

Individuals with autism-associated intellectual disability have an increased burden of *de novo* variants in *cis*-regulatory elements of glutamatergic neurons.

Eric A. Sosa, Dónal O'Shea, Samuel Rosean, Caroline Yoon, Jacob Stauber, Pablo E. Castillo, Cathal Seoighe, Melissa J. Fazzari and John M. Greally

Supplementary Figures

Supplementary Figure 1: Regulatory landscape enrichment analysis (RLEA) demonstrates global DNV enrichment across the *cis*-regulatory regions of diverse cell types.

Supplementary Figure 2: Comparison between the autism and neurotypical groups demonstrates that the properties of glutamatergic regulatory loci are the most distinguishing between groups.

Supplementary Figure 3: Comparison between low and typical IQ reveals significant differences in the genomic properties of brain cell types.

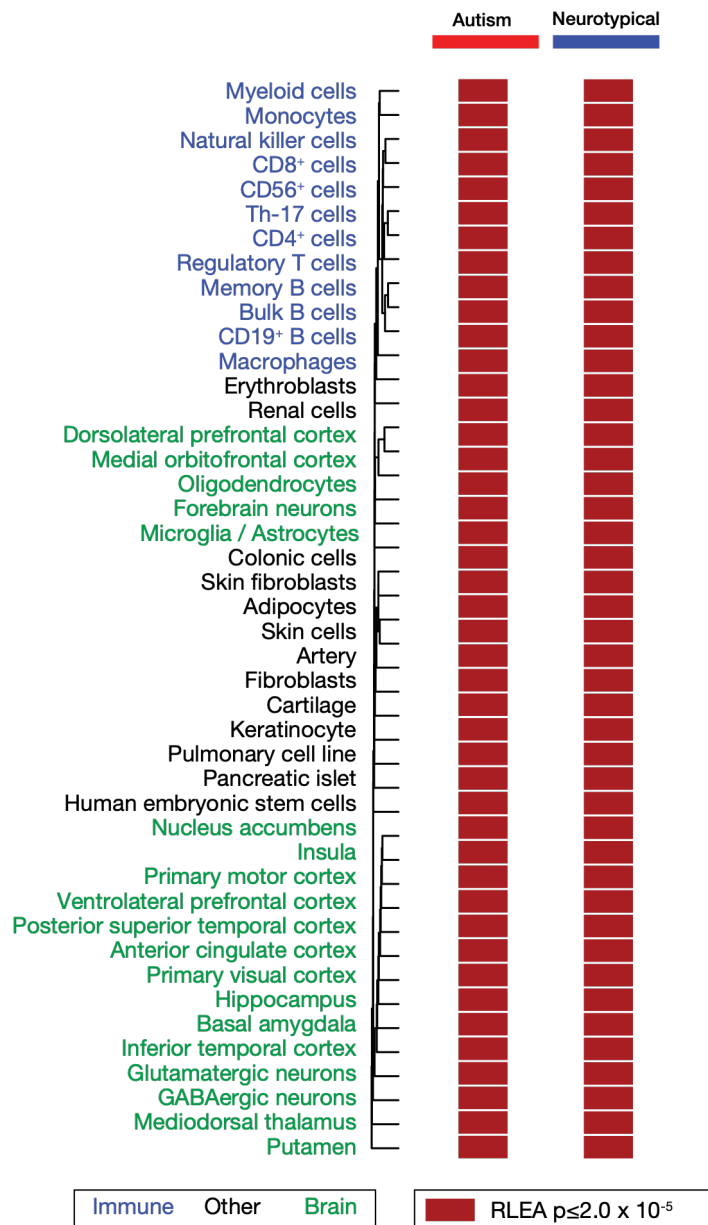
Supplementary Figure 4: Category-wide association study of the SSC cohort replicates the contributions of *de novo* protein truncating and missense variants in autism.

Supplementary Figure 5: Summarization of annotation categories used to analyse 597,439 variants by CWAS.

Supplementary Figure 6: CWAS reveals DNV enrichment at CREs of glutamatergic neurons near Brain-expressed genes in individuals with autism.

