Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Shared rare tandem-repeat expansion burden and autism-enriched developmental TAD-boundary-annotated deletion burden across autism and schizophrenia

Itaru Kushima

kushima.itaru.k0@f.mail.nagoya-u.ac.jp

Nagoya University Graduate School of Medicine https://orcid.org/0000-0002-0018-5178

Ran Wei

Japan Institute for Health Security

Ikuko Natsukari

Yakitsubeno-Michi Clinic

Hidekazu Kato

Nagoya University Graduate School of Medicine

Hiroki Kimura

Nagoya University Graduate School of Medicine

Yoshihiro Nawa

Nagoya University Graduate School of Medicine

Youta Torii

Nagoya University Graduate School of Medicine

Maeri Yamamoto

Nagoya University Graduate School of Medicine

Yu Hayashi

Nagoya University Graduate School of Medicine

Daisuke Mori

Nagoya University

Yuko Arioka

Nagoya University Graduate School of Medicine https://orcid.org/0000-0002-1792-8812

Hiroki Okumura

Nagoya University https://orcid.org/0000-0001-5506-9492

Sawako Furukawa

Tokyo Medical and Dental University (TMDU)

Junya Matsumoto

National Center of Neurology and Psychiatry https://orcid.org/0000-0003-4228-3208

Yuka Yasuda
National Institute of Mental Health
Satsuki Ito
National Center of Neurology and Psychiatry
Yuichiro Watanabe
Niigata University Graduate School of Medical and Dental Sciences
Ryo Morikawa
Niigata University Graduate School of Medical and Dental Sciences
Satoko Takano-Isozaki
Niigata University Graduate School of Medical and Dental Sciences
Jun Egawa
Niigata University Graduate School of Medical and Dental Sciences
Shuken Boku
Niigata University Graduate School of Medical and Dental Sciences
Yasue Horiuchi
Tokyo Metropolitan Institute of Medical Science
Makoto Arai
Tokyo Metropolitan Institute of Medical Science https://orcid.org/0000-0003-3400-9815
Kazuya Toriumi
Tokyo Metropolitan Institute of Medical Science https://orcid.org/0000-0002-8593-3269
Masahide Usami
Japan Institute for Health Security
Mizuki Hino
International Research Institute of Disaster Science, Tohoku University
Daisuke Fujikane
Department of Psychiatry, Gifu University Graduate School of Medicine
Masahito Nakataki
Tokushima University
Shuraku Son
Kyoto University
Toshio Munesue
Kanazawa University
Yugo Matsuo
Graduate School of Frontier Sciences, The University of Tokyo
Masahide Seki
The University of Tokyo, Graduate School of Frontier Sciences https://orcid.org/0000-0001-6091-8698
Yayoi Sekiya
Kyushu University
Shigeru Yokoyama

Kanazawa University

Toshiya Murai

Kyoto University https://orcid.org/0000-0001-6550-8188

Shusuke Numata

Tokushima University

Kazutaka Ohi

Department of Psychiatry, Gifu University Graduate School of Medicine https://orcid.org/0000-0001-9577-9640

Yasuto Kunii

Tohoku University

Hirotaka Kosaka

University of Fukui https://orcid.org/0000-0003-2210-5025

Masao Nagasaki

Division of Biomedical Information Analysis, Medical Research Center for High Depth Omics, Medical Institute of Bioregulation, Kyushu University https://orcid.org/0000-0002-4292-8785

Yutaka Suzuki

The University of Tokyo https://orcid.org/0000-0003-4852-1879

Norio Ozaki

Nagoya University

Masashi Ikeda

Nagoya University Graduate School of Medicine

Masanari Itokawa

Tokyo Metropolitan Institute of Medical Science

Ryota Hashimoto

National Center of Neurology and Psychiatry https://orcid.org/0000-0002-5941-4238

Yosuke Kawai

National Institute of Global Health and Medicine, Japan Institute for Health Security https://orcid.org/0000-0003-0666-1224

Katsushi Tokunaga

Japan Institute for Health Security

Article

Keywords:

Posted Date: June 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9869940/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Yes there is potential Competing Interest. I.K. has received research/grant support from Novartis Pharma, GlaxoSmithKline and Takeda Pharmaceutical Co., Ltd., all outside the submitted work. N.O. has received research support or speakers' honoraria from, or has served as a consultant to, Sumitomo Pharma, Otsuka, Viatris, Eisai, Mochida, Kyowa Pharmaceutical Industry, Nihon Medi-Physics, Nippon Chemiphar, Nippon Boehringer Ingelheim and SUSMED, all outside the submitted work. J.M. has received speakers' honoraria from Sumitomo Pharma, outside the submitted work. Y.Kun. has received research support or speakers' honoraria from, or has served as a consultant to, Takeda Pharmaceutical Co., Ltd., Sumitomo Pharma, Otsuka, Viatris and Lundbeck, all outside the submitted work. M.Ike. has received speakers' honoraria from Meiji Seika Pharma, Otsuka Pharmaceutical, Takeda Pharmaceutical, Lundbeck Japan, Mochida Pharmaceutical, Janssen Pharmaceutica, Mitsubishi Tanabe Pharma, MSD, Eisai, Kyowa Pharmaceutical, Towa Pharmaceutical and Viatris, all outside the submitted work. The remaining authors declare no competing interests.

Abstract

Autism spectrum disorder (ASD) and schizophrenia (SZ) share liability, but how whole-genome sequencing (WGS)-accessible mechanisms partition between shared and disorder-enriched architectures remains unclear. We performed harmonized PCR-free short-read WGS in 11,335 Japanese participants, testing rare genic-proximal tandem-repeat expansion (TRE) outlier burden and rare deletion/duplication burden at developmental cell-type-differential topologically associating domain (TAD)-boundary annotations. With two TRE callers, rare TRE outlier carriage was elevated in both disorders, with no between-disorder heterogeneity. By contrast, rare deletion burden at the aggregate developmental differential-boundary end-point was enriched in ASD (odds ratio (OR) = 1.43, $P = 2.9 \cdot 10^{-5}$), with concordant external support from two independent ASD cohorts (two-cohort inverse-variance-weighted OR = 1.25, two-sided $P = 2.2 \cdot 10^{-3}$). The ASD deletion signal exceeded matched static-boundary expectations, showed a larger per-bin effect at differential than static boundaries, persisted after exon exclusion and remained in joint-layer models. Duplications showed no comparable signal and SZ showed no significant primary boundary enrichment. Contextual rare coding, copy number variation and polygenic risk score analyses recovered established risk layers. These findings support layered cross-disorder architecture with a shared burden-level TRE signal and ASD-enriched deletion burden at developmental TAD-boundary annotations.

Introduction

Autism spectrum disorder (ASD) and schizophrenia (SZ) are highly heritable neurodevelopmental conditions with partially overlapping genetic liability[1]. Although their typical ages of onset differ, large-scale cross-disorder genomic studies have demonstrated shared genetic effects across psychiatric diagnoses, and complementary molecular studies have implicated convergent neuronal and synaptic biology[2–5]. A central unresolved question is therefore not whether ASD and SZ share genetic risk at established common-variant, rare-coding and pathogenic copy number variation (CNV) layers, but rather, how less-characterized whole-genome sequencing (WGS)-accessible mechanisms partition between shared and disorder-preferential liability.

Established genetic layers provide the reference point for this question. Genome-wide association studies (GWAS) have defined highly polygenic liability, most extensively for SZ, and cross-disorder analyses have shown partially shared common-variant effects[2–4, 6]. Large-scale exome studies have identified high-confidence risk genes in ASD and SZ[7–9], and rare or recurrent pathogenic CNVs further support a shared rare damaging axis across neurodevelopmental and psychiatric disorders[10–12]. Yet, these layers primarily capture common single-nucleotide polymorphisms, coding variation and gene-dosage alterations; they do not systematically resolve the two less-characterized WGS-accessible mechanisms tested here: rare tandem-repeat expansions (TREs) and rare structural-variant burden at annotations of developmental three-dimensional (3D) chromatin architecture in matched ASD–SZ designs.

The first of these mechanisms is TRE burden. Although repeat expansions are classically associated with Mendelian neurological disorders, short-read WGS now enables genome-wide screening for rare expansion outliers using approaches such as ExpansionHunter Denovo (EHdn) and STRling[13, 14]. Prior ASD studies have implicated multiple forms of tandem-repeat variation, including rare repeat expansions, de novo tandem-repeat mutations and short tandem repeat (STR) expansions involving neurodevelopmental gene programs[15–17], and genome-wide SZ studies have also reported rare TRE burden[18, 19]. However, these studies have not directly compared ASD and SZ under a harmonized design, and idiopathic SZ evidence has largely been based on European-ancestry (e.g. EHdn-derived analyses). Therefore, a matched ASD–SZ WGS design with independent callers and cross-fitted outlier definitions is needed to determine whether rare TRE burden represents shared liability across disorders.

The second mechanism concerns rare structural-variant burden at annotations of developmentally dynamic 3D chromatin architecture, particularly topologically associating domain (TAD)-boundary annotations. Most GWAS risk variants, including those for psychiatric disorders, map outside protein-coding sequences, and foetal brain chromatin maps have shown that distal regulatory contacts can connect noncoding variations to putative target genes[20, 21]. In principle, structural variants (SVs) can perturb this regulatory topology by disrupting boundary or regulatory elements, thereby altering insulation and enhancer–promoter specificity[22–25]. Prior WGS work in SZ has reported an increased burden of ultra-rare SVs at brain TAD boundaries using bulk Hi-C-derived maps, establishing a boundary-disruption hypothesis[26]. However, bulk-derived maps average chromatin structure across the mixed cell populations present in each sampled tissue and provide only tissue- or stage-level snapshots; they cannot determine whether rare SV burden localizes to boundaries that are dynamically remodelled across developmental cell states—indexed by brain region, developmental stage and cell type—rather than to generic boundaries shared across cell states. This distinction is important because neurodevelopmental liability may arise from perturbing architectural elements that are actively remodelled during lineage specification and cell-type maturation. A single-cell developmental 3D genome atlas of the human brain now enables rare SV burden to be tested at such developmentally dynamic, cell-type-resolved boundaries[27]. Whether rare SV burden preferentially localizes to these boundary annotations, and whether the signal is shared across ASD and SZ or disorder-preferential, remains unclear.

Here, we present a harmonized WGS study of ASD and SZ in Japanese cohorts (N = 11,335), generated using PCR-free Illumina short-read sequencing and analysed under a unified alignment, variant-calling, quality control and control framework. We focused on two primary WGS-accessible mechanisms: rare genic-proximal TRE outlier burden, assessed with two independent callers and cross-fitted outlier definitions, and rare deletion and duplication burden at developmentally dynamic, cell-type-resolved TAD-boundary annotations in the human brain. Rare coding-variant, CNV, gene-discovery and polygenic risk score (PRS) analyses were included as contextual reference layers rather than separate primary end-points. This matched ASD–SZ design enables a direct test of whether less-characterized WGS-

accessible mechanisms are shared across disorders or show disorder-preferential architecture in Japanese cohorts.

Results

Analytic overview

We analysed WGS data from 11,335 Japanese participants, comprising 597 individuals with ASD, 762 with SZ, 8,610 population-based controls and 1,366 family members contributing to family-based analyses (Fig. 1). The results are organized around two primary WGS-accessible end-points: rare genic-proximal TRE outlier burden, and rare deletion/duplication burden at developmentally dynamic, cell-type-resolved TAD-boundary annotations. Contextual analyses evaluated rare coding and CNV burden, de novo and gene-discovery evidence, PRS portability and joint-layer models.

Case–control analyses were performed in pedigree-aware, genetically independent Japanese-ancestry subsets with complete sex and principal component (PC) covariates (PC1–PC10). TRE burden, rare protein-truncating variant (PTV) burden and PRS analyses used the primary case–control subset (Controls 8,442; ASD 508; SZ 634), whereas rare CNV and TAD-boundary annotation analyses used the corresponding CNV/TAD case–control subset after exclusion of high-burden CNV outlier samples (Controls 8,372; ASD 505; SZ 627). Family-based de novo analyses used the trio subsets summarized in Table 1. All case–control regressions were adjusted for sex and PC1–PC10; TRE burden models were additionally adjusted for sequencing depth, and TAD-boundary models for SV-type-specific total span and gene-overlap burden. Module-specific denominators, statistical tests, multiple-testing procedures and sensitivity analyses are summarized in Supplementary Table 1.

