nature neuroscience

Genotype to phenotype relationships in autism spectrum disorders

Jonathan Chang[1,2,5], Sarah R Gilman[1,2,5], Andrew H Chiang[1,2], Stephan J Sanders[3,4] & Dennis Vitkup[1,2]

Autism spectrum disorders (ASDs) are characterized by phenotypic and genetic heterogeneity. Our analysis of functional networks perturbed in ASD suggests that both truncating and nontruncating *de novo* mutations contribute to autism, with a bias against truncating mutations in early embryonic development. We find that functional mutations are preferentially observed in genes likely to be haploinsufficient. Multiple cell types and brain areas are affected, but the impact of ASD mutations appears to be strongest in cortical interneurons, pyramidal neurons and the medium spiny neurons of the striatum, implicating cortical and corticostriatal brain circuits. In females, truncating ASD mutations on average affect genes with 50–100% higher brain expression than in males. Our results also suggest that truncating *de novo* mutations play a smaller role in the etiology of high-functioning ASD cases. Overall, we find that stronger functional insults usually lead to more severe intellectual, social and behavioral ASD phenotypes.

[1]Department of Biomedical Informatics, Columbia University, New York, New York, USA. [2]Department of Systems Biology, Center for Computational Biology and Bioinformatics, Columbia University, New York, New York, USA. [3]Department of Psychiatry, University of California, San Francisco, California, USA. [4]Department of Psychiatry, Department of Genetics, Yale University, New Haven, Connecticut, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to D.V. (dv2121@columbia.edu).

Received 12 August; accepted 26 November; published online 22 December 2014; doi:10.1038/nn.3907

Related publications:

Rare De Novo Variants Associated with Autism Implicate a Large Functional Network of Genes Involved in Formation and Function of Synapses
Sarah R. Gilman, Ivan Iossifov, Dan Levy, Michael Ronemus, Michel Wigler, Dennis Vitkup
Neuron, 70, 878-907 (2011)

Diverse types of genetic variation converge on functional gene networks involved in schizophrenia
Sarah R. Gilman, Jonathan Chang, Bin Xu, Tejdeep S Bawa, Joseph A Gogos, Maria Karayiorgou, Dennis Vitkup
Nature Neuroscience, 15, 1723-1728 (2012)

Autism Spectrum Disorders

- ASD now includes several conditions that used to be diagnosed separately: autistic disorder, pervasive developmental disorder not otherwise specified (PDD-NOS), Asperger syndrome, and childhood disintegrative disorder
- These disorders involve a wide range of cognitive and behavioral abnormalies: social deficits and communication difficulties, repetitive behaviors and interests, sensory issues, deviance in language development, poor motor skills' and in some cases, cognitive delays.
- About 1 percent of the world population has autism spectrum disorder. (CDC,2014)
- Prevalence in the United States is estimated at 1 in 68 births. (CDC, 2014)
- Prevalence of autism in U.S. children increased by 119.4 percent from 2000 (1 in 150) to 2010 (1 in 68). (CDC, 2014) Autism is the fastest-growing developmental disability. (CDC, 2008)
- Cost of lifelong care can be reduced by 2/3 with early diagnosis and intervention. (Autism Society estimate)
- ASD is almost 5 times more common among boys (1 in 42) than among girls (1 in 189)

http://www.autism-society.org/
http://www.cdc.gov/ncbddd/autism/

Data used in the study

- *de novo* **CNVs** and **SNVs** observed in autistic patients and their healthy siblings from Simons Simplex Collection (SSC)
  (Levy *et al., Neuron,* 70, 886, 2011; Iossifov *et al.*, *Neuron*, 74, 285,2012; O'Roak *et al.*, *Nature*, 485, 246, 2012; Sanders *et al.*, *Nature,* 485, 237, 2012)

  SSC – a permanent repository of genetic samples from **2600 simplex families**, each of which has one child affected with an autism spectrum disorder, and unaffected parents and siblings
  http://sfari.org/resources/simons-simplex-collection

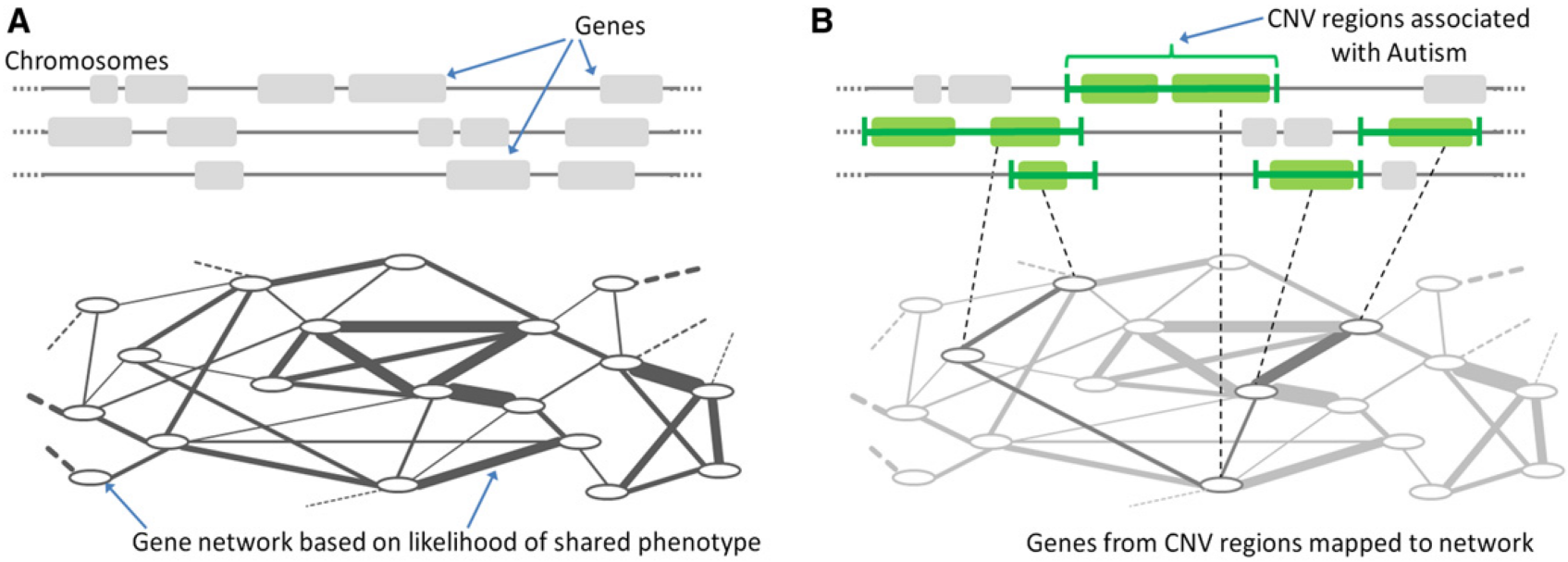
- **991** unique genes (**580** with *de novo* **SNVs** and **434** within *de novo* **CNVs**) from **624** independent loci

