LEAP-InovAND a multiscale resource to explore genetics, brain imaging and clinical data in autism

M. Fleury[1]*, Z. M. Mougin[1]*, C. S. Leblond[1], F. Cliquet[1], A. Mathieu[1], B. Oakley[2], A. Maruani[1, 3], J.-M. de Saint Agathe[1], S. Korkmaz[1], R. Holt[4], A. Vitrac[1], F. Amsellem[1], A. Batut[1], G. Botton-Amiot[1], T. Rolland[1, 5], E. Barthome[1], R. Bonicel[1,3], A. Beggiato[1, 3], N. Lemière[1], A.-C. Tabet[1,6], J. Levy[6], C. Lehoucq[1], A. Ayrolles[1, 3], C. A. Moreau[7], L. Colomar Molla[2], R. Toro[8], N. Traut[8], A. Lefebvre[9], D. L. Floris[10], G. Dumas[11, 12], G. Huguet[7], S. Bonnot-Briey[1, 13, 14, 15], F. Campana[16], J. Van Gils[17], V. Wang[1], J. Buratti[18], L. Jornea[19], S. Loiodice[19], S. Forlani[19], S. Fournier[16], D. Bacq-Daïan[20], R. Olaso[20], V. Meyer[20], B. Fin[20], AIMS2-2-Trials AReps group‡, A. Ghosh[20, 21, 22], M. Doherty[23], S. Li-Williams[4], InovAND group‡‡, EU-AIMS/AIMS-2-Trials LEAP group‡‡‡, J.-F. Deleuze[24], A. M. Persico[25, 26], M. Leboyer[27, 28], C. Gillberg[29 ,30, 31], S. Bölte[32, 33, 34], C. Beckmann[35, 36], C. Ecker[2, 37, 38], E. Jh Jones[39], V. Warrier[4, 40, 41], S. Baron-Cohen[4, 40, 41], T. Charman[2], E. Loth[2], J. K. Buitelaar[35], J. Tillmann[2, 42], R. Delorme[1, 3], D. Murphy[2], and T. Bourgeron[1].

Equally contributed*

AIMS-2-Trials AReps group‡: M. Doherty, S. Douglas, J. Fialová Kučerová, A. Ghosh, R. Holt, A. Middlecamp-Sommer, R. Plas, K. Van den Bosch and P. Violland

InovAND group‡‡: M. Alison, A. Amestoy, F. Amsellem, S. Antoun, A. Ayrolles, A. Beggiato, R. Bonicel, A. Bourmaud, T. Bourgeron, M. Bouvard, G. Botton-Amiot, F. Cliquet, F. De Roux, J.-M. de Saint Agathe, C. Delclaux, R. Delorme, P. Ellul, M. Elmaleh-Berges, M. Fleury, D. Germanaud, S. Guillemin-Crépon, S. Kutbari, C. Lehoucq, C. S. Leblond, A. Lefebvre, J. Levy, F.-X. Mauvais, A. Mathieu, A. Maruani, Z. Mougin, A.-C. Tabet, N. Traut, R. Toro and A. Vergnon

EU-AIMS/AIMS-2-Trials LEAP group‡‡‡: J. Ahmad, S. Ambrosino, B. Auyeung, T. Banaschewski, S. Baron-Cohen, S. Baumeister, C. F. Beckmann, S. Bölte, T. Bourgeron, C. Bours, M. Brammer, D. Brandeis, C. Brogna, Y. de Bruijn, J. K. Buitelaar, B. Chakrabarti, T. Charman, L. Colomar Molla, I. Cornelissen, D. Crawley, F. Dell' Acqua, G. Dumas, S. Durston, C. Ecker, J. Faulkner, V. Frouin, P. Garcés, D. Goyard, L. Ham, H. Hayward, J. Hipp, R. Holt, M. H. Johnson, E. J. H. Jones, P. Kundu, M.-C. Lai, X. Liogier D'Ardhuy, M. V. Lombardo, E. Loth, D. J. Lythgoe, R. Mandl, A. Marquand, L. Mason, M. Mennes, A. Meyer-Lindenberg, C. Moessnang, N. Mueller, D. G. M. Murphy, B. Oakley, L. O'Dwyer,

M. Oldehinkel, B. Oranje, G. Pandina, A. M. Persico, A. Rausch, B. Ruggeri, A. Ruigrok, J. Sabet, R. Sacco, A. San José Cáceres, E. Simonoff, W. Spooren, J. Tillmann, R. Toro, H. Tost, J. Waldman, S. C. R. Williams, C. Wooldridge, M. P. Zwiers, I. Ilioska and T. Mei

Affiliations

[1]Human Genetics and Cognitive Functions, Institut Pasteur, CNRS UMR3571, Institut Universitaire de France, Université Paris Cité, Paris, France
[2]Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, Psychology and Neuroscience, King's College London, Camberwell, London, United Kingdom.
[3]Department of Child and Adolescent Psychiatry, Robert Debré Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France
[4]Autism Research Centre, Department of Psychiatry, University of Cambridge
[5]CRBM, Univ Montpellier, CNRS, Montpellier, France
[6]Department of Genetics, Cytogenetics Unit, Robert Debré Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France
[7]Centre de recherche CHU Sainte Justine, Department of Psychiatry and Addictology, University of Montreal, Montreal, Quebec Canada
[8]Applied and Theoretical Neuroanatomy, Institut Pasteur, Paris, France
[9]Paris Saclay University, Neurospin, CEA Saclay, Service Avis et Expertise TND, Fondation Vallée, Gif-sur-Yvette, France
[10]Methods of Plasticity Research, Department of Psychology, University of Zürich, Zürich, Switzerland
[11]Département de Psychiatrie et d'Addictologie, Université de Montréal, Montréal, QC, Canada
[12]Mila - Quebec Artificial Intelligence Institute, Montreal, Canada, Department of Biomedical Engineering, McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, QC, Canada
[13]HANDI-VOICE, Paris, France
[14]PAARI et fédération AUTOP-H, Paris, France
[15]Autism-Europe, Brussels, Belgium
[16]Institut Pasteur Medical Center, Paris Cité University, Paris, France
[17]Department of Medical Genetics, University Hospital of Bordeaux and INSERM U1211, University of Bordeaux, Bordeaux, France
[18]Department of Genetics, Pitié-Salpêtrière Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France
[19]Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, APHP, Hôpital de la Pitié Salpêtrière, Paris, France
[20]Department of Neuroradiology, University Hospital Mannheim, Germany
[21]Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Germany
[22]University of Heidelberg, Germany
[23]University College Dublin, Dublin, Ireland
[24]Centre National de Recherche en Génomique Humaine, Institut de Biologie François Jacob, CEA, Université Paris-Saclay, Evry-Courcouronnes, France
[25]Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy
[26]Child & Adolescent Neuropsychiatry Program, Modena University Hospital, Modena, Italy
[27]Fondation FondaMental, French National Science Foundation, Créteil, France

[28]Institut National de la Santé et de la Recherche Médicale (INSERM), U955, Institut Mondor de Recherche Biomédicale, Psychiatrie Translationnelle, Créteil, France
[29]Gillberg Neuropsychiatry Centre, University of Gothenburg, Gothenburg 411 19, Sweden
[30]Kochi Gillberg Neuropsychiatry Centre, University Medical School, Kochi, Japan
[31]School of Health and Wellbeing, University of Glasgow, Glasgow, United Kingdom
[32]Department of Women's and Children's Health, Center of Neurodevelopmental Disorders (KIND), Centre for Psychiatry Research, Karolinska Institutet and Region Stockholm, Stockholm, Sweden
[33]Child and Adolescent Psychiatry, Stockholm Health Care Services, Region Stockholm, Stockholm, Sweden
[34]Curtin Autism Research Group, Curtin School of Allied Health, Curtin University, Perth, Australia
[35]Donders Institute, Radboud University, Nijmegen, Netherlands
[36]Centre for Functional MRI of the Brain, Nuffield Department of Clinical Neurosciences, Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, United Kingdom
[37]Department of Child and Adolescent Psychiatry, University Hospital of the Goethe University, Frankfurt am Main, Germany
[38]Cooperative Brain Imaging Center (CoBIC), Goethe University Frankfurt, Frankfurt am Main, Germany
[39]Centre for Brain & Cognitive Development, Birkbeck, University of London, London, United Kingdom
[40]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom
[41]Department of Psychology, University of Cambridge, Cambridge, United Kingdom
[42]Roche Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd., Basel, Switzerland

Abstract

Most current autism research focuses on categorical comparisons (e.g., autistic vs. neurotypical people) and usually examines only one biological domain (e.g., cognition, genetics, or brain imaging). Here, we present a resource integrating quantitative phenotypic data, whole genome sequencing, brain magnetic resonance imaging, and electroencephalography. A total of 2,061 autistic people, 2,551 relatives and 875 neurotypical people were recruited in Europe through LEAP-InovAND, including 2,531 participants with both clinical and genetic data, and 875 with additionally neuroimaging data (EEG and/or MRI). By stratifying participants according to autistic traits and cognitive abilities, we identified three clusters with different load of both rare and common variants, particularly in genes associated with synaptic and chromatin remodeling functions. Correspondingly, each cluster exhibits different anatomical and connectivity patterns in brain regions previously linked to sensory and social information processing. This resource is available to support research into the complex connections between genetics, brain structure/function, and autism.

Main

Autism is currently defined by persistent challenges in social communication and interaction, sensory processing, as well as restricted and stereotyped patterns of behavior, interests, or activities. Our understanding of autism etiology derives from various scientific fields and from research studies comparing a group of autistic people to neurotypical (NT) people [1]. Genetics plays a crucial role in autism, with an estimated heritability to be around 80% [2–4]. This influence arises from various genetic factors, including single *de novo* or ultra-rare variants and common variants found in the general population[5,6]. The genetic variants associated with autism can influence early brain development through multiple biological pathways, including the modification of chromatin remodeling complexes that regulate gene transcription by opening or closing genomic regions transcription [7]. As neuronal circuits mature, genetic variants can also impact activity-dependent signaling and strength of synaptic connections [8]. Neuroimaging studies have shown differences in brain anatomy/connectivity, particularly in regions involved in social perception, language, or sensory processing [9–12]. Autism can be seen as an "emergent" condition shaped by the individual genetic profiles and the environment over time [13].

Most of this knowledge comes from studies showing results obtained at a group level, but the fact remains that what is relevant for one autistic individual in terms of etiology and support needs might be completely irrelevant for another [1]. The conceptualization of the heterogeneity of autism has evolved over time. Wing and Gould were instrumental in defining autism as a spectrum [14], while Coleman, Fernell and Gillberg later proposed the term "autisms+" to reflect the notion of distinct etiological subgroups and the high prevalence of co-occurring medical conditions, such as intellectual developmental disorder (IDD), attention deficit hyperactivity disorder (ADHD) or epilepsy [15]. Recently, the Lancet Commission proposed the term "profound autism" to describe the subset of people who are none or minimally verbal with significant intellectual disability, and requiring 24/7 care [16].

Understanding autism diversity should help to elucidate the underlying biological mechanisms involved in its physiology and allow the development of more efficient personalized interventions. Previous data-driven stratification using genetic

[3], neuroimaging [12], or clinical data [6] have identified on average 2-4 more homogeneous subgroups of autistic people, but this subgrouping depends mainly on the methodology and the population included in these studies (see **Supplementary Table 11** for a review). Studies typically focus on either genetics, neuroimaging or clinics, limiting our understanding of how the subgroups differ in genetic architecture, brain structure/function, and behavior.

To address the need for integrative approaches across modalities, we combined data from the EU-AIMS/AIMS-2-Trials Longitudinal European Autism Project (LEAP) [17,18] and the Innovation for Autism and Neurodevelopmental Disorders (InovAND) cohort. By clustering clinical data, we identified three subgroups within the autism spectrum showing differences in the nature of the *de novo*/rare and common variants mutations as well as in their pattern of brain anatomy/function. This integrated approach not only enhances our understanding of autism's heterogeneity but also paves the way for more targeted and effective interventions, ultimately improving the quality of life for people on the autism spectrum.

Results

Overview of the LEAP-InovAND dataset

LEAP and InovAND are two distinct studies that recruited participants in different centers across Europe (**Fig. 1a**, **Supplementary Fig. 1**). We collected demographic and clinical data similarly to the Simons Foundation Powering Autism Research for Knowledge (SPARK) [19], [20]and Autism Brain Imaging Data Exchange (ABIDE), but our dataset includes a subset of people with clinical, genetic, neuroimaging data and bio-samples (**Fig. 1b**). The combined dataset encompasses 5,549 people (*i.e.,* 2,374 families) including 2,061 diagnosed with autism (DSM criteria) using gold standard (ADOS/ADI-R) instruments (**Fig. 1b; Supplementary Fig. 2**). The LEAP dataset includes 493 autistic people, 697 relatives (mostly parents) and 466 neurotypical (NT) people (**Supplementary Table 1**, **Supplementary Fig. 3**). The InovAND dataset includes 1,593 autistic people, 1,854 relatives (nuclear and extended family members) and 388 NT people (**Supplementary Table 2**, **Supplementary Fig. 3**). The clinical data collected by LEAP and InovAND show several similarities (**Supplementary Fig.**

3), such as comparable measured IQ scores among NT people and similar Vineland Adaptive Behavior Scales (VABS) scores among autistic people (**Supplementary Table 3** and **Supplementary Fig. 3**). However, regarding sex ratio, LEAP recruited a higher proportion of female autistic probands compared to InovAND, whereas the opposite pattern was observed among NT people (**Fig. 1c, Supplementary Tables 1 and 2**). Furthermore, autistic people from InovAND exhibited higher SRS-2 total t-scores (Mann-Whitney U test (MWU), $P_{MWU} = 1.62 \times 10^{-5}$) and lower full scale IQ scores ($P_{MWU} = 4.36 \times 10^{-35}$) than those from LEAP. We also collected data on parents and undiagnosed siblings, enriching the dataset for familial comparisons. Finally, 62 people with IDD who do not meet diagnostic criteria for autism were enrolled (36 for LEAP and 26 InovAND). All together the LEAP-InovAND dataset includes a fair representation of the whole continuum of autism in terms of autistic traits, cognition, gender, and age range (**Fig. 1e**).

