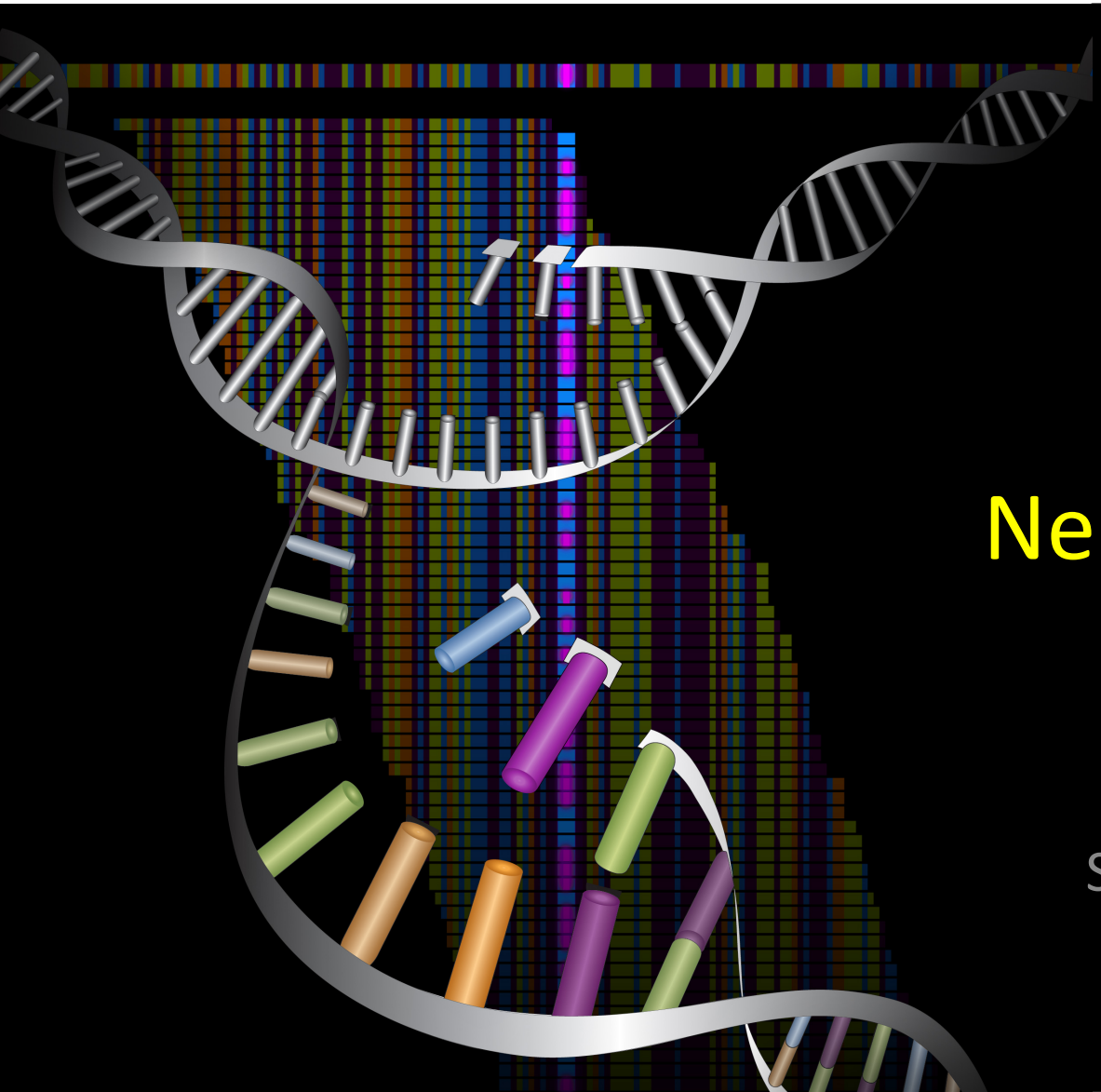
University of California
San Francisco



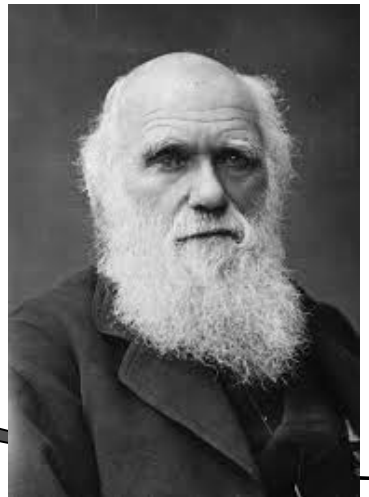
Neurogenetics II

Stephan J. Sanders
MD, PhD

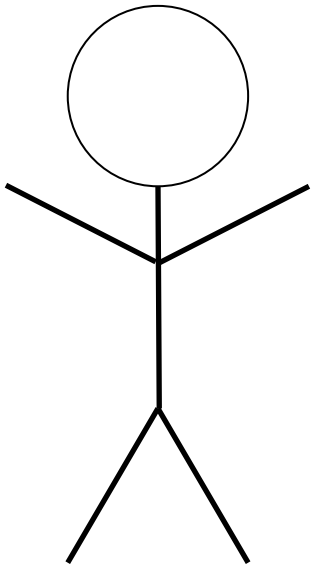
All types of genetics assess the flow of information in biological systems

Mendel's laws of inheritance

Genotype

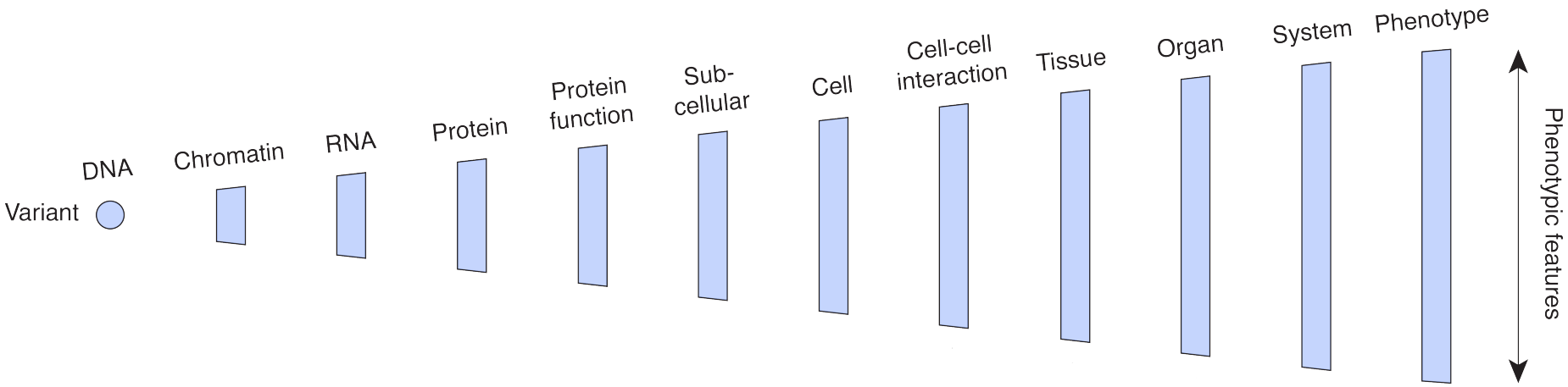


Darwin's theory of evolution by natural selection



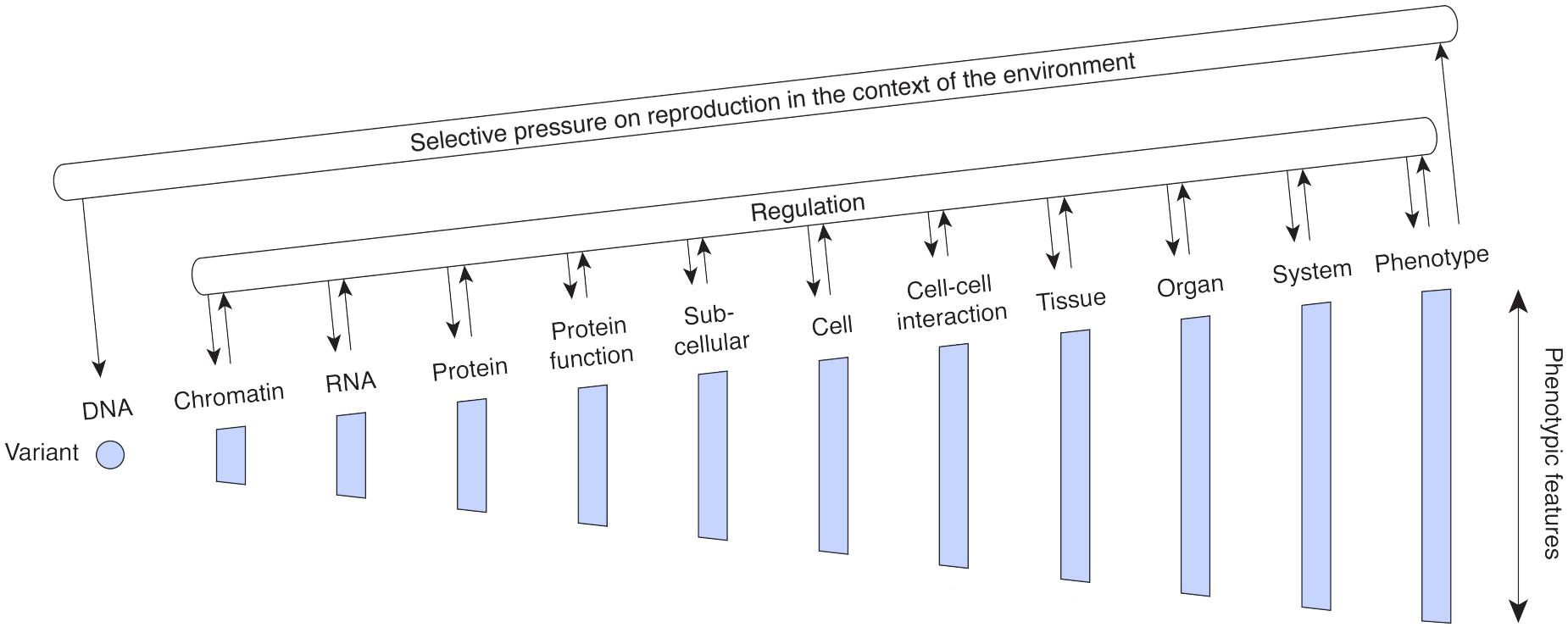
Phenotype

Genotypes are amplified to produce observable phenotypes



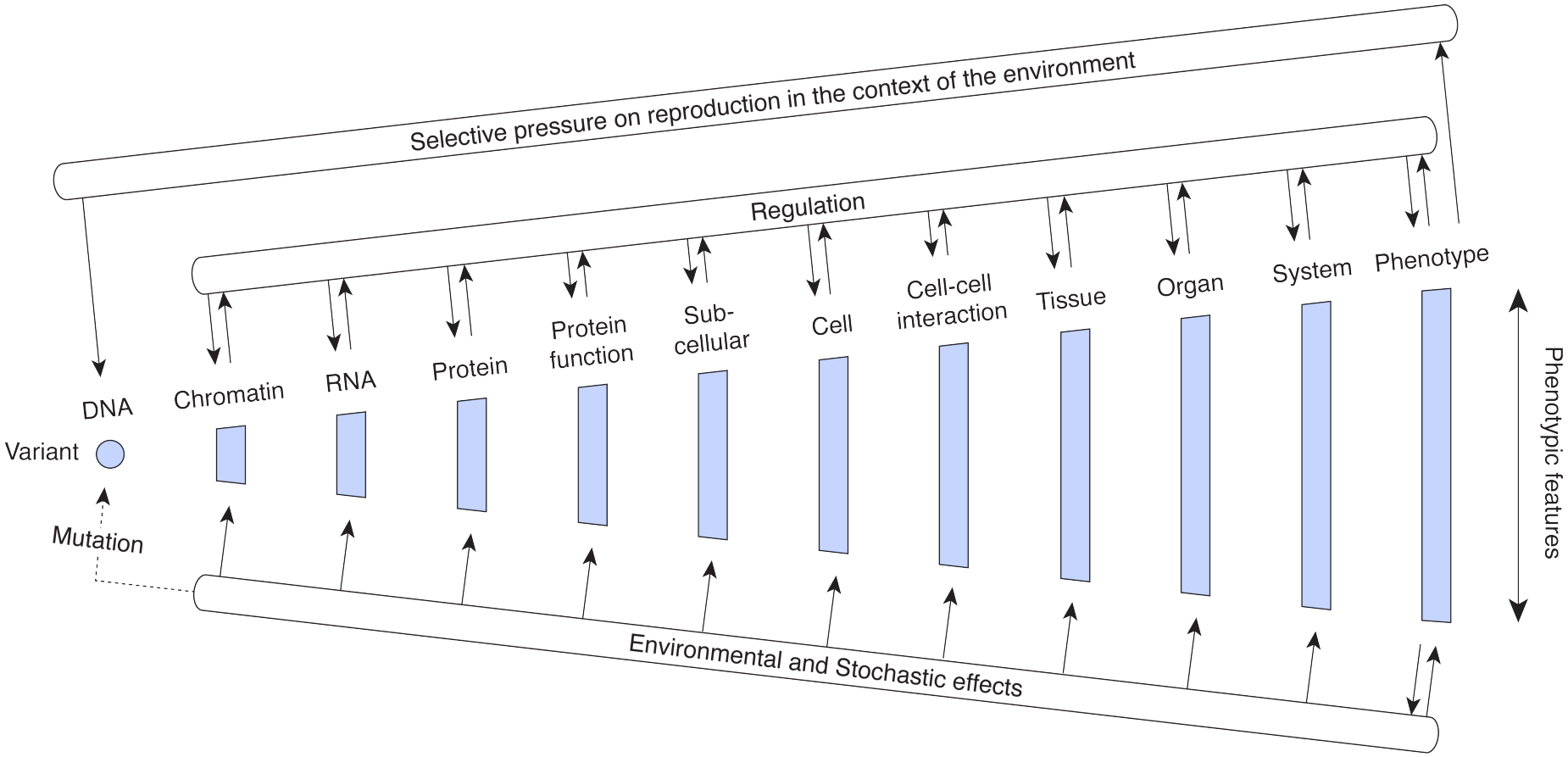
Sanders SJ, Curr Opin Genet Dev, 2015

Genotypes are amplified to produce observable phenotypes



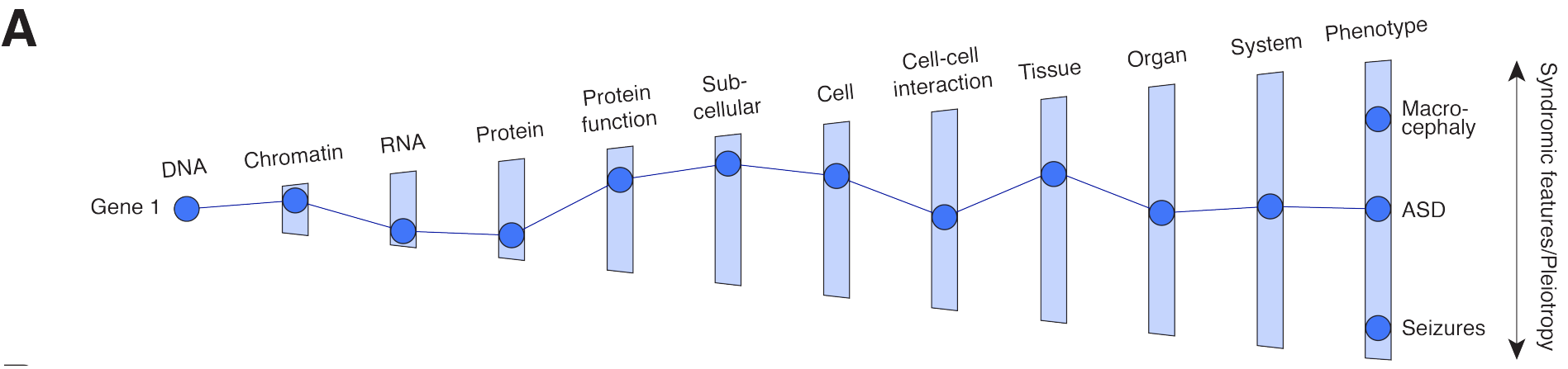
Sanders SJ, Curr Opin Genet Dev, 2015

Genotypes are amplified to produce observable phenotypes



Sanders SJ, Curr Opin Genet Dev, 2015

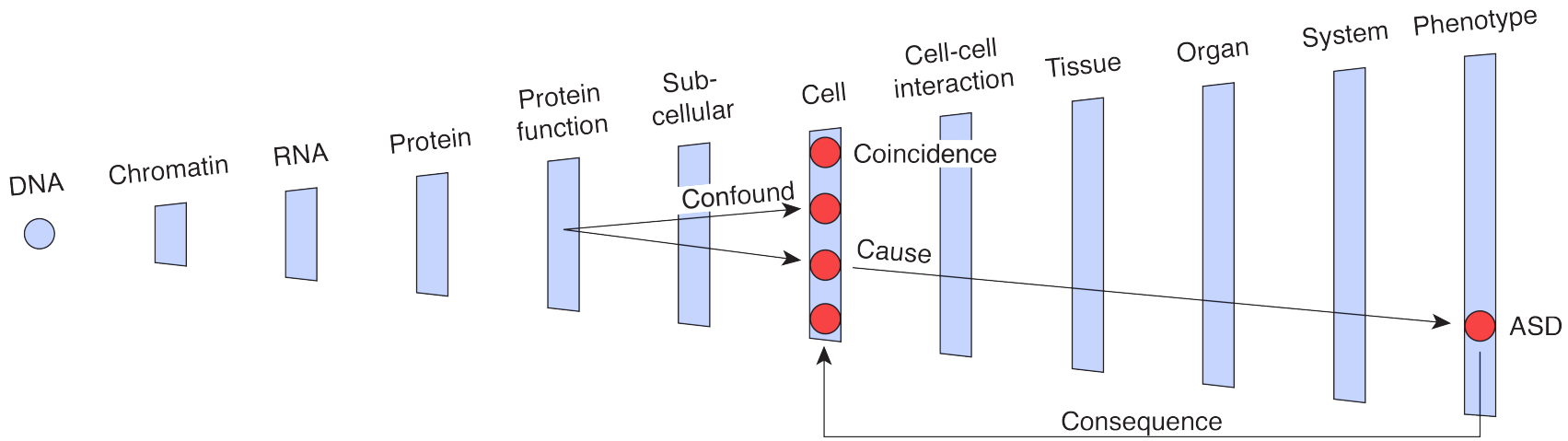
We aim to find the causal path between a genotype and phenotype



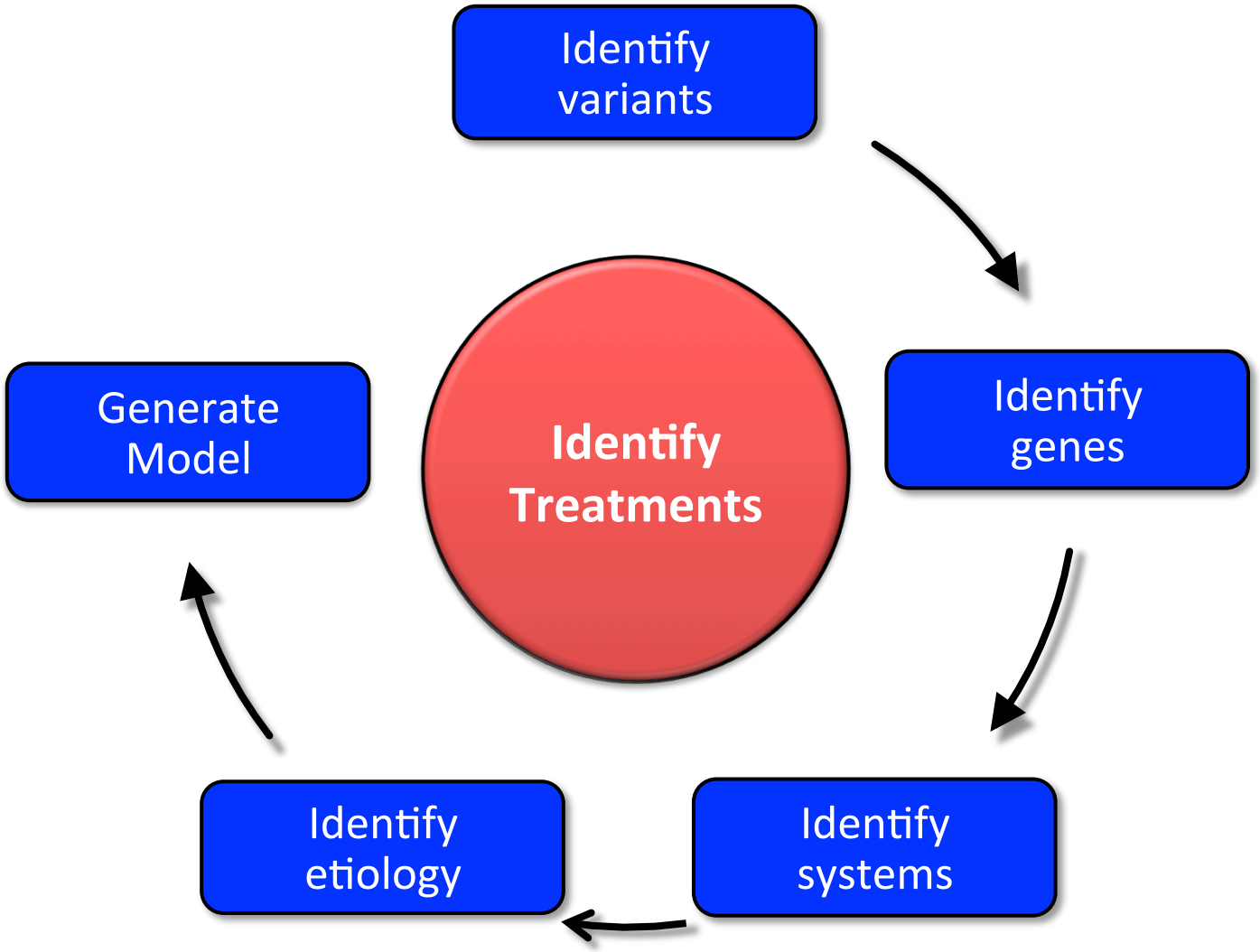
Sanders SJ, Curr Opin Genet Dev, 2015

Why consider genetics?

