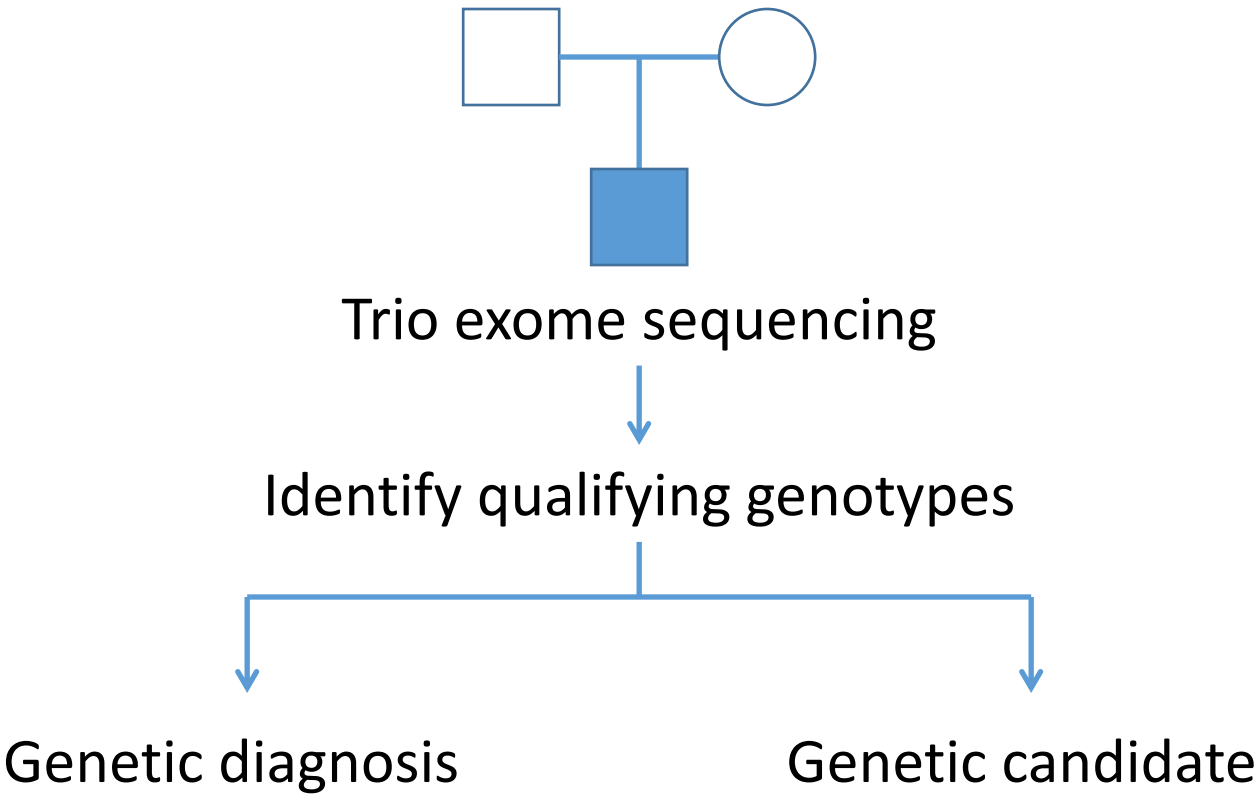
Clinical Sequencing for Rare Disease



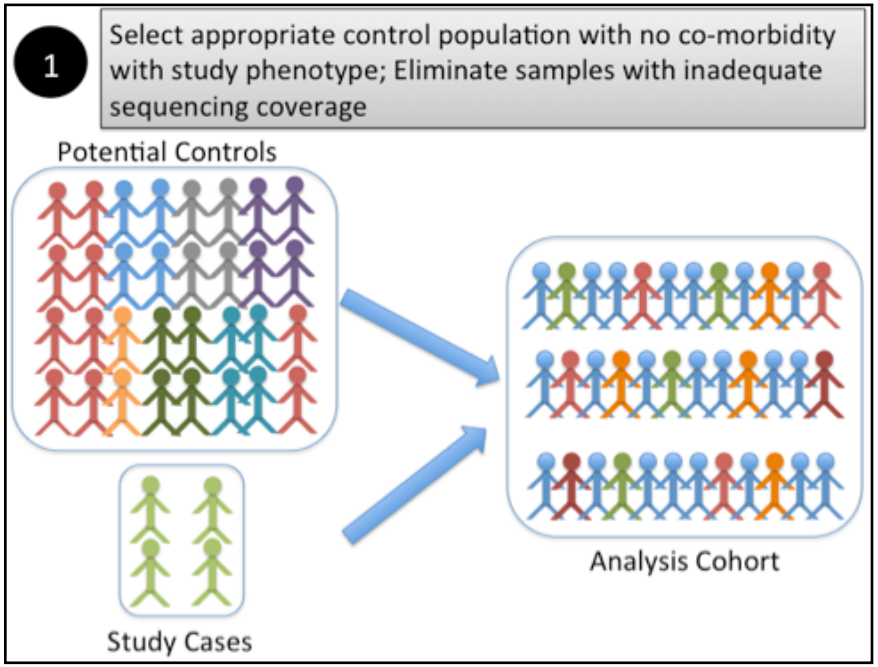
DIAGNOSTIC ANALYSIS FRAMEWORK

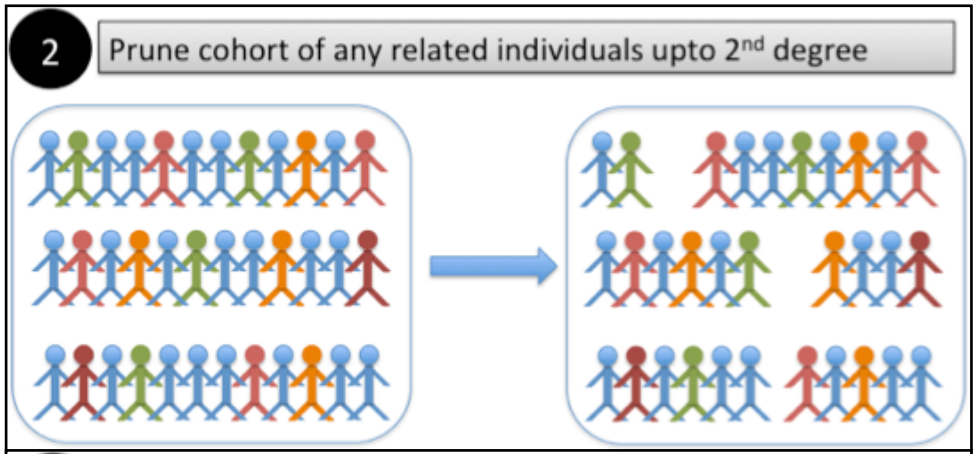
Identify rare and functional genetic variation in genes that have previous known association with disease. We look for variants that are