Genetic profile of the LEAP-InovAND participants

Following quality control and data harmonization procedures (see **Online Methods**), we analyzed whole genome sequencing (WGS) data through single nucleotide variants (SNV), small insertions/deletions (indel) and structural variants (SV) such as copy-number variants (CNV). The dataset contains WGS data for 2,531 people, including 800 with autism and their 1,210 relatives, 41 people with IDD only, and 480 NT people.

De novo variants*:* In 377 families with parental WGS data available, we could identify on average 77 *de novo* SNV (range 37-126) per participant. Consistent with previous studies [21], we observed a positive correlation between paternal age and the number of *de novo* SNV, with no significant excess in autistic people compared to their siblings (**Supplementary Fig. 9**). However, when focusing on predicted deleterious *de novo* variants in genes intolerant to protein-truncating variants (hereafter called "constrained genes" see **Online Methods**), autistic people were significantly more likely to carry such deleterious variants compared to their siblings (**Fig. 2a;** autism = 14.9%, undiagnosed siblings/NT = 3.9%, OR = 4.25, Fisher's exact test false discovery rate (FDR) corrected (Fisher FDR) $P_{Fisher\ FDR} = 8.1 \times 10^{-5}$). The enrichment was even higher for variants in high-confidence neurodevelopmental genes (HCNDD; autism = 11.50%, undiagnosed siblings/NT = 1.14%, OR = 11.30; $P_{Fisher\ FDR} = 7.0 \times 10^{-6}$), for autism-linked list of genes (*e.g.*, SFARI1/SPARK lists; autism = 8.89%, undiagnosed siblings/NT=0.57%, OR = 16.94, $P_{Fisher\ FDR} = 3.4 \times 10^{-5}$), and for genes with definitive and strong EAGLE[22] scores, a gene-level metric for autism gene evidence (autism = 4.85%, undiagnosed siblings/NT = 0%, OR = inf). Notably, genes involved in synaptic function (SynGO[23]) and/or chromatin remodeling/transcription (ChromEpiTF), two biological pathways previously implicated in autism [24], were enriched in *de novo* variants (SynGO: autism = 8.89%, undiagnosed siblings/NT = 0.57%, OR = 16.94, $P_{Fisher\ FDR} = 3.4 \times 10^{-5}$; ChromEpiTF: autism = 8.82%, undiagnosed siblings/NT = 1.70%, OR = 4.85, $P_{Fisher\ FDR} = 0.0032$). We confirmed all *de novo* variants in HCNDD genes using PCR and Sanger sequencing (**Supplementary Fig. 9**). Using the SPARK as a replication cohort, trends are similar for all the gene sets studied above (**Supplementary Fig. 15a**).

High impact ultra-rare variants: the definition of the high impact ultra-rare variants (hiURV) is summarized in **Fig. 2b and Supplementary Fig. 6**. Briefly, these variants can be *de novo* or inherited, but they must have minor allele frequencies (MAF) of less than 0.1% in the general population and be predicted to have a deleterious impact on the function of the protein (see **Online Methods)**. Here, we focused our analysis on loss-of-function (LoF) variations and deletions (DEL) within "constrained genes". Compared to NT people, autistic people were more likely to carry at least one hiURV in HCNDD genes (**Fig. 2b**, autism = 15.57%, NT = 4.73%, OR = 3.53; $P_{Fisher\ FDR}$ = 0.035), autism-linked genes (SPARK/SFARI1; autism = 9.15%, NT = 1.85%, OR = 4.54, $P_{Fisher\ FDR}$ = 1.2 x 10^{-5}) and EAGLE genes (autism = 3.96%, NT = 0.41%, OR = 8.23, $P_{Fisher\ FDR}$ = 0.0018). Genes involved in synaptic function and/or chromatin remodeling/transcription were also enriched for hiURV (SynGO: autism = 14,71%, NT = 7.41%, OR = 2.05, $P_{Fisher\ FDR}$ = 0.0023; ChromEpiTF: autism = 12.73%, NT = 6.58%, OR = 2.12, $P_{Fisher\ FDR}$ = 0.0023). Similar results are obtained if deleterious missense variants are added and if genes less constrained are included (**Supplementary Fig. 11f**). Using SPARK as a replication cohort, similar trends are observed (**Supplementary Fig. 15b**).

Genetic diagnostic yield: Returnable clinically relevant variants are filtered through a very strict pipeline (see **Supplementary Fig. 5**) that retains only functional variants in genes/*loci*/aneuploidies previously known as causes of autism/neurodevelopmental disorders (NDD). We found these variants in 13.7% of the autistic people (n=110/800) and in 39% of the people with IDD only (n=16/41) (**Fig. 2c**). In autistic people, co-occurring IDD raises the diagnostic yield to 32.1%, while its absence lowers it to 7.6% (**Supplementary Fig. 12b**). *SHANK3* is the most frequently implicated gene in the cohort, encompassing eight distinct families (in one family, two sons carried a *SHANK3* LoF variant transmitted from a mother mosaic for this variant). These nine people carrying *SHANK3* variants accounted for 4.7% of autistic people with IDD (n = 190). *SHANK3* was followed in frequency by recurrent CNV involving the 15q and 16p loci (n = 5 each) and by *SHANK2* (n = 4). (**Fig. 2c**).

Polygenic Scores (PGS): Common variants such as single nucleotide polymorphism (SNP) also contribute to autism and to a range of other traits and neuropsychiatric

conditions, including post-traumatic stress disorder (PTSD), ADHD, epilepsy, cognitive performance (as assessed by IQ scores), and various dimensions of subjective well-being (**Supplementary Figs. 16 and 17**). Although each SNP exerts only a small effect on a person's phenotype, their accumulation in the person can be calculated through PGS. PGS were calculated for each participant in the LEAP-InovAND cohort across multiple traits (see **Online Method**). As expected, autism PGS showed the strongest difference between autistic people and NT (autism *vs.* NT; OR = 2.88, $P_{MWU} = 1.5 \times 10^{-9}$), followed by PTSD (OR = 1.95, $P_{MWU} = 0.0002$), anxiety (OR = 1.81, $P_{MWU} = 0.0007$), depression (OR = 1.63, $P_{MWU} = 0.0052$), ADHD (OR = 1.54, $P_{MWU} = 0.01$), and neuroticism (OR = 1.44, $P_{MWU} = 0.04$) (**Fig. 2d**). Analyses restricted to people of European ancestry yielded comparable results (**Supplementary Fig. 15a**). Replication analyses conducted using the independent SPARK cohort (**Supplementary Fig. 16a**) corroborated these findings. In the LEAP-InovAND and the SPARK cohorts, the "felt loved as a child" PGS displayed the lowest odds ratio among autistic people (LEAP-InovAND: autism *vs.* NT OR = 0.54, $P_{MWU} = 0.0012$; SPARK: autism *vs.* undiagnosed siblings OR = 0.86, $P_{MWU} = 0.0089$). The concordance of the PGS odds ratio across both cohorts (**Supplementary Fig. 16b**) underscores the similar contribution of common variants in autistic people from independent samples.

Brain imaging profile of the LEAP-InovAND participants

Magnetic resonance imaging (MRI) and electroencephalography (EEG) data were collected across 6 centers and 8 MRI scanners (6 from the LEAP study and 2 from InovAND). Following quality control, standard preprocessing and data harmonization procedures (see **Online Methods**), we evaluated diagnostic differences in cortical morphology, functional connectivity, and electrophysiological features.

Structural MRI (sMRI) analysis: The *cortex* was parcellated into 62 regions of interest (ROI) using the Desikan–Killiany (DK) atlas [25] (see **Online Methods**). Compared to NT (n = 349) autistic people (n = 776) exhibited a selective pattern of increased cortical thickness, primarily in temporo-parietal and occipital regions (**Fig. 3a**). The largest effect was observed in the left superior temporal gyrus (STG; t = 5.37, Bonferroni-adjusted p-value (BON) $P_{BON} = 7.6 \times 10^{-6}$), with additional increases were observed in the right inferior parietal lobe (t = 3.54, $P_{BON} = 0.029$), right lateral occipital cortex (t =

3.42, P_{BON} = 0.046), left rostral middle frontal (t = 3.67, P_{BON} = 0.017) and right superior temporal (t = 3.6, P_{BON} = 0.02). No surface area differences survived correction for multiple comparisons (**Fig. 3b**). Among subcortical volumes (**Fig. 3c**), only the right lateral ventricles were enlarged in the autism group (t = 3.26, P_{BON} = 0.032).

*Resting state functional MRI (rsfMRI) analysis***:** Functional connectivity was computed between 115 brain regions for 396 autistic people 296 NT people. We then averaged the voxel-wise time series within each parcel and computed Pearson's correlation between the time course of all pairs of parcels for each participant as described in [26](see **Online Methods**). Group differences were assessed at the edge level, and only connections surviving FDR correction (P_{FDR} < 0.05) were retained for subsequent analyses. To summarize effects at the network level, we quantified edges showing hypoconnectivity or hyperconnectivity and report both the raw number of significant edges (N_{raw}) and the normalized proportion of significant edges relative to all possible connections within/between each network pair (N_{norm}). Group contrasts revealed extensive underconnectivity in autism. The highest fraction of hypoconnected edges occurred between somatomotor and visual networks (N_{norm} = 0.22; *e.g.*, 22% of possible edges) (**Fig. 3d-e**), followed by visual–dorsal attention (N_{norm} = 0.17), and somatomotor-dorsal attention (N_{norm} = 0.12). Default mode connections to sensory–attentional systems were also reduced, notably default mode-visual (N_{raw} = 36, N_{norm} = 0.088) and default mode-dorsal attention (N_{raw} = 29, N_{norm} = 0.081). Within network hypoconnectivity peaked in somatomotor (N_{raw} = 27, N_{norm} = 0.29), with smaller within network reductions in visual (N_{raw} = 10, N_{norm} = 0.074) and default mode (N_{raw} = 19, N_{norm} = 0.069). In contrast, hyperconnectivity (**Fig. 3f-g**) was sparse and centered on the default mode network's (DMN) coupling with subcortical and control/attention systems (<4% of possible edges), including default mode-thalamus, default mode-frontoparietal, and default mode-ventral attention. Few within network hyperconnectivity effects were observed.

Resting state EEG (rsEEG) analysis: EEG spectral features were compared between 406 autistic people and 217 NT. *Alpha* peak frequency (6-13 Hz) measured in occipital regions was comparable between autistic (n = 406) and NT (n = 217) participants (**Supplementary Fig. 20**), and spectral power across the five canonical frequency

bands (*delta*, *theta*, *alpha*, *beta*, and *gamma*) did not differ significantly. This absence of robust EEG differences, at the group level, aligns with recent population-scale reports[27].

Clustering analysis on clinical profiles of people

As an example of the analytical potential of the LEAP-InovAND dataset, we conducted a subgrouping analysis of 1,023 unrelated people (609 autistic people, 31 people with IDD, and 383 NT) using the Social Responsiveness Scale total (SRS-2 total t-score) and the measured IQ scores (see **Supplementary Section 6**). The SRS-2 t-score is a widely used, standardized questionnaire that quantifies autism-related traits dimensionally, allowing for more nuanced analyses beyond binary diagnostic categories [28]. Feature selection for subgrouping balanced multiple considerations: principal component analysis on all available clinical measures showed that IQ and SRS-2 total t-scores were among the top contributors to the first two principal components (capturing 78.6% of the total variance, **Supplementary Fig. 21**), while also maximizing sample size. This 2-variable combination retained 1,023 participants compared to 511 participants when including VABS composite scores (See **Supplementary Section 6**, **Supplementary Table 12**). Hierarchical clustering based on measured IQ and SRS-2 total t-score, revealed an optimal three-cluster solution as determined by comprehensive fit indices analysis (*e.g.*, silhouette, variance ratio score, gap statistic, pseudo-F) (**Fig. 4a**; **Supplementary Table 13**; **Supplementary Fig. 21 and 22**). Cluster 1 (C1) includes 451 people; cluster 2 (C2) comprises 379 people; and cluster 3 (C3) consists of 193 people (**Fig. 4b**). These clusters demonstrate distinct profiles across clinical presentation, genetic architecture, and neuroimaging features, suggesting meaningful stratification within the autism spectrum.

Clinical profiles: Cluster C1 gathers people with medium-to-high IQ scores and low levels of autistic traits (*i.e.*, SRS-2 total t-scores below the cut-off for autism). This cluster includes primarily NT people (n = 354; 92.4% of all NT; 78.5% of people in C1), along with a subset of autistic people without IDD (n = 97; 15.9% of all autistic people; 21.5% of C1). Cluster C2 includes people with medium-to-high IQ and elevated SRS-2 total t-scores and is composed almost entirely of autistic people

without IDD (n = 360 58.9% of all autistic people; 95.0% of C2). Cluster C3 includes people with low IQ and high SRS-2 scores, with most of the people in this cluster are autistic (n = 152; 25.0% of all autistic people; 78.7% of C3), with approximately half also meeting criteria for IDD (n = 79). Notably, almost all people with a clinical diagnosis of IDD only are also assigned to C3 (n = 30; 96.8% of all people with IDD; 15.5% of C3). Beyond IQ and autistic traits, we observed clinical differences in the autistic people across clusters, particularly with an increased prevalence of clinical diagnosis of ADHD in C2 and C3 compared to C1 (C2: $P_{Fisher} = 1.7 \times 10^{-17}$; C3: $P_{Fisher} = 1.3 \times 10^{-11}$). Similarly, autistic people within C2 and C3 (**Fig. 4c**) presented with higher sensory atypicality (measured by the short sensory profile, SSP) compared to those within C1 (C2 *vs.* C1: $P_{MWU} = 2.2 \times 10^{-10}$; C3 *vs.* C1: $P_{MWU} = 3.0 \times 10^{-9}$). Additionally, autistic people within C2 and C3 displayed lower adaptive functioning (assessed via the VABS) compared to those within C1 (C2 *vs.* C1: $P_{MWU} = 3.3 \times 10^{-7}$, C3 *vs.* C1: $P_{MWU} = 7.1 \times 10^{-19}$). Developmental milestones analysis revealed that autistic people in all clusters showed delayed acquisition of early motor and language milestones (e.g., age at onset of walking, first words, first phrases) compared to NT people, but with a gradient of severity, with the most pronounced delays observed for autistic people from C3 (see **Supplementary Fig. 23** for detailed statistics). Interestingly, no significant sex differences were observed between the autistic people in C1, C2 and C3 ($\chi^2 = 1.5$, $P_{Chi\ squared} = 0.47$).

