Study design: analyses, samples, & technologies

Study design working group
(Jeff Barrett, Nancy Cox, Teri Manolio, Ben Voight)

NHGRI Causality Workshop, September 13, 2012

Caveats

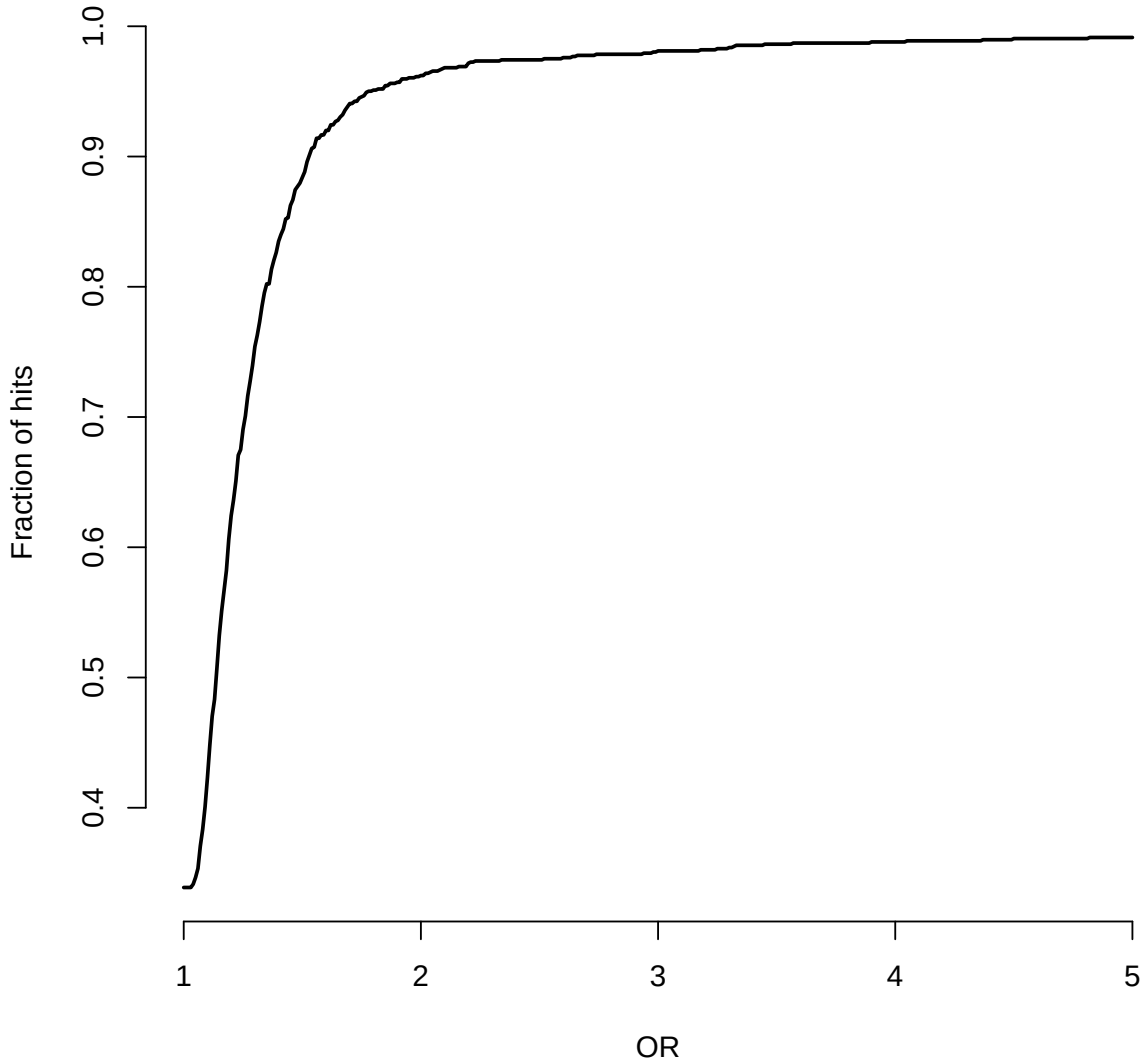
- Questions raised here pertain more to research than clinical setting.
- Focus on design of *genetic* aspect of study

