Genetic Architecture in Autism Spectrum Disorders

Danny Antaki, Sebat Lab
Biomedical Sciences Graduate Program
University of California, San Diego

Thesis Committee Meeting May 2, 2018

Mini CV

- Publications
  - Brandler, W. M.*, Antaki, D.*, Gujral, M.*, et al. Paternally inherited cis-regulatory structural variants are associated with autism. *Science*, 2018
  - Antaki, D., Brandler, W. M., Sebat J. SV^2: Accurate structural variation genotyping and de novo mutation detection from whole genomes. *Bioinformatics*, 2018
  - Brandler, W. M.*, Antaki, D.*, Gujral, M.*, et al. Frequency and complexity of de novo structural mutation in autism. *The American Journal of Human Genetics,* 2016
  - Breuss, M., Kleiber, M., George, R. D., Antaki, D., et al. Quantification of autism recurrence risk by direct assessment of paternal sperm mosaicism. *bioRxiv*, 2017
- Consortium Publications
  - Chaisson, M. J., et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *bioRxiv,* 2017 (1000 Genomes)
  - Marshall, C. R., et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature Genetics,* 2017 (PGC)
  - Sudmant, P. H., et al. "An integrated map of structural variation in 2,504 human genomes." *Nature,* 2015 (1000 Genomes)
- Posters
  - SV^2: Accurate Structural Variation Genotyping and De Novo Mutation Detection for Whole Genomes. *American Society of Human Genetics*, Orlando FL 2017
  - Whole Genome Sequencing Identifies Complex and Balanced De Novo Structural Variation in Autism. World Congress of Psychiatric Genetics, Toronto Canada 2015
  - Whole Genome Sequencing Identifies Complex and Balanced De Novo Structural Variation in Autism. American Society of Human Genetics, Baltimore MD 2015
  - Discovery, validation and genotyping of CNVs by analysis of genome sequence and microarray. American Society of Human Genetics, San Diego CA 2014
- Training Programs
  - Genetics Training Program 2014-2017 NIH predoctoral training grant T32 GM008666

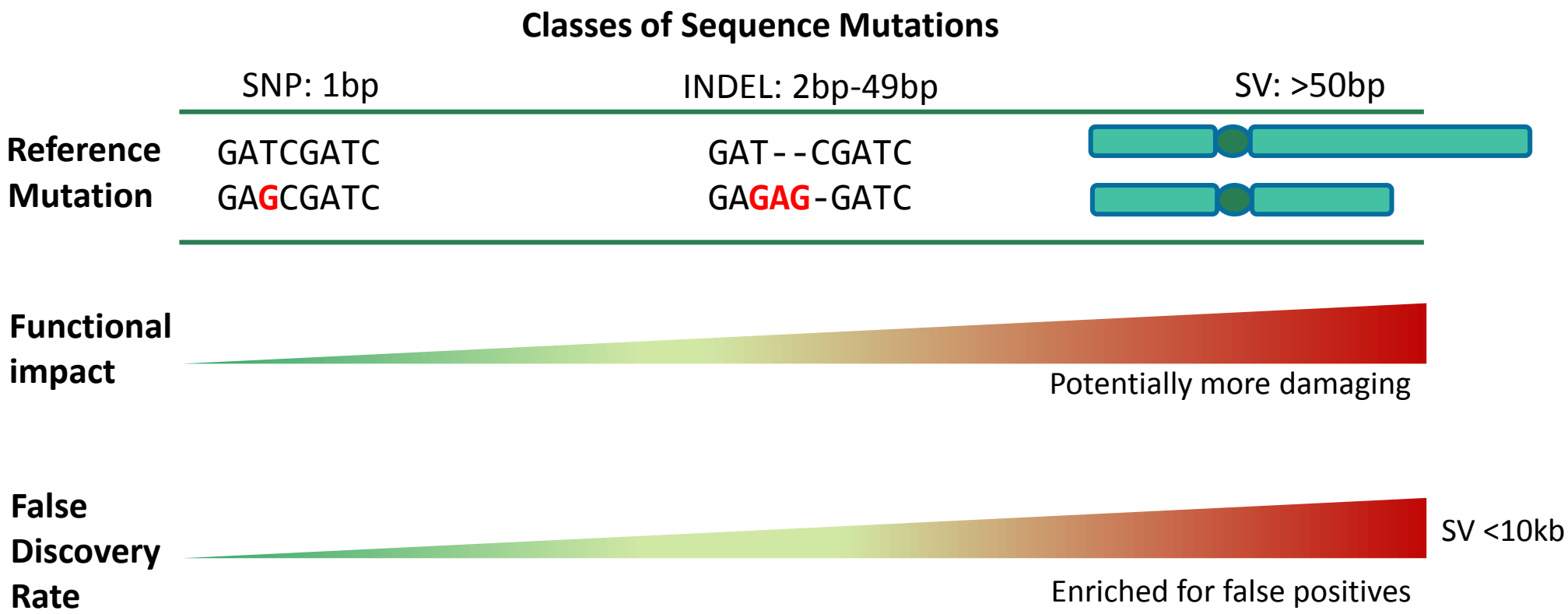
Aims

1. Refine Structural Variant (SV) calling in whole genome sequences (WGS)
   - SV^2: genotyping SVs with machine learning

2. Assay the burden of de novo SV in Autism Spectrum Disorders (ASD)
   - WGS offers optimal resolution compared to other methods

3. Interrogate the role of rare inherited variants in ASD
   - Identifying genetic risk stratified by parent of origin

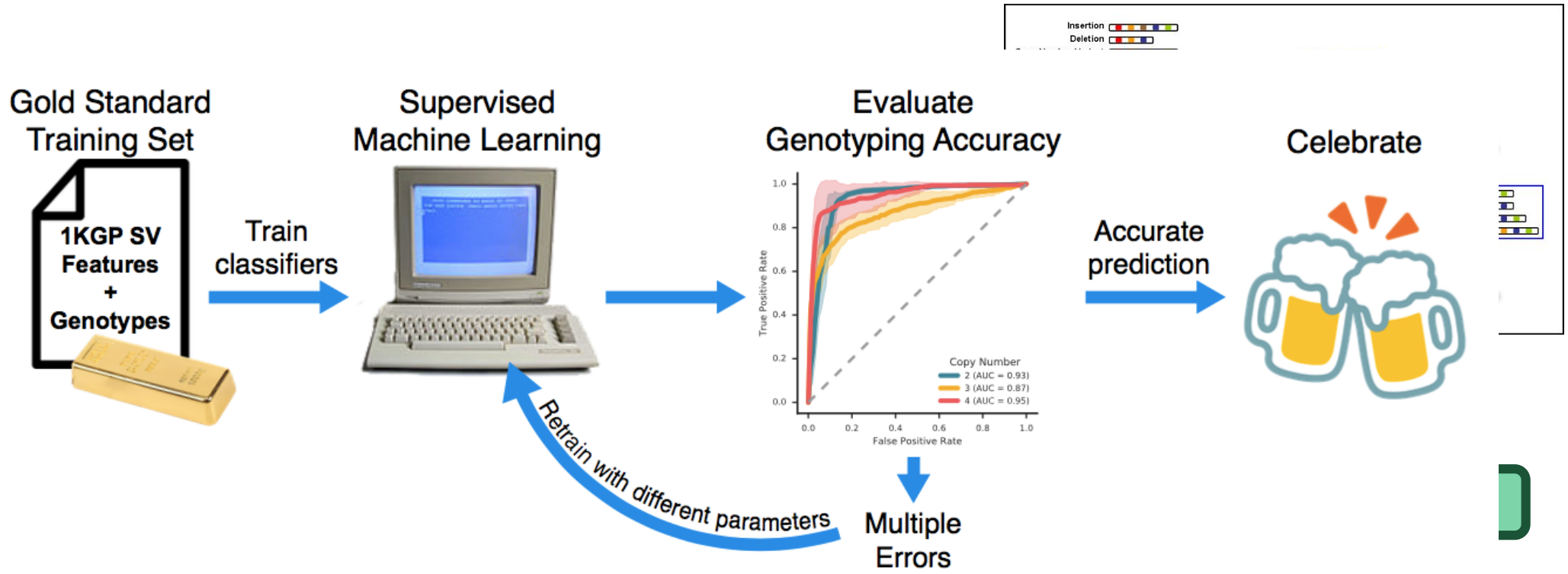
Rationale for Structural Variants

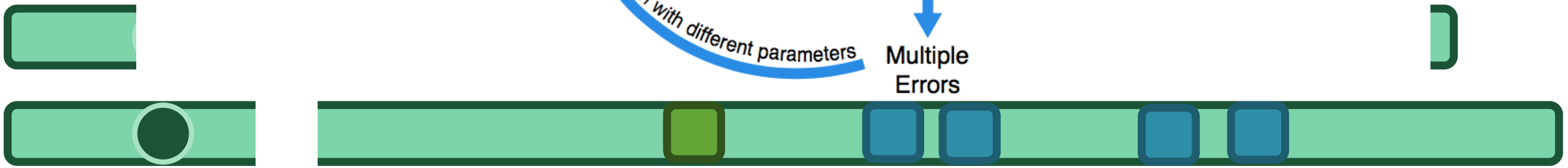


SVs are more likely to elicit a functional change but detection carries the cost of a high FDR

