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.

# CD3 on CD39-Positive Resting CD4 Regulatory T Cells and ALS Risk: Genetic Localization to the CD247 Region

**Rongrong Du**
Chinese PLA General Hospital
**Hongfen Wang**
Chinese PLA General Hospital
**Fei Yang**
Chinese PLA General Hospital
**Qiang Shi**
Chinese PLA General Hospital
**Xusheng Huang**
Chinese PLA General Hospital
**Mao Li**
limao301shennei@163.com

Chinese PLA General Hospital

**Research Article**

**Keywords:** amyotrophic lateral sclerosis, regulatory T cells, CD247, CD3, Mendelian randomization, colocalization

**Posted Date:** November 5th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11169945/v1

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

**Additional Declarations:** No competing interests reported.

# Abstract

Previous studies implicated CD3-related T-cell traits in amyotrophic lateral sclerosis (ALS), but the regional genetic basis and cellular scope of these associations remain unclear. We examined the previously reported association involving CD3 on CD39-positive resting CD4 regulatory T cells (Tregs) within an exploratory 73-trait screen. Using an ALS genome-wide association study of 27,205 cases and 110,881 controls, we combined two-sample Mendelian randomization (MR), regional colocalization, locus-deletion analyses and comparisons across related Treg phenotypes. Higher genetically predicted CD3 on CD39-positive resting CD4 Tregs was inversely associated with ALS (odds ratio 0.923, 95% confidence interval 0.880–0.969; P = 0.00115). The 73-trait false discovery rate was 0.084 under uniform variance-floor inference and 0.042 under the original model, with unchanged phenotype-specific estimates. Colocalization supported a shared signal in the CD247 region (default posterior probability 0.983). Removing rs1723016 or the entire region attenuated the association, and related CD39-positive Treg gates showed similar regional dependence. Secondary multivariable, directionality, and expression analyses did not materially alter the regional interpretation. These findings localize a previously reported CD3–ALS association to the CD247 region and prioritize T-cell receptor–CD3 biology for cell-specific functional investigation. Screen-wide evidence remained sensitive to the variance specification, and the mediating cell type and mechanism require independent study.


# Introduction

Amyotrophic lateral sclerosis (ALS) is characterized by progressive motor-neuron loss, but its clinical course varies considerably between patients [1, 2]. Immune dysfunction may contribute to this variation [3]. In particular, impaired regulatory T-cell (Treg) suppressive function has been associated with faster disease progression, providing a rationale for early studies of autologous Treg infusion [4, 5]. These observations concern patients with established disease. Whether inherited differences in specific T-cell traits also influence susceptibility to ALS is a separate question.

Recent therapeutic studies have renewed interest in Treg biology while illustrating the complexity of translating immune findings into clinical benefit. In the 220-participant MIROCALS trial, low-dose interleukin-2 did not significantly improve survival in the unadjusted primary analysis, although adjusted and biomarker-stratified analyses suggested heterogeneous responses [6]. A six-participant study of allogeneic Tregs reported preliminary safety and uncontrolled functional observations [7], and subsequent work examined the expansion of patient-derived Tregs and their activity in preclinical models [8]. Single-cell and spatial studies also indicate that immune states vary with disease progression, subtype and tissue [9, 10]. Against this background, studying a defined cellular measurement may help distinguish the immune features relevant to ALS susceptibility from the changes that accompany established disease.

We examined CD3 on CD39-positive resting CD4 Tregs, a flow-cytometric measure of CD3 within a specified cell population. CD39 defines the gate in which CD3 was measured; the phenotype does not

directly measure Treg abundance or suppressive function. CD39 participates in extracellular nucleotide metabolism, whereas CD3 forms part of the T-cell receptor (TCR) complex [11, 12]. CD247 encodes its zeta signaling chain [12]. Localizing the genetic association could therefore help distinguish receptor-related variation from broader changes in Treg biology.

Lu et al. surveyed 244 T-cell traits and reported an association between this same CD3 phenotype and ALS using a single genetic variant in the 2018 outcome dataset [13]. Broader immune-cell Mendelian randomization (MR) studies also implicated CD3-related traits [14]. These findings provided a specific association requiring regional clarification. Using the larger 2021 ALS genome-wide association study (GWAS), we asked whether the CD3 and ALS associations share a regional genetic signal, whether the association persists after removal of the implicated region, and whether related CD3 measurements across Treg gates show similar regional dependence. The focal phenotype was followed up within an exploratory screen. Our purpose was to refine the biological interpretation of the earlier association and guide targeted experimental investigation.

## Methods

# Design and data sources

We conducted a two-sample summary-data MR study following Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization (STROBE-MR) reporting principles [15]. The study was not preregistered. Original marker-name selection and instrument thresholds were retained; uniform variance-floor inference, joint linkage disequilibrium (LD) verification, related-gate comparisons and revised expression analyses were post hoc. The original multiple-testing family was unchanged.

Orrù et al. profiled 731 immune-cell traits in a predominantly Sardinian European cohort [16]. Marker-name rules selected 74 Treg-related or associated phenotypes with harmonized data. One retained only one single-nucleotide polymorphism (SNP) instrument and was outside the original 73-trait inverse-variance-weighted (IVW) family, comprising 63 regulatory and 10 non-regulatory T-cell gates. Thus, the selected panel was not exclusively Treg-specific. Selection rules, phenotype names, accessions and instrument counts are supplied in Online Resources 1 and 2.

The primary outcome was the European 2021 ALS GWAS (ebi-a-GCST90027164), with 27,205 cases and 110,881 controls [17]. Historical GCST005647 data included 20,806 cases and 59,804 controls and contributed participants to the later GWAS [18]; these outcomes were compared but not meta-analyzed. Exposure–outcome participant overlap could not be quantified. Table 1 summarizes all data sources and analytical roles.