- DNA contains all the information necessary to build a human brain (with the right context)
- Majority of neuropsychiatric disorders are highly heritable
- DNA variation precedes disease onset
  - Genetic association must be causal



Genetics provides a path to identifying etiology and treatment

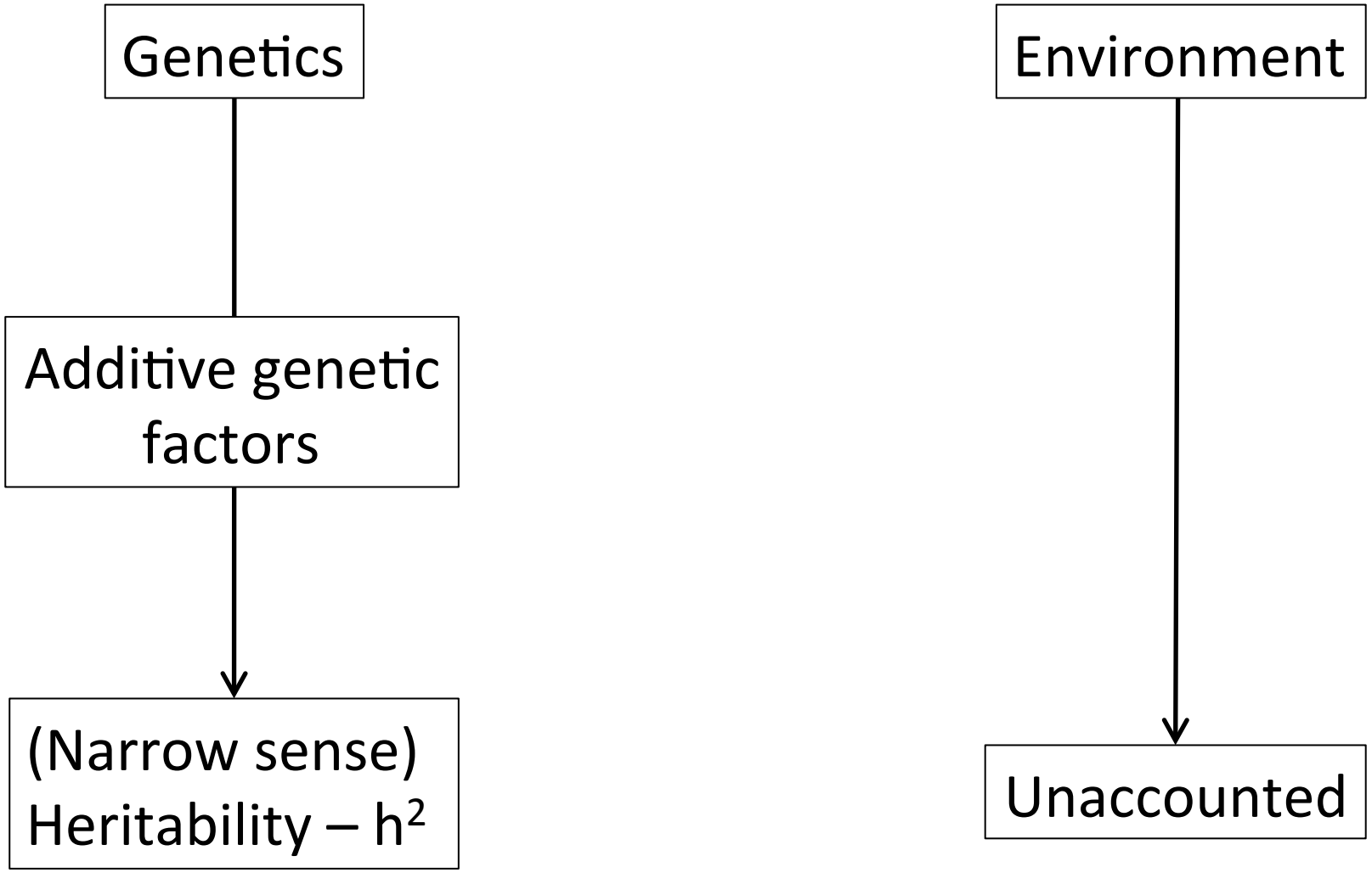


Heritability

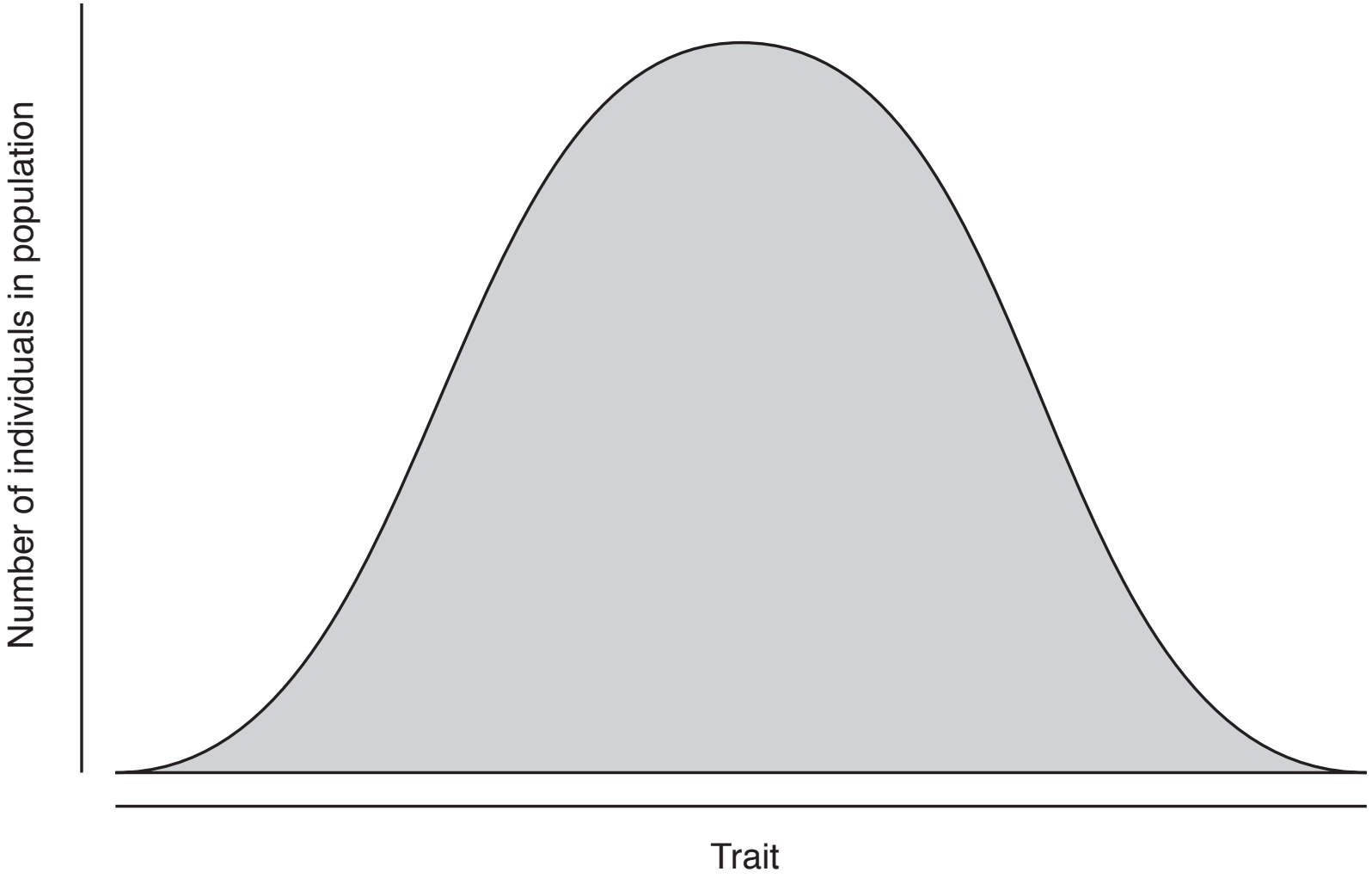
Heritability assess how much of the variance in a trait is inherited

- Twin studies:
  - Monozygotic twins vs. Dizygotic twins
  - 100% vs. 50% shared genetic factors
  - Similar degree of environmental sharing
- Family studies:
  - Sibling incidence vs. population incidence
  - 50% vs. ~0% shared genetic factors
  - Similar environmental exposure...
- SNP-based heritability:
  - Correlation of genotypes with phenotype

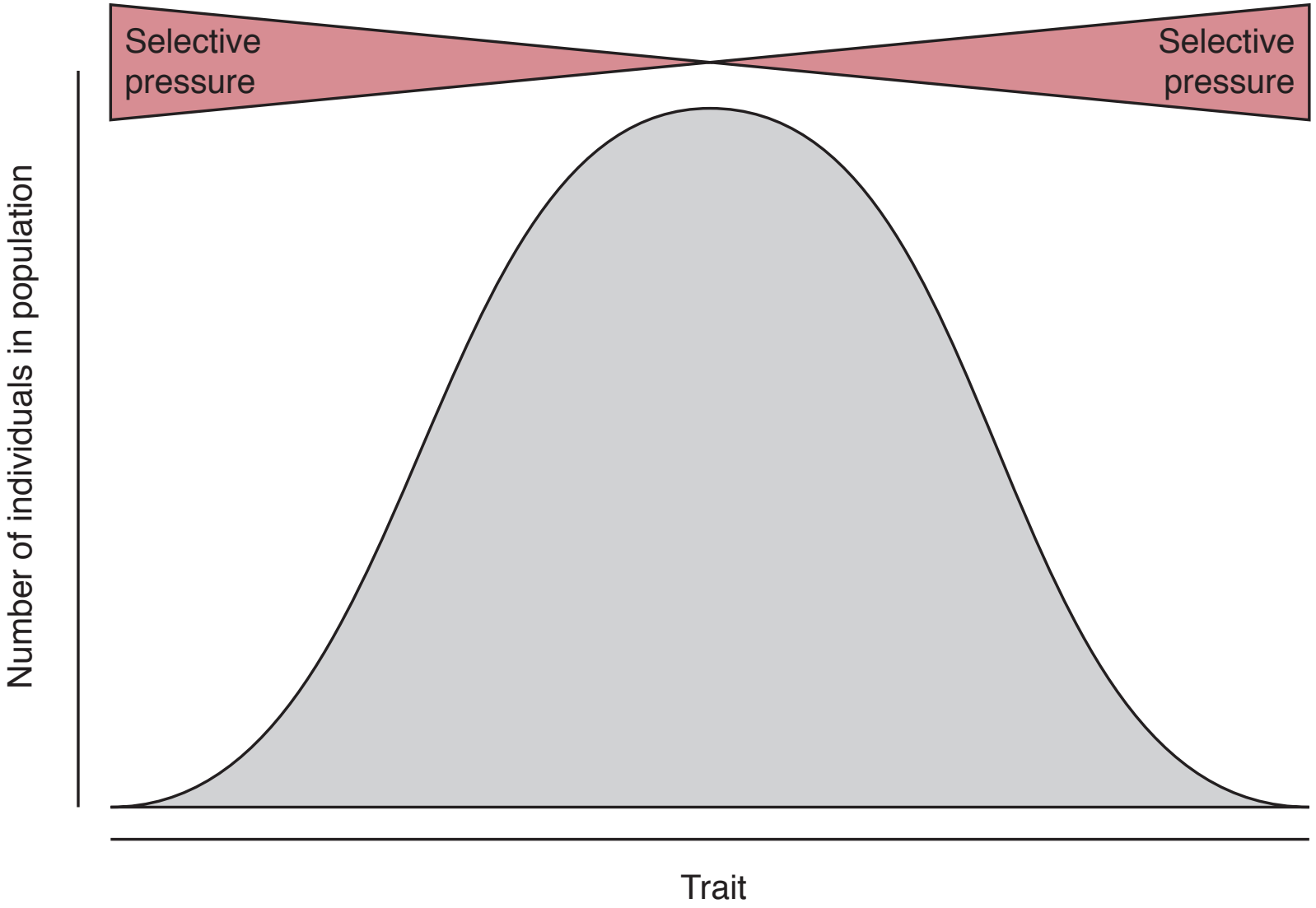
Heritability estimates the role that additive genetic factors play in a trait