Shared rare genic-proximal TRE outlier burden in ASD and SZ

We tested rare genic-proximal TRE outlier burden in the primary pedigree-aware case–control subset (N = 9,584; Controls 8,442; ASD 508; SZ 634) using two complementary short-read callers: EHdn, which detects anchored in-repeat reads without requiring a predefined locus catalogue, and STRling, which discovers and genotypes STR loci genome-wide. To avoid reuse of controls for both threshold definition and association testing, outlier and rarity thresholds were defined by fivefold cross-fitting in training controls and applied to held-out samples exactly once. We pre-specified carriage of at least one rare TRE outlier as the primary end-point because EHdn and STRling differ in locus opportunity, and treated per-individual outlier counts as a secondary end-point.

Both callers showed elevated rare TRE outlier carriage in ASD and SZ relative to controls (Fig. 2a). In **EHdn, rare TRE outlier carriage was increased in ASD (odds ratio (OR) = 1.37, 95% confidence interval (CI) 1.07–1.75, P = 0.014) and SZ (OR = 1.61, 95% CI 1.30–1.99, P = 1.0 ´ 10^{-5}**; Fig. 2b). **STRling showed concordant enrichment in the same case–control subset (ASD OR = 1.38, 95% CI 1.07–1.77, P = 0.012; SZ OR = 1.46, 95% CI 1.17–1.82, P = 7.9 ´ 10^{-4}). Secondary Poisson count models were directionally consistent for both callers (**Fig. 2c; Supplementary Table 2). Among 321 individuals in

whom both callers detected at least one rare TRE outlier, 188 (58.6%) shared at least one overlapping outlier locus, with 96% canonical-motif concordance among overlapping pairs. Direct ASD-vs-SZ comparison showed no significant between-disorder difference in rare TRE outlier carriage for either caller, and Cochran's Q testing detected no heterogeneity between the ASD-vs-control and SZ-vs-control effect sizes (I^2 = 0%; Supplementary Table 3). Together, these results support shared rare genic-proximal TRE burden across ASD and SZ rather than a disorder-skewed TRE signal.

To evaluate orthogonal support for the short-read TRE calls, we assessed selected case probands carrying rare TRE outlier calls using two independent long-read sequencing platforms: PacBio HiFi (PacBio) in six ASD probands and Oxford Nanopore Technologies PromethION (ONT) in 60 probands (33 ASD and 27 SZ). IGV-based motif-concordance inspection supported 73 of 82 evaluated short-read rare TRE outlier calls across both platforms (PacBio 6 of 6 (100%); ONT 67 of 76 (88.2%); combined 89.0%; Supplementary Table 4; Supplementary Fig. 1), providing motif-concordant long-read support for the evaluated short-read calls.

ASD-enriched rare deletion burden at developmental cell-type-differential TAD-boundary annotations

Next, we tested whether rare deletions and duplications preferentially overlap published Level 2 (L2) developmental cell-type-differential TAD-boundary annotations from the Heffel et al.[27] single-cell 3D chromatin atlas of the developing human brain, reasoning that these annotations index dynamic architectural elements not resolved by bulk-boundary maps. The primary architecture-level end-point was rare deletion or duplication burden at the aggregate Diff_any union, comprising 4,980 unique 25-kb bins differential in at least one of the 10 analysed L2 boundary classes. Class-level L2 enrichment was treated as secondary localization, and non-differential static boundary bins served as a specificity comparator. Discovery analyses used the Japanese WGS CNV/TAD case–control subset and the B' gene-overlap-adjusted model, which adjusted for sex, PC1–PC10, SV-type-specific total span and gene-overlap burden. External support was evaluated in an extended Japanese case–control array comparative genomic hybridization (arrayCGH) cohort[10] and the MSSNG ASD WGS cohort[28].

In the discovery WGS cohort, rare deletion burden at the aggregate Diff_any primary end-point was significantly enriched in ASD (B' logistic OR = 1.43, 95% CI 1.21–1.70, Wald P = 2.9×10^{-5}; Benjamini–Hochberg false discovery rate (BH-FDR)[29] q = 1.2×10^{-4} across the four primary aggregate tests; Fig. 3a; Supplementary Table 5, Panel A). The other three primary aggregate tests were null: ASD duplications (OR = 0.95, P = 0.55), SZ deletions (OR = 1.12, P = 0.23) and SZ duplications (OR = 0.96, P = 0.62), indicating an ASD-enriched, deletion-specific architecture-level signal. Sensitivity analyses under Patterns B and C were broadly supportive, although Pattern C attenuated the class-level pattern (Supplementary Table 5, Panels B and C). Secondary class-level localization showed broad distribution rather than confinement to a single lineage: 9 of 10 L2 classes reached BH-FDR <0.05 in ASD deletion analyses (Fig. 3b; Supplementary Table 5, Panel A).

The aggregate ASD deletion signal was externally supported in two independent ASD cohorts. At the Diff_any end-point, the Japanese arrayCGH and MSSNG WGS cohorts both showed positive effects (arrayCGH OR = 1.20, P = 0.019; MSSNG OR = 1.65, P = 0.015), with a two-cohort inverse-variance-weighted (IVW) estimate of OR = 1.25 (two-sided P = 2.2×10^{-3}; Fig. 3a; Supplementary Table 6, Panel A). At the secondary class level, all nine discovery-positive classes were directionally concordant across the external cohorts, and seven met the one-sided BH-FDR <0.05 replication threshold in the two-cohort IVW meta-analysis. Hippocampal (HPC) Exc-DG and HPC Inh-MGE remained directionally positive but did not pass the confirmatory threshold.

Next, we asked whether the ASD deletion enrichment was specific to developmental cell-type-differential boundaries, as opposed to reflecting generic boundary overlap. In discovery-internal matched-static resampling using the same 4,980-bin Diff_any universe, aggregate differential-boundary enrichment exceeded property-matched static-boundary expectations at both the per-individual level (observed OR = 1.43 vs. matched-static median OR = 1.30; empirical P = 0.016) and event level (observed event-level OR = 2.12 vs. matched-static mean OR = 1.42; empirical P = 4.0×10^{-4}; Fig. 3c; Supplementary Table 6, Panel B). In the external two-cohort IVW specificity analysis, Diff_any bins showed per-bin enrichment for rare ASD deletions (IVW OR = 1.25 per overlapping bin, 95% CI 1.08–1.44, one-sided P = 1.1×10^{-3}), whereas static bins showed only a minimal per-bin effect (IVW OR = 1.015, 95% CI 1.004–1.027; Fig. 3d; Supplementary Table 6, Panel C). These analyses support preferential localization of the ASD deletion signal to genomic intervals annotated as developmental cell-type-differential TAD boundaries.

The boundary-annotated deletion signal was not explained solely by direct coding-exon disruption. In exon-exclusion analyses, exon-free boundary-bin burden was enriched in discovery (OR = 1.47, BH-FDR = 0.016) and showed one-sided replication support in MSSNG (OR = 1.99, Holm-adjusted one-sided P = 1.1×10^{-2}). In the joint model including exon-free and exon-overlapping boundary-bin burden simultaneously, the exon-free component retained one-sided replication support in MSSNG (OR = 1.74, Holm-adjusted one-sided P = 0.030), with directional but non-significant support in discovery (OR = 1.33, P = 0.075; Supplementary Table 5, Panel D; Supplementary Table 6, Panel D). These results indicate that the ASD deletion signal is not simply a subset of direct coding-exon disruption.

We also evaluated ASD–SZ asymmetry at the same architecture-level end-point. Rare deletion exposure at the aggregate Diff_any union was higher in ASD than in SZ (B' logistic OR = 1.31, 95% CI 1.05–1.64, Wald P = 0.017; BH-FDR q = 0.034 across two ASD-vs-SZ tests), whereas duplications showed no difference (OR = 1.01, P = 0.95). The external arrayCGH ASD-vs-SZ analysis was directionally consistent but not significant, and MSSNG lacked SZ samples. Thus, the aggregate ASD>SZ deletion contrast is best interpreted as a discovery-level cross-disorder asymmetry with directional but incomplete external support (Supplementary Table 6, Panel E).

At the domain scale, ASD deletions contributing to the Diff_any-boundary signal preferentially implicated constrained genes (Genome Aggregation Database (gnomAD)[30] loss-of-function observed/expected upper bound fraction (LOEUF) <0.35) within flanking Diff_any-bounded TAD-scale regulatory domains. In

a fixed-effect three-cohort IVW analysis, the constraint OR was 1.26 per constrained gene (two-sided P = 3.5×10^{-3}; Fig. 3e; Supplementary Table 6, Panel G). In an external two-cohort IVW analysis, the constrained-vs-unconstrained within-domain contrast and direct-overlap-exclusion sensitivity supported the same interpretation, and cutoff sensitivities using LOEUF <0.6 and the probability of being loss-of-function intolerant (pLI) >0.9 were directionally consistent (Supplementary Fig. 2; Supplementary Table 6, Panels G–H). Fig. 3f shows representative ASD deletion loci near constrained neurodevelopmental genes, including SMARCA2, NFIB and AUTS2, illustrating deletion–differential-boundary–gene configurations rather than formal locus-level gene discovery.

Additional robustness and comparator analyses supported the architecture-level interpretation. Standard and strict leave-one-class-out analyses, top 1–5 high-burden sample exclusion and top 1–5 deletion-event exclusion all retained the aggregate ASD-DEL signal (Supplementary Table 6, Panel L). As a complementary comparator rather than a scale-matched null, rare CNV burden at 40-kb TAD-boundary bins from Halvorsen et al.[26] was analysed in parallel for adult and foetal cortical boundaries. ASD-DEL estimates were directionally positive but non-significant (adult OR = 1.35, P = 0.15; foetal OR = 1.44, P = 0.087), whereas SZ-DEL and DUP signals were null. ASD–SZ heterogeneity was non-significant for adult DEL and nominally suggestive but not corrected-significant for foetal DEL (adult OR_diff = 1.41, P = 0.23; foetal OR_diff = 1.77, P = 0.058; Supplementary Table 6, Panel M). Thus, the Halvorsen bulk cortical comparator provided directional but non-confirmatory support and did not reproduce the primary developmental Diff_any signal at the current sample size and annotation resolution, although power and annotation-resolution differences may contribute to this contrast. Distance-window, static-bounded and cohort-composition robustness analyses are reported in Supplementary Table 6, Panels I–K. Aggregate Diff_any-DEL boundary burden was further evaluated within a joint-layer logistic regression alongside TRE, rare coding, CNV, non-allelic homologous recombination (NAHR)-CNV and PRS layers (see joint-layer integration, below).

These robustness analyses indicated that the aggregate ASD Diff_any-DEL signal was not attributable to direct exon loss, a single L2 boundary class, high-burden CNV outlier samples or a small number of influential deletion events. Together, these analyses indicate ASD-enriched rare deletion burden at genomic intervals annotated as developmental cell-type-differential TAD boundaries, supported by external ASD cohorts, preferential Diff_any localization, exon-exclusion analyses, discovery-level ASD>SZ asymmetry and constrained-domain annotation.

Contextual reference layers and joint-layer integration

Next, we evaluated established rare damaging and common-variant layers as contextual reference analyses rather than separate primary end-points. Rare-variant burden was tested in three externally defined risk-gene sets—ASD genes from Fu et al.[9], SZ genes from Chick et al.[31] and neurodevelopmental disorder (NDD) genes from Fu et al.[9]—using Firth logistic regression[32] adjusted for sex and PC1–PC10. Rare PTV analyses used the primary case–control subset (ASD n = 508, SZ n = 634, Controls n = 8,442), whereas rare CNV analyses used the corresponding CNV/TAD case–control subset after excluding high-

burden CNV outlier samples (ASD n = 505, SZ n = 627, Controls n = 8,372). For rare PTVs (gnomAD[30] allele frequency (AF) <0.001; LOEUF <0.6), disease-aligned enrichments were observed across externally defined gene sets (Fig. 4), with the strongest signal for SZ-discovered genes among SZ cases (1.6% vs. 0.2%; adjusted OR = 8.75, 95% CI 3.92–19.57; P = 1.2×10^{-7}; BH-FDR <0.001). Sensitivity analyses using stricter LOEUF thresholds preserved effect directions with larger effect sizes. For rare CNVs, ASD gene-set deletions were enriched in ASD (OR = 7.60; P = 5.5×10^{-4}), and NAHR-mediated genomic disorder CNVs were enriched in both disorders, with the strongest signal for SZ-associated loci in SZ cases (OR = 10.73, 95% CI 5.57–20.70; P = 1.4×10^{-12}). De novo carrier rates were descriptively higher in ASD but did not survive FDR correction in ASD–SZ comparisons (Supplementary Table 7; Supplementary Fig. 3). These analyses recovered the expected shared rare damaging axis across ASD and SZ.