Genetic profiles: We identified significant differences in the genetic architecture of the participants across the three clusters. In total, 22.7% of autistic people in C3 received a genetic diagnosis with a returnable variant linked to autism/NDD, versus 6.1% of autistic people in C2 and 7.5% of autistic people in C1 (C1 *vs.* C2: $P_{Fisher} = 0.012$; C1 *vs.* C3: $P_{Fisher} = 1.3 \times 10^{-5}$) (**Fig. 5a**). Regarding rare variants, for the constrained HCNDD gene list, all clusters of autistic people exhibit a significantly higher variant burden relative to the NT (C1: OR = 4.4, $P_{Fisher\ FDR} = 0.0056$; C2: OR = 3.9, $P_{Fisher\ FDR} = 2.7 \times 10^{-4}$; C3: OR = 7.7, $P_{Fisher\ FDR} = 1.7 \times 10^{-7}$; IDD: OR = 10, $P_{Fisher\ FDR} = 0.0013$). The same is true for the constrained SPARK/SFARI1 gene list (C1: OR = 4.6, $P_{Fisher\ FDR} = 0.0027$; C2: OR = 3.4, $P_{Fisher\text{-}FDR} = 0.023$; C3: OR = 6.8, $P_{Fisher\ FDR} = 0.0003$). Interestingly, only autistic people in cluster C3 display a significant association for the constrained EAGLE definitive and strong gene set, as well as for the SynGO and ChromEpiTF gene

lists (EAGLE d&s: OR = 6.7, $P_{Fisher\ FDR}$ = 0.037; SynGO: OR = 2.7, $P_{Fisher\ FDR}$ = 0.0064; ChromEpiTF: OR = 3.2, $P_{Fisher\ FDR}$ = 0.048). Focusing on the cell types of the prefrontal cortex and expression patterns during neurodevelopment, only autistic people from cluster C3 showed significant OR compared with NT, mostly for genes expressed early during brain development such as the medial ganglionic eminences (MGE) inhibitory interneurons and the principal excitatory neurons (**Supplementary Fig. 13**).

Compared to NT, PGS analysis revealed increased OR for autism, ADHD, and major depressive disorder (MDD/depression) in autistic people from clusters C1 (autism OR = 3.2, P_{MWU} = 0.0063; ADHD OR = 2.5, P_{MWU} = 0.035; MDD OR = 2.8, P_{MWU} = 0.02) and C2 (autism OR = 2.6, P_{MWU} = 0.0002; ADHD OR = 2.4, P_{MWU} = 0.0007; MDD OR = 2.0, P_{MWU} = 0.0052), but not in C3. Notably, as mentioned, the autism PGS in C3 did not differ significantly between autistic people and NT (autism OR = 1.8, P_{MWU} = 0.1). Compared to NT people, elevated PGS for PTSD was observed in both C2 (PTSD OR = 3.4, P_{MWU} = 1.0 x 10^{-6}) and C3 (PTSD OR = 2.7, P_{MWU} = 0.0021), with anxiety PGS specifically increased in C2 (anxiety OR = 2.3, P_{MWU} = 0.001). Finally, the PGS related to intelligence was significantly reduced in autistic people from C3 compared to NT (intelligence OR = 0.4, P_{MWU} = 0.0041), an effect not seen in C1 or C2 (**Fig. 5b**). In the overall cluster population, there was a positive correlation between the PGS for intelligence and the measured IQ score (r = 0.217, $P_{Pearson}$ = 1.0 x 10^{-6}; **Fig. 5c**). Notably, several autistic people in cluster C3 carried highly deleterious rare variants in SynGO and ChromEpiTF genes along with low intelligence PGS scores (**Supplementary Fig. 24**), suggesting cumulative effects of rare and common variants in this subgroup. Finally, we examined the associations between IQ scores and rare variants in genes based on their level of intolerance to LoF variants (LOEUF scores, **Fig. 5d**). A negative correlation was observed for all protein-coding genes (r = -0.08, P_{perm} = 0.044), but the strongest association was found between higher constraints within SynGO genes and lower IQ (r = -0.24, P_{perm} = 0.0002).

Brain structural/functional profiles of people based on clustering:

sMRI: Autistic people from cluster C2 showed greater cortical thickness than NT in the left STG (t = 3.89, P_{FDR} = 0.008) (**Fig. 6a**). In contrast, autistic people in C3 exhibited

thinner cortex than NT in the right precentral gyrus (t = -4.06, P_{FDR} = 0.004), right temporal pole (t = -3.68, P_{FDR} = 0.008), and left precentral gyrus (t = -3.19, P_{FDR} = 0.026). Cluster C3 also showed a thicker cortex than NT in the right pericalcarine visual cortex (t = 3.63, P_{FDR} = 0.008) and left pericalcarine visual cortex (t = 3.13, P_{FDR} = 0.026). Autistic people in C3 showed reduced surface area compared to NT in the left pars orbitalis (t = -3.39, P_{unc} = 0.001, P_{FDR} = 0.06). For subcortical volumes, Autistic people in C1 showed larger right lateral ventricle volume compared to NT (t = 3.64, P_{FDR} = 0.009).

rsfMRI: Autistic people from cluster C2 showed more inter-network hypoconnectivity than NT in the default mode and visual networks (N_{raw} = 48, N_{norm} = 0.12), followed by dorsal attention with default mode (N_{raw} = 30, N_{norm} = 0.08) and ventral attention with somatomotor (N_{raw} = 32, N_{norm} = 0.19) (**Fig. 6b**). Hyperconnectivity in autistic people in C2 *vs.* NT was widespread, notably involving thalamo-cortical and limbic–cortical coupling: thalamus with default mode (N_{raw} = 27, N_{norm} = 0.07), default mode with ventral attention (N_{raw} = 13, N_{norm} = 0.05), limbic with visual (N_{raw} =15, N_{norm} = 0.13), and frontoparietal with somatomotor (N_{raw} = 11, N_{norm} = 0.06). For autistic people in C3 *vs.* NT, hypoconnectivity concentrated between default mode and visual (N_{raw} = 31, N_{norm} = 0.08) and between dorsal attention and somatomotor (N_{raw} =13, N_{norm} = 0.06), while hyperconnectivity featured thalamus–default mode (N_{raw} = 24, N_{norm} = 0.06), default mode–frontoparietal (N_{raw} = 19, N_{norm} = 0.06). In contrast, autistic people in C1 *vs.* NT exhibited sparse effects overall; hypoconnectivity was limited (*e.g.,* thalamus–visual, N_{raw} = 5, N_{norm} = 0.02), and hyperconnectivity was modest, with small increases in default mode–dorsal attention (N_{raw} = 4, N_{norm} = 0.01) and default mode-somatomotor (N_{raw} = 5, N_{norm} = 0.015).

rsEEG: The examination of the autism clusters compared to NT people, revealed distinct EEG patterns. Autistic people in C3 (n = 51), showed significantly lower *alpha* peak frequency in occipital region compared to NT (t = -3.47, P_{FDR} = 0.0009) (**Supplementary Fig. 26**). In contrast, autistic people in C1 and C2 showed no significant differences from NT.

Correlation between genes, brain structure/function and IQ

For all carriers of LoF variants, we examined if there was a link between higher constraint on the mutated gene (indicated by higher -log(LOEUF) score [29]) and higher cortical thickness (**Fig. 7a-c**). We found two positive correlations: the left STG showed significantly increased thickness associated with higher constraint scores in ChromEpiTF factor genes ($r = 0.215$, $P_{perm} = 0.02$) and the left lateral occipital ($r = 0.239$, $P_{perm} = 0.002$). To showcase the diversity of clinical, genetic and neuroimaging data, we displayed the IQ score and the SRS-2 total t-score, highlighting one autistic person in each cluster, carrying predicted deleterious variants in either SynGO or ChromEpiTF genes (**Fig. 7d**).

Discussion

This study represents a major collaborative effort that brought together a diverse group of researchers and experts by experience from across Europe, including psychologists, psychiatrists, geneticists, neuroscientists, and, importantly, autistic people, relatives and caregivers (see **Supplementary Section 1 and Section 8** on participatory research and the frequently asked questions). This integrative framework has allowed us to investigate how genetic variation may shape brain structure, and in turn, how these brain differences relate to autistic traits and developmental outcomes.

In the first part of this study, we demonstrated that this new LEAP-InovAND dataset closely resembles previously established datasets in terms of clinical features, brain imaging, and genetics. This similarity confirms shared biological pathways while also highlighting differences in brain anatomy/function between autistic and NT people. In the second part, we found that group-level distinctions can be more precisely attributed to specific clinical subgroups.

For instance, previous work, including our own, has shown that the genetic architecture of autism involves a complex interplay of *de novo* variants, rare and common inherited variants [2–6]. In the LEAP-InovAND dataset, we found that *de novo*/rare variants in constrained genes were significantly more frequent in autistic people compared to NT people. This enrichment was even more pronounced when variants disrupted coding sequences of genes previously implicated in autism/NDD.

These variants were enriched in autistic people with co-occurring cognitive challenges (**Supplementary Fig. 12**), although they were not exclusive to this subgroup. The returnable genetic variants in autistic people were enriched with genes previously associated with mild-to-moderate IDD (*e.g.*, *ADGRL1*, *ITSN1*, *KDM5B*, *QRICH1*, *CHD5*, *CHD8*), highlighting the overlap of autism with mild IDD. Interestingly, several of these genes have been challenging to implicate in NDD due to their wide clinical spectrum and incomplete penetrance (*ADGRL1*, *ITSN1*, *KDM5B*, *PSMD11*)[30–33]. One example is the gene coding for the presynaptic protein BSN, recently implicated in a wide spectrum of NDD [34], for which the only participant with a truncating variant in the LEAP-InovAND cohort was a 12-year-old NT female with some autistic traits according to her SRS-2 t-score at 57. Our cohort that includes phenotypic explorations of NT people provides a valuable resource to better understand some wide-spectrum genetic conditions on the edge of autism.

We also confirmed previous findings implicating genes involved in chromatin remodeling, transcriptional regulation, and synaptic function in autism [8]. Genes expressed during prenatal development (*e.g.,* chromatin-related genes) and those expressed later in development (*e.g.,* synaptic genes) were similarly represented across subgroups of autistic people. These results suggest that the presence of certain genetic variants alone is insufficient to fully explain the phenotypic variability in autism, pointing to the influence of additional genetic, environmental, or developmental modifiers [5].

We then examined the contribution of common genetic variants by calculating PGS for a range of traits related to cognition, psychiatric outcomes, and well-being. Autistic people without IDD (*i.e.*, cluster C1 and C2) tended to have on average slightly higher autism PGS scores compared to autism with IDD (*i.e.*, cluster C3) [3].

We found that cluster C3 of autistic people with low IQ exhibited significantly lower intelligence PGS compared to other clusters. This same cluster was also enriched for *de novo*/rare variants affecting constrained genes and genes previously implicated in NDD. Importantly, both carriers and non-carriers of these rare variants within the C3

cluster showed similarly reduced intelligence PGS (**Supplementary Fig. 24**). These findings support a model in which cognitive function in autism can be shaped by the presence of a single variant, but also by the cumulative impact of rare and common genetic variants [35].

Finally, we explored PGS associated with traits such as anxiety, MDD, PTSD, and subjective well-being. PGS for anxiety and MDD were significantly higher in autistic people compared to NT. Of particular interest was the PGS derived from a genome wide association study (GWAS) on the question "*Did you feel loved as a child?*", a proxy for social-affective well-being. Autistic people had on average lower PGS for this trait compared to NT people. While the construct captured by this question is broad and somewhat imprecise, this result raises intriguing questions about the relationship between social-affective genetic profiles and autism. It remains to be determined which specific genetic variants underlie this measure and why they appear less frequently in autistic populations. Notably, this PGS is also significantly lower in the autistic people from the SPARK cohort compared to their non-autistic siblings (**Supplementary Fig. 16**). Further research is needed to dissect the biological underpinnings of this and related social-emotional traits, and how they intersect with neurodevelopmental trajectories in autism.

Our neuro-imaging analyses replicate some of the findings reported in previous brain imaging dataset [20,36]. We observed an increased cortical thickness in the STG and the right pericalcarine visual *cortex,* especially in autistic people from cluster C2 and C3 respectively. Our findings are aligned with the literature in the field and provide potential insight into the neural substrates of social and cognitive skill and sensory process in autism [36]. The STG has been repeatedly identified as a key region involved in social perception [37], including gaze direction, facial expressions, speech, prosody, and multisensory integration [37–39]. Hemispheric specialization within the STG is well established: the left STG is primarily involved in semantic and linguistic processing, while the right STG is more engaged in prosody and the pragmatic aspects of language [40]. In line with this functional asymmetry, our results reveal a significant increase in cortical thickness of the left STG in autistic people compared to NT.