Table 1
Data sources and analytical roles

| Layer | Source or accession | Sample context | Role |
|---|---|---|---|
| Immune phenotypes | Orrù 2020; 731 traits, 74 selected, 73 multi-SNP IVW traits | Predominantly Sardinian European; phenotype-specific N | Screen and related-gate analyses |
| Primary ALS | GCST90027164 / ebi-a-GCST90027164 | 27,205 cases; 110,881 controls | Primary susceptibility outcome |
| Historical ALS | GCST005647 | 20,806 cases; 59,804 controls | Overlapping historical sensitivity |
| FinnGen motor neuron disease | finn-b-G6_ALS | 219 cases; 111,621 controls | Exploratory broader endpoint |
| Blood CD247 eQTL | eqtl-a-ENSG00000198821 | Up to 31,684; selected-SNP N 31,300–31,681 | Four selected instruments; no regional eQTL colocalization |
| Reference genotypes | 1000 Genomes Phase 3 EUR via Ensembl GRCh37 | 503 participants; 30 SNPs | Joint LD and dosage sensitivity |
| Blood expression | GSE234297 | 96 ALS; 48 controls | Welch comparison |
| PBMC expression | GSE288704 | 47 ALS; 19 controls matched to metadata | Welch comparison |
| Motor cortex expression | GSE308462, BA4 only | 29 ALS; 5 controls, unique donors | Age- and sex-adjusted sensitivity |

**Notes**: The principal immune GWAS sample sizes were 2,649 for CD3, 3,437 for CD39 percentage, and 3,435 for CD25. Case counts refer to the specified data versions, not newer releases. Historical ALS data overlap the primary outcome; FinnGen participant independence was not established. PBMC, peripheral blood mononuclear cells; BA4, primary motor cortex.

## Instruments and statistical inference

The principal exposure was "CD3 on CD39 + resting CD4 regulatory T cell" (ebi-a-GCST90001852; N = 2,649), termed CD3 on CD39-positive resting CD4 Tregs throughout. The other two originally selected follow-up traits—CD39-positive secreting Treg percentage and CD25 on secreting Tregs—were retained as distinct exposures in supplementary comparisons and multivariable MR.

Instruments were selected at $P < 5 \times 10^{-6}$, clumped at $r^2 < 0.001$ within 10,000 kb using the 1000 Genomes Phase 3 European reference through OpenGWAS/TwoSampleMR, and retained when $F = \text{beta}^2/\text{SE}^2$ exceeded 10 [19]. Exact variants were extracted without proxies. Allele harmonization excluded incompatible and ambiguous palindromic variants; coordinates refer to GRCh37. Forward odds ratios concern ALS case status per unit of higher genetically predicted, inverse-normalized immune

phenotype. Causal interpretation requires instrument relevance, independence from confounders and exclusion of pathways outside the modeled exposure.

The main IVW analysis used inverse outcome-variance weights and a residual-variance floor of one. The original multiplicative random-effects IVW estimator implemented in TwoSampleMR (mr_ivw_mre) was retained as a comparison using the same data. Equations and implementation details are provided in Online Resource 1.

We used two-sided normal P values and 95% confidence intervals; single-instrument analyses used first-order Wald ratios. Exposure-association uncertainty remains a limitation of these estimators.

Benjamini–Hochberg correction was applied to the same 73 P values under each variance convention; a false discovery rate (FDR) < 0.05 defined screen-wide evidence. The three originally selected phenotypes remained in follow-up. Sensitivities included $P < 5 \times 10^{-8}$ instruments, leave-one-out analysis, and deletion of rs1723016 or the entire CD247 window. Original MR-Egger, weighted median and mode, heterogeneity, intercept and MR-PRESSO outputs were retained as secondary diagnostics [20]. Non-significant diagnostic tests were not treated as proof of valid instruments.

## Directionality and multivariable analyses

Reverse MR used 11 ALS instruments with $P < 5 \times 10^{-8}$ and $F > 10$, the same variance floor, and Benjamini–Hochberg and Bonferroni corrections across three tests. Effects are immune-phenotype beta coefficients per unit higher genetically predicted log odds of ALS.

Steiger sensitivity used liability-scale ALS correlations with effect-allele frequencies and case/control counts [21]. The reference prevalence was 5 per 100,000, motivated by a recent global estimate [22]. Alternative prevalences, sample-size conventions and an illustrative measurement-reliability grid were examined; full specifications and exclusions are given in Online Resource 1.

Multivariable MR included CD3, CD39-positive secreting Treg percentage and CD25 using 30 complete-case SNPs [23]. Joint LD was verified using complete genotypes from 503 participants in the 1000 Genomes European reference, with both PLINK haplotype clumping and genotype-dosage pruning [24, 25]. Both approaches used $r^2 = 0.001$ within 10 Mb and prioritized the minimum exposure-association P value. Because the exposure GWAS shared participants, SNP–exposure covariances were incorporated using jointly-null genome-wide estimates. Sanderson–Windmeijer conditional F statistics assessed instrument strength [26]. Full implementation details and alternative specifications are provided in Online Resource 1.

## Regional colocalization and related phenotypes

Approximate Bayes factor colocalization used ± 500-kb regions around lead immune variants [27]. Quantitative traits used sdY = 1 and sample sizes of 2,649, 3,437 and 3,435 for CD3, CD39 percentage and CD25, respectively; ALS used case fraction 27,205/138,086. Alignment retained trait-specific allele frequencies. Default priors were p1 = p2 = $10^{-4}$ and p12 = $10^{-5}$, with sensitivity to all three priors and per-variant shared-signal posterior estimates. The model assumes one causal variant per trait per region. Full specifications, allele-quality controls and prior grids are provided in Online Resource 1.

All eight CD3-measured Treg gates in the original screen were compared using the same variance floor and 73-trait FDR values, before and after deletion of chr1:166,933,056–167,933,056. Shared instruments were tabulated. Comparisons were descriptive because between-gate covariance was unavailable; the panel contained no CD3-measured non-Treg comparator.

## Expression context and reproducibility

Blood CD247 expression quantitative trait locus (eQTL) analysis used eqtl-a-ENSG00000198821, with four genome-wide-significant clumped instruments, two within the CD247 region [28]. We assessed uniform IVW estimates, weight shares and dominant-variant removal. Regional eQTL–ALS colocalization was not performed from these selected instruments.