Complementary gene-discovery integrations placed the Japanese WGS data within established rare-coding frameworks. For ASD, TADA-based integration[33] across external and study-specific de novo single nucleotide variant (SNV)/insertion/deletion (indel) and CNV evidence yielded 98 genes at FDR ≤ 0.001 and 267 genes at FDR ≤ 0.05 in the combined evidence framework (Supplementary Fig. 4a; Supplementary Table 8; Supplementary Fig. 5). For SZ, adding the Japanese WGS case–control stratum and de novo stratum to the Chick et al.[31] rare-coding framework yielded 15 exome-wide-significant genes and 30 genes at BH-FDR <0.05, including ZMYM2 newly reaching exome-wide significance after Japanese-stratum addition (Supplementary Fig. 4b; Supplementary Table 9). These analyses were used to contextualize the WGS findings within broader rare-variant discovery resources, not as independent primary end-points.

We also evaluated common-variant PRS portability for SZ and ASD in the same Japanese cohort under matched genotype quality control (QC), covariates and controls using PRS-CSx[34]. For SZ, trans-ancestry discovery resources were available (PGC3 wave3 EUR + EAS; N» 155,000)[6], whereas for ASD, only European-ancestry summary statistics were available (iPSYCH/PGC; N» 46,000)[35]. SZ PRS was robustly associated with case status (OR per SD = 1.90, 95% CI 1.75–2.07; P = 2.61×10^{-49}), with substantial predictive gain (DR^2 = 0.062; DAUC = 0.111) and strong decile-based risk stratification (top-decile OR = 7.31, P = 1.64×10^{-18}). By contrast, ASD PRS showed little detectable portability (OR per SD = 1.05, P = 0.31; DR^2 = 0.0003; top-vs-bottom decile OR = 1.26, P = 0.30). This within-cohort contrast is consistent with known ancestry-related limitations of PRS portability[36] and indicates that portability is strongly influenced by the ancestry composition, diversity and sample size of the available discovery resources (Supplementary Fig. 6; Supplementary Table 10).

Finally, we fitted joint-layer logistic models as an integrative sensitivity analysis to test whether the primary TRE and boundary-annotated deletion signals were retained after adjustment for established genetic risk layers (Fig. 5; Supplementary Table 11). In ASD, aggregate Diff_any TAD-boundary deletion burden remained associated after joint adjustment for rare TRE carriage, ASD PRS, rare ASD-gene PTV burden, exonic ASD-gene deletion burden and ASD-associated NAHR-mediated CNV carrier status (full-model Diff_any-DEL OR = 1.39–1.40 across callers; added-last LRT P = 4.2×10^{-5} for EHdn and 3.0×10^{-5}

for STRling; EHdn progression shown in Fig. 5a, with full caller-specific estimates in Supplementary Table 11). In SZ, the corresponding Diff_any-DEL estimate was smaller and did not reach significance (full-model OR» 1.12; LRT P = 0.22; Fig. 5b). A shared-control multinomial logistic model formally contrasted the aggregate Diff_any-DEL coefficient and supported a larger ASD effect (b_DiffAny,ASD – b_DiffAny,SZ = 0.24, two-sided Wald P = 0.028; Fig. 5d), with stable estimates in caller-specific and TRE-adjusted sensitivity models. The ASD Diff_any signal also remained associated after simultaneous adjustment for static-boundary deletion burden (Diff_any OR = 1.31, P» 6–7 ´ 10^{-3}; Static OR» 1.01; Fig. 5e). Rare TRE carriage remained elevated in both disorders after multi-layer adjustment, whereas disorder-tailored PRS, PTV, CNV and NAHR-CNV layers showed the expected disease-aligned effects (Fig. 5c).

Together, the contextual and joint-layer analyses indicate that the primary TRE and developmental TAD-boundary signals are not simply recapitulations of established coding, CNV or common-variant risk layers.

Discussion

In this study, we used harmonized Japanese WGS to compare ASD and SZ under a matched cross-disorder framework focused on two less-characterized WGS-accessible mechanisms. Rare genic-proximal TRE outlier burden was elevated in both disorders, supporting a shared burden-level TRE signal. By contrast, rare deletion burden at developmental cell-type-differential TAD-boundary annotations showed an ASD-preferential pattern, with external ASD support, preferential Diff_any localization, exon-exclusion sensitivity and retention in joint-layer models. Established rare coding, CNV, gene-discovery and PRS analyses served as contextual reference layers, supporting the interpretation that the primary TRE and TAD-boundary signals were not simply restatements of known coding, CNV or common-variant risk architecture.

The TRE results support a cross-disorder burden-level signal. Using two independent short-read frameworks with distinct assumptions—EHdn, which detects anchored in-repeat reads without a predefined catalogue, and STRling, which discovers and genotypes STR loci genome-wide—elevated rare TRE outlier carriage was observed in both ASD and SZ[13, 14]. The use of cross-fitted thresholding reduced the risk of control-reuse inflation, and direct ASD–SZ comparison and Cochran's Q testing showed no detectable between-disorder heterogeneity. This supports a shared rather than disorder-skewed burden-level TRE signal. Concordant enrichment across callers and long-read motif-concordance support for the evaluated short-read calls further support this interpretation, although the long-read assessment was performed in selected case carriers rather than in a population-random validation sample.

These findings extend prior ASD and SZ TRE studies by directly comparing both disorders in a harmonized Japanese WGS design. Prior ASD studies have implicated multiple forms of tandem-repeat variation, including rare repeat expansions, de novo tandem-repeat mutations and STR expansions involving neurodevelopmental gene programs[15–17], and idiopathic SZ studies have reported rare TRE

burden primarily in European-ancestry EHdn-derived analyses[18, 19]. Our contribution is to show concordant TRE burden across ASD and SZ using two independent callers, cross-fitted outlier definitions and orthogonal long-read support. At the same time, the TRE signal should be interpreted at the aggregate burden level. Short-read approaches are sensitive to local mappability and alignment context and do not uniformly resolve expanded allele lengths or identify the specific causal repeat loci or mechanisms underlying the burden shift[13, 14].

The second primary finding was ASD-enriched rare deletion burden at developmental cell-type-differential TAD-boundary annotations. Prior WGS work established a boundary-disruption hypothesis in SZ using bulk Hi-C-derived brain TAD boundaries[26]. By using the Heffel developmental single-cell 3D genome atlas, our analysis tested whether rare SV burden localizes to developmentally dynamic, cell-type-resolved boundary elements rather than to generic boundary bins[27]. The ASD deletion signal was detected at the aggregate Diff_any end-point, externally supported in Japanese arrayCGH and MSSNG ASD cohorts, and distributed across multiple L2 boundary classes rather than confined to a single lineage[10, 28]. This pattern is consistent with the idea that ASD liability may involve rare deletions overlapping architectural annotations active during developmental lineage specification and cell-type maturation.

Several analyses support the specificity of the TAD-boundary signal. Matched-static resampling indicated that the ASD deletion enrichment exceeded property-matched static-boundary expectations, and external Diff_any-vs-Static testing showed a substantially larger per-bin effect at developmental differential than at static boundaries. Exon-exclusion analyses further showed that the signal persisted when boundary bins overlapping protein-coding exons were removed and retained one-sided replication support in MSSNG after joint modelling of exon-free and exon-overlapping boundary-bin burden. Because the same deletions can also remove regulatory elements, alter local dosage and change the gene context of flanking domains, these analyses alone do not demonstrate chromatin-boundary dysfunction in carrier cells. However, they argue against a simple explanation in which the association is merely a subset of direct protein-coding exon disruption. Together with direct-overlap exclusion and aggregate robustness analyses excluding individual L2 classes, high-burden samples and high-burden deletion events (Supplementary Table 6, Panels H and L), these results support the interpretation that the ASD Diff_any signal is not driven by direct exon loss, a single boundary class or a small number of influential CNV carriers.

The deletion specificity is biologically plausible but should not be overinterpreted. Deletions can directly remove boundary-associated sequence elements, CTCF-bound insulation elements or adjacent regulatory sequences, thereby changing local chromatin organization and enhancer–promoter specificity[22–25]. Duplications can also affect 3D genome organization, including through neo-TAD formation and altered domain architecture[22–25], but such effects may not be captured by a simple boundary-overlap burden framework. Therefore, in this study, the absence of a comparable duplication

signal supports deletion-specific enrichment under the present annotation and modelling strategy, but does not exclude other duplication-mediated 3D genome mechanisms.

At the domain scale, ASD deletions overlapping Diff_any-boundary annotations were associated with increased constrained-gene burden within flanking Diff_any-bounded TAD-scale regulatory domains in a fixed-effect three-cohort IVW analysis. This domain-level pattern was supported by an independent two-cohort external IVW estimate, a constrained-vs-unconstrained within-domain contrast and exact direct-overlap-exclusion sensitivity, although the discovery WGS component alone was directionally positive but not significant. Because chromatin conformation and expression were not directly measured in carrier cells, these findings do not prove regulatory dysregulation, but do link ASD boundary-annotated deletions to constrained neurodevelopmental gene architecture.

The ASD–SZ contrast at developmental TAD boundaries should be interpreted as ASD-preferential rather than ASD-specific. In the Japanese discovery WGS cohort, the aggregate ASD > SZ deletion contrast was significant, and the joint-layer multinomial model supported a larger ASD TAD coefficient. However, external cross-disorder support remains incomplete: the arrayCGH ASD–SZ comparison was directionally consistent but not formally significant, and MSSNG lacks SZ samples. Moreover, prior SZ boundary-disruption work used adult and foetal bulk Hi-C maps in a different cohort and annotation framework[26], whereas our analysis used developmental cell-type-differential boundaries from the prefrontal cortex and hippocampus[27]. These differences in ancestry, brain region, developmental context, assay resolution and boundary definition mean that our study is not a direct replication of the Halvorsen bulk-boundary analysis, but rather a developmental cell-type-resolved extension of the same boundary-disruption hypothesis. A lower-magnitude or differently contextualized SZ contribution remains possible. This ASD-preferential boundary pattern should not be interpreted as weaker rare-variant liability in SZ overall because the contextual CNV, rare-coding and PRS layers recovered strong SZ-aligned effects (Supplementary Tables 7 and 10); rather, it suggests that developmental differential-boundary deletion burden may represent a variant-class-specific architecture that is more prominent in ASD.

The contextual reference layers recovered expected features of ASD and SZ genetic architecture. Rare PTV and CNV burden analyses supported a shared rare damaging axis across externally defined ASD, SZ and NDD gene sets, consistent with large-scale exome and CNV studies[7–10, 31]. Gene-discovery integrations placed the Japanese WGS data within established rare-coding frameworks rather than serving as independent primary end-points. PRS analyses provided a within-cohort contrast: SZ PRS showed robust portability in Japanese samples when trans-ancestry discovery resources were available, whereas ASD PRS showed little detectable portability under currently available European-ancestry discovery resources[6, 34–36]. These results confirm that established coding, CNV and common-variant layers are present in the cohort and support the interpretation that the TRE and developmental TAD-boundary signals represent additional WGS-accessible layers.

Joint-layer models test conditional retention rather than causal independence. Rare TRE carriage and ASD Diff_any-boundary deletion burden remained associated after adjustment for measured coding, CNV, NAHR-CNV and PRS layers, indicating that these signals are not simply recapitulations of established risk layers. The TAD-boundary signal, however, should be interpreted as an annotation-specific component of rare deletion/CNV burden, not as a variant class independent of CNVs. These models also do not establish mediation or interaction among layers because the sample size limits fine-grained interaction testing and some layers, particularly disorder-specific exonic CNV strata, are sparse.

Several limitations should be considered. First, although the cohort is large for short-read WGS, it remains underpowered for locus-level TRE discovery, individual boundary-overlapping deletion events and interaction analyses. The TRE analyses support aggregate burden but not specific causal repeat loci. Second, long-read evaluation was restricted to selected case probands carrying short-read TRE outlier calls; broader orthogonal validation in randomly sampled cases and controls would quantify caller-level sensitivity and specificity more accurately. Third, the TAD-boundary analyses depend on a single developmental single-cell 3D genome atlas generated from a limited set of brain regions and developmental stages and may miss disorder-relevant boundaries in unprofiled cell types, brain regions or developmental windows[27]. Fourth, the external ASD replication framework was strong, but cross-disorder replication of the ASD > SZ boundary asymmetry remains incomplete because MSSNG lacks SZ samples and the arrayCGH ASD–SZ contrast was directional but not formally significant. Fifth, population-based controls were older than cases and not evaluated with structured psychiatric interviews. Although controls were curated using available diagnostic records and questionnaire information, and rare germline variants are constitutionally present from birth, residual ascertainment differences could influence absolute burden estimates. This concern is mitigated, but not eliminated, by harmonized PCR-free WGS generation, ancestry adjustment, module-specific burden covariates, matched-static specificity analyses and external ASD cohorts. Finally, chromatin conformation, gene expression and cellular phenotypes were not measured in carriers, so the inferred regulatory consequences of boundary-annotated deletions remain mechanistic hypotheses.

