CCDG NEUROPSYCHIATRIC WORKING GROUP

NEW YORK GENOME CENTER

Bustamante (Stanford)
Buyske (Rutgers)
Daly (Broad)
Darnell (NYGC)
Eichler (U W/NYGC)
Goldstein (CUMC/Broad)
Hall (Wash U)
Iossifov (CSHL/NYGC)
Locke (Wash U)
Matise (Rutgers)
Neale (Broad)
Pickrell (NYGC)
Sabo (Baylor)
Turner (U W/NYGC)
Moore Vogel (NYGC)
Wigler (CSHL/NYGC)
Zody (NYGC)

APRIL 13, 2016

NEUROPSYCHIATRIC GENETICS

Center For Common Disease Genetics

- Comprehensive Approach to Disease
  - Genes
  - Variants/Gene

MISSING HERITABILITY

WGS & GENOMIC FRONTIERS

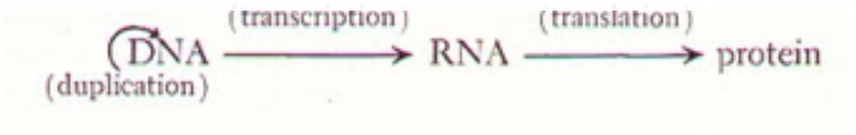
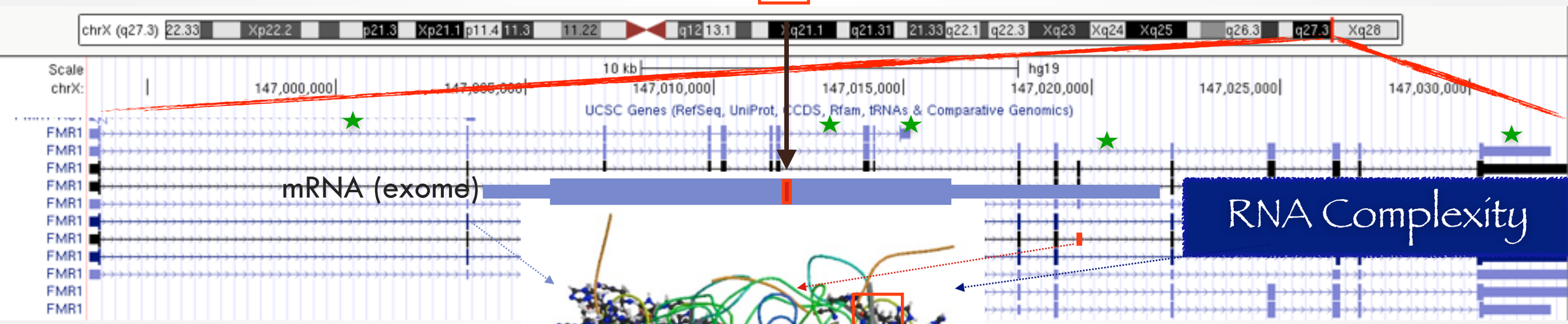


Fig. 2. Watson's version of the Central Dogma. This figure is taken from the first edition of The Molecular Biology of the Gene (p. 298).

```
> range=chrX:146993469-147032647 FMR1
ACTTCCGGTGGAGGGCCGCCTCTGAGCGGGCGGCGGGCCGACGGCGAGCGCGGGCGGCGGCGGTGACGGAGGCGCCGCTGCCAGGGGGCGTGCGGCAGCG
CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGGGCCTCGAGCGCCCGCAGCCCACCTCTCGGGGGCGGG
CTCCCGGCGCTAGCAGGGCTGAAGAGAAGATGGAGGAGCTGGTGGTGGAAGTGCGGGGCTCCAATGGCGCTTTCTACAAGGTACTTGGCTCTAGGGCAGG
```



Lappalainen T, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013; 501: 506–511.

NEW YORK GENOME CENTER

Examples of non-coding mutations in disease (partial list)

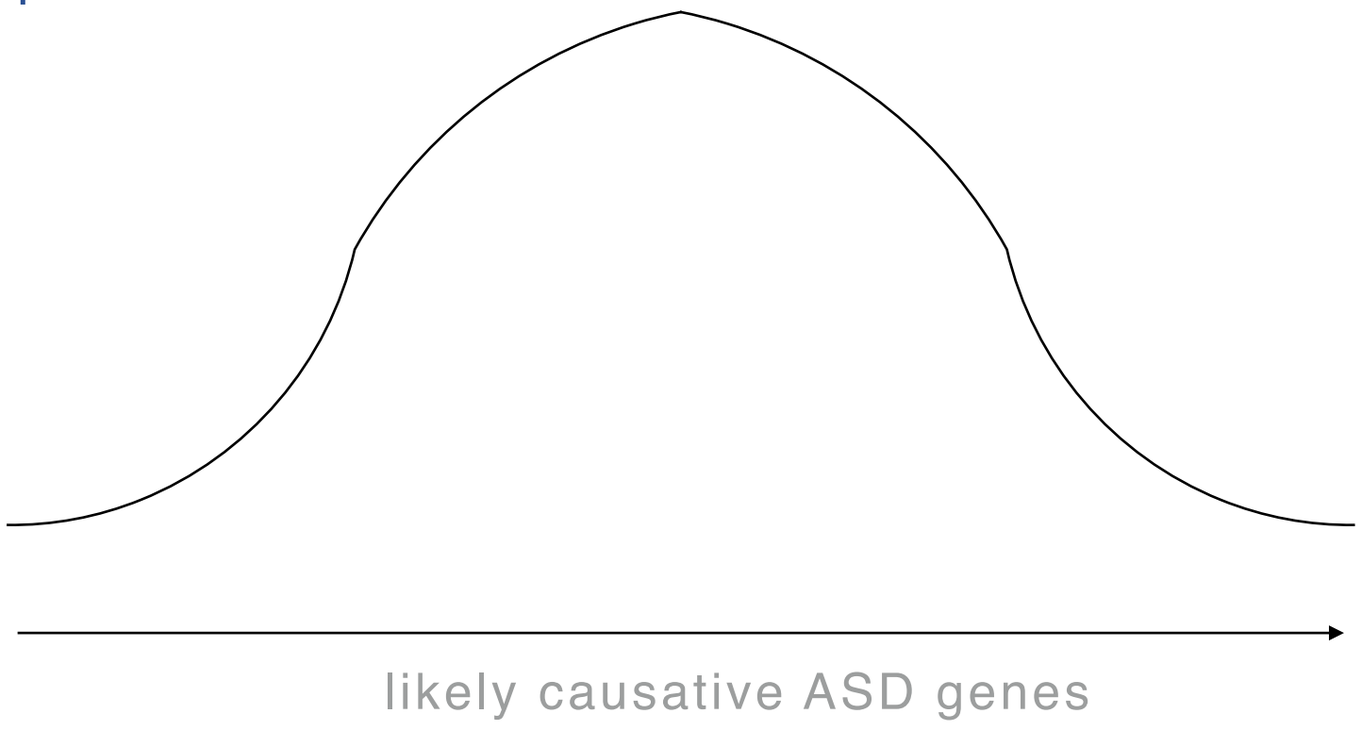
Disease or Phenotype	Mutation	Gene Affected	Result of Mutation	Reference
Aniridia	Translocation breakpoint affecting tissue specific enhancers	*PAX6*	Misregulation of gene	Kleinjan et al. 2001
Bernard-Soulier Syndrome	Point mutation in promoter	*GPIBB*	Reduced expression of gene	Ludlow et al. 1996
Cleft Palate	Microdeletions, translocation breakpoints, and point mutations in enhancer	*SOX9*	Altered binding of transcription factor to enhancer	Benko et al. 2009
Deafness	Deletions and rearrangements of otic vesicle enhancer	*POU3F4*	Misregulation of gene	de Kok et al. 1995
Deafness	Point mutations in seed sequence of miRNA	*miR-96*	Haploinsufficiency of miRNA	Mencia et al. 2009
Feingold Syndrome	Deletions of polycistronic miRNA cluster	*miR-17~92*	Haploinsufficiency of miRNA	de Pontual et al. 2011
Fragile X Syndrome	Expanded CGG repeat in 5' UTR (FMR1 full mutation)	*FMR1*	Silencing of gene	Verkerk et al. 1991
Hirschsprung Disease	Point mutations in enhancer	*RET*	Altered expression levels of gene	Grice et al. 2005
Muscularity in sheep	Point mutations in 3' UTR	*GDF8*	Point mutation creates a new miRNA target site	Clop et al. 2006
Preaxial Polydactyly	Point mutations in enhancer	*SHH*	Abnormal location for expression of gene	Lettice et al. 2003
α-Thalassaemia	Point mutation in polyadenylation signal	*HBA2*	Reduced expression of gene	Higgs et al. 1983
Fragile X-associated tremor/ ataxia Syndrome (FXTAS)	Expanded CGG repeat in 5' UTR (FMR1 premutation)	*FMR1*	Overproduction of abnormal FMR1 mRNA	Tassone et al. 2004
Frontotemporal Dementia and Amyotrophic Lateral Sclerosis	G4C2 repeat expansion in first intron	*C9orf72*	Creating of a novel protein called poly(GA)	Dejesus-Hernandez et al. 2011, Xi et al. 2013
Leukemia	indels upstream of gene create novel transcription factor binding sites	*TAL1*	Creating of super-enhancers	Mansour et al. 2014
Breast cancer	miRNA	*miR-335, miR96, miR-126*	Metastases	Tavazoie & Massague, 2008