Our findings also point to structural variation in the visual *cortex* as a potentially underappreciated feature of autism. Specifically, we observed increased cortical thickness in the visual pericalcarine region among autistic people from cluster C3 with lower IQ, along with a negative correlation between this *cortex* thickness and IQ across autistic people (**Supplementary Fig. 18**). This association has been largely overlooked in prior research, likely due to IQ thresholds in commonly used cohorts such as ABIDE, which predominantly include people with full-scale IQ above 70. The relevance of visual *cortex* involvement is further supported by recent transcriptomic data from postmortem tissue, showing that the primary visual *cortex* exhibits the largest differential gene expression between autistic and neurotypical people [41]. This molecular finding aligns with our neuroimaging data and supports the notion that early sensory *cortices* may be more central to autism biology than previously assumed [42].

Our finding of reduced rsfMRI connectivity between the visual and somatomotor networks aligns with recent studies [43,44] reporting lower connectivity among visual association, somatosensory, and motor networks in autism [43,45]. Decreased somatomotor connectivity has been linked to altered primary sensory processing [44], potentially impairing the ability to filter relevant sensory inputs and generate appropriate motor responses, thereby weakening integration between the somatomotor, salience, and dorsal attention networks [46]. Within the default mode network, we observed hypoconnectivity of the right temporal region, accompanied by widespread hyperconnectivity both within the DMN and between the DMN and other large-scale systems, notably the frontoparietal and visual networks. These findings are consistent with previous reports using independent datasets, including the ABIDE dataset [43,44,47].

Consistent with prior work, our group-level analyses revealed no significant differences in resting-state EEG *alpha* peak frequency or spectral power between autistic and neurotypical people, suggesting no global electrophysiological signature of autism [48]. In contrast, autistic people in cluster C3 showed a significantly lower *alpha* peak frequency compared to NT people. Because *alpha* peak frequency indexes cortical excitability and processing speed, this reduction may reflect slower

neural dynamics or altered thalamocortical coupling in C3 which align with evidence linking reduced *alpha* activity and/or *alpha* peak frequency to cognitive impairment and lower non-verbal IQ in autism [49,50].

The study offers important strengths, though it is also subject to several limitations. A defining feature of this dataset is its comprehensive collection of multiscale data, ranging from WGS to brain imaging and clinical assessments. The inclusion of NT people, undiagnosed relatives, and numerous families with both biological parents enable robust investigation of *de novo* variants and the interplay between rare and common genetic variants within familial contexts. An additional strength is the inclusion of brain imaging data from autistic people with IDD, a population often underrepresented in previous neuroimaging datasets. Another notable advantage is the harmonization of data across multiple sites and across datasets, including LEAP and InovAND. These harmonization protocols are now being adopted by several European and international research initiatives, highlighting the broader utility and reproducibility of our multidimensional approach [51]. However, the collection of such a large multiscale dataset also presents certain limitations. One is related to our clustering analysis, which is constrained by the number of people in each cluster. For instance, we were unable to incorporate adaptive functioning measures, such as the VABS, despite their potential to significantly enhance subgroup differentiation, particularly within clusters C1 and C3. VABS data were available for only approximately 500 people, which we deemed insufficient for reliable subgroup characterization, especially given that some clusters included fewer than 100 people. A second limitation concerns the interpretation of increased cortical thickness observed in the temporal and visual *cortices*. These findings should be approached with caution due to the age range of our cohort, which spans primarily from 6 to 18 years. From a developmental standpoint, the observed increases in cortical thickness may reflect delayed cortical thinning, a process that typically begins before age 6 [52]. To fully understand these developmental trajectories, future studies should aim to include younger people or a follow-up cohort. This would allow for characterization of earlier stages of brain maturation and provide a more comprehensive understanding of neurodevelopment in autism. Rather than interpreting results in terms of absolute increases or decreases in cortical thickness, future research should adopt a growth

chart-based framework, leveraging age-specific developmental norms to better assess deviations from typical cortical development [53–55]. Several ongoing initiatives are now focused on collecting longitudinal data beginning in early childhood [53]. As part of this effort, our consortium will contribute to a new longitudinal cohort PIP (Preschooler Brain Imaging and Behavior Project; https://www.aims-2-trials.eu/pip/) focusing on children aged 3-6 years. This dataset will enable deeper exploration of early developmental trajectories and help clarify the neural mechanisms underlying autism from a much earlier age.

Our results also open new perspectives for understanding the biological and developmental foundations of autism in its multidimensionality. Understanding the complex interplay between genetic variation, brain structure/function, and developmental trajectories in autism remains a major challenge. Comprehensive insights will ultimately require large, multiscale datasets that integrate genomic, neuroimaging, and clinical data across diverse populations. While the current dataset represents a significant step in this direction, enabling analyses across these interconnected levels, further expansion will be essential to capture the full spectrum of variability. Nonetheless, our findings offer preliminary insights into how genetic variations, spanning *de novo*/rare, and common variants, relate to structural/functional brain differences and clinical profiles in autism. By bridging molecular and neural phenotypes with cognitive and behavioral outcomes, this work helps to decipher the heterogeneity of autism and may in future inform the personalized support that each autistic person and their family wants and needs.

Online Methods

Dataset and participants

For the LEAP study, participants were recruited between January 2014 and March 2017 across six European centers: Institute of Psychiatry, Psychology and Neuroscience, King's College London (IoPPN/KCL, UK), Autism Research Centre, University of Cambridge (UCAM, UK), University Medical Centre Utrecht (UMCU, Netherlands), Radboud University Nijmegen Medical Centre (RUNMC, Netherlands), Central Institute of Mental Health (CIMH, Germany) and the University Campus Bio-Medico (UCBM) in Rome, Italy. At each site, an independent ethics committee approved the study. All participants (where appropriate) and their parent/legal guardians provided written informed consents for research and data sharing [17,18]. C07-InovAND participants were recruited between 2007 and 2016 in four French hospitals: Hopital Robert-Debré (Paris), Hopital Albert Chenevier (Creteil), Hopital Alpes Isère (Grenoble) and Hopital Charles Perrens (Bordeaux). For both studies, all participants and their parent/legal guardians provided written informed consent upon inclusion.

Genetics

DNA isolation and sequencing: DNA was extracted from whole blood/saliva, using manufacturer's protocols on a Tecan Freedom EVO®-HSM workstation and a Maxwell® RSC 48 Instrument, and sequenced by the CNRGH using short read Illumina technology at a read depth of 30× on average. A total of 2,560 samples were sequenced, and reads were aligned to GRCh37.

Variant detection: Individual gVCF were called using DeepVariant and genotyped jointly using Glnexus. Detected SNVs/indel of interest were validated by visualization of the CRAM files and then by Sanger sequencing. For SV, we first ran WisecondorX and ClinSV. A total of 2,189 variants were manually visualized and further used to create a random forest model to discriminate true SV from noise. For short tandem repeats (STR), we ran ExpansionHunter using the loci dictionary from stripy.org and validated using targeted PCR. The list of genes carrying variants analyzed in this study is available in **Supplementary Data 1**.

Ancestry: We computed the admixture of the cohort along with the Human Genome Diversity Project. We then predicted the superpopulation (EUR, AFR, EAS, CSA, MID, OCE, AME) with a random forest model using admixture results and known data from the Human Genome Diversity Project.

Polygenic scores: We used summary statistics from different GWAS and used SBayesRC to compute the PGS. All individual PGS were corrected for ancestry using the first 10 principal components.

Statistical analysis: Differences between categorical variables were assessed using the Chi-square test, with effect sizes quantified by Cramer's V. When necessary, Fisher's exact approach was applied to adjust for small sample sizes or sparse data. Comparisons between independent quantitative variables were performed using the Mann-Whitney-Wilcoxon test; effect sizes were estimated using the rank biserial correlation when appropriate. Statistical significance was defined as $p < 0.05$.

For genetic data, associations between the presence of genetic variants and categorical variables were assessed using Fisher's exact test, with false discovery rate (FDR, one-sided independent t-tests adjusted for multiple comparisons using Benjamini–Hochberg correction) correction. Odds ratios (ORs) with 95% confidence intervals (CI95%) were calculated using the NT group or, when appropriate, the combined NT/sibling group as the reference category.

Gene lists: The list of genes used in this study is available on our website GeneTrek (https://genetrek.pasteur.fr/) and in **Supplementary Data 1**. The notion of gene constraint is based on the union of GnomAD v4 for autosomes, GnomAD v2 for X chromosome and sHET [35].

Replication in the SPARK cohort: We replicated the genetic analyses in the SPARK cohort using whole exome sequence (WES) data from the iWES v.3 dataset.

Brain Imaging analyses

MRI Acquisition

For the LEAP cohort, Magnetic resonance imaging (MRI) scans were acquired across five European research sites (University of Cambridge, King's College London, Central Institute of Mental Health Mannheim, Radboud University Nijmegen Medical Centre,

and University Medical Center Utrecht) using a harmonized protocol based on ADNI-2/GO recommendations. High-resolution T1-weighted structural images were acquired using a magnetization-prepared rapid gradient-echo (MPRAGE) sequence (1.2 mm isotropic resolution). Resting-state functional MRI (rsfMRI) scans were acquired using single-echo echo-planar imaging (EPI) with full brain coverage (see **Supplementary Tables 4 and 5** for acquisition parameters). In the InovAND cohort, MRI data were collected at a single site (Robert Debré Hospital, Paris) between 2010 and 2024, using three Philips scanners (1.5T and 3T field strengths). T1-weighted images were acquired with Fast Field Echo (FFE) or Turbo Field Echo (TFE) sequences (voxel size 1 mm isotropic, TR = 7–25 ms, TE = 3–5.6 ms, FOV = 240–260 mm). Updated 3D-TFE protocols with compressed sensing and SENSE acceleration achieved voxel resolutions up to 0.9 mm isotropic. Resting-state fMRI was acquired with single-echo EPI sequences (TR = 2000–2500 ms; TE = 30–35 ms; voxel size = 3 × 3 × 3 mm). Detailed sequence parameters are listed in **Supplementary Table 6 and 7**.

Structural MRI data preprocessing

T1-weighted structural images were processed using FreeSurfer (v6.0.0) Cliquez ou appuyez ici pour entrer du texte. with the full *recon-all* pipeline. Quality control of cortical and subcortical segmentations was performed manually by trained raters blinded to diagnosis and scans failing manual QC were excluded. Cortical thickness (CT) and surface area (SA) were extracted from 68 cortical regions (Desikan–Killiany atlas [25]) and volumes from 14 subcortical regions and lateral ventricles. Cortical thickness and surface area were extracted for 68 cortical regions according to the Desikan–Killiany atlas, while volumes were obtained for 14 subcortical regions and the lateral ventricles. Global metrics, including mean cortical thickness, total surface area, intracranial volume, total gray matter, and cerebrospinal fluid volume, were also derived. For each regional measure, residuals were computed after modeling diagnosis, age, sex, age squared, MRI site, and intracranial volume. Site-related variability was mitigated using ComBat [57] harmonization. Residuals were z-scored relative to neurotypical controls, and group differences were assessed using two-sided Welch's t-tests with Bonferroni correction ($P_{BON} < 0.05$).

fMRI preprocessing and analyses

Preprocessing of rsfMRI data was performed using fMRIPrep (v23.0.0; [58]) and XCP-D (v0.8.0; [59]). Steps included realignment, slice-timing correction, coregistration to the participant's T1-weighted image using FLIRT, normalization to MNI152NLin2009cAsym space, and resampling to 2 mm isotropic resolution. Spatial smoothing was applied using a 6 mm FWHM Gaussian kernel. Physiological and motion confound regression was conducted using the 36p+Despike model. This included removal of six motion parameters and their derivatives, global signal, white matter and CSF signals, their squares, and quadratic terms, followed by bandpass filtering (0.01–0.08 Hz). Volumes with excessive motion were scrubbed when mean framewise displacement (FD) exceeded 0.5 mm or when more than 20% of volumes exceeded the FD threshold. Participants not meeting these motion criteria were excluded. Mean FD values per group are reported in **Supplementary Fig. 19a**. Functional connectivity (FC) was estimated at the ROI level using the Schaefer 100-parcel cortical atlas [60] combined with 15 subcortical ROIs from CIT168 subcortical atlas [61]. For each participant, mean time series were extracted per ROI, and pairwise Pearson correlations were computed to generate a 115 × 115 symmetric connectivity matrix. Correlation coefficients were Fisher z-transformed prior to group-level analyses. FC matrices were standardized by subtracting the median and dividing by the interquartile range (IQR) to ensure comparable scaling across participants. For each connectivity feature, residuals were computed after modeling diagnosis, age, sex, age squared, MRI site. To mitigate site-related variability, ComBat harmonization was applied to all FC features [57]. The efficiency of harmonization was confirmed using one-vs-one SVM classification of acquisition site before and after ComBat (**Supplementary Fig. 19b**). Between-group and between-cluster differences were evaluated for each connection using two-sided Welch's t-tests, with FDR correction ($q < 0.05$). Connections were classified as hyper- or hypoconnected based on the direction of the t-statistic, and network-level summaries were derived by counting altered edges within and between Yeo 7-network systems [43].

Resting state EEG acquisition

Five sites acquired EEG data at baseline. In the LEAP cohort, 4 minutes of resting state EEG were recorded per participant (2 minutes with eyes open and 2 minutes with eyes closed) with 30s blocks alternatively, using 10–20 layout caps with 60–70 electrodes. In the InovAND cohort, all subjects were recorded in resting-state with eyes open and closed for periods of 3 to 5 minutes using a 129-GSN hydrocel Geodesic montage. Sampling was at 1,000 Hz, and the reference electrode was the Cz position.

Resting-state EEG data preprocessing

Preprocessing was done using the MNE-Python software [62], according to the OHBM COBIDAS MEEG good practice recommendations [63]. EEG signals were band-pass filtered at 1–100 Hz (FIR) and notched at 50 Hz to suppress line noise. Robust referencing and bad-channel handling were performed per subject with PyPREP, and bad channels were interpolated with spherical splines before setting a common average reference. We fit ICA using FastICA to the epoched data and classified components with ICLabel [64]; components labeled as non-brain sources (e.g., eye, muscle, heart, line noise, channel noise) were excluded and the cleaned solution was applied to the data. Next, EEG signals were segmented into 4-second epochs. The Autoreject toolbox [65] was used to interpolate or reject bad epochs. Then, we extracted the spectral features across all bands (*theta*, *delta*, *alpha*, *beta,* and *gamma*). Alpha peak frequency was identified as the maximum amplitude within the 6–15 Hz range. Further methodological details are provided in **Supplementary Section 5**.

