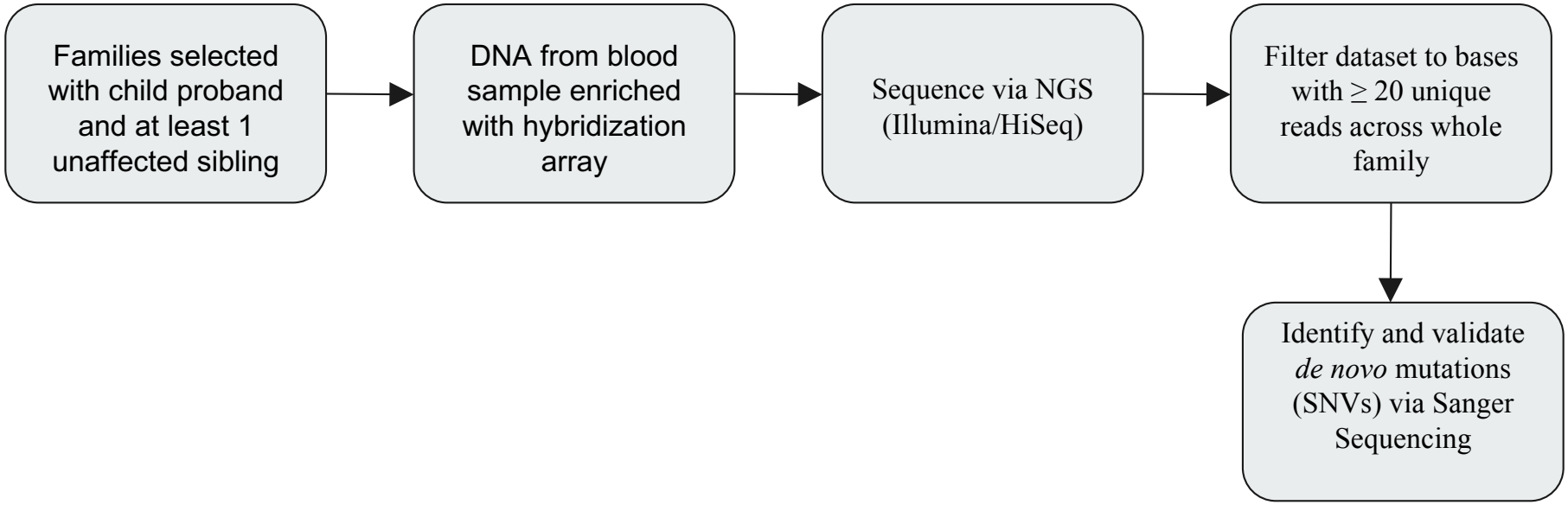
De Novo mutations revealed by whole-exome sequencing are strongly associated with autism (Sanders et al., Nature 2012)

Presented by In Ae (Sophia) Cho, Sidonie K Kilpatrick, Austin Crinklaw

Introduction

- Authors set out to establish links between *de novo* mutations and autism spectrum disorders (ASD)
- Big Question: How can high-throughput sequencing technology aid in identifying and diagnosing genetic risk factors for disease phenotypes
- Idea: Sequence whole-exomes of select individuals and isolate mutations that have strong correlation with ASD