NEUROPSYCHIATRIC GENETICS

Center For Common Disease Genetics

- Autism as Exemplar project
  - Large-scale family based sequencing as approach to comprehensiveness.
    - Architecture 1: large case-control studies; WES, WGS
    - Architecture 2: de novo presentation in context of quads/trios
    - Architecture 3: multiply affected individuals within well-phenotyped family architecture

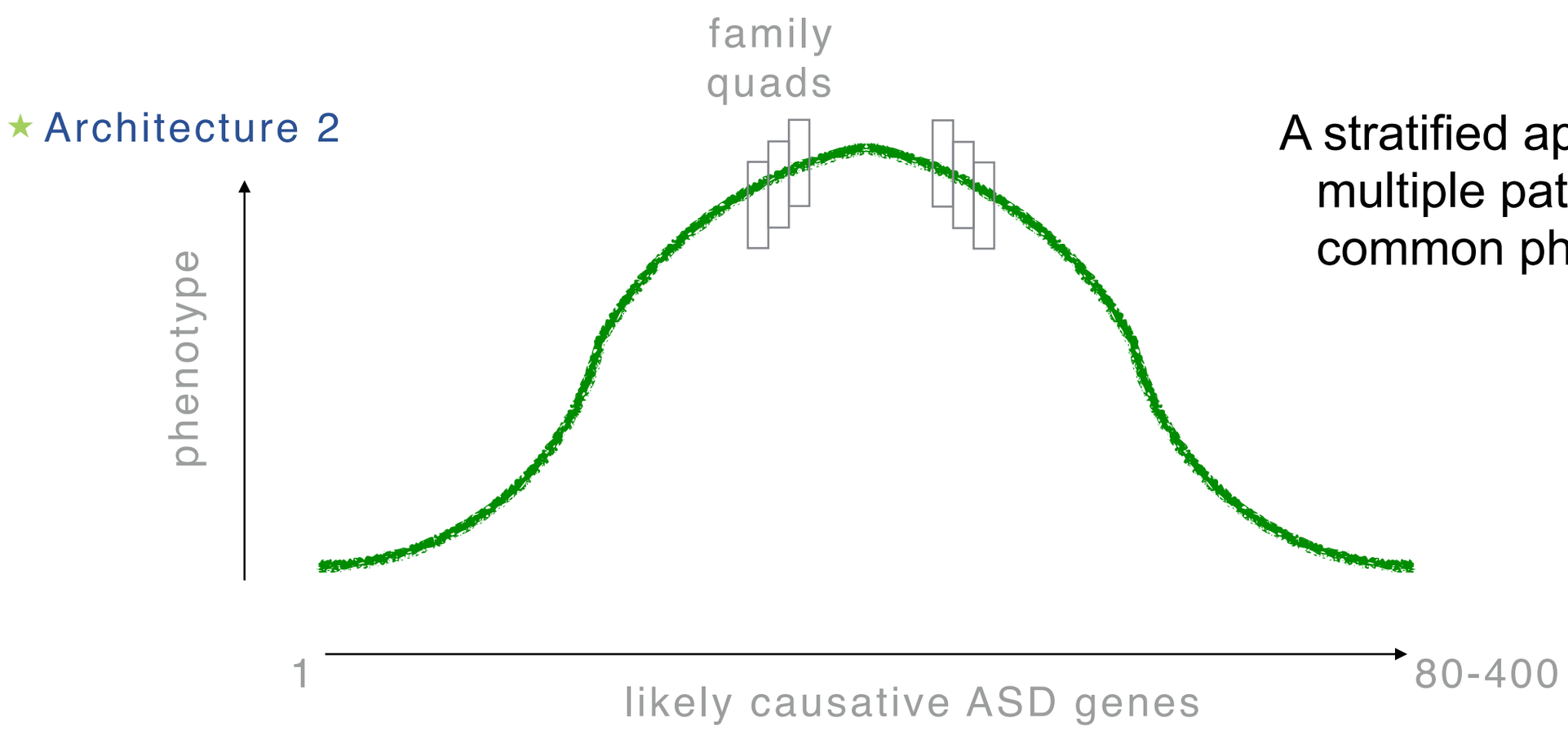
★ Architecture 1



NEUROPSYCHIATRIC PROJECTS

Autism As Exemplar Project

- NYGC Center for Common Disease Genomics
  - Large-scale family based sequencing of autism WGS.
    - Architecture 2: de novo presentation in context of quads/trios



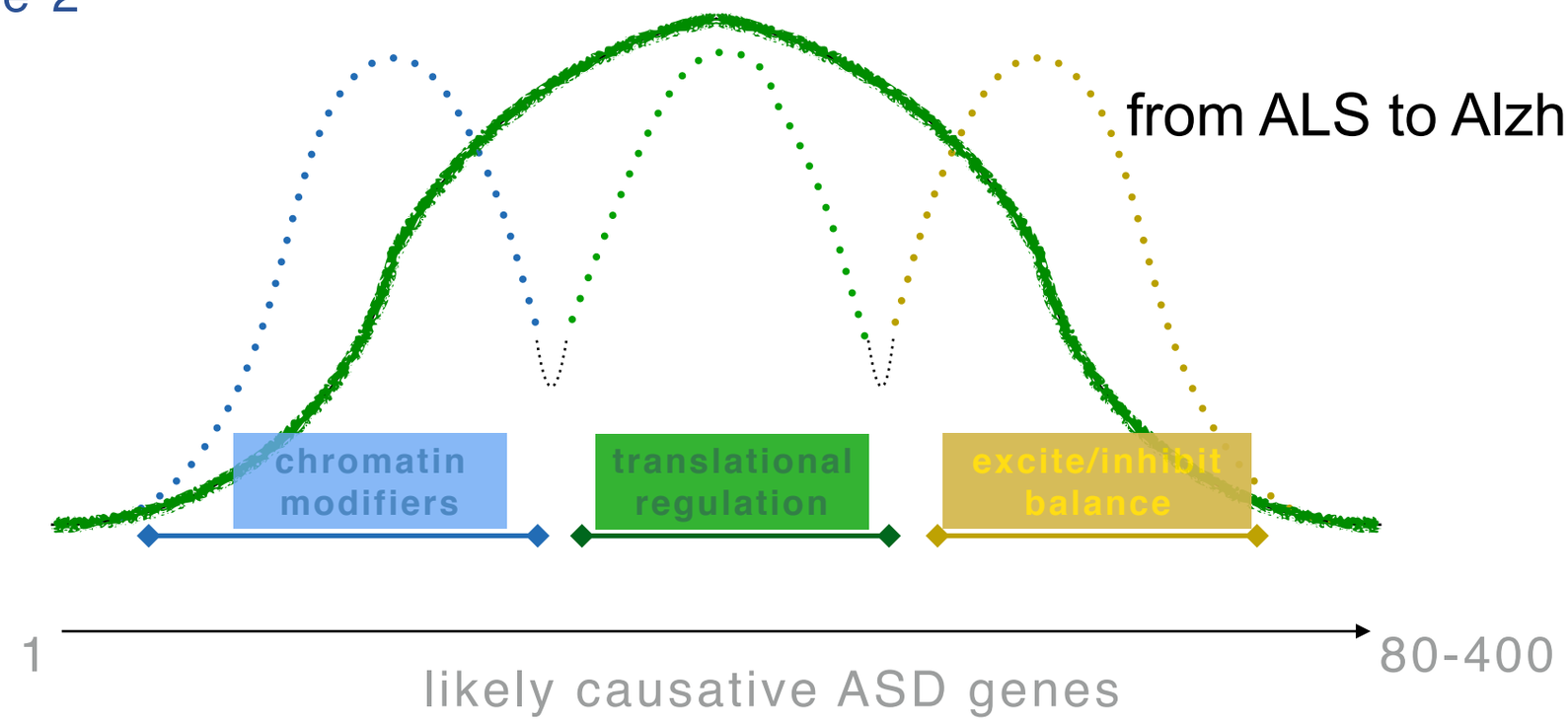
A stratified approach:
multiple pathways,
common phenotype