Statistical Correction: For each feature (power band × brain region × laterality), linear regression was performed to remove effects of age, age-squared, and sex. ComBat harmonization was then applied to the regression residuals to remove site and cohort effects, with age, sex, and age-squared included as covariates. ComBat-adjusted residuals were z-scored based on the mean and standard deviation of the TD control group. Subjects with any feature value exceeding 10 standard deviations (z-score > 10) were excluded as outliers. Between-group differences in corrected PSD measures and alpha peak frequencies were assessed using two-sided Welch's t-tests with FDR correction ($q < 0.05$) applied independently for each cohort and for the combined dataset.

Clustering analyses

We began with a broad set of clinical measures, including IQ subscales (full-scale, verbal, non-verbal), SRS-2 total t-score, Vineland composite and domain scores, RBS-R, and SSP. PCA was conducted after excluding people with missing data and relatives (n = 380) and standardizing all features. The first two components explained most of the variance (>75%), with IQ and SRS contributing most strongly whereas Vineland scores contributed substantially to PC1 but minimally to PC2 (**Supplementary Fig. 21**). Restricting analyses to IQ and SRS also maximized sample size (n = 1,026 vs. ~500 with VABS; **Supplementary Table 12**). For both statistical and practical reasons, clustering focused on these two dimensions, Full methodological details are provided in **Supplementary Section 6**.

To determine the optimal number of clusters, we evaluated k = 2–10 using a multi-criterion approach. NbClust, which computes 26 internal validation indices, identified three clusters as the best partition (10/26 indices), with a secondary peak at k = 8. Additional model-based and partition-based indices (AIC, BIC, ICL, Silhouette coefficient, Variance Ratio Score, Davies–Bouldin index, Pseudo-F statistic) and the Gap statistic also favored k = 3 (**Supplementary Table 13**).

Acknowledgments

The authors thank all the participants of the LEAP and InovAND cohorts for their involvement in the research on autism. This work was funded by Institut Pasteur, Université Paris Cité, the Simons Foundation Autism Research Initiative (SFARI award #240059), the Bettencourt-Schueller Foundation, the GenMed Labex, and AIMS-2-TRIALS, which received support from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777394 for the project AIMS-2-TRIALS. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA and AUTISM SPEAKS, Autistica, SFARI, and the Inception program (Investissement d'Avenir grant ANR-16-CONV-0005). This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant 847818 (CANDY), and from Horizon Europe under grant 101057385 (R2D2-MH). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. This work benefited from the DNA & cell bank core facility, at the ICM-Paris Brain Institute. This work received support from the French government, managed by the National Research Agency (Agence Nationale de la Recherche), under the France 2030 program, reference ANR-23-IAHU-0010. Part of this work was funded by a grant from the Conseil Régional d'Ile de France (grant number EX024087).

Data Availability

The datasets generated and analyzed in this study—including all modalities from the LEAP (EU-AIMS / AIMS-2-TRIALS) and InovAND cohorts are securely stored on ELIXIR-LU servers, part of the European infrastructure for life science information, based at the Luxembourg Centre for Systems Biomedicine (LCSB) and supported by the Luxembourg National Data Service (LNDS) (https://elixir-luxembourg.org/). The LEAP dataset (clinical, cognitive, eye-tracking, neuroimaging, and genetic data) is available via the ELIXIR-Luxembourg data catalog at https://datacatalogue.elixir-luxembourg.org/e/study/ELU-4-4D20BB-1. The InovAND dataset will be made available through the same repository upon publication. Access to both datasets is granted upon reasonable request following review and approval by the Data Access Committee, including scientific leads, ethics experts, and Autism community representatives. Requests should be submitted via the data catalog website with a detailed project proposal describing the intended use of the data and must be aligned with General Data Protection Regulation (GDPR) requirements as well as the AIMS-2-TRIALS consortium or InovAND data sharing policies, respectively.

Code availability

A GitHub repository will be made available upon publication. Analysis scripts can be obtained from the corresponding authors upon reasonable request.

References

1. Heraty, S. *et al.* Bridge-building between communities: Imagining the future of biomedical autism research. *Cell* **186**, 3747–3752 (2023).
2. Sandin, S. *et al.* The Heritability of Autism Spectrum Disorder. *JAMA* **318**, 1182–1184 (2017).
3. Warrier, V. *et al.* Genetic correlates of phenotypic heterogeneity in autism. *Nat Genet* **54**, 1293–1304 (2022).
4. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431–444 (2019).
5. Rolland, T. *et al.* Phenotypic effects of genetic variants associated with autism. *Nat Med* **29**, 1671–1680 (2023).
6. Litman, A. *et al.* Decomposition of phenotypic heterogeneity in autism reveals underlying genetic programs. *Nat Genet* https://doi.org/10.1038/S41588-025-02224-Z (2025) doi:10.1038/S41588-025-02224-Z.
7. Mossink, B., Negwer, M., Schubert, D. & Nadif Kasri, N. The emerging role of chromatin remodelers in neurodevelopmental disorders: a developmental perspective. *Cell Mol Life Sci* **78**, 2517–2563 (2021).
8. Bourgeron, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat Rev Neurosci* **16**, 551–563 (2015).
9. Gandal, M. J. *et al.* Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* **611**, 532–539 (2022).
10. Yap, C. X. *et al.* Brain cell-type shifts in Alzheimer's disease, autism, and schizophrenia interrogated using methylomics and genetics. *Sci Adv* **10**, (2024).
11. Wagstyl, K. *et al.* Transcriptional cartography integrates multiscale biology of the human cortex. *Elife* **12**, (2024).
12. Leyhausen, J. *et al.* Parsing Autism Heterogeneity: Transcriptomic Subgrouping of Imaging-Derived Phenotypes in Autism. *Biol Psychiatry Cogn Neurosci Neuroimaging* https://doi.org/10.1016/J.BPSC.2025.07.001 (2025) doi:10.1016/J.BPSC.2025.07.001.
13. Green, J. Autism as emergent and transactional. *Front Psychiatry* **13**, (2022).
14. Wing, L. & Gould, J. Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. *J Autism Dev Disord* **9**, 11–29 (1979).
15. Gillberg, C. & Fernell, E. Autism plus versus autism pure. *J Autism Dev Disord* **44**, 3274–3276 (2014).
16. Lord, C. *et al.* The Lancet Commission on the future of care and clinical research in autism. *Lancet* **399**, 271–334 (2022).
17. Charman, T. *et al.* The EU-AIMS Longitudinal European Autism Project (LEAP): Clinical characterisation. *Mol Autism* **8**, 1–21 (2017).
18. Loth, E. *et al.* The EU-AIMS Longitudinal European Autism Project (LEAP): design and methodologies to identify and validate stratification biomarkers for autism spectrum disorders. *Molecular Autism 2017 8:1* **8**, 1–19 (2017).
19. Feliciano, P. *et al.* SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. *Neuron* **97**, 488–493 (2018).
20. Di Martino, A. *et al.* The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry 2014 19:6* **19**, 659–667 (2013).

21. Taylor, J. L. *et al.* Paternal-age-related de novo mutations and risk for five disorders. *Nat Commun* **10**, (2019).
22. Schaaf, C. P. *et al.* A framework for an evidence-based gene list relevant to autism spectrum disorder. *Nat Rev Genet* **21**, 367–376 (2020).
23. Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217-234.e4 (2019).
24. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
25. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
26. Ilioska, I. *et al.* Connectome-wide Mega-analysis Reveals Robust Patterns of Atypical Functional Connectivity in Autism. *Biol Psychiatry* **94**, 29–39 (2023).
27. Garcés, P. *et al.* Resting state EEG power spectrum and functional connectivity in autism: a cross-sectional analysis. *Mol Autism* **13**, 1–16 (2022).
28. Constantino, J. N. Social Responsiveness Scale. *Encyclopedia of Autism Spectrum Disorders* 4457–4467 (2021) doi:10.1007/978-3-319-91280-6_296.
29. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886–D894 (2019).
30. Borroto, M. C. *et al.* A Genotype/Phenotype Study of KDM5B-Associated Disorders Suggests a Pathogenic Effect of Dominantly Inherited Missense Variants. *Genes (Basel)* **15**, (2024).
31. Bruel, A. L. *et al.* ITSN1: a novel candidate gene involved in autosomal dominant neurodevelopmental disorder spectrum. *Eur J Hum Genet* **30**, 111–116 (2022).
32. Vitobello, A. *et al.* ADGRL1 haploinsufficiency causes a variable spectrum of neurodevelopmental disorders in humans and alters synaptic activity and behavior in a mouse model. *Am J Hum Genet* **109**, 1436–1457 (2022).
33. Deb, W. *et al.* PSMD11 loss-of-function variants correlate with a neurobehavioral phenotype, obesity, and increased interferon response. *Am J Hum Genet* **111**, 1352–1369 (2024).
34. Guzman, S. G. *et al.* Variants in BSN, encoding the presynaptic protein Bassoon, result in a distinct neurodevelopmental disorder with a broad phenotypic range. *The American Journal of Human Genetics* **112**, 1415–1429 (2025).
35. Urpa, L. *et al.* Evidence for the additivity of rare and common variant burden throughout the spectrum of intellectual disability. *Eur J Hum Genet* **32**, 576–583 (2024).
36. Bedford, S. A. *et al.* Brain-Charting Autism and Attention-Deficit/Hyperactivity Disorder Reveals Distinct and Overlapping Neurobiology. *Biol Psychiatry* **97**, 517–530 (2025).
37. Zilbovicius, M. *et al.* Autism, the superior temporal sulcus and social perception. *Trends Neurosci* **29**, 359–366 (2006).
38. Beauchamp, M. S. The social mysteries of the superior temporal sulcus. *Trends Cogn Sci* **19**, 489–490 (2015).
39. Deen, B., Koldewyn, K., Kanwisher, N. & Saxe, R. Functional Organization of Social Perception and Cognition in the Superior Temporal Sulcus. *Cereb Cortex* **25**, 4596–4609 (2015).

40. Kausel, L., Michon, M., Soto-Icaza, P. & Aboitiz, F. A multimodal interface for speech perception: the role of the left superior temporal sulcus in social cognition and autism. *Cereb Cortex* **34**, 84–93 (2024).
41. Gandal, M. J. *et al.* Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* **611**, 532–539 (2022).
42. Pretzsch, C. M. *et al.* Patterns of Brain Maturation in Autism and Their Molecular Associations. *JAMA Psychiatry* **81**, (2024).
43. Ilioska, I. *et al.* Connectome-wide Mega-analysis Reveals Robust Patterns of Atypical Functional Connectivity in Autism. *Biol Psychiatry* **94**, 29–39 (2023).
44. Buch, A. M. *et al.* Molecular and network-level mechanisms explaining individual differences in autism spectrum disorder. *Nature Neuroscience 2023 26:4* **26**, 650–663 (2023).
45. Oldehinkel, M. *et al.* Altered Connectivity Between Cerebellum, Visual, and Sensory-Motor Networks in Autism Spectrum Disorder: Results from the EU-AIMS Longitudinal European Autism Project. *Biol Psychiatry Cogn Neurosci Neuroimaging* **4**, 260–270 (2019).
46. Wigham, S., Rodgers, J., South, M., McConachie, H. & Freeston, M. The interplay between sensory processing abnormalities, intolerance of uncertainty, anxiety and restricted and repetitive behaviours in autism spectrum disorder. *J Autism Dev Disord* **45**, 943–952 (2015).
47. Ren, P. *et al.* Stratifying ASD and characterizing the functional connectivity of subtypes in resting-state fMRI. *Behavioural Brain Research* **449**, 114458 (2023).
48. Li, Q. *et al.* Intellectually able adults with autism spectrum disorder show typical resting-state EEG activity. *Scientific Reports 2022 12:1* **12**, 1–14 (2022).
49. Javitt, D. C. *et al.* A roadmap for development of neuro-oscillations as translational biomarkers for treatment development in neuropsychopharmacology. *Neuropsychopharmacology 2020 45:9* **45**, 1411–1422 (2020).
50. Dickinson, A., DiStefano, C., Senturk, D. & Jeste, S. S. Peak alpha frequency is a neural marker of cognitive function across the autism spectrum. *European Journal of Neuroscience* **47**, 643–651 (2018).
51. Bloomfield, M. *et al.* European Autism GEnomics Registry (EAGER): protocol for a multicentre cohort study and registry. *BMJ Open* **14**, (2024).
52. Gilmore, J. H., Knickmeyer, R. C. & Gao, W. Imaging structural and functional brain development in early childhood. *Nat Rev Neurosci* **19**, 123–137 (2018).
53. Bethlehem, R. A. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
54. Marquand, A. F., Rezek, I., Buitelaar, J. & Beckmann, C. F. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biol Psychiatry* **80**, 552–561 (2016).
55. Marquand, A. F. *et al.* Conceptualizing mental disorders as deviations from normative functioning. *Mol Psychiatry* **24**, 1415–1424 (2019).
56. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774 (2012).
57. Yu, M. *et al.* Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Hum Brain Mapp* **39**, 4213–4227 (2018).
58. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods 2018 16:1* **16**, 111–116 (2018).

59. Mehta, K. *et al.* XCP-D: A Robust Pipeline for the post-processing of fMRI data. *bioRxiv* 2023.11.20.567926 (2023) doi:10.1101/2023.11.20.567926.
60. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb Cortex* **28**, 3095–3114 (2018).
61. Pauli, W. M., Nili, A. N. & Michael Tyszka, J. A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Scientific Data 2018 5:1* **5**, 1–13 (2018).
62. Gramfort, A. *et al.* MEG and EEG data analysis with MNE-Python. *Front Neurosci* **7**, 70133 (2013).
63. Pernet, C. *et al.* Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research. *Nature Neuroscience 2020 23:12* **23**, 1473–1483 (2020).
64. Mayeli, A., Zotev, V., Refai, H. & Bodurka, J. Real-time EEG artifact correction during fMRI using ICA. *J Neurosci Methods* **274**, 27–37 (2016).
65. Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F. & Gramfort, A. Autoreject: Automated artifact rejection for MEG and EEG data. *Neuroimage* **159**, 417–429 (2017).
66. Kim, S. W. *et al.* Whole genome sequencing analysis identifies sex differences of familial pattern contributing to phenotypic diversity in autism. *Genome Med* **16**, 114 (2024).