We examined 12 genes in three public expression datasets. Welch tests compared log2(CPM + 1) in GSE234297 (96 ALS cases, 48 controls) [29] and log2(normalized count + 1) in metadata-matched GSE288704 (47 cases, 19 controls; three unmatched columns excluded). GSE308462 analyses used a post hoc BA4 subset of 29 ALS and five control donors, one sample per person [30]. Models of log2(TPM + 1) adjusted for age and sex, with HC3 standard errors and Student-t inference. Benjamini–Hochberg correction covered each 12-gene contrast and all 36 comparisons. Sample reconciliation and confounding limitations are detailed in Online Resource 1.

Analyses were conducted in R using TwoSampleMR and in Python using NumPy, SciPy, pandas, and statsmodels; selected calculations, including conditional instrument strength and liability-scale correlation conversion, were cross-checked in R. Figures were generated programmatically with Matplotlib. Exact software versions, analysis-ready inputs, code, and result tables are provided in Online Resources 1–3. Generative AI tools were used to assist with analytical code generation, literature organization, and manuscript revision. All analyses, interpretations, and reported results were reviewed and verified by the authors.

## Results

# The focal CD3 phenotype showed a nominal inverse association with ALS

Eight instruments for CD3 on CD39-positive resting CD4 Tregs yielded an ALS odds ratio of 0.923 (95% CI 0.880–0.969; P = 0.00115; Fig. 1). Under the uniform variance-floor specification, the FDR across the original 73-trait screen was 0.084. The original model gave FDR = 0.042 with identical CD3 effect, standard error and unadjusted P value. Its own standard error did not change because Q = 10.59 exceeded seven degrees of freedom; changes in other phenotypes altered the full-screen ranking and BH adjustment. Thus, the nominal association persisted, while screen-wide significance was sensitive to the variance convention (Fig. S1; Online Resource 2). No phenotype passed the 73-trait FDR threshold under uniform variance-floor inference.

The other original follow-up traits had estimates of 1.034 for CD39-positive secreting Treg percentage (95% CI 1.006–1.063; P = 0.0177; FDR = 0.323) and 0.980 for CD25 on secreting Tregs (95% CI 0.949–1.012; P = 0.227; FDR = 0.752). Minimum marginal F statistics exceeded 21 for all three selected phenotypes. Full estimates, original-model comparisons and diagnostics are supplied in Online Resource 2.

## The association was concentrated in the CD247 region

At the stricter instrument threshold of $P < 5 \times 10^{-8}$, only rs1723016 remained, with a Wald-ratio odds ratio of 0.851 (95% CI 0.793–0.912; $P = 5.32 \times 10^{-6}$). Removing this variant from the relaxed set attenuated the estimate to 0.959 (95% CI 0.915–1.005; P = 0.0814). Excluding the full CD247 region left five instruments and an estimate of 0.966 (95% CI 0.918–1.015; P = 0.172). The attenuation after lead-variant and regional deletion supported concentration of the association in the CD247 region (Fig. 1).

# Regional colocalization supported a shared signal at the CD247 locus

The CD3 region contained 3,175 aligned variants. The default posterior probability for a shared signal was 0.983, compared with 0.012 for two distinct signals. The shared-signal probability ranged from 0.856 to 0.997 across the p12 sensitivity grid (Fig. 2). The broader p1/p2 perturbations yielded probabilities from 0.632 to 0.999; support was weaker when stronger trait-specific priors were combined with the smallest shared-variant prior (Online Resource 2). The leading shared-posterior variant was rs1723018, with posterior probability approximately 0.290 conditional on sharing. Seven variants formed the 95% credible set conditional on sharing, including rs1723016, indicating that the regional signal could not be resolved to a unique variant.

CD39 percentage and CD25 had default shared-signal probabilities of 0.00566 and 0.0122, respectively, compared with 0.983 for CD3. Their regional data provided markedly weaker support for sharing, with some hypotheses favoring association with only one trait or neither trait (Online Resource 2). Integrated results for the focal CD3–ALS association are summarized in Table 2.

Table 2
Integrated analyses of the CD3–ALS association at the CD247 region

| Analysis | SNPs | ALS OR (95% CI) | Unadjusted P | Contribution to interpretation |
|---|---|---|---|---|
| Main MR | 8 | 0.923 (0.880–0.969) | 0.00115 | Phenotype-level association |
| Strict-threshold sensitivity | 1 | 0.851 (0.793–0.912) | $5.32 \times 10^{-6}$ | Estimate retained at strict threshold |
| Without rs1723016 | 7 | 0.959 (0.915–1.005) | 0.0814 | Association attenuated |
| Without CD247 region | 5 | 0.966 (0.918–1.015) | 0.172 | Association attenuated |
| Multivariable joint LD | 30 | 0.933 (0.900–0.968) | 0.000226 | LD-checked conditional estimate |
| Multivariable dosage LD | 26 | 0.933 (0.899–0.968) | 0.000208 | Consistent across LD definitions |

**Notes**: Main-screen FDR = 0.084 across 73 traits. Original-model FDR = 0.042 is shown as a method comparison in Fig. S1 and Online Resource 2. Strict-threshold, deletion and multivariable P values are exploratory and do not replace the screen-wide correction. Multivariable models also include CD39-positive secreting Treg percentage and CD25 on secreting Tregs; conditional F for CD25 remains below 10. All intervals use uniform variance-floor normal inference except the single-SNP Wald ratio. Regional colocalization: 3,175 aligned variants; default PP.H4 = 0.9834; seven variants in the 95% credible set conditional on sharing. PP.H4 denotes the posterior probability of a shared causal variant under the specified colocalization model and priors.

# Related Treg gates showed similar regional dependence

CD3 on CD39-positive resting, activated, and secreting Treg gates showed nominal inverse estimates of 0.923, 0.950, and 0.955, respectively. Their full-screen FDR values were 0.084, 0.317, and 0.317. Removing the CD247 region moved the corresponding estimates toward the null, to 0.966, 0.984, and 1.003. Six CD3 gates shared rs2949661 in their retained instrument sets. The broader resting-Treg CD3 phenotype had an estimate above one, so there was no uniform direction across all Treg measurements. Regional dependence extended across several related CD39-positive gates, whereas the broader Treg measurements differed in direction (Fig. 3a).