- Temporal and spatial brain expression data from the Human Brain Transcriptome database (HBT, Kang *et al.*, *Nature*, 478, 483, 2011) hbatlas.org

- Expression profiles of 25 cell lines of mouse CNS (Doyle *et al. Cell*, 135, 749, 2008)

NETwork-Based Analysis of Genetic associations (NETBAG)

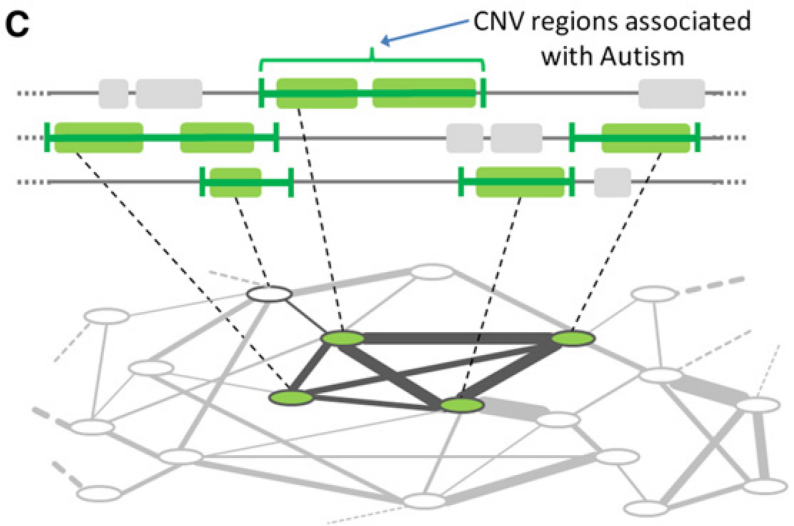
de novo **CNVs** and **SNVs**



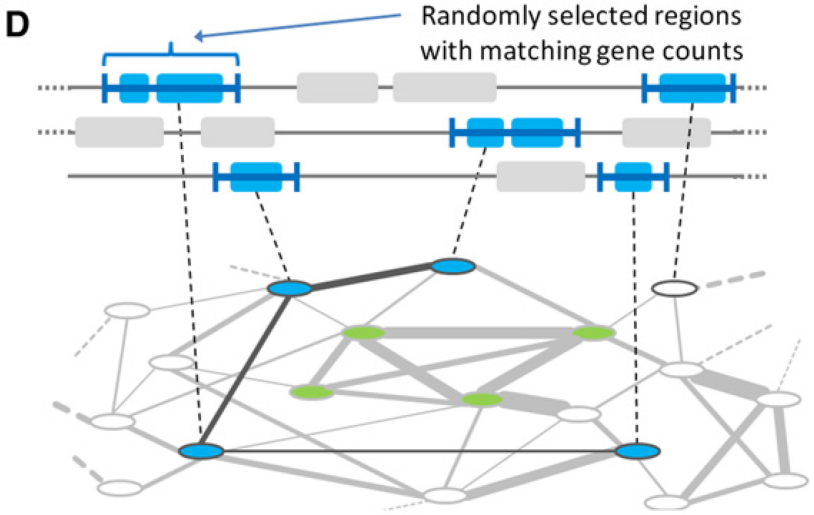
Naïve Bayesian integration of protein function descriptors:
- shared GO annotations;
- shared pathways in KEGG;
- protein domains from the InterPro database;
- tissue expression from the TiGER database;
- direct protein-protein interactions;
- shared interaction partners in databases;
- phylogenetic profiles;
- chromosomal co-clustering across genomes

Gilman, *et al., Neuron*, 70, 878-907 (2011)