In summary, harmonized Japanese WGS of ASD and SZ resolves a layered rare-variant architecture in which a shared burden-level TRE signal coexists with ASD-preferential rare deletion burden at developmental cell-type-differential TAD-boundary annotations. These findings support a model in which some WGS-accessible mechanisms, such as rare genic-proximal TRE burden, contribute in parallel across ASD and SZ, whereas others, particularly rare deletion burden at developmental cell-type-differential TAD-boundary annotations, show disorder-preferential architecture. This framework illustrates how matched cross-disorder WGS can separate convergent and asymmetric components of psychiatric genetic liability beyond established common-variant, coding and CNV layers.

Methods

Study cohorts, WGS and harmonized analysis design

We analysed WGS data from Japanese ASD and SZ cohorts together with population-based controls from the National Center Biobank Network (NCBN)[37, 38]. ASD and SZ were diagnosed by experienced psychiatrists according to Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5) criteria; Autism Diagnostic Interview-Revised/Autism Diagnostic Observation Schedule for ASD and Structured Clinical Interview for DSM for SZ were available in subsets, and SZ recruitment included the Japanese Schizophrenia Sib-Pair Linkage Group[39]. PCR-free Illumina NovaSeq PE150 WGS was aligned to GRCh38 and processed through genomic variant call format joint genotyping and variant quality score recalibration. Rare CNVs were called from read depth using CNVkit[40], annotated with AnnotSV[41], restricted to autosomal events ≥ 25 kb and validated against independent NimbleGen array CGH data[10] (Supplementary Table 12). Sequencing technical quality metrics are summarized across diagnostic groups in Supplementary Table 13. Within Japanese WGS case–control modules, analyses were restricted to verified Japanese ancestry and adjusted for sex and PC1–PC10; module-specific independent subsets and multiple-testing procedures are detailed below and in the Supplementary Methods.

NCBN controls were curated for psychiatric genetics applications by excluding individuals with primary 10th revision of the International Classification of Diseases F00–F99 or Q00–Q99 diagnoses based on available medical record and questionnaire data. Module-specific subsets accommodated pedigree structure, sequencing-depth availability, high-burden CNV outlier exclusion and CNV/SV call filtering while preserving a common ancestry, covariate and control framework.

Genic-proximal TRE outlier burden

We quantified rare genic-proximal TRE outlier burden with EHdn[13] and STRling[14] in the primary pedigree-aware case–control subset (ASD n = 508; SZ n = 634; Controls n = 8,442), requiring complete sequencing depth, sex and PC1–PC10 covariates. Fivefold cross-fitting defined outlier and rarity thresholds in training controls and evaluated each held-out sample once. Because EHdn and STRling differ in locus opportunity, carriage of ≥ 1 rare TRE outlier was pre-specified as the primary cross-tool end-point and per-individual outlier counts as a secondary measure.

For EHdn, we profiled each sample on GRCh38 with repeat-unit length restricted to 3–8 bp, restricted loci to genic-proximal regions (gene bodies ± 1 kb), canonicalized motifs (minimal rotation/reverse complement), and clustered loci within motif by coordinate proximity (1,000 bp) to accommodate EHdn coordinate wobble. To maximize specificity, we applied three additional filters: (i) exclusion of artefact-prone motifs (trinucleotides AAC, AAG, ACT, ATC; telomeric AACCCT; centromeric AATGG), (ii) removal of loci overlapping a curated blacklist of problematic genomic regions and (iii) a recurrent-hit filter removing clusters observed in more than 10 samples. Anchored in-repeat read (IRR) counts were depth-normalized to 40· using per-sample depth estimated from alignment idxstats. Outlier events were defined in each fold as $z > 5.0$ for depth-normalized IRR signals with a minimum raw IRR threshold ($IRR_{raw} \geq 10$). Rare clusters were defined by training-control carrier frequency < 0.001. Per sample, we

computed carriage of at least one rare TRE outlier as the pre-specified primary end-point and per-individual outlier counts as a secondary measure. Primary association testing used covariate-adjusted logistic regression for carriage of at least one rare TRE outlier (depth, sex, PC1–PC10); secondary Poisson regression for per-individual outlier counts used an exposure offset of log(number of observed clusters).

For STRling, we first generated genome-wide candidate bounds by running extract per sample and merge per chromosome, then built a fixed “genic in-bounds” panel by intersecting joint bounds with the same genic-proximal BED-format interval file and restricting repeat-unit length to 3–8 bp, yielding a fixed panel of 3,471 loci shared across individuals. We ran call per case–control sample using this panel and enforced strict in-bounds outputs by filtering genotype records to the exact bounds key set. Outliers were computed using strling-outliers.py, defining events as $z > 5.0$, adjusted $P \leq 0.05$ and sum_str_counts ≥ 1. Because STRling’s outlier z-score is derived from depth-normalized STR read counts ($\log_2[(\text{sum_str_counts} + 1) / \text{depth}]$), loci with sum_str_counts = 0 carry no direct STR read evidence and their outlier status reflects local depth variation rather than genuine expansion; requiring sum_str_counts ≥ 1 removes these depth-driven false positives from the burden count. Under cross-fitting, rare loci were defined by training-control carrier frequency < 0.001; the exposure denominator was constant (n = 3,471). Association testing used logistic regression for carriage of at least one rare TRE outlier and Poisson regression for per-individual outlier counts (with offset), adjusting for depth, sex and PC1–PC10.

To quantify cross-caller concordance, we compared per-individual outlier loci between EHdn and STRling. For each individual carrying at least one rare TRE outlier from both callers, we assessed whether any EHdn outlier cluster overlapped a STRling outlier locus within ± 1,000 bp and whether the canonical motif matched. Among 321 individuals in whom both callers independently detected at least one rare TRE outlier, 188 (58.6%) shared at least one outlier locus at overlapping coordinates (ASD 52.0%; SZ 50.0%; controls 60.5%), with 96% canonical-motif concordance among overlapping pairs. Cross-disorder homogeneity of TRE burden was assessed by direct ASD-vs-SZ logistic regression using the same covariate framework and by Cochran’s Q statistic on the two disorder-specific log-ORs from case–control analyses.

To assess motif-concordant long-read support for the short-read TRE calls, we performed orthogonal evaluation using two independent long-read sequencing platforms—PacBio and ONT—in 66 case probands carrying short-read rare TRE outlier calls (6 PacBio probands, all with ASD; 60 ONT probands, 33 ASD and 27 SZ). A short-read rare TRE outlier locus was considered long-read supported when one or more long reads showed a motif-concordant insertion or soft-clipped repeat-containing segment at the candidate locus. Full procedures and per-locus results are provided in the Supplementary Methods and Supplementary Table 4.

Cell-type-resolved TAD-boundary SV-burden analysis

We used the Heffel et al.[27] developmental single-cell 3D genome atlas (> 53,000 snm3C-seq[42] profiles across second trimester, third trimester, infancy and adulthood in PFC and HPC). Domain boundaries were defined at 25-kb resolution. Published L2 differential-boundary files provided the primary cell-type-differential boundary sets; static controls were raw non-imputed boundary bins with boundary value = 2 in at least two atlas observation units and not overlapping any L2 differential boundary.

For primary analyses, we used published L2 differential domain boundaries because they balance cell-type resolution, biological interpretability and multiple-testing burden. Of 11 L2 sets, HPC Astro was excluded a priori for < 500 bins, leaving 10 classes across HPC and prefrontal excitatory, inhibitory and astrocyte lineages. After high-confidence raw-boundary intersection, per-class bin counts ranged from 522 to 1,401 and collapsed to a 4,980-bin Diff_any union.

Rare deletions and duplications were filtered using the structural-variant QC framework described above. Common CNV sites (≥ 0.1% cohort frequency) were removed by one-directional overlap masking. Three SV-filter configurations were evaluated: Pattern A (NAHR-mediated genomic disorder loci excluded; primary), Pattern B (NAHR-mediated loci included) and Pattern C (NAHR-excluded and SV length ≤ 1 Mb; secondary sensitivity analysis). An SV was counted as overlapping a boundary bin only when it covered ≥ 10% of the 25-kb bin, a threshold consistent with prior WGS-based analyses of ultra-rare SV burden at brain regulatory annotations[26]; overlapping bins were then collapsed to unique genomic bins per individual. We organized the TAD-boundary analysis around an architecture-level primary end-point, defined as rare deletion (or duplication) burden at the aggregate Diff_any union—the set of unique 25-kb bins annotated as differential in at least one of the 10 analysed L2 cell-type boundary classes (4,980 unique bins after cross-class deduplication). A bin was counted once per individual regardless of its membership in multiple L2 classes. Class-level L2 enrichment analyses were performed as a secondary localization family to map how the aggregate signal was distributed across developmental boundary classes.

Samples exceeding the 99th percentile of total rare CNV count were excluded from the analytic sample to prevent high-burden outliers from inflating the results. Primary association testing used the gene-overlap-adjusted B' logistic regression model, adjusting for sex, PC1–PC10, total deletion span (log1p) and total gene-overlap burden (log1p) computed separately for deletions and duplications, providing a conservative test of boundary-bin burden beyond general genomic and genic SV load (full-model specification in the Supplementary Methods). The primary comparisons were ASD vs. controls (n_complete = 8,877) and SZ vs. controls (n_complete = 8,999). Per-class between-disorder heterogeneity was tested by multinomial logistic regression with shared controls (Control = baseline; ASD = 1; SZ = 2) under the same B' covariates; class-specific Wald P-values were two-sided. Class-level global heterogeneity was summarized by a correlation-adjusted Stouffer Z-method (Brown variance correction) and a one-sided sign test for directional consistency across the 10 classes. Benjamini–Hochberg correction was applied hierarchically across the four primary aggregate tests (ASD/SZ · DEL/DUP at the Diff_any union) and the 10 secondary class-level tests per comparison · SV-type stratum

(Supplementary Methods). Aggregate-signal robustness was assessed by four sensitivity layers (standard and strict leave-one-class-out; top 1–5 sample and event exclusion; Supplementary Methods).

Matched-static specificity analysis. To test specificity to cell-type-differential boundaries, we compared the observed OR for each boundary class against a null distribution of ORs from property-matched static boundaries. Static boundaries were defined as boundaries not annotated as differential in the Heffel framework. For each differential boundary class, we resampled 10,000 matched sets of static boundaries, matching each differential 25-kb bin to static bins on the same chromosome with concordant gene-density bin and segmental-duplication bin (exact match). When the exact-match candidate pool fell below a minimum threshold, matching was progressively relaxed to chromosome + gene-density bin, and then to chromosome only; within each candidate pool, distance-weighted sampling prioritized candidates by the similarity of gene_density and segdup_pct. The same B' logistic regression model was applied to each resampled static set, and the empirical P-value was computed with a pseudo-count upper-tail adjustment as (1 + number of resampled ORs ≥ observed OR) / (1 + 10,000), as detailed in the Supplementary Methods.

ASD-vs-SZ heterogeneity analysis. ASD-vs-SZ heterogeneity was assessed at two levels. At the aggregate level, B' logistic regression contrasted ASD vs. SZ at the aggregate Diff_any union under the same B' covariates. At the per-class level, multinomial logistic regression (Controls = baseline; ASD = 1; SZ = 2) used two-sided Wald tests for δ = β_ASD − β_SZ and a correlation-adjusted Stouffer Z plus sign test for ASD > SZ; the per-class family was treated as independent of the class-level ASD-vs-Control localization family. External replication used only the arrayCGH cohort (MSSNG lacks SZ); see Supplementary Tables 5 and 6, Panel E.

Boundary-type specificity analysis (Diff_any vs. Static). To test whether the ASD deletion signal preferentially localized to cell-type-differential rather than generic boundaries, we classified 25-kb bins into two mutually exclusive categories: Diff_any, the union of published L2 differential-boundary bins across the 10 classes; and static, boundary bins non-differential in all 10 classes, defined from the raw (non-imputed) boundary matrix as bins with value ≥ 2 in at least two atlas observation units and excluding any published L2 differential boundary. The same B' logistic model (case ~ n_boundary_k + Sex + PC1–PC10 + log1p(total_bases_svt) + log1p(total_gene_svt)) was applied to each category's per-bin count. Effect estimates, one-sided P-values and between-cohort heterogeneity were assessed by two-cohort (arrayCGH + MSSNG) external IVW fixed-effect meta-analysis with Cochran's Q and I^2.