Exome Sequencing and Variant Identification

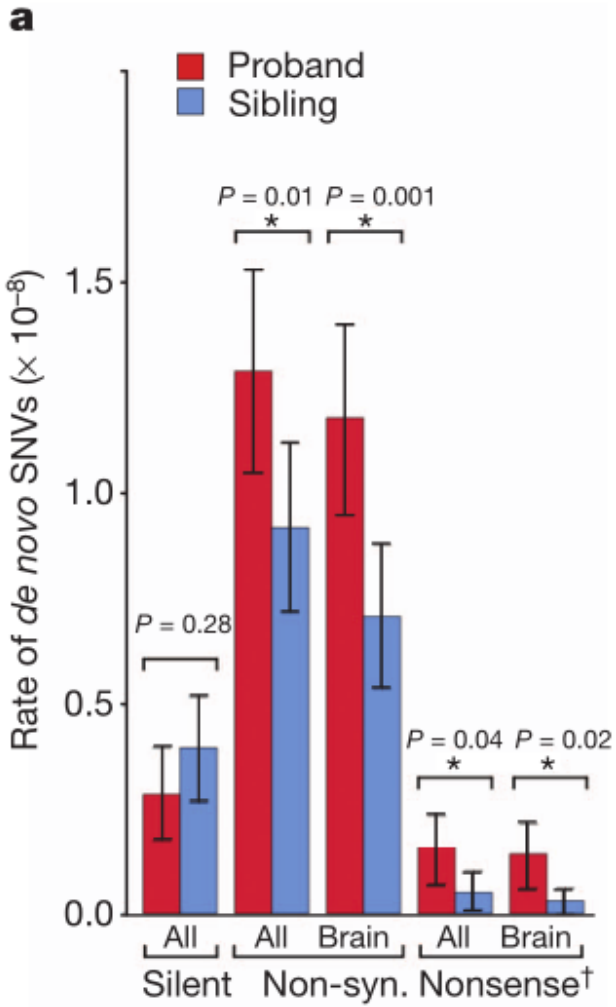


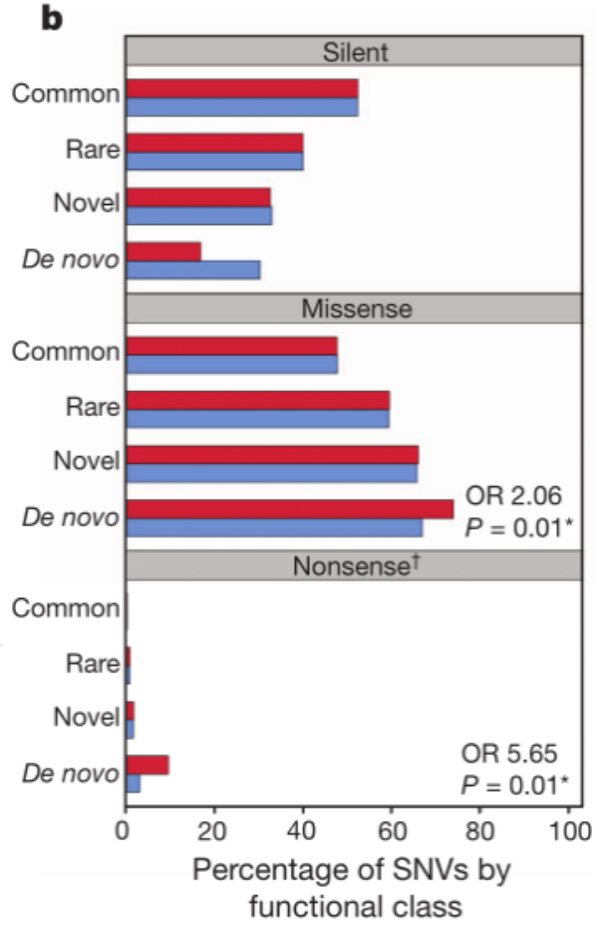
Data Filtration and Classification

- Variants classified as novel only if a single allele present in parent, none seen in control exomes
- Variants analyzed against RefSeq hg18 gene database (for multiple isoforms, highest impact isoform was selected)
- Genes were classified as brain expressed using the Stanley Neuropathology Consortium Database (SNCID)

Interpreting Results

- No evidence of systematic bias in variant detection between individual with ASD and unaffected siblings
    - Siblings were controls
- Compared silent *de novo,* non-coding *de novo,* and novel transmitted variants
- Non-synonymous *de novo* SNVs in proband > unaffected siblings





- 200 probands and their unaffected sibling for all mutation types and allele frequencies
- Proportion of transmitted variants in brain-expressed genes were the same except for missense and nonsense *de novo*.

