNVs

S3 Table: CNVs within SFARI, pLI > 0.9 an Brain genes

S4 Table: CMC-emmax gene-wise association test from WES

S5 Table: SKAT-O gene-wise association test from WES

S6 Table: b.score single variant association test from WES

S7 Table: rare LGD or MIS30 SNV within SFARI, pLI and Brain genes (< 1% gnomAD)

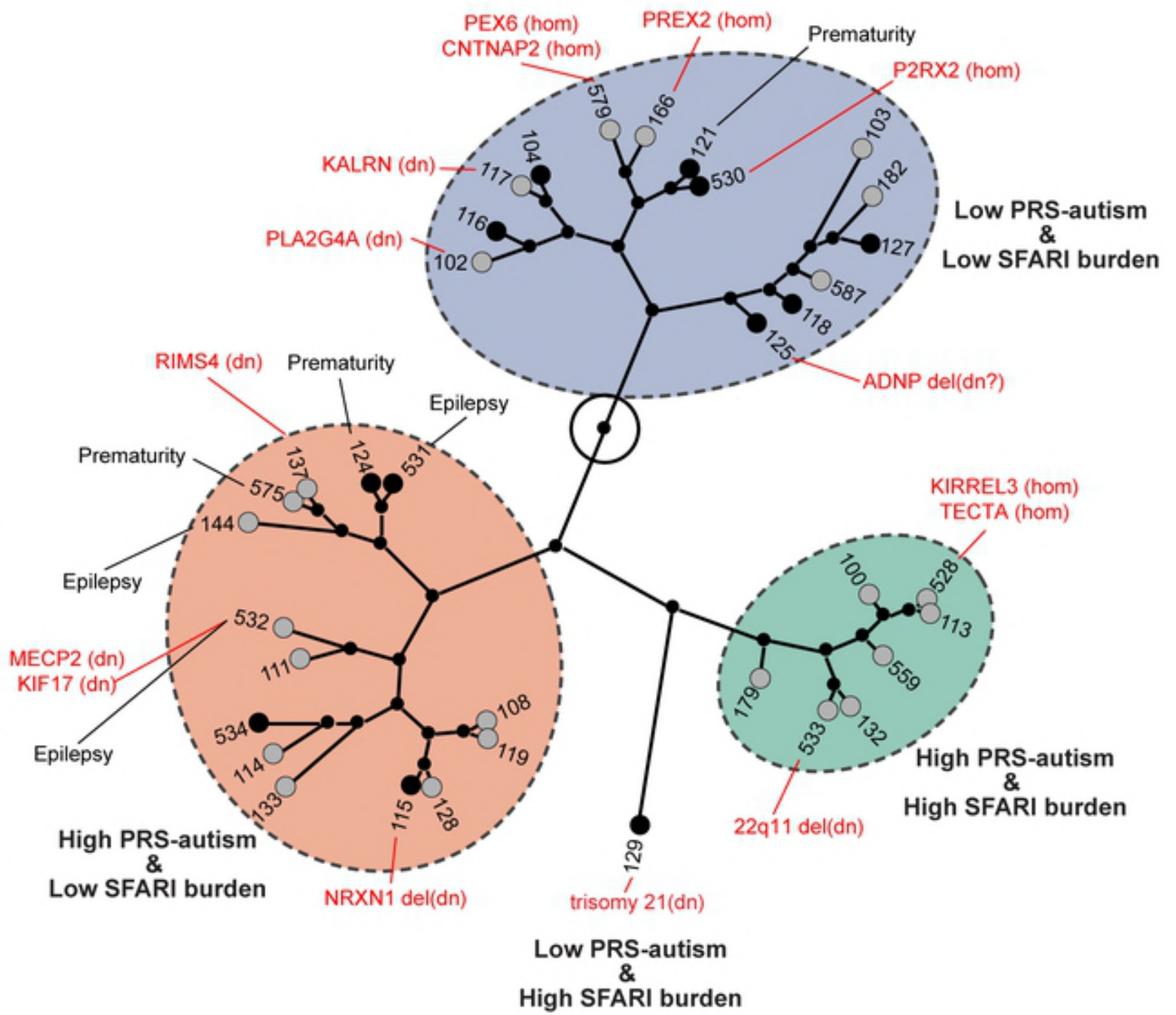
S8 Table: recessive SNVs

S9 Table: GWAS from SNP genotyping data

S10 Table: MAGMA Gene-based association study from SNP genotyping data

S11 Table: MAGMA Gene-set-based association study from SNP genotyping data

Patient	ID	Sex	PRS-autism	SFARI
PN400100	○	M		
PN400528	○	F		
PN400113	○	M		
PN400559	○	M		
PN400132	○	M		
PN400533	○	F		
PN400179	○	M		
PN400129	●	M		
PN400108	○	M		
PN400119	○	F		
PN400128	○	M		
PN400115	●	M		
PN400133	○	M		
PN400114	○	M		
PN400534	●	F		
PN400111	○	M		
PN400532	○	F		
PN400144	○	M		
PN400575	○	F		
PN400137	○	M		
PN400124	●	M		
PN400531	●	F		
PN400102	○	M		
PN400116	●	M		
PN400117	○	M		
PN400104	●	M		
PN400579	○	M		
PN400166	○	M		
PN400121	●	F		
PN400530	●	M		
PN400103	○	M		
PN400182	○	M		
PN400127	●	M		
PN400587	○	M		
PN400118	●	M		
PN400125	●	M		