Aggregate exon-exclusion sensitivity analysis. To assess whether the boundary signal was confounded by direct protein-coding exon disruption, we used a two-test exon-exclusion family in discovery WGS and MSSNG[28]. A1 (primary) restricted the exposure to boundary bins not overlapping protein-coding exons (GENCODE v46, merged per chromosome); A2 (joint co-exposure analysis) jointly modelled exon-free and exon-overlapping boundary-bin counts as simultaneous co-exposures. The discovery cohort was fitted using Firth-penalized B'-adjusted logistic regression[32] for A1 and standard B'-adjusted joint logistic regression for A2; the MSSNG cohort was fitted using a GEE[43] Binomial Independence model with

family-level clustering for both tests. Holm correction was applied over the two MSSNG one-sided replication P-values to control the family-wise error rate at the replication stage[44]. arrayCGH was not used as a formal exon-exclusion replication platform because clinical aCGH probe designs preferentially cover coding exons and have limited sensitivity for exon-poor and intergenic CNVs[45]. Full operational definitions, model specifications and collinearity diagnostics are provided in the Supplementary Methods (Section 12.11).

Halvorsen 40-kb TAD-boundary comparator (parallel adult and foetal cortical analyses) and joint-layer model. For comparison with the bulk-boundary framework of Halvorsen et al.[26], we used the 40-kb TAD-boundary annotation published by Halvorsen et al. (TAD bins derived from Giusti-Rodríguez/Sullivan cortical Hi-C data[46]), which marks bins observed as TAD boundaries in adult cortex Hi-C and foetal cortex Hi-C. After excluding low-quality, centromeric and blacklisted bins, autosome-restricted bins were lifted over from hg19 to GRCh38, retaining 1,948 adult and 1,792 foetal boundary bins, which were analysed in parallel (bins flagged in both tissues contribute to both analyses). Rare Pattern A boundary-overlapping CNVs were tested under the same SV-type-specific B' covariate framework for each of {adult, foetal} · {DEL, DUP} · {ASD vs. Controls, SZ vs. Controls} plus ASD-vs-SZ heterogeneity (Supplementary Methods § 12.14; Supplementary Table 6, Panel M). Aggregate Diff_any-DEL burden was also evaluated in nested joint-layer logistic models including rare TRE carriage, disorder-matched PRS, rare PTV burden, exonic risk-gene deletion burden and NAHR-mediated CNV carrier status; added-last likelihood-ratio tests assessed per-layer significance (Supplementary Methods § 12.13–12.14).

Replication. External ASD replication combined estimates from an extended Japanese case–control arrayCGH cohort[10] and family-clustered WGS estimates from the MSSNG ASD WGS cohort[28] using fixed-effect IVW meta-analysis, with harmonized low-mappability and exclusion-region filtering across cohorts (samples overlapping with the discovery WGS dataset were excluded). The Japanese arrayCGH analysis (838 cases, 3,583 controls) used a B' logistic regression fitted within each platform stratum and combined by within-array IVW; the MSSNG analysis (5,955 ASD probands and 511 unaffected siblings from 4,978 families) used GEE logistic regression with family-level clustering, robust sandwich standard errors and independence working correlation; full rationale and convergence diagnostics are detailed in the Supplementary Methods. For the nine discovery-positive classes, two-cohort cross-replication was assessed by one-sided BH-FDR < 0.05; the discovery-non-significant class and negative controls used two-sided tests (Supplementary Table 6, Panel A). Because the external replication framework included two cohorts with different ascertainment and assay designs, fixed-effect IVW was used as the primary pre-specified summary, with Cochran's Q and I^2 reported in Supplementary Table 6 for transparency. For the aggregate Diff_any boundary-burden end-point, external support was summarized using the independent two-cohort IVW (arrayCGH + MSSNG) excluding the discovery WGS cohort; for the domain-level constraint-enrichment analysis (Supplementary Methods § 12.12), the primary summary was the three-cohort IVW (WGS + arrayCGH + MSSNG) because the discovery WGS component was directional but underpowered, and the analysis was interpreted as a domain-level biological support layer rather than an independent replication end-point.

In the arrayCGH cohort, we additionally assessed SZ-vs-control deletion burden and ASD-vs-SZ heterogeneity. For the SZ analysis, 2,683 SZ cases (after excluding samples overlapping with the discovery WGS SZ dataset) were compared against the same 3,583 controls using a B' logistic regression model adjusting for sex, log1p(total_bases), log1p(total_gene) and platform in the pooled sample; platform-stratified IVW results are reported in Supplementary Table 6, Panel E as a sensitivity analysis. For heterogeneity testing, a multinomial logistic regression model with three-level outcome (Control = 0, ASD = 1, SZ = 2) and B' covariates (sex, log1p(total_bases), log1p(total_gene), platform) was fitted to the pooled sample (n = 7,104). The heterogeneity statistic z_het = (β_ASD − β_SZ) / SE_joint was computed as in the discovery analysis. Class-specific heterogeneity P-values were derived from two-sided Wald tests, whereas class-level global directional consistency was summarized by a correlation-adjusted Stouffer Z-method with Brown's variance correction[47] and a one-sided sign test, both specified for the directional alternative ASD > SZ. Sensitivity analyses fitted the same MNLogit model within each platform stratum separately (Agilent-only, NimbleGen-only), without the platform covariate.

Exploratory locus illustration. To select representative loci for genomic illustration (Fig. 3f; Supplementary Table 6, Panel F), we mapped rare ASD deletions that overlapped Diff_any bins (25-kb bins annotated as differential in at least one of the 10 analysed L2 classes) to TAD domains reconstructed from the Heffel L2 boundary data. For each L2 class, consensus boundary bins were defined as bins in which at least two of the class's constituent observation units (region · stage · cell-type groups) had a raw boundary value ≥ 2; TAD domains were then defined as the genomic intervals between adjacent consensus boundary bins on each chromosome. Representative loci shown in Fig. 3f were selected based on interpretable deletion–boundary–gene configurations near constrained neurodevelopmental genes or gene clusters overlapping curated ASD/NDD and/or SFARI 1–2 gene sets (LOEUF ≤ 0.4). This selection was for exploratory illustration only; no formal gene-burden significance test, systematic candidate-gene ranking or network-based priority scoring was performed, and the locus-level examples do not constitute independent proof of causality at any single locus.

Contextual analyses. Rare protein-truncating variants and CNV burden were evaluated within externally defined ASD, SZ and NDD gene sets. Rare PTV analyses (gnomAD[30] AF < 0.001 with MANE Select transcript annotation[48]; primary LOEUF < 0.6) used Firth logistic regression[32]. Rare CNV analyses evaluated exonic deletions in the same gene sets (excluding common CNV sites at AF ≥ 1%); NAHR-mediated genomic disorder loci were analysed separately. De novo SNVs/indels in trios were used as a descriptive catalogue and for ASD–SZ carrier-rate comparisons (Supplementary Methods). Gene-discovery analyses for ASD integrated de novo SNVs/indels from five sequencing cohorts—whole-exome data from the Autism Sequencing Consortium and Simons Powering Autism Research (SPARK) (Fu et al.[9]); whole-genome data from our study, the Korean cohort (Kim et al.[49]), and MSSNG (Trost et al.[28])—totalling 21,135 probands and 5,993 siblings, using the TADA Bayesian framework[9, 33]. Transmitted and case–control SNV/indel evidence was sourced from Fu et al.[9], with de novo CNVs additionally drawn from our study and MSSNG[28] (Supplementary Methods). SZ gene discovery extended the Cochran–Mantel–Haenszel-based case–control meta-analysis of Chick et al.[31] with our Japanese case–control

and de novo evidence; the SZ gene-discovery meta-analysis used a more restricted unrelated orphan case–control stratum (367 SZ cases and 8,610 controls) plus 324 SZ trios, whereas primary WGS burden analyses used the broader pedigree-aware case–control subset of 634 SZ cases. Common-variant polygenic risk portability was evaluated using PRS-CSx[34]. Full operational definitions, sample inclusion criteria and model specifications are provided in the Supplementary Methods.

Declarations

Data availability

Individual-level WGS data for the ASD and SZ cohorts and their family members generated in this study have been submitted for controlled-access deposition in the Japanese Genotype–Phenotype Archive, hosted by the NBDC Human Database; the accession number will be provided once repository processing is complete. Access to individual-level data will require approval by the relevant data access committee. For the National Center Biobank Network (NCBN) control samples, allele and genotype frequency data are available in the NBDC Human Database under accession hum0331. Individual-level genomic data for NCBN samples are available upon review and approval by the participating biobanks. Summary statistics and processed results supporting the findings of this study are provided in the Supplementary Tables, and source data underlying Figs. 1–5 are provided with this paper as source data.

MSSNG ASD WGS data used for external replication and gene-discovery analyses were accessed through the Autism Speaks Genomics Program under an approved data access application and data use agreement. Access to MSSNG requires submission of a Researcher Application and signed Genomics Program Access Agreement through the Autism Speaks Genomics Application System, review by the Autism Speaks Coordinating Center and approval by the independent Data Access Compliance Office. The authors are not permitted to redistribute individual-level MSSNG genomic data; processed non-identifying cohort-level results derived from MSSNG are provided in the Supplementary Tables and source data where applicable.

Public external resources used in this study include gnomAD AF data and constraint metrics, PGC3 schizophrenia GWAS summary statistics, iPSYCH/PGC ASD GWAS summary statistics, the Heffel et al.[27] single-cell developmental brain 3D genome atlas generated by snm3C-seq (GEO accession GSE213950), GENCODE v46, MANE Select transcript annotations and the GRCh38 human reference genome. Adult and foetal cortical 40-kb TAD-boundary annotations from Halvorsen et al.[26] were obtained from the corresponding publication/source files; processed boundary annotation files and cohort-level analysis outputs required to reproduce the main reported burden analyses are provided in the Supplementary Tables, source data and/or the accompanying code repository, subject to restrictions on individual-level genomic data.

Code availability

Code supporting all primary and supporting analyses is available at https://github.com/KushimaItaru/ASD-SZ-WGS-Japan. The repository contains, at the time of submission, the modules covering: (i) the TRE cross-fitting and rare-burden pipeline (tre/; entry-point wrapper scripts for ExpansionHunter Denovo and STRling, together with cross-fitting, rare-frequency thresholding and burden-regression helper scripts); (ii) the TAD-boundary SV-burden analysis with a two-tier end-point structure: an architecture-level primary end-point (tad/16_aggregate_primary_sensitivity_v1; aggregate Diff_any boundary-bin four-test family ASD/SZ ´ DEL/DUP, the aggregate ASD-vs-SZ contrast and four-layer aggregate-signal robustness comprising standard and strict leave-one-class-out plus top-N high-burden sample and event exclusion), a shared upstream pipeline constructing the boundary universe and per-individual burden (tad/01_heffel_boundary_master through tad/05_wgs_sample_burden), and sensitivity layers comprising matched-static resampling with chromosome ´ gene density ´ segmental-duplication matching (tad/07_wgs_matched_static), Diff_any-vs-Static per-bin specificity testing (tad/11_diff_all_vs_static) and exon-exclusion two-test analyses across discovery WGS and MSSNG (tad/12_exon_exclusion_wgs, tad/13_exon_exclusion_mssng); (iii) the external replication and IVW meta-analysis pipeline for the arrayCGH and MSSNG cohorts (tad/08_arraycgh_sample_burden, tad/09_mssng_sample_burden, tad/10_replication_2way_meta; within-array IVW, GEE family-clustered logistic regression, fixed-effect three-cohort IVW combining discovery WGS, arrayCGH, and MSSNG and the Diff_any-bounded TAD-scale regulatory-domain constraint-enrichment analysis in tad/14_constraint_enrichment_v1); (iv) the secondary class-level localization (tad/06_wgs_primary_L2; per-class B' logistic regression and per-class multinomial heterogeneity testing with shared-control covariance and correlation-adjusted Stouffer–Brown class-level global directional testing[47]); and (v) numerical-reproducibility verification scripts that re-derive the primary L2 burden numbers reported in the manuscript (tad/99_verify_vs_draft). The integrative joint-layer logistic regression for ASD and SZ is available at joint_layer/asd_sz_layered_logistic_v6.R. Reproducing the full analyses additionally requires access to the cohort-specific input data and reference resources used in this study. The following published tools were used: ExpansionHunter Denovo[13] (https://github.com/Illumina/ExpansionHunterDenovo), STRling[14] (https://github.com/quinlan-lab/STRling), CNVkit v0.9.9 (https://github.com/etal/cnvkit), AnnotSV v3.2.2[41] (https://www.lbgi.fr/AnnotSV/), PRS-CSx v1.1.0[34] (https://github.com/getian107/PRScsx), PLINK2 (https://www.cog-genomics.org/plink/2.0/) and scHiCluster + TopDom (boundary-bin annotation pipeline applied to the Heffel et al.[27] developmental 3D genome atlas; original atlas processing as published). Detailed parameters and version information are provided in the Methods and **Supplementary Methods**.