# Multivariable estimates were consistent across LD specifications

Joint PLINK clumping retained all 30 SNPs; the maximum haplotype $r^2$ among the 31 same-chromosome pairs within 10 Mb was 0.000918. The uniform-variance CD3 multivariable estimate was 0.933 (95% CI 0.900–0.968; $P = 2.26 \times 10^{-4}$), with conditional F statistics of 11.34, 24.06, and 6.81 for CD3, CD39 percentage, and CD25. Nine within-window pairs exceeded $r^2 = 0.001$ when measured using genotype dosages. Dosage-based pruning retained 26 SNPs and gave a CD3 estimate of 0.933 (95% CI 0.899–

0.968; P = 2.08 × $10^{-4}$), with conditional F values of 13.07, 23.68, and 7.85. Thus, direction and magnitude were stable to the two LD definitions, but conditional weakness remained (Fig. 3b; Online Resource 2).

# Historical outcomes and directionality

In GCST005647, the nine available relaxed-threshold instruments gave an odds ratio of 0.971 (95% CI 0.927–1.016; P = 0.199). Restriction to the same eight instruments retained in the primary outcome gave 0.967 (95% CI 0.921–1.016; P = 0.182). The strict single-variant estimate was 0.918 (95% CI 0.844–0.999; P = 0.0471), but removing the CD247 window yielded 0.997 (P = 0.908). These are overlapping historical comparisons, not independent replication. An exploratory FinnGen motor-neuron-disease analysis, using the corrected source counts of 219 cases and 111,621 controls, was imprecise and not significant (nine-SNP CD3 estimate 0.775, 95% CI 0.549–1.092; P = 0.145; Online Resource 1) [31]. Its broader endpoint and unverified participant independence limit its role.

Reverse MR gave beta − 0.185 for ALS liability to the principal CD3 phenotype (95% CI − 0.386 to 0.015; P = 0.0699; FDR = 0.105). The CD39 percentage reverse estimate was positive but did not survive correction (P = 0.0400; FDR = 0.105); CD25 was also unsupported (P = 0.430). The estimates are shown in Fig. S2. Steiger variance ordering favored the forward direction across all prevalence and sample-size specifications. It also persisted across all 260 admissible reliability-grid combinations; 40 combinations implied invalid explained variances and were excluded. These sensitivity results describe variance ordering rather than verify the IV assumptions.

# CD247 eQTL and tissue-expression context

The four-instrument blood CD247 eQTL estimate was inverse (odds ratio 0.884, 95% CI 0.837–0.933; P = 7.21 × $10^{-6}$), but rs2949669 contributed 88.1% of the IVW weight. Removing it yielded 0.898 (95% CI 0.768–1.050; P = 0.177). The two-instrument cis-region analysis was similarly dominated by the same SNP. The estimate was therefore concentrated in the dominant variant.

In the revised expression comparisons, CD247 was nominally lower in GSE234297 (difference − 0.233; P = 0.0493), but the within-contrast FDR was 0.592. The matched GSE288704 estimate was − 0.179 (P = 0.196; FDR = 0.337), and the donor-level BA4 estimate was 0.076 (P = 0.179; FDR = 0.807). None passed correction across all 36 comparisons either. These comparisons provided tissue-expression context without establishing differential CD247 expression or expression-mediated ALS risk (Online Resource 1, Fig. S3; Online Resource 2).

## Discussion

The genetic association between CD3 on CD39-positive resting CD4 Tregs and ALS was concentrated in the CD247 region. The larger ALS dataset showed a nominal inverse phenotype-level association, regional colocalization favored a shared signal, and removal of the region attenuated the MR estimate. Related CD39-positive Treg gates showed similar dependence. Locus concentration is the principal

observation motivating regional localization, rather than independent confirmation of a phenotype-wide causal effect. Together, these findings refine the earlier association and prioritize TCR–CD3 biology for investigation across defined cell populations.

Lu et al. provided the direct starting point for this study [13]. Their supplementary table reported an odds ratio of 0.918 (95% CI 0.844–0.999; unadjusted P = 0.0471) for the same phenotype, based on rs1723016 and the 2018 ALS outcome; our historical single-variant analysis reproduced that estimate. The larger 2021 outcome gave a more precise inverse estimate, but shared immune GWAS data and overlapping ALS participants preclude an independent-replication interpretation. The added value lies in regional colocalization, lead-variant and region deletion, and comparison across related gates. Pang et al. examined CD4 cytotoxic T-cell genes using single-cell and eQTL-MR analyses [32]; the present flow-cytometric phenotype addresses a distinct question. The phenotype-specific association was unchanged by the variance convention, whereas its screen-wide interpretation was model-sensitive (FDR 0.042 versus 0.084).

CD247 encodes CD3-zeta, a signaling component of the TCR–CD3 complex [12], providing a biological rationale for investigating this region. The source phenotype measures surface CD3 within a CD39-positive gate, rather than CD247 transcription, zeta-chain phosphorylation or suppressive function. Regional sharing and locus dependence therefore focus the mechanistic question on receptor biology while leaving the mediating gene and pathway unresolved. Horizontal pleiotropy remains an alternative to mediation through the measured phenotype. The eQTL estimate was dominated by one variant, and bulk-tissue comparisons supplied no corrected evidence of differential CD247 expression. These findings support targeted molecular measurements rather than an expression-mediated account of ALS risk.

The cellular comparisons further define that investigation. CD39-positive resting, activated and secreting Treg gates showed similar regional dependence, whereas the broader resting-Treg CD3 measurement had the opposite direction. These measures are not interchangeable, although unavailable covariance prevented a formal between-gate comparison. CD39 positivity defines the measured population and does not identify an effect of CD39 enzymatic activity. Follow-up work could combine surface CD3, intracellular CD3-zeta and suppression assays in sorted CD39-positive Treg subsets and matched conventional CD4/CD8 cells, with genotype information where available. Clinical Treg trials concern treatment after ALS onset [6, 7], while expanded-cell and single-cell/spatial studies provide functional and tissue context [8–10, 30]. Genetic susceptibility and treatment response are different outcomes; equivalence between inherited and therapeutic perturbations, including their dose and timing, remains to be tested.