Detecting genotyping errors with machine learning

- Tra





Homozygous Deletion: 0 copies

Heterozygous Deletion: 1 copy

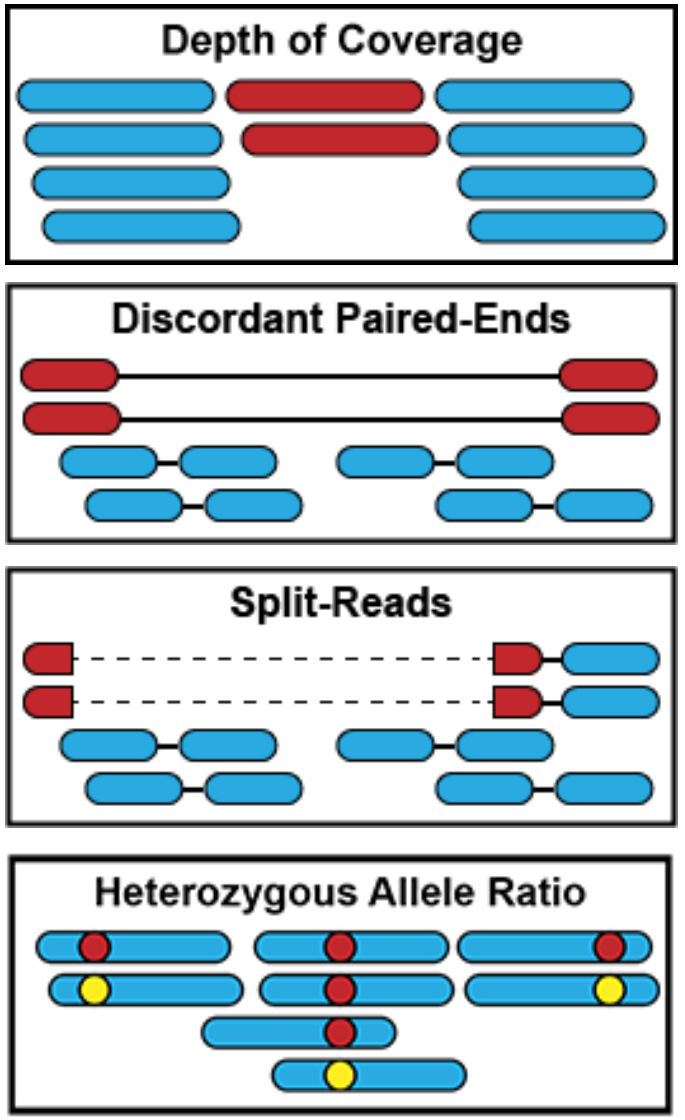
Homozygous neutral : 2 copies

Heterozygous Duplication : 3 copies

Homozygous Duplication : 4 copies

Features of SV in Next-Gen Sequencing

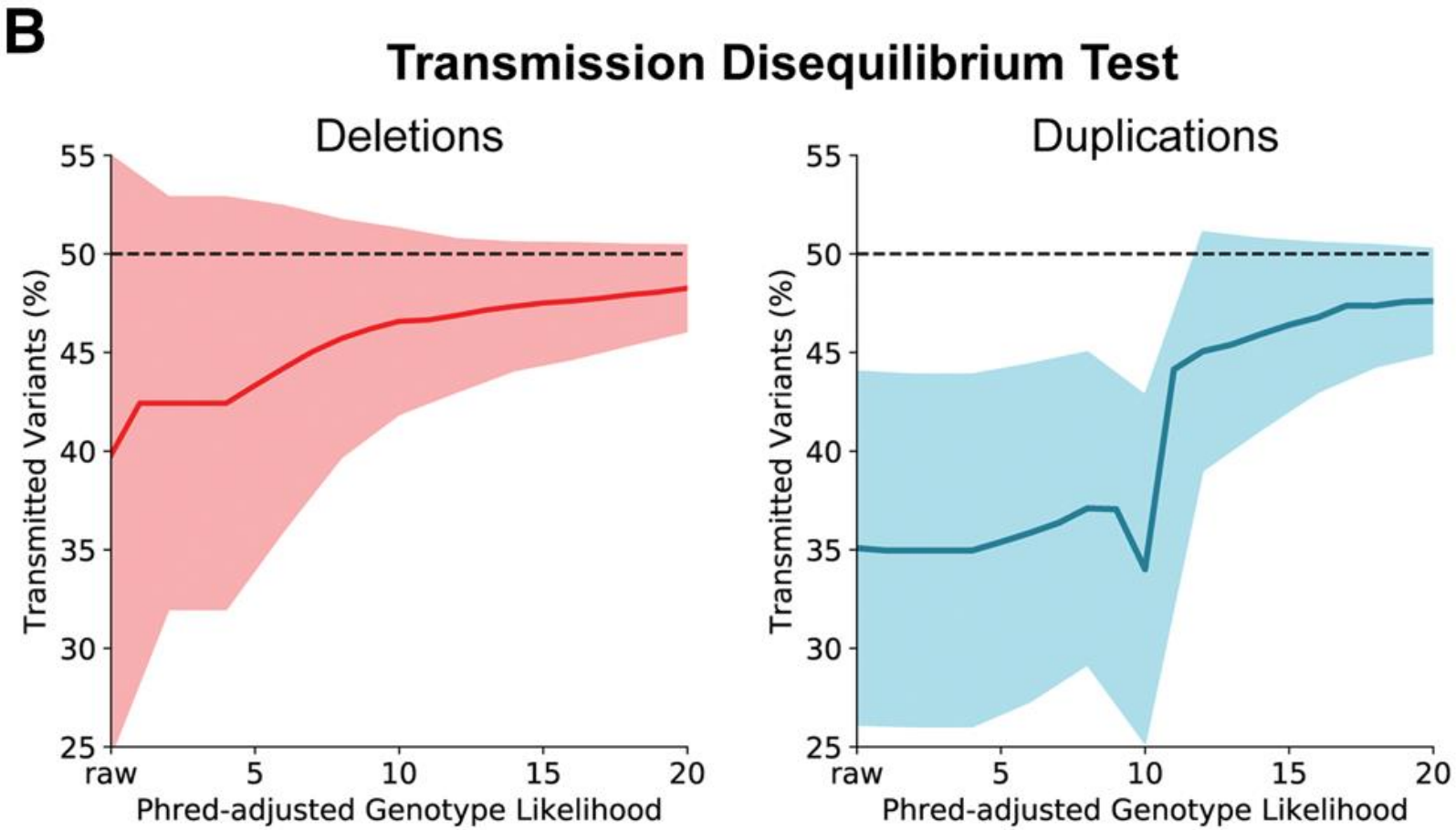
- Coverage: Number of reads spanning the SV
- Discordant Paired-Ends: abnormal insert sizes
- Split-Reads: reads that span SV breakpoints
- Heterozygous Allele Ratio:
  - Duplications only



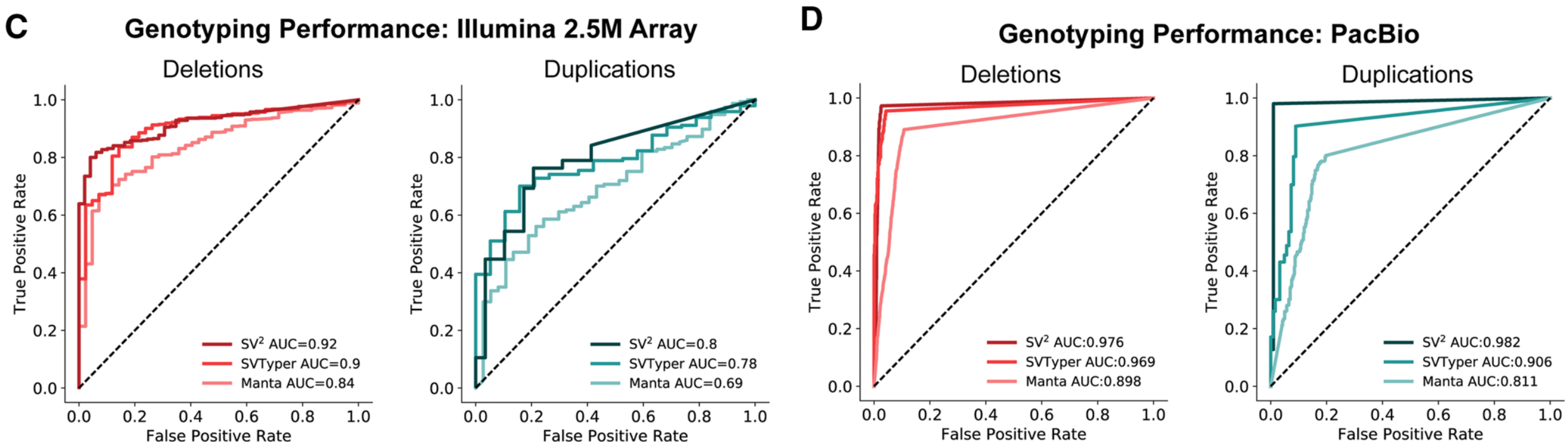
Genotyping Performance

Trio-based validation:
- Transmission Disequilibrium Test
  - deviations from the expected transmission rate (50%) are due to genotyping errors
- False positives in parents

Orthogonal validation:
- Microarray: N=57
  - SVTookit validated WGS SV genotypes
- PacBio SMRT: N=9
- False negatives in offspring
  - Split-reads or CIGAR string



Comparing SV^2 to other probabilistic models

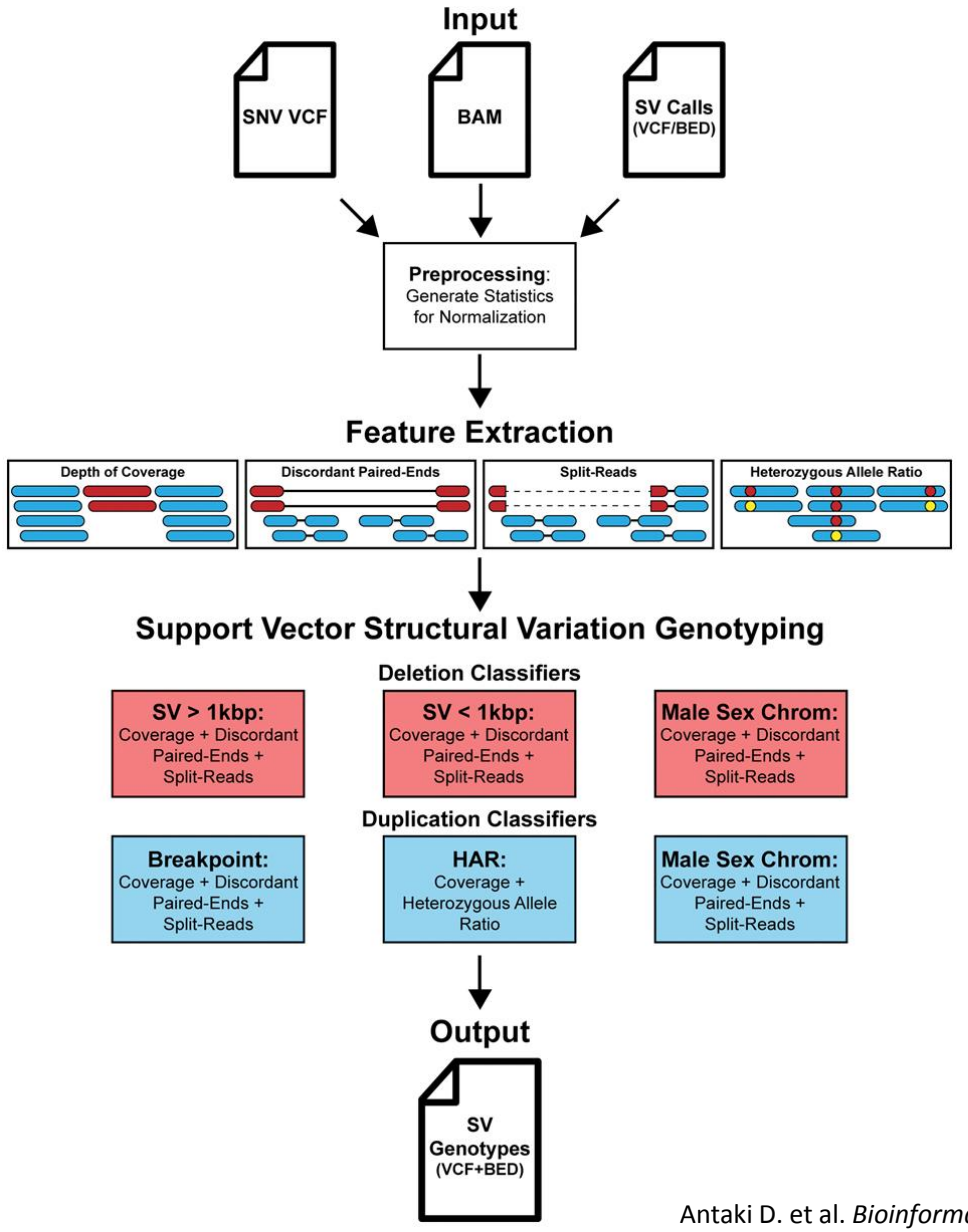


SVTyper and Manta utilize paired-end features and probabilistic models

SV^2 leverages paired-end, coverage, and heterozygous allele ratios and was trained on real data

SV² Overview

- Low FDR
  - 1.2% Deletions
  - 4.4% Duplications
- Low Genotyping Error rates
  - ~48% Transmission Rates
- Outperforms other methods
- Freely available on github
  - Python/Cython
  - `github.com/dantaki/SV2`



Antaki D. et al. *Bioinformatics* 2018

Aims

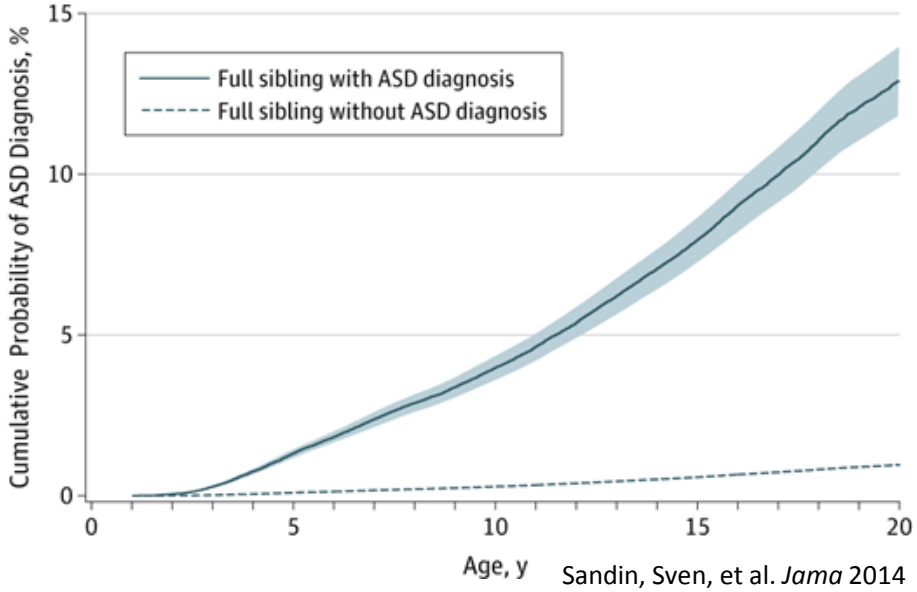
1. Refine Structural Variant (SV) calling in whole genome sequences (WGS)
   - SV^2: genotyping SVs with machine learning

2. Assay the burden of de novo SV in Autism Spectrum Disorders (ASD)
   - WGS offers optimal resolution compared to other methods