Use of large language models. Details of the use of large language models (ChatGPT, OpenAI; Claude, Anthropic) in this study are provided in Supplementary Methods §15.

Acknowledgements

This work was supported by the Japan Agency for Medical Research and Development (AMED) (grant numbers JP23ak0101215, JP24zf0127011, JP22gm1410011, JP23gm1910005, JP23ek0109678, JP25wm0625518 (to N.O.), JP21km0405216 (to I.K.), JP22tm0424222 (to I.K., R.W., D.M., Y.A., H.O., S.F., M.Nag., Y.S., M.Ito., R.H., Y.Kaw. and K.T.)), the Japan Society for the Promotion of Science (JSPS KAKENHI grant numbers 21K07543 (to I.K.), 21H00194 (to I.K.), 25K22598 (to N.O.), and 26K02355 (to N.O.)), and the SENSHIN Medical Research Foundation (to I.K.). J.M. and R.H. were additionally supported by AMED (grant numbers JP21wm0425012 and JP25wm0625521). Y.Kun. was supported by JSPS KAKENHI (grant number JP24K02376) and AMED (grant number JP26wm0625126). M.Nag. was supported by JST NBDC (grant numbers JPMJND2302 and JPMJND2601), JSPS KAKENHI (grant number JP21H02681), the Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures and the High Performance Computing Infrastructure in Japan (Project IDs jh200047-NWH, jh210018-NWH, jh220014, jh230016, jh240015, jh250010, and jh260014), the MEXT Cooperative Research Project Program (Medical Research Center Initiative for High Depth Omics, and CURE:JPMXP1323015486 for the Medical Institute of Bioregulation, Kyushu University), the Pan-Omics Data-Driven Innovation Research Center, Kyushu University, and AMED (grant numbers JP20km0405205, JP20km0405501, JP21wm0425009, JP21ek0109548, JP22fk0210111, JP23fk0210138, JP23ek0109672, JP23ek0109675, JP23ek0210194, JP24gm2010001, JP25fk0310536, JP25fk0310535, JP25wm0625519 and JP25kk0305031). Computations were partially performed using the OASIS infrastructure at the Medical Institute of Bioregulation, Kyushu University, and the NIG supercomputer at the ROIS National Institute of Genetics. We thank the Japanese Schizophrenia Sib-Pair Linkage Group (JSSLG) for schizophrenia case sample collection during their genome-wide linkage studies, a subset of which was used in the present analyses. Part of the Japanese ASD and SZ cohorts were obtained from research to elucidate the pathophysiology of mental disorders and support their disorders through the construction and utilization of mental disorder databases (https://byoutai.ncnp.go.jp/database/).

Author contributions

I.K. conceptualized and designed the study, developed the analytical framework, performed the formal analyses (rare coding and CNV burden, TRE outlier burden, TAD-boundary SV-burden analysis, gene discovery and PRS analysis), interpreted the results, prepared the figures and wrote and critically revised the manuscript. R.W. performed variant calling and quality control from raw sequencing data. Y.S., M.Sek., Y.Ma., Y.Sek. and M.Nag. provided sequencing infrastructure and contributed to data generation. I.N., H.Kat., H.Kim., Y.N., Y.T., M.Y., Y.Ha., D.M., Y.A., H.O., S.F., J.M., Y.Y., S.I., Y.W., R.M., S.T.I., J.E., S.B., Y.Ho., M.A., K.Tor., M.Usa., M.Hi., D.F., M.Nak., S.S., T.Mu., S.Y., T.Mr., S.N., K.O., Y.Kun., H.Ko., N.O., M.Ike., M.It. and R.H. contributed to participant recruitment and sample collection. I.K., Y.Kaw. and K.T. supervised the study and acquired funding. All authors reviewed and approved the final manuscript.

Competing interests

I.K. has received research/grant support from Novartis Pharma, GlaxoSmithKline and Takeda Pharmaceutical Co., Ltd., all outside the submitted work. N.O. has received research support or speakers'

honoraria from, or has served as a consultant to, Sumitomo Pharma, Otsuka, Viatris, Eisai, Mochida, Kyowa Pharmaceutical Industry, Nihon Medi-Physics, Nippon Chemiphar, Nippon Boehringer Ingelheim and SUSMED, all outside the submitted work. J.M. has received speakers' honoraria from Sumitomo Pharma, outside the submitted work. Y.Kun. has received research support or speakers' honoraria from, or has served as a consultant to, Takeda Pharmaceutical Co., Ltd., Sumitomo Pharma, Otsuka, Viatris and Lundbeck, all outside the submitted work. M.Ike. has received speakers' honoraria from Meiji Seika Pharma, Otsuka Pharmaceutical, Takeda Pharmaceutical, Lundbeck Japan, Mochida Pharmaceutical, Janssen Pharmaceutica, Mitsubishi Tanabe Pharma, MSD, Eisai, Kyowa Pharmaceutical, Towa Pharmaceutical and Viatris, all outside the submitted work. The remaining authors declare no competing interests.

References

1. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet* **13**, 537-551 (2012).
2. Cross-Disorder Group of the Psychiatric Genomics C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-1379 (2013).
3. Brainstorm C*, et al.* Analysis of shared heritability in common disorders of the brain. *Science* **360**, (2018).
4. Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address pmhe, Cross-Disorder Group of the Psychiatric Genomics C. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482 e1411 (2019).
5. Gandal MJ*, et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693-697 (2018).
6. Trubetskoy V*, et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502-508 (2022).
7. Satterstrom FK*, et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584.e523 (2020).
8. Singh T*, et al.* Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* **604**, 509-516 (2022).
9. Fu JM*, et al.* Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nature Genetics* **54**, 1320-1331 (2022).
10. Kushima I*, et al.* Cross-Disorder Analysis of Genic and Regulatory Copy Number Variations in Bipolar Disorder, Schizophrenia, and Autism Spectrum Disorder. *Biological Psychiatry* **92**, 362-374 (2022).
11. Marshall CR*, et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature Genetics* **49**, 27-35 (2017).
12. Sanders SJ*, et al.* Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215-1233 (2015).

13. Dolzhenko E*, et al.* ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biology* **21**, (2020).
14. Dashnow H*, et al.* STRling: a k-mer counting approach that detects short tandem repeat expansions at known and novel loci. *Genome Biology* **23**, (2022).
15. Trost B*, et al.* Genome-wide detection of tandem DNA repeats that are expanded in autism. *Nature* **586**, 80-86 (2020).
16. Kim JH*, et al.* Short tandem repeat expansions in cortical layer-specific genes implicate in phenotypic severity and adaptability of autism spectrum disorder. *Psychiatry Clin Neurosci* **78**, 405-415 (2024).
17. Mitra I*, et al.* Patterns of de novo tandem repeat mutations and their role in autism. *Nature* **589**, 246-250 (2021).
18. Wen J*, et al.* Rare tandem repeat expansions associate with genes involved in synaptic and neuronal signaling functions in schizophrenia. *Mol Psychiatry* **28**, 475-482 (2023).
19. Mojarad BA*, et al.* Genome-wide tandem repeat expansions contribute to schizophrenia risk. *Mol Psychiatry* **27**, 3692-3698 (2022).
20. Maurano MT*, et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-1195 (2012).
21. Won H*, et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527 (2016).
22. Dixon JR*, et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).
23. Franke M*, et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265-269 (2016).
24. Spielmann M, Lupianez DG, Mundlos S. Structural variation in the 3D genome. *Nat Rev Genet* **19**, 453-467 (2018).
25. Lupiáñez DG*, et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).
26. Halvorsen M*, et al.* Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia. *Nature Communications* **11**, (2020).
27. Heffel MG*, et al.* Temporally distinct 3D multi-omic dynamics in the developing human brain. *Nature* **635**, 481-489 (2024).
28. Trost B*, et al.* Genomic architecture of autism from comprehensive whole-genome sequence annotation. *Cell* **185**, 4409-4427 e4418 (2022).
29. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **57**, 289-300 (1995).

30. Karczewski KJ, *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
31. Chick SL, *et al.* Whole-exome sequencing analysis identifies risk genes for schizophrenia. *Nature Communications* **16**, (2025).
32. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27-38 (1993).
33. He X, *et al.* Integrated Model of De Novo and Inherited Genetic Variants Yields Greater Power to Identify Risk Genes. *PLoS Genetics* **9**, e1003671 (2013).
34. Ruan Y, *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nature Genetics* **54**, 573-580 (2022).
35. Grove J, *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics* **51**, 431-444 (2019).
36. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics* **51**, 584-591 (2019).
37. Omae Y, Goto Y-I, Tokunaga K. National Center Biobank Network. *Human Genome Variation* **9**, (2022).
38. Kawai Y, *et al.* Exploring the genetic diversity of the Japanese population: Insights from a large-scale whole genome sequencing analysis. *PLOS Genetics* **19**, e1010625 (2023).
39. Arinami T, *et al.* Genomewide High-Density SNP Linkage Analysis of 236 Japanese Families Supports the Existence of Schizophrenia Susceptibility Loci on Chromosomes 1p, 14q, and 20p. *The American Journal of Human Genetics* **77**, 937-944 (2005).
40. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* **12**, e1004873 (2016).
41. Geoffroy V, *et al.* AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics* **34**, 3572-3574 (2018).
42. Lee DS, *et al.* Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat Methods* **16**, 999-1006 (2019).
43. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *biometrika*, 13-22 (1986).
44. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70 (1979).
45. Boone PM, *et al.* Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat* **31**, 1326-1342 (2010).
46. Giusti-Rodríguez P, *et al.* Using three-dimensional regulatory chromatin interactions from adult and fetal cortex to interpret genetic results for psychiatric disorders and cognitive traits. *BioRxiv*, 406330 (2018).
47. Brown MB. 400: A method for combining non-independent, one-sided tests of significance. *Biometrics*, 987-992 (1975).

48. Morales J, *et al.* A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. *Nature* **604**, 310-315 (2022).
49. Kim S-W, *et al.* Whole genome sequencing analysis identifies sex differences of familial pattern contributing to phenotypic diversity in autism. *Genome Medicine* **16**, (2024).

Table

Table 1 | Demographic characteristics and analytic sample sizes of the study cohorts. Data are shown for ASD and SZ probands/cases and population controls in the full study cohort, followed by analysis-specific sample sizes for the primary case–control, CNV/TAD case–control and trio subsets. Family members used for family-based inference are not shown as a separate demographic column. Age is reported as mean ± SD (years). †Used for TRE burden, rare PTV burden and PRS analyses after module-specific quality control and pedigree-aware deduplication/relatedness pruning. ‡Used for rare CNV burden and TAD-boundary SV-burden analyses after exclusion of high-burden CNV outlier samples. §N indicates available trio probands; families indicates independent families. After retaining one proband per family and applying joint-call quality and ancestry QC, 376 ASD and 293 SZ probands were included in de novo SNV/indel carrier-rate analyses.