The modest immune-GWAS sample and relaxed $P < 5 \times 10^{-6}$ threshold leave susceptibility to selection bias, winner's curse and weak instruments [33]. Exposure–outcome overlap was unquantified, and conditional weakness of CD25 limits independent-effect interpretation in multivariable MR. Steiger and pleiotropy diagnostics assess, rather than verify, instrument validity, so residual bias cannot be

quantified. Regional inference depends on the one-causal-variant assumption of colocalization and the choice of priors. European reference LD may differ from Sardinian LD, and checks on selected instruments cannot replace dense regional analyses allowing multiple signals. Cellular specificity also remains unresolved because the marker-name screen was not exclusively Treg-specific and no matched CD3-measured non-Treg comparator was available within the selected panel. External validation was limited by the absence of an adequately powered independent genetic dataset and functional perturbation. Small expression comparisons remain vulnerable to cell composition, batch and tissue-related confounding. Transferability beyond predominantly European populations, across ages or across levels of CD3 expression remains uncertain.

## Conclusion

CD3 on CD39-positive resting CD4 Tregs showed a nominal inverse association with ALS that was strongly concentrated in the CD247 region. Regional colocalization and similar locus dependence across related CD39-positive Treg gates refine the earlier association and prioritize TCR–CD3 biology for cell-specific investigation. Screen-wide evidence remains model-sensitive; independent genetic and cell-specific functional studies are needed to define the mediating cell type and mechanism.

## Declarations

# Competing Interests:

The authors have no relevant financial or non-financial interests to disclose.

# Ethics Approval and Consent to Participate:

Our analysis used publicly available GWAS summary statistics, deposited expression data, and public reference genotypes. No new participants were recruited and no new data were collected directly from participants; therefore, no additional ethical approval was required for this secondary analysis. Ethical approval and informed consent for the original studies were obtained by the respective investigators and are documented in the corresponding publications or databases.

**Consent to publish:**

Not applicable. No individual participant data, images, videos, or other identifiable personal information are reported in this manuscript.

# Funding:

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

# Author Contribution

R.D. and H.W. performed the analyses and drafted the manuscript. F.Y. assisted with data preparation. Q.S. and X.H. contributed to interpretation of the results and critical revision of the manuscript. M.L. supervised the study and revised the manuscript. All authors read and approved the final manuscript.

# Acknowledgement

The authors thank the participants and investigators of the source GWAS, FinnGen, eQTLGen, the 1000 Genomes Project, and the Gene Expression Omnibus studies whose publicly available data made this work possible.

# Data Availability

Immune-cell and ALS GWAS summary statistics are available through the GWAS Catalog (https://www.ebi.ac.uk/gwas/) and OpenGWAS (https://opengwas.io/) under the accessions listed in Table 1 and Online Resource 2. Expression data are available from the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under accessions GSE234297, GSE288704, and GSE308462. Public reference genotype data were obtained from the 1000 Genomes Project. Analysis-ready inputs, machine-readable result tables, and numerical outputs supporting this study are provided in Online Resources 2 and 3. Access to the original source datasets remains subject to the terms of the respective repositories.

# Code Availability:

Analysis code, software requirements, and instructions for reproducing the reported computational analyses and figures from frozen analysis-ready inputs are provided in Online Resource 3. Procedures for source-data acquisition and the original genome-wide clumping are documented separately.

## References

1. van Es MA, Hardiman O, Chio A, Al-Chalabi A, Pasterkamp RJ, Veldink JH et al (2017) Amyotrophic lateral sclerosis. Lancet 390:2084–2098. https://doi.org/10.1016/S0140-6736(17)31287-4
2. Taylor JP, Brown RH Jr, Cleveland DW (2016) Decoding ALS: from genes to mechanism. Nature 539:197–206. https://doi.org/10.1038/nature20413
3. McCauley ME, Baloh RH (2019) Inflammation in ALS/FTD pathogenesis. Acta Neuropathol 137:715–730. https://doi.org/10.1007/s00401-018-1933-9
4. Beers DR, Zhao W, Wang J, Zhang X, Wen S, Neal D et al (2017) ALS patients' regulatory T lymphocytes are dysfunctional, and correlate with disease progression rate and severity. JCI Insight 2:e89530. https://doi.org/10.1172/jci.insight.89530

5. Thonhoff JR, Beers DR, Zhao W, Pleitez M, Simpson EP, Berry JD et al (2018) Expanded autologous regulatory T-lymphocyte infusions in ALS: A phase I, first-in-human study. Neurol Neuroimmunol Neuroinflamm 5:e465. https://doi.org/10.1212/NXI.0000000000000465
6. Bensimon G, Leigh PN, Tree T et al (2025) Efficacy and safety of low-dose IL-2 as an add-on therapy to riluzole (MIROCALS): a phase 2b, double-blind, randomised, placebo-controlled trial. Lancet 405:1837–1850. https://doi.org/10.1016/S0140-6736(25)00262-4
7. Shneider NA, Nesta AV, Rifai OM et al (2025) Clinical Safety and Preliminary Efficacy of Regulatory T Cells for ALS. NEJM Evid 4:EVIDoa2400249. https://doi.org/10.1056/EVIDoa2400249
8. You Y, Zhu H, Zhang Q et al (2026) Development and efficacy of ex vivo expanded autologous regulatory T cells for the treatment of amyotrophic lateral sclerosis. Front Immunol 17:1854252. https://doi.org/10.3389/fimmu.2026.1854252
9. Itou T, Fujita K, Okuzono Y et al (2024) Th17 and effector CD8 T cells relate to disease progression in amyotrophic lateral sclerosis: a case control study. J Neuroinflammation 21:331. https://doi.org/10.1186/s12974-024-03327-w
10. Zhang Z, van Olst L, Alessandrini F et al (2026) Integrated single-cell and spatial transcriptomic profiling in ALS uncovers peripheral-to-central immune infiltration and reprogramming. Nat Neurosci 29:1735–1748. https://doi.org/10.1038/s41593-026-02300-5
11. Shevyrev D, Tereshchenko V (2020) Treg heterogeneity, function, and homeostasis. Front Immunol 10:3100. https://doi.org/10.3389/fimmu.2019.03100
12. Dong D, Zheng L, Lin J, Zhang B, Zhu Y, Li N et al (2019) Structural basis of assembly of the human T cell receptor-CD3 complex. Nature 573:546–552. https://doi.org/10.1038/s41586-019-1537-0
13. Lu T, Luo L, Yang J, Cheng X, Sun J (2024) Circulating Levels of T-Cell Traits and the Risk of Amyotrophic Lateral Sclerosis: A Mendelian Randomization Study. Mol Neurobiol 61:10529–10537. https://doi.org/10.1007/s12035-024-04226-0
14. Tang C, Lei X, Ding Y et al (2024) Causal relationship between immune cells and neurodegenerative diseases: a two-sample Mendelian randomisation study. Front Immunol 15:1339649. https://doi.org/10.3389/fimmu.2024.1339649
15. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA et al (2021) Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. JAMA 326:1614–1621. https://doi.org/10.1001/jama.2021.18236
16. Orrù V, Steri M, Sidore C, Marongiu M, Serra V, Olla S et al (2020) Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. Nat Genet 52:1036–1045. https://doi.org/10.1038/s41588-020-0684-4
17. van Rheenen W, van der Spek RAA, Bakker MK, van Vugt JJFA, Hop PJ, Zwamborn RAJ et al (2021) Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. Nat Genet 53:1636–1648. https://doi.org/10.1038/s41588-021-00973-1

18. Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R et al (2018) Genome-wide analyses identify KIF5A as a novel ALS gene. Neuron 97:1268–1283e6. https://doi.org/10.1016/j.neuron.2018.02.027
19. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D et al (2018) The MR-Base platform supports systematic causal inference across the human phenome. eLife 7:e34408. https://doi.org/10.7554/eLife.34408
20. Verbanck M, Chen CY, Neale B, Do R (2018) Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 50:693–698. https://doi.org/10.1038/s41588-018-0099-7
21. Hemani G, Tilling K, Davey Smith G (2017) Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet 13:e1007081. https://doi.org/10.1371/journal.pgen.1007081
22. Liu RY, Su WM, Duan QQ, Wen XJ, He SY, Zhang N et al (2026) Global epidemiology of amyotrophic lateral sclerosis: a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry. https://doi.org/10.1136/jnnp-2026-338675
23. Sanderson E, Davey Smith G, Windmeijer F, Bowden J (2019) An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. Int J Epidemiol 48:713–727. https://doi.org/10.1093/ije/dyy262
24. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7. https://doi.org/10.1186/s13742-015-0047-8
25. 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. Nature 526:68–74. https://doi.org/10.1038/nature15393
26. Sanderson E, Spiller W, Bowden J (2021) Testing and correcting for weak and pleiotropic instruments in two-sample multivariable Mendelian randomization. Stat Med 40:5434–5452. https://doi.org/10.1002/sim.9133
27. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C et al (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 10:e1004383. https://doi.org/10.1371/journal.pgen.1004383
28. Vosa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B et al (2021) Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet 53:1300–1310. https://doi.org/10.1038/s41588-021-00913-z
29. Grima N, Liu S, Southwood D, Henden L, Smith A, Lee A et al (2023) RNA sequencing of peripheral blood in amyotrophic lateral sclerosis reveals distinct molecular subtypes: Considerations for biomarker discovery. Neuropathol Appl Neurobiol 49:e12943. https://doi.org/10.1111/nan.12943
30. O'Neill K, Shaw R, Bolger I, Consortium NYGCALS, Tam OH, Phatnani H, Hammell MG (2025) ALS molecular subtypes are a combination of cellular and pathological features learned by deep multiomics classifiers. Cell Rep 44:115402. https://doi.org/10.1016/j.celrep.2025.115402

31. Kurki MI, Karjalainen J, Palta P, Sipila TP, Kristiansson K, Donner KM et al (2023) FinnGen provides genetic insights from a well-phenotyped isolated population. Nature 613:508–518. https://doi.org/10.1038/s41586-022-05473-8
32. Pang XY, Wang HF, Bai JM, Huang XS (2026) Integration of single-cell RNA sequencing and Mendelian randomization analysis for identifying potential immune therapeutic targets in amyotrophic lateral sclerosis. Biomed Environ Sci 39:327–341. https://doi.org/10.3967/bes2026.011
33. Forde A, Hemani G, Ferguson J (2026) Simulated sample splitting approach to address biases due to instrument selection and participant overlap in two-sample Mendelian Randomization studies. PLoS Genet 22:e1011949. https://doi.org/10.1371/journal.pgen.1011949