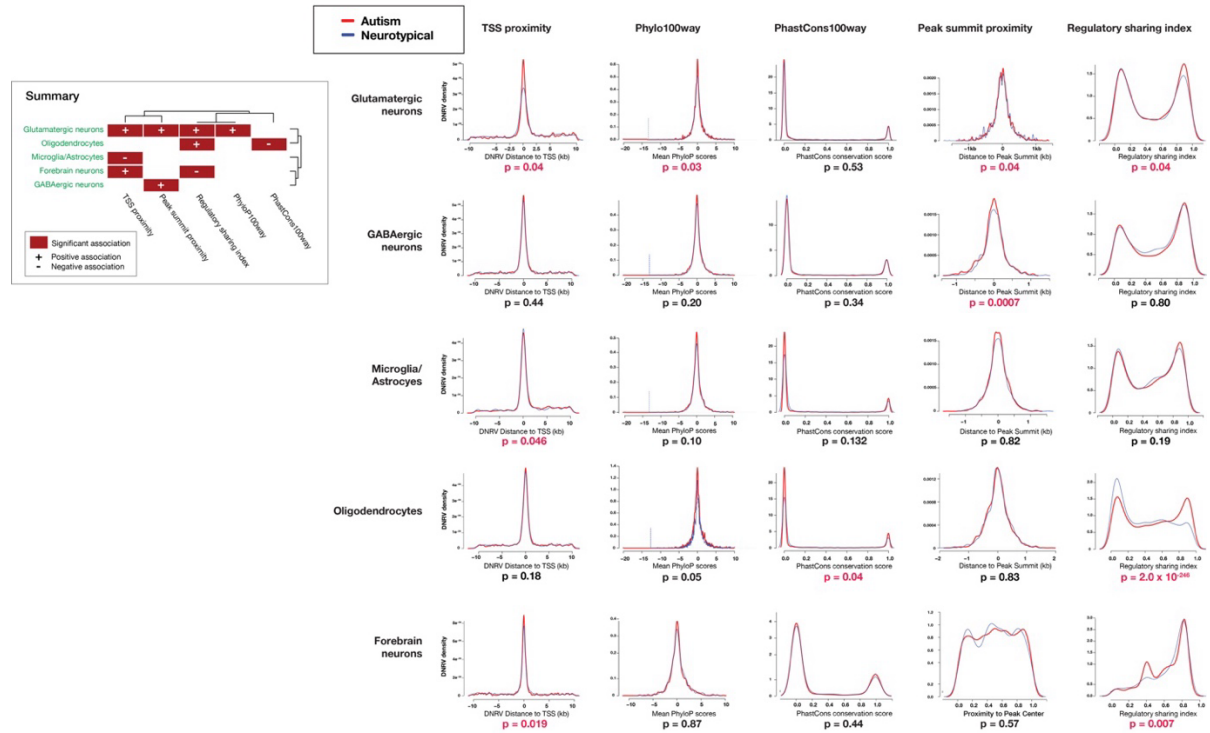
Supplementary Figure 7: DNV counts per individual, per cohort and by autism status.