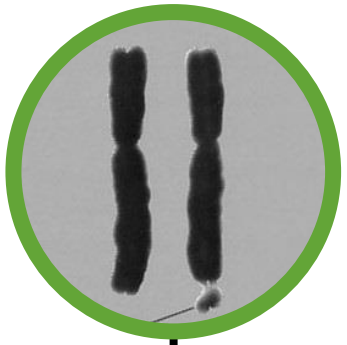
3. Interrogate the role of rare inherited variants in ASD
   - Identifying genetic risk stratified by parent of origin

- Autism is a heterogeneous neurodevelopmental disorder
  - Impaired social interaction and restricted behavior
  - 1% Prevalence, 4:1 Male:Female
- Sporadic or familial
  - Siblings have increased risk
- Genetic liability is significant
  - MZ twin concordance 30-99% (0.6-0.8 heritability)

Wikipedia Commons

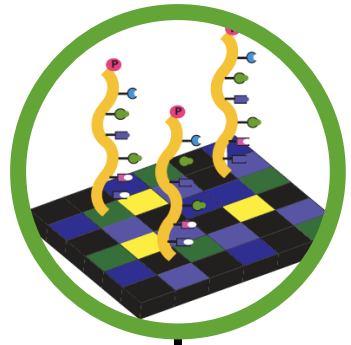


Sandin, Sven, et al. *Jama* 2014



First genetic microscopic (>10Mb)

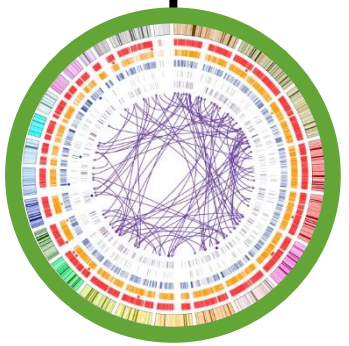
Exome studies found loss of function and missense SNPs+INDELs (1-50bp)



1960's | 1970's | 1980's | 1990's | 2000's | 2010's

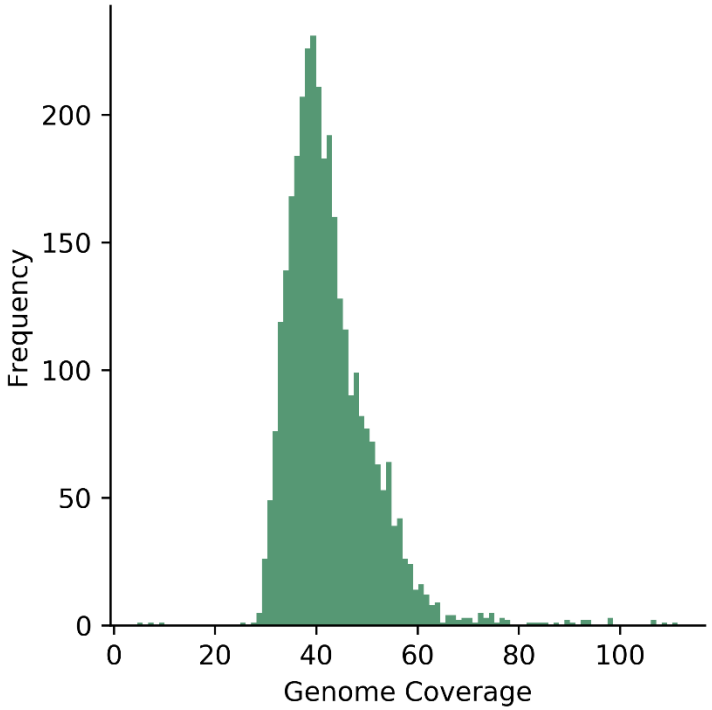
Microarrary microscopic (100kb-10M

Whole genome sequencing resolved complex de novo structural changes (50bp-10Mb)



WGS in 3,169 whole genomes

- 829 families
  - 880 ASD offspring
  - 630 Control siblings
- 2 Cohorts
  - REACH: local sources ASD patients (Rady Children's Hospital, 311 families)
  - Simons Simplex Consortium (518 families)
    - Screened negative for de novo LoF or large SV
- Illumina paired-end WGS (42x)
- SV calling: ForestSV, LUMPY, Manta
  - Genotype refinement: SV^2
- Filters
  - 66% overlap to segmental duplications, STRs, unmappable loci, assembly gaps
  - SV^2 strict filters for de novo SVs

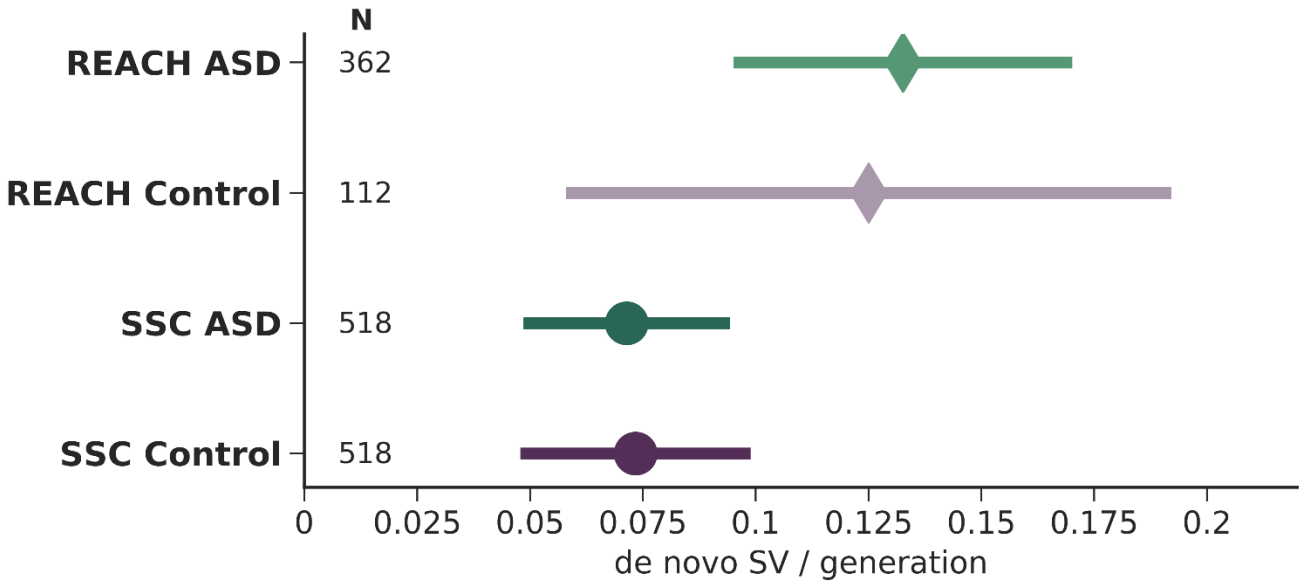


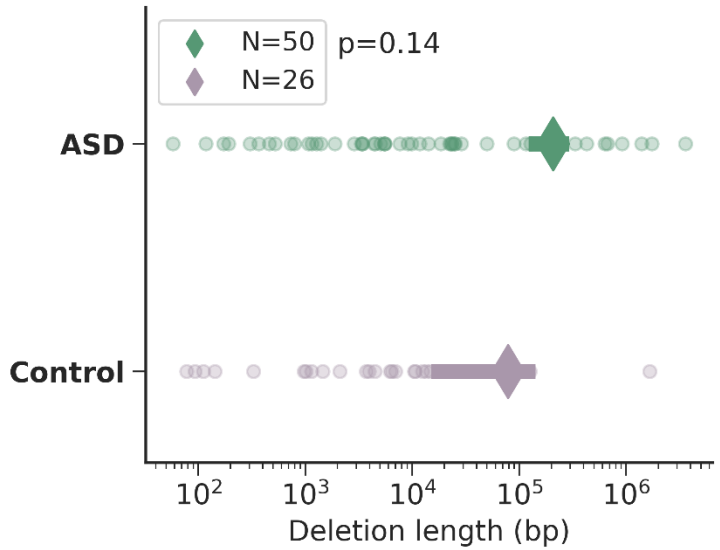
WGS better estimates the de novo SV rate

- Previous estimates of de novo SV were limited to microarrays
  - SV > 20kb
  - Copy Number Variants
- WGS better estimates the de novo rate in controls
  - SVs 100bp-10kb
  - Complex and Unbalanced SVs

Sample group	*n*	CNVs de novo	Ratio
Simplex autism	118	12	0.102
Multiplex autism	77	2	0.026
Simplex + multiplex	195	14	0.072
Controls	196	2	0.010

Sebat J., et al. *Science* 2007





De novo SVs are enriched for exons in ASD

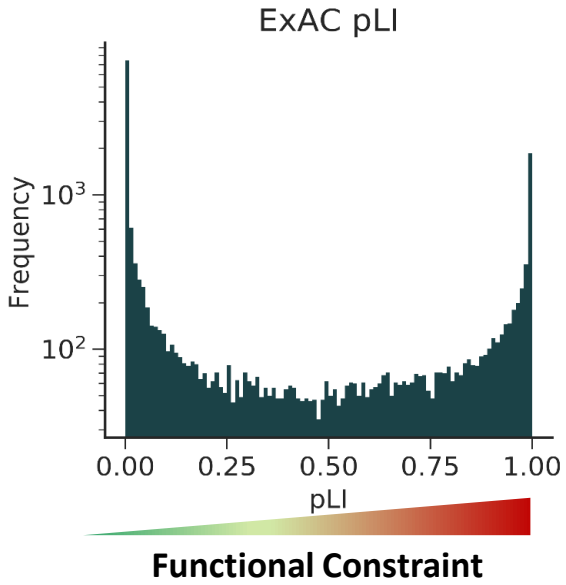
de novo Exon Burden

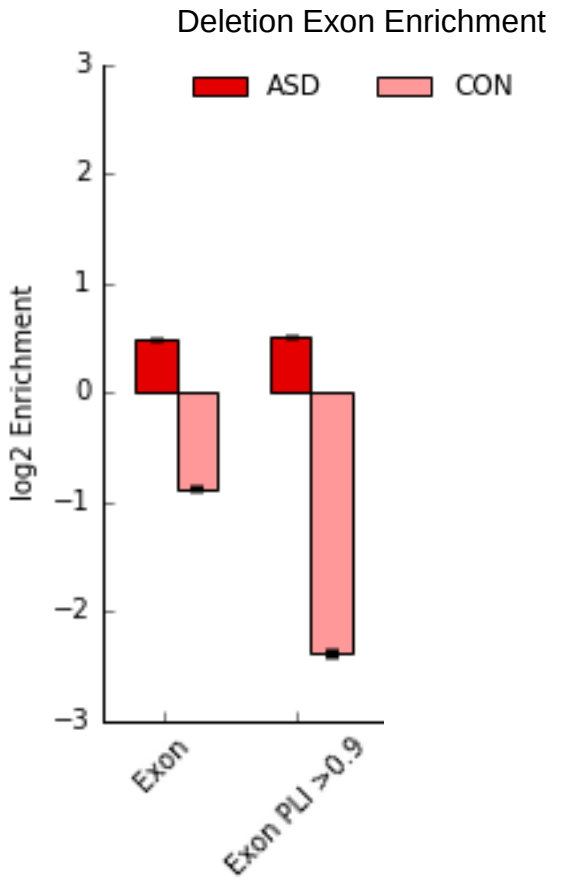
Functional Constraint

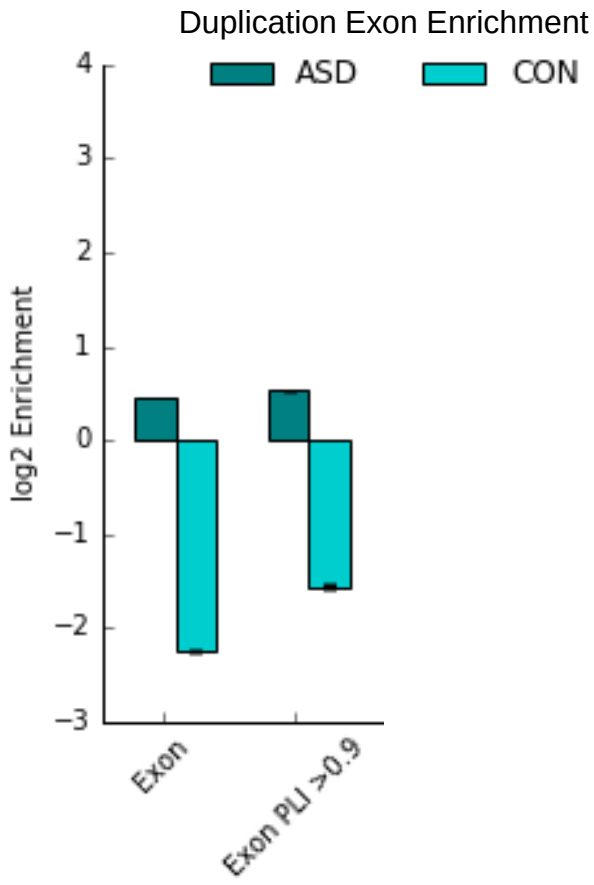
pLI	ASD Rate	Control Rate	Pval
0	4.2	0.2	0.004
0.9	1.72	0.12	0.005
0.99	1.03	0.09	0.01