	ASD	SZ	Controls
Study cohort, N	597	762	8,610
Male, n (%)	435 (72.9)	401 (52.6)	4,226 (49.1)
Age, mean ± s.d.	20.1 ± 10.5	41.4 ± 17.2	59.2 ± 19.3
Case–control subset for TRE/PTV/PRS, N†	508	634	8,442
CNV/TAD case–control subset, N‡	505	627	8,372
Trio subset, N (families)§	436 (392)	324 (302)	—

Figures

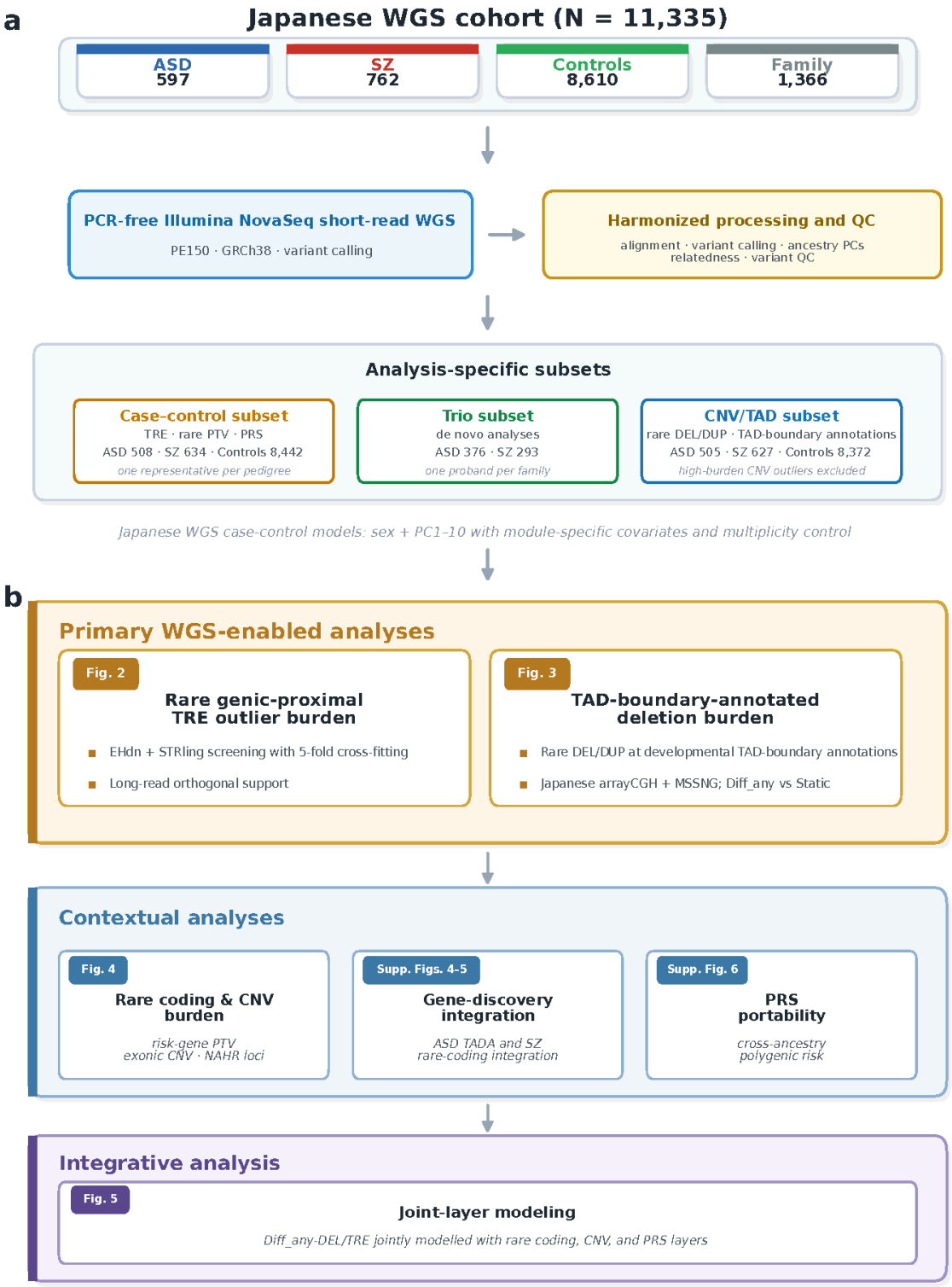


Figure 1

Study design and harmonized analytic framework.

a, Overview of the Japanese WGS cohort and harmonized pipeline. Participants (N = 11,335; ASD n = 597, SZ n = 762, controls n = 8,610, family members n = 1,366) were analysed from PCR-free Illumina NovaSeq short-read WGS data aligned to GRCh38 and processed through unified quality control. Analysis-specific

subsets were constructed to ensure statistical independence and complete covariate availability: a primary case–control subset for TRE burden, rare PTV burden and PRS (ASD 508; SZ 634; Controls 8,442), a trio subset (ASD 436 trios from 392 families; SZ 324 trios from 302 families) for de novo analyses (after retaining one proband per family and applying joint-call quality and ancestry QC, 376 ASD and 293 SZ probands with complete trio data were included in de novo SNV/indel carrier-rate analyses) and a CNV/TAD case–control subset for rare CNV burden and TAD-boundary SV-burden analyses after excluding high-burden CNV outlier samples (ASD 505; SZ 627; Controls 8,372). Japanese WGS case–control models included sex and PC1–PC10, with module-specific covariates and multiple-testing procedures as described in the Methods. **b,** Analytic architecture. The study centres on two primary WGS-enabled analyses: genic-proximal TRE outlier burden using two independent callers with cross-fitting (**Fig. 2**) and rare deletion burden at developmental cell-type-differential TAD-boundary annotations, with external two-cohort (Japanese arrayCGH + MSSNG) IVW support and per-bin boundary-type specificity testing (**Fig. 3**). Rare coding/CNV burden (**Fig. 4**), gene-discovery integration (**Supplementary Figs. 4 and 5**) and PRS portability (**Supplementary Fig. 6**) serve as contextual layers, whereas joint-layer modelling integrates the primary TAD/TRE signals with rare coding, CNV and PRS layers (**Fig. 5**).

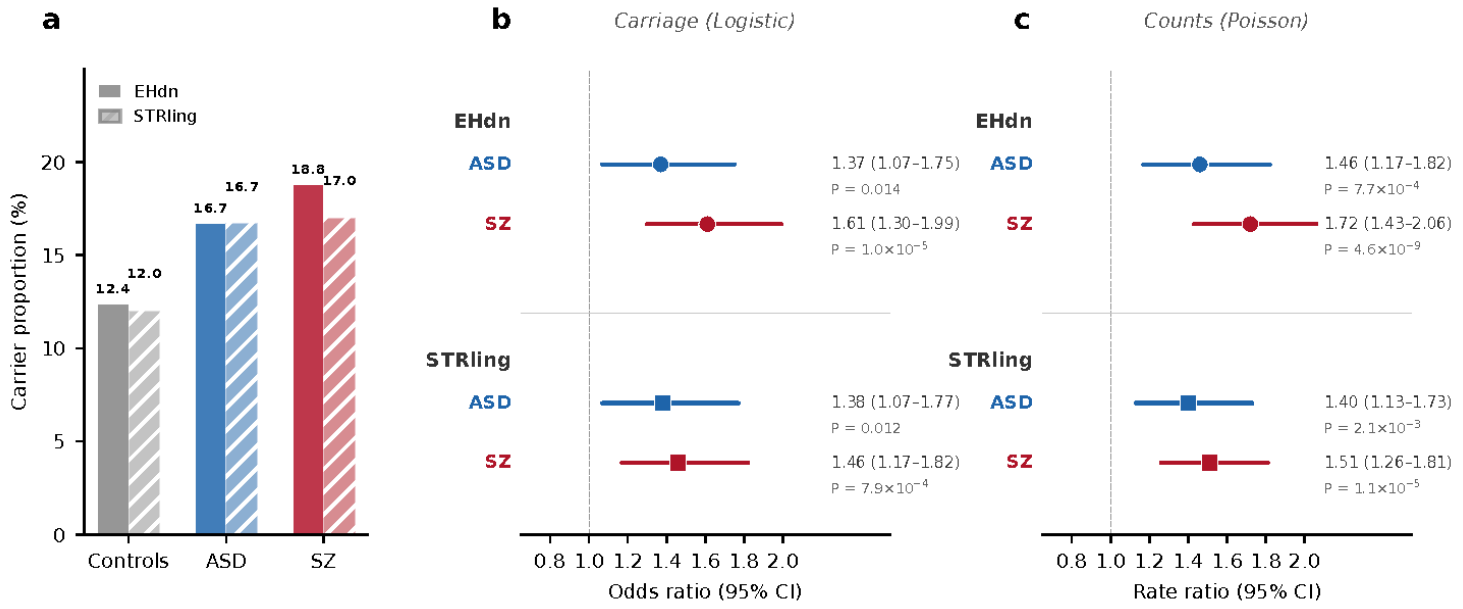


Figure 2

Shared rare genic-proximal TRE outlier burden in ASD and SZ.

Rare genic-proximal TRE outlier burden assessed using two independent callers (EHdn and STRling) with fivefold cross-fitting. **a,** Proportion of individuals carrying at least one rare TRE outlier. **b,** Covariate-adjusted odds ratios (ORs) for carriage of ≥1 rare TRE outlier in ASD and SZ compared with controls. **c,** Covariate-adjusted rate ratios (RRs) from secondary Poisson regression on per-individual rare TRE outlier counts. All models were adjusted for sequencing depth, sex and ancestry principal components (PC1–PC10). Error bars indicate 95% confidence intervals. Among individuals with at least one rare TRE outlier detected by both callers, 58.6% shared at least one overlapping outlier locus, with 96% canonical-motif

agreement among overlapping pairs; independent long-read sequencing provided motif-concordant support for the tested TRE outlier calls. See **Supplementary Tables 2–4** for detailed statistics, cross-disorder comparisons and long-read support.

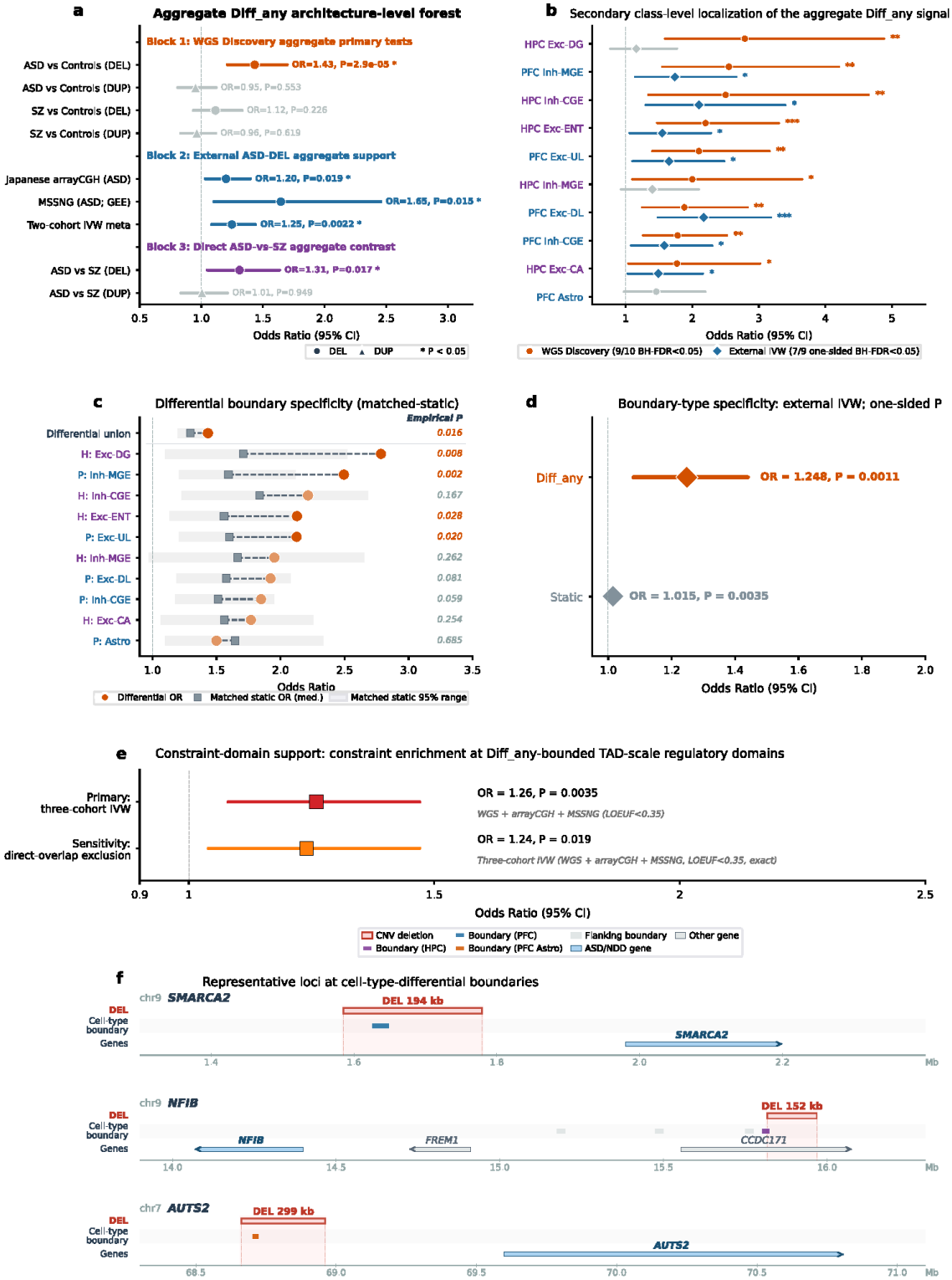


Figure 3

ASD-enriched rare deletion burden at developmental cell-type-differential TAD-boundary annotations.