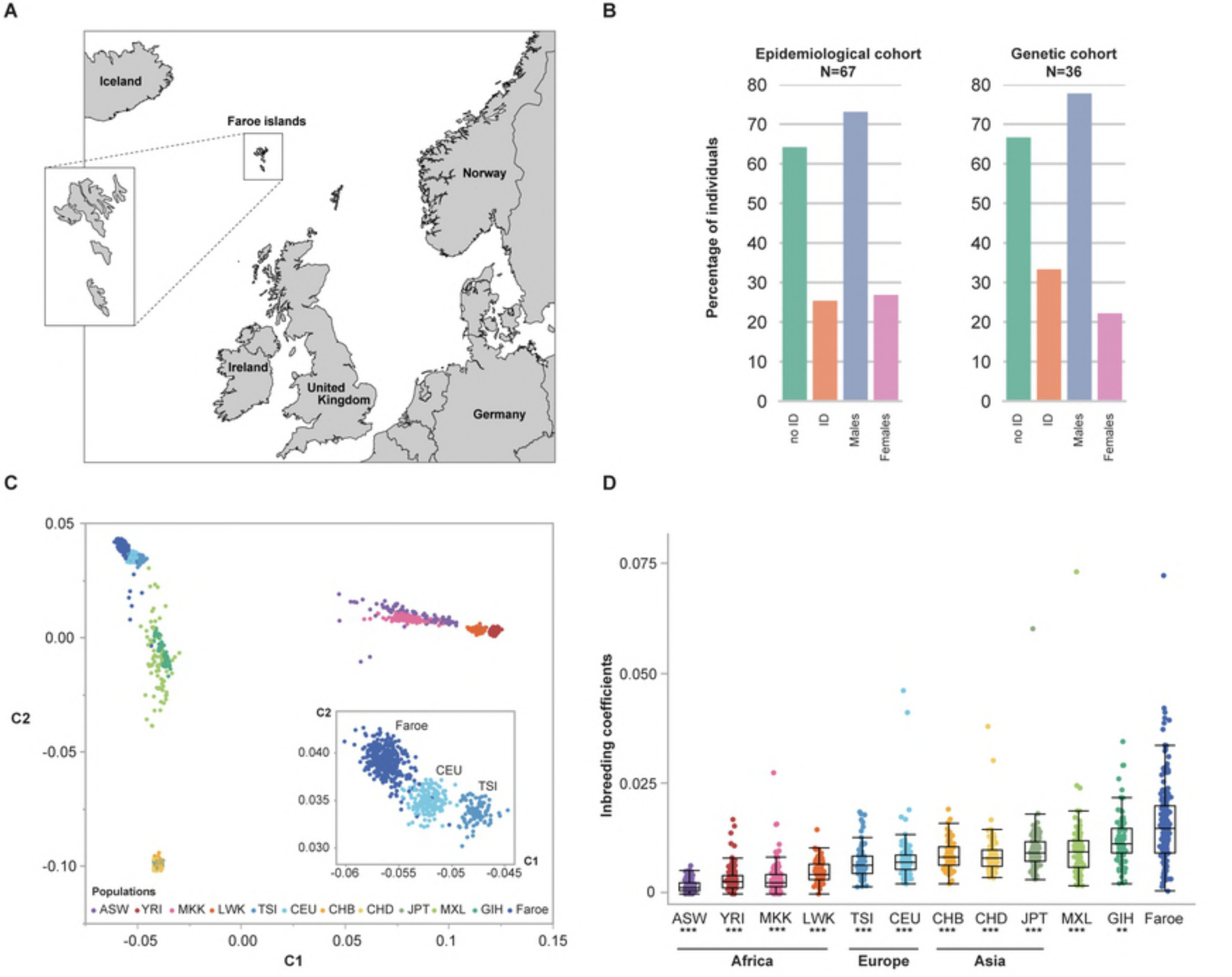


— Clinical profile
— Genetic profile

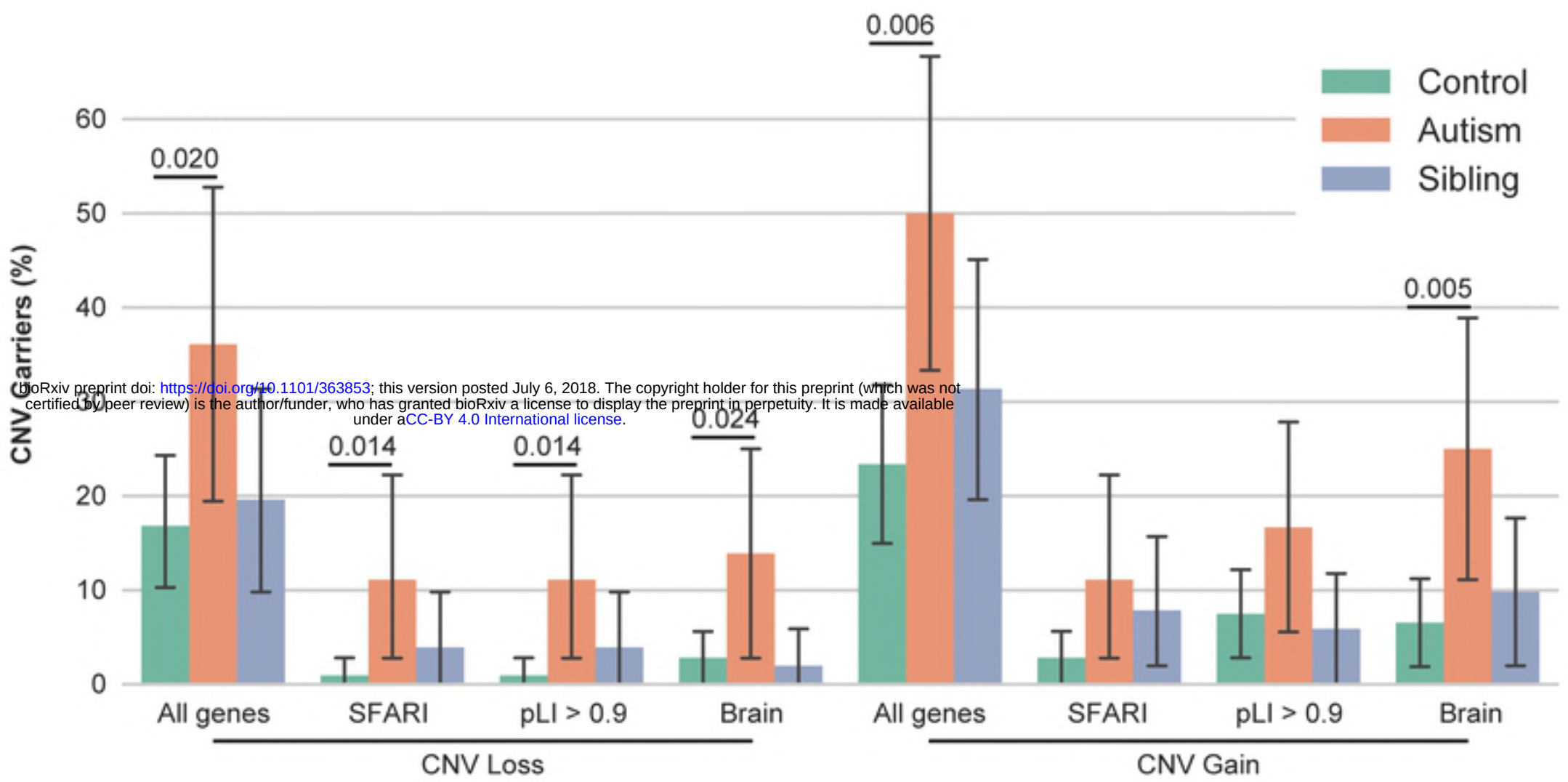
● ID
○ no ID

(dn): *de novo*
(dn?): unknown inheritance
(hom): homozygous