Probability of being loss of function intolerant (pLI)

Measure of haploinsufficiency from observed and expected LoFs in Exomes



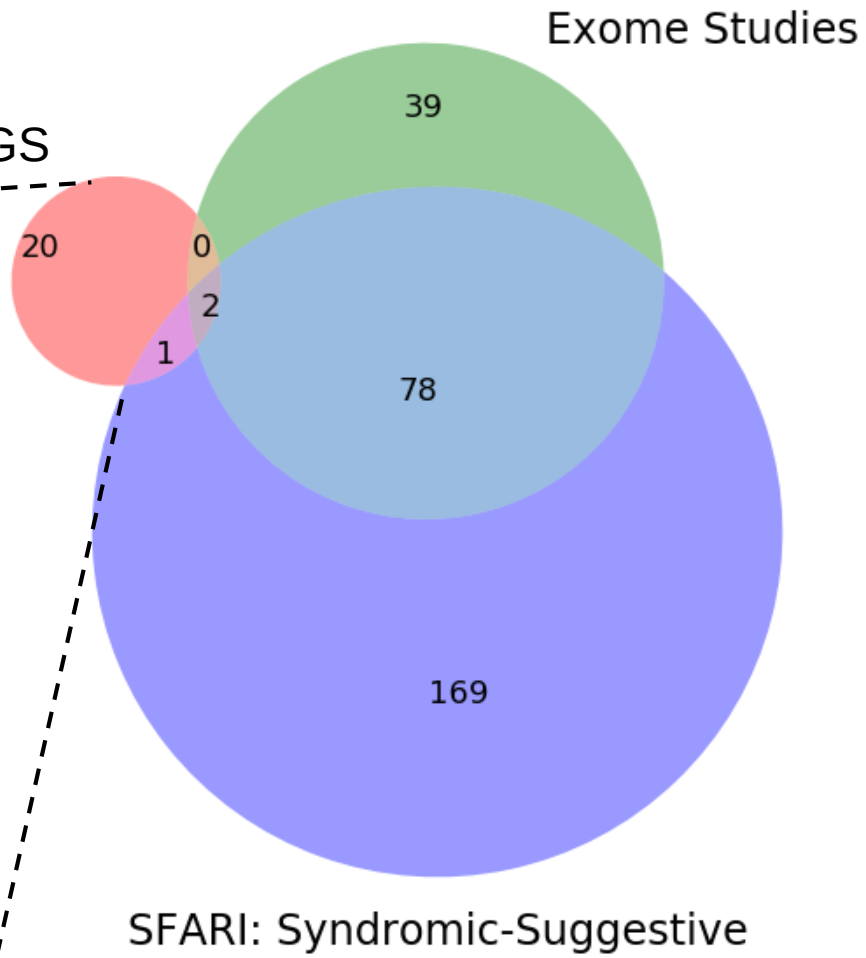




WGS resolves novel single gene hits

- 20 genes in cases were novel
  - Previously implicated:
    - MACROD2,DMD,CHD2

Gene	GO Term	Description
CACNG2	GO:0005245	voltage-gated calcium channel activity
CANX	GO:0035255	ionotropic glutamate receptor binding
COL5A2	GO:1903225	negative regulation of endodermal cell differentiation
CSMD1	GO:0042593	glucose homeostasis
DDX43	GO:0004004	ATP-dependent RNA helicase activity
F2R	GO:0051928	positive regulation of calcium ion transport
FAF1	GO:0043130	ubiquitin binding
FGD6	GO:0043547	positive regulation of GTPase activity
HERC4	GO:0042787	protein ubiquitination
LNPEP	GO:0060395	SMAD protein signal transduction
PIGL	GO:0006506	GPI anchor biosynthetic process
PPP1R9A	GO:0060999	positive regulation of dendritic spine development
SAE1	GO:0016925	protein sumoylation
SH3TC1	-	-
SLC4A5	GO:0008509	anion transmembrane transporter activity
SNX24	GO:0030659	cytoplasmic vesicle membrane
TAOK1	GO:0004702	receptor signaling protein serine/threonine kinase activity
TESC	GO:0004860	protein kinase inhibitor activity
UBE3B	GO:0004842	ubiquitin-protein transferase activity
ZNF462	GO:0003677	DNA binding

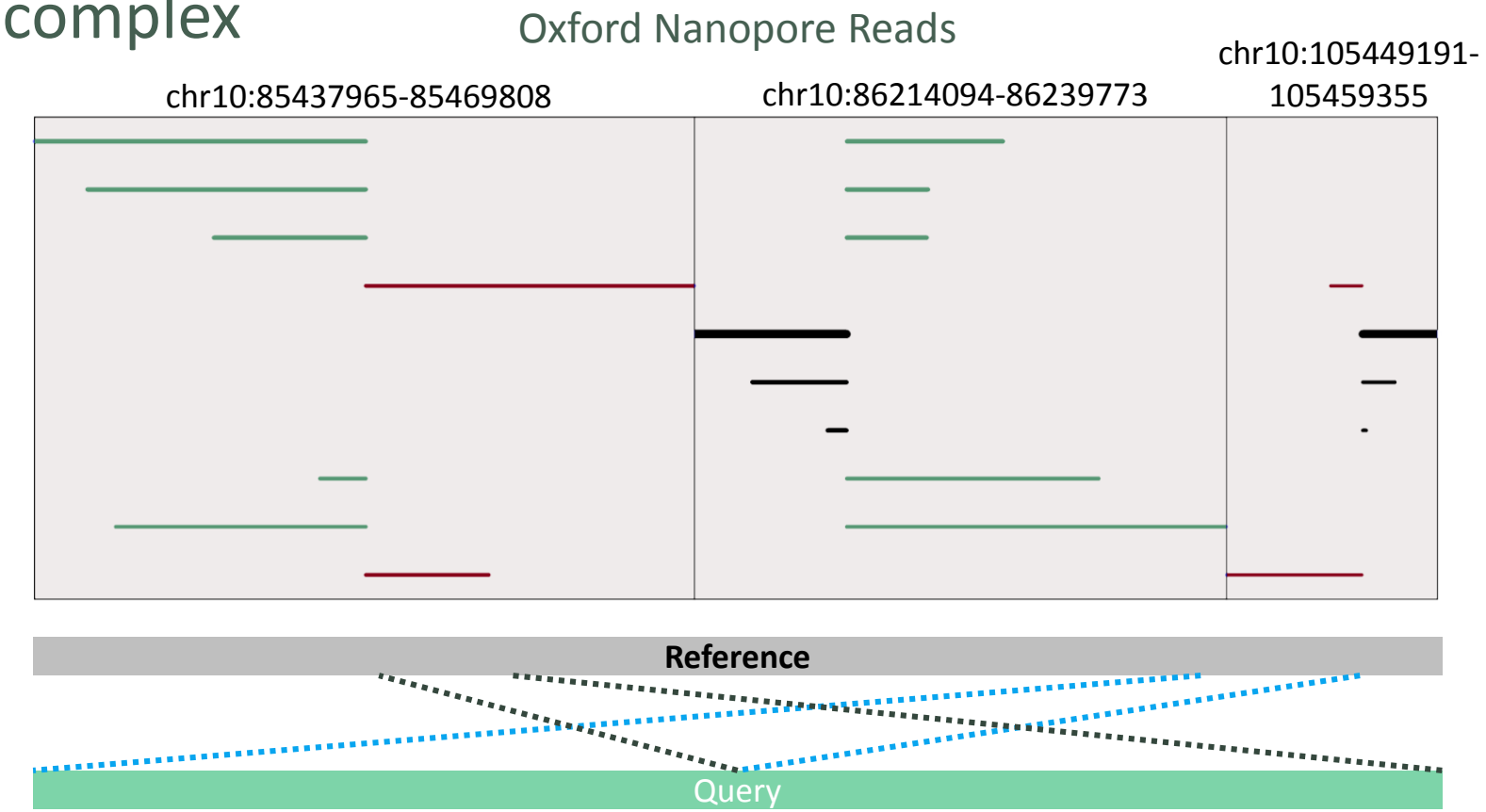


Identifying Complex and Unbalanced de novo SV

- WGS can identify Insertions, Inversions, and other complex rearrangements
  - 2 INV
  - 8 Complex
  - 32 INS

ABCD

A*CB*D



Aim2 Overview

- De novo SV rate is the same for ASD and control
- ASD de novo SV are enriched for haploinsufficient exons
- WGS can resolve SV better than previous methods

Brandler, W. M.*, Antaki, D.*, Gujral, M.*, et al. Frequency and complexity of de novo structural mutation in autism. *The American Journal of Human Genetics,* 2016

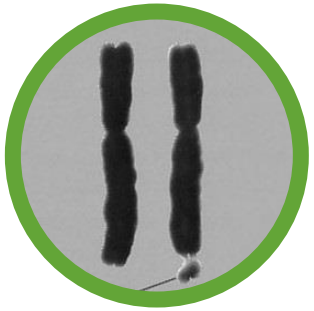
Aims

1. Refine Structural Variant (SV) calling in whole genome sequences (WGS)
   - SV^2: genotyping SVs with machine learning

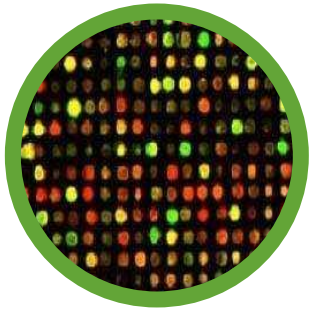
2. Assay the burden of de novo SV in Autism Spectrum Disorders (ASD)
   - WGS offers the best resolution

3. Interrogate the role of rare inherited variants in ASD
   - Parent of origin effects

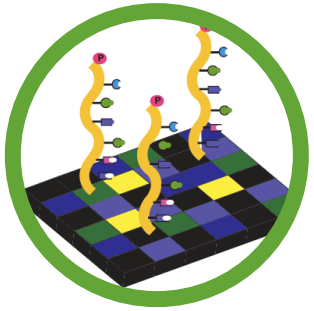
Explaining Missing Heritability in Autism



Method	Class of Variant	Limitations
Cyt	**15-25% of cases have a casual variant** *~60-100 genes associated with ASD*	>10Mb



Mic	**What have previous studies missed?** *Cis-Regulatory Elements*	.0kb – 10Mb Not enough power for ommon association



Targeted Sequencing (Exomes)	Copy Number Variants SNPs + INDELs	Limited to target. Imprecise breakpoints

Accounting for Missing Heritability

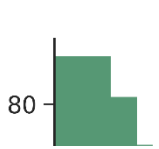
- Exome sequencing neglects cis-regulatory elements
  - Promoters
  - UTRs
  - Trai

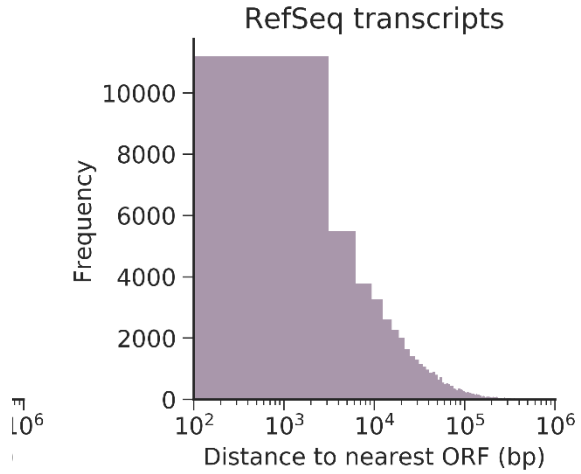
Are Cis-Regulatory Structural Variants (CRE-SV) associated with Autism?