NEUROPSYCHIATRIC PROJECTS

Autism As Exemplar Project

- NYGC Center for Common Disease Genomics
  - Large-scale family based sequencing of autism WGS.
    - Architecture 2: de novo presentation in context of quads/trios

★ Architecture 2



Approach extends far beyond autism:

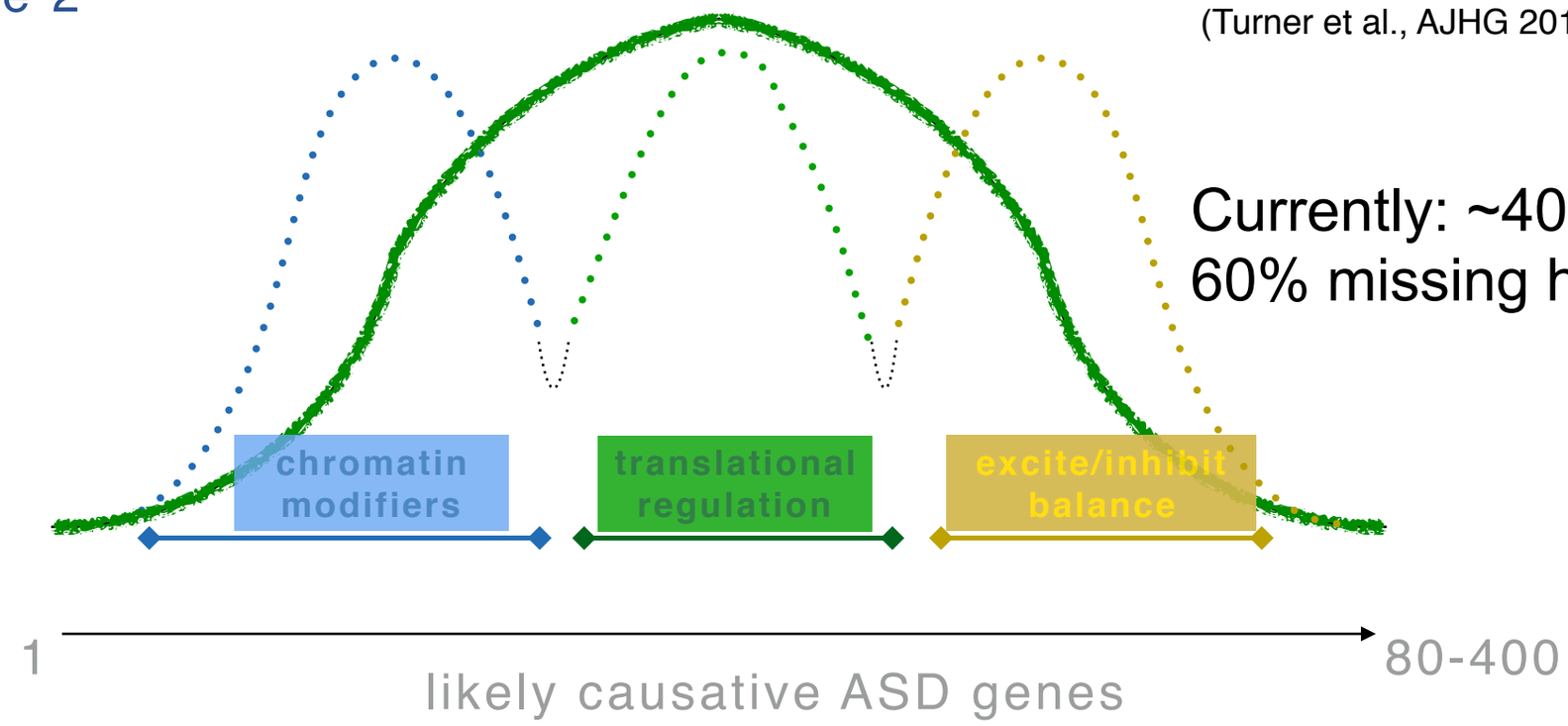
from ALS to Alzheimer's Dz

NEUROPSYCHIATRIC PROJECTS

Autism As Exemplar Project

- NYGC Center for Common Disease Genomics
  - Large-scale family based sequencing of autism WGS.
    - Architecture 2: de novo presentation in context of quads/trios

★ Architecture 2



WGS vs. WES, yield is 24% for genic variants
(Turner et al., AJHG 2016)

Currently: ~40% cases
60% missing heritability

NEUROPSYCHIATRIC PROJECTS

Autism As Exemplar Project

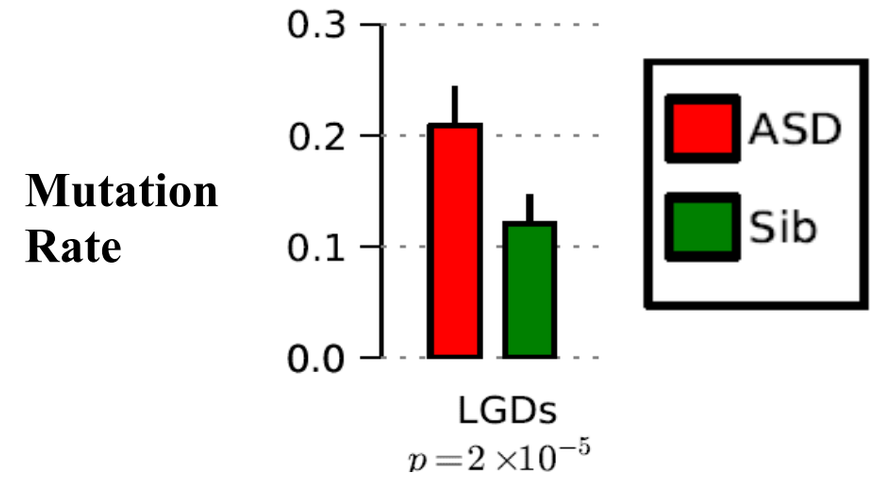
- Large-scale family based sequencing as approach to comprehensiveness.
  - Architecture 3: multiply affected individuals within well-phenotyped family architecture.
  - AGRE (NIHM/Geschwind)

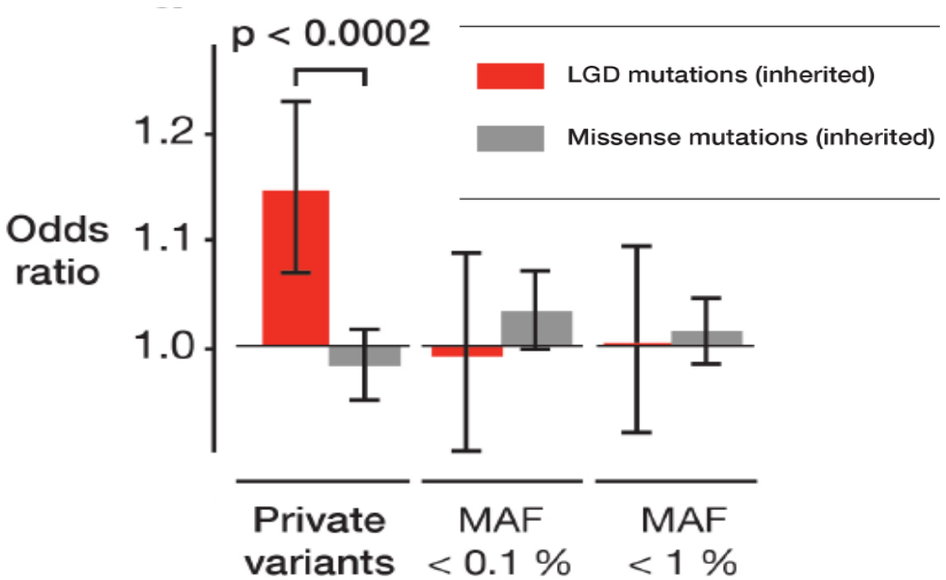
★ Architecture 3:

Number of affected individuals per family (may include MZ twins):		
Families with **two** affected individuals		1135
Families with **three** affected individuals		189
Families with **four** affected individuals		19
Families with **five** affected individuals		8
Twin Family Pedigrees		**323**
	Dizygotic	204
	Monozygotic	118
Triplet Family Pedigrees		15
Quadruplet Family Pedigrees (Monozygotic)		1