a *De novo* variants (DNVs) within ATAC-seq peaks



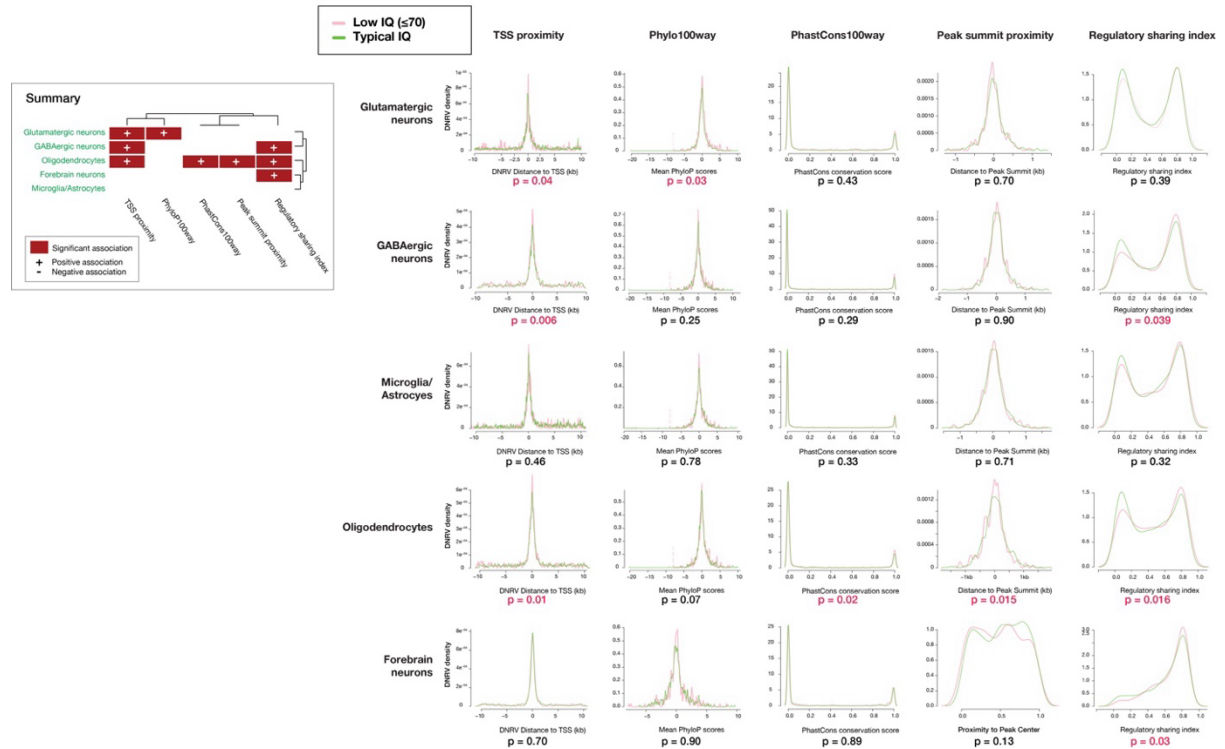
Supplementary Figure 1: Regulatory landscape enrichment analysis (RLEA) demonstrates global DNV enrichment across the *cis*-regulatory regions of diverse cell types. The RLEA approach was used to test for non-random over-representation of DNVs in open chromatin regions of 44 cell types. Every cell type revealed significant enrichment for DNVs in their open chromatin regions, independent of neurodevelopmental status (threshold $p \leq 2.0 \times 10^{-5}$). Statistical significance was assessed using the RLEA enrichment framework with multiple-testing correction (significance threshold $p \leq 2.0 \times 10^{-5}$).

Properties of *de novo* variants (DNVs) in promoter-proximal *cis*-regulatory elements (ppCREs)