- Frequency distribution of brain-expressed missense *de novo* SNVs
- Red= probands
- Blue = siblings
- Data suggests that multiple *de novo* SNVs in a single individual do not confirm ASD risk
    - Conclude this because data follows Poisson distribution

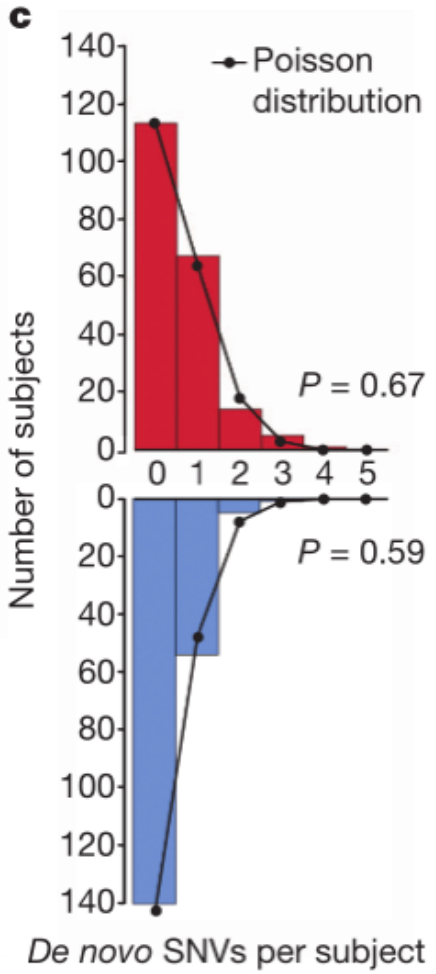


Table 1 | Distribution of SNVs between probands and siblings

Category	Total number of SNVs*		SNVs per subject		Per base SNV rate (x10^{-8})		*P*†	Odds ratio (95% CI)‡
	Pro *N* = 200	Sib *N* = 200	Pro *N* = 200	Sib *N* = 200	Pro *N* = 200	Sib *N* = 200		
De novo								
				All genes				
All	154	125 §	0.77	0.63	1.58	1.31	0.09	NA
Silent	29	39	0.15	0.20	0.29	0.40	0.28	NA
All non-synonymous	125	87	0.63	0.44	1.29	0.92	**0.01**	1.93 (1.11–3.36)
Missense	110	82	0.55	0.41	1.13	0.86	0.05	1.80 (1.03–3.16)
Nonsense/splice site	15	5	0.08	0.03	0.16	0.05	**0.04**	4.03 (1.32-12.4)
				Brain-expressed genes				
All	137	96	0.69	0.48	1.41	1.01	**0.01**	NA
Silent	23	30	0.12	0.15	0.24	0.31	0.41	NA
All non-synonymous	114	67	0.57	0.34	1.18	0.71	**0.001**	2.22 (1.19–4.13)
Missense	101	64	0.51	0.32	1.04	0.68	**0.005**	2.06 (1.10–3.85)
Nonsense/splice site	13	3	0.07	0.02	0.14	0.03	**0.02**	5.65 (1.44–22.2)
Novel transmitted								
				All genes				
All	26,565	26,542	133	133	277	277	0.92	NA
Silent	8,567	8,642	43	43	90	91	0.57	NA
All non-synonymous	17,998	17,900	90	90	188	187	0.61	1.01 (0.98–1.05)
Missense	17,348	17,250	87	86	181	180	0.60	1.01 (0.98–1.05)
Nonsense/splice site	650	650	3.3	3.3	7	7	1.00	1.01 (0.90–1.13)
				Brain-expressed genes				
All	20,942	20,982	105	105	219	220	0.85	NA
Silent	6,884	6,981	34	35	72	74	0.42	NA
All non-synonymous	14,058	14,001	70	70	147	146	0.74	1.02 (0.98–1.06)
Missense	13,588	13,525	68	68	142	141	0.71	1.02 (0.98–1.06)
Nonsense/splice site	470	476	2.3	2.4	5	5	0.87	1.00 (0.88–1.14)

* An additional 15 *de novo* variants were seen in the probands of 25 trio families; all were missense and 14 were brain-expressed.