AUTISM

Genomic Architecture





Type I: Excess de novo mutations LGD and missense

21% of autism families

Iossifov et al., Nature, 2014

Type II; Inherited LGD transmitted preferentially from mothers to sons based on assessment of simplex quads

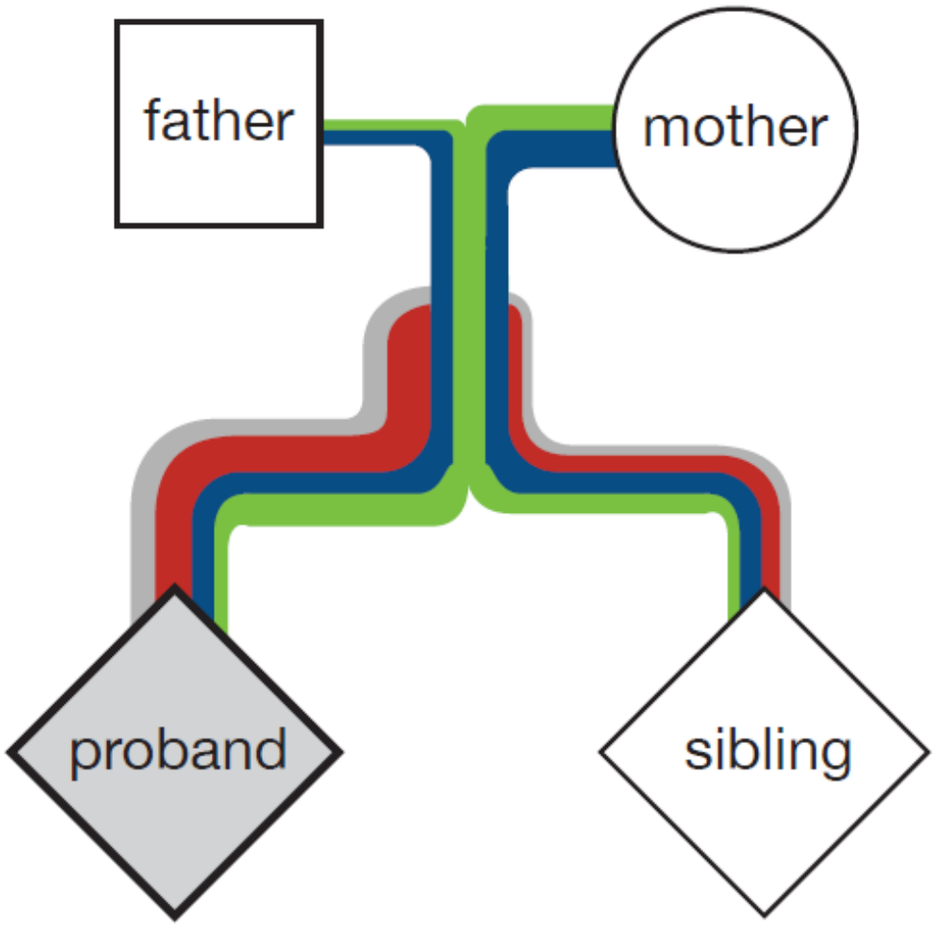
8% of families (attributable fraction)

Krumm et al., Nat. Genet., 2015

Multiplex families?

AUTISM

Integrated Genetic Model of Simplex Autism

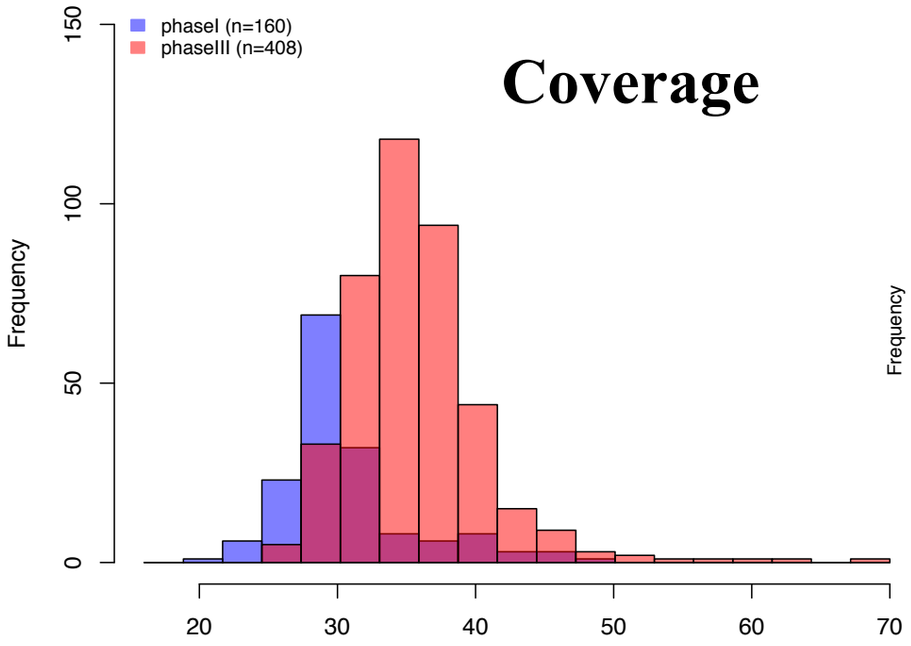


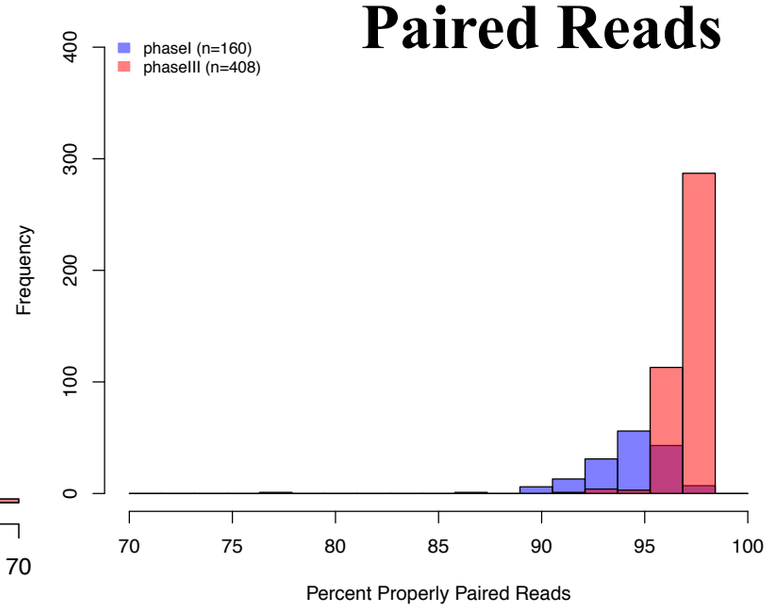
	Odds Ratio	P value
de novo SNV (disruptive)	1.68	$< 1 \times 10^{-5}$
de novo CNV	1.95	0.0003
Inherited CNVs - Maternal	**1.52**	**0.0005**
Inherited CNVs - Paternal	1.12	0.33
Inherited SNVs - Maternal	**1.11**	**0.0002**
Inherited SNVs - Paternal	1.06	0.18

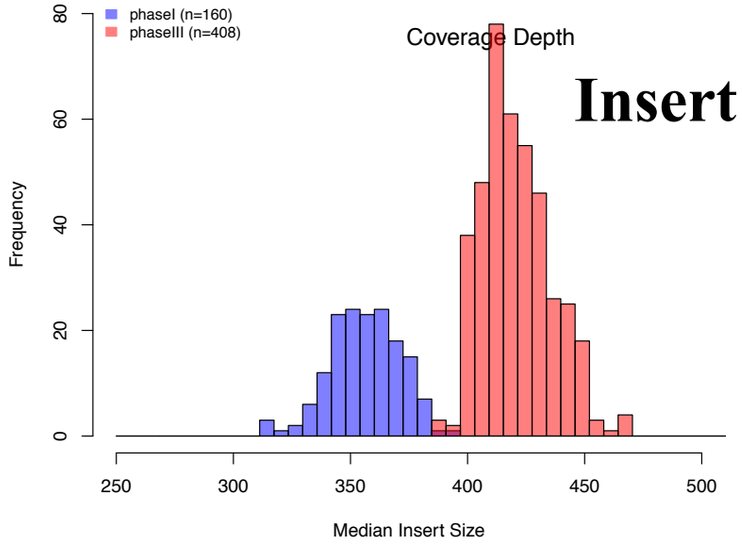
- Significant excess of CNVs in autism probands—when transmitted bias comes from mothers
- Excess of de novo gene-disruptive SNVs and indels—paternal germline
- Excess of private LGDs in conserved genes transmitted preferentially from mothers to sons—an explanation for the male bias