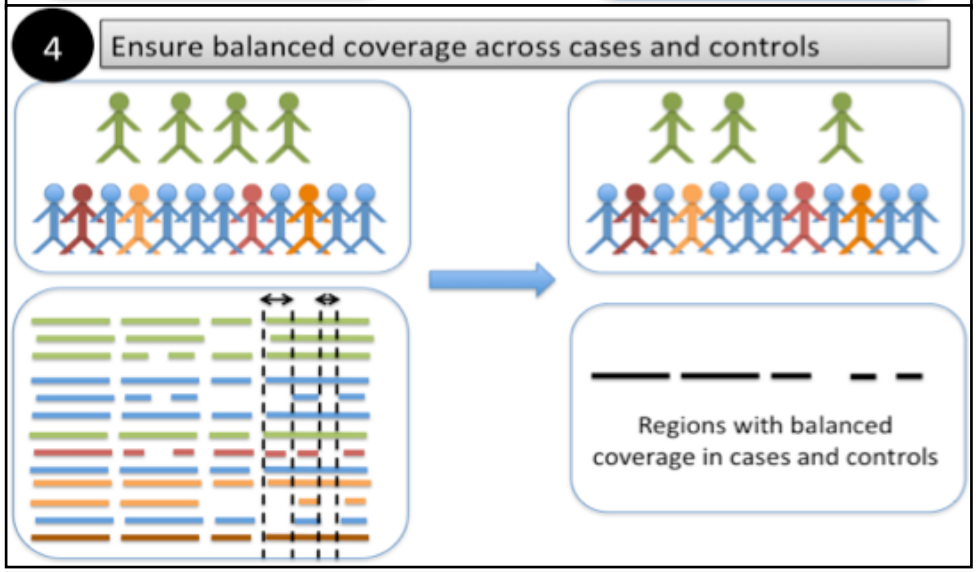
- High quality
  - MQ>40, QD>2, QUAL>30
- Extremely rare
  - Variants that are represented maximally up to 5 time across internal and external (EVS, ExAC) controls
- Present in KnownVar - ClinVar, HGMD
  - Previously reported pathogenic
  - At same or adjacent genomic sites
- LoF in genes in KnownVar
  - LoF - Nonsense, splice donor/acceptor, frameshift
  - Haploinsufficient gene
    - Clinvar reported pathogenic LoF variants
    - ClinGen classification of haploinsufficient
- LoF in LoF Depleted genes
  - pLI score > 0.9

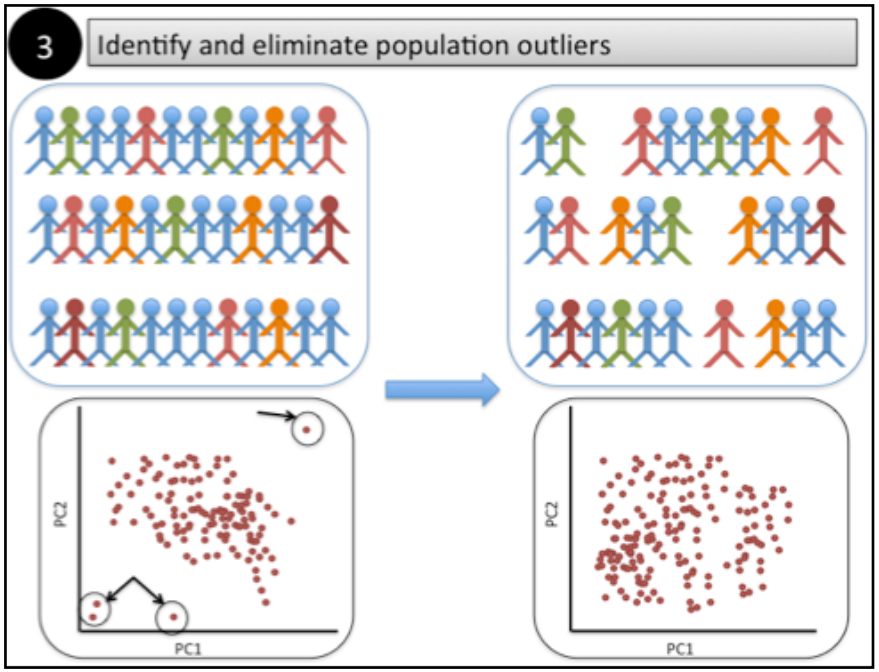
A remarkably successful clinical test

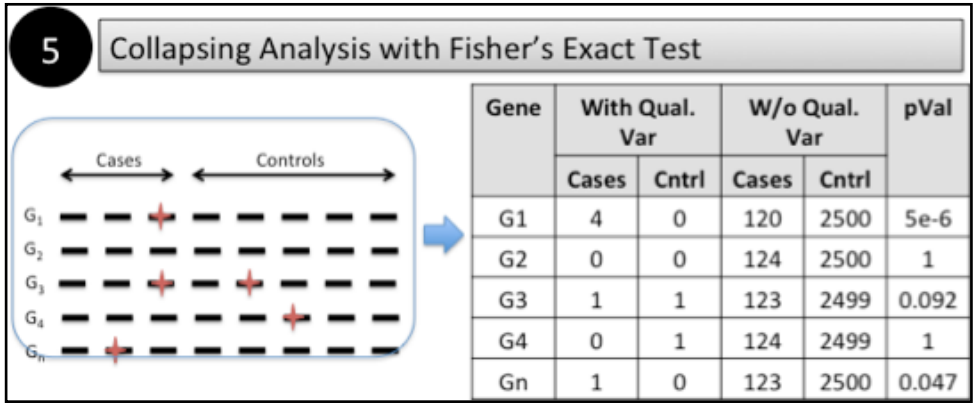
Study	Journal	N	Ascertainment	% resolved
Need 2012	J Med Genet	12	Mixture	50%
Yang 2013	NEJM	250	80% Neuro	25%
Calvo 2012	Sci Transl Med	42	Mitochondrial	24%
DeLigt 2013	NEJM	100	Severe ID	16%
Zhu 2014	Genetics in Medicine	119	Mixture	24%
Srivastava 2014	Annals of Neuro	78	Neuro	41%
Yang 2014	JAMA	2,000	Mixture	25%
Lee 2014	JAMA	814	Mixture	26%
Soden 2014	Sci Transl Med	119	Neuro	45%
Combined	**-**	**3,534**	**Mixture**	**26%**











Gene	With Qual. Var		W/o Qual. Var		pVal
	Cases	Cntrl	Cases	Cntrl	
G1	4	0	120	2500	5e-6
G2	0	0	124	2500	1
G3	1	1	123	2499	0.092
G4	0	1	124	2499	1
Gn	1	0	123	2500	0.047