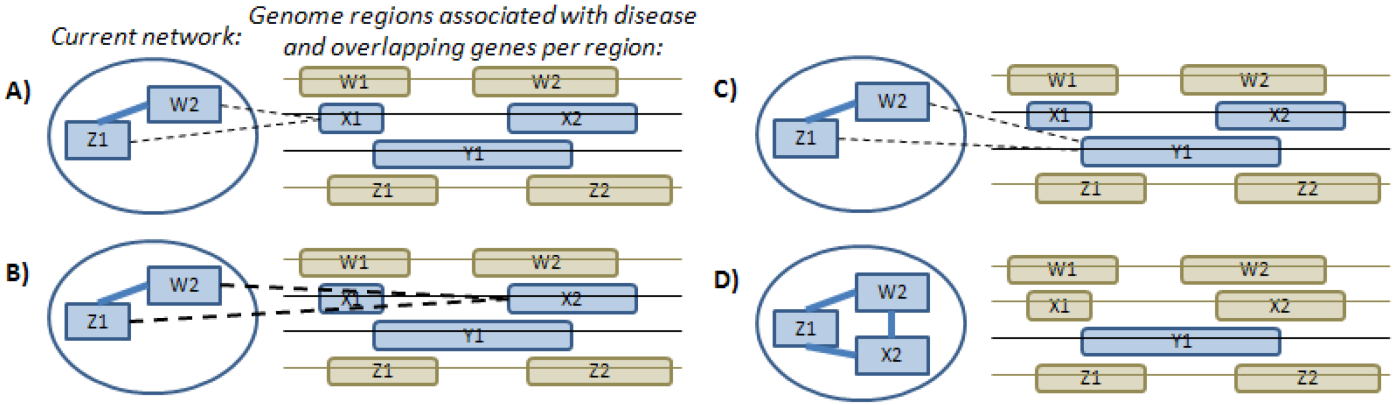
NETBAG search algorithm



Cluster or sub-cluster with strong connections selected

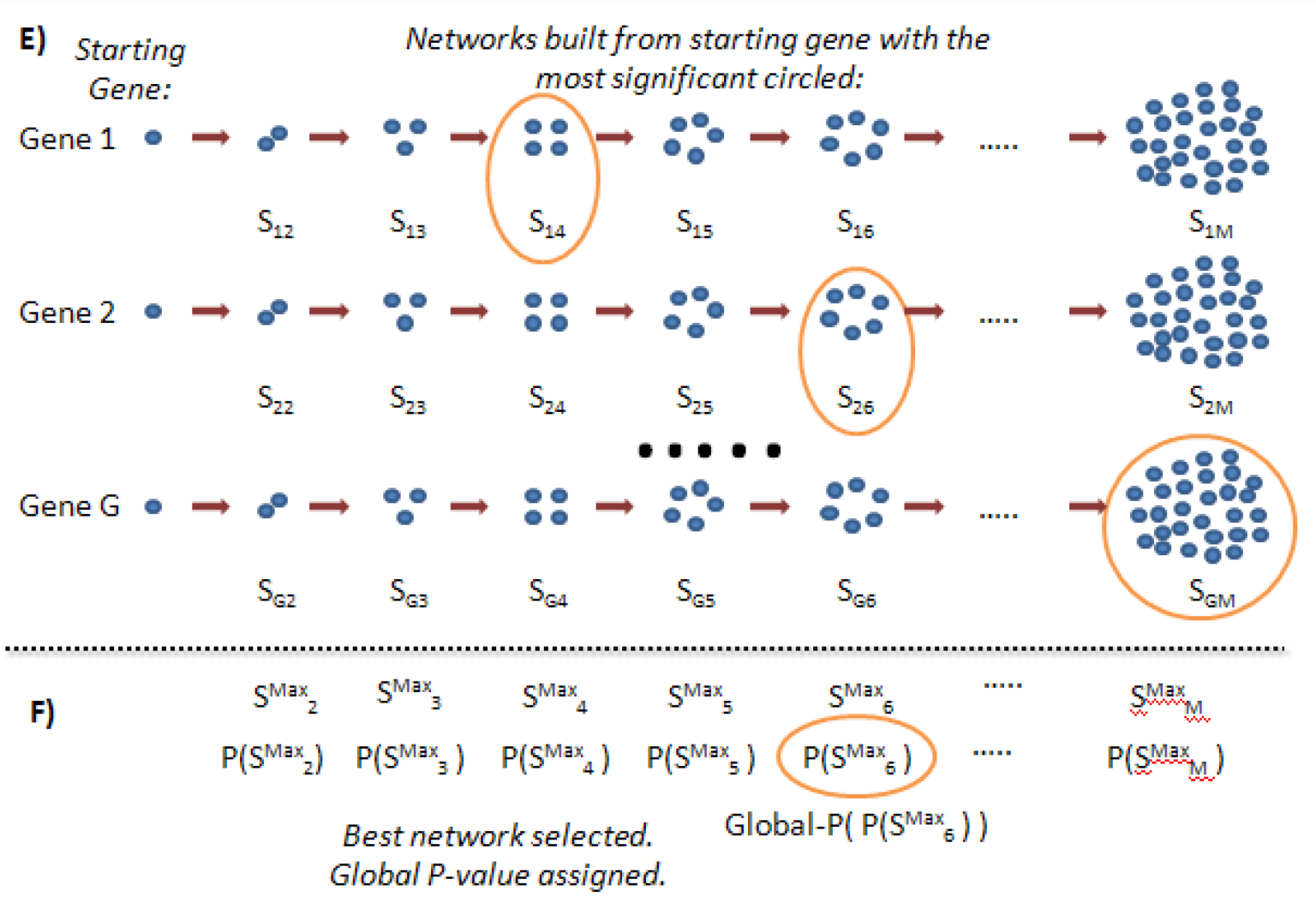


Cluster p-value determined by comparison with clusters built from randomly selected regions



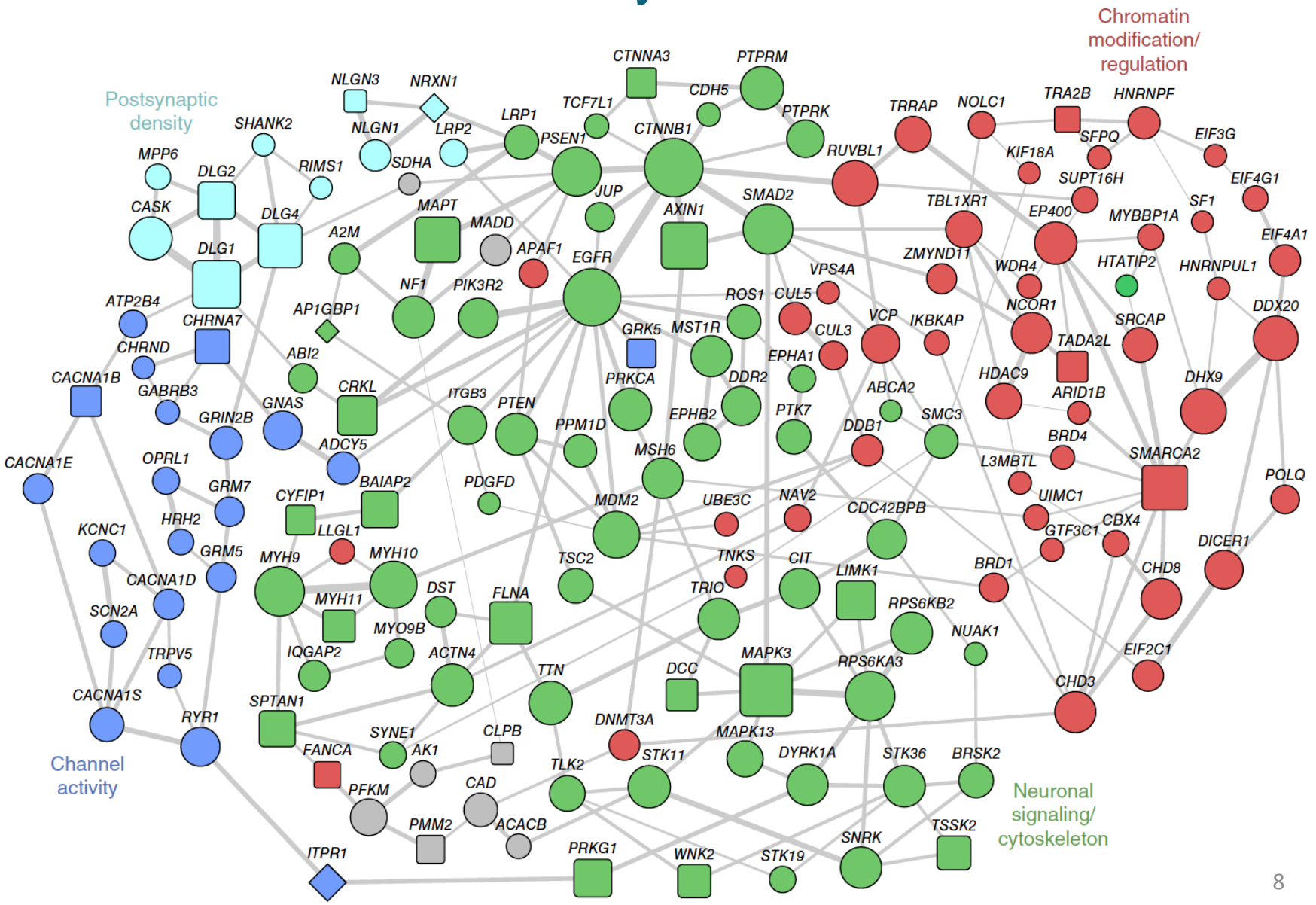
Gilman, *et al., Neuron*, 70, 878-907 (2011)