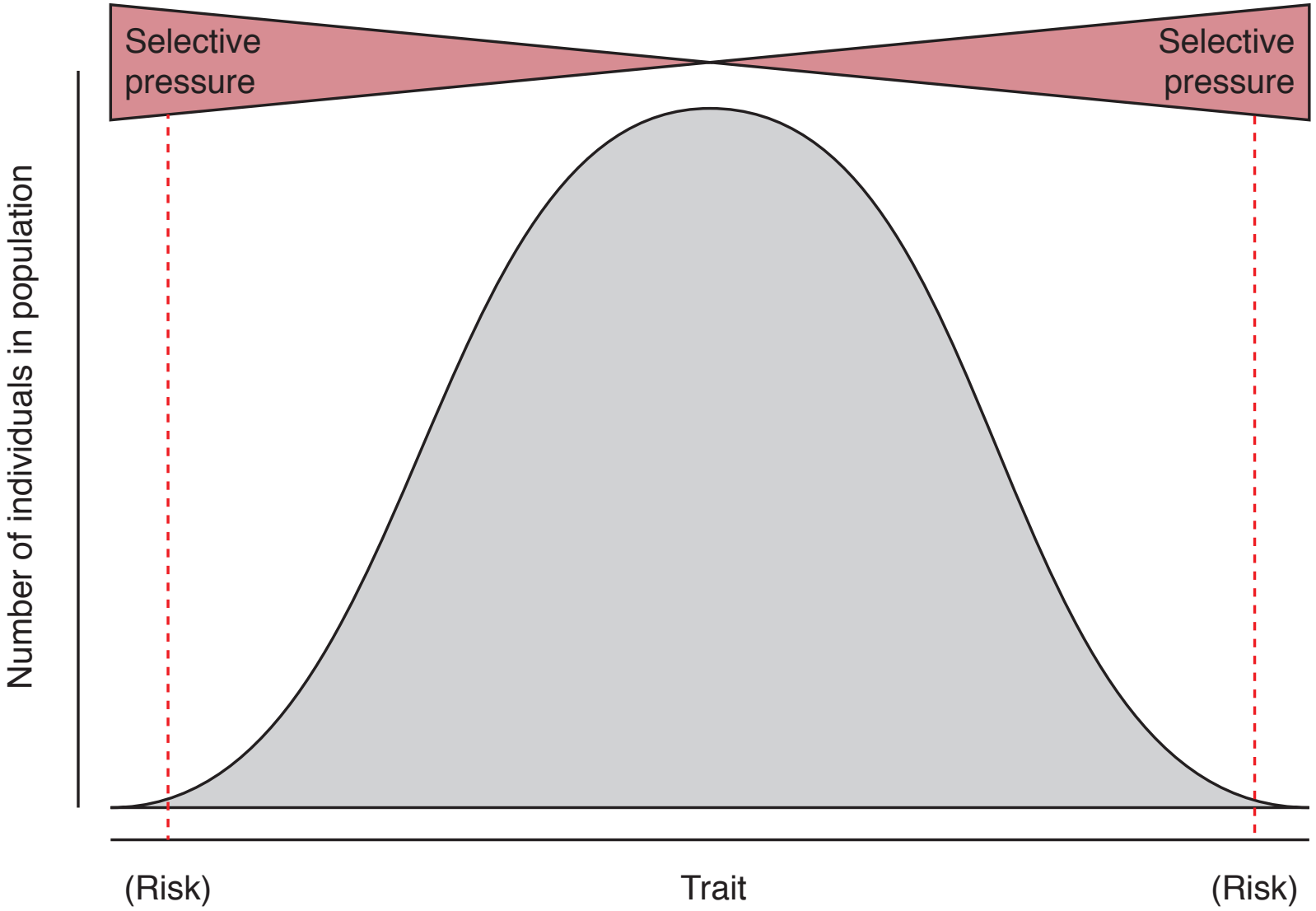
Quantitative models of human traits



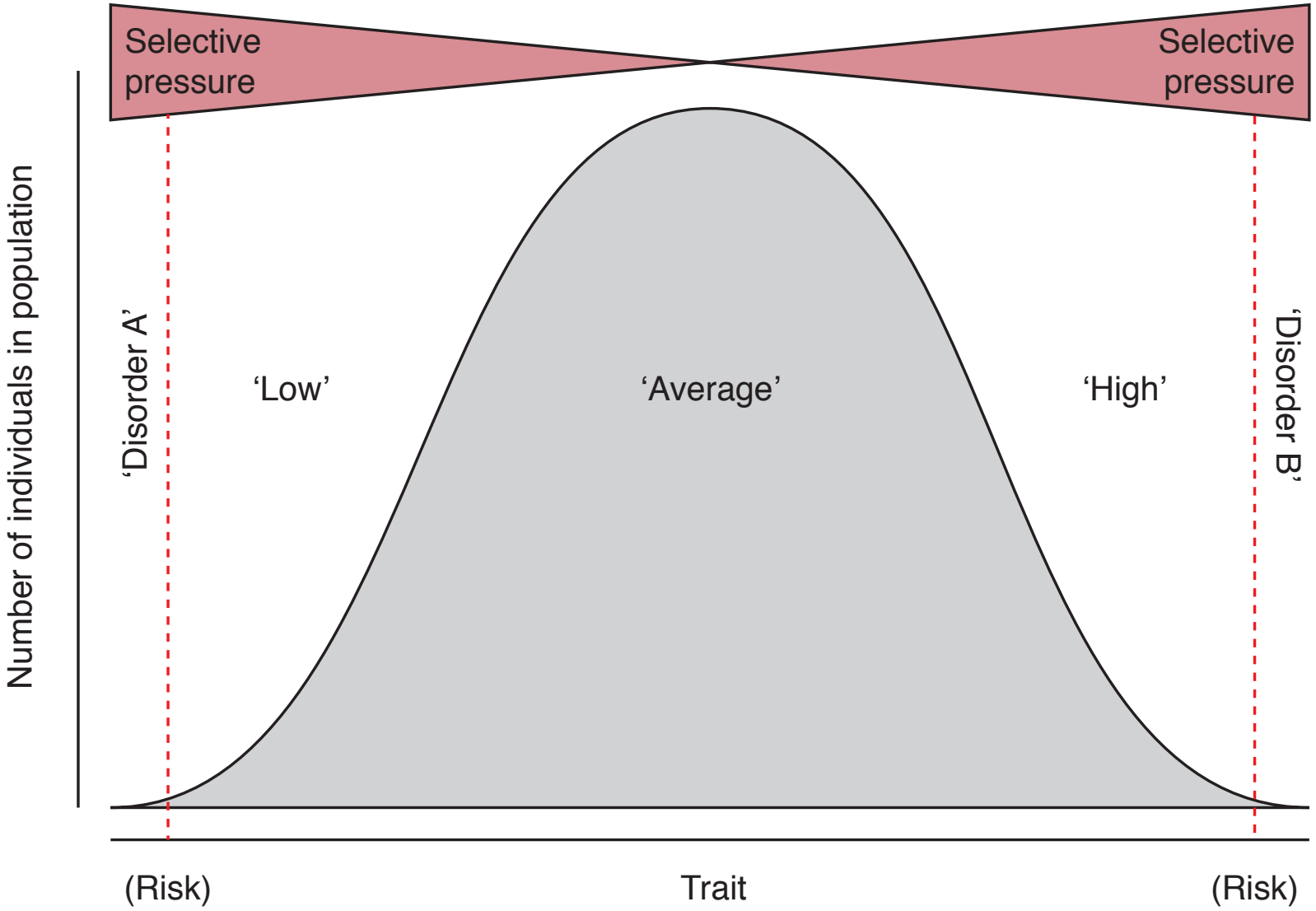
Quantitative models of human traits



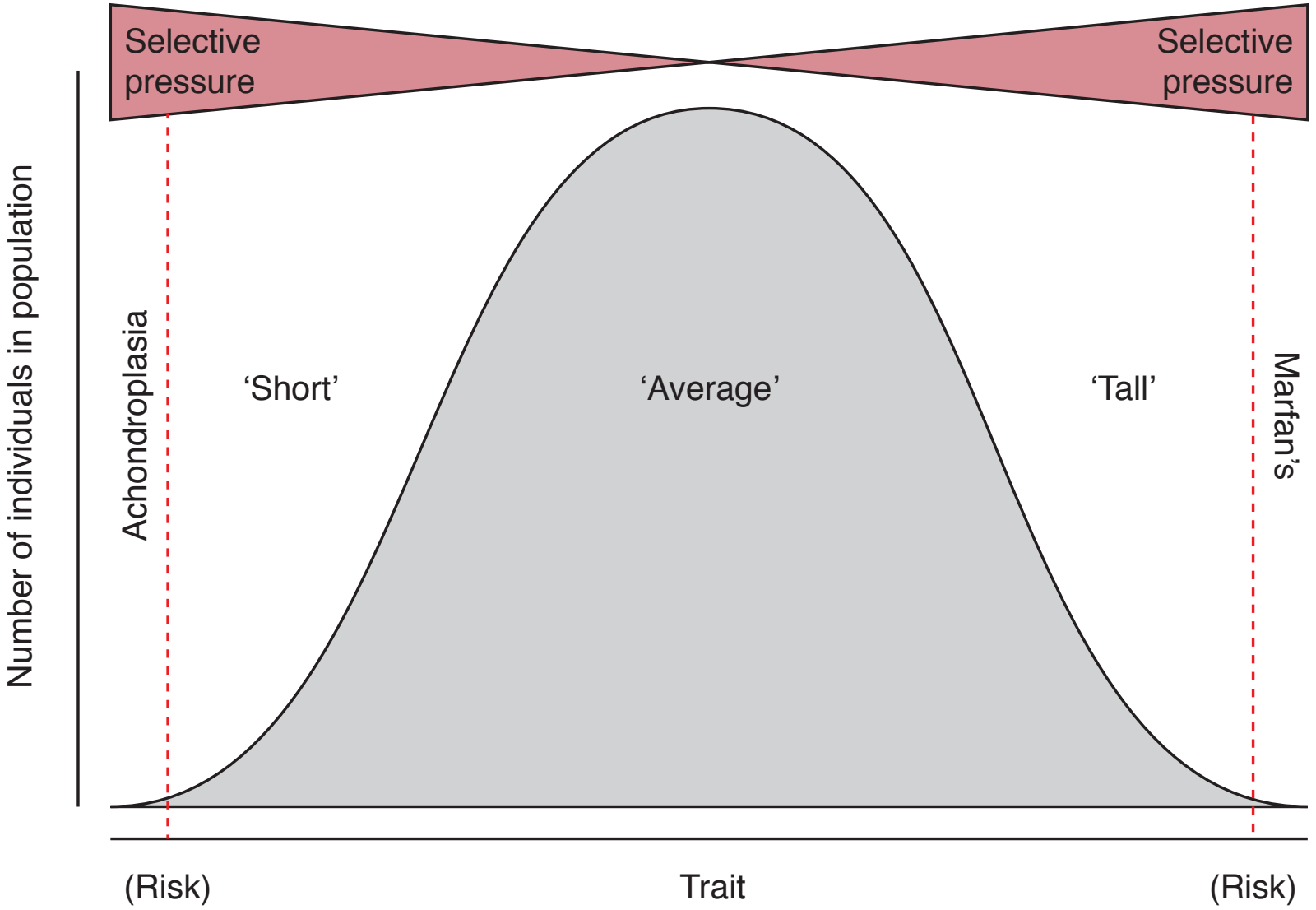
Quantitative models of human traits



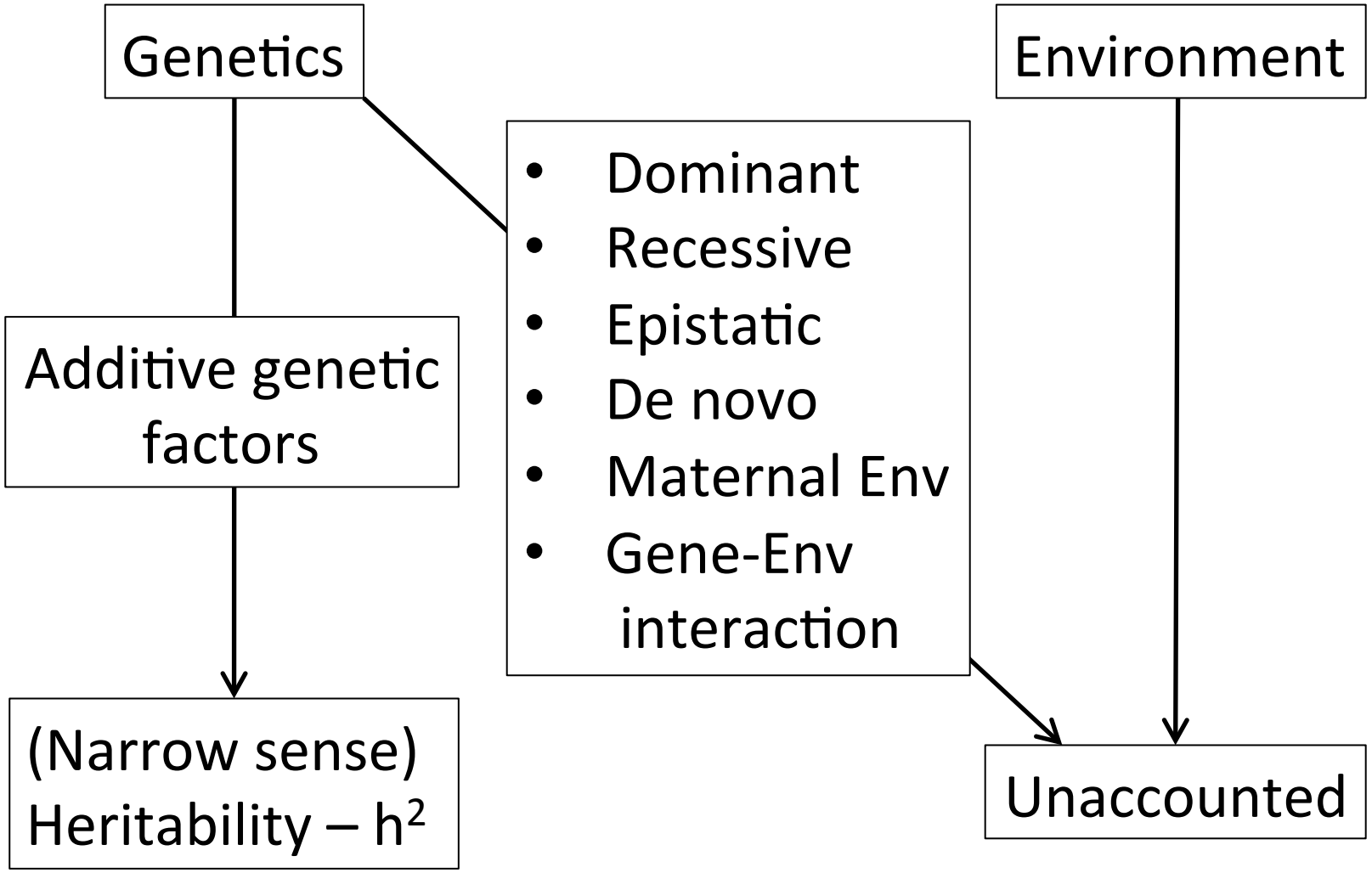
Quantitative models of human traits



Human diseases are the extremes of physiological traits



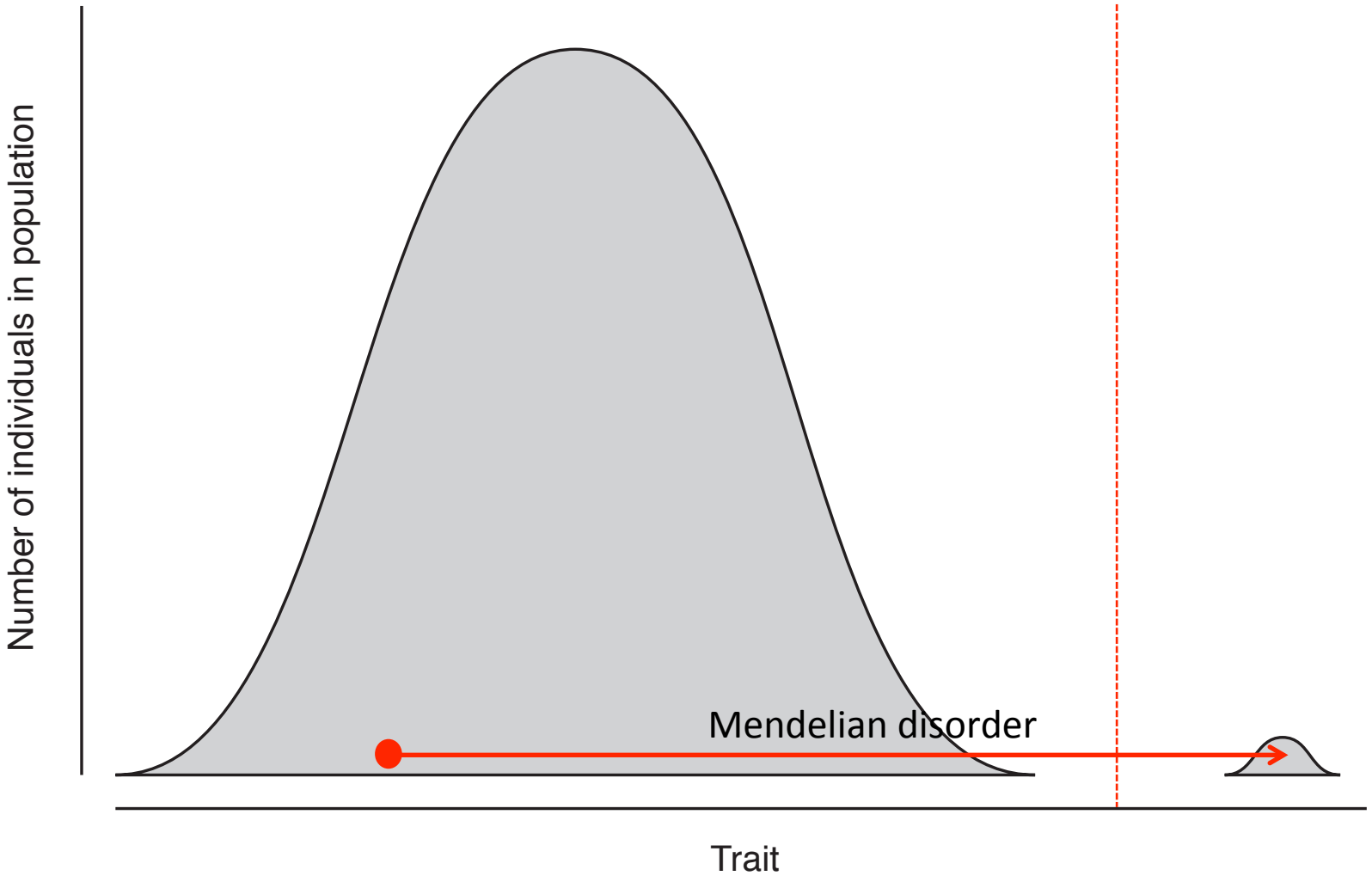
Heritability sets a lower threshold for the role of genetics in a disorder



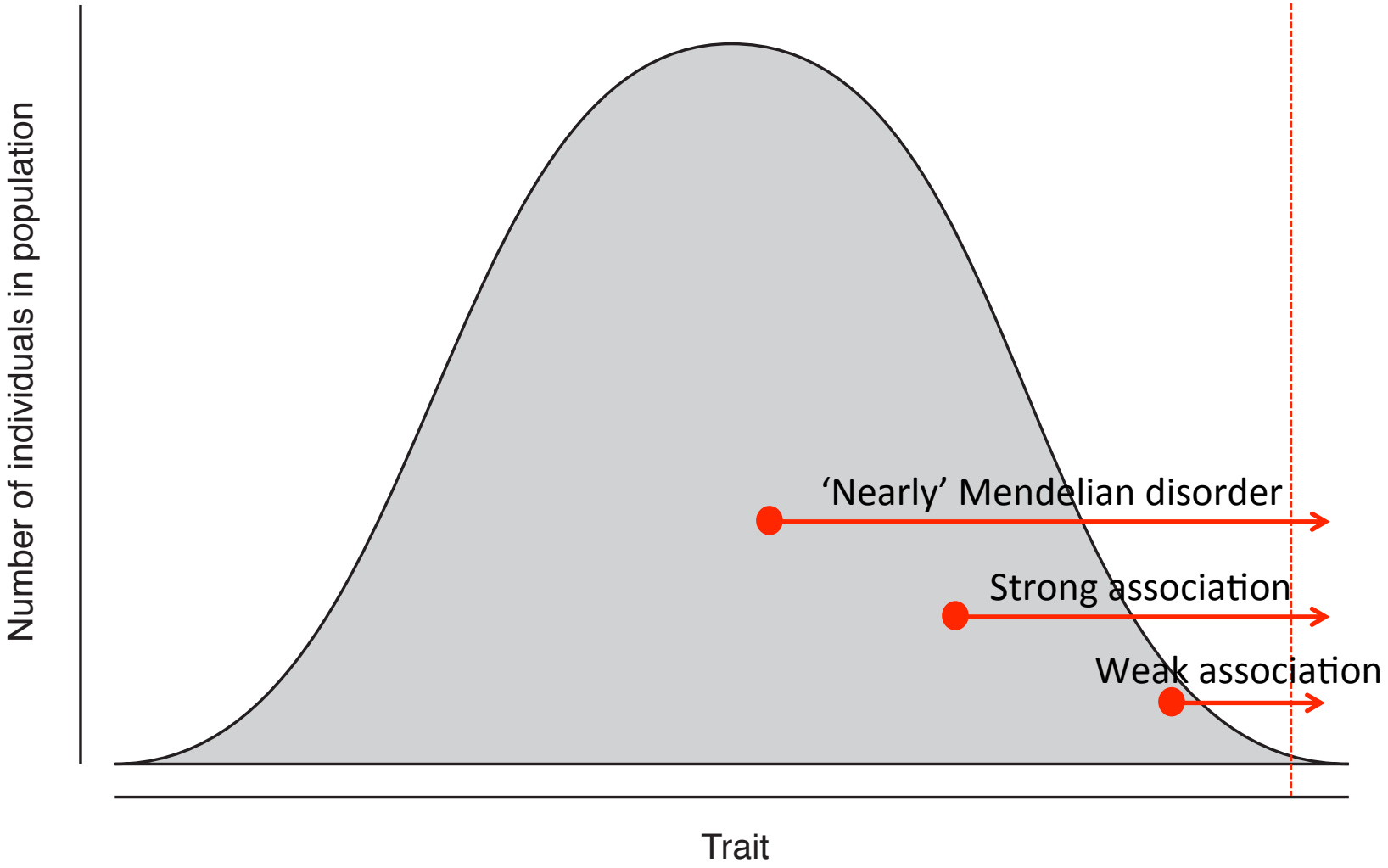
Heritability estimates of common neuropsychiatric disorders

Disorder / Trait	Heritability
IQ	50%
Autism	50-90%
Schizophrenia	60-90%
Alzheimer's	60-80%
Parkinson’s	40%
Multiple Sclerosis	64%
Fronto-temporal dementia	75-86%

The magnitude of effect of a genetic variant determines the inheritance pattern



The magnitude of effect of a genetic variant determines the inheritance pattern



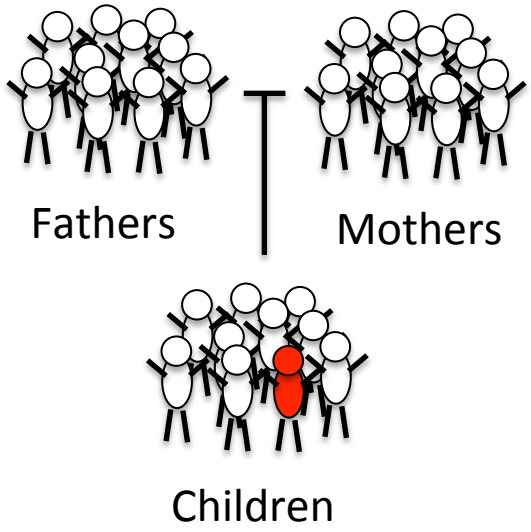
Methods of gene discovery

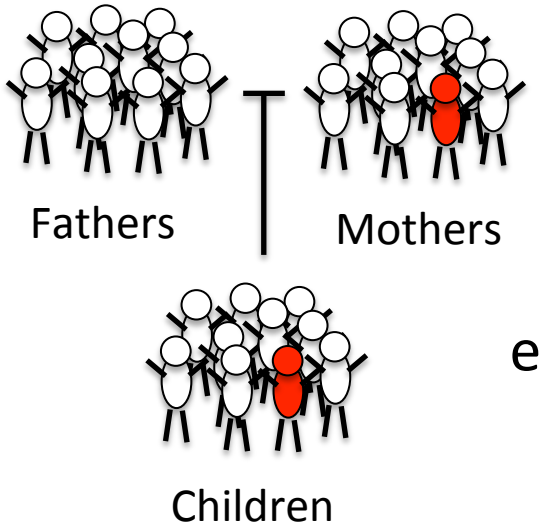
Genomic variation affects varying numbers of nucleotides

Single nucleotide variant (SNV)	Insertion/deletion (indel)	Copy number variant (CNV)
A C A C A C T	A C A C A C T	
↓	↓	↓
A C A T A C T	A C A × C T	
	or	or
	A C A C A C A C T	
0bp	1-1,000bp	>1,000bp