Datasets

N = 650 GGE with epilepsy family history

N = 1,213 Non-acquired focal epilepsies (NAFE)
N = 543 NAFE with epilepsy family history

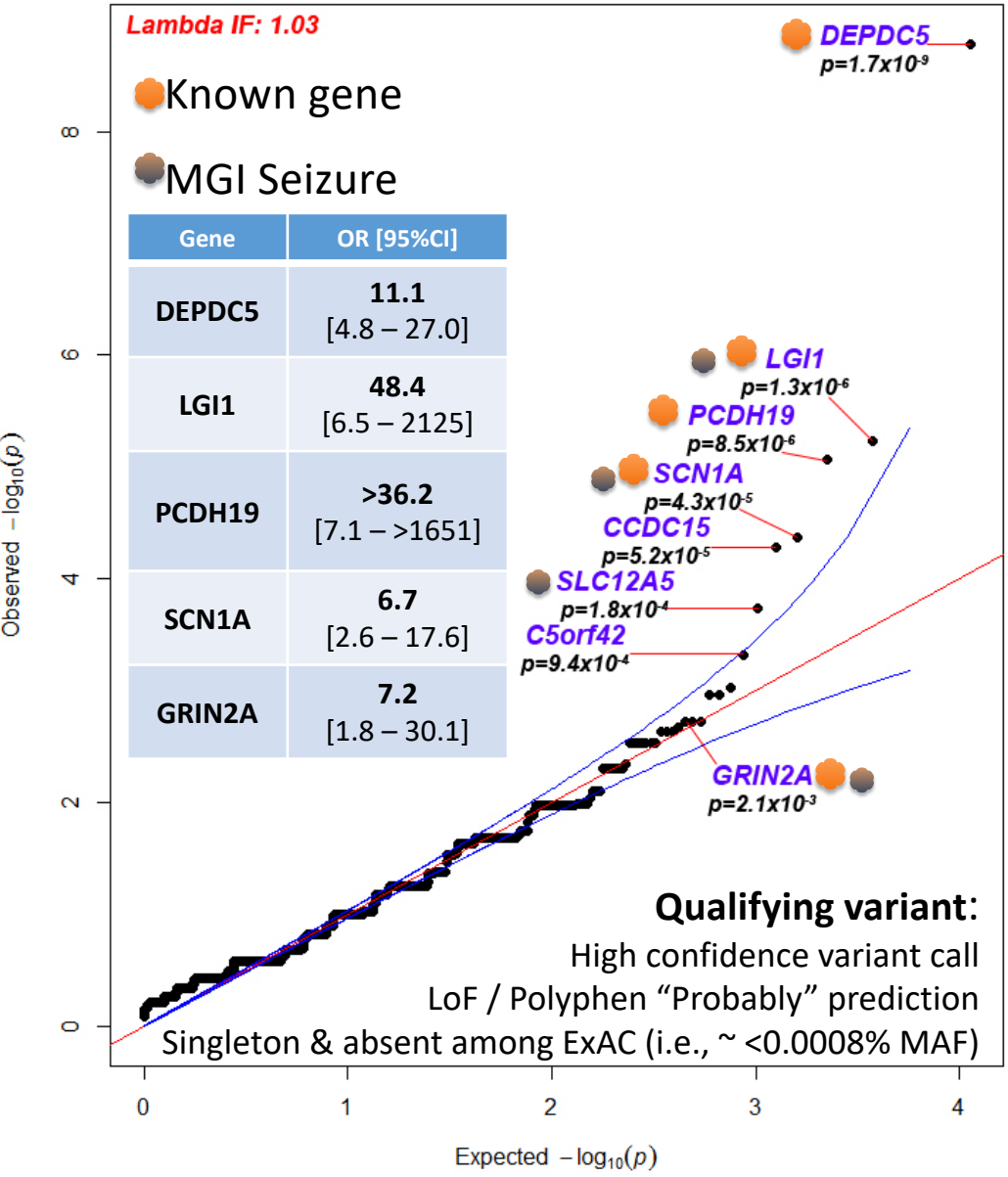
N = 3,422 IGM controls
Controls have not been ascertained for epilepsy, neuropsychiatric, neurodevelopmental or undiagnosed congenital disorders

Analyses restricted to individuals of European genetic ancestry

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

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

NAFE Fam Hx + (586 vs 3,503)



Gene	OR [95%CI]
DEPDC5	11.1 [4.8 – 27.0]
LGI1	48.4 [6.5 – 2125]
PCDH19	>36.2 [7.1 – >1651]
SCN1A	6.7 [2.6 – 17.6]
GRIN2A	7.2 [1.8 – 30.1]

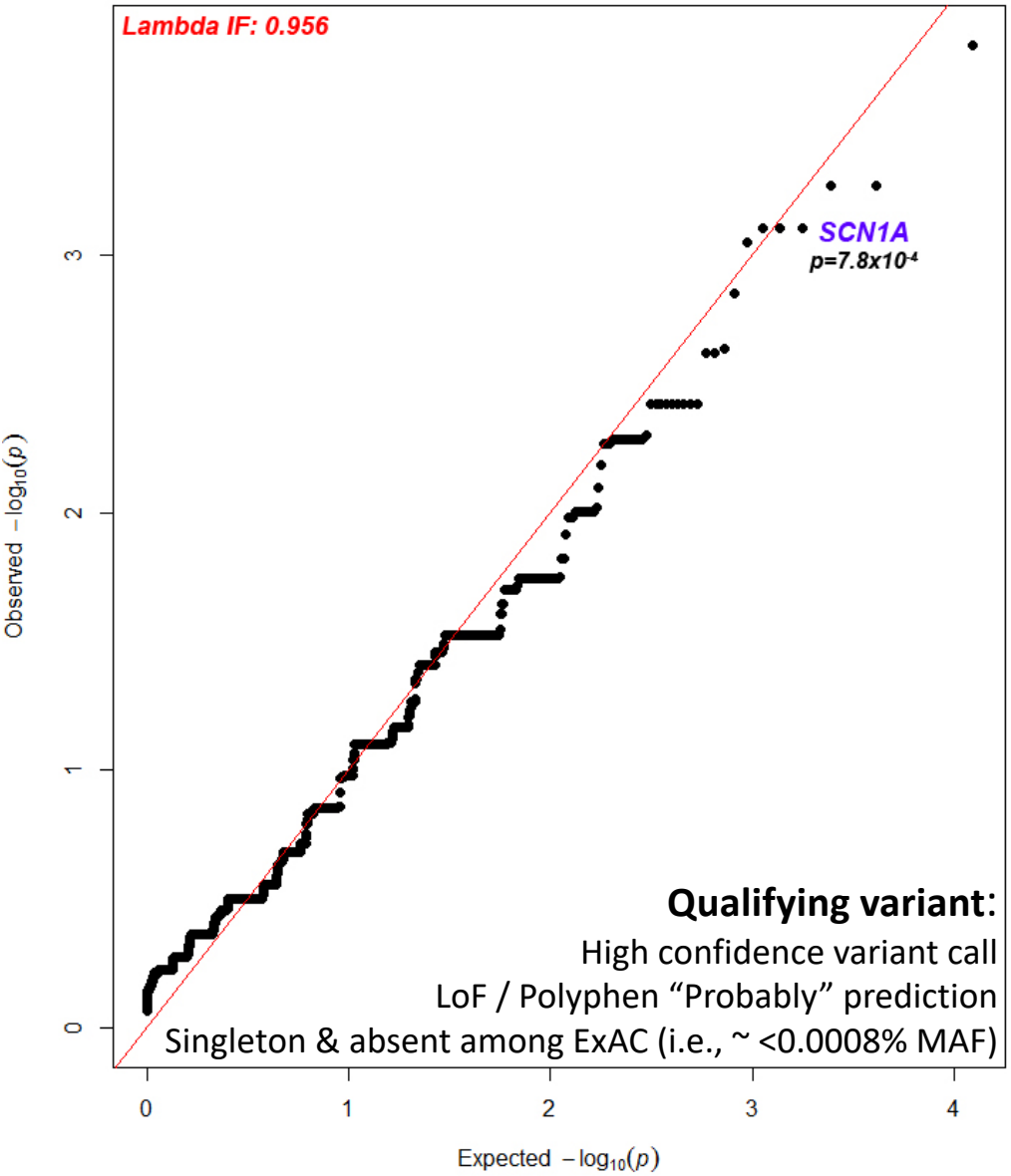
Qualifying variant:
High confidence variant call
LoF / Polyphen "Probably" prediction
Singleton & absent among ExAC (i.e., ~ <0.0008% MAF)