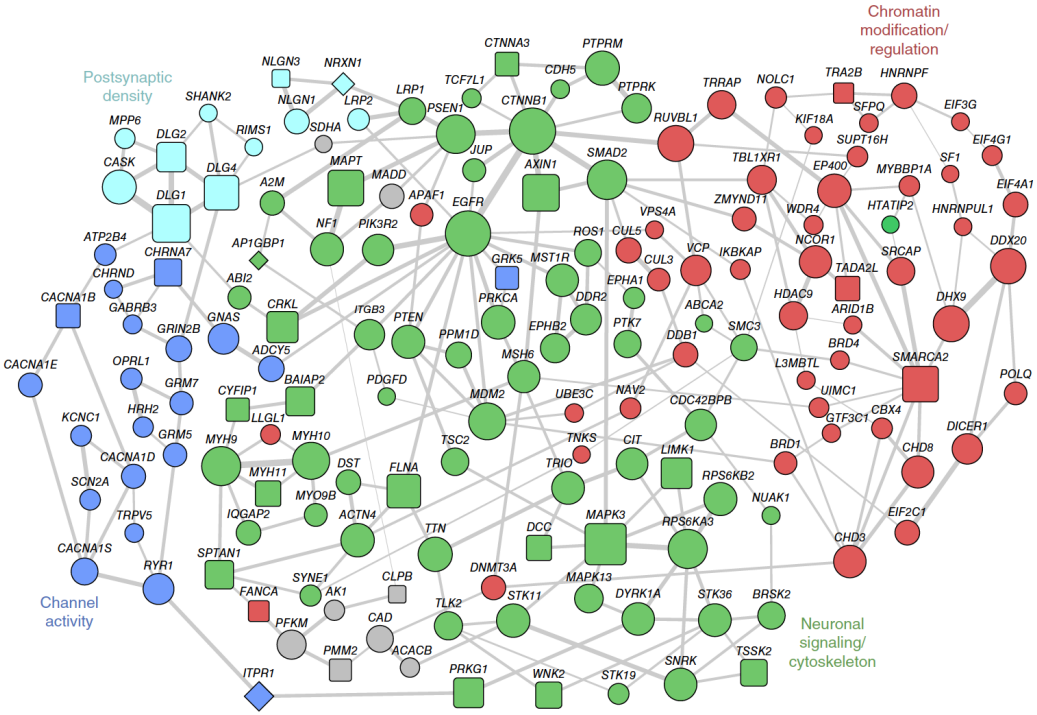
NETBAG search algorithm



Gilman, *et al., Neuron*, 70, 878-907 (2011)

The network based on ASD-associated *de novo* SNVs & CNVs by NETBAG+





- The NETBAG+ search revealed a functional network containing **159** genes (*P* = 0.036), of which **131** genes were affected by *de novo* **SNVs** and **31** by *de novo* **CNVs**
- *Node sizes* are proportional to the overall network score and *edge widths* are proportional to the likelihood that the corresponding gene pair contributes to the same genetic phenotype
- *Circles* represent genes from SNVs, *squares* represent genes from CNVs and *diamonds* represent genes affected by both
- No significant networks were detected using genes associated with the **368** non-synonymous *de novo* mutations identified in siblings
- Divided into four cohesive functional clusters (*node colors*) using average linkage hierarchical clustering (Biological functions of these clusters obtained by DAVID)

GO terms enriched in implicated network

Table 1 GO terms associated with the implicated network

GO ID	Ontology	Term	*P* value
GO:0045202	CC	Synapse	3×10^{-7}
GO:0016568	BP	Chromatin modification	6×10^{-6}
GO:0051015	MF	Actin filament binding	7×10^{-6}
GO:0005262	MF	Calcium channel activity	1×10^{-5}
GO:0031252	CC	Cell leading edge	2×10^{-5}
GO:0004386	MF	Helicase activity	3×10^{-5}
GO:0015629	CC	Actin cytoskeleton	6×10^{-5}
GO:0030036	BP	Actin cytoskeleton organization	7×10^{-5}
GO:0016887	MF	ATPase activity	7×10^{-5}
GO:0005913	CC	Cell-cell adherens junction	8×10^{-5}
GO:0045211	CC	Postsynaptic membrane	8×10^{-5}
GO:0044456	CC	Synapse part	0.0001
GO:0007611	BP	Learning or memory	0.0001
GO:0030029	BP	Actin filament-based process	0.0001
GO:0005938	CC	Cell cortex	0.0001
GO:0005911	CC	Cell-cell junction	0.0002
GO:0014069	CC	Postsynaptic density	0.002
GO:0043005	CC	Neuron projection	0.004
GO:0030425	CC	Dendrite	0.004

GO terms enriched in the implicated network, as identified by DAVID david.abcc.ncifcrf.gov. The ontology column indicates GO domain: **BP** for biological process, **MF** for molecular function and **CC** for cellular component.