Clinical data		Autism	Autism without IDD	Autism with IDD	Relatives	Neurotypicals		Probands with autism *vs.* Neurotypical probands	Probands with autism *vs.* Relative siblings	Neurotypical probands *vs.* Relative siblings	Probands with autism without IDD *vs.* Probands with autism with IDD
Sex ratio	*n*	2061	961	528	2551	875		1729 *vs.* 592	1729 *vs.* 333	592 *vs.* 333	728 *vs.* 465
	M:F	2.92:1	2.31:1	2.91:1	0.88:1	1.17:1	Cramér's V; p-value	**0.21; 4.78e-24**	**0.24; 1.03e-26**	0.06; 5.85e-02	0.02; 4.57e-01
Age at inclusion	*n*	1692	923	276	1339	654		1424 *vs.* 540	1424 *vs.* 202	540 *vs.* 202	699 *vs.* 248
	mean ±sd	13.58 ±8.9	16.24 ±9.2	14.6 ±7.4	42.13 ±15.3	21.43 ±11.9	r; p-value	**-0.31; 9.29e-42**	**-0.06; 8.99e-03**	**-0.24; 3.67e-11**	**-0.09; 6.18e-03**
SRS-2 total t-score ↗	*n*	845	720	108	914	488		639 *vs.* 412	639 *vs.* 123	412 *vs.* 123	541 *vs.* 89
	mean ±sd	71.71 ±12.4	71.39 ±12.5	75.26 ±10.6	47.23 ±9.3	44.9 ±6.3	r; p-value	**0.80; 1.28e-149**	**0.52; 9.87e-47**	**0.09; 4.74e-02**	**0.12; 3.02e-03**
Full scale IQ ↘	*n*	1183	809	345	368	507		985 *vs.* 445	985 *vs.* 88	445 *vs.* 88	637 *vs.* 319
	mean ±sd	84.28 ±31.5	101.07 ±17.7	44.17 ±18.6	112.6 ±15.5	108.8 ±14.5	r; p-value	**-0.43; 5.41e-60**	**-0.29; 7.87e-21**	0.07; 8.52e-02	**-0.82; 1.85e-140**
Verbal IQ ↘	*n*	0 3	805	173	58	331		832 *vs.* 321	832 *vs.* 19	321 *vs.* 19	640 *vs.* 159
	mean ±sd	93.53 ±26.3	101.44 ±21.1	58.34 ±17.3	114.09 ±18.8	108.14 ±15.8	r; p-value	**-0.31; 1.15e-25**	**-0.14; 7.12e-05**	0.09; 9.12e-02	**-0.63; 5.73e-72**
Non verbal IQ ↘	*n*	1166	864	265	351	414		945 *vs.* 374	945 *vs.* 91	374 *vs.* 91	670 *vs.* 238
	mean ±sd	90.71 ±24.2	99.78 ±19.1	60.37 ±11.3	112.19 ±16.2	108.62 ±15.8	r; p-value	**-0.36; 3.76e-38**	**-0.24; 9.00e-15**	0.03; 5.08e-01	**-0.71; 8.00e-102**
VABS II - ABC score ↘	*n*	473	410	63	8	90		453 *vs.* 88	453 *vs.* 2	88 *vs.* 2	391 *vs.* 62
	mean ±sd	70.67 ±14.9	73.09 ±13.4	54.92 ±14.4	93.62 ±18.4	103.52 ±12.7	r; p-value	**-0.59; 2.49e-42**	NA	NA	**-0.42; 8.29e-19**
RBS-R total ↗	*n*	682	554	110	486	347		607 *vs.* 304	607 *vs.* 118	304 *vs.* 118	494 *vs.* 101
	mean ±sd	19.91 ±17.2	18.73 ±17.0	25.57 ±17.0	4.07 ±8.3	1.81 ±3.5	r; p-value	**0.70; 1.14e-99**	**0.45; 2.35e-34**	**0.15; 6.06e-04**	**0.18; 1.45e-05**
SSP total ↘	*n*	505	430	73	208	257		453 *vs.* 237	453 *vs.* 47	237 *vs.* 47	382 *vs.* 70
	mean ±sd	141.59 ±27.8	142.73 ±28.5	136.04 ±23.1	177.7 ±17.4	180.45 ±15.7	r; p-value	**-0.63; 1.65e-56**	**-0.29; 3.40e-10**	**-0.15; 2.70e-02**	**-0.11; 2.40e-02**

Table 1: **Clinical characteristics of the people enrolled in LEAP-InovAND**. Continuous variables were compared using the Mann-Whitney-Wilcoxon test due to non-normal distribution associated with rank-biserial correlation for the effect size, while categorical variables were analyzed using the Chi-square (χ^2) test associated with Cramér's V for the effect size. Statistically significant differences between groups were highlighted where applicable. The interpretation of effect sizes is as follows; absolute value near or below 0.1: small effect, absolute value near 0.3: middle effect, absolute value near or greater than 0.5: large effect. A positive effect size indicates that the first group has a higher score than the second group, whereas a negative effect size indicates that the second group has a higher score than the first. The arrows next to the test names indicate the direction in which symptom severity is highest. IDD: intellectual developmental disorder, IQ: intelligence quotient, NA: not applicable, RBS-R: repetitive behavior scale-revised, SRS-2: social responsiveness scale 2nd edition, SSP: short sensory profile, NT: neurotypical people, VABS II - ABC: vineland adaptative behaviors scale 2nd edition - communication, daily living skills and socialization

FIGURES

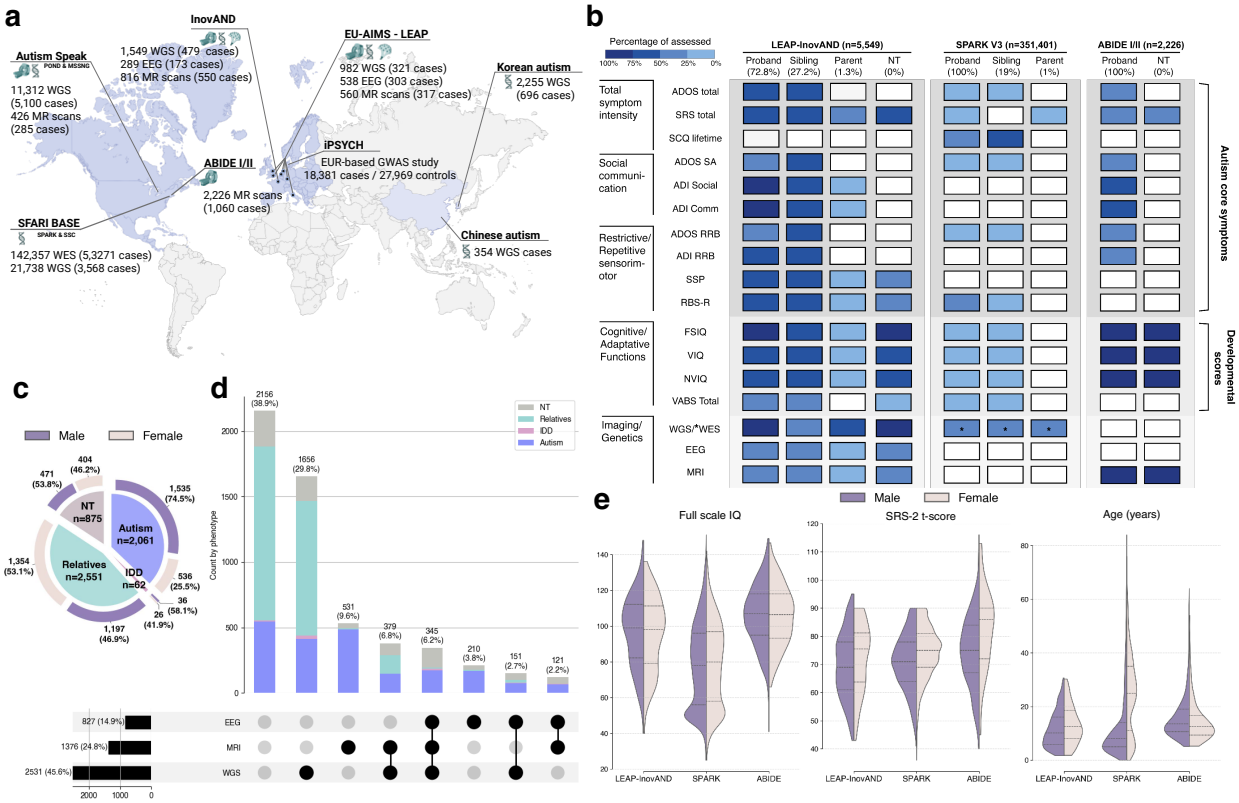

Figure 1. Overview of LEAP-InovAND cohort. a. Worldwide distribution of the major autism cohorts with >300 people with WGS or MRI data. For each dataset, sample sizes are indicated for WGS, and where available, MRI and EEG data, along with the number of autistic people. **b.** Percentage of assessed clinical variables across the LEAP-InovAND cohort, stratified by participant group: proband (autistic people; percentage shown below label), siblings, parents, and neurotypical people (NT). Clinical domains include core autism symptoms, developmental scores, and availability of imaging and genetics. The proband is the person in a family who is first included in the cohort. **c.** Composition of the full study samples (N = 5,549), represented as a nested pie chart showing the proportion of autistic and IDD people, NT, and undiagnosed relatives. Outer rings denote sex distribution (male *vs.* female). **d.** UpSet plot illustrating the overlap in available multimodal data (WGS, MRI, EEG), stratified by group (autism, IDD, NT, relatives). **e.** Distribution of full-scale IQ, Social Responsiveness Scale 2 (SRS-2) total t-scores, and age across our cohort, SPARK, and ABIDE datasets. Fig. inspired by Kim *et al.* [66].

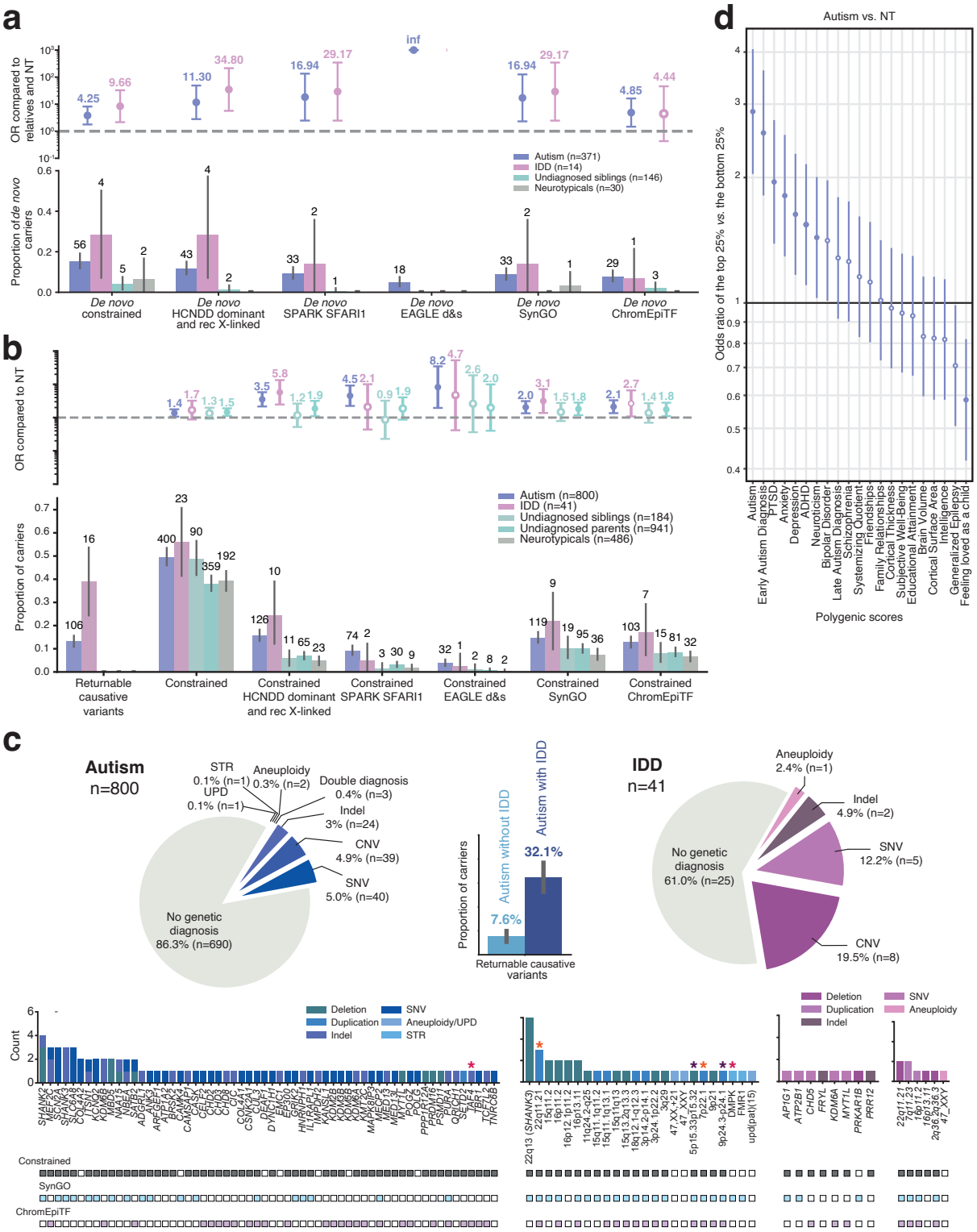

Figure 2. Genetic profile of the participant of the LEAP-InovAND cohort. a. Genetic study of *de novo* variants (deletions, loss of function and deleterious missenses) found in participants with both parents with WGS data. The bottom panel represents the proportion of *de novo* variant carriers according to the different gene lists used. The figure represents the number of carriers per population and per gene list. The top panel represents the odd ratio (OR) based on the proportions of the bottom panel. For *de novo* variations, OR (with CI95%) are calculated by comparing the frequency of *de novo* variants in autistic people/IDD to the pooled frequencies in undiagnosed siblings and NT. **b.** Analysis of returnable genetic diagnosis and rare deleterious variants, specifically deletions and predicted loss-of-function variants (hiURV), in predefined lists of constrained genes of interest. For this analysis, ORs (with CI95%) are computed and corrected (FDR, one-sided independent t-tests adjusted for multiple comparisons using Benjamini–Hochberg correction, difference is significant when the circle is full) compared to that of NT**. c.** Summary of genetic diagnostic findings in the autistic and IDD populations. The upper panel shows

two pie charts illustrating the proportion of the different types of variants classified as diagnostic. Among autistic people with IDD, 4.7% carry variants affecting *SHANK3* or the 22q13 *locus*. These cases include 9 people with 8 distinct variants (8 distinct families). The lower panel displays the distribution of genes and loci with their respective counts, highlighting their overlap with constrained gene lists, SynGO, and ChromEpiTF datasets. Three cases of dual genetic diagnoses were identified; colored stars (in three distinct colors) indicate variants carried by the same person. **d.** PGS for different GWAS compared to NT, representing the OR for the top 25% versus the bottom 25% of the PGS distribution. The difference is significant when the circle is full. ADHD: attention-deficit/hyperactivity disorder, HCNDD: high confidence NDD, IDD: intellectual developmental disorder, NT: neurotypical, PGS: polygenic score, PTSD: post-traumatic stress disorder, STR: short tandem repeats, UPD: uniparental disomy.