HGNC	RVIS%	Qual Case	Case Freq	Qual Ctrl	Ctrl Freq	FET p- value
DEPDC5	**6.6%**	**18**	**3.1%**	**10**	**0.3%**	**$1.7x10^{-9}$**
LGI1	**14.4%**	**8**	**1.4%**	**1**	**0.03%**	**$1.3x10^{-6}$**
PCDH19	**10.4%**	**6**	**1.0%**	**0**	**0%**	**$8.5x10^{-6}$**
SCN1A	**4.0%**	**11**	**1.9%**	**10**	**0.3%**	**$4.3x10^{-5}$**
CCDC15	16.0%	6	1.0%	1	0.03%	$5.2x10^{-5}$
SLC12A5	4.5%	6	1.0%	2	0.06%	$1.8x10^{-4}$
C5orf42	19.9%	7	1.2%	6	0.2%	$9.4x10^{-4}$
TRPM5	11.3%	6	1.0%	4	0.1%	0.001
ADCY10	83.8%	6	1.0%	4	0.1%	0.001
C9orf3	45.6%	4	0.7%	1	0.03%	0.002

Summary:
Four of the 30 known genes occupy genome-wide ranks [1-4], ***$p=6x10^{-12}$***

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

IGE/GGE (733 vs 3,503)



HGNC	RVIS%	Qual Case	Case Freq	Qual Ctrl	Ctrl Freq	FET p-value
RTFDC1	28.9%	5	0.7%	0	0%	$1.5x10^{-4}$
COPB1	6.7%	6	0.8%	2	0.06%	$5.4x10^{-4}$
PNPLA1	93.6%	6	0.8%	2	0.06%	$5.4x10^{-4}$
SCN1A	**4.0%**	**10**	**1.4%**	**10**	**0.3%**	**$7.8x10^{-4}$**
CACNA1B	3.0%	7	1.0%	4	0.1%	$7.8x10^{-4}$
WDR83	33.2%	5	0.7%	1	0.03%	$7.9x10^{-4}$
SLC1A7	24.7%	4	0.6%	0	0%	$8.9x10^{-4}$
PARD3B	62.8%	6	0.8%	3	0.09%	0.001
FAT4	21.8%	15	2.1%	25	0.7%	0.002
ATXN1	20.9%	5	0.7%	2	0.06%	0.002

Summary:

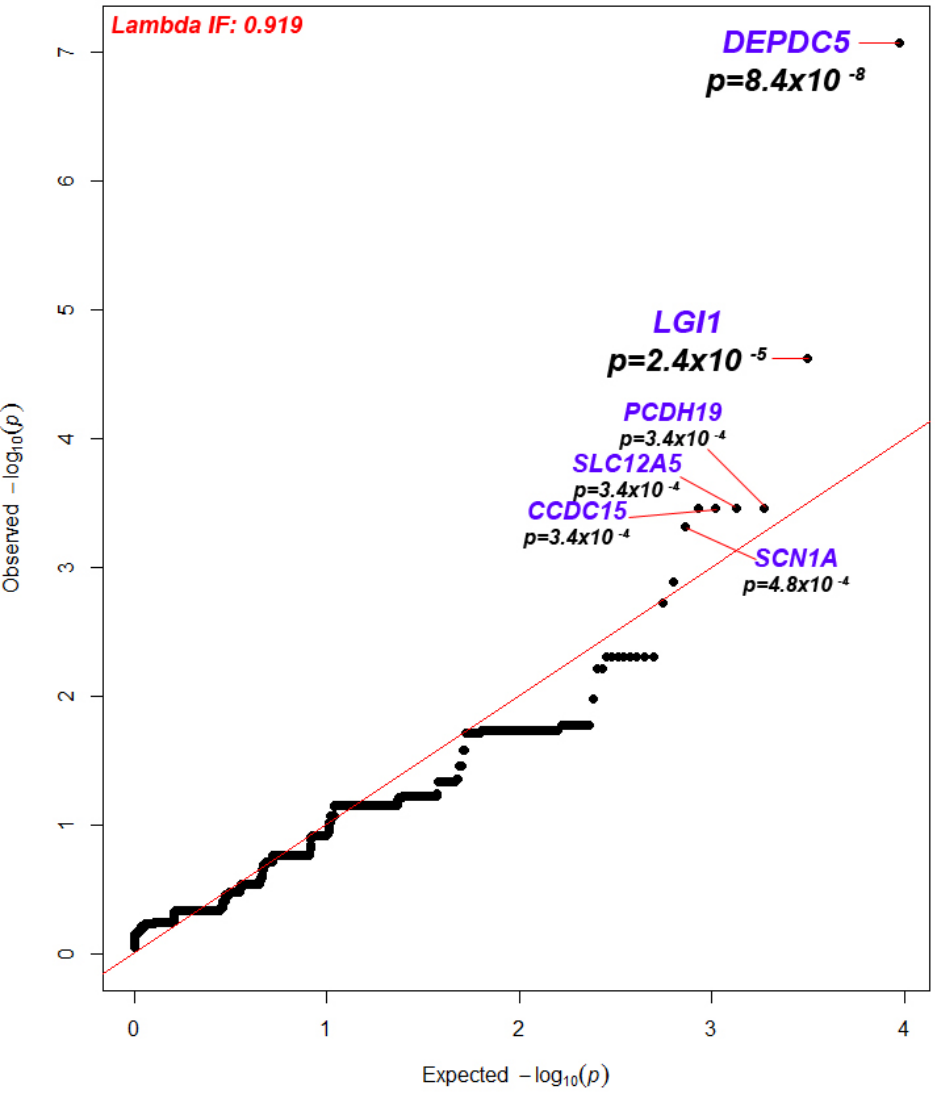
No single gene is genome-wide significant:

Adjusted alpha $p=4x10^{-6}$

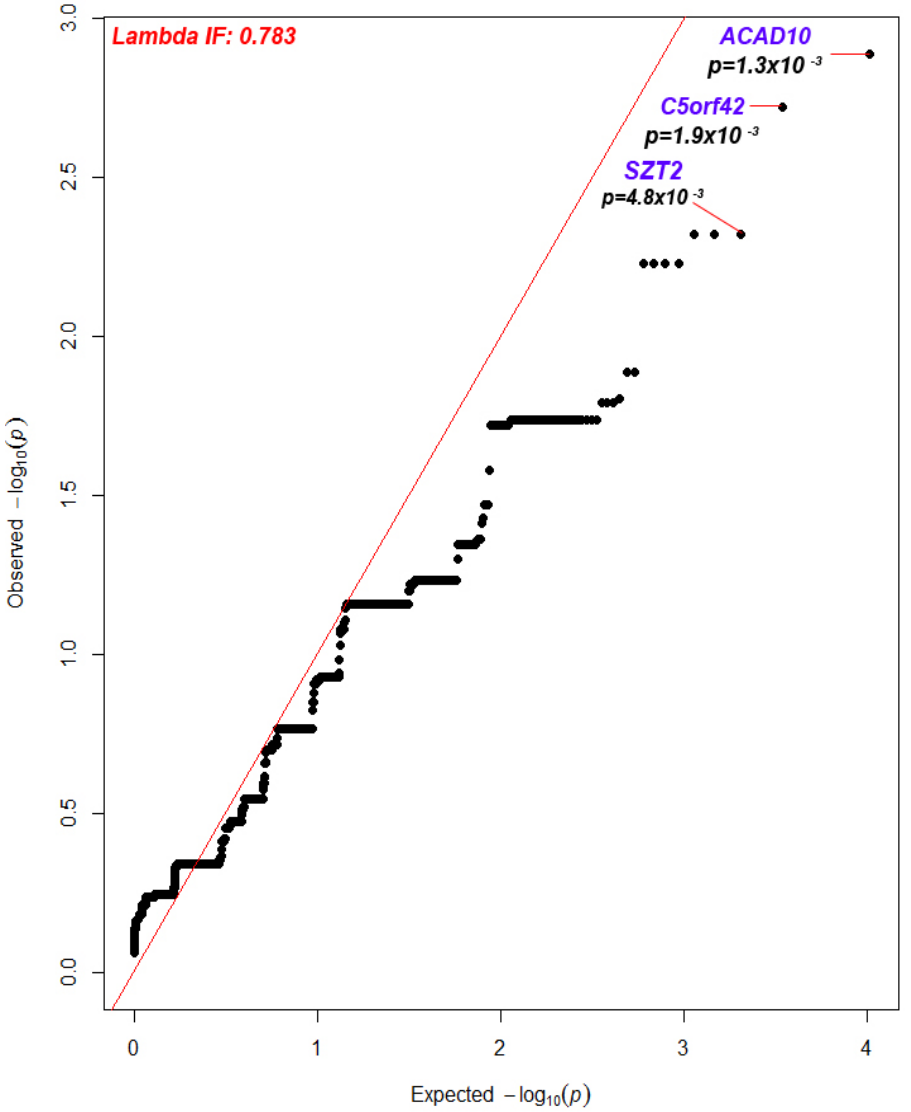
Interpretation:

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

Family History (586 vs 1,621)

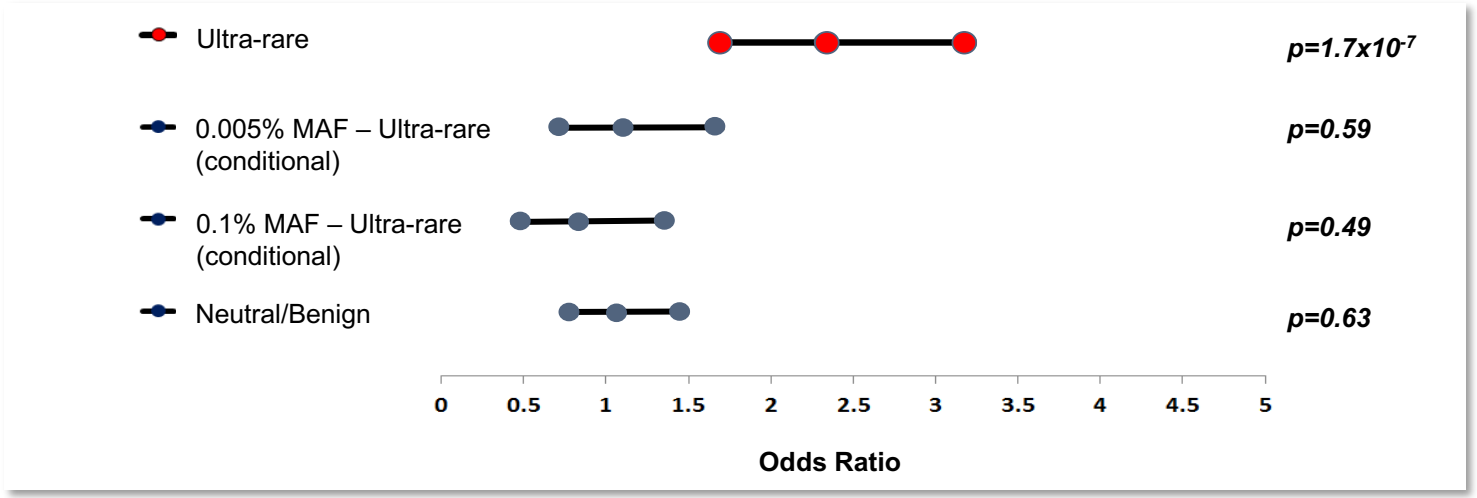


Sporadic NAFE (658 vs 1,882)