The pattern of inheritance is also important in describing DNA variants

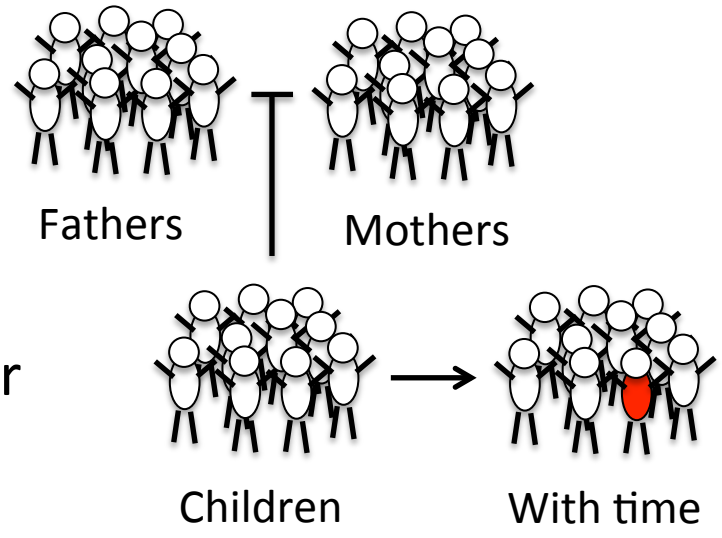
***De novo* germline mutation**
e.g. Down's syndrome

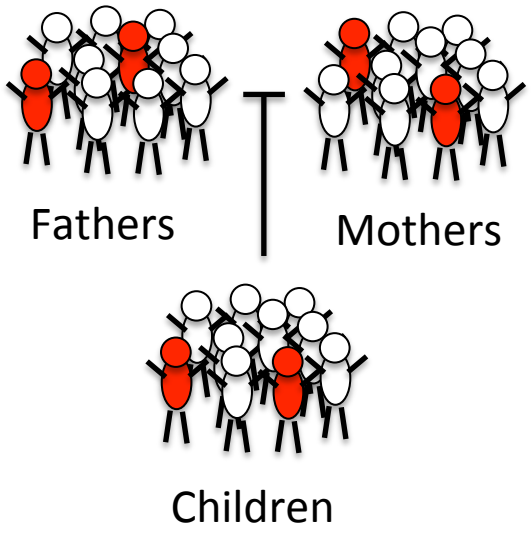




Rare inherited variant
e.g. Marfan's syndrome

***De novo* somatic mutation**
e.g. Cancer





Common inherited variant
e.g. Alzheimer's

Every individual has ~ 3.5 million genetic variants; most are common

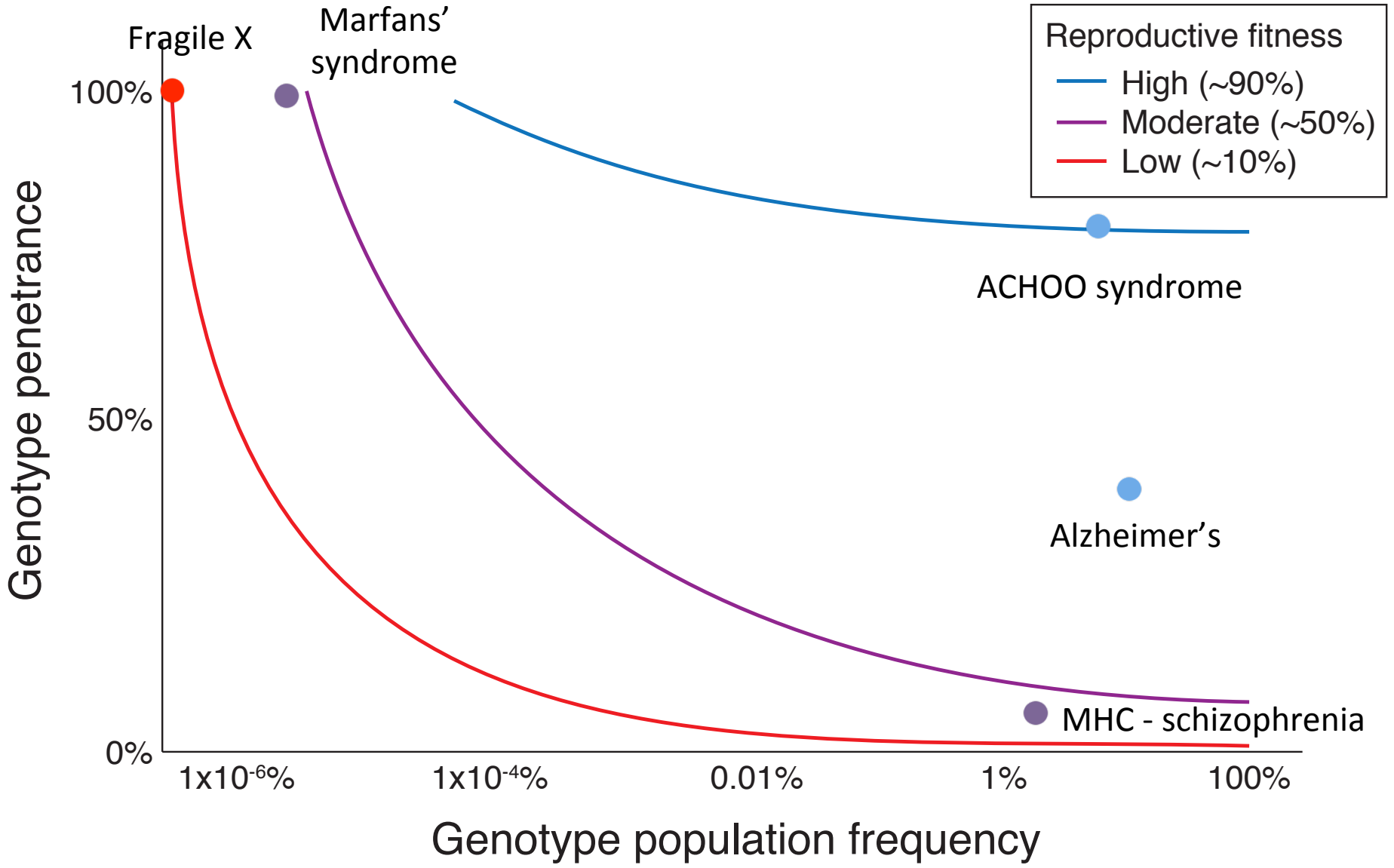
Variation in one individual

Frequency

Size

	De novo	Rare Inherited	Common Inherited
SNV 1bp	70	150,000	3 million
Indel 1-1,000bp	5	15,000	250,000
SV/CNV >1,000bp	1	150	2,000

Common variants have low effect size in disorders that reduce reproductive fitness



Targeted vs. Genome-wide

Genome-wide

- Look everywhere and see what sticks out

- Hypothesis generating
  - Can find 'new' biology
- Data can be used for multiple studies
- Correcting for multiple comparisons limits discovery

Targeted

- Look in specific places

- Hypothesis following
  - Can validate findings
- Efficient if hypothesis is correct
- Power estimate can be hard
- Fewer comparisons to correct for (?)

Genome-wide methods for finding different types of genetic variation

Frequency →

Size ↓

	De novo	Rare Inherited		Common Inherited
SNV 1bp	Exome Genome	Linkage	Exome Genome	Microarray (GWAS)
Indel 1-1,000bp	Exome Genome	Linkage	Exome Genome	Microarray (GWAS)
SV/CNV >1,000bp	Karyotype Microarray Genome	Linkage	Karyotype Microarray Genome	???

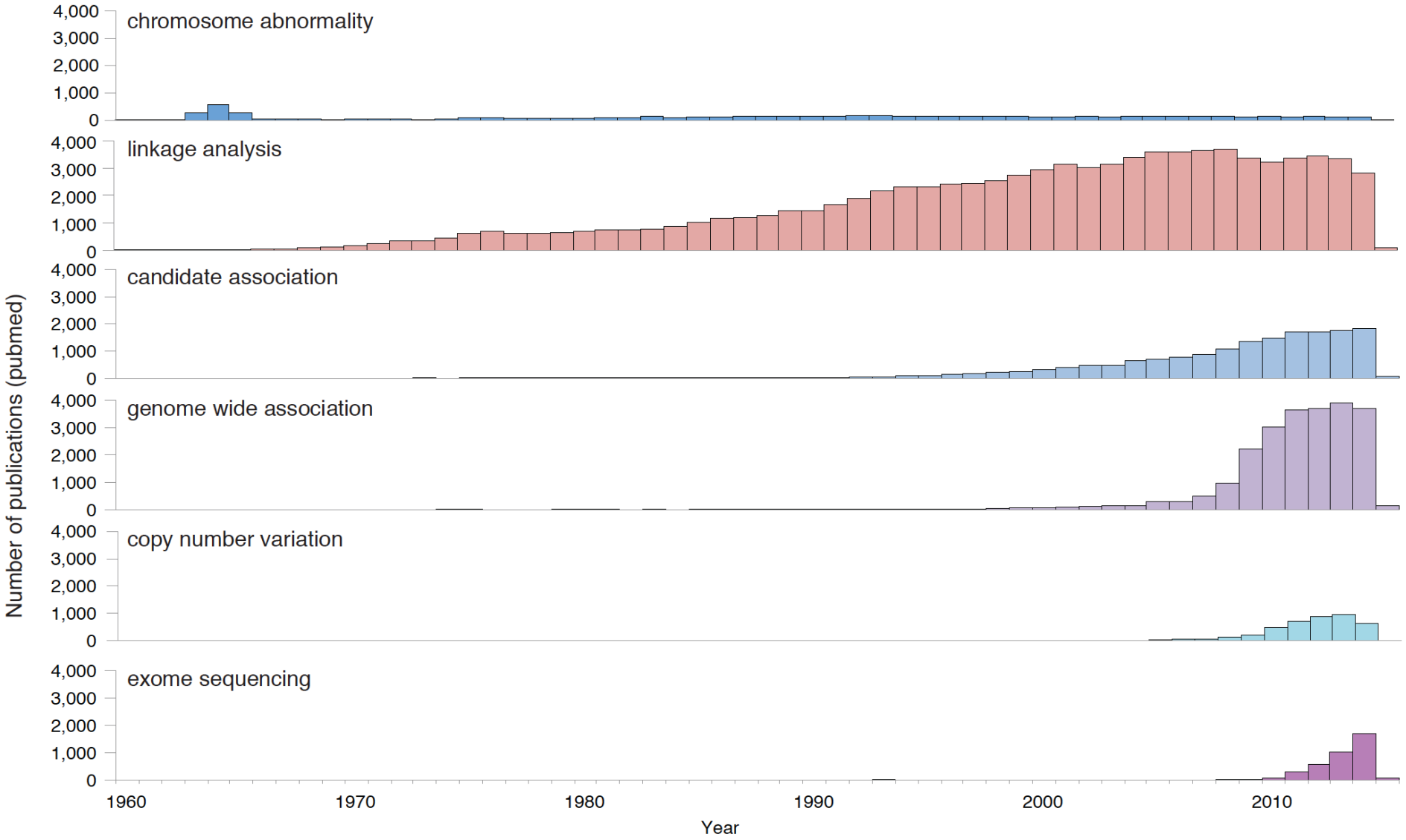
Targeted methods for finding different types of genetic variation

Frequency →

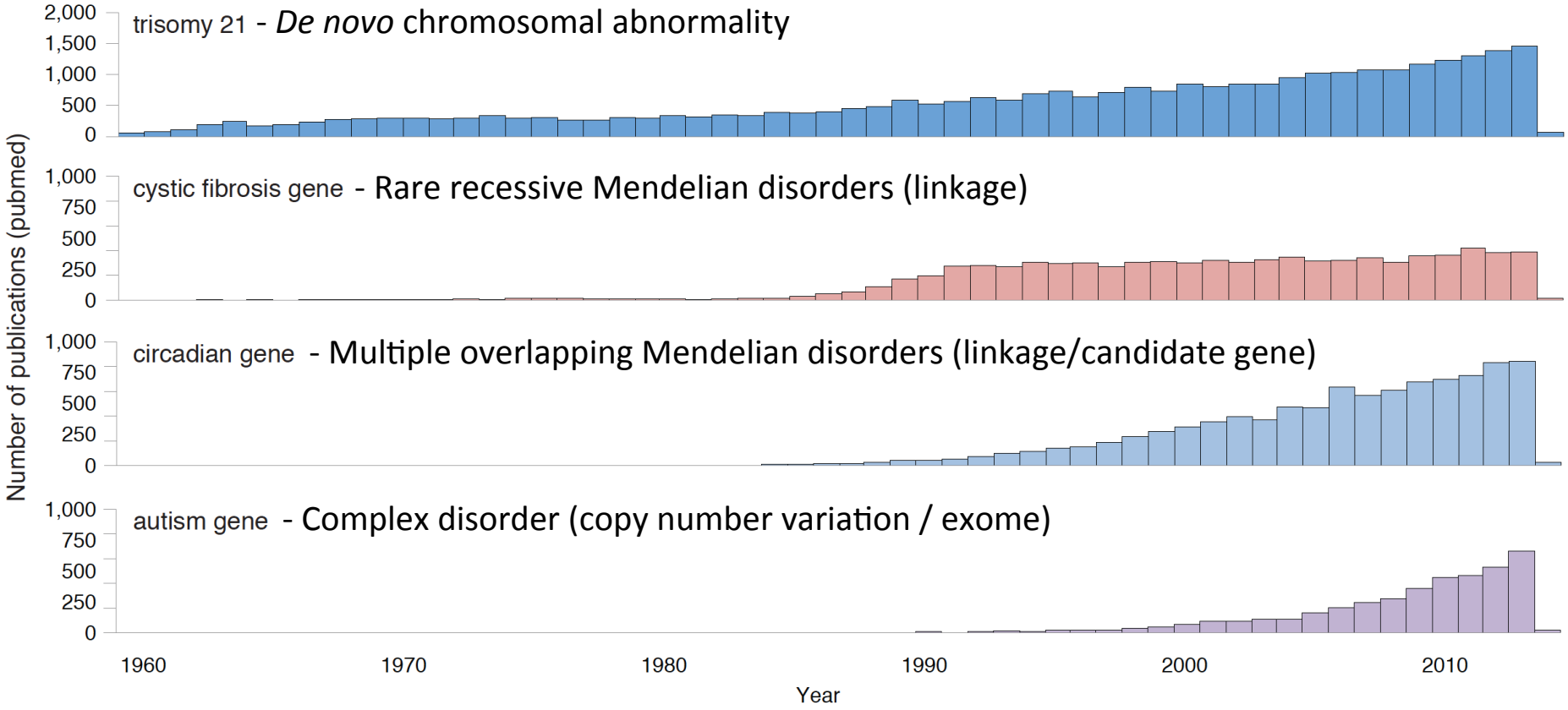
Size ↓

	De novo	Rare Inherited	Common Inherited
SNV 1bp	Candidate gene PCR	Candidate gene PCR	Candidate gene PCR
Indel 1-1,000bp	Candidate gene PCR	Candidate gene PCR	Candidate gene PCR
SV/CNV >1,000bp	FISH	FISH	Candidate region qPCR

Publications by year using different strategies for gene discovery



Publications by year for disorders with different patterns of causation



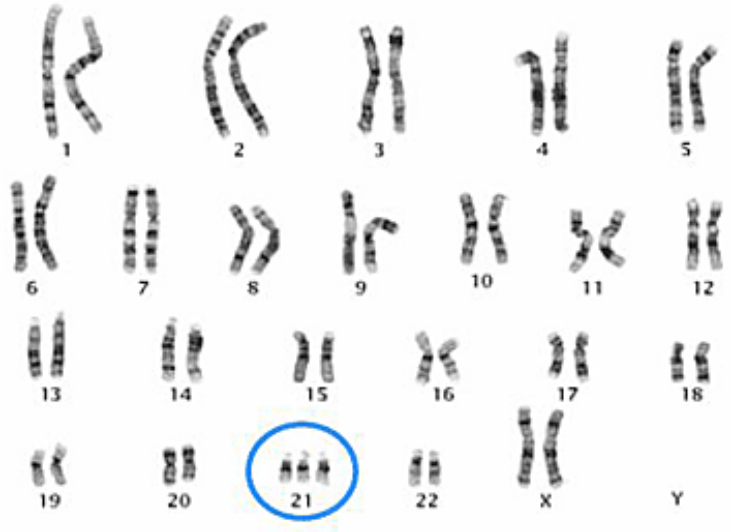
Chromosomal Abnormalities

Karyotype analysis only finds the largest structural variants

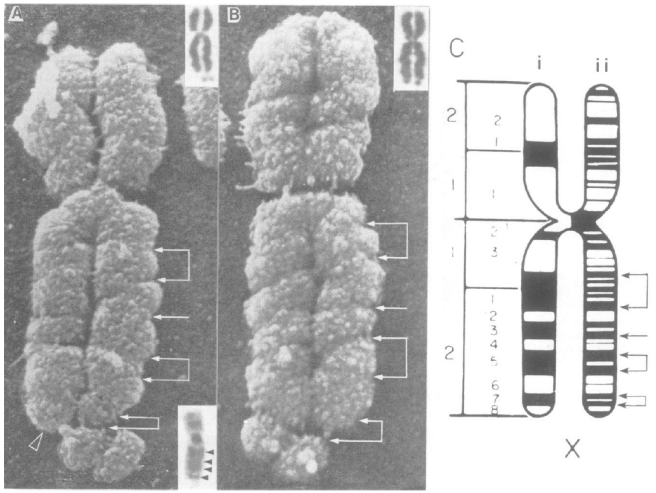
- Karyotype analysis was the first 'genome-wide' technology
- Good for large variants causing serious disease
- Low resolution (≥3Mbp)

- FISH is a high-resolution targeted approach
  - Too slow and expensive multiple regions

Karyotype analysis shows trisomy 21 and fragile X in autism families



Rare, *de novo* chromosomal Abnormality (~1/600)



Harrison *et al. J. Med. Gen.*, 1983

Rare, *de novo* indel (~1/4,000) (triplet repeat expansion)

~3% have autism

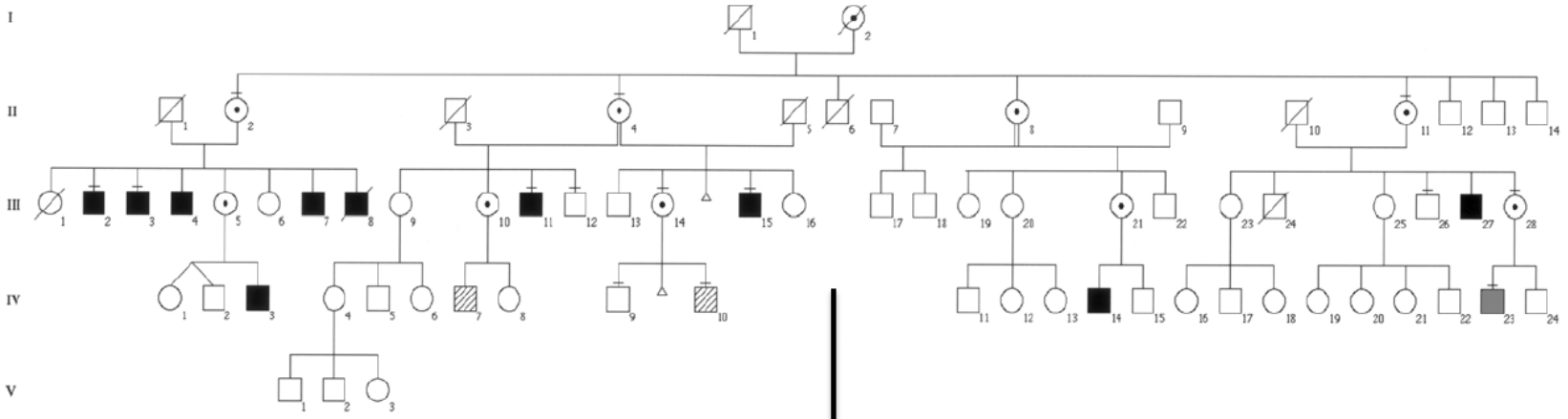
~50% have autism

Linkage

Linkage identifies rare Mendelian disorders

- Efficient approach for finding highly penetrant rare, inherited variants
- Requires large families with multiple affected/ unaffected
- Identifies a (large) region likely to contain the causative variant
- Modern tools have made it much easier
  - Human genome project
  - SNP microarrays
  - exome sequencing