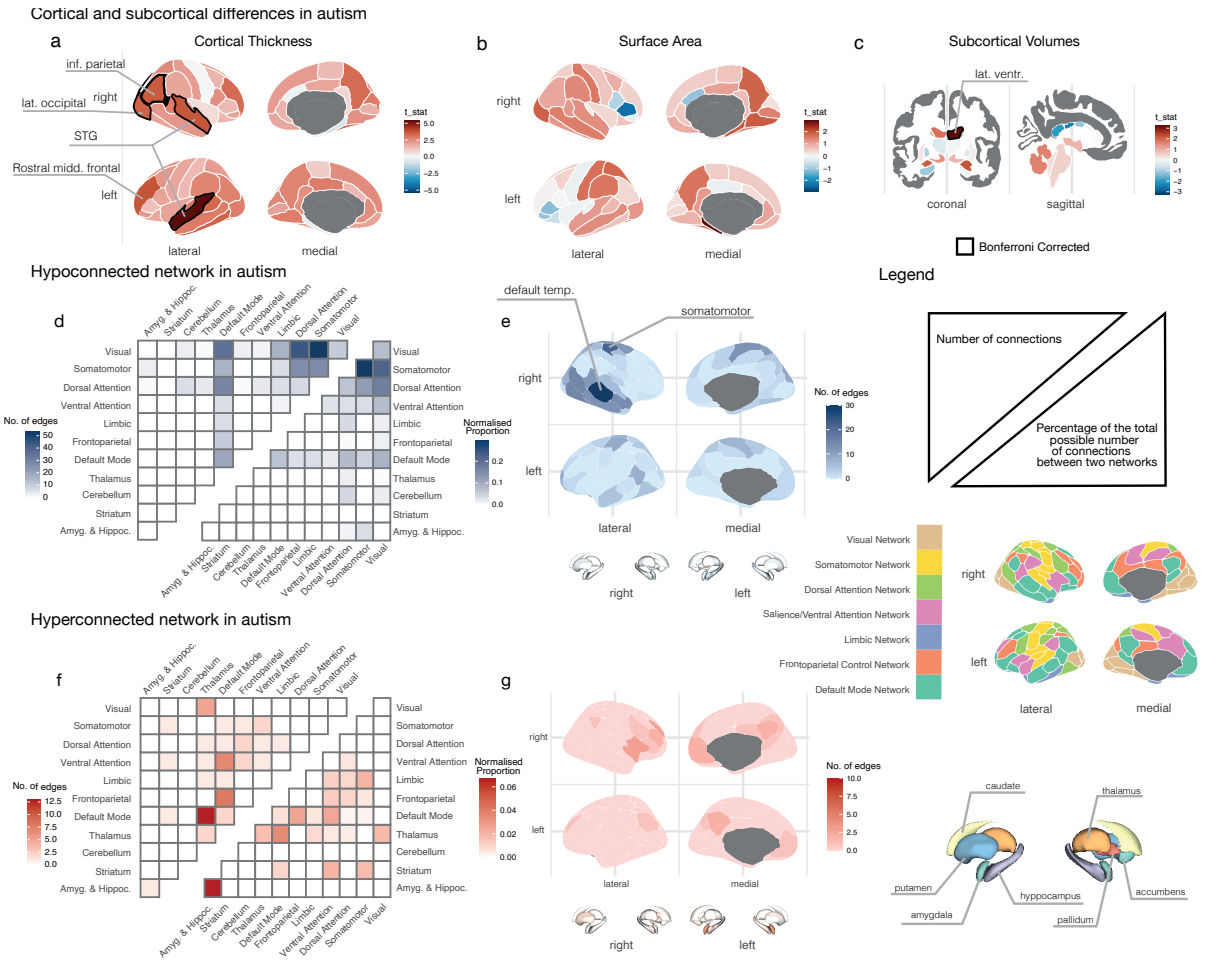

Figure 3. Anatomical and functional cortical and subcortical differences in autistic people compared to neurotypical people (NT). Cortical and subcortical anatomical differences assessed via vertex-wise t-tests on cortical thickness (**a.**), surface area (**b.**) and subcortical volume (**c.**) between autistic (n = 776) and NT people (n = 349). Brain maps show t-statistics, thresholded using Bonferroni correction ($q < 0.05$), projected onto the Desikan-Killiany (cortical) and aseg (subcortical) atlases. Functional connectivity differences across cortical and subcortical regions using the Schaefer-100 parcellation and CIT168 subcortical atlas between autistic (n = 396) and NT (n = 296) people. **d-g.** Connectivity matrix displaying the number of significantly altered connections (edges) between and within functional networks (FDR-corrected). The lower triangle of the matrix shows the proportion of increased connections falling within and between each of the networks, normalized by the total number of possible connections within or between the corresponding networks; the upper triangle shows the raw count of edges for hypoconnectivity (**d.**) and hyperconnectivity (**f.**). Cortical and subcortical maps showing the degree of each region significant in hypoconnectivity (**e.**) and hyperconnectivity (**g.**), colored by the number of altered edges per region. Note: STG: Superior Temporal Gyrus; inf.: Inferior; Mid.: Middle; Temp.: Temporal; Lat. Ventr.: Lateral Ventricle; DMN: Default Mode Network.

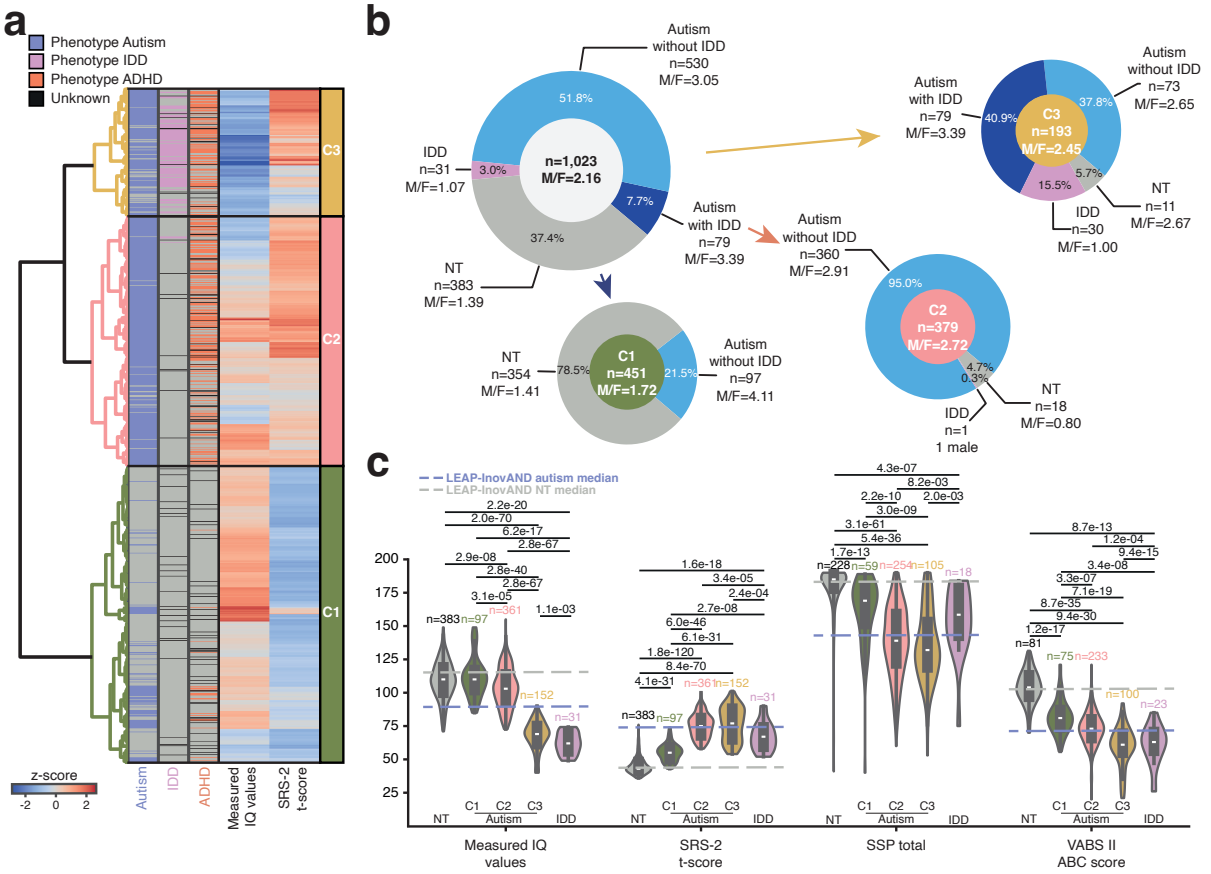

Figure 4. **Hierarchical clustering of 1,023 people based on measured IQ values and SRS-2 t-score**. **a.** Dendrogram generated using hierarchical clustering with Ward's method on standardized behavioral measures, revealing three distinct clusters: C1 (n = 451), C2 (n = 379), and C3 (n = 193). These clusters represent the autism-derived subgroups only and do not include NT or IDD only people in downstream visualizations (*e.g.*, panel **c.**). **b.** Circle plots showing the count and proportion of people within each cluster presenting with specific categorical characteristics, including autism without IDD, autism with IDD, IDD, NT and sex ratio. **c.** Cluster-wise distributions of clinical scores including full-scale IQ, SRS 2 total t-score, short sensory profile (SSP) total score, and Vineland Adaptive Behavior Composite II (VABS-II) ABC score. Blue and grey dashed lines indicate the median scores of autistic people (with or without IDD) and neurotypical (NT) people in all the cohort, respectively. The presented statistics are Mann-Whitney-Wilcoxon test results. ADHD: attention-deficit/hyperactivity disorder, IDD: intellectual developmental disorder, NT: neurotypical.

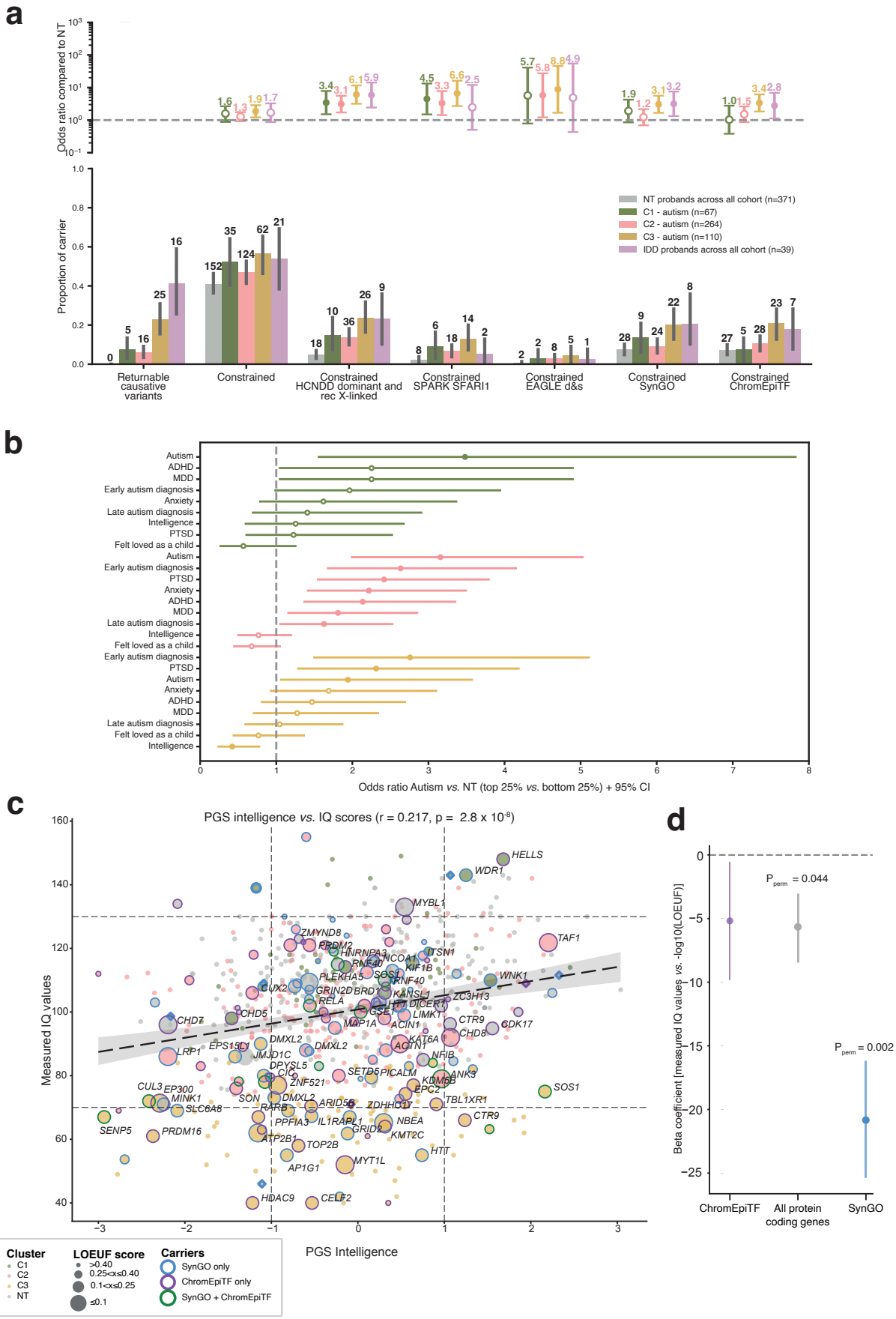

Figure 5. hiURV and common variants analysis of the different clusters. a. Analysis of returnable genetic diagnosis and rare deleterious variants (hiURV), specifically deletions and predicted loss-of-function variants, in predefined lists of constrained genes of interest. The figure shows the number of deleterious variant carriers,