- Microarrays cannot resolve smaller structural variants (50bp-10kb)
  - 37% of non-o within 10kb c

No de novo CRE-SV

moderate effect



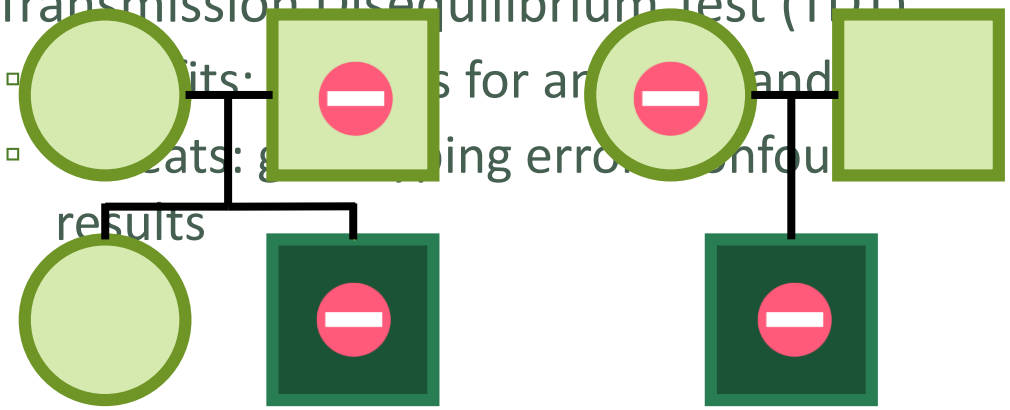


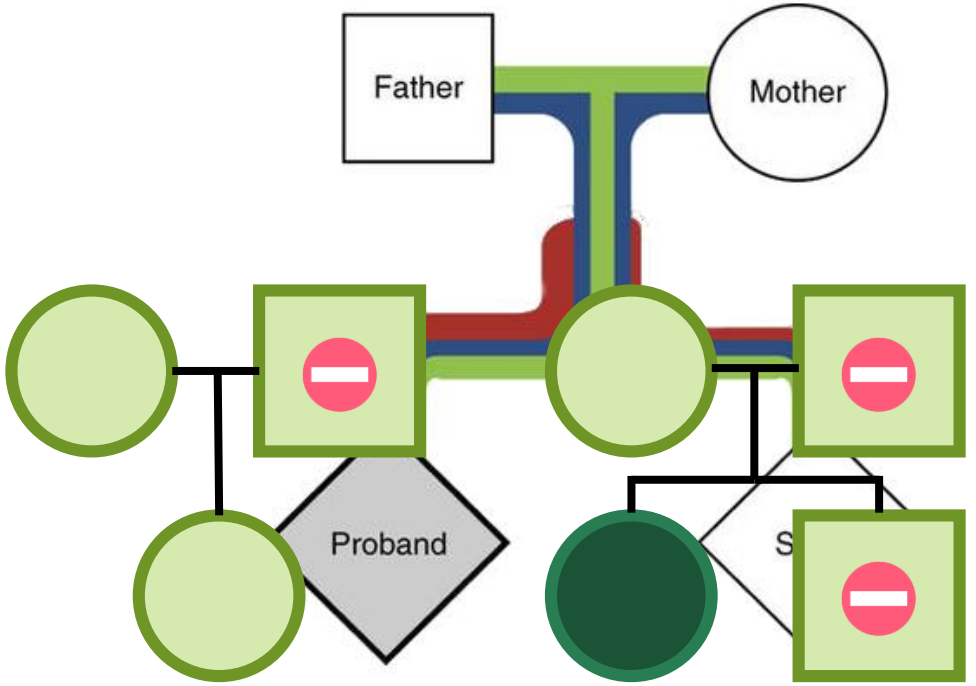
pLI: probability that a given gene is intolerant to loss of function mutations (ExAC).

Ascertaining Rare Inherited CRE-SV risk

- Group-wise Additive model
  - Bin variants by functional annotation
  - Test the transmission rate for binned variants for all offspring in the cohort
- Transmission Disequilibrium Test (TDT)
  - Benefits: [illegible]s for ar[illegible]and
  - Caveats: g[illegible]ping erro[illegible]nfou[illegible] results

Are CRE-SVs transmitted to cases more often than controls?





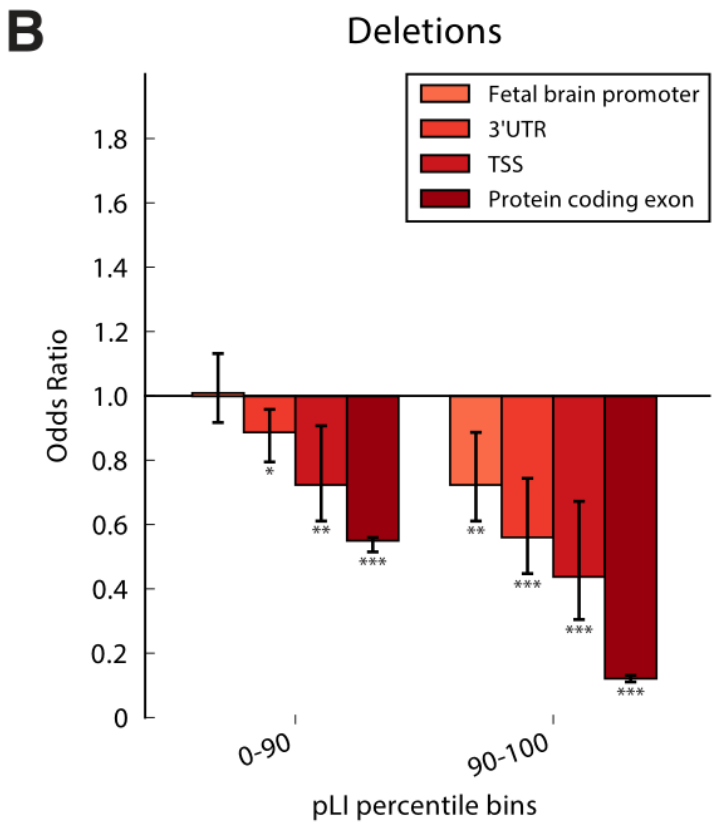
The affected child (proband) inherits more damaging rare mutations (red), than the control sibling

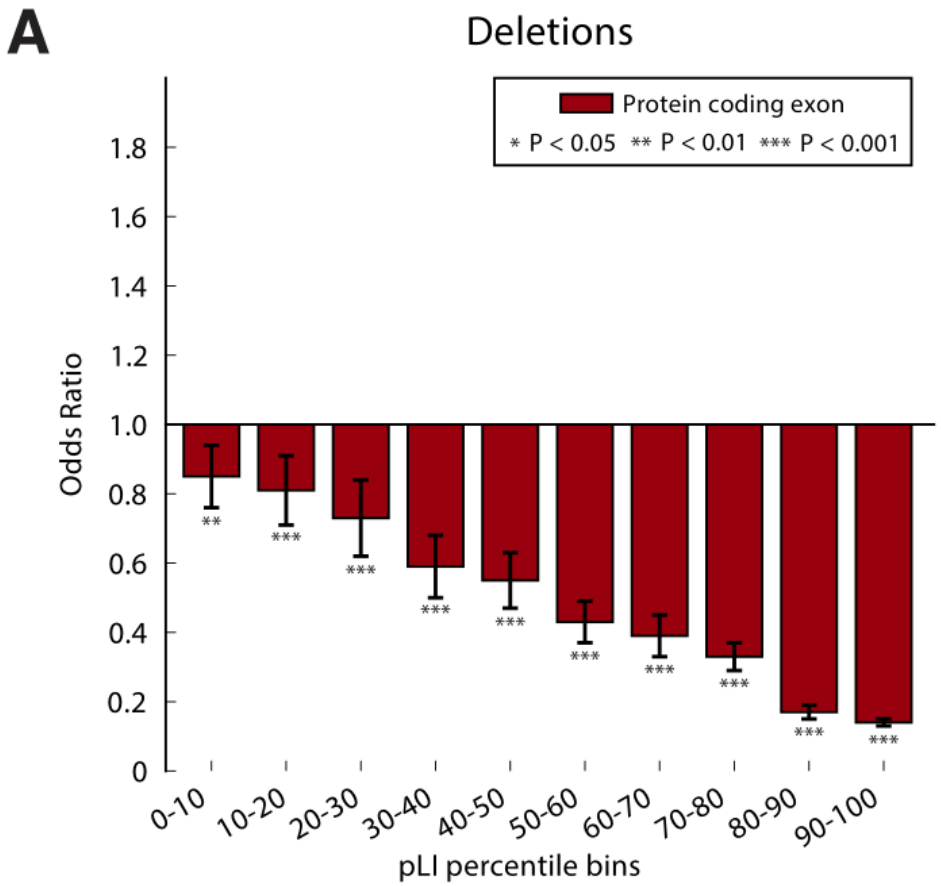
Phenotype	Transmitted	Not Transmitted	Transmission Rate
Autism	2	1	66%
Control	1	2	33%

[illegible]mm et al. Nat Gen 2015

Intolerant Functional Elements are Rarely Deleted

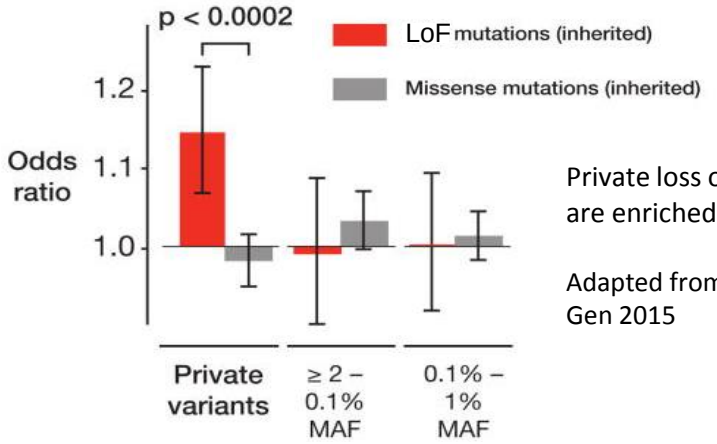
- Genes with strong negative selection (pLI >0.9)
- Limit to deletions, easier to interpret





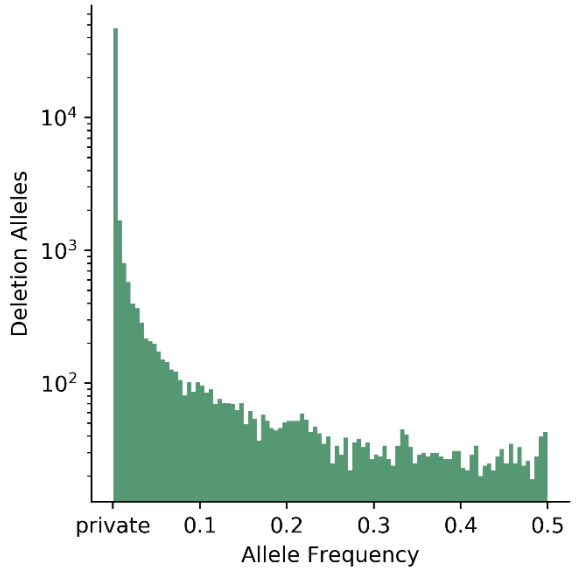
Private Variants are Enriched for Risk

- Private variants (singletons) are young mutations
- Not enough time for natural selection
- Most variants are rare
  - More transmission events to test



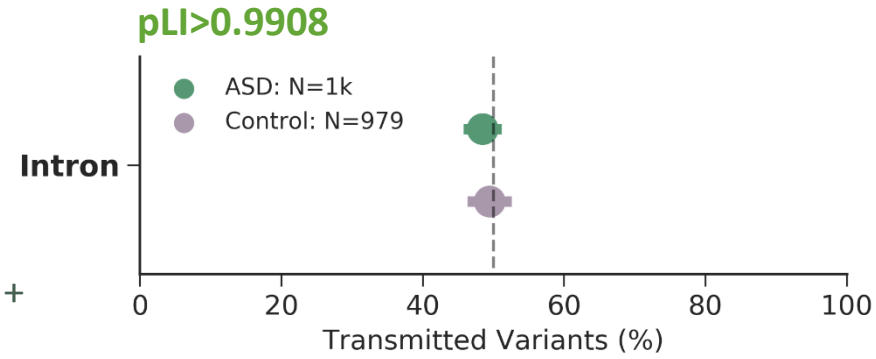
Private loss of function (LoF) SNPs are enriched Autistic children.