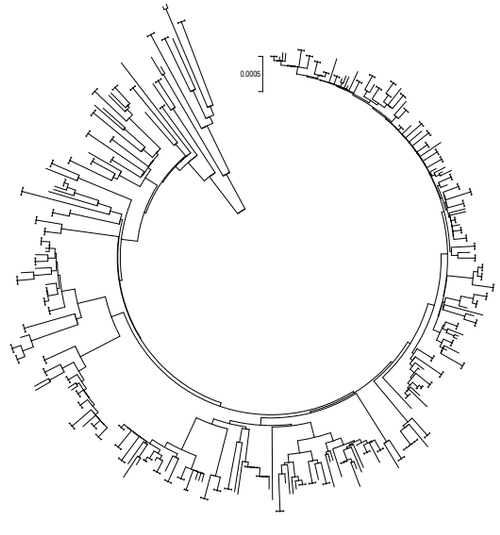
AUTISM

Quality of NYGC Genomes









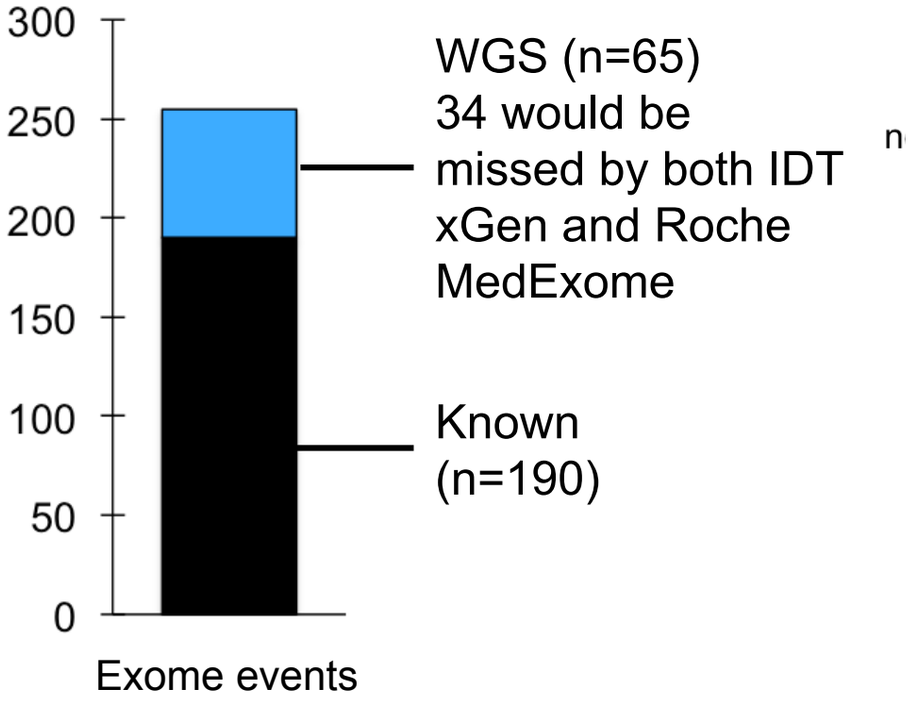
100% maternal-child transmission of mitochondria

AUTISM

WGS vs WES

de novo SNV/indel in the exome

- 102 new phase III families
  - 102 probands
  - 88 siblings*



* 14 families were only run as trios in the WES data

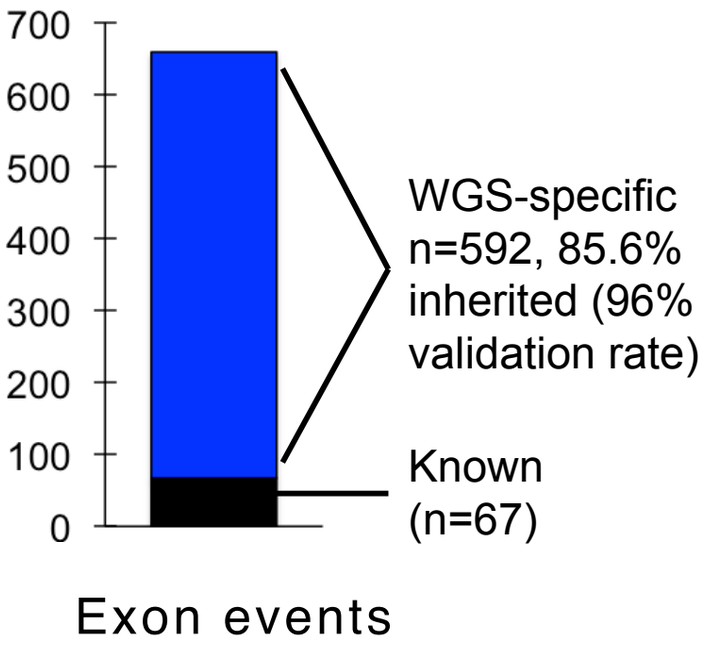
AUTISM

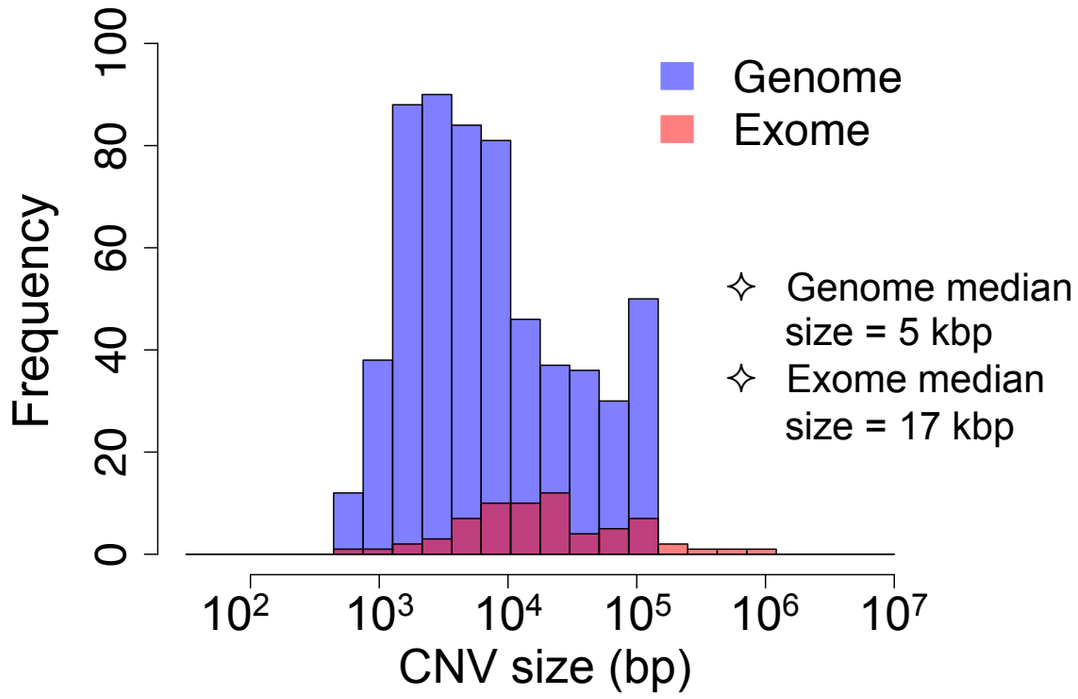
WGS vs WES

de novo/ inherited CNVs in the exome

COMPARISONS WITH:
X HMM
CONIFER

- 102 new phase III families
  - 102 probands
  - 88 siblings*



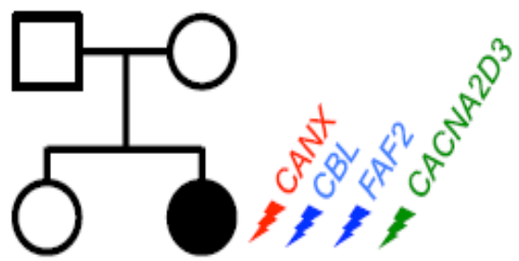


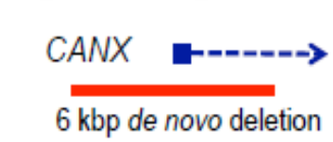
- In WGS specific regions there were 9 and 3 novel exonic de novo or private inherited deletions in SFARI genes in probands and siblings, respectively (p=0.11)
  - Proband (*BRAF, CSMD1, DMPK [2], INPP1, NRXN1, SLC27A4, EHMT1, KCNT1*)
  - Sibling (*PTPN11, EHMT1, KCNT1*)