1. Analyses

2. Samples

3. Technologies

Use the existing body of knowledge



NHGRI GWAS catalog

Use the existing body of knowledge II

LETTER

doi:10.1038/nature11011

Patterns and rates of exonic *de novo* mutations in autism spectrum disorders

Benjamin M. Neale[1,2], Yan Kou[3,4], Li Liu[5], Avi Ma'ayan[3], Kaitlin E. Samocha[1,2], Aniko Sabo[6], Chiao-Feng Lin[7], Christine Stevens[2], Li-San Wang[7], Vladimir Makarov[4,8], Paz Polak[2,9], Seungtai Yoon[4,8], Jared Maguire[2], Emily L. Crawford[10], Nicholas G. Campbell[10], Evan T. Geller[7], Otto Valladares[7], Chad Schafer[5], Han Liu[11], Tuo Zhao[11], Guiqing Cai[4,8], Jayon Lihm[4,8], Ruth Dannenfelser[3], Omar Jabado[12], Zuleyma Peralta[12], Uma Nagaswamy[6], Donna Muzny[6], Jeffrey G. Reid[6], Irene Newsham[6], Yuanqing Wu[6], Lora Lewis[6], Yi Han[6], Benjamin F. Voight[2,13], Elaine Lim[1,2], Elizabeth Rossin[1,2], Andrew Kirby[1,2], Jason Flannick[2], Menachem Fromer[1,2], Khalid Shakir[2], Tim Fennell[2], Kiran Garimella[2], Eric Banks[2], Ryan Poplin[2], Stacey Gabriel[2], Mark DePristo[2], Jack R. Wimbish[14], Braden E. Boone[14], Shawn E. Levy[14], Catalina Betancur[15], Shamil Sunyaev[2,9], Eric Boerwinkle[6,16], Joseph D. Buxbaum[4,8,12,17], Edwin H. Cook Jr[18], Bernie Devlin[19], Richard A. Gibbs[6], Kathryn Roeder[5], Gerard D. Schellenberg[7], James S. Sutcliffe[10] & Mark J. Daly[1,2]

LETTER

d

De novo mutations revealed by whole-exom sequencing are strongly associated with aut

Stephan J. Sanders[1], Michael T. Murtha[1], Abha R. Gupta[2]*, John D. Murdoch[1]*, Melanie J. Raubeson[1]*, A. Je A. Gulhan Ercan-Sencicek[1]*, Nicholas M. DiLullo[1]*, Neelroop N. Parikshak[3], Jason L. Stein[3], Michael F. Walke Nicole A. Teran[1], Youeun Song[1], Paul El-Fishawy[1], Ryan C. Murtha[1], Murim Choi[4], John D. Overton[4], Rober Nicholas J. Carriero[5], Kyle A. Meyer[6], Kaya Bilguvar[7], Shrikant M. Mane[8], Nenad Sestan[6], Richard P. Lifton Kathryn Roeder[9], Daniel H. Geschwind[3], Bernie Devlin[10] & Matthew W. State[1]

LETTER

doi:10.1038/nature10989

Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations

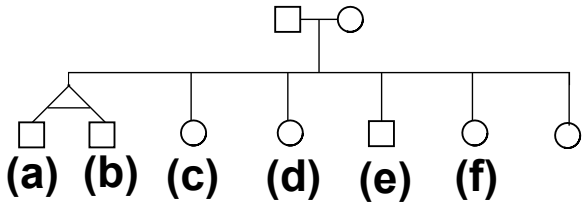
Brian J. O'Roak[1], Laura Vives[1], Santhosh Girirajan[1], Emre Karakoc[1], Niklas Krumm[1], Bradley P. Coe[1], Roie Levy[1], Arthur Ko[1], Choli Lee[1], Joshua D. Smith[1], Emily H. Turner[1], Ian B. Stanaway[1], Benjamin Vernot[1], Maika Malig[1], Carl Baker[1], Beau Reilly[2], Joshua M. Akey[1], Elhanan Borenstein[1,3,4], Mark J. Rieder[1], Deborah A. Nickerson[1], Raphael Bernier[2], Jay Shendure[1] & Evan E. Eichler[1,5]