# Figures

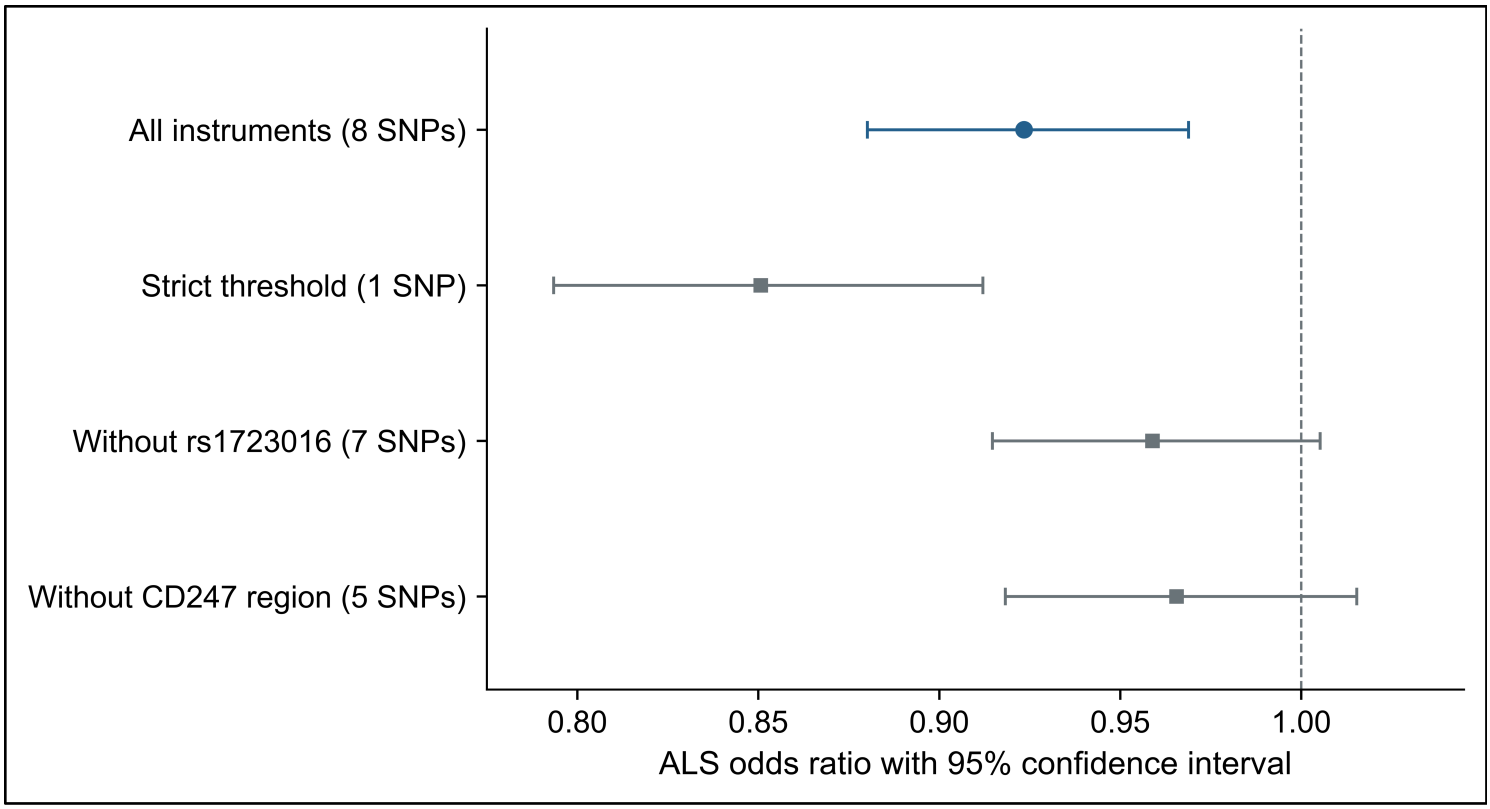



**Figure 1**

Locus dependence of the CD3–ALS association: attenuation after lead-variant and CD247-region deletion. Points and bars show odds ratios and 95% confidence intervals in the 2021 ALS GWAS. The main estimate uses eight instruments selected at $P < 5 \times 10^{-6}$; the strict $P < 5 \times 10^{-8}$ analysis retains rs1723016 alone. Deletion analyses remove either rs1723016 or chr1:166,933,056–167,933,056 (GRCh37). Multi-instrument intervals use the uniform variance floor; the single-instrument interval uses the Wald ratio. The main unadjusted P value is 0.00115 and the 73-trait FDR is 0.084. The strict and

deletion analyses are post hoc and are not independent replications. The dashed line marks the null odds ratio