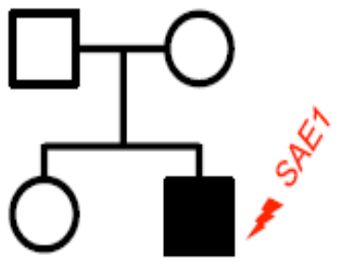
* 14 families were only run as trios in the WES data

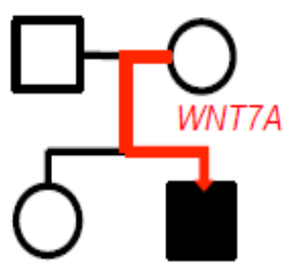
AUTISM

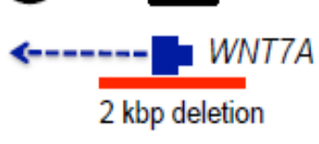
De novo and Private Disruption of Noncoding Regulatory Sequences

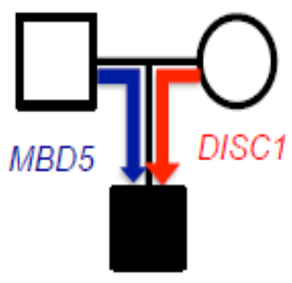


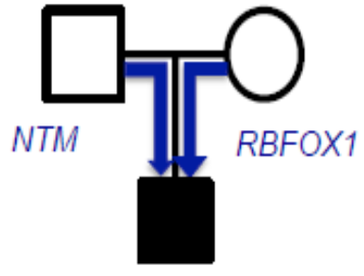


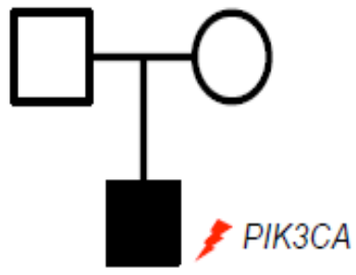












AUTISM

Noncoding regulatory mutations burden

(n=57 autism prior genes, Turner et al., 2016)

Variant category	*Autism counts (n=156)*	*Control counts (n=143)*	*Fisher 1-sided p-value*	*Fisher OR*
Total	7364	6375	-	-
Noncoding	1215	1096	0.87	0.95
Noncoding regulatory (d=10 kbp)	8	1	0.03	6.90
Noncoding regulatory (d=25 kbp)	9	1	0.02	7.80
Noncoding regulatory (d=50 kbp)	10	2	0.04	4.33
Noncoding regulatory (d=100 kbp)	13	4	0.05	2.82
Noncoding regulatory (d=500 kbp)	36	18	0.04	1.73
Noncoding regulatory (d=1 Mbp)	47	33	0.21	1.23

Enrichment of noncoding, regulatory mutations in probands

SNP events phase I-III + CNV events phase I-III

d = distance considered on either side of the gene

Turner et al, unpublished

AUTISM
Update

Total samples in process: 1,976

Current stages:

QC	1,484
Prep	72
Sequencing	420

NEUROPSYCHIATRIC PROJECTS

Autism & Epilepsy Overview

- Autism spectrum disorders (ASD) & Epilepsy show substantial comorbidity
  - Epilepsy rates in ASD are 20-40%
  - WES data show overlap
- Hypothesis: studying each disorder in parallel will enrich understanding of the other

AUTISM & EPILEPSY

Platform Synergy; NYGC Thoughts

Combining methods to determine functional significance

- Leveraging RVIS to develop NCRVIS (Goldstein)
- fitCons and fGWAS to incorporate fitness (Siepel and Pickrell)
- Integrate functional data - ENCODE, Epigenetics Roadmap, GTEx, DNA-ChIP, RNA CLIP, Alternative Splicing and Polyadenlylation

.

Comprehensiveness

- WES vs WGS; needs to be done to find out where value lies, how to mine data.
- Tx factors or miRNAs affecting multiple genes.
- Getting good intolerance score to apply to WGS.
- Scale
  - Platform comparisons; allows further apples:apples analyses (Eichler/NYGC Turner et al)

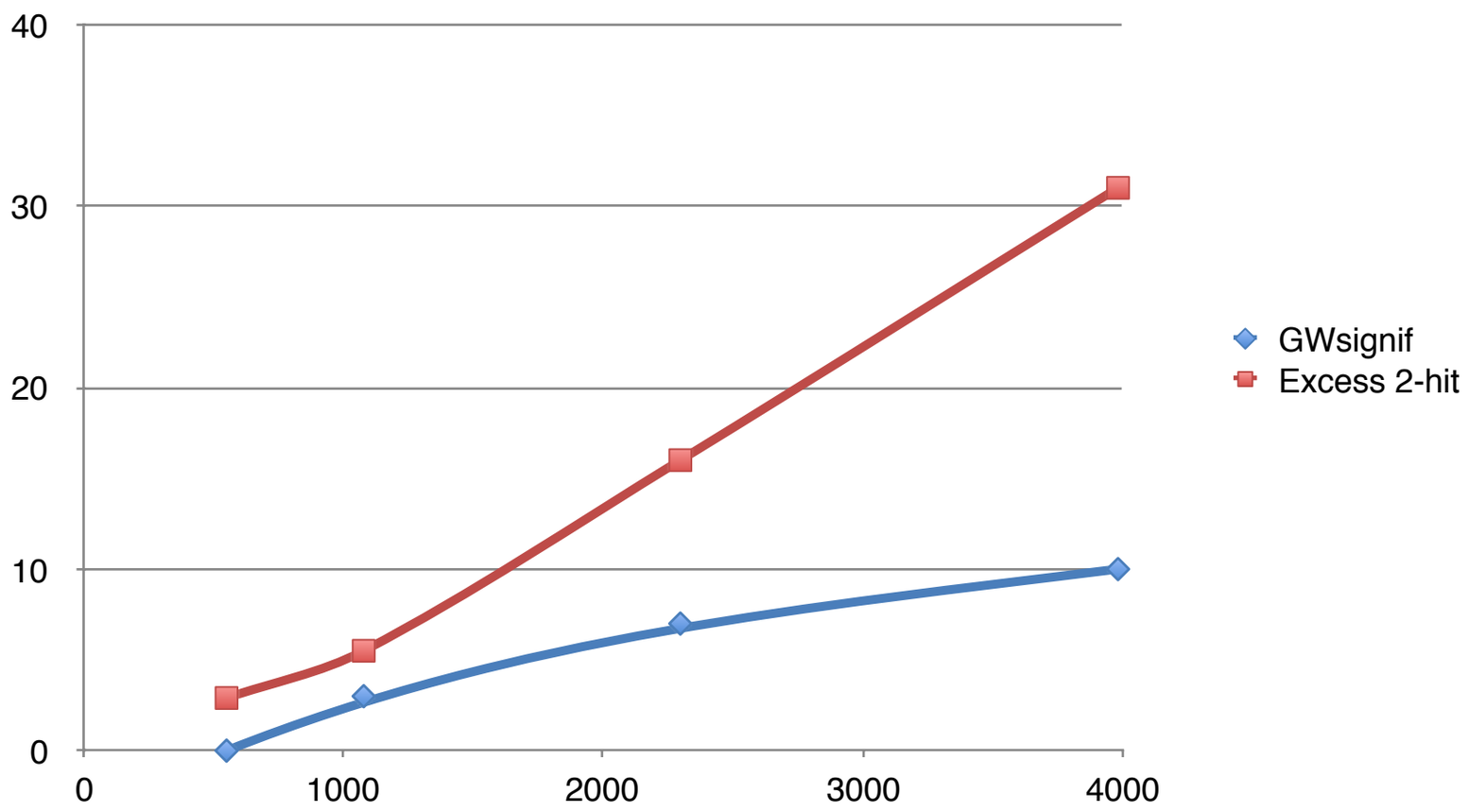
Broad Institute NP Production
Mark Daly

- Proposal – continue with exomes
- 2/3 Epilepsy, 1/3 Autism
- Y1 (ends Nov 2016) – expect ~8500 exomes
- Will take advantage of integration with massive scale-up of exome sequencing in schizophrenia and autism at Broad

Why exomes in autism

- Converting the long FDR lists from *de novo* publications into certain risk or non-risk genes
  - Need certain gene-phenotype connections between before functional/clinical studies undertaken
- Discovery of additional genes
  - Discovery via *de novo* mutation not yet plateauing
  - Genes with s < .1 will only be flagged by case/control and inherited variation with much larger samples
- Providing the pointers around which the WGS effort can focus
  - Early studies and theoretical power calculations suggest this is required

Exome discovery not plateauing



Each successive year (2012-2015) published exome studies demonstrates

- Increase in genomewide significant genes ($p < 2.5 \times 10^{-6}$)
- Larger increase in excess of 2-hit genes over chance / low FDR genes that need to be converted into true hits and non-hits

NEUROPSYCHIATRIC PROJECTS

Epilepsy

- Cost/benefit value discussion WES up front, WGS perhaps later (scale, scope)
  - ability to detect from proband only in many cases
    - DG: explanation … presume this is because de novo stand out from control population
  - Does this require stratification of gene analysis
    - by selection coefficient, RVIS?