Enrichment of qualifying variants among 43 known epilepsy genes

GGE



NAFE

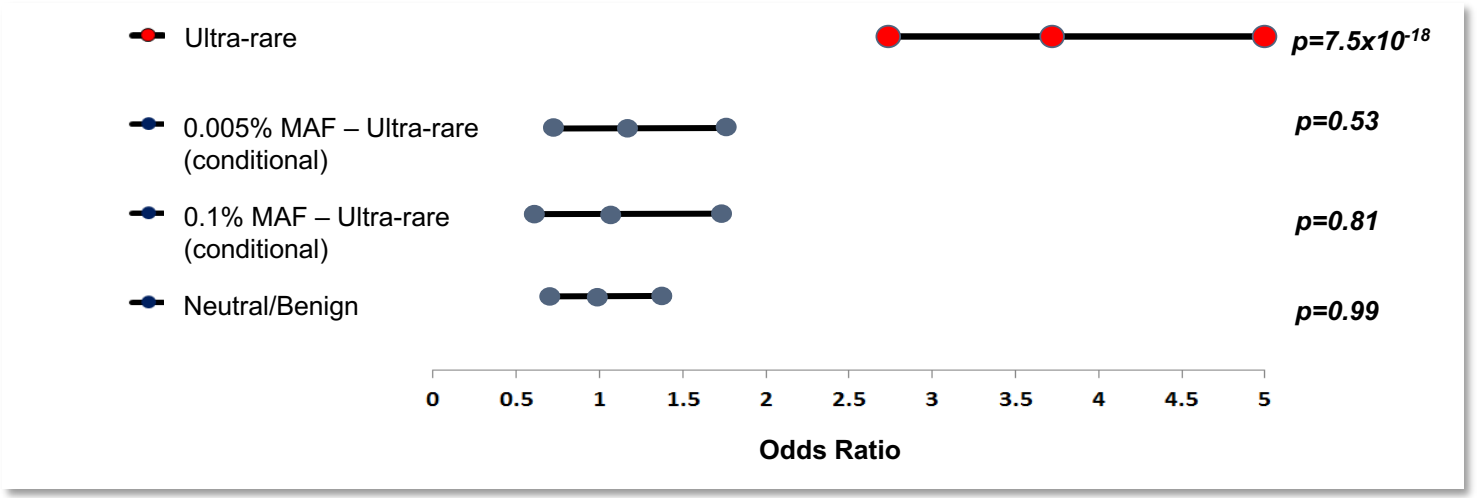
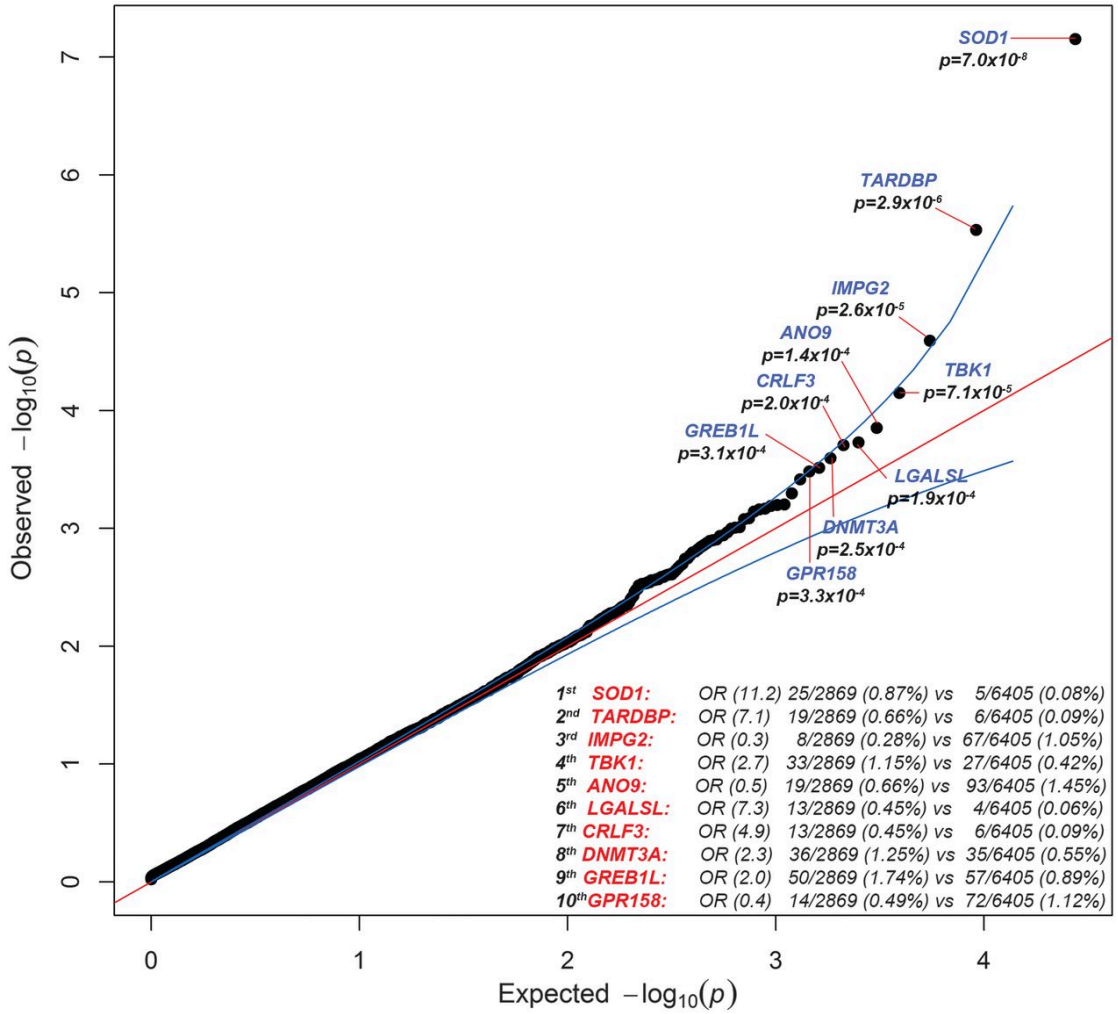


Fig. 1 Quantile-quantile plot of discovery results for dominant coding model.Results for the analysis of 2869 case and 6405 control exomes are shown; 16,491 covered genes passed quality control with more than one case or control carrier for this test