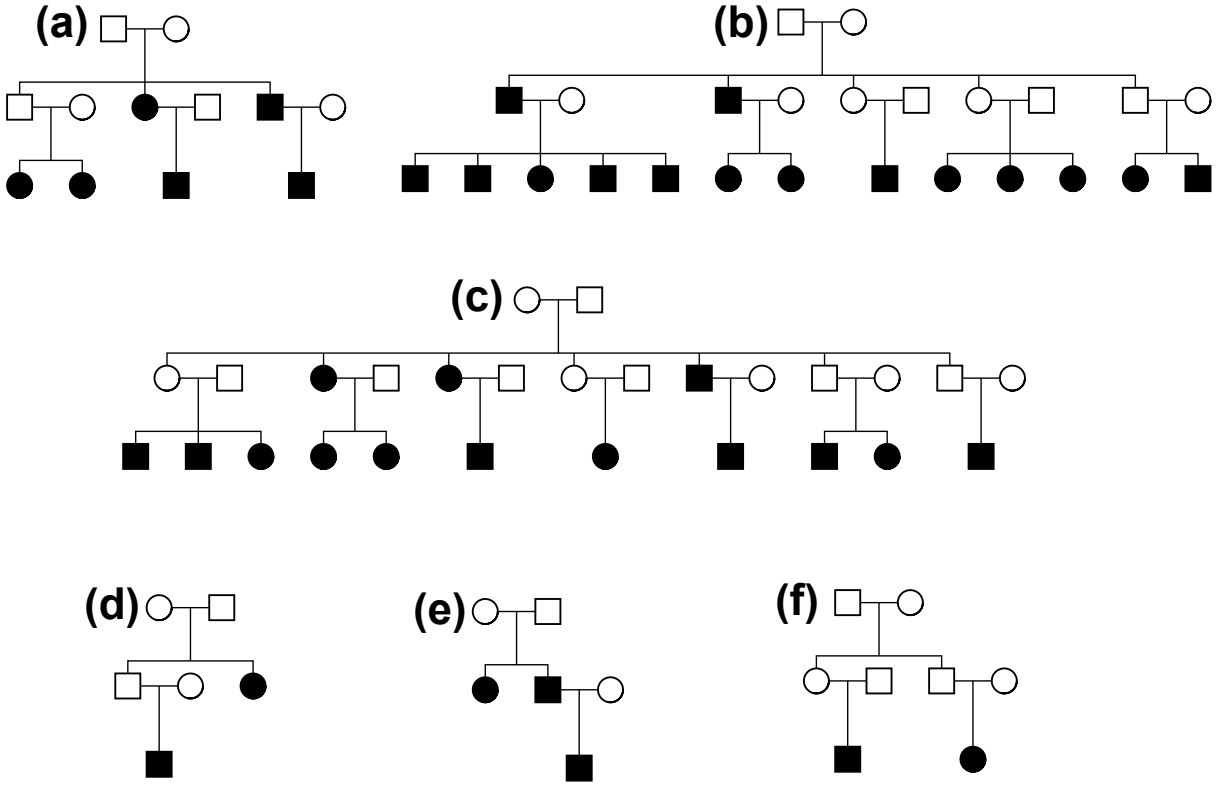
Neuron
Article

De Novo Gene Disruptions in Children on the Autistic Spectrum

Ivan Iossifov,[1,6] Michael Ronemus,[1,6] Dan Levy,[1] Zihua Wang,[1] Inessa Hakker,[1] Julie Rosenbaum,[1] Boris Yamrom,[1] Yoon-ha Lee,[1] Giuseppe Narzisi,[1] Anthony Leotta,[1] Jude Kendall,[1] Ewa Grabowska,[1] Beicong Ma,[1] Steven Marks,[1] Linda Rodgers,[1] Asya Stepansky,[1] Jennifer Troge,[1] Peter Andrews,[1] Mitchell Bekritsky,[1] Kith