EPILEPSY

Update

Approved: 7,500 cases and 3,500 controls
6,000-7000 samples will be sent to Broad soon

EPILEPSY

Update

WES, Probands vs. Control; Trios
Rationale
- Large S
- Proven value; review of WES
  - Prior Team Data:
    - Wigler
    - Eichler
    - Daly

Missing Heritability 75%

NEUROPSYCHIATRIC PROJECTS

Epilepsy

N = 650 GGE with epilepsy family history

N = 1,213 Non-acquired focal epilepsies (NAFE)

N = 543 NAFE with epilepsy family history

N = 3,422 IGM controls

Controls have not been ascertained for epilepsy, neuropsychiatric, neurodevelopmental or undiagnosed congenital disorders.

Analyses restricted to individuals of European genetic ancestry

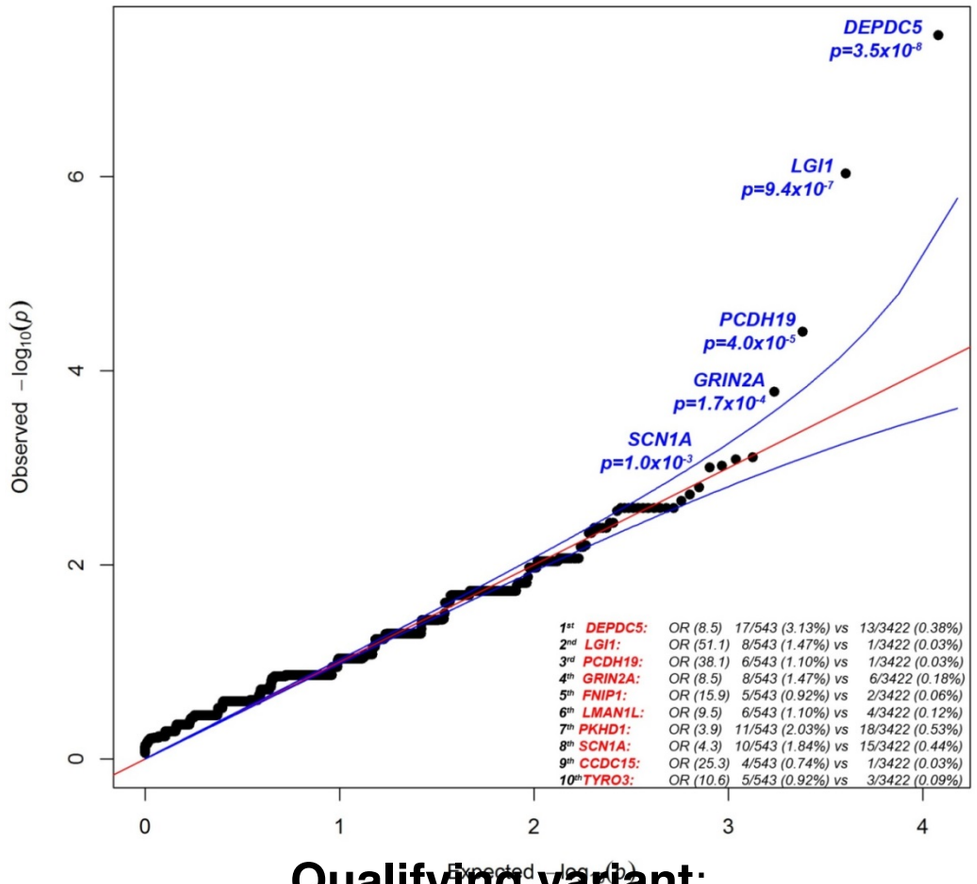
Above summaries include only samples passing sequencing and bioinformatic QC, known and cryptic relatedness testing, and have >85% of the CCDS sequence (~33Mb) covered at least 10-fold.

NEUROPSYCHIATRIC PROJECTS

Epilepsy - Preliminary Data

Do patients with epilepsy have more 'qualifying variants' in gene X than general control

NAFE Fam Hx + (**543** vs 3,422)



HGNC	RVIS%	Qual Case	Case Freq	Qual Ctrl	Ctrl Freq	FET p- value
DEPDC5	6.7%	17	3.13%	13	0.38%	**3.52E-08**
LGI1	8.8%	8	1.47%	1	0.03%	**9.37E-07**
PCDH19	5.3%	6	1.10%	1	0.03%	**3.98E-05**
GRIN2A	1.2%	8	1.47%	6	0.18%	**1.65E-04**
FNIP1	11.2%	5	0.92%	2	0.06%	**7.83E-04**
LMAN1L	78.1%	6	1.10%	4	0.12%	**8.27E-04**
PKHD1	67.4%	11	2.03%	18	0.53%	**9.59E-04**
SCN1A	2.4%	10	1.84%	15	0.44%	**0.001**
CCDC15	93.6%	4	0.74%	1	0.03%	**0.0016**
TYRO3	10.6%	5	0.92%	3	0.09%	**0.0019**

Qualifying variant:
High confidence variant call
LoF / Polyphen "Probably" prediction
Ultra-rare and absent among EVS and ExAC
(i.e. ~<0.0008% MAF)

Summary:
Four of the 43 known genes occupy genome-wide ranks [1-4], ***p=2.7x10^{-11}***

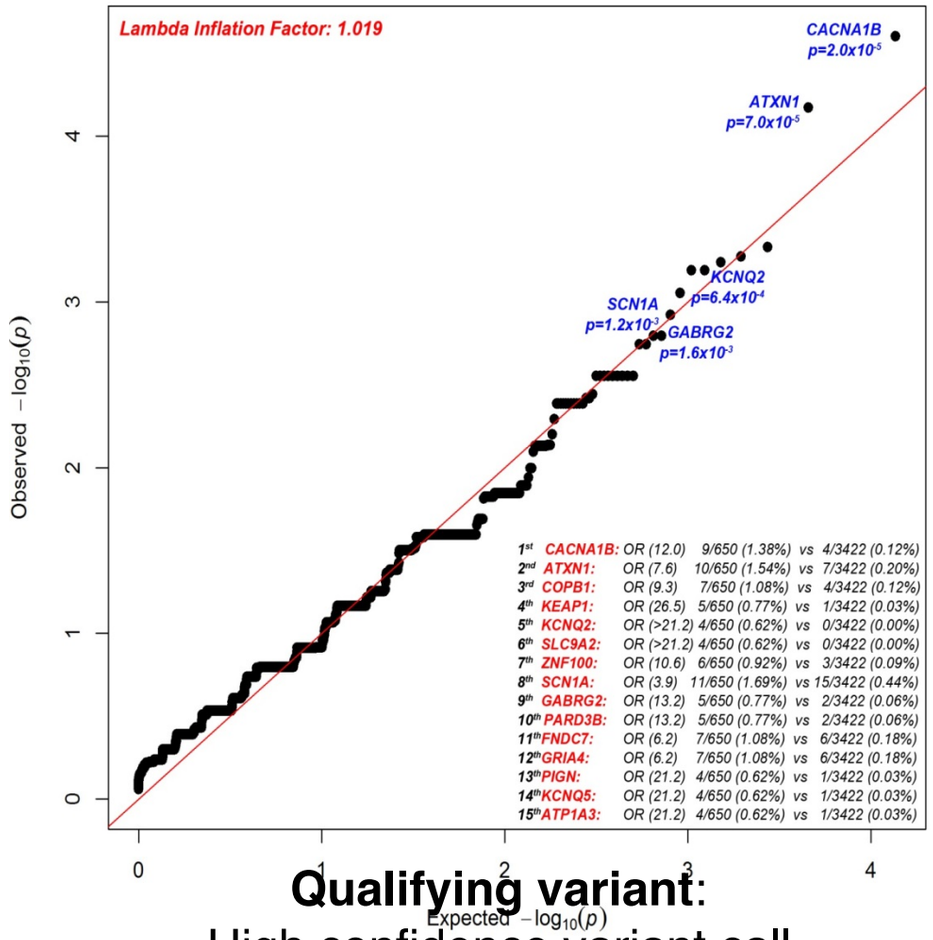
Interpretation:
Compelling evidence of lower locus heterogeneity for NAFE, relative to GGE. This suggests potentially better genetic tractability for focal epilepsies.