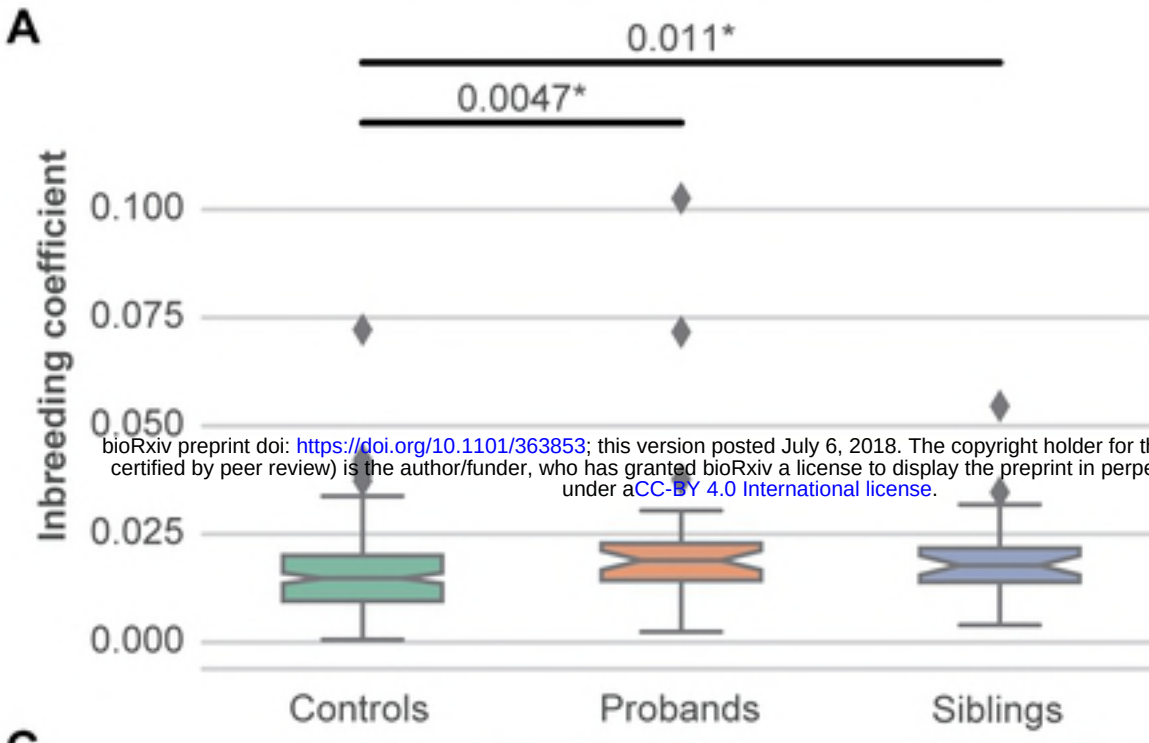
SFARI:	PRS-autism:
0	0.00015
13	0.00097

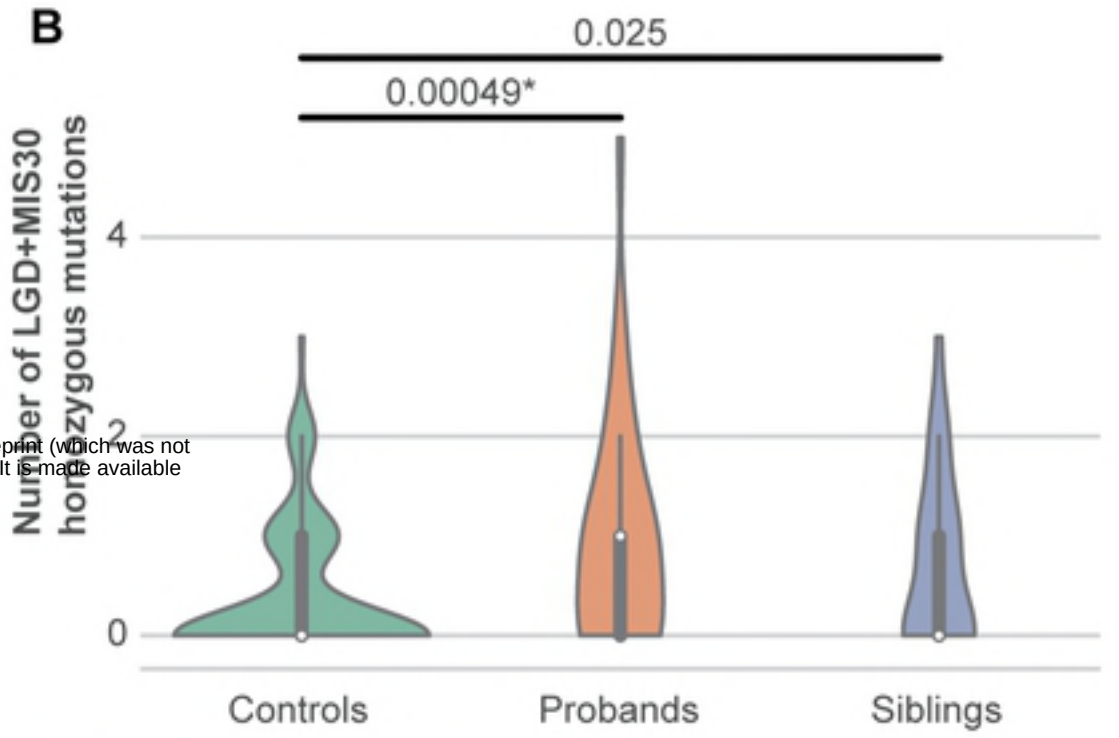

A
Iceland
Faroe islands
Norway
Ireland
United Kingdom
Germany
B
Epidemiological cohort
N=67
Genetic cohort
N=36
Percentage of individuals
80
70
60
50
40
30
20
10
0
no ID
ID
Males
Females
C
C2
C1
0.05
0.00
-0.05
-0.10
-0.05
0.00
0.05
0.1
0.15
Faroe
CEU
TSI
0.040
0.035
0.030
-0.06
-0.055
-0.05
-0.045
Populations
ASW YRI MKK LWK TSI CEU CHB CHD JPT MXL GIH Faroe
D
Inbreeding coefficients
0.075
0.050
0.025
0
ASW YRI MKK LWK TSI CEU CHB CHD JPT MXL GIH Faroe
***
**
Africa
Europe
Asia