Pradhan,[1] Elena Ghiban,[1] Melissa Kramer,[1] Jennifer Parla,[1] Ryan Demeter,[2] Lucinda L. Fulton,[2] Robert S. Fulton,[2] Vincent J. Magrini,[2] Kenny Ye,[3] Jennifer C. Darnell,[4] Robert B. Darnell,[4,5] Elaine R. Mardis,[2] Richard K. Wilson,[2] Michael C. Schatz,[1] W. Richard McCombie,[1] and Michael Wigler[1,*]

Assumptions: Mendelian forms of complex disease?

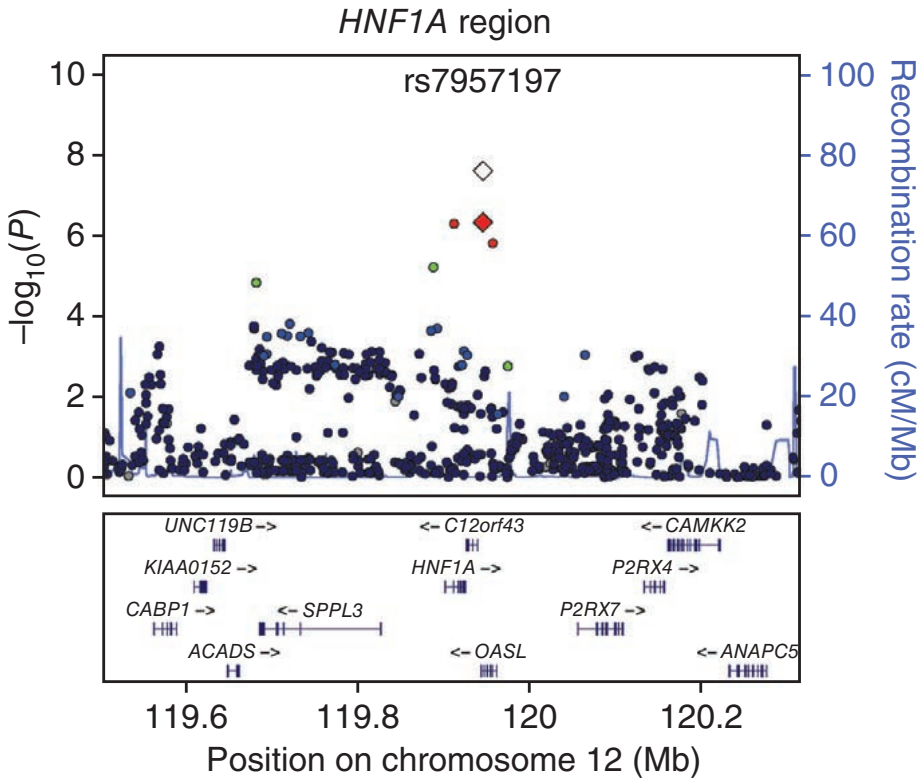




Do we care which *variant* is causal?