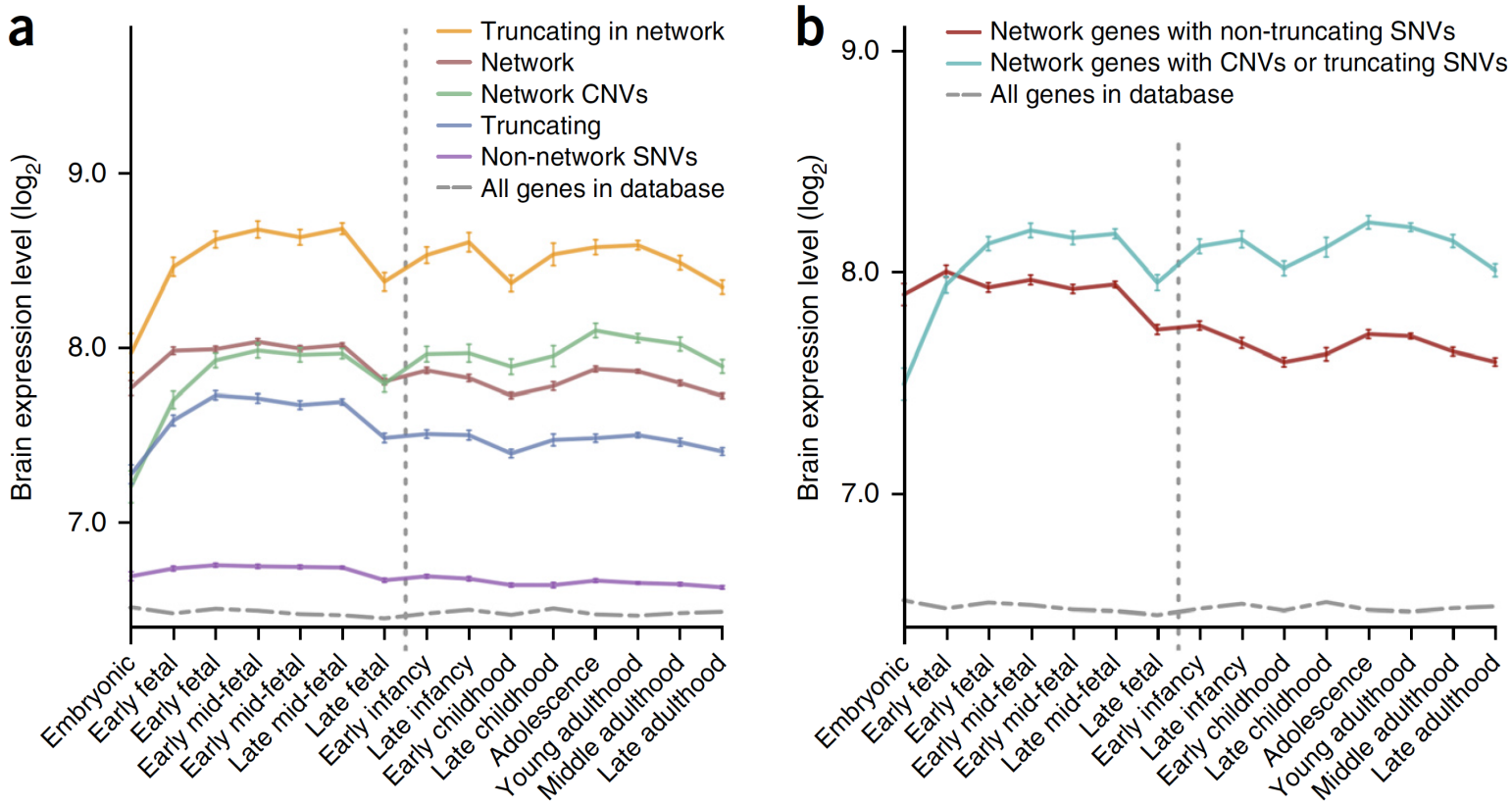
Association of implicated genes with functional subsets

- Enrichment of causal genes in the implicated network: 11 genes identified in previous ASD studies with recurrent truncating mutations in them (ADNP, ANK2, ARID1B, CHD8, CUL3, DYRK1A, GRIN2B,K ATNAL2, POGZ, SCN2A and TBR1). 6 of 11 genes with recurrent truncating mutations are in the network
- Genes in the network were significantly more likely to be haploinsufficient than genes that were not selected by NETBAG+ (median probabilities are 0.57 and 0.32)
- Ave GERP scores (characterizing the severity of SNV mutations) of network genes was higher than the average GERP score of genes not selected by NETBAG+ (4.0 and 3.3)
- PSD (Postsynaptic density) genes and FMRP (fragile X mental retardation protein) target genes enrichment:

Overlap between ASD gene sets and additional FMRP targets or PSD proteins

Gene set	Number of genes	Ascano *et al.* FMRP targets		Bayes *et al.* PSD proteins	
		Expected : observed	*P* value	Expected : observed	*P* value
Network genes	159	17.5 : 32 (1 : 1.83)	0.0008	8.52 : 29 (1 : 3.4)	$7x10^{-9}$
Truncating SNV genes	108	11.9 : 13 (1 : 1.09)	0.8	5.79 : 11 (1 : 1.9)	0.05
Network non-truncating SNV genes	138	15.2 : 25 (1 : 1.65)	0.013	7.4 : 22 (1 : 2.97)	$5x10^{-6}$
Neuronal signaling/cytoskeleton cluster genes	69	7.59 : 13 (1 : 1.71)	0.05	3.7 : 15 (1 : 4.06)	$3x10^{-6}$
Chromatin modification/regulation cluster genes	50	5.5 : 12 (1 : 2.18)	0.01	2.68 : 2 (1 : 0.746)	1
Postsynaptic density cluster genes	11	1.21 : 2 (1 : 1.65)	0.3	0.59 : 7 (1 : 11.9)	$3x10^{-7}$
Channel activity cluster genes	21	2.31 : 1 (1 : 0.43)	0.7	1.13 : 3 (1 : 2.66)	0.1
Non-network SNV genes	449	49.4 : 45 (1 : 0.91)	0.5	24.1 : 14 (1 : 0.582)	0.04
Sibling SNV genes	355	39.1 : 47 (1 : 1.20)	0.2	19 : 20 (1 : 1.05)	0.8