† The *P* values compare the number of variants between probands and siblings using a two-tailed binomial exact test (Supplementary Information); *P* values below 0.05 are highlighted in bold.

‡ The odds ratio calculates the proportion of variants in a specific category to silent variants and then compares these ratios in probands versus siblings. NA, not applicable.

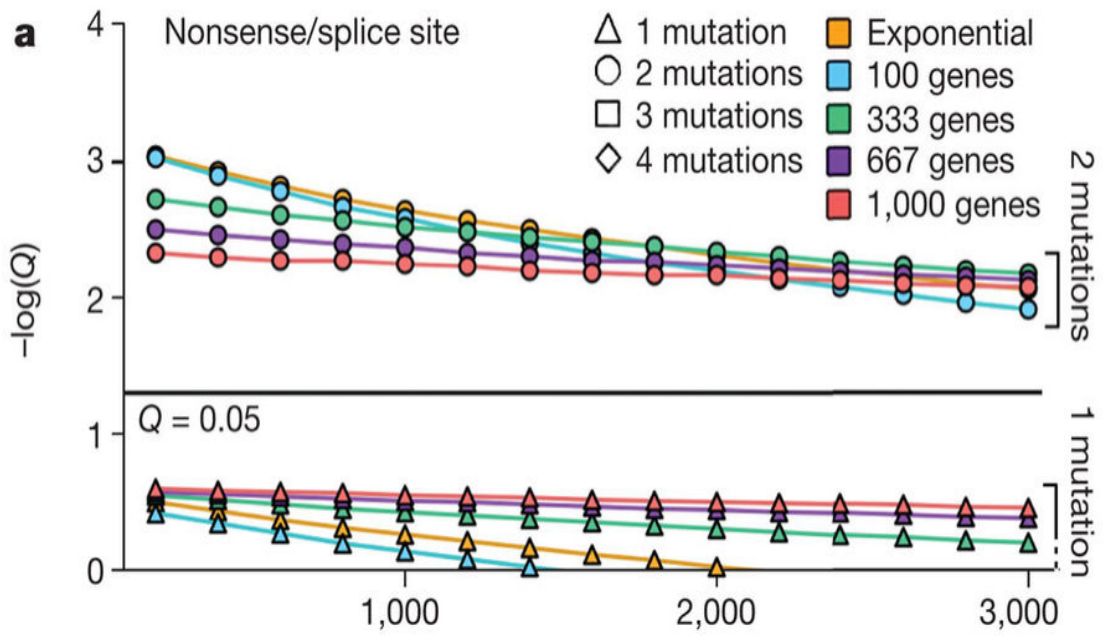
§ The sum of silent and non-synonymous variants is 126, however one nonsense and two silent *de novo* variants were indentified in *KANK1* in a single sibling, suggesting a single gene conversion event. This event contributed a maximum count of one to any analysis.

Figure 2: Identification of multiple *de novo* mutations in the same gene reliably distinguishes risk-associated mutations.