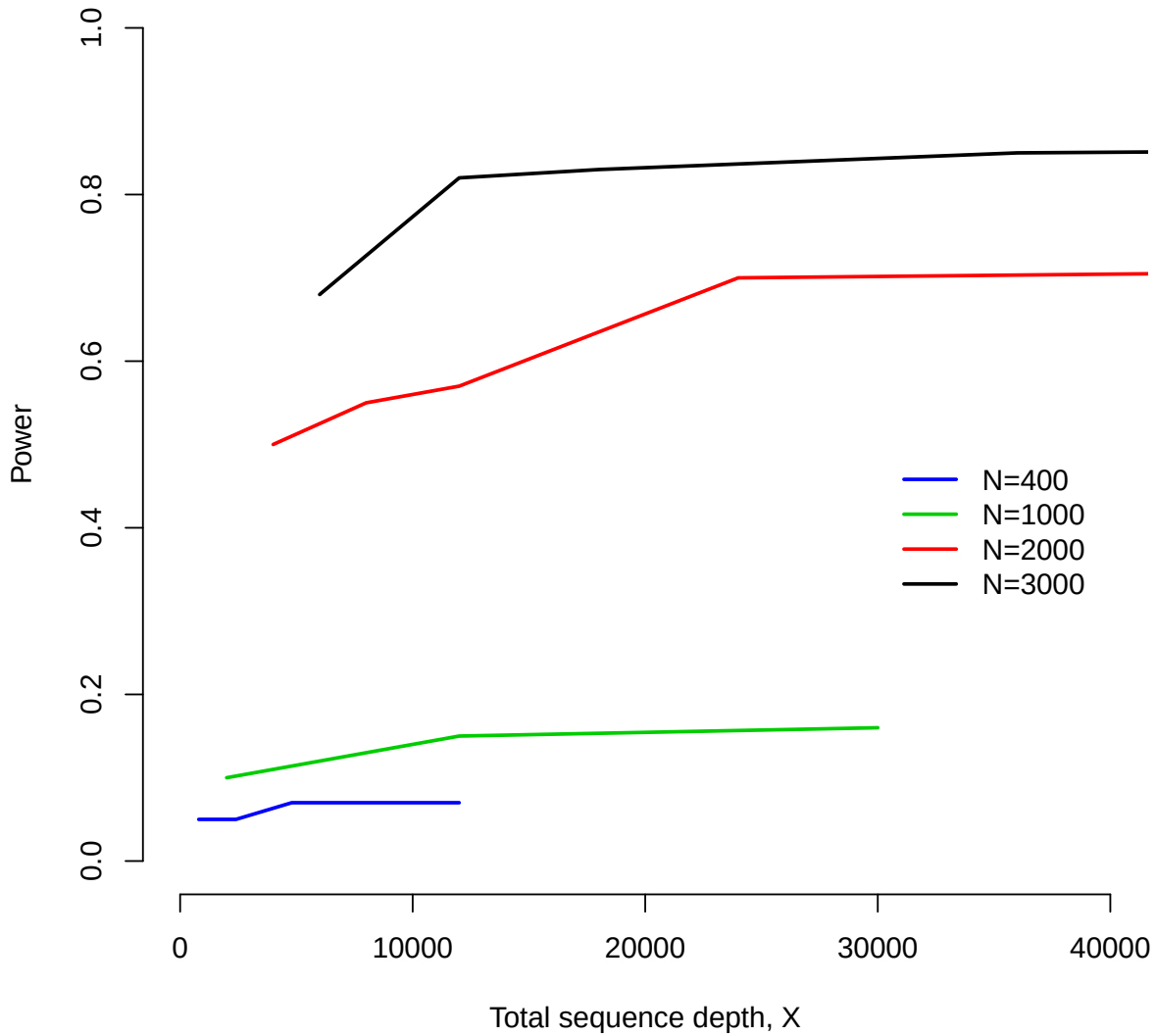
If we know the relevant gene, as well as the mode of biological action (e.g. reducing function of the gene increases risk of disease), do we necessarily care what variant is responsible?

Do we care which *variant* is causal?