stratified by cluster and gene list. Results are displayed as counts with corresponding 95% CIs. ORs are presented with their 95% CIs; filled circles denote results that remain significant after FDR correction using the Benjamini–Hochberg method. **b.** PGS for different GWAS compared to NT, representing the OR for the top 25% versus the bottom 25% of the PGS distribution. The difference is significant when the circle is full. **c.** A correlation plot illustrating the relationship between measured IQ, the PGS for intelligence and the hiURV found across people in different clusters. The size of each dot represents the lowest LOEUF score, and people within SynGO, ChromEpiTF and both gene lists. The different gene lists are highlighted with outlines. **d.** Beta coefficients and standard errors from linear regression of measured IQ on –log10(LOEUF) for protein-coding genes, ChromEpiTF, and SynGO. For each gene set, permutation p-values are shown for Pearson correlation between constrained gene scores [-log10(LOEUF)] and measured IQ. Statistical significance was assessed using 1,000 permutation iterations with gene-set–specific null hypotheses: random reassignment of gene scores for protein-coding genes. HCNDD: high confidence NDD genes (only dominant inheritances and recessive X-linked in male), SFARI1/SPARK: union of the SFARI1 and SPARK returnable genes, SynGO and ChromEpiTF are described in **Online Methods**. ADHD: attention-deficit/hyperactivity disorder, EAGLE d&s: EAGLE definitive and strong, HCNDD: high confidence NDD, IDD: intellectual developmental disorder, NT: neurotypical, PGS: polygenic score, PTSD: post-traumatic stress disorder.

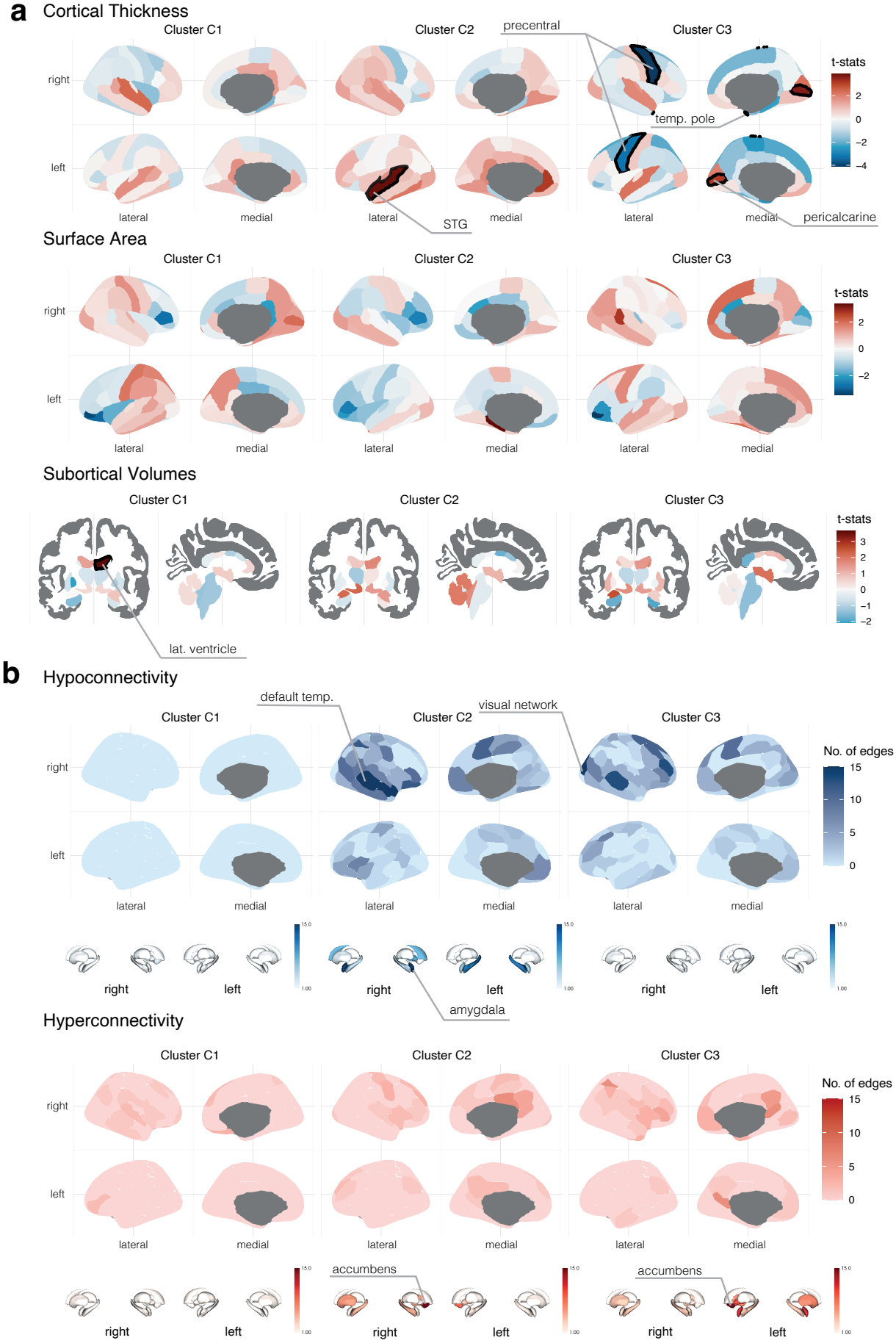
a
Cortical Thickness
Cluster C1
Cluster C2
Cluster C3
precentral
temp. pole
STG
pericalcarine
right
left
lateral
medial
t-stats
2
0
−2
−4
Surface Area
Cluster C1
Cluster C2
Cluster C3
t-stats
2
0
−2
Subortical Volumes
Cluster C1
Cluster C2
Cluster C3
lat. ventricle
t-stats
3
2
1
0
−1
−2
b
Hypoconnectivity
default temp.
visual network
Cluster C1
Cluster C2
Cluster C3
No. of edges
15
10
5
0
amygdala
Hyperconnectivity
Cluster C1
Cluster C2
Cluster C3
No. of edges
15
10
5
0
accumbens
accumbens

Figure 6. Structural and functional MRI differences between autism in the clusters and neurotypical a. Cortical and subcortical anatomical differences between clusters and neurotypicals (C1: n = 303; C2: n = 208; C3: n = 82) were assessed using vertex-wise t-tests on cortical thickness (left) and volumetric comparisons of subcortical structures (right). t-statistic maps are displayed on the Desikan–Killiany (cortical) and aseg (subcortical) atlases, thresholded using false discovery rate (FDR) correction ($q < 0.05$). **b.** Functional connectivity differences between clusters and neurotypicals were analyzed using the Schaefer-100 parcellation and CIT168 subcortical atlas. Brain maps show significant alterations in connectivity strength across regions. Regions exhibiting significant hypoconnectivity (top) and hyperconnectivity (bottom) are colored according to the number of significantly decreased or increased connections per region (unc. $p < 0.05$). Note: Lat.: Lateral.; Temp.: Temporal; STG: Superior Temporal Gyrus.

.

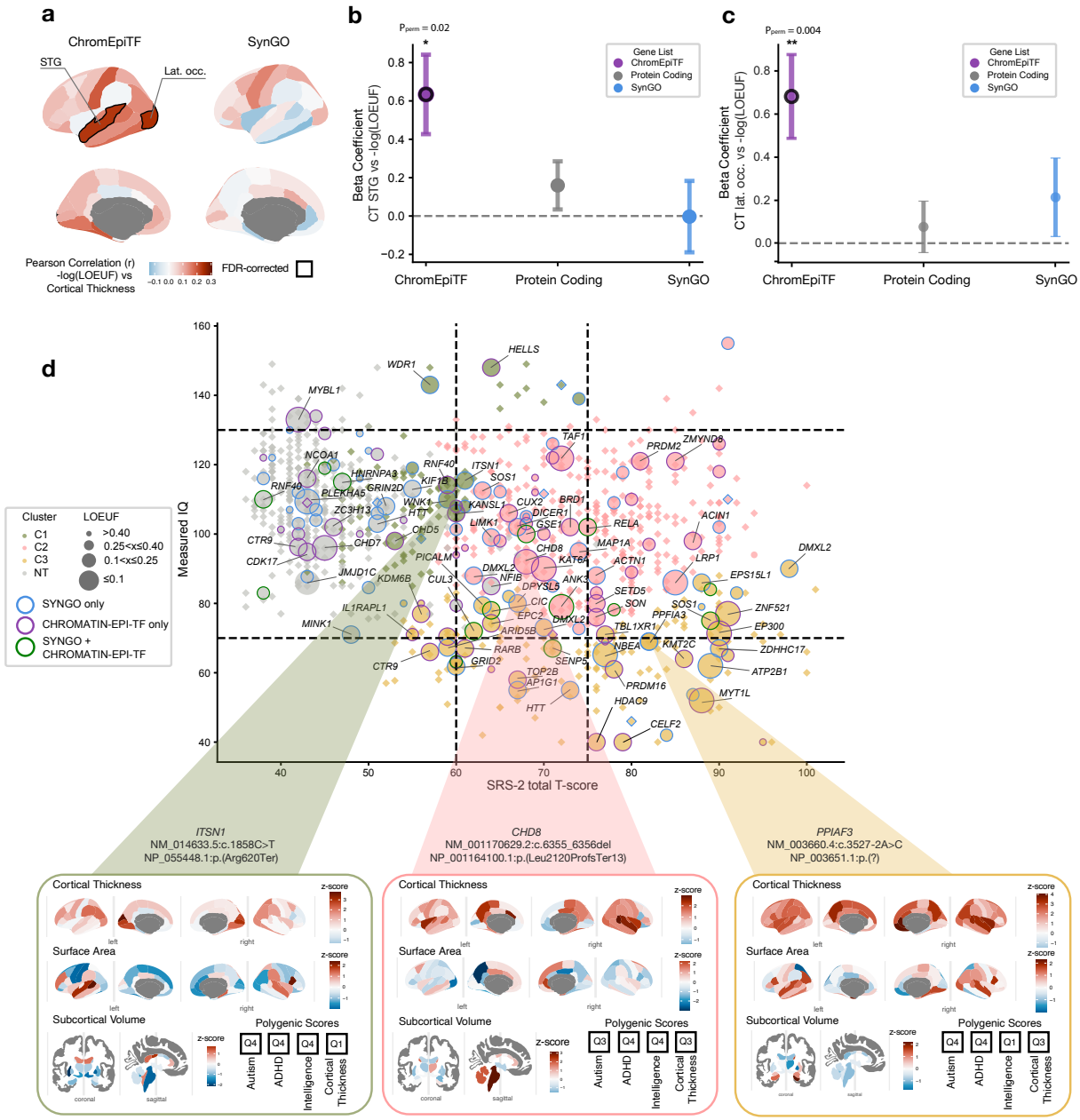

Figure 7. Relationship between genetic variants, brain anatomy and clinical features a. Pearson correlation were computed between constrained gene scores (-log10LOEUF) from SynGO and ChromEpiTF and cortical thickness (CT) in the 68 Desikan–Killiany atlas. LOEUF is a population-based measure of the intolerance of each gene to LoF variants; higher $-\log_{10}$(LOEUF) values indicate stronger constraint. **b-c.** Beta coefficients and standard errors from linear regression of CT on -log10(LOEUF) for protein-coding genes, ChromEpiTF (n = 204 carriers), and SynGO (n = 226 carriers). For each gene sets, permutation p-values are displayed. Pearson correlation between constrained gene scores [-log10(LOEUF)] and CT are illustrated for two representative regions: left superior temporal gyrus (STG) and left lateral occipital cortex (lat. occ.). Statistical significance was evaluated through permutation tests (1,000 iterations) with gene list-specific null hypotheses: random assignment for protein-coding genes. **d.** Scatter plot of measured IQ and SRS-2 total t-score colored by the clustering of the people. People with hiURV in SynGO, ChromEpiTF or both are highlighted with outlines. The size of each dot is proportional to the intolerance to LoF variants (a LOEUF value < 0.6 indicates high intolerance to LoF variants as described in GnomADv4.1). Dashed lines represent the standard clinical thresholds. A focus has been made on three different people (one in each cluster), all three carrying a hiURV in specific genes with their sMRI z-score values and 4 different PGS.