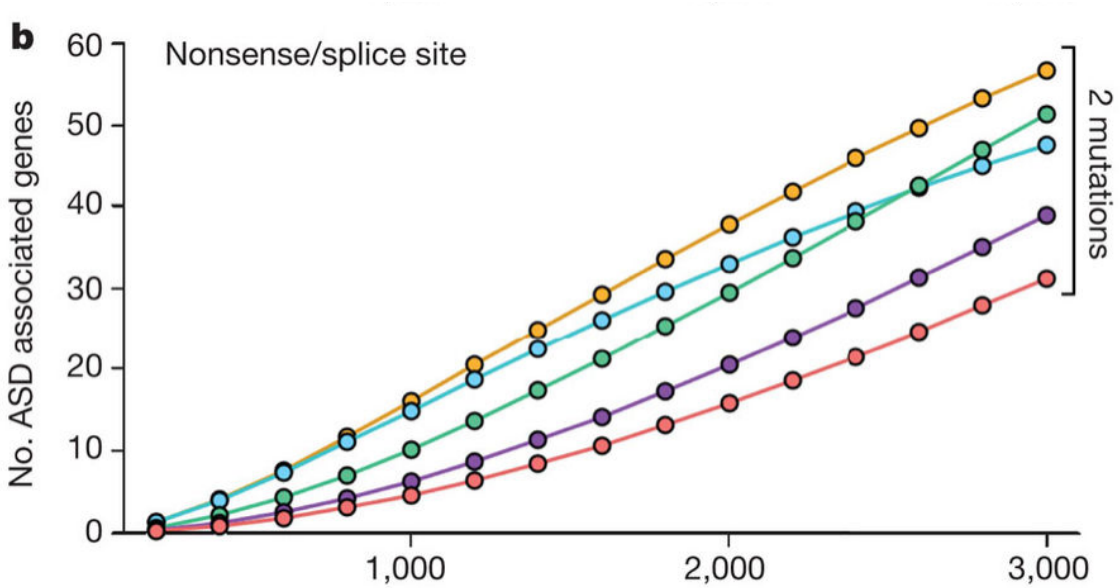
Hypothesis: estimating the probability of observing multiple independent *de novo* SNVs in the same gene in unrelated individuals would provide a more powerful statistical approach to identifying ASD-risk genes than the alternative of comparing mutation counts in affected versus unaffected individuals.

Experiments used: simulation model that focused on *de novo* SNVs in brain expression gene

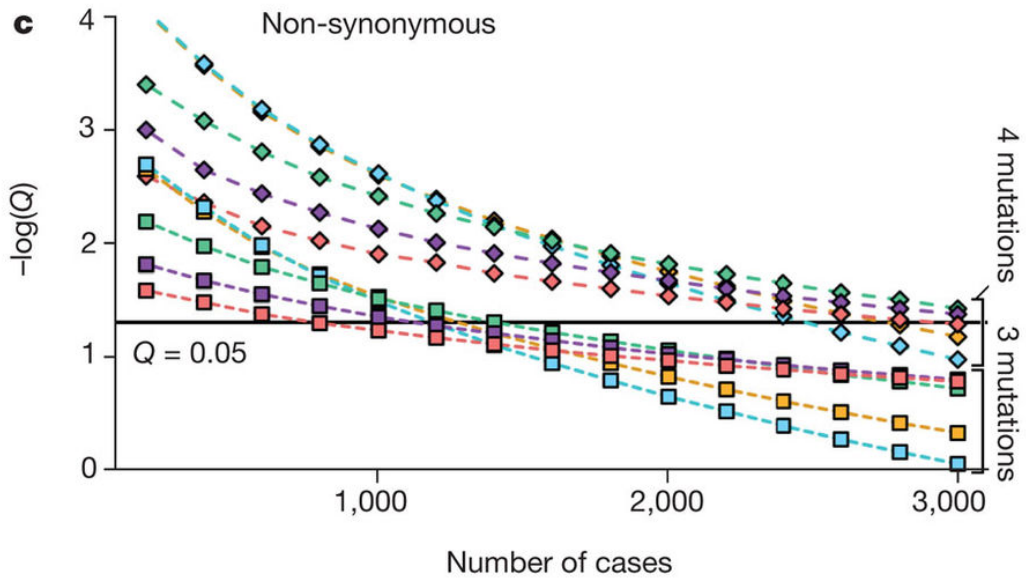
- Observing multiple independent de novo events in an ASD risk-conferring gene
- Gene size and GC content (derived from the full set of brain expressed RefSeq genes) of brain-expressing gene used to evaluate the number of genes contributing to ASD showing two or more variants of the specified type