Single example of a gene found using linkage in autism



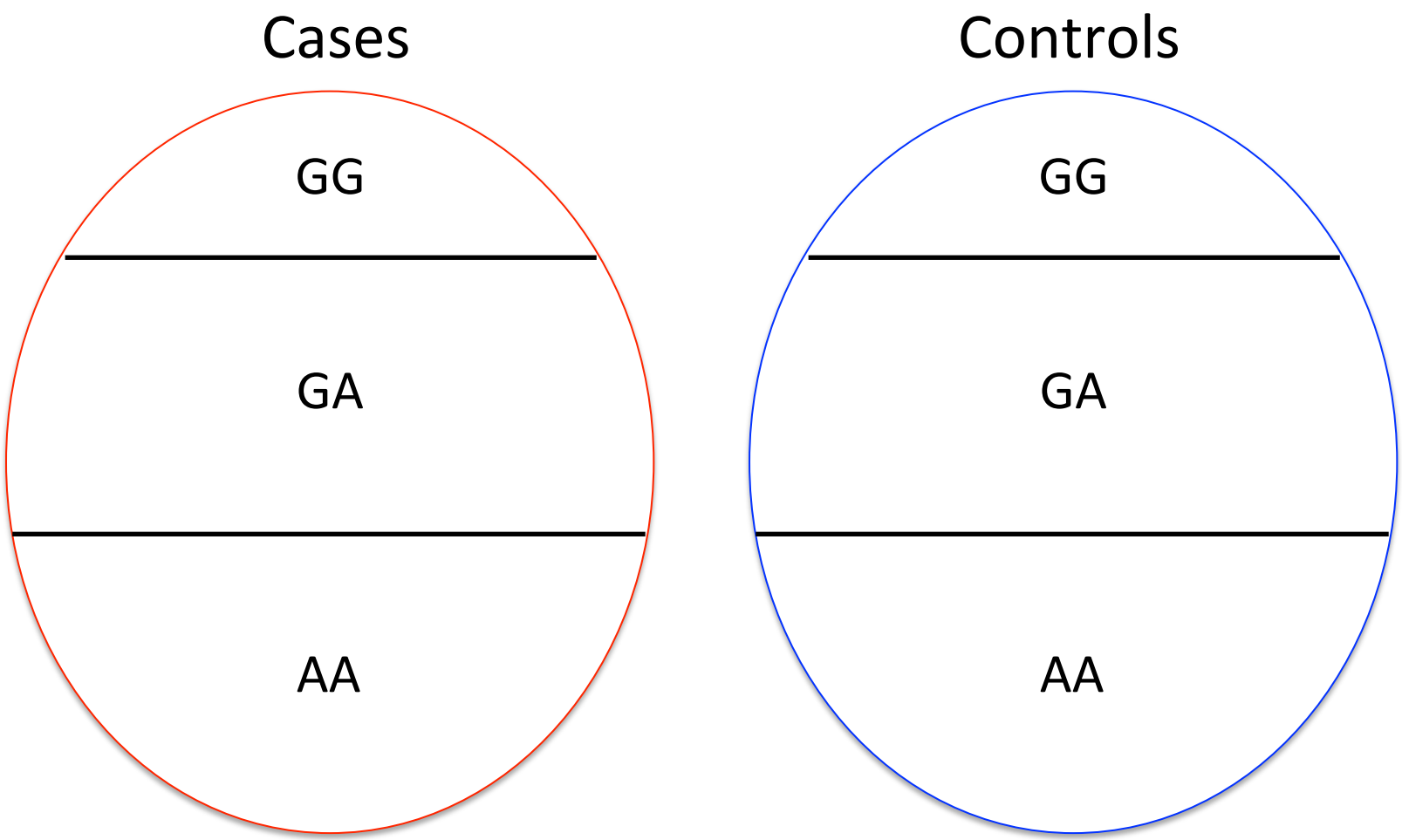
Rare 2bp deletion in exon of NLGN4X (two families)
- Frameshift, disrupts gene function

Autism is a disorder of the synapse

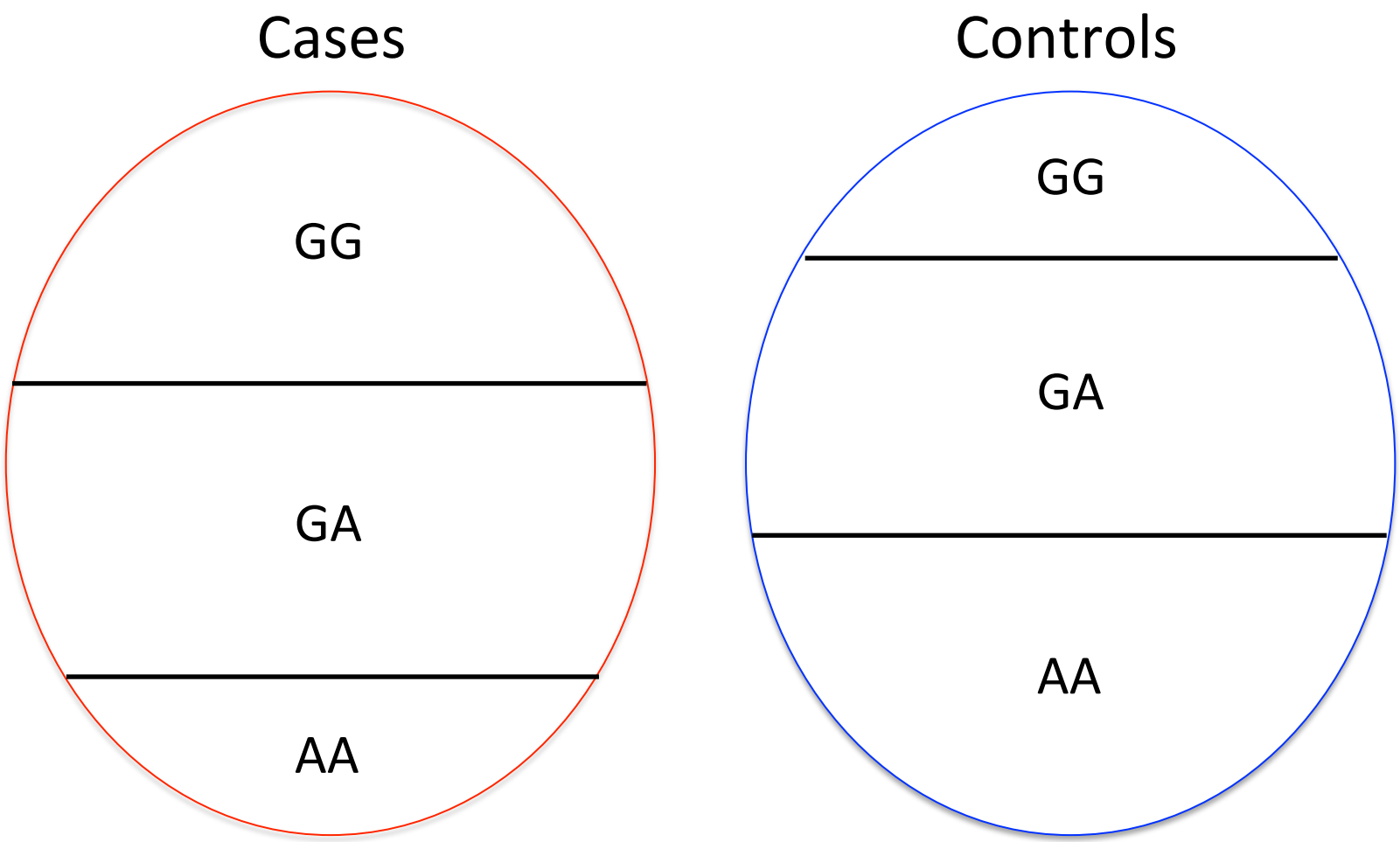
Laumonnier *et al. AJHG.*, 2004

Genetic Association

We expect alleles to be present at the same rate in cases and controls

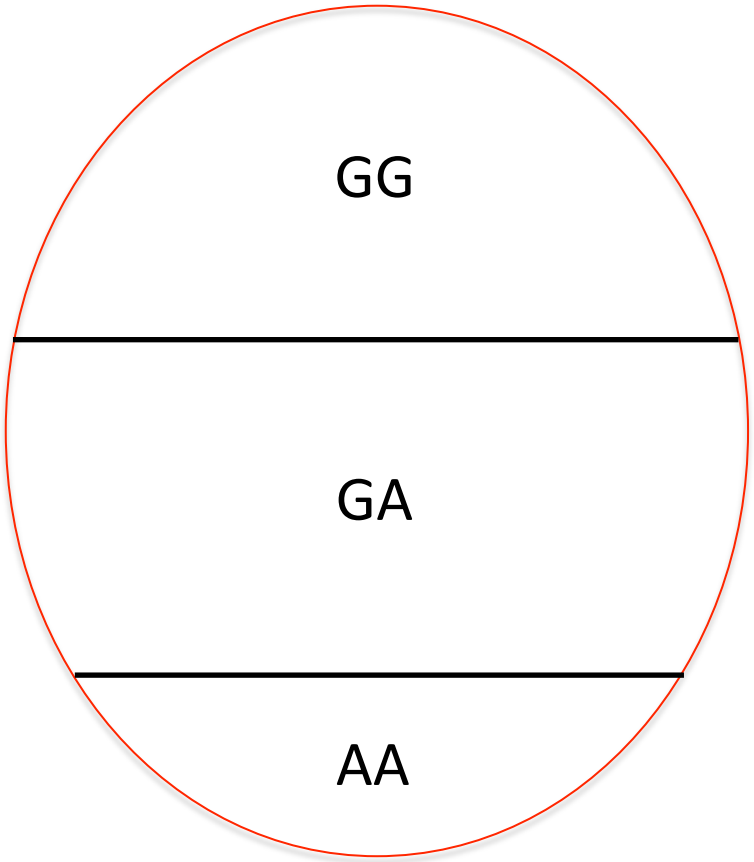


If an allele is more frequent in cases than controls it is ‘associated’

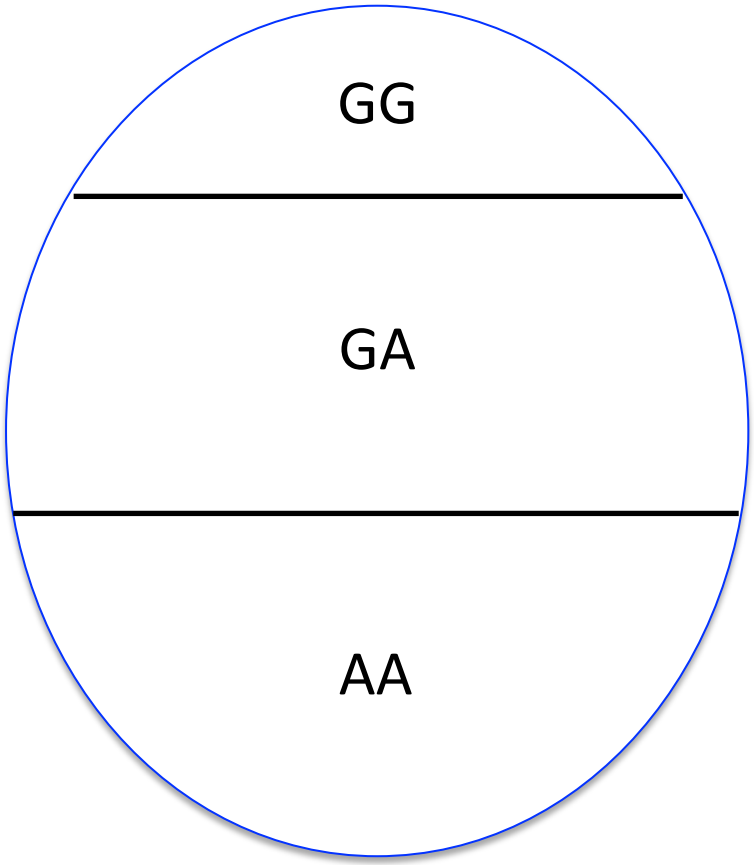


Data must be cleaned to exclude false association (e.g. ancestry)

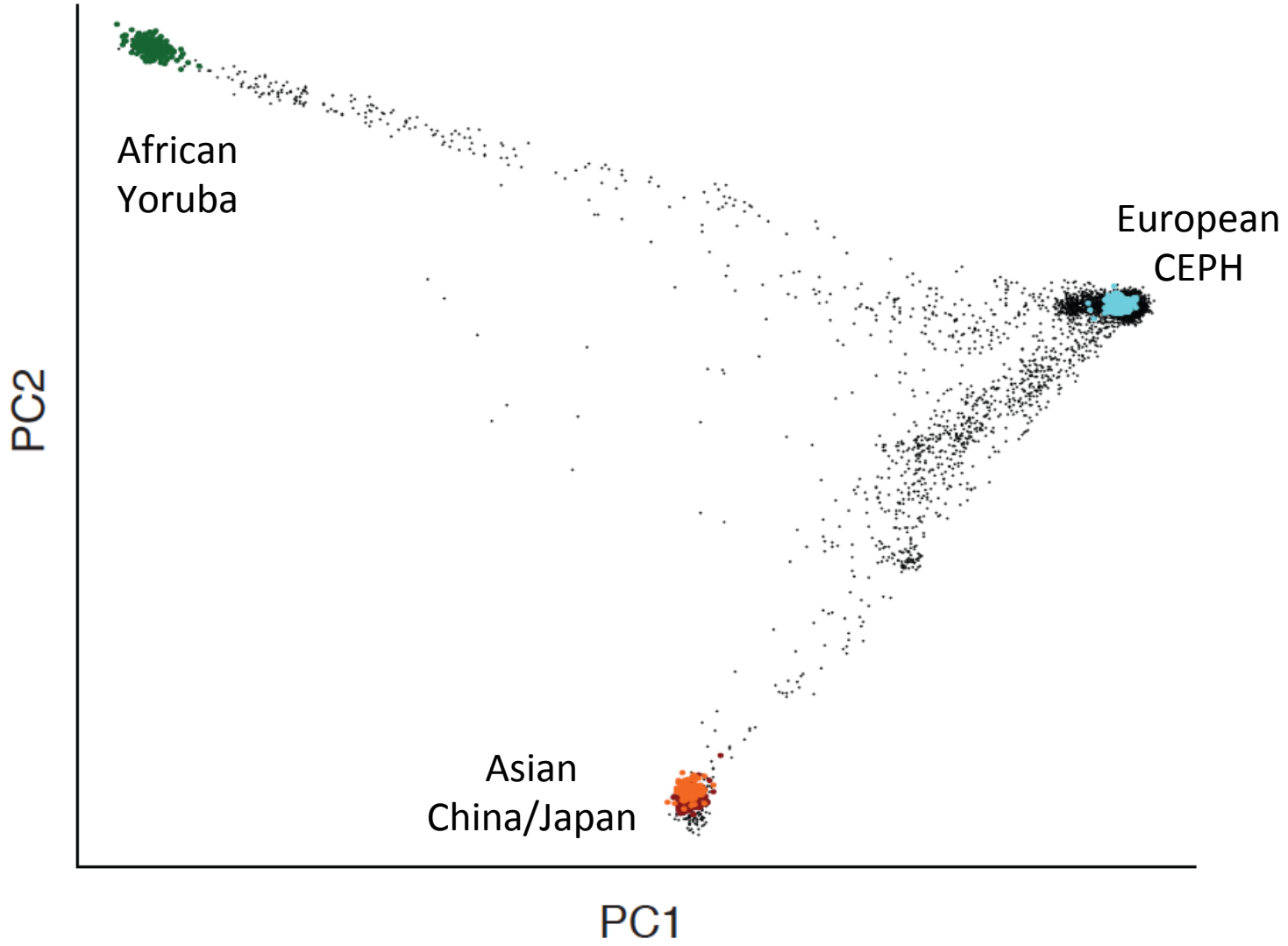
1,021 asthma cases from UCSF Benioff



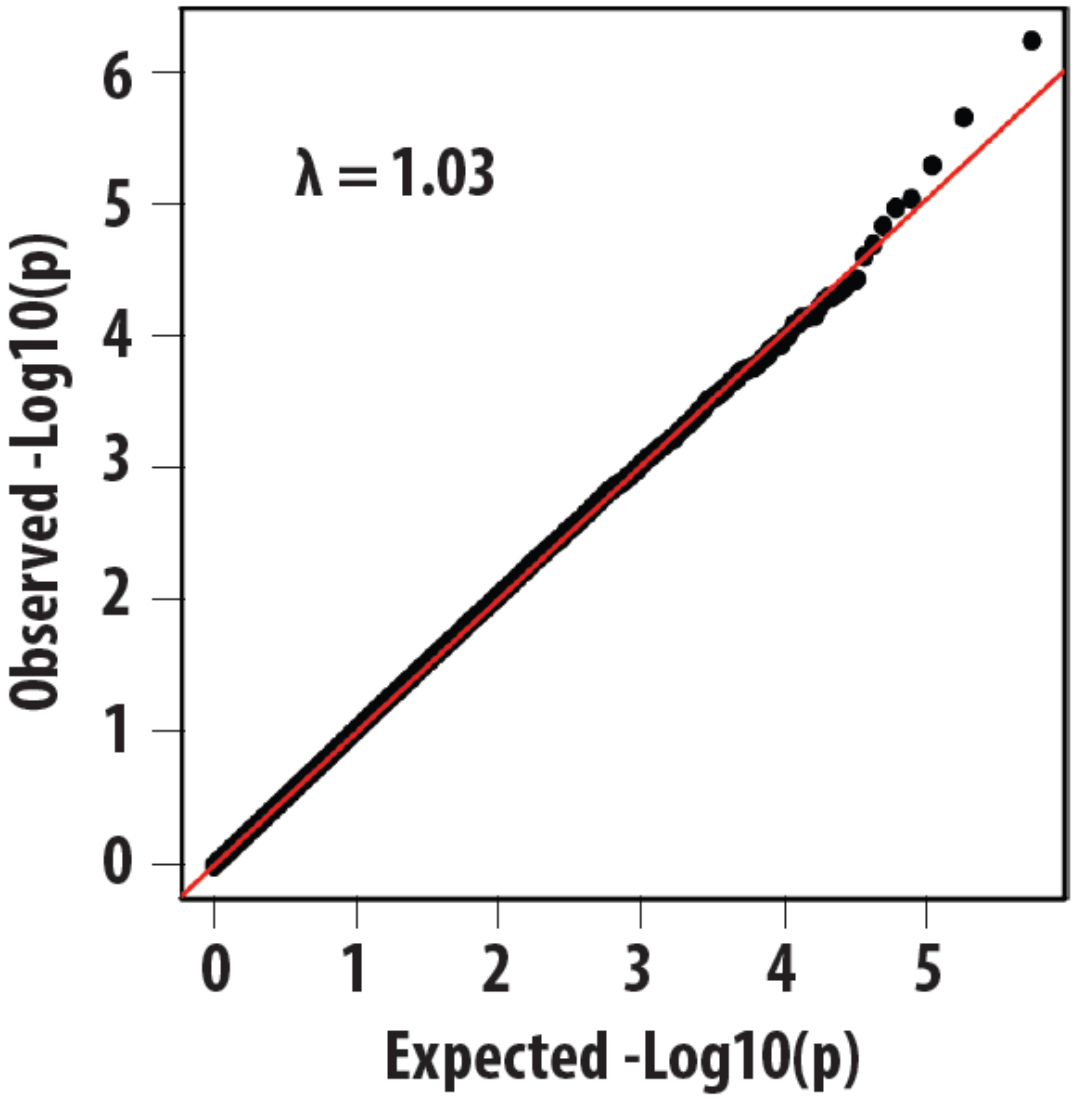
2,321 controls from European collaborator