"At some loci, particularly those near *HNF1A*, *HMGA2* and *KLF14*, existing biology, coupled with phenotypic and expression data presented here, highlight the named genes as prime candidates for mediating the susceptibility effect." (Voight *et al. Nat Genet.* 2010)

Sample size is still king



Adapted from Li *et al. Genome Research*, 2011

Larger samples in Mendelian sequencing also useful

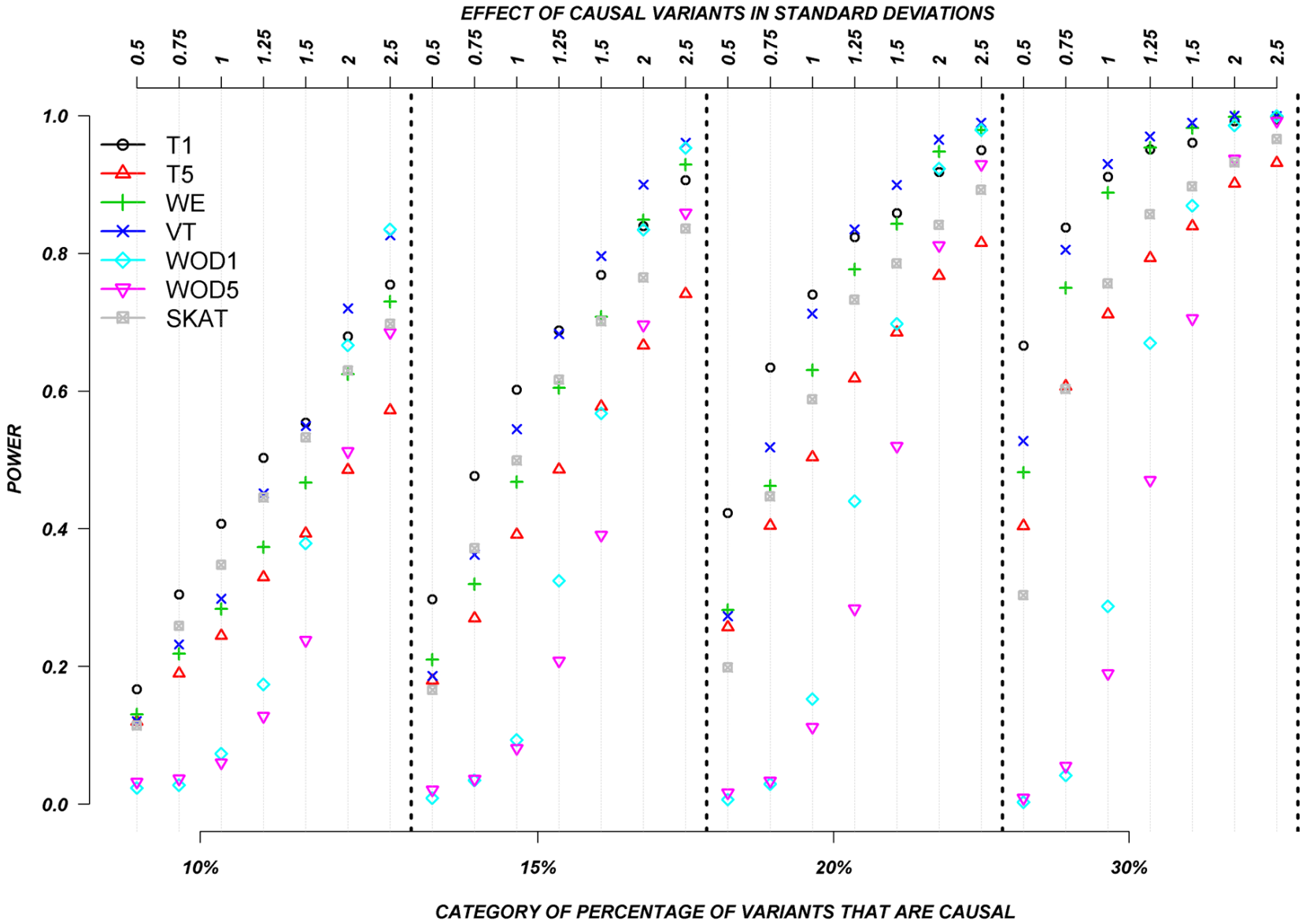
# genes in which each affected has at least one...	FSS24895	FSS24895 FSS10208	FSS24895 FSS10208 FSS10066	FSS24895 FSS10208 FSS10066 FSS22194	ANY 3 OF 4 FSS24895 FSS10208 FSS10066 FSS22194
nonsynonymous cSNP, splice site variant or coding indel (NS/SS/I)	4,510	3,284	2,765	2,479	3,768
NS/SS/I not in dbSNP	513	128	71	53	119
NS/SS/I not in 8 HapMap exomes	799	168	53	21	160
NS/SS/I neither in dbSNP nor 8 HapMap exomes	360	38	8	1 (MYH3)	22
... AND predicted to be damaging	160	10	2	1 (MYH3)	3