Adapted from Krumm et al. Nat Gen 2015

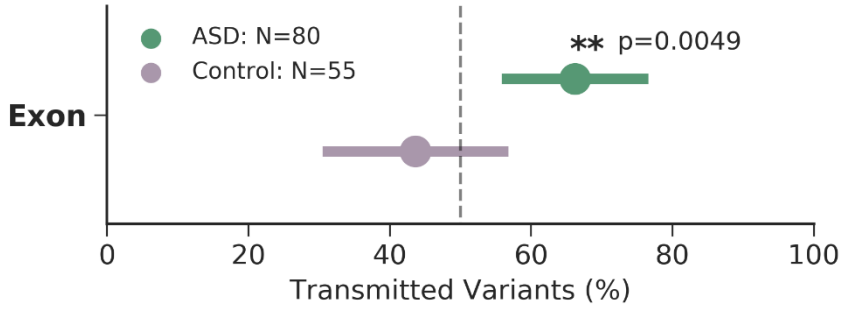


Inheritance of private variants within functionally constrained genes

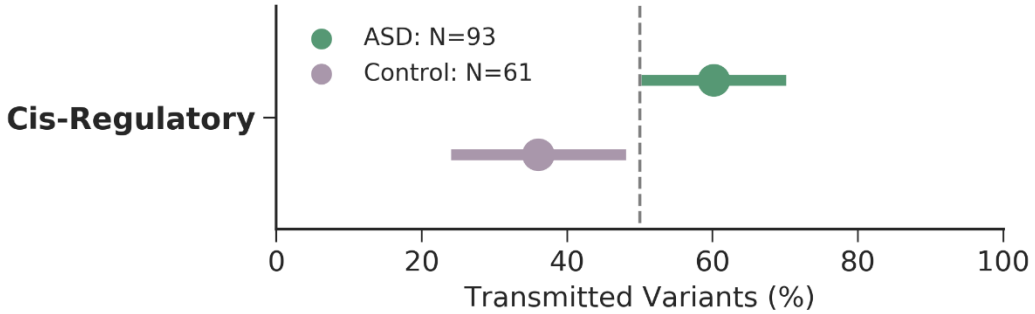
- SVs within introns did not
- exhibit transmission bias
  - Error bars are 95% CI (binomial proportion)
  - N = number of independent assortment events (transmitted + non-transmitted)



- SVs overlapping exons were over-transmitted to cases
  - Concordant with exome studies

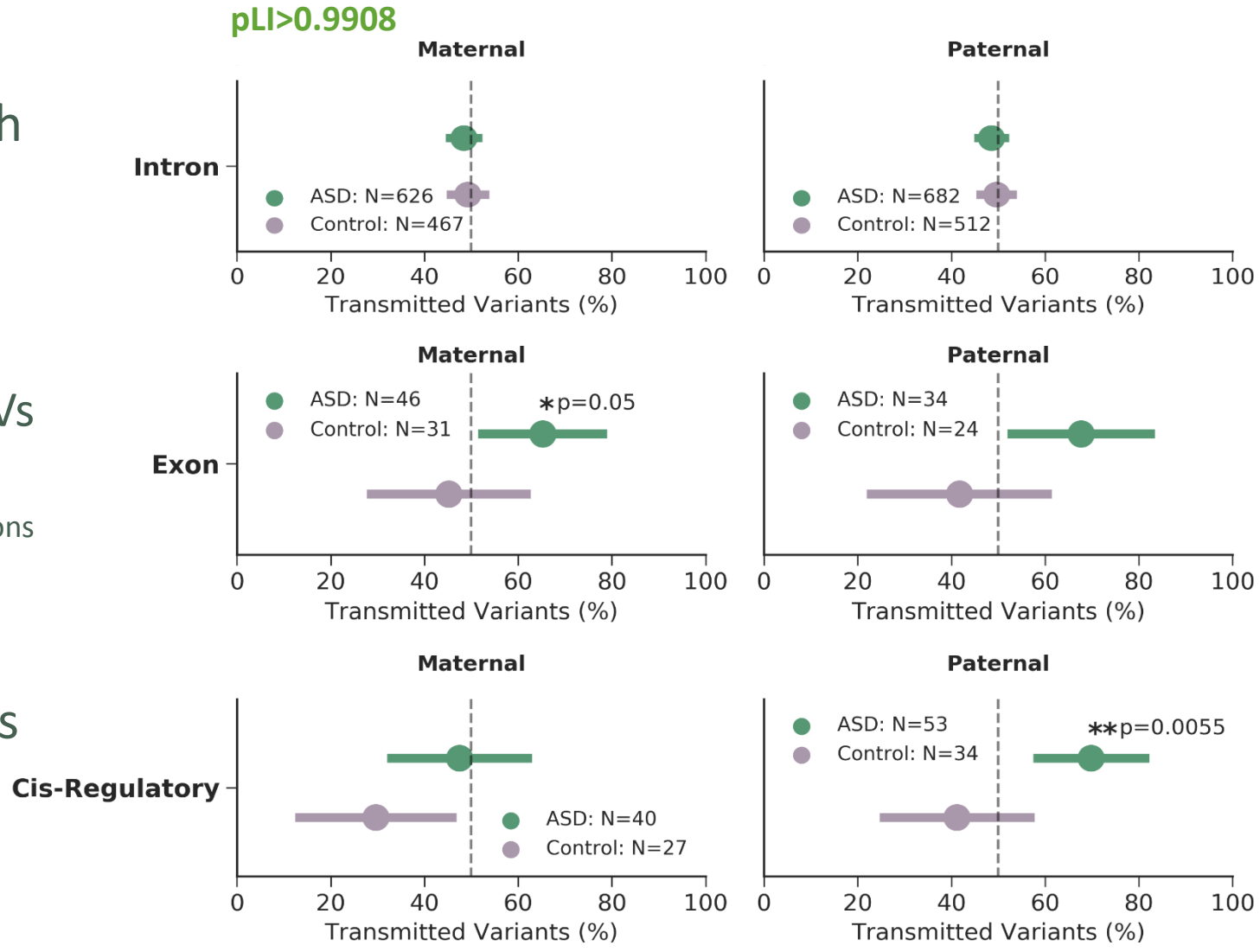


- CRE-SVs were slightly over-transmitted to cases



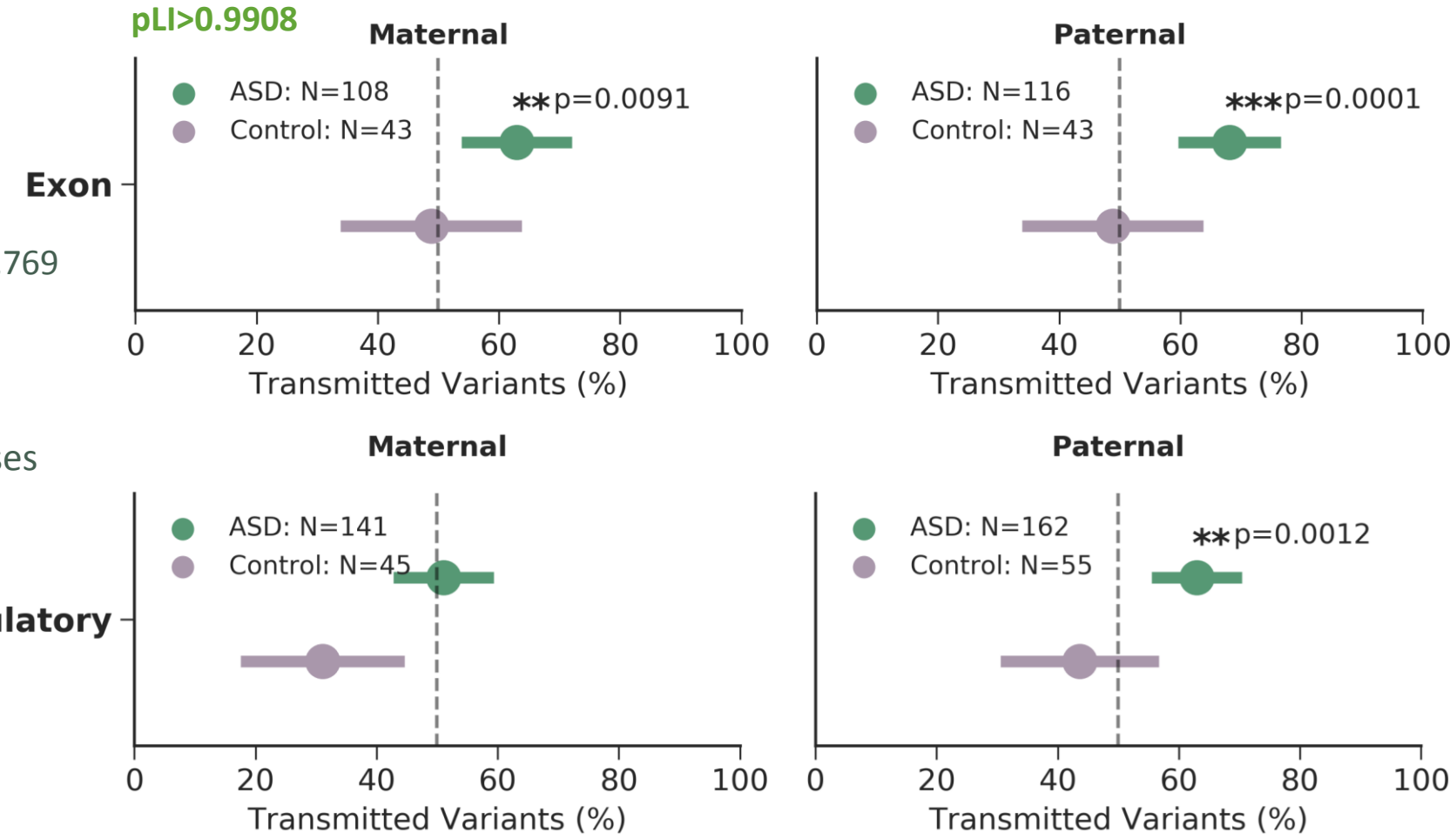
CRE-SVs exhibit a parent of origin effect

- No parent of origin effect with private SVs within introns
- Mothers over-transmit exonic SVs to cases
  - Females have higher tolerance for LoF mutations
- Fathers over-transmit CRE-SVs to cases
  - Under-transmission to controls



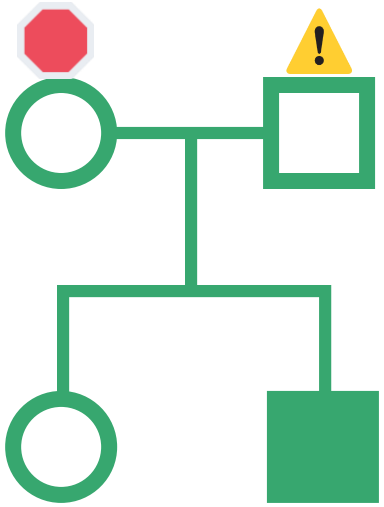
Replicating the initial findings

- 9,274 whole genomes
  - 2,600 families
    - 2,859 cases
    - 1,214 controls
  - MSSNG cohort (Stephen Scherer): 3,769
  - SSC phase 2: 2,336
- Exonic SVs are over-transmitted to cases
- Rare CRE-SVs exhibit a paternal origin effect



Overview and future directions

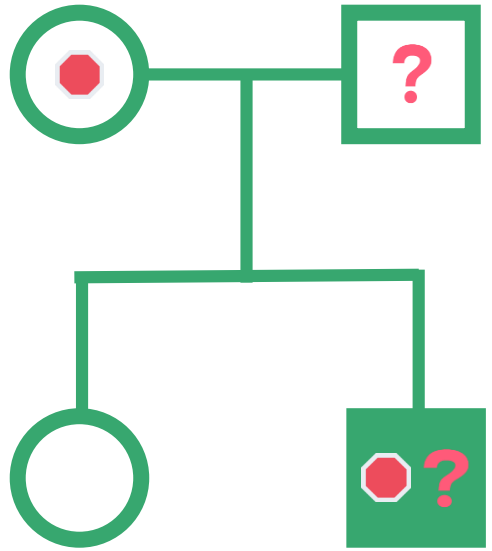
- CRE-SVs exhibit a significant paternal origin effect
  - ~0.8% of cases have a paternally inherited CRE-SV
    - Rare inherited variants thought to contribute ~3-4% to ASD
- Autism is more complex than previously thought
  - Fathers also contribute inherited risk
- Bilineal two-hit model
  - Inherited maternal LoF of large effect
  - Inherited paternal CRE-SV of moderate effect



Brandler, W.M.*, Antaki, D.*, Gujral, M.*, et al. Paternally inherited cis-regulatory structural variants are associated with autism. *Science* 2018.