Ancestry can be determined and corrected using principal component analysis



A Q-Q plot can be used to assess the degree of population stratification



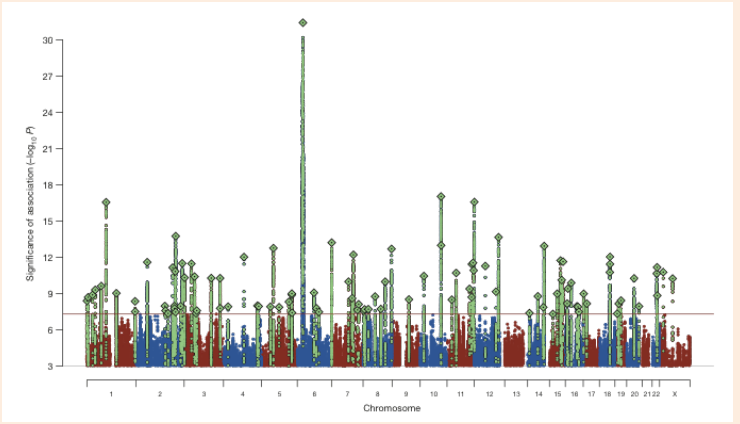
Candidate gene studies

Candidate gene methods work well if you know the underlying biology

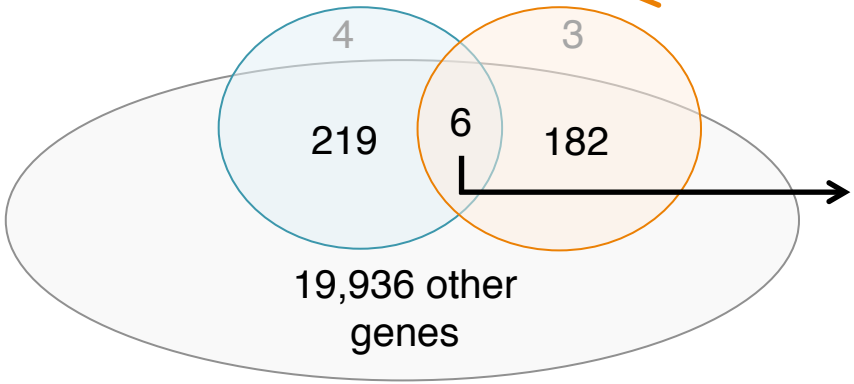
- Assess association in genes expected to be involved in biology
- Efficient approach if biology understood
  - Outliers for circadian rhythms
- Inefficient approach if biology not understood
  - Autism
  - Schizophrenia
- Limited ability to expand understanding of etiology

Candidate gene studies were of limited benefit in schizophrenia

Candidate gene studies:
229 genes from 190 studies



GWAS in 2014:
128 variants that map to 191 genes



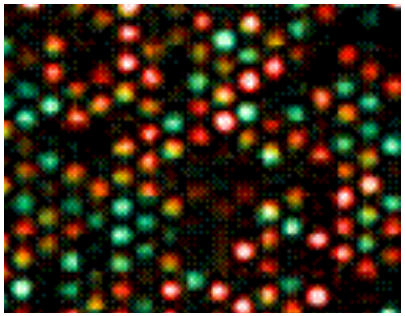
6 correct candidates out of 229
- 2.6% chance of being right
- 1.0% chance of choosing a correct gene at random (p=0.02)

SNP genotyping and GWAS

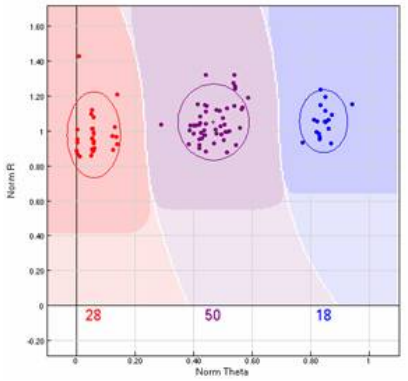
Genotyping arrays detect common inherited variants

Whole-blood derived genomic DNA

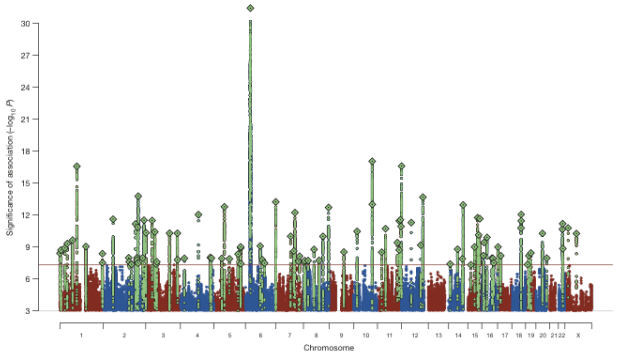
Hybridization



Measure intensity



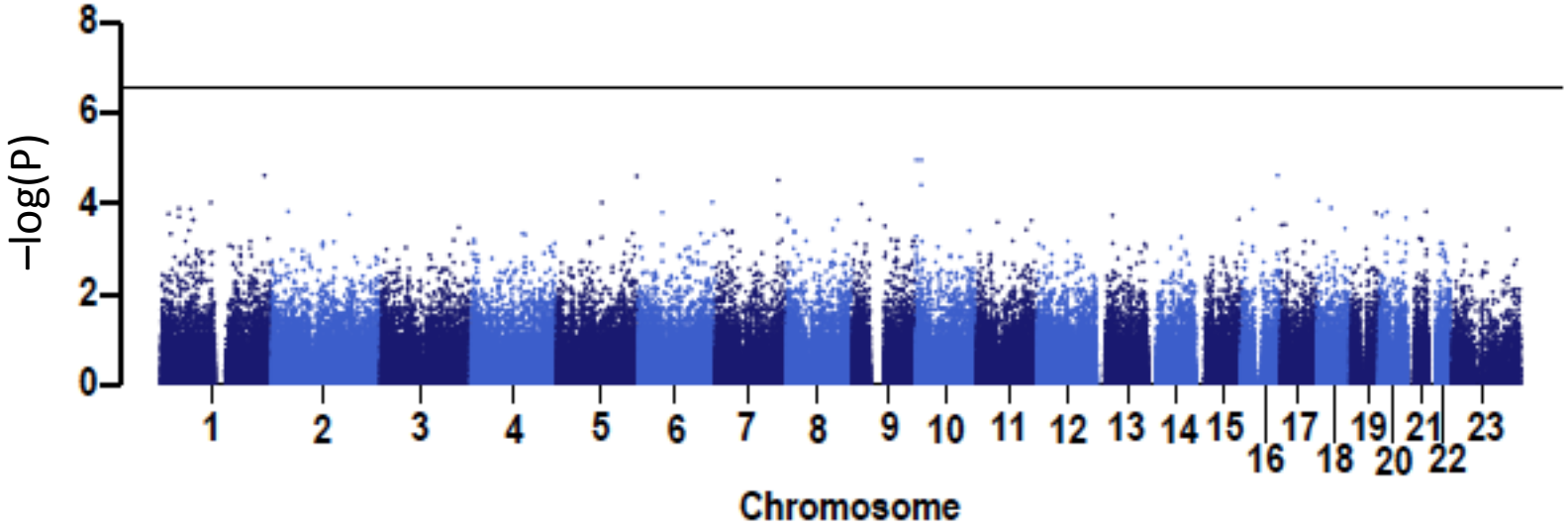
Clustering and genotyping



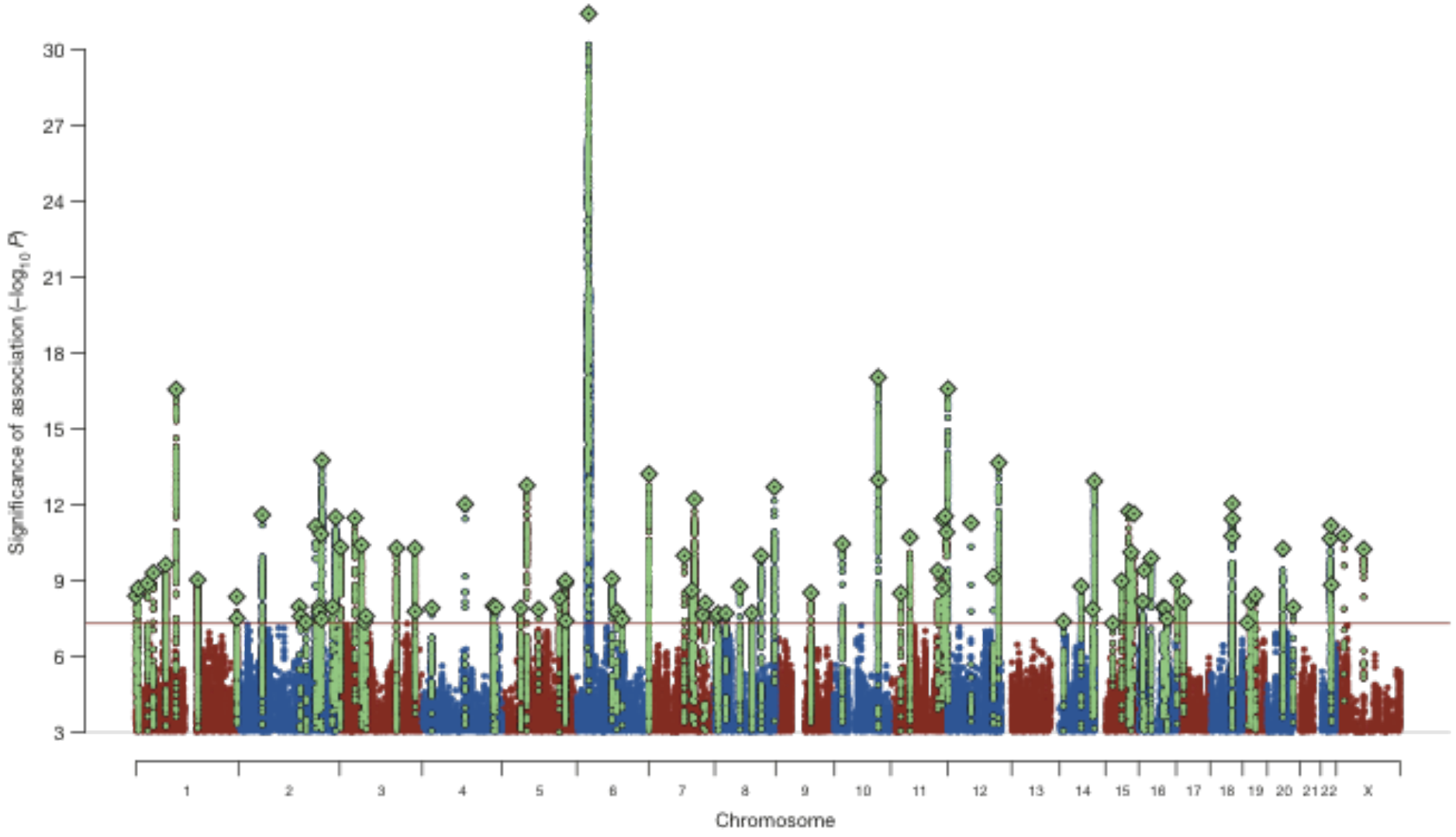
GWAS

GWAS are simply the association for millions of SNPs instead of just one

- Genome-wide analysis of SNP Genotypes
- Extensive data cleaning
- Association test on each remaining SNP
- Plot –log(P) value

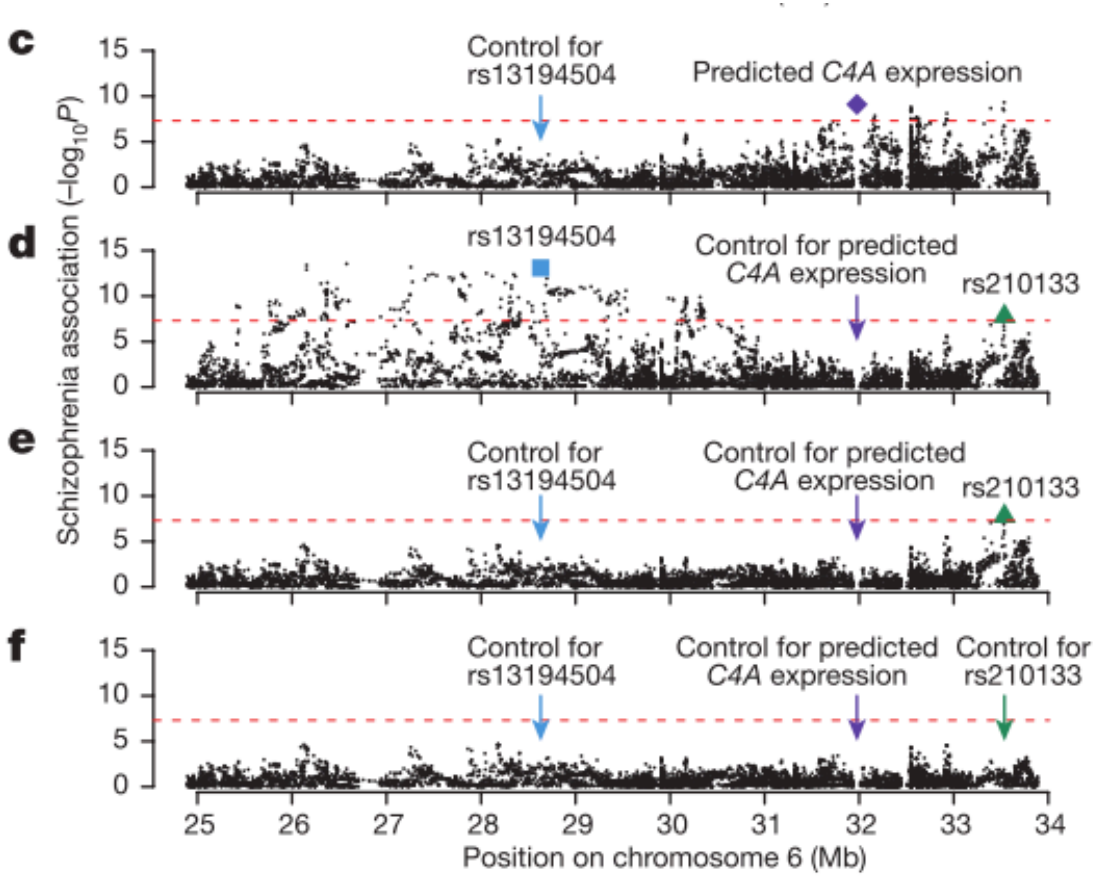


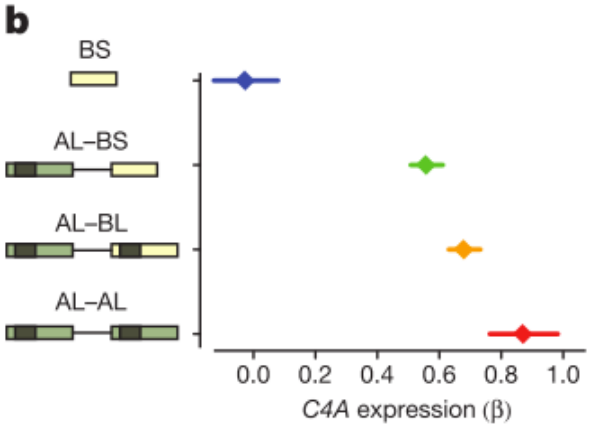
GWAS of 36,989 cases and 113,075 controls finds 108 loci in schizophrenia

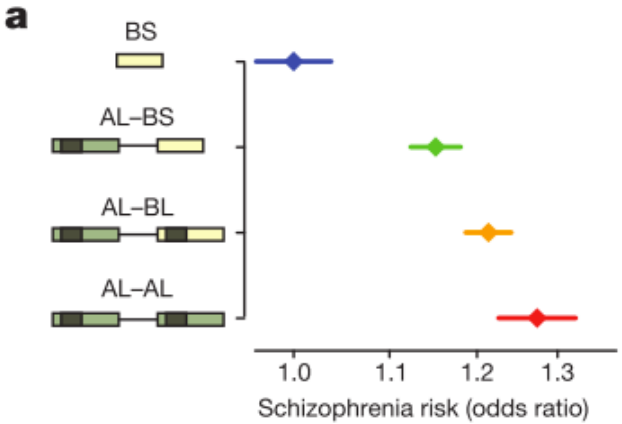


Ripke S *et al.* Nature Genetics., 2014

Identifying the gene contributing the GWAS risk can be complicated







Sekar A *et al.* Nature., 2015

The success of GWAS varies by cohort size and disorder

Disorder	Cases	Controls	Loci	Reference
ASD	2,576	Pseudocontrols	**0**	Chaste et al, Biological Psychiatry 2015
ASD	2,705	Pseudocontrols	**0**	Anney et al, Hum Mol Genet, 2012
Schizophrenia	36,989	113,075	**108**	Ripke et al, Nature 2014
Depression	8,534	8,523	**2**	Cai et al, Nature 2015
Depression	121,380	338,101	**15**	Hyde et al, Nature Genetics 2016
Bipolar disorder	11,974	51,792	**2**	Sklar et al, Nature Genetics 2011
Multiple sclerosis	9,772	17,376	**87**	Sawcer et al, Nature 2011
Multiple sclerosis	29,300	50,794	**103**	Beecham et al, Nature Genetics 2013
Parkinson's disease	19,061	100,833	**26**	Nalls et al, Nature Genetics 2014
Alzheimer's disease	25,580	48,466	**20**	Lambert et al, Nature Genetics 2013

Interlude: Genome Browser

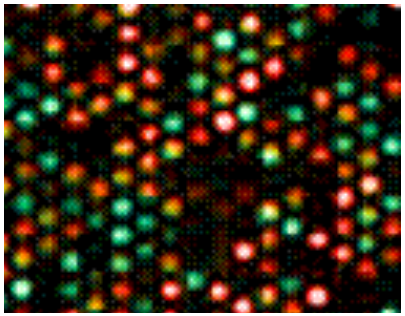
http://genome.ucsc.edu

Copy number variation

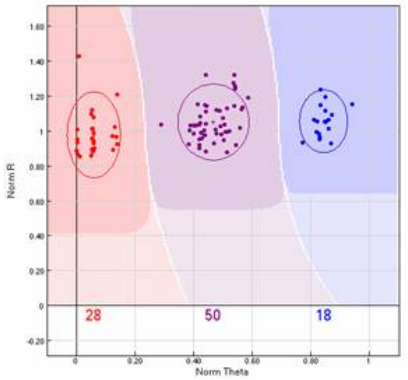
Genotyping arrays detect *de novo* copy number variants (CNVs)