Ng *et al. Nature*, 2009

Power should govern study design

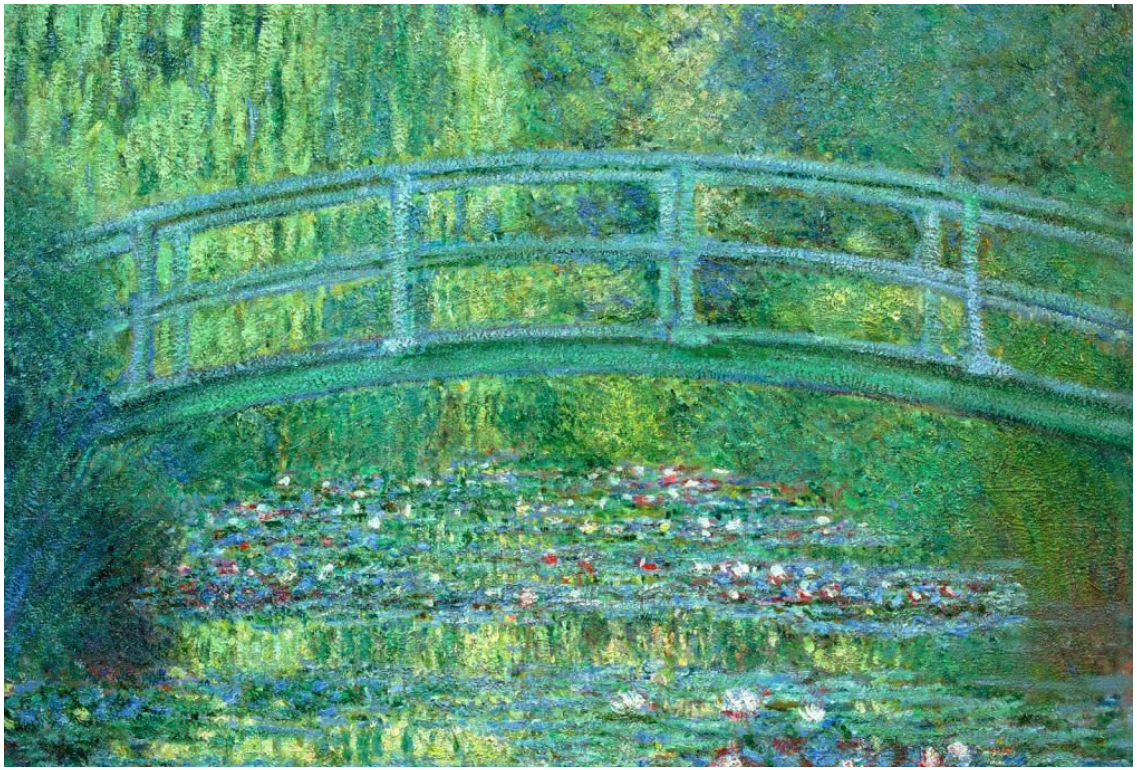
$$\mathbb{E}[\chi^2] \propto N\gamma^2 p(1-p)r^2$$