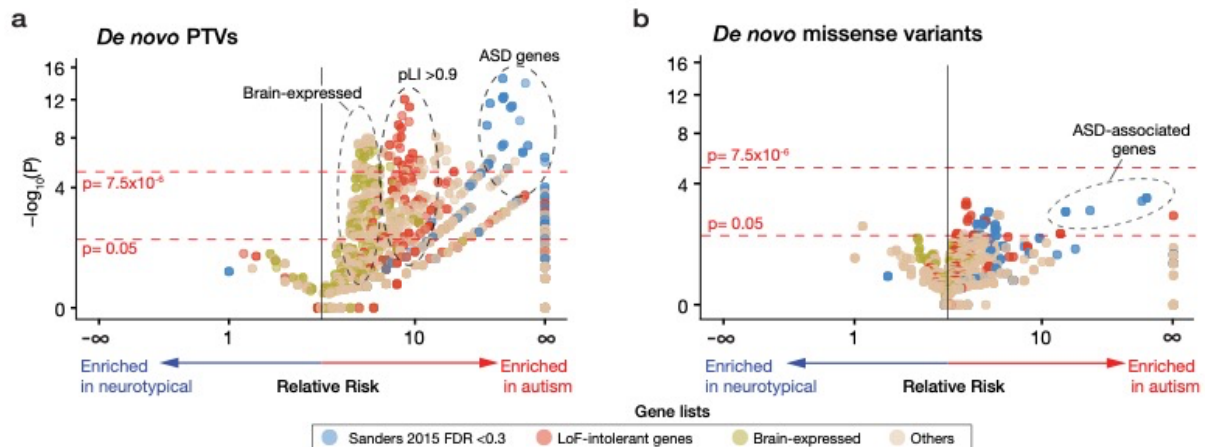


Supplementary Figure 2: Comparison between the autism and neurotypical groups demonstrates that the properties of glutamatergic regulatory loci are the most distinguishing between groups. The TSS proximity, peak summit proximity, conservation status (PhyloP and PhastCons100way) and regulatory sharing index from regulatory loci in microglia/astrocytes, oligodendrocytes, glutamatergic, GABAergic, and forebrain neurons were compared between the autism (red) and neurotypical group (blue). Glutamatergic neuron genomic properties were significantly different between neurotypical and individuals with autism in all categories except PhastCons100way. Statistical significance was determined using a two-tailed Student's ($p < 0.05$). A Wilcoxon-Mann-Whitney test was used to assess significance in data with bimodal distributions (PhastCons and regulatory sharing index).

Properties of *de novo* variants (DNVs) in promoter-proximal *cis*-regulatory elements (ppCREs)



Supplementary Figure 3: Comparison between low and typical IQ reveals significant differences in the genomic properties of brain cell types. The TSS proximity, peak summit proximity, conservation status (PhyloP and PhastCons100way) and regulatory sharing index from regulatory loci in microglia/astrocytes, oligodendrocytes, glutamatergic, GABAergic, and forebrain neurons were compared between low (pink) and typical IQ autism (green). Oligodendrocyte genomic properties were significantly different between both autism groups in all descriptive properties except PhyloP conservation score ($p: 0.07$). Statistical significance was determined using a two-tailed Student's ($p < 0.05$). A Wilcoxon-Mann-Whitney test was used to assess significance in data with bimodal distributions (PhastCons and regulatory sharing index).



Supplementary Figure 4: Category-wide association study of the SSC cohort replicates the contributions of *de novo* protein truncating and missense variants in autism. (a) CWAS was performed to replicate the previous findings of An *et al.*, 2019 using 1,902 families from the SSC cohort. *De novo* protein truncating variants (PTVs) are enriched across brain expressed, autism, and evolutionarily conserved (pLI > 0.9) genes in individuals with autism (right side). The horizontal dashed red lines indicate the nominal significance threshold ($p < 0.05$) and the multiple testing-corrected significance threshold ($p < 7.5 \times 10^{-6}$), calculated from the effective number of independent CWAS categories. PTVs were most enriched near “ASD genes” which are represented in the Sanders risk genes TADA (FDR < 0.3) gene list. (b) The equivalent CWAS is shown for *de novo* missense variants. Although *de novo* missense variants were similarly enriched in individuals with autism ($p < 0.05$), they did not surpass the significance threshold after correcting for multiple testing.

12 GENCODE annotation categories

Category		Variants
Any		597,439
Coding	Missense	5,731
	Silent	2,211
	Loss of function	1,022
	Frameshift	481
	In-frame	146
Non-coding	Intronic	327,069
	Intergenic	182,709
	Promoter	34,317
	Untranslated regions	7,618
	Other non-coding exons	4,088
	Splice-site	35

3 variant types

Category	Variants
All	597,439
Single nucleotide variants	539,286
Insertion-deletions	58,154

6 conservation categories

Category	Variants
All	597,439
PhastCons100way ≥ 0.2	73,040
PhyloP100way ≥ 2	18,267
Runs of contiguous constraint (RoCCs)	6,663
Unannotated intergenic constrained regions (UNICORNS)	3,344
Zoonomia ultra-conserved elements (zooUCEs)	28