Whole-blood derived genomic DNA

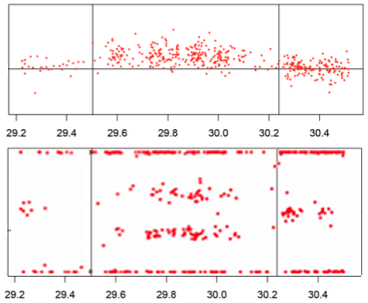
Hybridization



Measure intensity



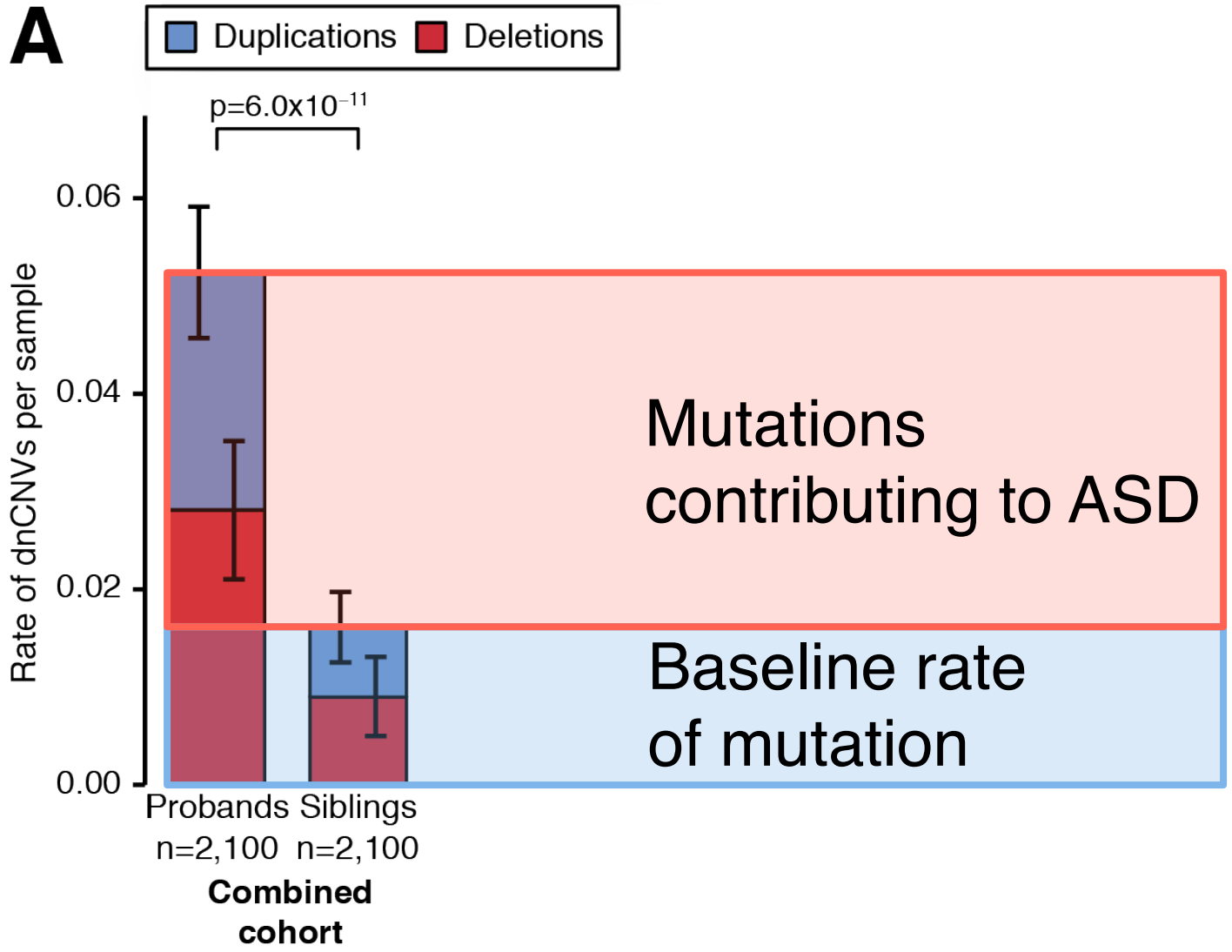
Clustering and genotyping



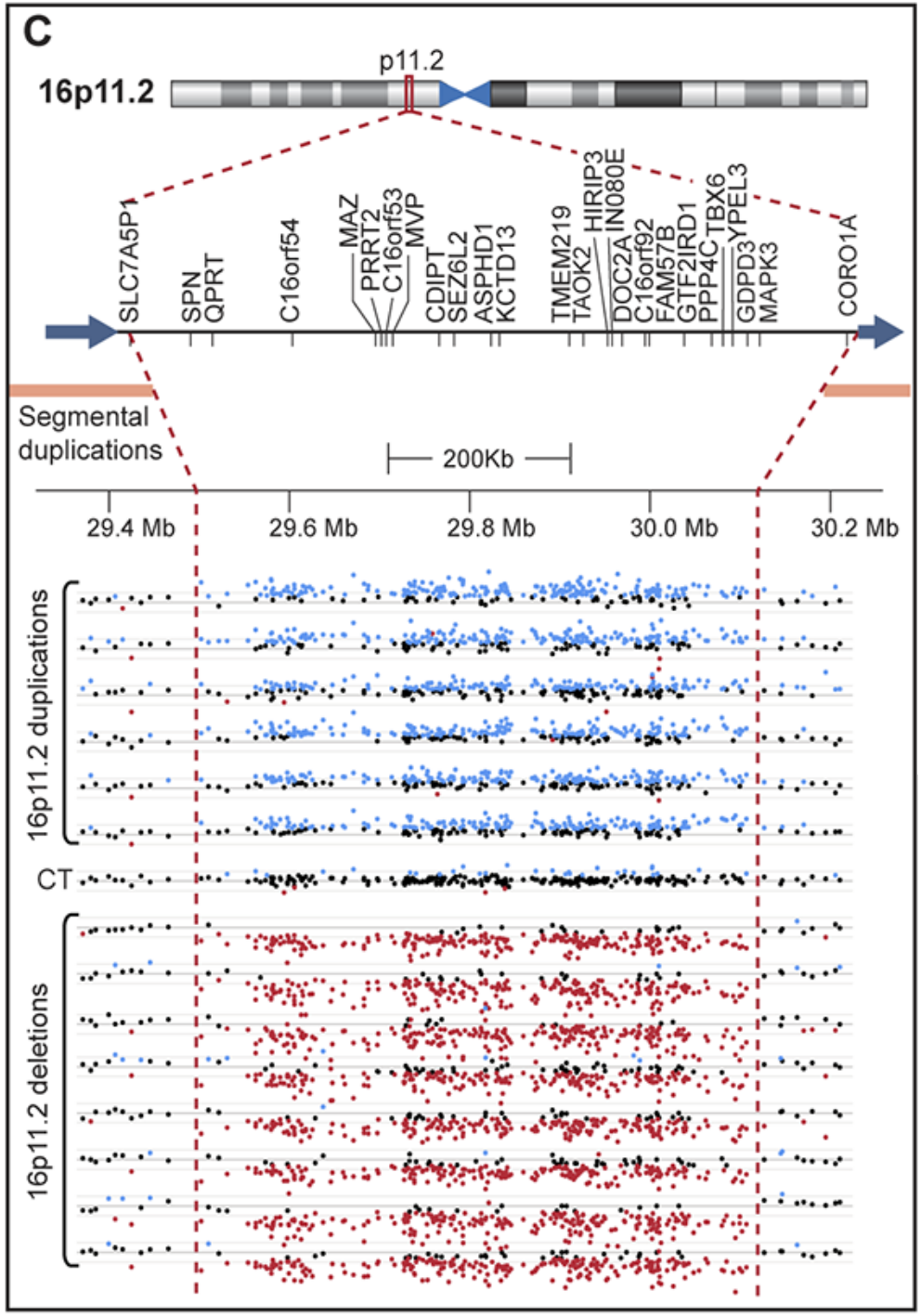
De novo CNVs

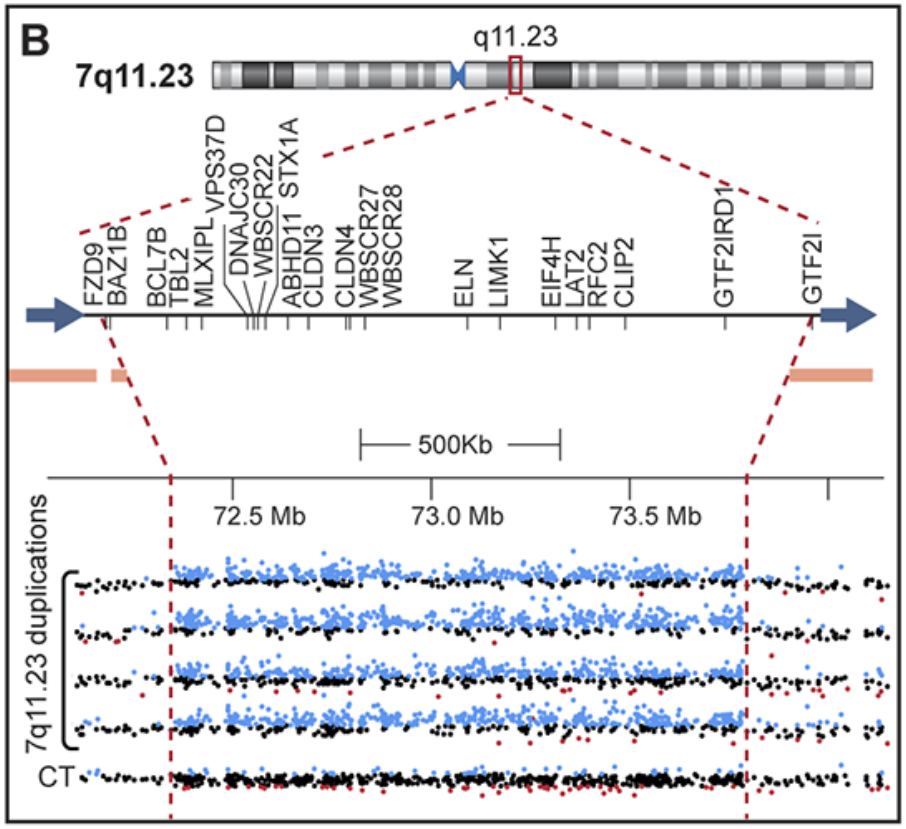
qPCR Confirmation

De novo CNVs are associated with ASD



Sanders *et al.* Neuron, 2015





7q11.23 duplications are associated with ASD

7q11.23 deletion -> Williams Syndrome

- Hypersociable personality

7q11.23 duplication -> Social impairment

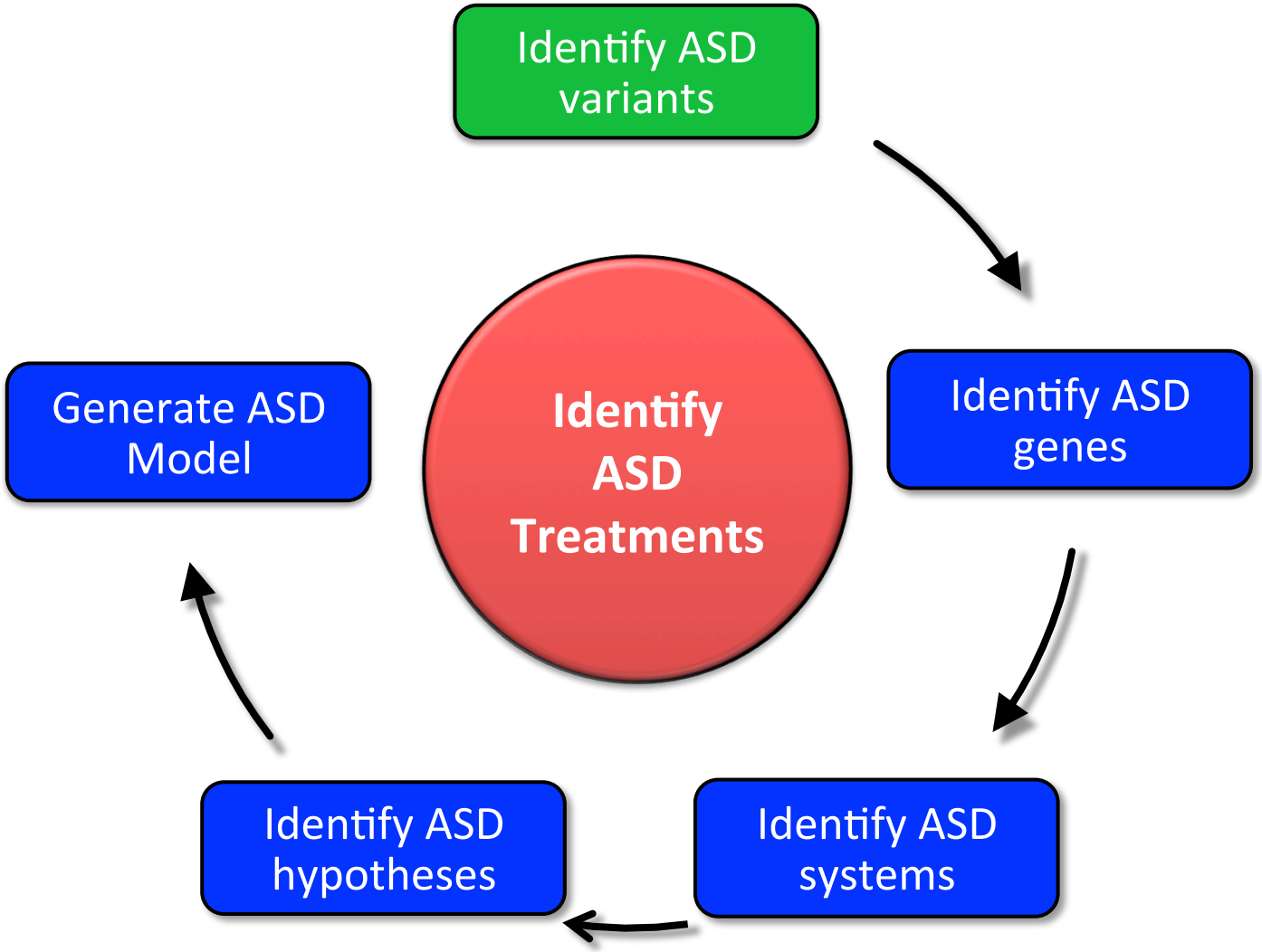
- Hyposociable personality

A few CNV loci meet genome-wide significance

Disorder	Cases	Controls	Loci	Reference
ASD	4,687	De novo	**8**	Sanders et al, Neuron 2015
Schizophrenia	21,094	20,227	**8**	Marshall et al, Nature Genetics 2017
Depression	3,106	3,158	**0**	Rucker et al, Biol Psychiatry 2015
Bipolar disorder	2,591	8,842	**1**	Green et al, Mol Psychiatry 2014

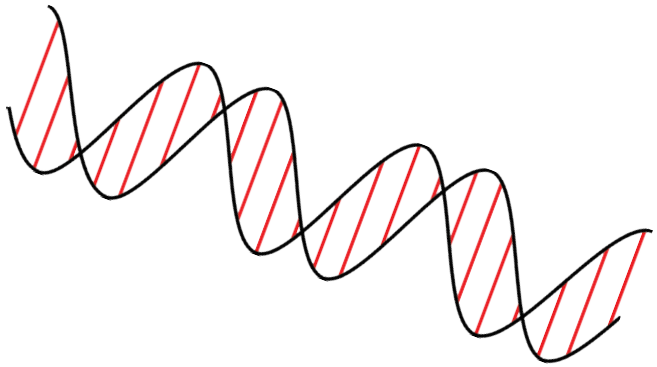
No large-scale CNV analyses published in multiple sclerosis, Parkinson's, or Alzheimer's

De novo CNVs show that ASD variants can be identified



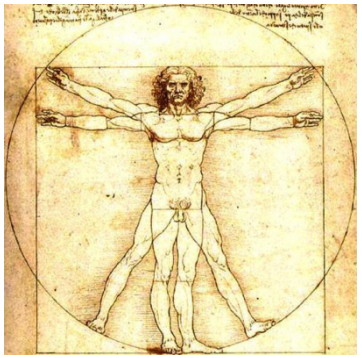
Exome sequencing

Next generation sequencing detects *de novo* single nucleotide variants (SNVs)