NEUROPSYCHIATRIC PROJECTS

Epilepsy - Preliminary Data

Do patients with epilepsy have more 'qualifying variants' in gene X than general control

GGE (650 vs 3,422)



Qualifying variant:
High confidence variant call
LoF / Polyphen "Probably" prediction
Ultra-rare and absent among EVS and ExAC
(i.e. ~<0.0008% MAF)

HGNC	RVIS%	Qual Case	Case Freq	Qual Ctrl	Ctrl Freq	FET p-value
CACNA1B	0.8%	9	1.38%	4	0.12%	**$2.5x10^{-5}$**
ATXN1	8.9%	10	1.54%	7	0.20%	**$6.7x10^{-5}$**
COPB1	24.9%	7	1.08%	4	0.12%	**$4.7x10^{-4}$**
KEAP1	8.8%	5	0.77%	1	0.03%	**$5.3x10^{-4}$**
KCNQ2	5.9%	4	0.62%	0	0%	**$6.4x10^{-4}$**
SLC9A2	4.0%	4	0.62%	0	0%	**$6.4x10^{-4}$**
ZNF100	69.2%	6	0.92%	3	0.09%	**$8.8x10^{-4}$**
SCN1A	2.4%	11	1.69%	15	0.44%	**0.0012**
GABRG2	10.5%	5	0.77%	2	0.06%	**0.0016**
PARD3B	73.9%	5	0.77%	2	0.06%	**0.0016**
GRIA4	3.1%	7	1.08%	6	0.18%	**0.0018**

Summary:
No single gene is genome-wide significant:
Adjusted alpha $p=4x10^{-6}$

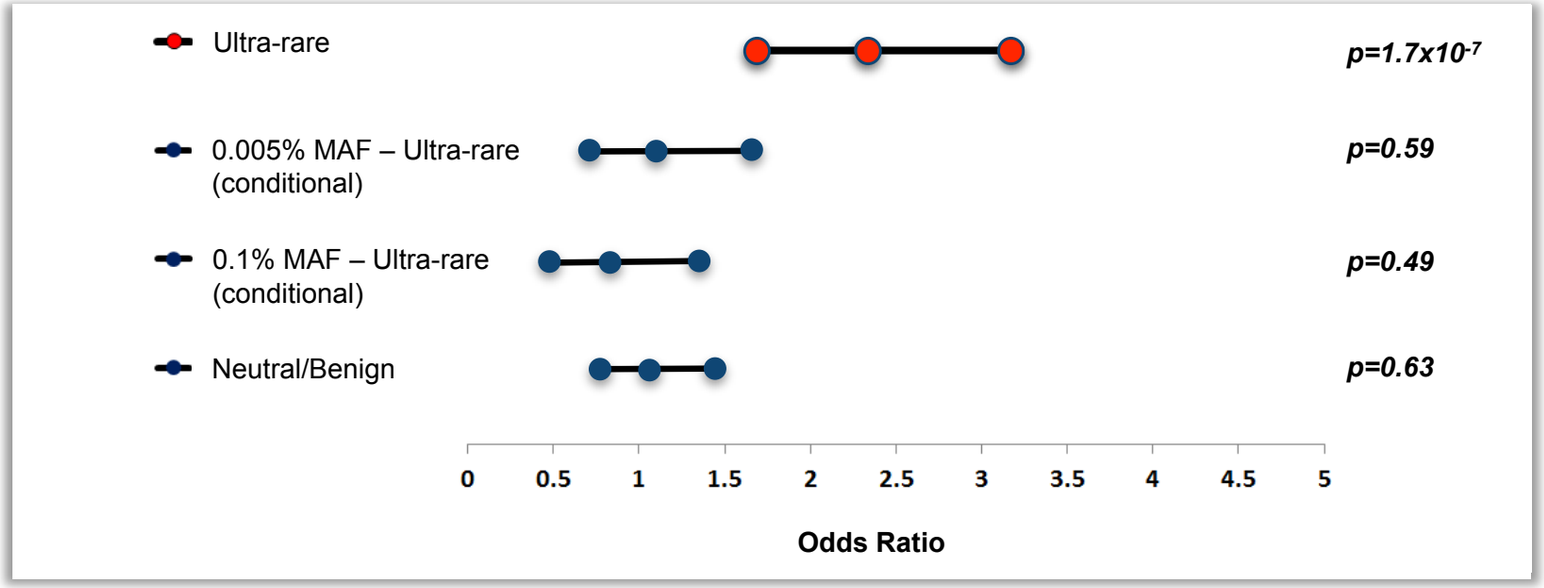
Interpretation:
Single genes do not account for a high proportion of GGE risk. Likely due to high genetic and/or phenotypic heterogeneity.

NEUROPSYCHIATRIC PROJECTS

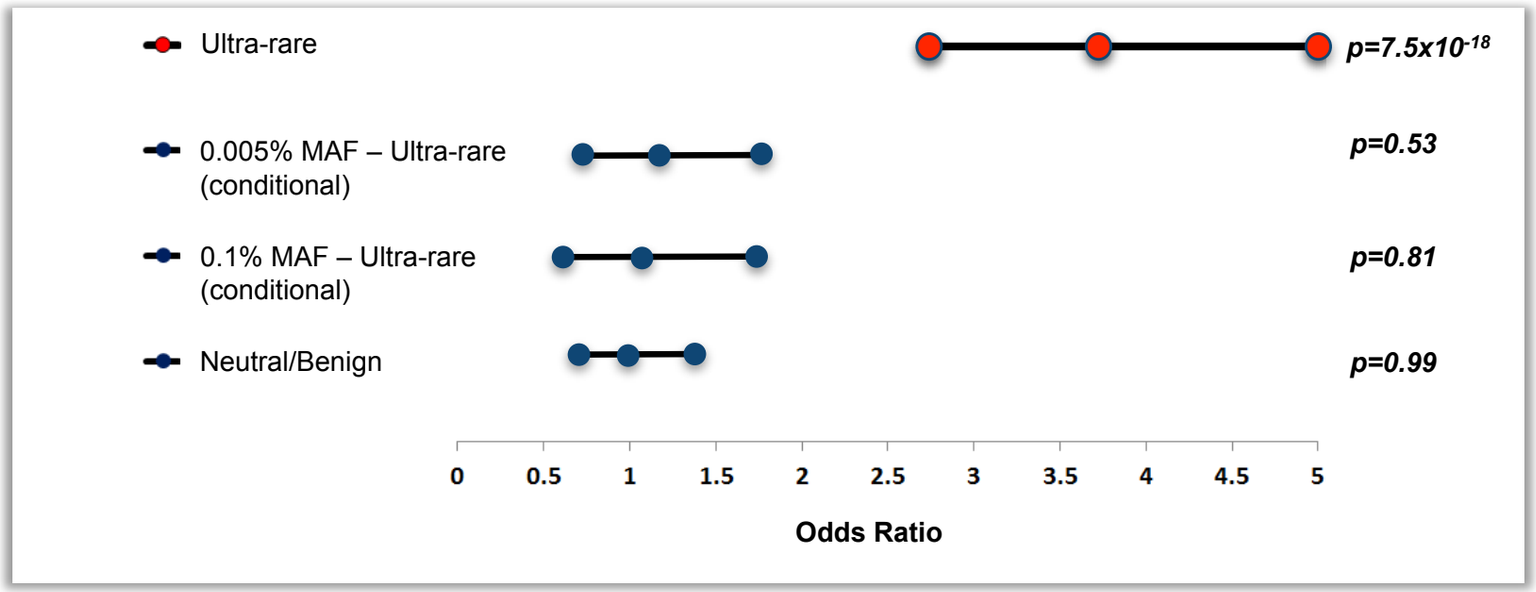
Epilepsy - Preliminary Data

Enrichment of qualifying variants among 43 known epilepsy genes

GGE



NAFE



AUTISM & EPILEPSY

Platform Synergy; Broad Thoughts

We discover genes through exomes, we complete allelic series with non-coding

- We succeed at exome point mutations because
  - A) We know the precise location of exons (**~1%** of genome)
  - B) We know "the code" (i.e., can recognize the **~1%** of variants that are truncating vs. the 99% that are missense/synonymous)
- Non-coding point mutations at an enormous disadvantage
  - Even if there are similar high-impact variants, we don't recognize them (*recall E. Lander talk Tuesday morning*)
- Non-coding CNVs avoid these limitations
  - Taking out large regions, removing critical variants without us needing to recognize them at nucleotide level

As in Mendelian disease: Continued exome discoveries will help target the non-coding CNV analyses (ala Tychele's paper) and eventually some day non-coding point mutations

DISCUSSION