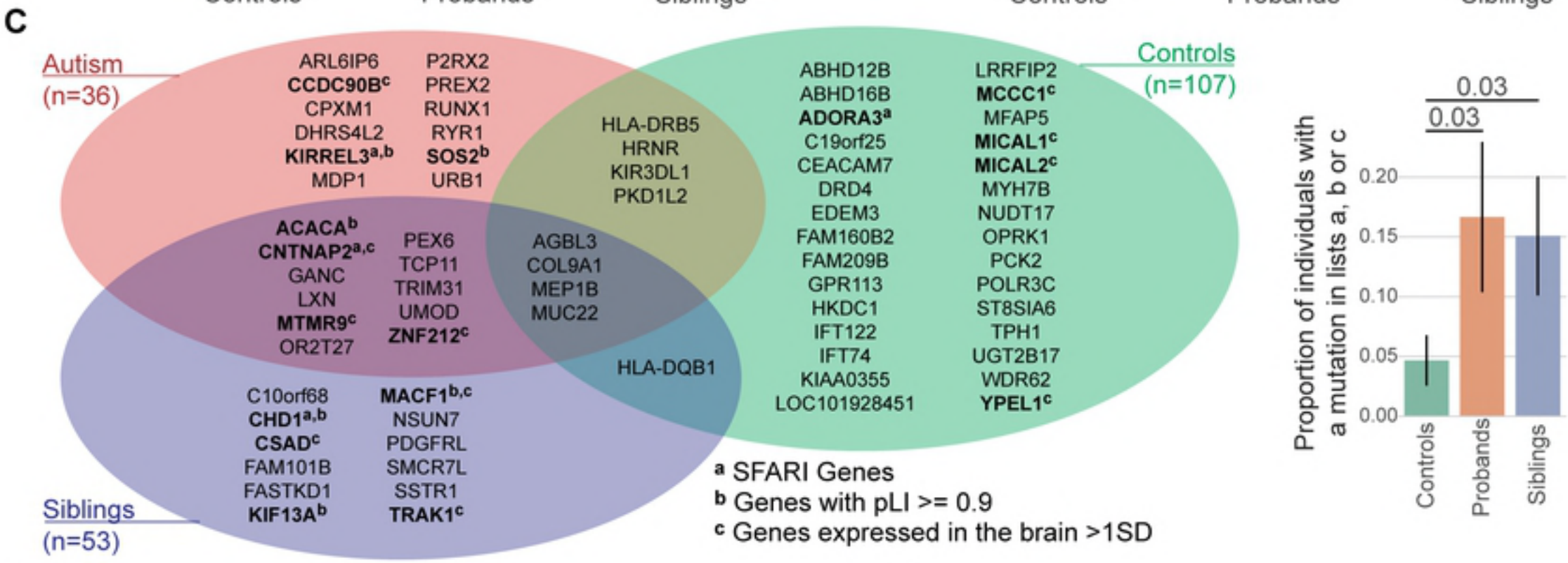
A

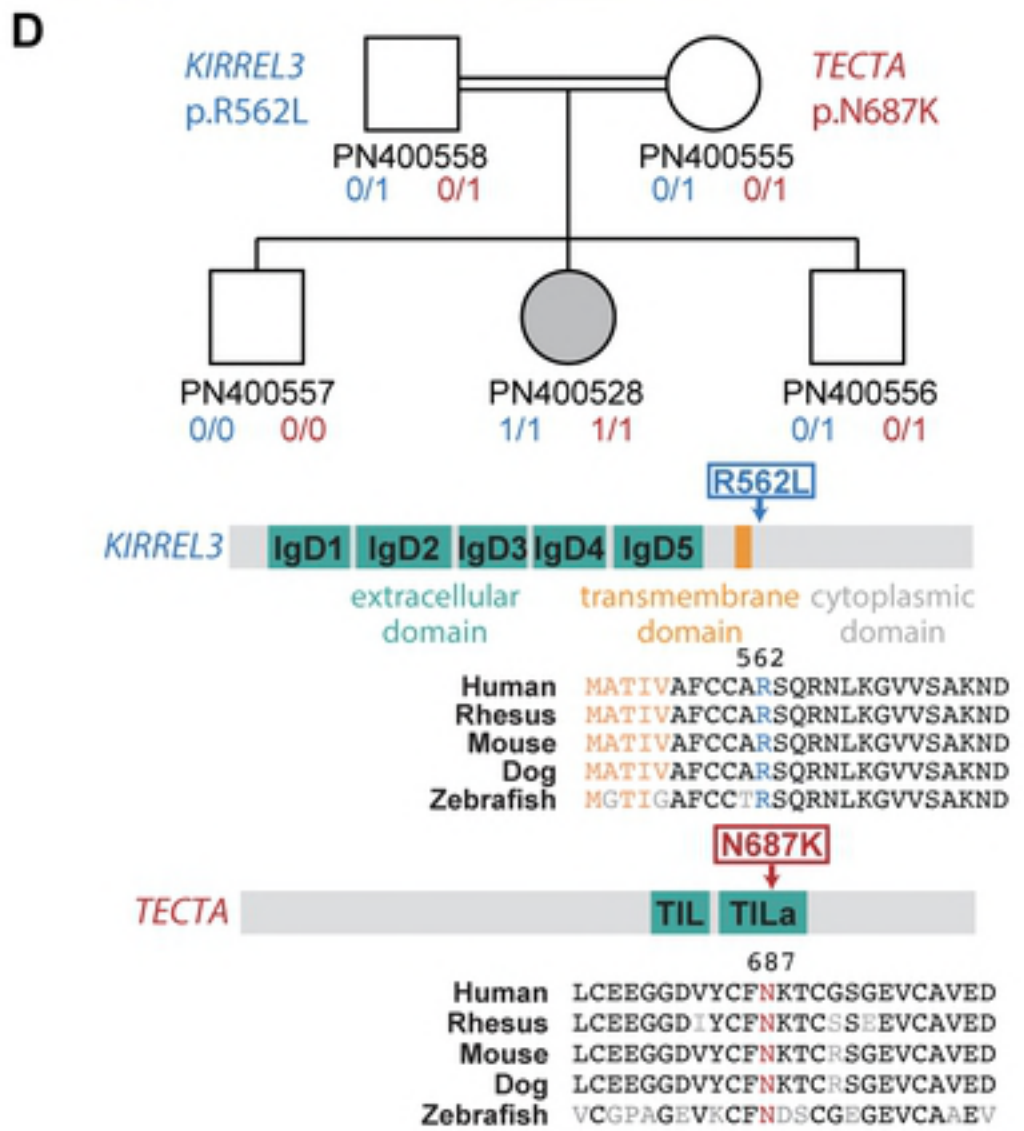


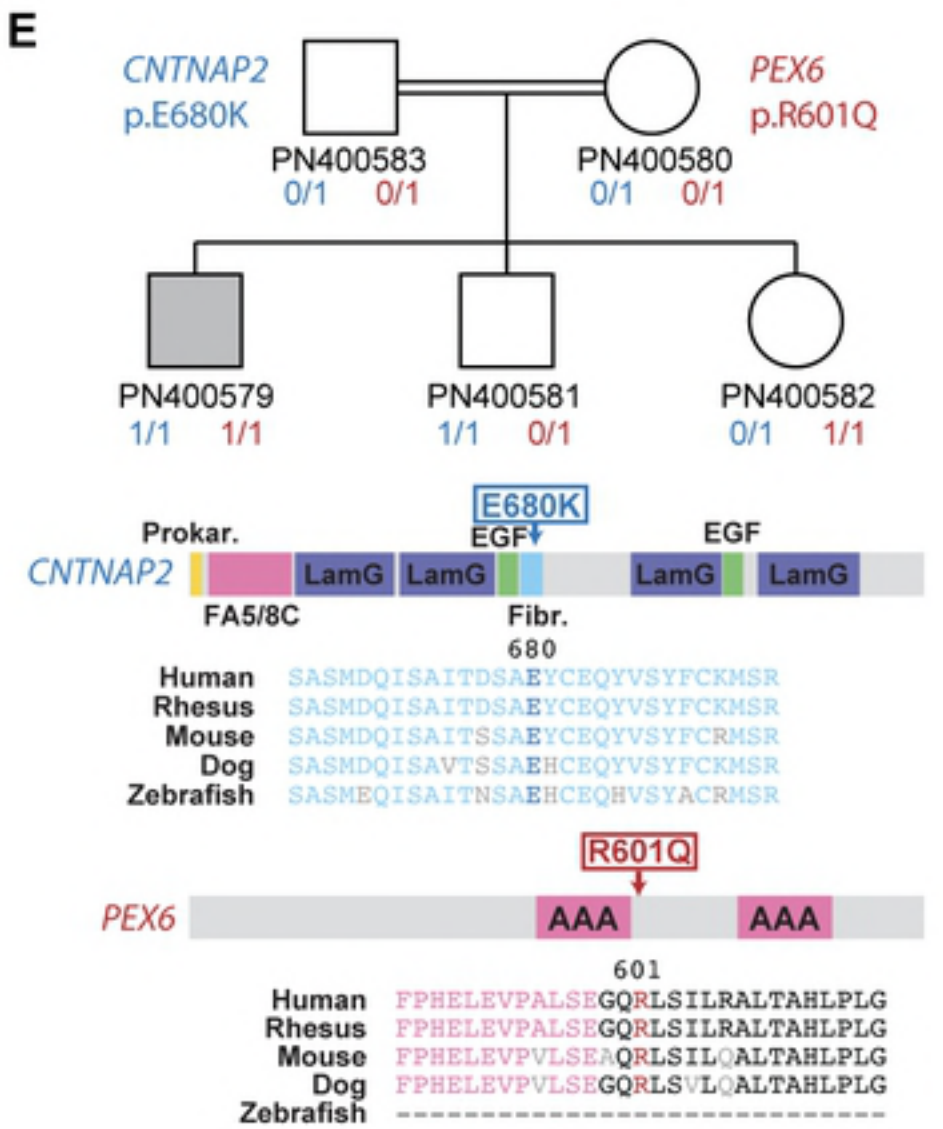