Elizabeth T. Cirulli et al. Science 2015;347:1436-1441

Sample Comparison

01/17 Petrovski Paper

- 262 IPF cases (Duke)
- 4,141 Controls

Updated Results

- 372 IPF cases (110 new CUMC cases)
- 8,168 Controls

Acknowledgements

Duke cases

Scott Palmer

CUMC cases

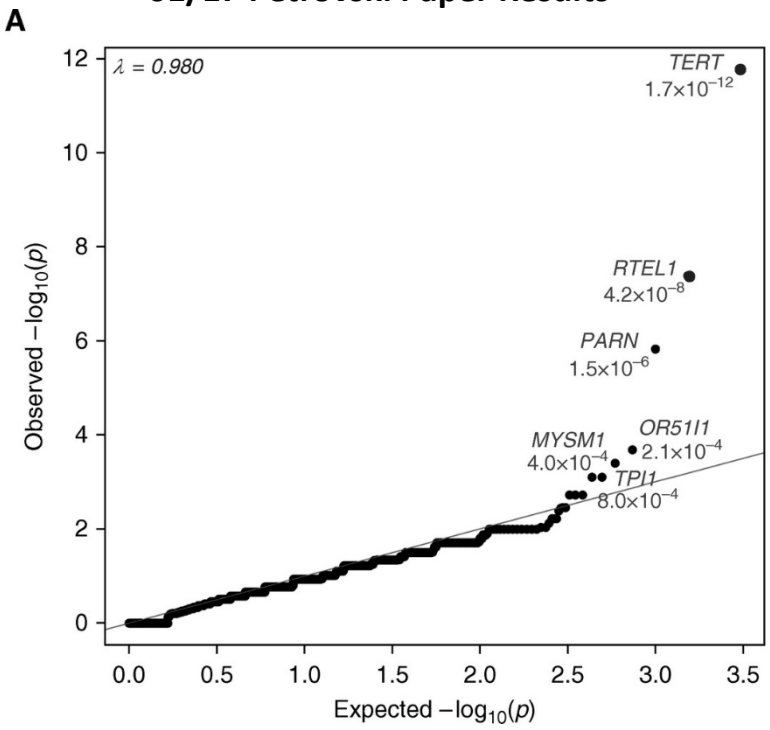
Dave Lederer

Purnema Madahar

Functional Model Comparison

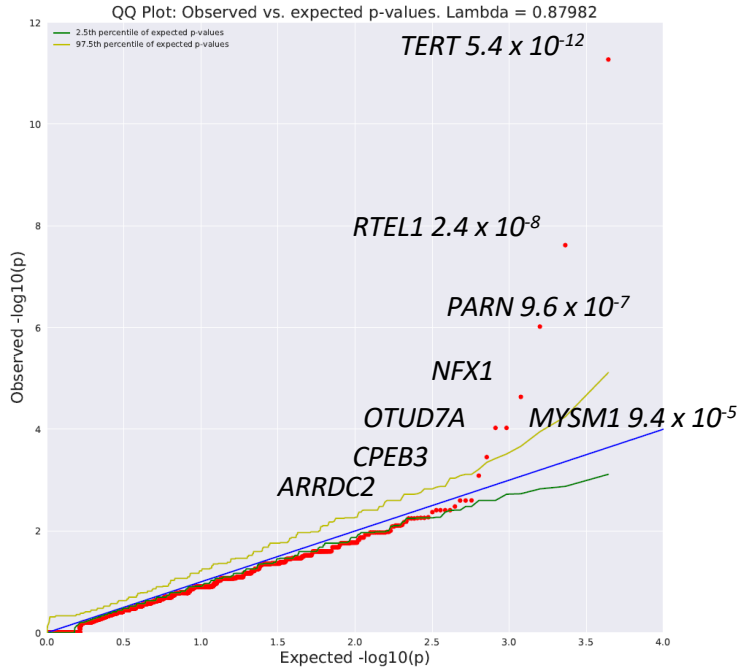
- Loo AF = 0.05%, ExAC AF = 0, EVS AF = 0
- Polyphen Humdiv probably damaging

01/17 Petrovski Paper Results



Gene0	P-Value	Qualified case freq	Qualified ctrl freq
'TERT'	1.7E-12	5.0%	0.1%
'RTEL1'	4.2E-08	2.3%	0%
'PARN'	1.5E-06	2.7%	0.1%

Updated Results



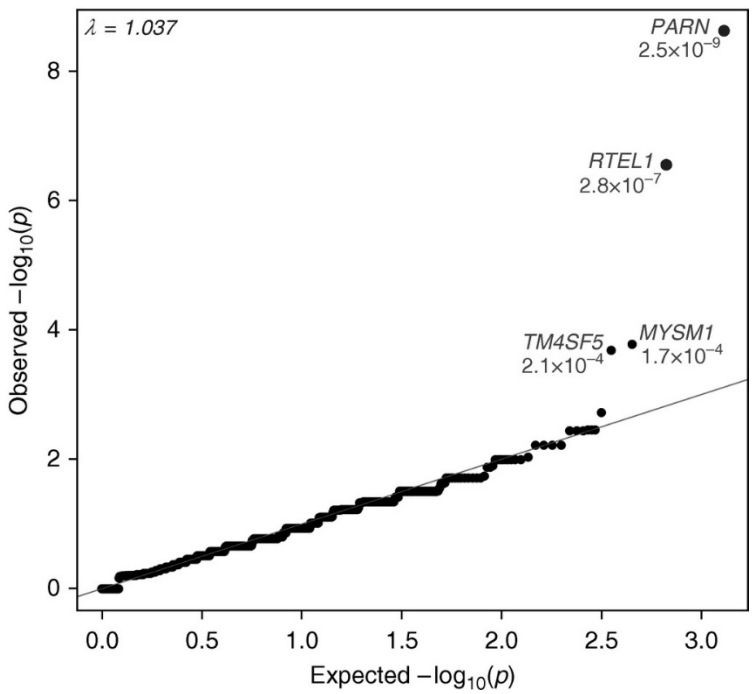
Gene	P-value	Unique Variants	Qualified case freq	Qualified ctrl freq
'TERT'	5.36E-12	27	3.76%	0.20%
'RTEL1'	2.40E-08	33	2.96%	0.23%
'PARN'	9.56E-07	16	2.15%	0.15%
'NFX1'	2.31E-05	24	2.15%	0.26%
'OTUD7A'	9.38E-05	11	1.34%	0.09%
'MYSM1'	9.38E-05	12	1.34%	0.09%
'CPEB3'	3.52E-04	20	1.61%	0.21%
'ARRDC2'	8.17E-04	13	1.34%	0.16%