Temporal Expression Patterns

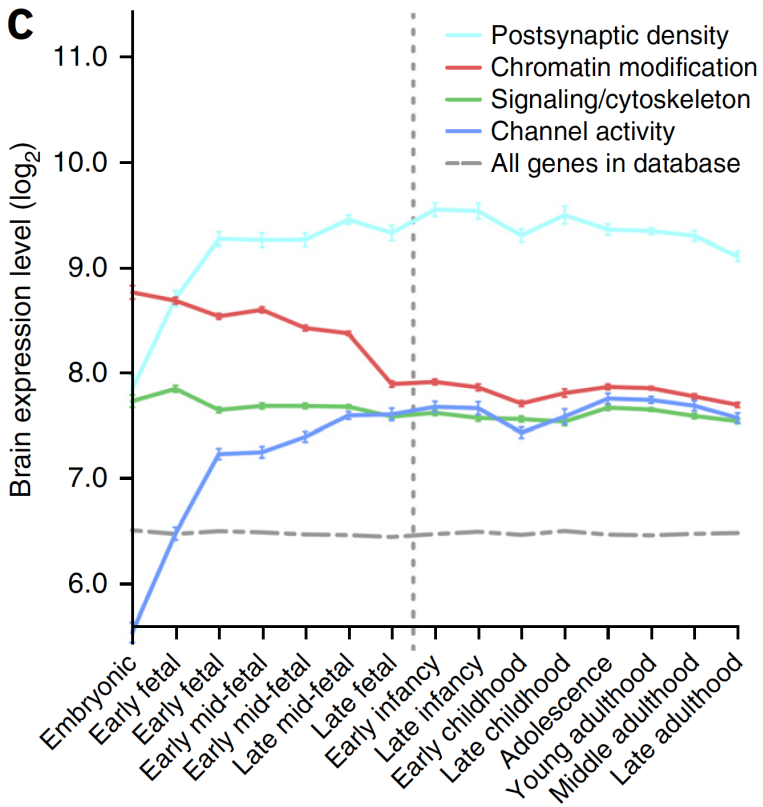


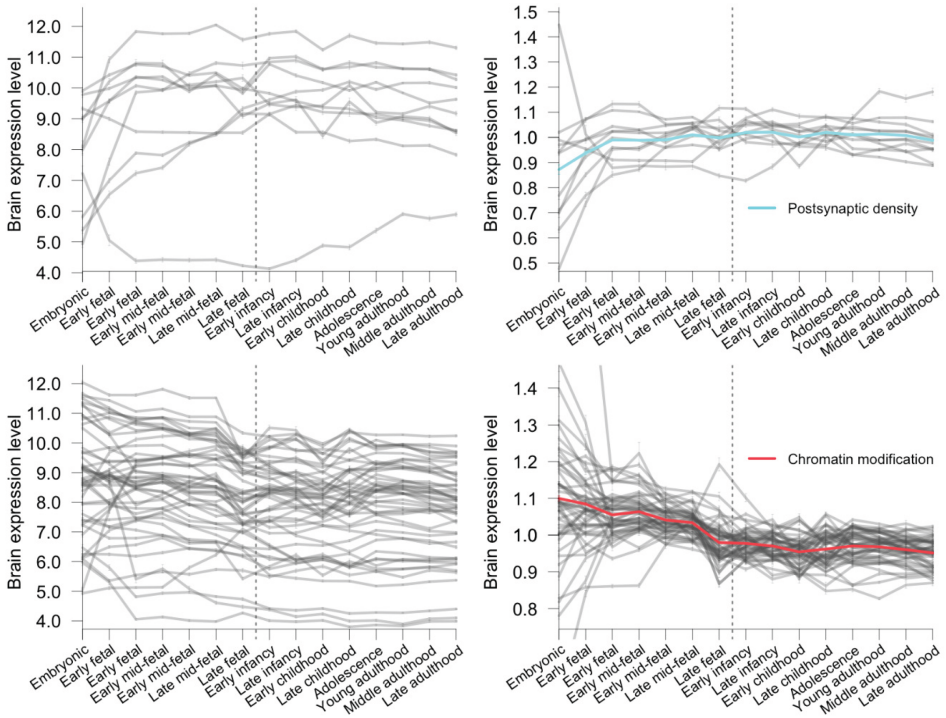
-- Expression data obtained from Human Brain Transcriptome database

-- Further analysis is focused on genes **implicated by the network** and genes affected by all de novo **truncating mutations** from other studies

-- The bias against mutations in the embryonic period was absent for non-synonymous SNV genes not selected in the network and for genes with *de novo* SNVs in siblings

Temporal Expression Patterns: network clusters



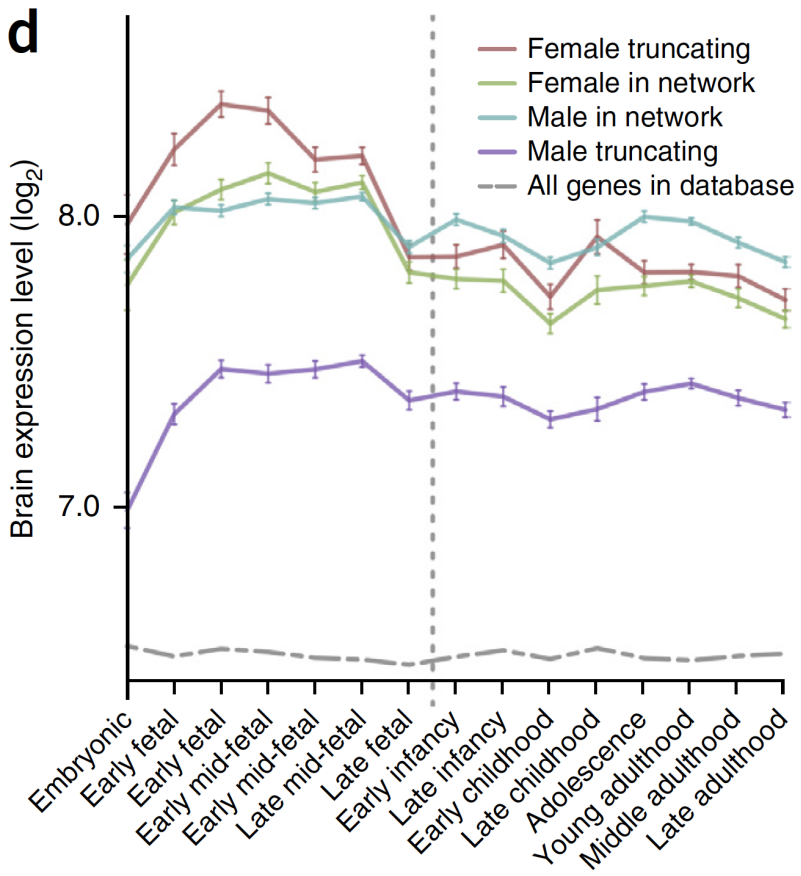


-- PSD and channel activity clusters show rise during early fetal development – start of synaptogenesis

-- Chromatin modification/regulation – high embryonic and fetal expression – peak of neuronal proliferation and differentiation

-- Signalling/cytoskeleton – constant across developmental stages – consistent with important roles of signalling and structural genes across all developmental stages

Temporal Expression Patterns: female/male



-- A high male-to-female incidence ratio, (> 4:1) for high-functioning individuals, is one of the most consistent and notable findings in ASDs

--Shows further evidence that this bias may be a result of a female protective effect that requires a higher threshold of genetic insults to trigger ASDs in females than in males