Power should govern study design



Mendelian sequencing studies need a minimum sample size

Mendelian sequencing studies need a minimum sample size



Technological holes

Technological holes

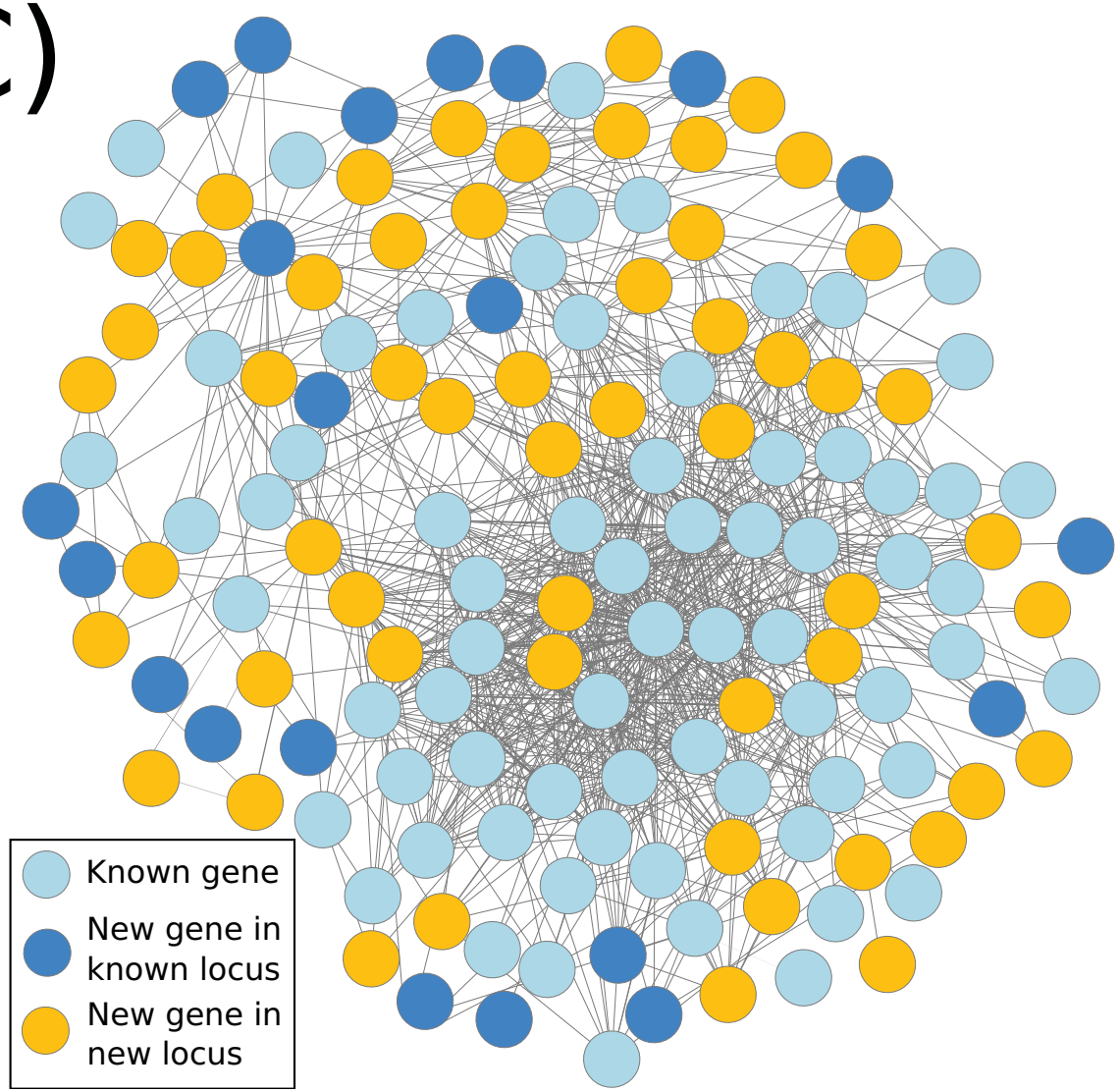
- Sequencing one gene (or a handful) in 1000s of individuals

Technological holes

- Sequencing one gene (or a handful) in 1000s of individuals
- Genotyping 10^4 – 10^6 variants in a million individuals

Balancing the role of genotyping and sequencing

C)



Exomes are obsolete