Lof Model Comparison

- Loo AF = 0.1%, ExAC AF = 0.1% EVS AF = 0.1%

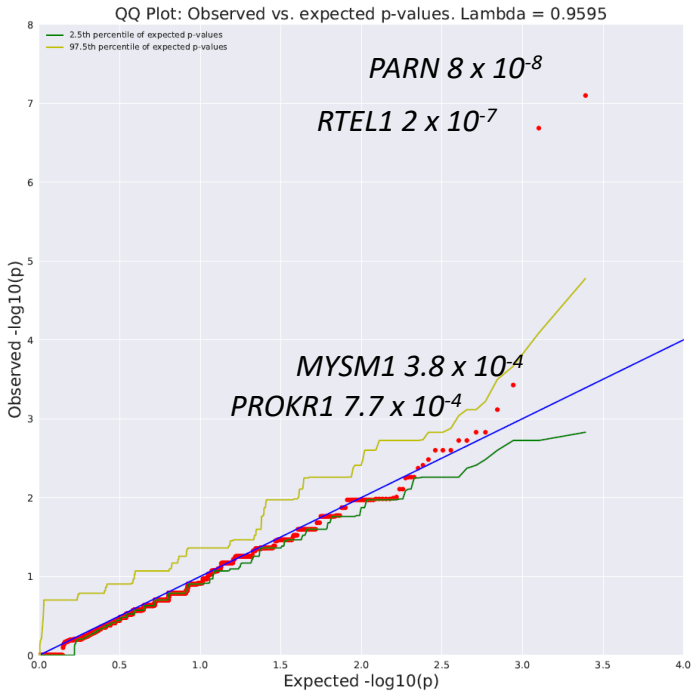
01/17 Petrovski Paper Results

B



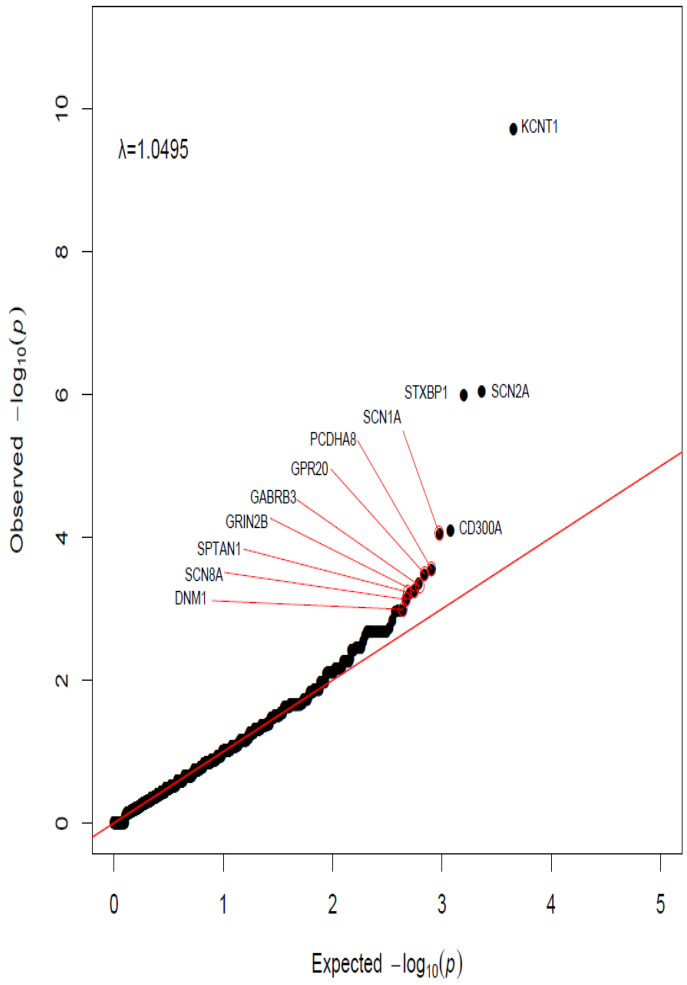
Gene	P-Value	Qualified case freq	Qualified ctrl freq
'PARN'	2.5E-09	2.7%	0%
'RTEL1'	2.8E-07	2.3%	0.02%

Updated Results



Gene	P-value	Unique Variants	Qualified case freq	Qualified ctrl freq
'PARN'	7.99E-08	6	1.93%	0.02%
'RTEL1'	2.07E-07	10	2.21%	0.06%
'MYSM1'	3.75E-04	5	1.10%	0.02%
'PROKR1'	7.68E-04	4	0.83%	0.02%

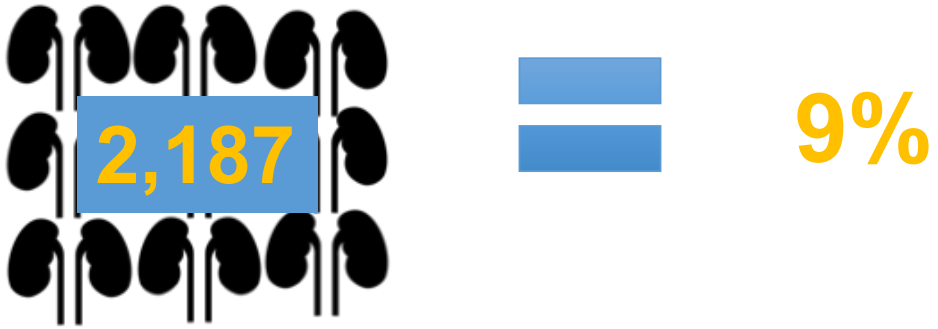
509 vs. 9866
Probably damaging missense + LoF
(IGM cases only; FET)



Top 20	Fet P
KCNT1	1.88E-10
SCN2A	8.99E-07
STXBP1	1E-06
CD300A	7.94E-05
SCN1A	9.2E-05
PCDHA8	0.000291
GPR20	0.000342
GABRB3	0.000453
GRIN2B	0.000555
SPTAN1	0.000617
SCN8A	0.000757
DNM1	0.0011
MYT1	0.0011
RASGRP3	0.0011
CUL4A	0.0011
RGS14	0.0014
LENG8	0.0018
FBXO33	0.0021
ACAP3	0.0021
GABBR2	0.0021

.

Sequencing in Kidney Diseases



65/2,187 genetic diagnosis of Alport Syndrome, only 42% were clinical recognized as having Alport Syndrome

- **51 year old Male with "CKD of unknown etiology”**
- **Causal variant in CLCN5, resulting in a genetic diagnosis of Dent disease 1**
- **Genetic diagnosis led to targeted therapy (thiazide diuretics and high citrate diet to help decrease hypercalciuria) and informed family counseling and testing of male relatives with CKD**

Sequencing in Liver Diseases

- Physician taking care of this patient: “I had a feeling that I was missing something with this kid but I didn’t know what more to do…”

What does it all mean?

- Missing heritability
- Architecture (rare and common variation not part of a continuum?)
- Implications for disease biology ?
- Open questions
  - What modifies the large effect mutations?
  - What is the explanation for the widespread signals throughout the genome?