Rare SVs were tested for enrichment at 10 L2 cell-type-differential developmental TAD boundary classes derived from a single-cell developmental brain 3D genome atlas[27]. Boundary classes span hippocampal (HPC; purple) and prefrontal cortical (PFC; blue) lineages. **a,** Aggregate Diff_any architecture-level forest. Block 1, discovery WGS aggregate tests (ASD/SZ ´ DEL/DUP; B'-model OR ± 95% CI; Pattern A). Block 2, External ASD-DEL support (arrayCGH; MSSNG GEE; two-cohort IVW). Block 3, ASD-vs-SZ aggregate contrast (DEL/DUP). Asterisks denote P < 0.05; heterogeneity statistics in Supplementary Table 6, Panel A. **b,** Secondary class-level localization. Per-class ASD-DEL ORs in discovery WGS (left; BH-FDR asterisks) and external two-cohort IVW (right; arrayCGH n = 838/3,583; MSSNG n = 5,955 probands/511 siblings from 4,978 families; one-sided BH-FDR over the nine discovery-positive classes). Nine of 10 reached BH-FDR <0.05 in discovery (only PFC Astro did not pass); seven of nine replicated at one-sided BH-FDR <0.05. Asterisks indicate discovery BH-FDR or external one-sided BH-FDR levels as applicable (*q < 0.05, **q < 0.01, ***q < 0.001). **c,** Differential-boundary specificity: observed carrier-level differential-boundary ORs (B' logistic; orange circles) compared with matched static-boundary OR distributions (grey squares indicate median; shaded bars indicate 95% range). Properties were matched at the L2 level (chromosome, gene density, segmental-duplication content). Event-level matched-static sensitivity is reported in Supplementary Table 6, Panel B. d, Boundary-type specificity (two-cohort external IVW meta; one-sided P): per-bin ORs and 95% CIs for Diff_any (union of bins differential in ≥1 L2 class) and Static (non-differential in all 10 classes). Diff_any showed substantially larger per-bin enrichment than Static (Supplementary Table 6, Panel C). **e,** Constraint enrichment at Diff_any-bounded TAD-scale regulatory domains. Three-cohort IVW (primary) and direct-overlap-exclusion sensitivity for LOEUF<0.35 constrained-gene burden in flanking Diff_any-bounded TAD-scale regulatory domains; two-sided P. External two-cohort IVW and additional constraint sensitivities are reported in Supplementary Table 6, Panels G and H; exon-exclusion analyses are reported in Panel D. **f,** Representative loci at cell-type-differential boundaries. SMARCA2, NFIB and AUTS2 (constrained ASD/NDD genes) are each flanked by L2-class-specific boundaries overlapped by rare 152–299 kb ASD deletions; full coordinates and L2-class assignments in Supplementary Table 6, Panel F.

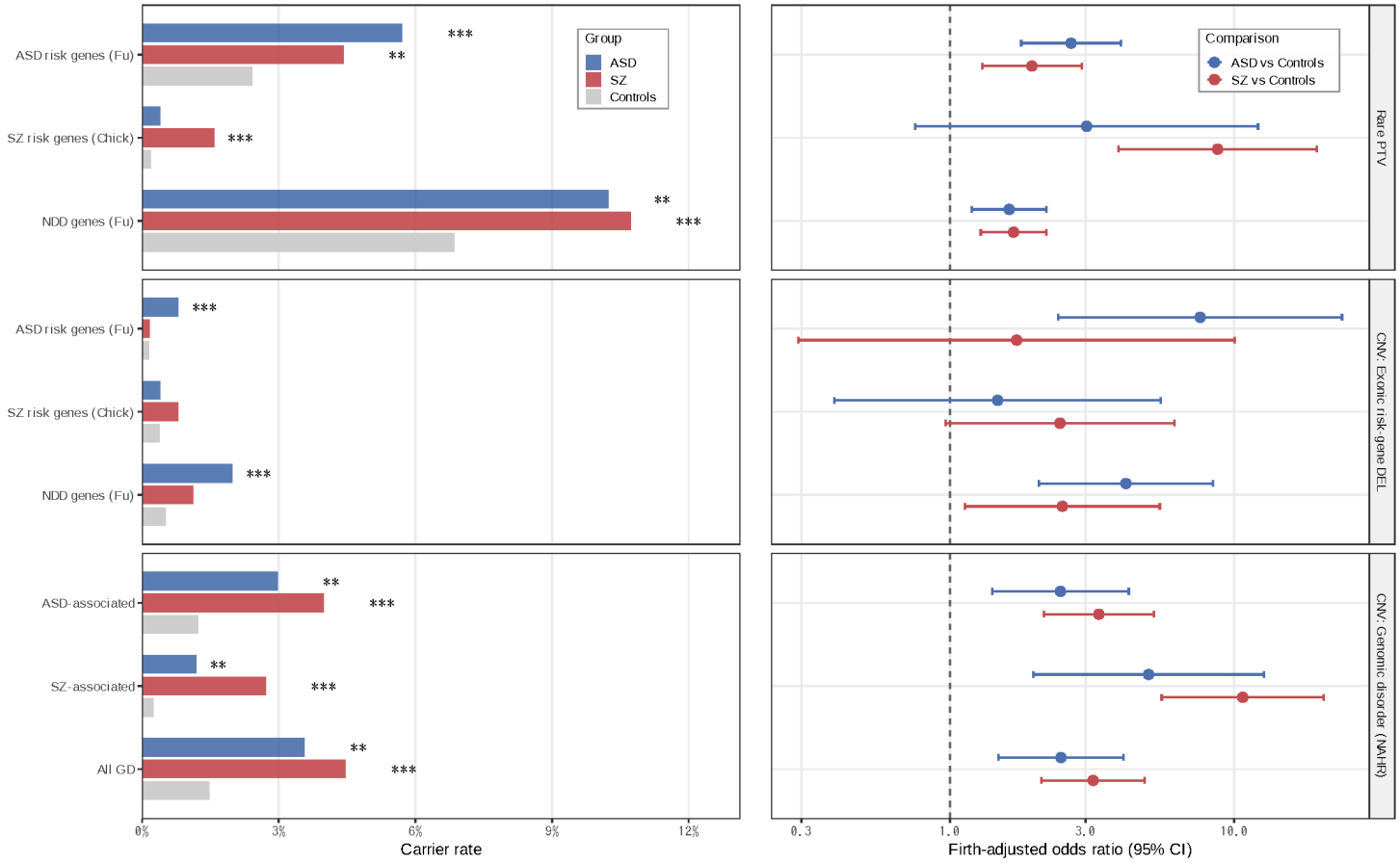


Figure 4

Rare coding and CNV burden in externally defined neuropsychiatric risk-gene sets.

Left panels show carrier rates for Controls (grey), ASD (blue) and SZ (red); right panels show adjusted odds ratios (Firth logistic regression) with 95% confidence intervals on a $\log_{10}$ scale. Rows from top: rare protein-truncating variants (PTVs; LOEUF <0.6), rare exonic risk-gene deletions (CNV DEL) and NAHR-mediated genomic disorder CNVs. Asterisks on carrier-rate bars indicate significant enrichment vs. Controls (**FDR q < 0.01, ***FDR q < 0.001). All models adjusted for sex and PC1–10. See **Supplementary Table 7**, Panels A and B, for detailed statistics.

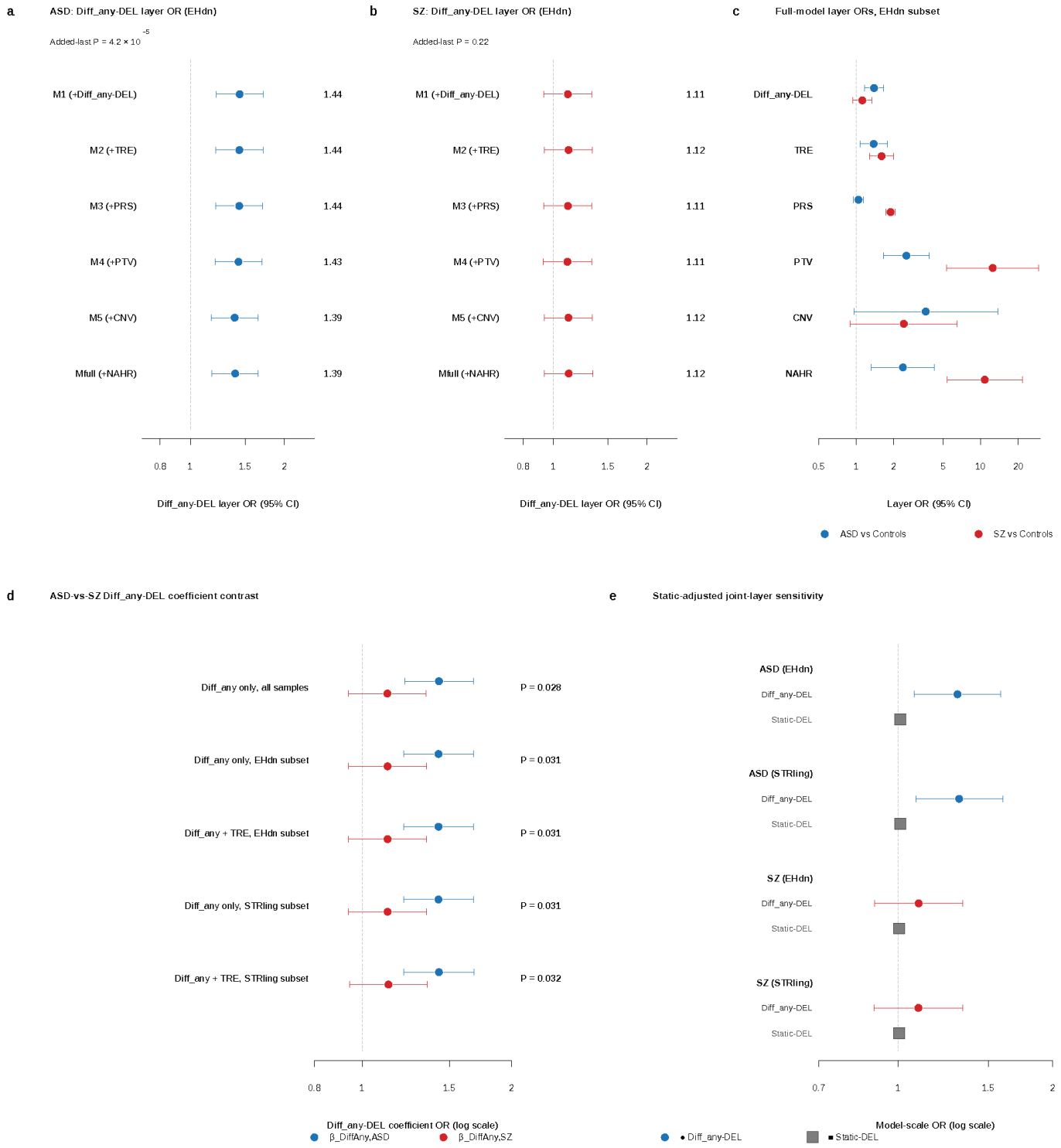


Figure 5

Joint-layer models support an ASD-preferential aggregate differential-boundary deletion signal. Forest plots summarizing the joint-layer integrative sensitivity analyses (**Supplementary Table 11**). Blue indicates ASD and red indicates SZ; grey indicates the static-boundary deletion term in panel e. All points show odds ratios with 95% confidence intervals on a log scale. **a**,**b**, Aggregate Diff_any TAD-boundary deletion burden across nested logistic models for ASD and SZ using EHdn-based TRE carriage. Models

sequentially added TRE, disorder-matched PRS, rare PTV burden, exonic risk-gene deletion burden and NAHR-mediated recurrent CNV carrier status to a B'-aligned base model. **c**, Full-model layer ORs. Diff_any-DEL and TRE are comparably defined across disorders; PRS, PTV, CNV and NAHR layers are disorder-tailored and illustrate within-disorder alignment with established risk architecture. **d**, Multinomial logistic regression with shared controls formally contrasting the aggregate Diff_any-DEL coefficient between ASD and SZ. **e**, Static-boundary adjustment sensitivity; coloured points indicate Diff_any-boundary deletion burden and grey points indicate static-boundary deletion burden in the same model. Exact estimates, model definitions, sample sizes (joint subset N = 8,877 for ASD; N = 8,999 for SZ; panel d uses the shared-control multinomial subset, N = 9,504 with Controls 8,372, ASD 505 and SZ 627) and STRling-based replication are in **Supplementary Table 11**; full methodology in Supplementary Methods §12.13.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable9v1.xlsx
- SupplementaryTable5v51.xlsx
- SupplementaryTable7v6.xlsx
- SupplementaryTable8v5.xlsx
- SupplementaryTable6v51.xlsx
- Supplementary05092026v237.docx
- SupplementaryTable11jointlayeredv06.xlsx