B

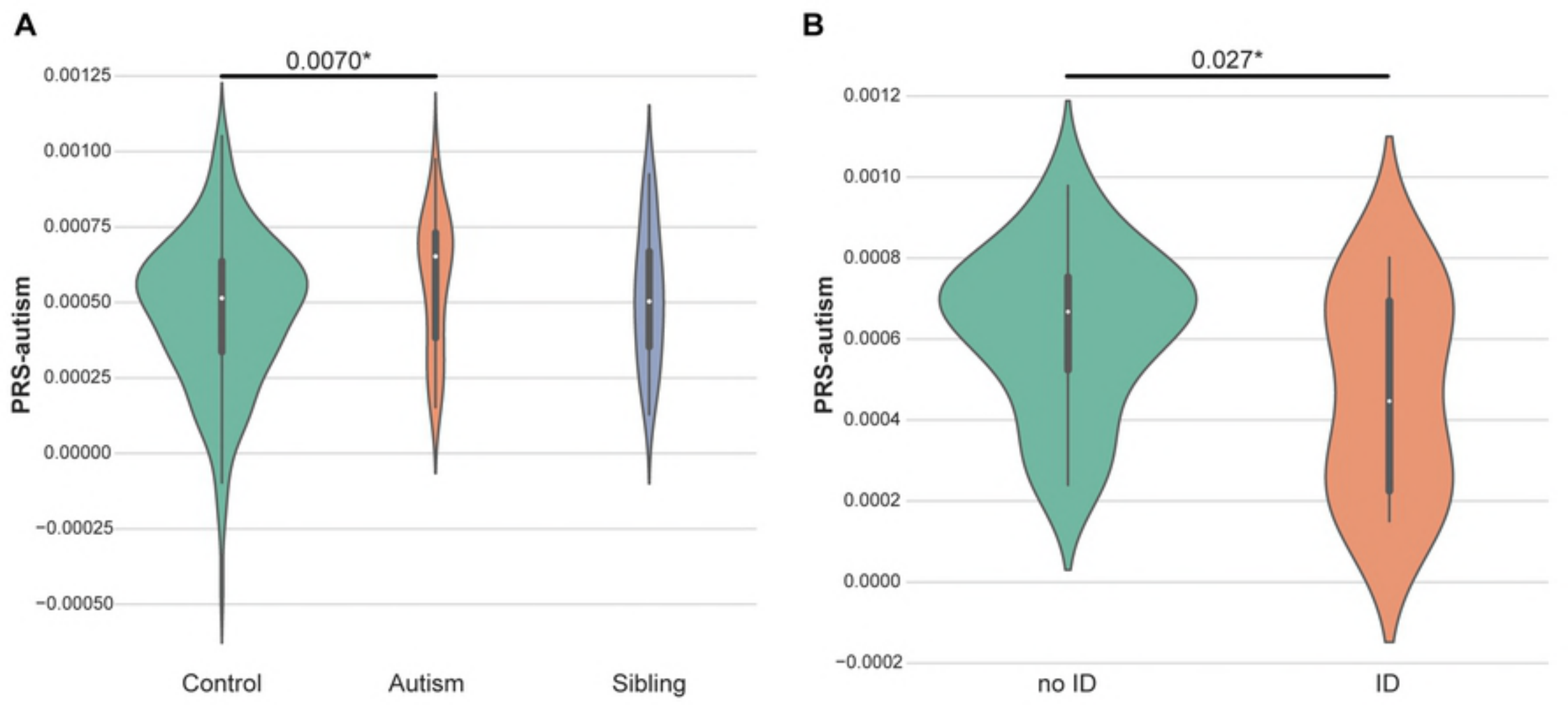











A
0.0070*
0.00125
0.00100
0.00075
0.00050
0.00025
0.00000
−0.00025
−0.00050
PRS-autism
Control
Autism
Sibling
B
0.027*
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
−0.0002
PRS-autism
no ID
ID