19 gene lists

Category		Variants
Any		597,439
GENCODE	Protein-coding	292,037
	LincRNA	101,458
	Pseudogene	2,876
Autism-associated gene lists	Brain-expressed genes	236,408
	LoF-intolerant genes	78,445
	Postsynaptic density genes	34,942
	CHD8 target genes	30,671
	FMRP target genes	29,754
	Mid-fetal co-expression genes	11,795
	Fu 2023 ASD/DD	10,600
	HC-NDD	8,148
	Fu 2023 ASD	2,399
	DDD genes	1,640
	Sanders 2015 FDR <0.3	1,093
	SPARK 2023 genes	991
	Satterstrom 2020	710
	Fu 2023 ASD	548
	SPARK 2022 genes	442

25 functional annotation categories

Category		Variants
Any		597,439
Brain	Fetal brain (male)	17,295
	Fetal brain (female)	16,178
	Glutamatergic neurons	16,034
	Astrocytes (primary cells)	15,807
	Neuronal progenitors	14,347
	Midfetal brain ATAC-seq	14,066
	Microglia/astrocytes	13,365
	GABAergic neurons	13,128
	Oligodendrocytes	11,979
	Hippocampus	6,598
	Midfetal brain H3K27ac (15gw)	5,784
	Midfetal brain H3K27ac (18gw)	5,715
Non-brain	Midfetal brain H3K27ac (17gw)	4,502
	Basal amygdala	4,146
	Forebrain neurons	3,789
	hESCs	14,479
	CD4+ cells	8,693
	Natural killer cells	8,032
	Regulatory T cells	6,607
	Monocytes	6,010
	Fibroblasts	5,050
	Fantom enhancers	4,469
	Vista enhancers	450
	Human-accelerated regions	147

517,104 combinations
(12 x 3 x 6 x 19 x 42 x 3)
e.g. Conserved SNVs
in glutamatergic regulatory regions
of developmental delay genes

288,000 non-redundant
combinations

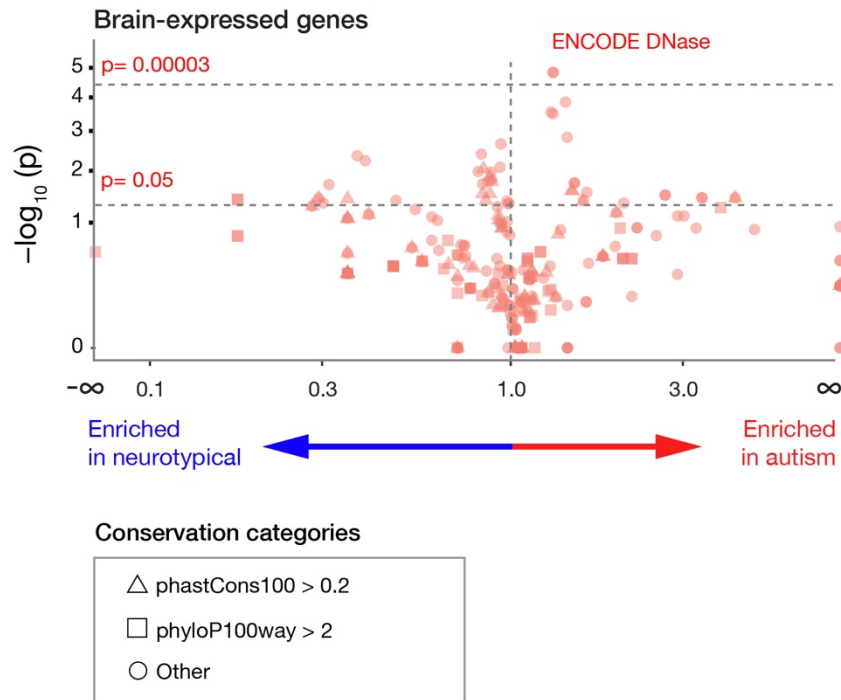
3 descriptive categories

Category	Variants
All	597,439
Near TSS (± 10 kb)	105,360
AlphaMissense ≥ 0.5	1,688