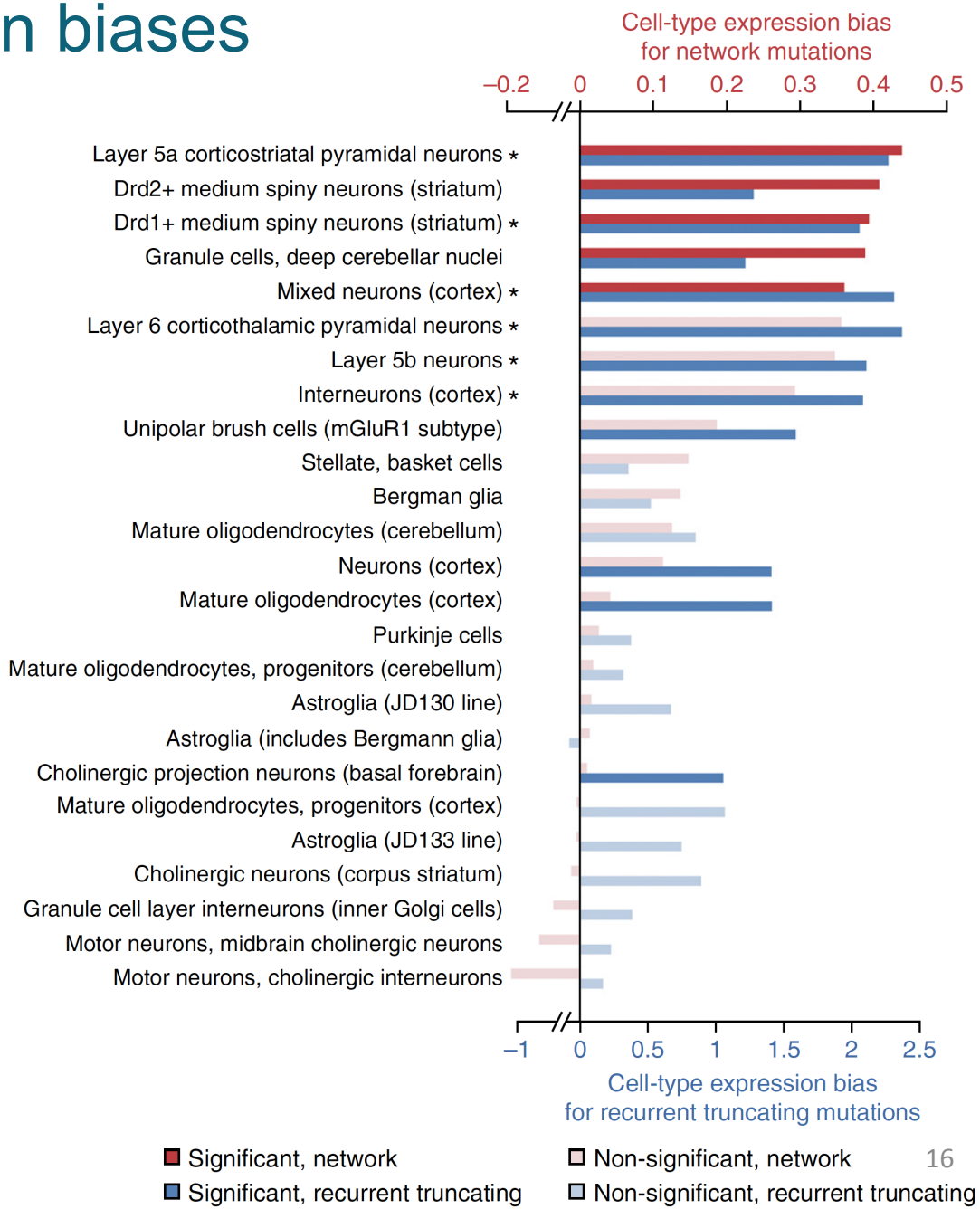
Prenatal *vs* postnatal brain expression

Table 3 Prenatal versus postnatal brain expression biases across brain regions

	Network genes			Truncating SNVs			Sibling SNVs		
Brain region	Bias	S.e.m.	*P* value	Bias	S.e.m.	*P* value	Bias	S.e.m.	*P* value
Amygdala	0.25	0.08	1×10^{-12}	0.26	0.10	2×10^{-6}	0.02	0.05	1.0
Striatum/BG	0.24	0.08	2×10^{-12}	0.22	0.10	3×10^{-5}	0.04	0.05	1.0
Cerebellum	0.23	0.08	3×10^{-9}	0.17	0.10	4×10^{-3}	0.07	0.05	0.1
Occipital	0.17	0.08	3×10^{-7}	0.15	0.10	3×10^{-2}	0.04	0.05	1.0
Parietal	0.16	0.08	3×10^{-11}	0.17	0.10	3×10^{-4}	0.05	0.05	0.7
Temporal	0.16	0.07	2×10^{-17}	0.19	0.09	2×10^{-8}	0.05	0.05	1.0
Hippocampus	0.15	0.07	1×10^{-5}	0.18	0.09	4×10^{-4}	0.02	0.05	1.0
Thalamus	0.14	0.08	1×10^{-4}	0.22	0.09	1×10^{-5}	0.03	0.05	1.0
Frontal	0.12	0.08	1×10^{-16}	0.17	0.09	4×10^{-10}	0.03	0.05	1.0

The biases were calculated using human expression data obtained from the HBT database. To quantify the expression bias for each brain region, we calculated the difference between the average $\log_2$ prenatal expression and the average $\log_2$ postnatal expression, such that positive values in the table indicate higher expression levels in the prenatal periods. The significances of the expression biases were evaluated using the Wilcoxon rank-sum one-tail test and corrected using the Bonferroni procedure. Bias s.e.m. was calculated separately for each brain region. BG, basal ganglia.