- Two independent nonsense/splice site *de novo* mutations in the same brain-expressed gene among unrelated probands
- Q (false discovery rate) = 0.005
- P (probability) = 0.008
- two independent nonsense/splice site *de novo* variants in a brain-expressed gene provides significant evidence for ASD association for all models



- the number of genes that will be found to carry two or more nonsense/splice site *de novo* mutations for a sample of a given size (specified on the *x* axis).
- 25-50 additional ADS-risked genes
- sequencing of the 2,648 SSC



- The simulation model for non-synonymous *de novo* mutations.
- The identification of three or more independent non-synonymous *de novo* mutations in a brain-expressed gene
- Significant evidence for ASD association ($P < 0.05$; $Q < 0.05$)

Further Data Analysis

- Significant association of SCN2A, KATNAL2, and CHD8 predict 25-50 additional ASD-risk genes sequencing 2,648 SSC families
- Non-synonymous *de novo* SNVs are associated with risk
- Rare *de novo* copy number variations to the risk for ASD
- Multiple independent *de novo* mutations in brain-expressed genes are associated with ASD
- Multiple independent *de novo* single nucleotide variants in the same gene among unrelated probands reliably identifies risk alleles

Table 2 | Loss of function mutations in probands

Gene symbol	Gene name	Mutation type
ADAM33	ADAM metallopeptidase domain 33	Nonsense
CSDE1	cold shock domain containing E1, RNA-binding	Nonsense
EPHB2	EPH receptor B2	Nonsense
FAM8A1	family with sequence similarity 8, member A1	Nonsense
FREM3	FRAS1 related extracellular matrix 3	Nonsense
MPHOSPH8	M-phase phosphoprotein 8	Nonsense
PPM1D	protein phosphatase, Mg^{2+}/Mn^{2+} dependent 1D	Nonsense
RAB2A	RAB2A, member RAS oncogene family	Nonsense
SCN2A	sodium channel, voltage-gated, type II, α subunit	Nonsense
SCN2A	sodium channel, voltage-gated, type II, α subunit	Nonsense
BTN1A1	butyrophilin, subfamily 1, member A1	Splice site
FCRL6	Fc receptor-like 6	Splice site
KATNAL2	katanin p60 subunit A-like 2	Splice site
NAPRT1	nicotinate phosphoribosyltransferase domain containing 1	Splice site
RNF38	ring finger protein 38	Splice site
SCP2	sterol carrier protein 2	Frameshift*
SHANK2	SH3 and multiple ankyrin repeat domains 2	Frameshift*

* Frameshift *de novo* variants are not included in any of the reported case-control comparisons (Supplementary Information).

- SCN2A is the only nonsense variant gene which two unrelated probands carried
- SCN2A mutation associated with seizures, explaining link between ASD and seizure symptoms
- KATNAL2, and CHD8 loss of function mutation has correlation with ASD in further studies (Stessman et al. 2017, Cotney et al. 2014 respectively)

Do Other Factors Explain Findings?

- Rates of SNVs were higher correlating with paternal age, but accounting for age did not change results
- No significant relationship between proband IQ and rates of *de novo* SNVs was observed
- Similarly, no significant relationship between proband sex and rates of *de novo* SNVs was observed

Conclusion and Future Direction

- Rates of *de novo* SNVs are highly correlated with ASD, particularly those found in brain expressed genes
- Differential whole-exome analysis can provide insights into genetic risk factors in ASD
- Since Publishing: Other scientists have integrated gene co-expression analysis to find further differences in the transcriptomes of those with ASD versus controls (Gupta et al. 2014)