Exploring the Inherited Bilineal Model

- REACH + SSC1 + SSC2 + SCC3
  - SSC phase 4 to 6 processing
- Condition on maternal LoF
  - What is the remaining risk from father?
    - To sons or daughters?
  - Test inheritance of rare SNPs, INDELs, and SVs
    - Functional constraint
      - pLI (Coding)
      - CDTS + Ubiquitous Enhancers (Noncoding)

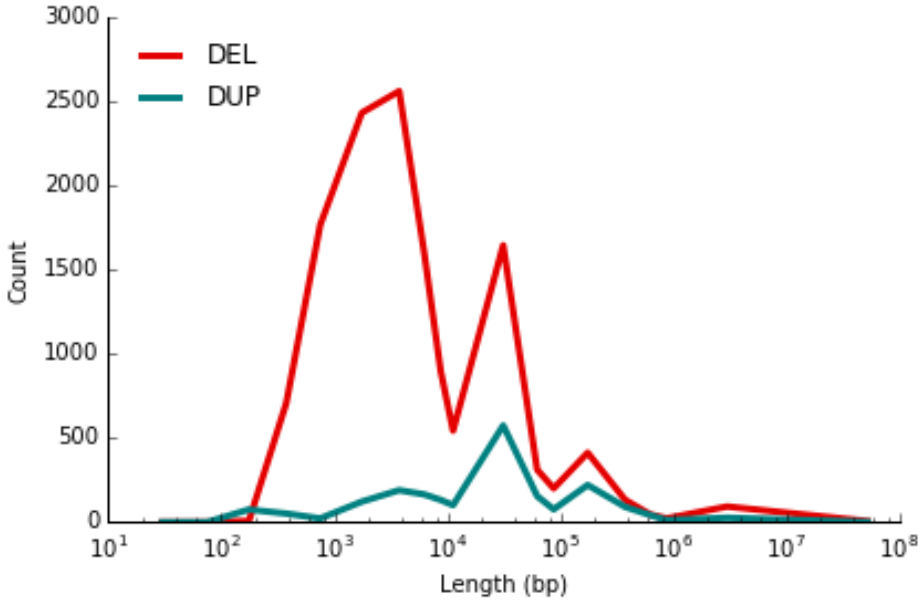


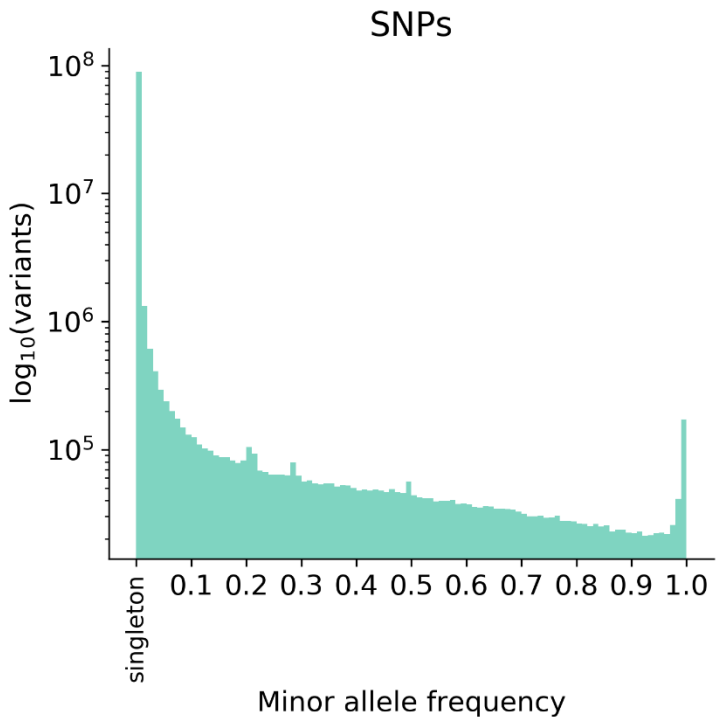
Methodology

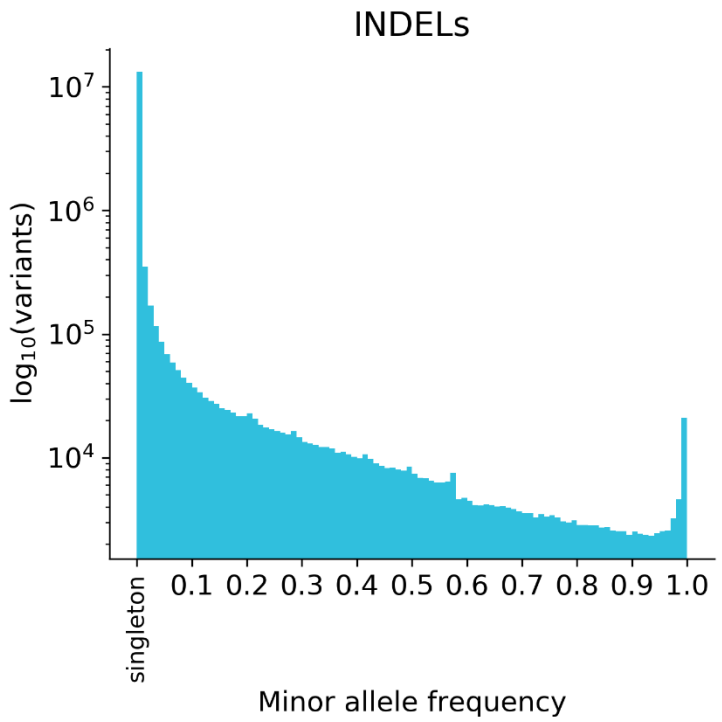
- WGS in 7,716 individuals
  - 2,005 families
    - 2,059 ASD
    - 1,808 control
- Call SNPs , INDELs, SVs
  - Flag damaging de novo variants
  - Flag damaging inherited variants
- Perform TDT on conditioned trios to ascertain inherited risk
  - Group variants according to functional constraint

Variant QC

- 3.6M SNPs per individual
- 773k INDELs per individual
- 3.7k SV per individual



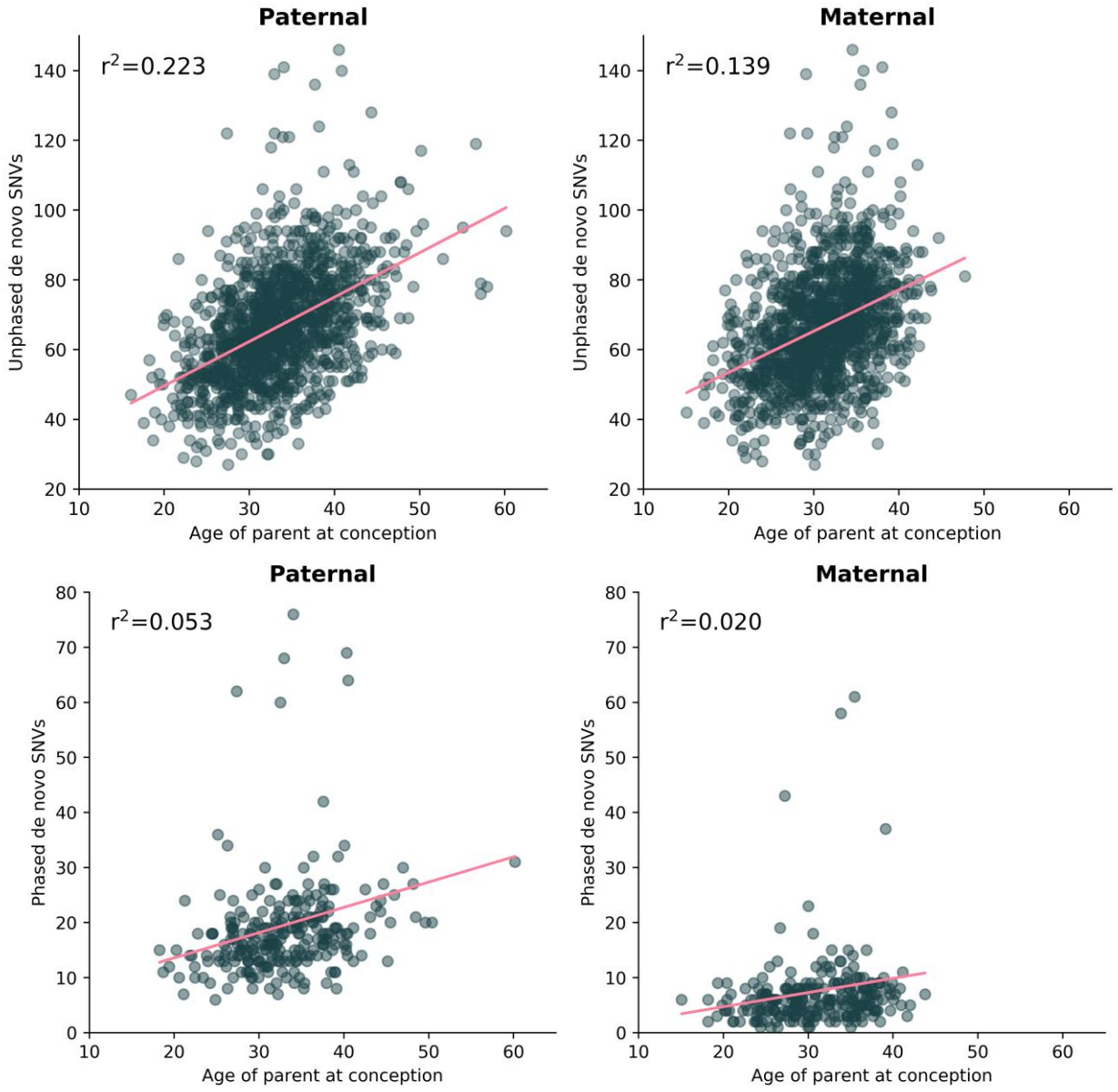




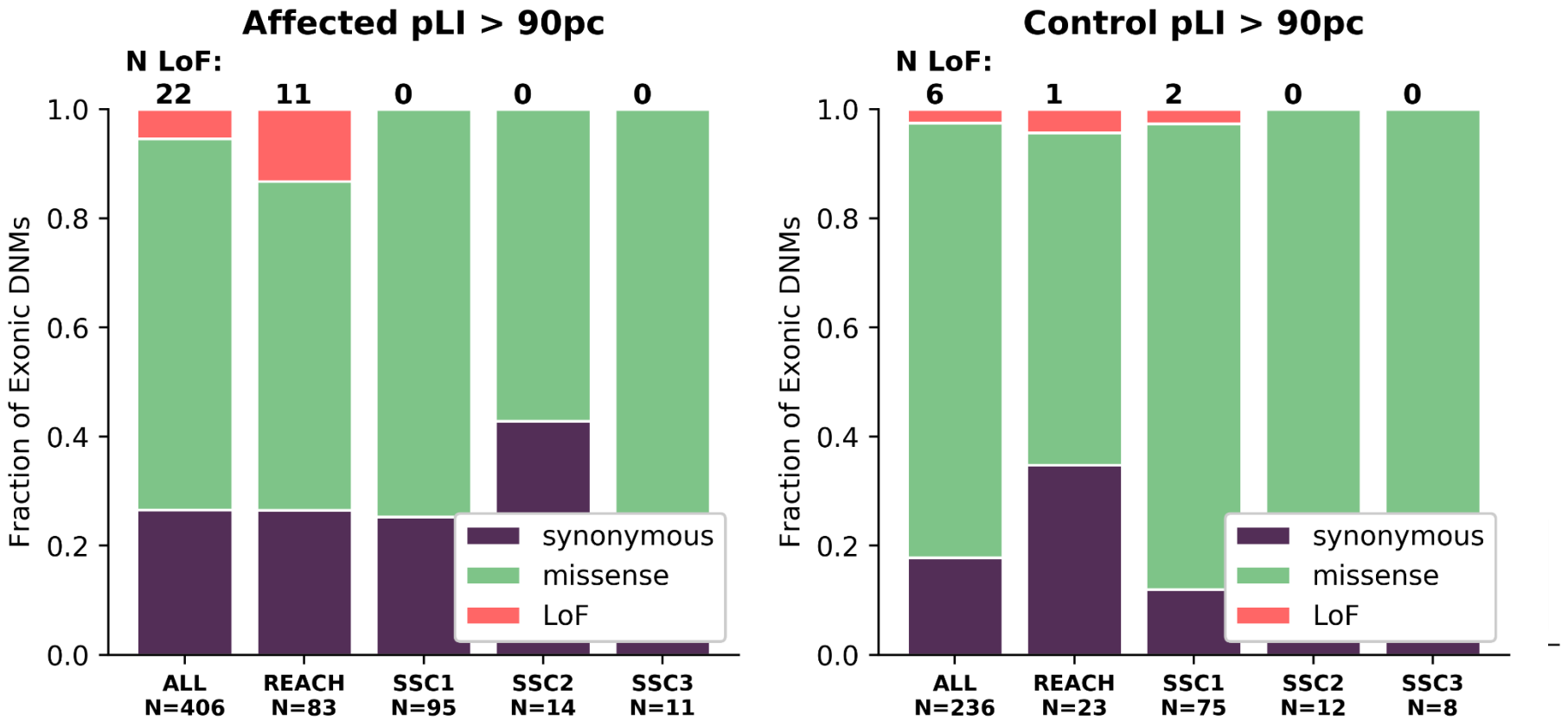
De novo SNP

- ForestDNM

- Filters
  - Common variants (>1% AF) 1000 genomes phase 3
  - Segmental Duplications
  - 0% AF in parents