Cell-type expression biases

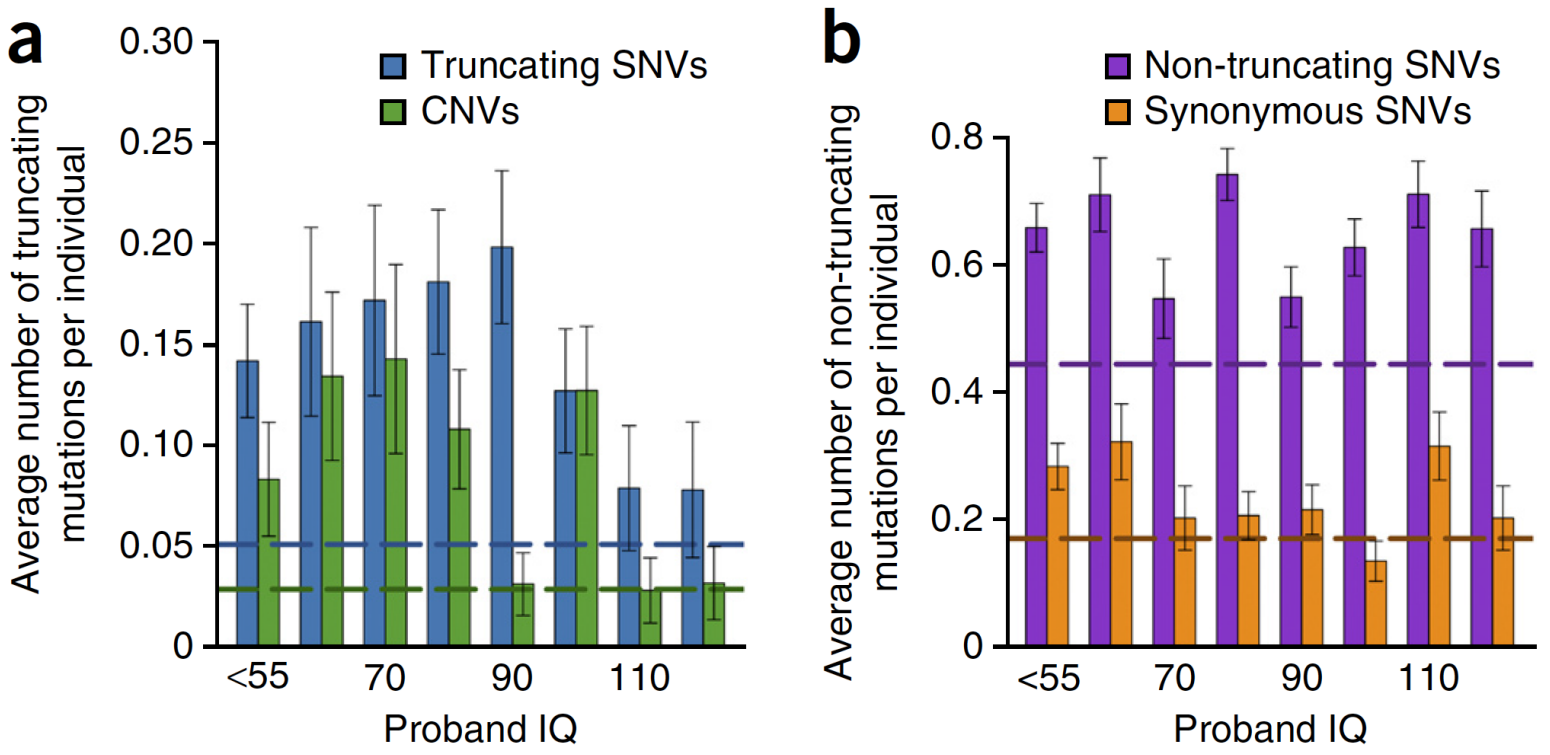


-- Proband vs unaffected sibling expression biases across implicated network genes and for 11 genes with recurrent truncating SNVs.

--The difference between the average $\log_2$ expression of mouse orthologs for implicated human genes and the average $\log_2$ expression of mouse ortholog for human genes with de novo SNVs in unaffected siblings

--using mouse CNS cell-specific data from Doyle *et al. Cell*, 135, 749, 2008

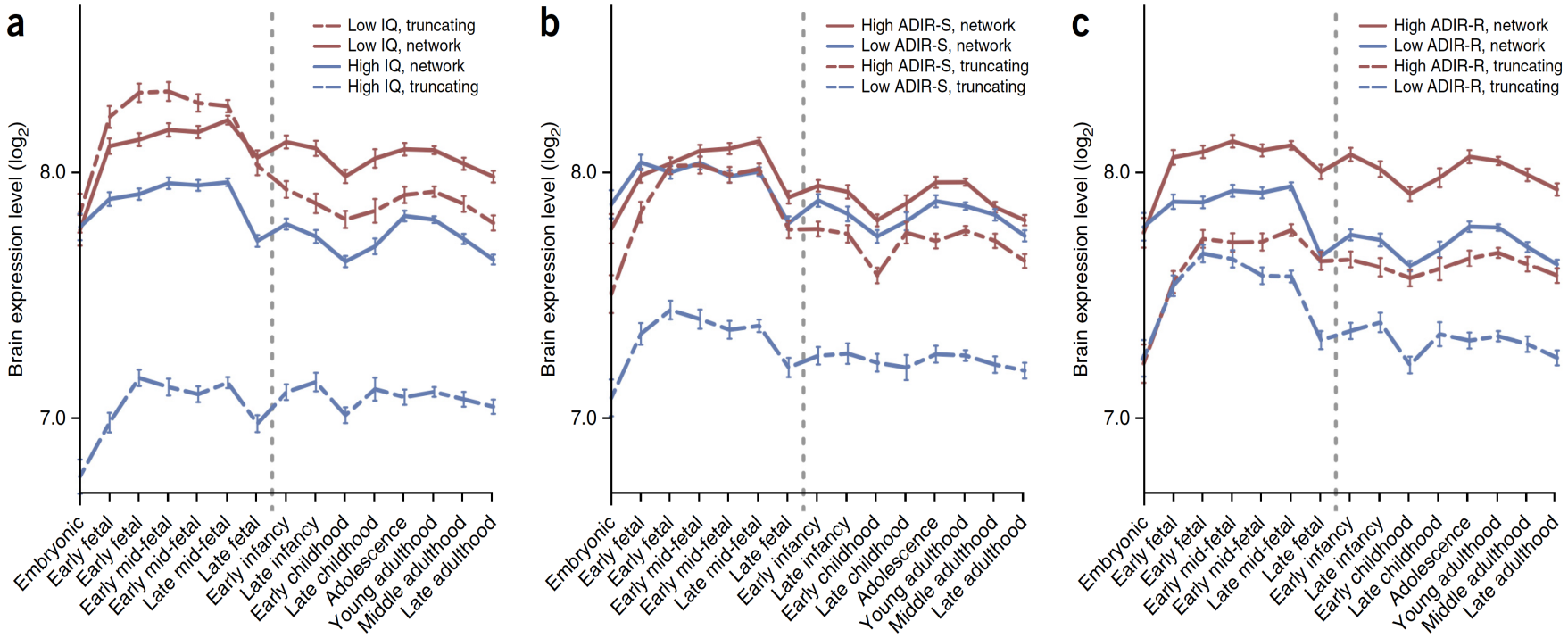
Average numbers of *de novo* mutations per individuals with different IQs



--The average number of truncating SNVs for probands with IQ <100 was about twice higher than probands with IQ>=100

--The average IQs for probands with CNV deletions and duplications, were 63.8 and 83.9, respectively (Wilcoxon rank-sum one-tail test, $P = 6 \times 10^{-3}$)

Average expression levels of implicated genes identified in phenotypic subsets



--**ADIR (Autism Diagnostic Interview Revised)** scores are based on structured interviews with proband parents and reflect patterns in reciprocal social interactions (ADIR-S) and repetitive/restrictive behaviors (ADIR-R)

Conclusions/Discussion

Pathophysiological heterogeneity of ASD is matched by the diversity of genetic and functional insults associated with the disorder

Affected genes are likely to be important in multiple stages of neurogenesis, neuron mobility, synaptogenesis and brain function

Implicated genes are active across multiple cell types; interneurons, pyramidal neurons and medium spiny neuron of the striatum seem to be among strongly affected ones

The study shows that in addition to truncating de novo mutations in ASD, some non-truncating de novo missense mutations also contribute to the disorder

Functional characteristics of affected genes, such as brain expression levels, are likely to influence the observed phenotypic consequences

Casual mutations target haploinsufficient genes: genes in the implicated network were significantly more likely to be haploinsufficient

Further evidence that stronger functional insults (perturbation of genes with higher brain expression) are associated with female autistic phenotypes

Further evidence that truncating mutations play a smaller role in high-functioning cases