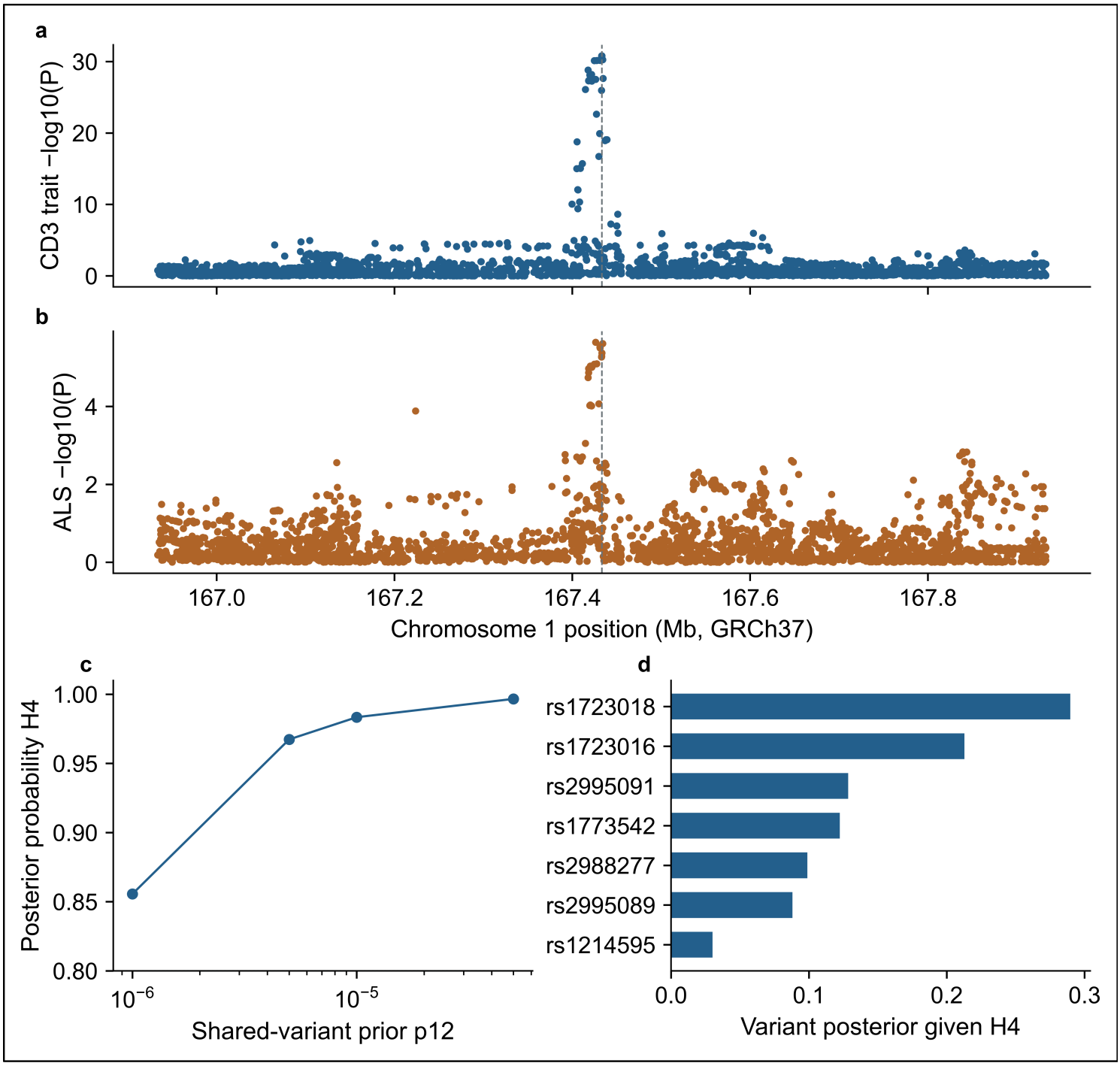



**Figure 2**

Regional colocalization supports a shared CD3–ALS genetic signal in the CD247 region. a Immune-phenotype and b ALS association P values across the 1-Mb region in GRCh37 coordinates. Dashed vertical lines identify rs1723018; all 3,175 aligned variants are shown. c Shared-signal posterior probabilities under four p12 values with p1 = p2 = $10^{-4}$. Wider p1/p2 sensitivity is reported in Online

Resource 2. d The seven variants forming the 95% credible set conditional on a shared signal. Colocalization does not distinguish mediation from horizontal pleiotropy or identify a unique causal gene

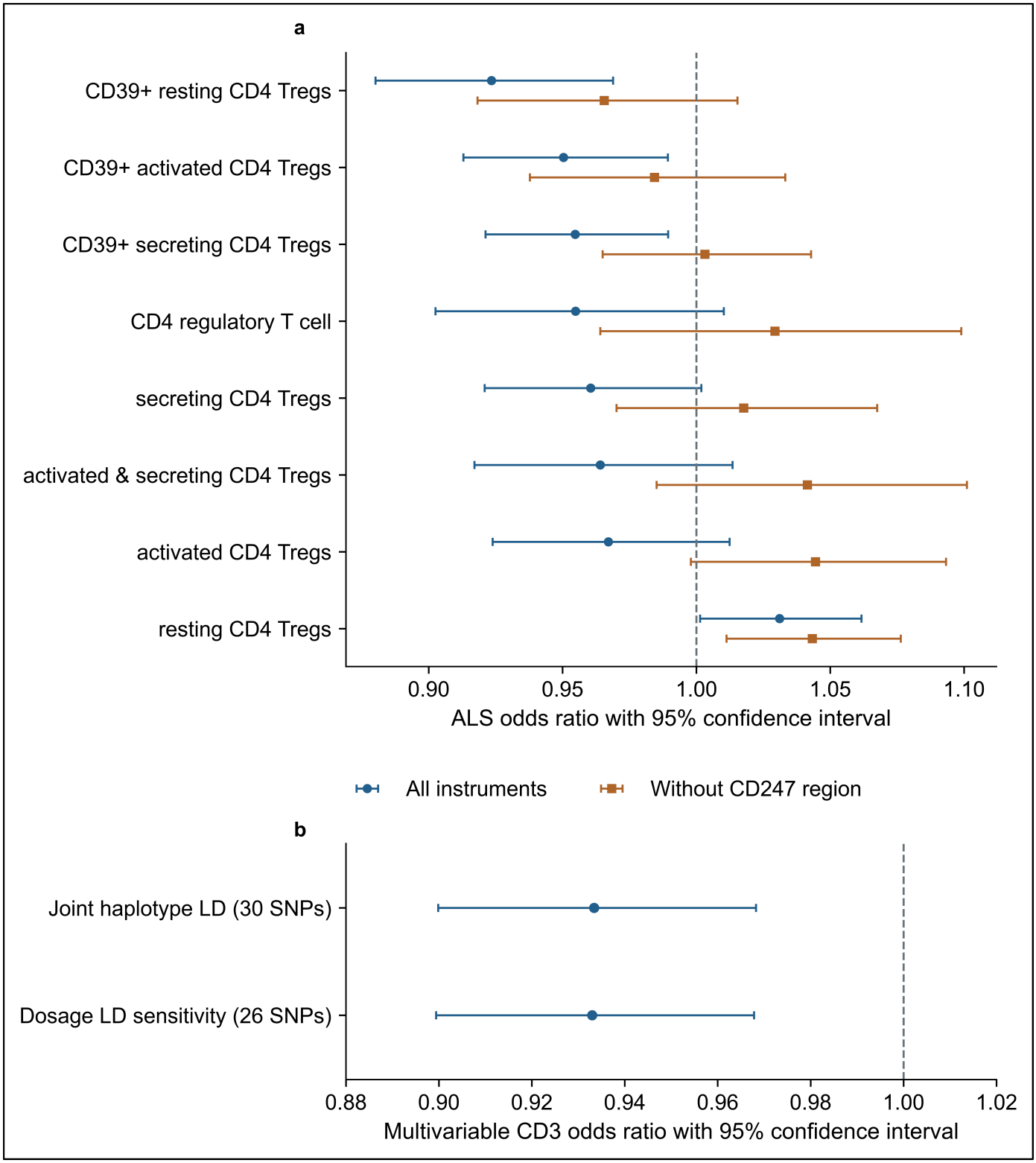



**Figure 3**

Cellular scope of CD247-region dependence across related CD3 phenotypes. a All eight CD3 measurements in the original 73-trait family, before and after CD247-region removal. The labels identify

the source flow-cytometric gates; the focal CD39-positive resting CD4 Treg gate is included alongside the related gates. All-instrument circles and region-excluded squares show 95% confidence intervals; no between-gate significance test was performed. b Estimates for CD3 on CD39-positive resting CD4 Tregs after joint haplotype-LD clumping and genotype-dosage LD pruning, with variance-floor normal 95% confidence intervals. The model also includes CD39-positive secreting Treg percentage and CD25 on secreting Tregs. CD25 remained conditionally weak in both models

## Supplementary Files

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

- ESM1.pdf
- ESM2.zip
- ESM3.zip