SSC samples are depleted for LoF de novo SNPs

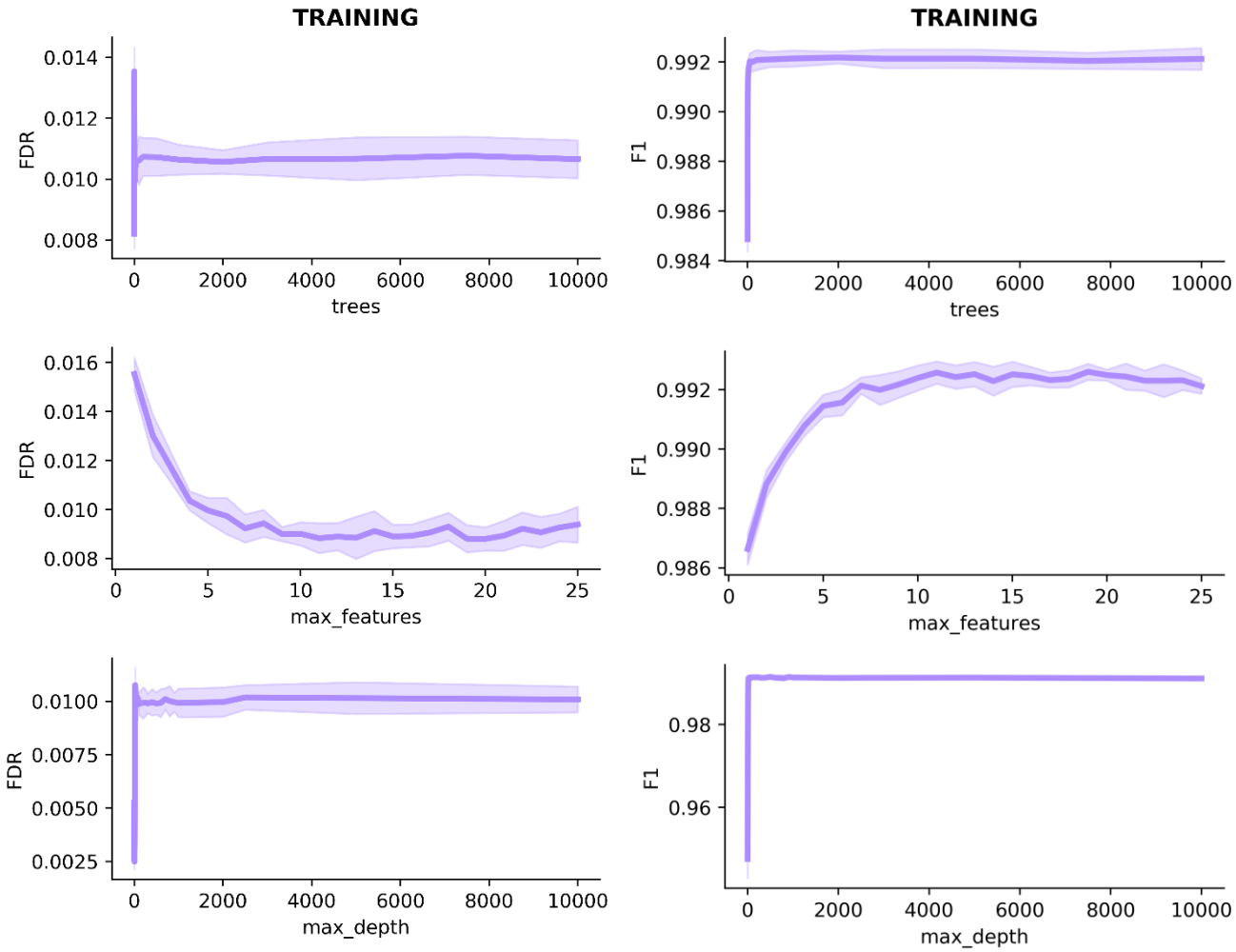


De novo INDELs (Work in Progress)

- Developing a machine learning classifier for de novo INDELs
  - Similar to ForestDNM
- Training set: 11 pairs of MZ twins
  - True Positives: Inherited private variants genotyped in swapped parents
  - False Positives: Discordant de novos
- Test set: 98 offspring
  - True Positives: Inherited private variants in swapped parents
  - False Positives: Apparent de novos
- Validation set: 15 offspring
  - True Positives: Validated de novo INDELs in Single Molecule Reads
  - False Positives: WGS de novos not validated with long reads

Initial Performance Results

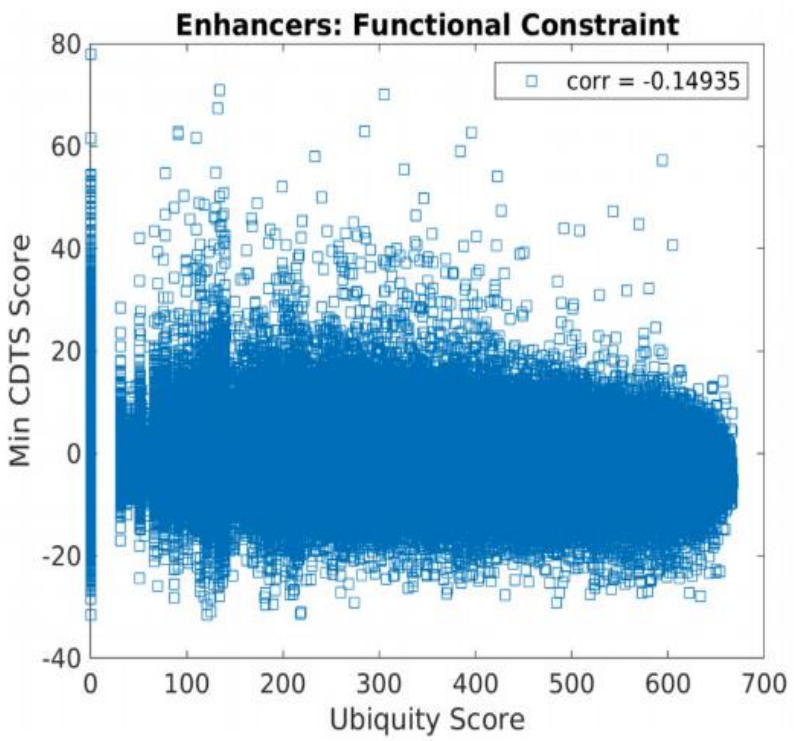
- 10x cross validation
- Parameter Sweeps
  - FDR below 2%
- Grid Search
  - selecting the best parameters
- Test set performance
  - Aojie Lian

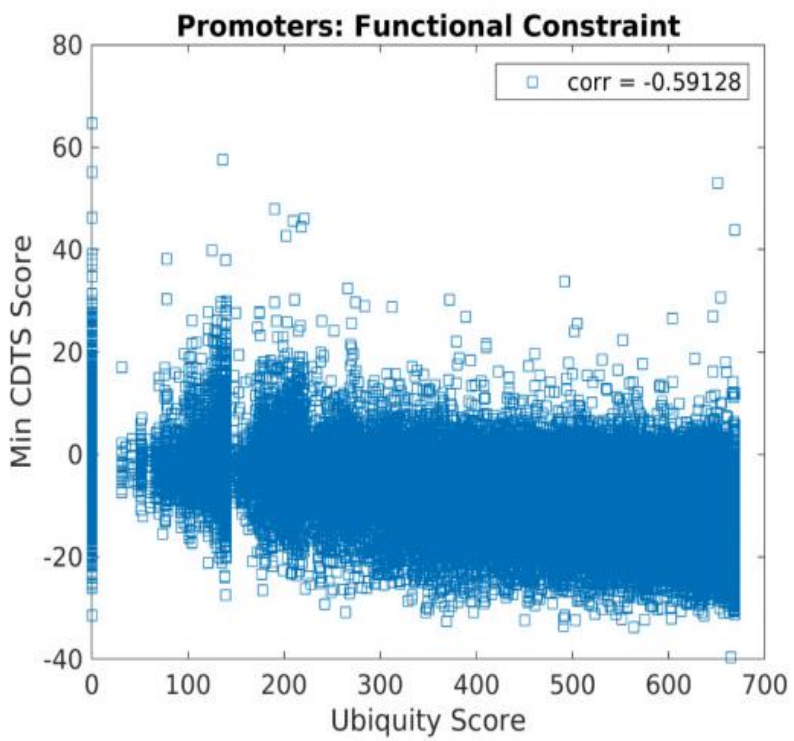


CDTS (Di Iulio J. 2018)

Prateek Tandon

- Functional constraint score of the genome
  - 10bp intervals
  - WGS SNP mutation
- Private deletions are depleted in constrained regions
- CDTS can help bin constrained cis-regulatory elements





Aim 3 Checklist

	SNP	INDEL	SV
Variant Calling	✓	✓	✓
De novo Annotation	✓	Evaluating classifier	SSC2 + SSC3 need annotations
Private Variant Annotation	✓	✓	SSC2 + SSC3
TDT	✓	✓	Software framework only needs some adaptions
Conditioned TDT	Need to bin offspring by risk		

Acknowledgements

Sebat Lab

- **William Brandler**
- **Madhu Gujral**
- Prateek Tandon
- Oanh Hong
- Morgan Kleiber
- Tim Chapman
- Michelle Maile
- Omar Shanta
- Adam Koch
- Aojie Lian

Stephen Scherer (MSSNG)
Rady Children's Hospital
REACH project
Human Longevity Institute
Simons Simplex Project
1000 Genomes Project

San Diego Supercomputer Center

Thesis Committee

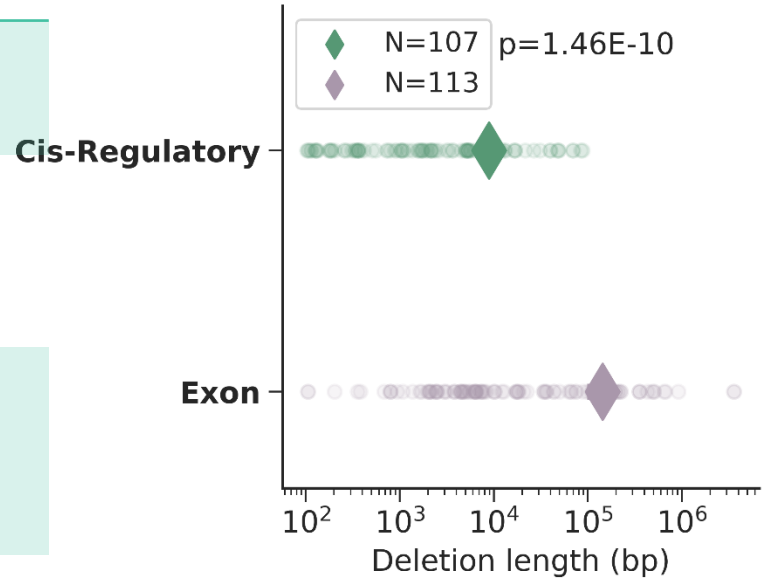
- Jonathan Sebat: chair
- Joseph Gleeson
- Alysson Muotri
- Abraham Palmer
- Nicholas Schork

Funding:
NIH predoctoral training grant T32 GM008666 (UCSD Genetics Training Program)
Amazon Web Services Research Grant

Sebat Lab funding:
NIH: MH076431, MH113715
Beyster Family Foundation

Recurrent CRE-SVs

Gene	Cytoband	Annotation	Syndrome / ASD finding
CNTN4	3p26.3-p26.2	Axon-assoc. cell adhesion molecule	3p deletion/ASD CNV
RAF1	3p25.2	Proto-oncogene, serine/threonine Kinase	Noonan/ASD CNV
MEST	7q32.2	hydrolase activity, paternally expressed	Silver-Russell /7q ASD susceptibility locus
LEO1	15q21.2	Comp. of PAF1 complex, assoc. w/ RNA pol II	./2 de novo LoF in ASD exome studies



CRE-SVs are significantly smaller than exonic SVs