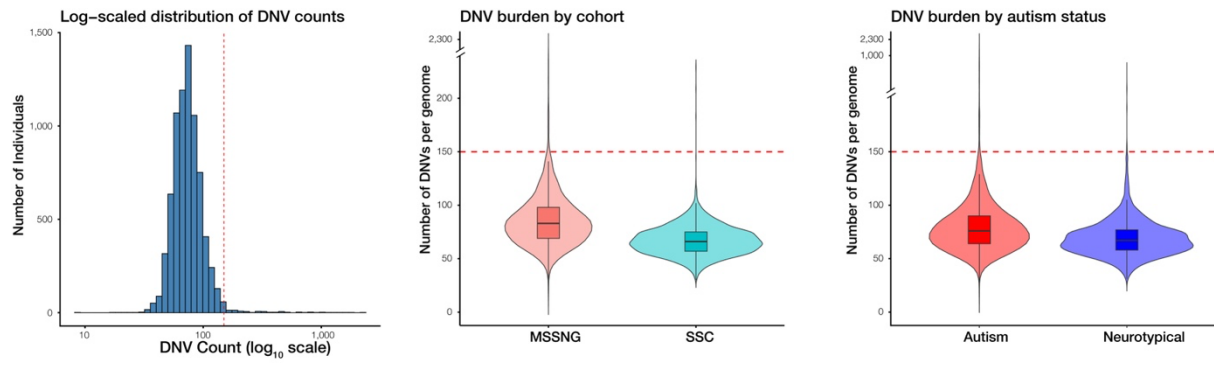
17 functional single-cell annotation categories

Category		Variants
Any		597,439
Glutamatergic	Intratelencephalic projecting neurons, cortical layer 2/3 (IT-L 2/3 types 1-6)	34,536
	Intratelencephalic projecting neurons, cortical layer 5 - type 1-4 (ITL5 1-4)	16,185
	Intratelencephalic projecting neurons, cortical layer 6 - subclass 1 - type 1 (ITL6)	12,486
	Intratelencephalic projecting neurons from primary visual cortex - type 1-3 (ITV1C 1-3)	11,114
	Intratelencephalic projecting neurons, cortical layer 4 - type 1/2	9,415
	L6 Corticothalamic neurons (CT 1/2)	5,885
	Intratelencephalic projecting neurons, cortical layer 6B - type 1/2 (L6B)	5,579
	Intratelencephalic projecting neurons, cortical layer 2/3 - type 4	5,117
	Intratelencephalic projecting neurons, cortical layer 4/5 like - type 1/2 (ITL45 1-2)	5,021
	Amygdala glutamatergic neurons (AMY)	3,838
	Cerebellar neurons (CBGRC)	3,832
	Near-projecting neurons - type 1/2/3 (NP types 1-3)	3,716
	Entorhinal cortex glutamatergic neurons (ERC 1/2)	3,219
	Piriform cortex glutamatergic neurons (PIR)	1,804
	Temporal pole glutamatergic neurons (TP)	732
	Cholinergic neurons (CHO)	487
	Extratelencephalic projecting neurons (ET)	471
Subicular cortex glutamatergic neurons (SUB)	185	

Supplementary Figure 5: Summarization of annotation categories used to analyse 597,439 variants by CWAS. Seven groups of annotations were defined composed of 25 functional annotations from cell/tissue types (purple) or pseudobulked single nuclear ATAC-seq from 18 cell subtypes (dark pink), 19 gene lists (green), 12 GENCODE annotations (orange), 6 conservation measures (blue), 3 descriptive parameters (yellow), 3 variant types (light pink). These variables provided 517,104 combinations, of which 288,000 were unique.



Supplementary Figure 6: CWAS reveals DNV enrichment at CREs of glutamatergic neurons near Brain-expressed genes in individuals with autism. DNVs in ppCREs defined by ENCODE DNase hypersensitivity data (red points) near known brain-expressed genes are significantly enriched in the autism group. Statistical significance is shown as $-\log_{10}(p)$, with red lines indicating the nominal ($p \leq 0.05$) and multiple testing-corrected ($p \leq 1 \times 10^{-4}$) significance thresholds.



Supplementary Figure 7: DNV counts per individual, per cohort and by autism status. In the left panel the strongly right-skewed DNV count per individual is shown, and the 150 DNV threshold, representing a $\sim 3 \times$ IQR cutoff. There was no significant difference of DNV burden by cohort or autism status.