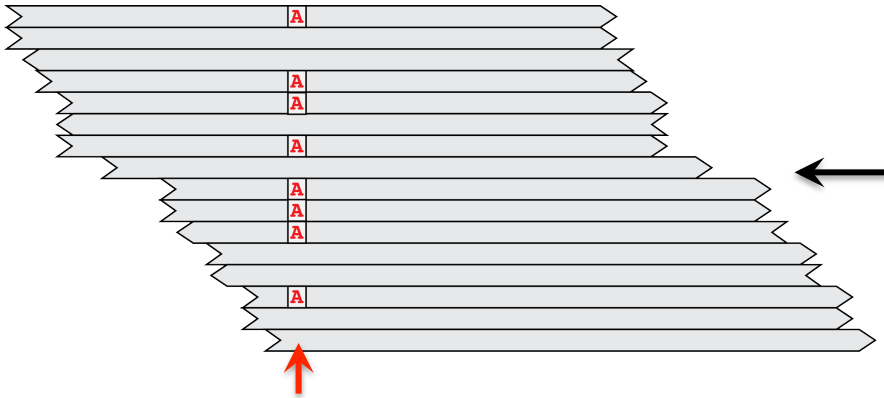


Whole-blood derived genomic DNA

High-throughput sequencing

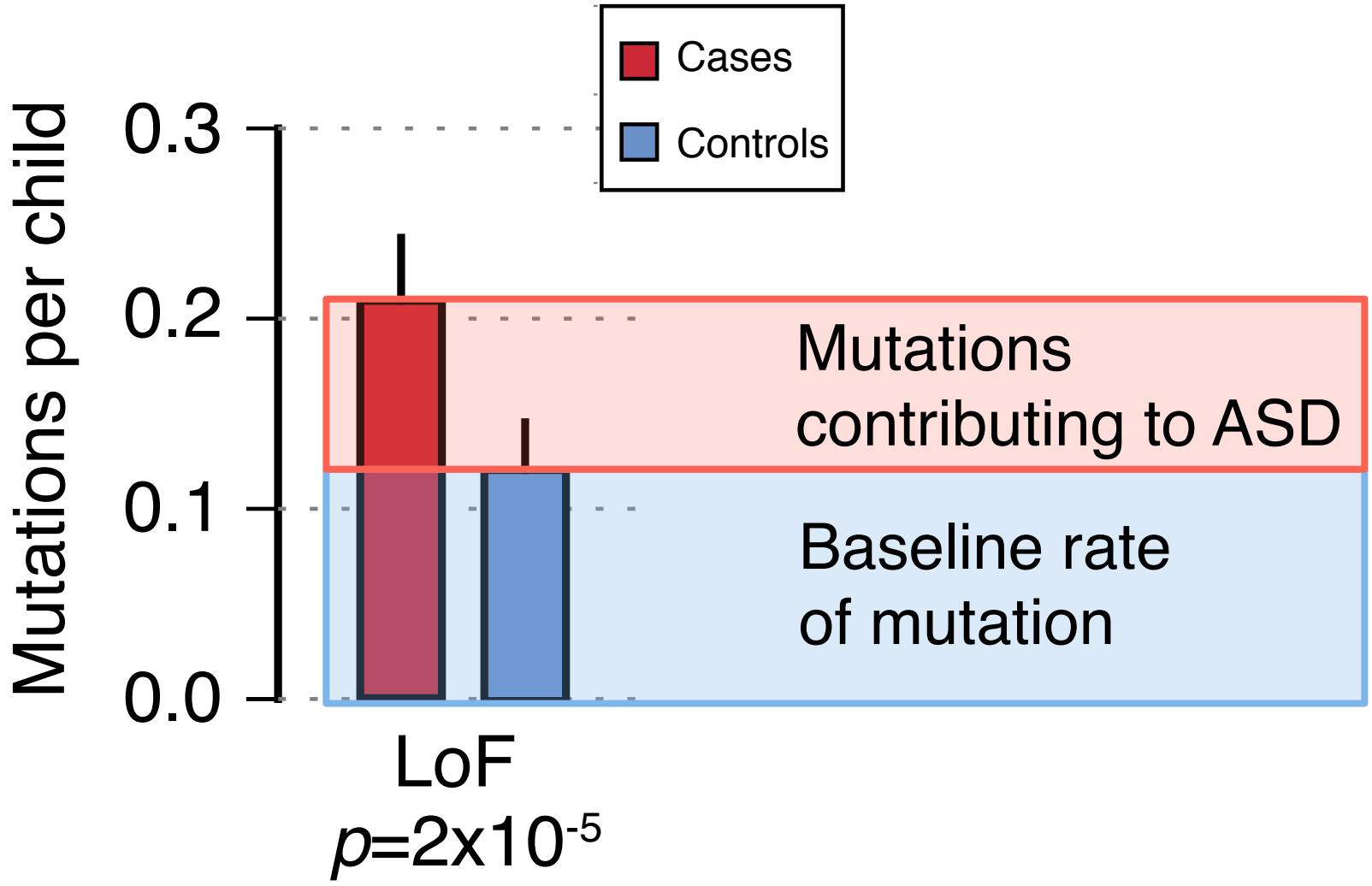


Aligned to human genome



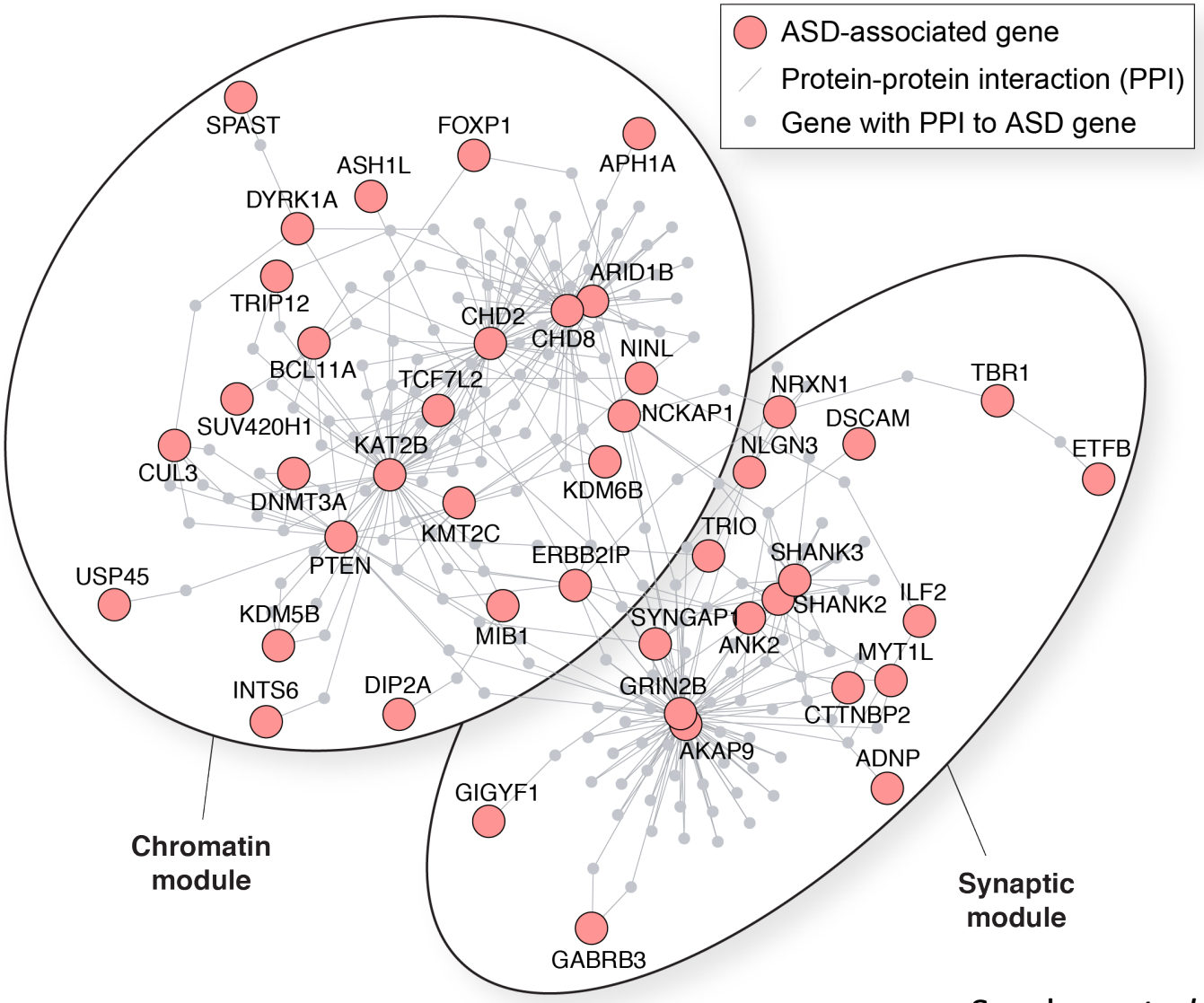
Variants detected

An increased rate of *de novo* LoF mutations in cases shows association with ASD



Adapted from Iossifov et al, Nature, 2014

The 65 ASD risk genes converge on chromatin and synaptic networks

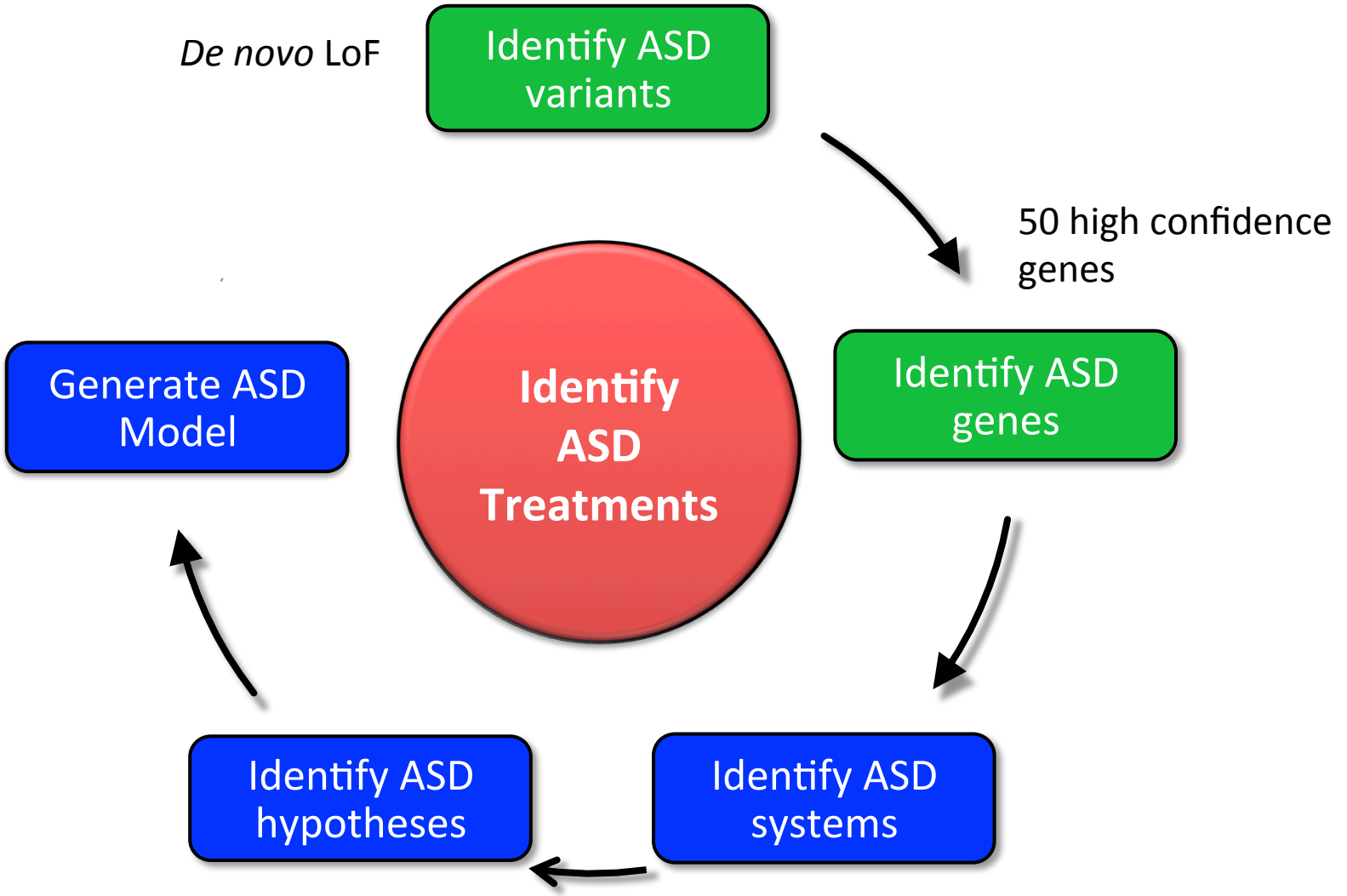


Sanders *et al.* Neuron, 2015

Exome sequencing has identified multiple genes in early onset disorders

Disorder	Cases	Controls	Loci	Reference
Developmental Delay	4,293	De novo	**94**	McRae et al, BioRXiv
ASD	5,563	De novo	**65**	Sanders et al, Neuron 2015
Schizophrenia	617	De novo	**0**	Fromer et al, Nature 2014
Schizophrenia	2,536	2,543	**0**	Purcell et al, Nature 2014

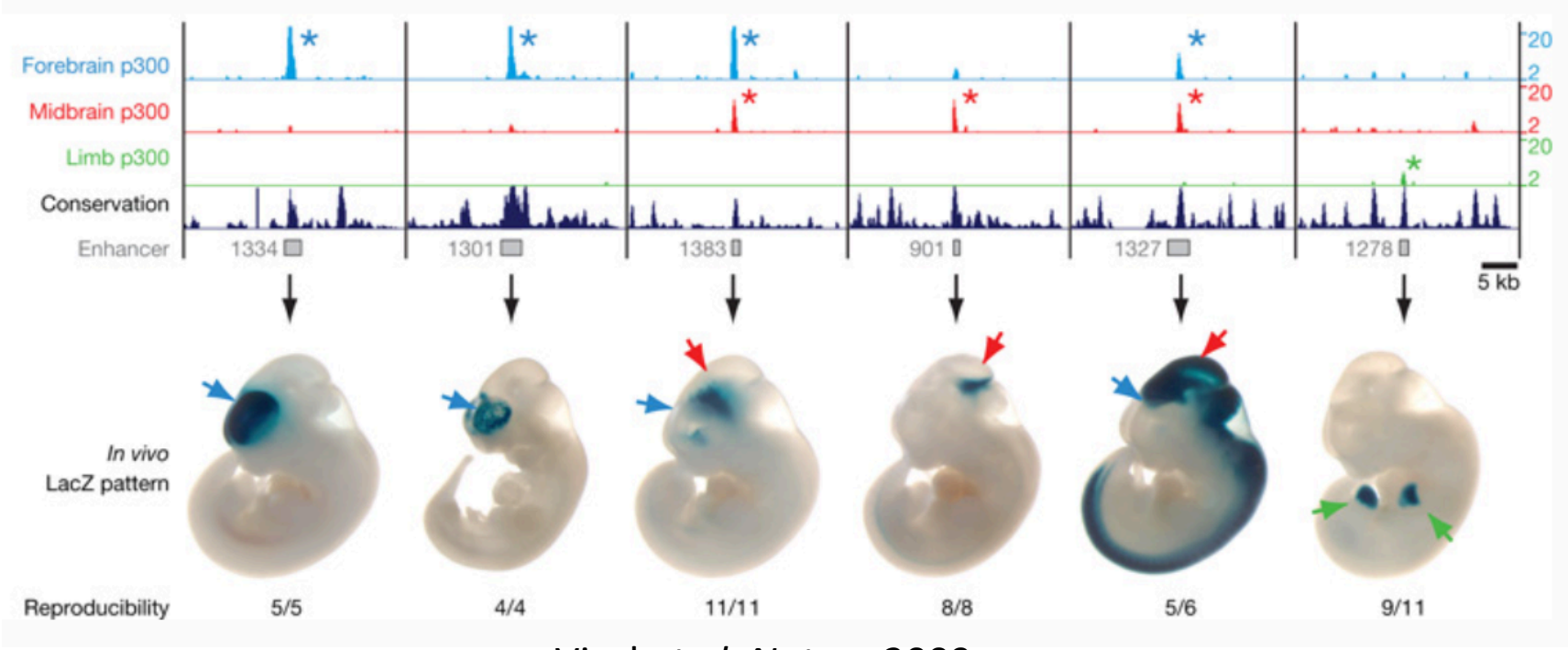
De novo SNVs can be used to identify ASD genes with high confidence



Whole genome sequencing

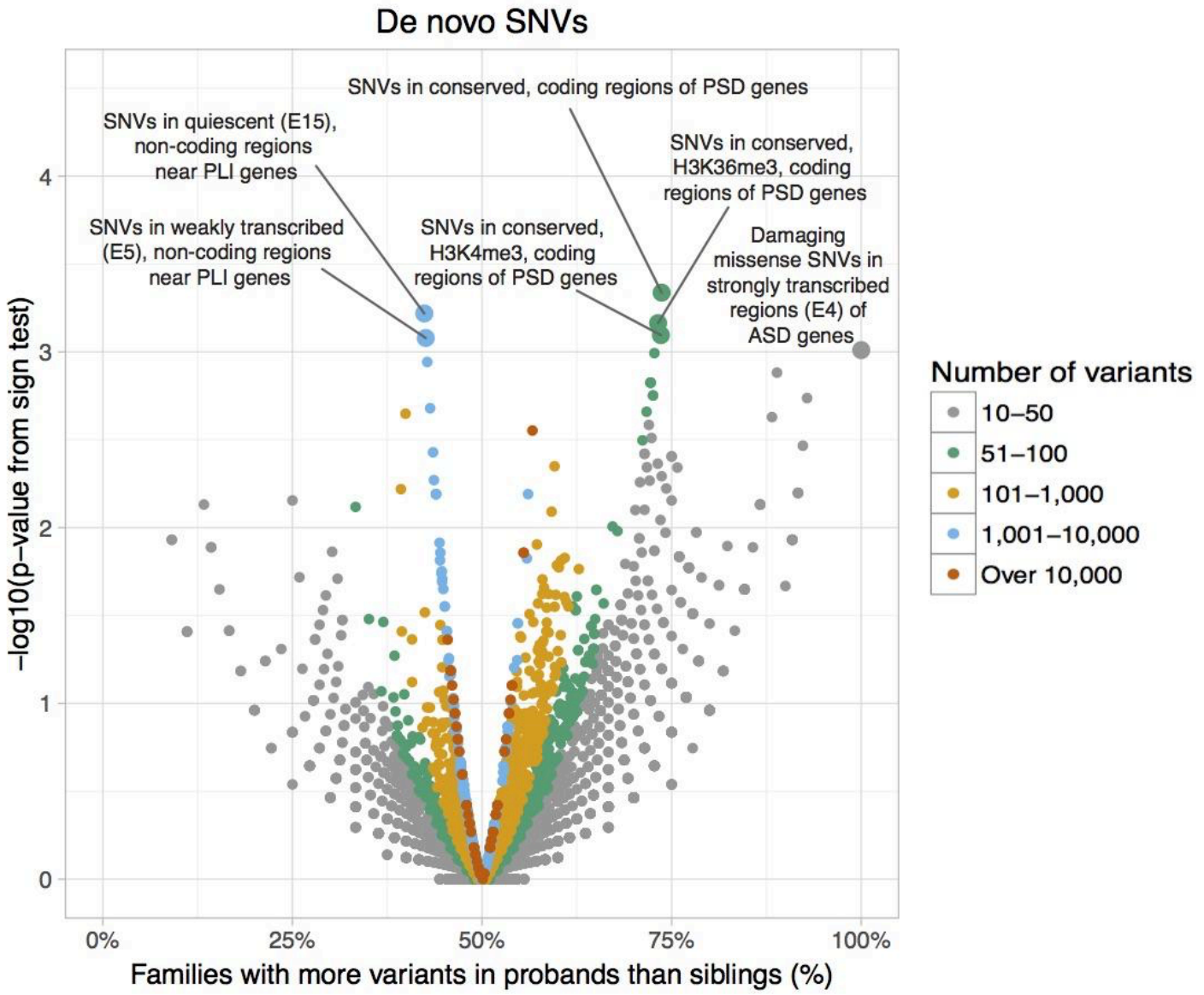
Whole genome analysis has the potential to identify rare non-coding variants

- When in development
- Where in the brain
- Which cell type



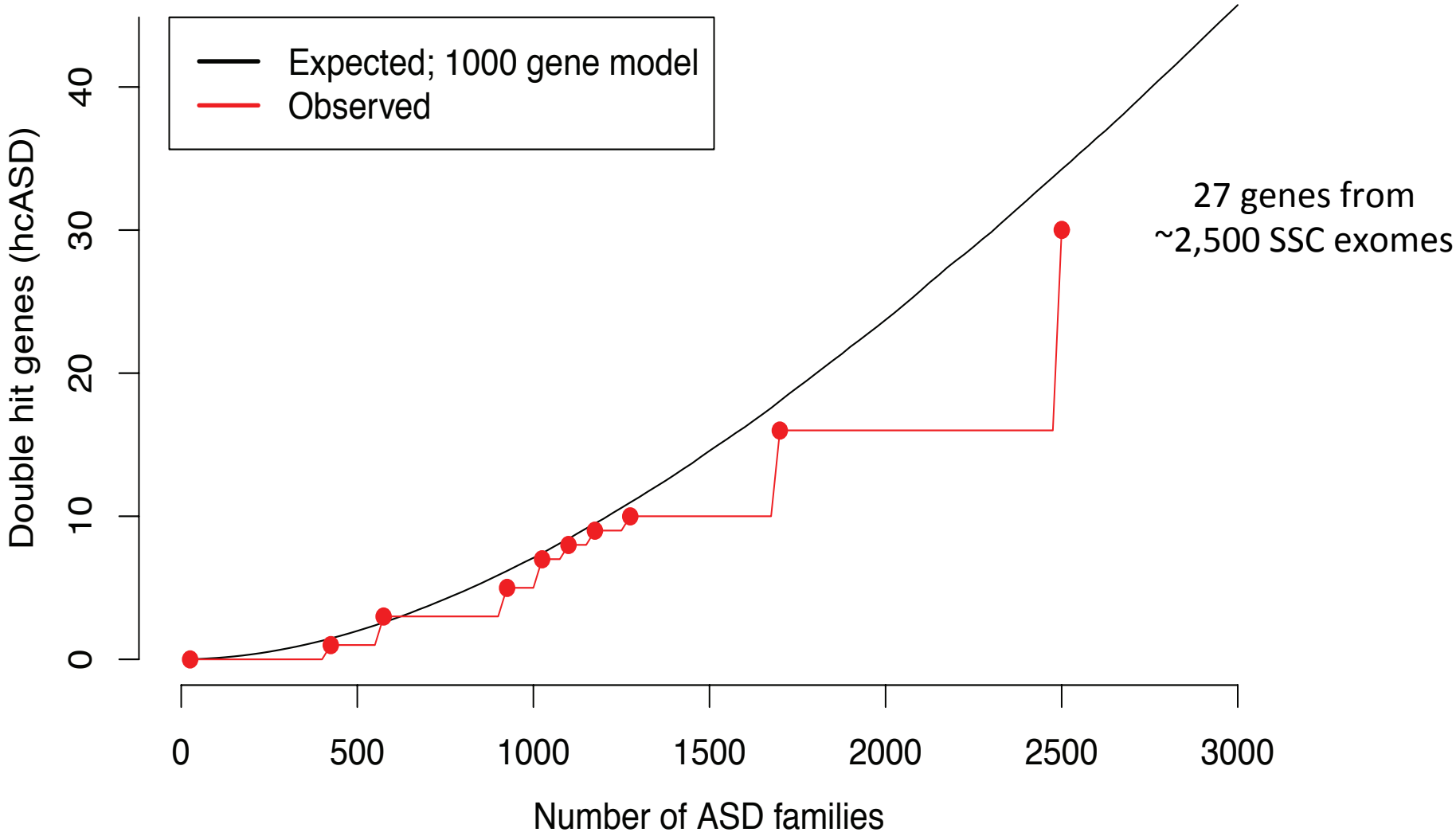
Visel *et al. Nature* 2009

No clear non-coding signal in 519 ASD families, but more samples are coming

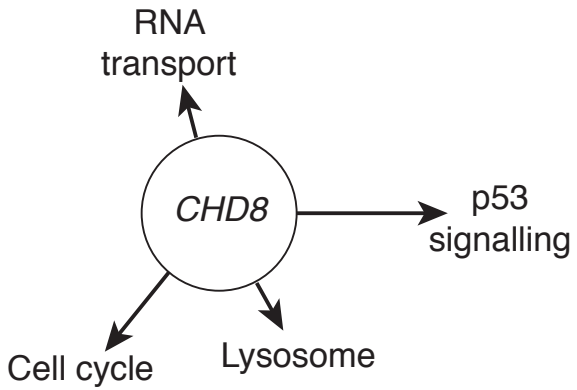


Systems analysis

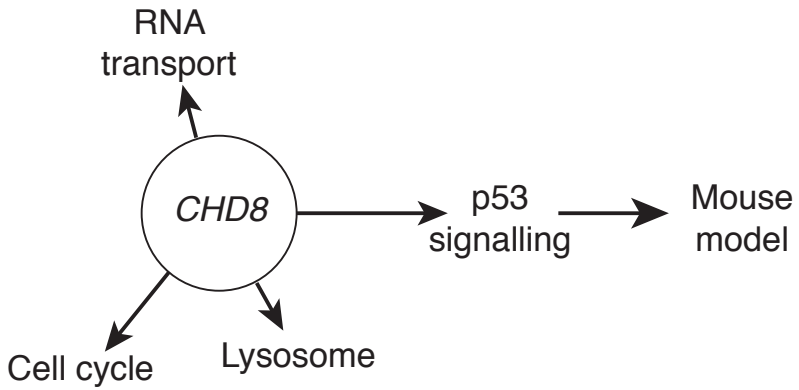
Exome data fits 1,000 gene model of ASD causation



Having 1,000 genes may be a benefit, not a disadvantage

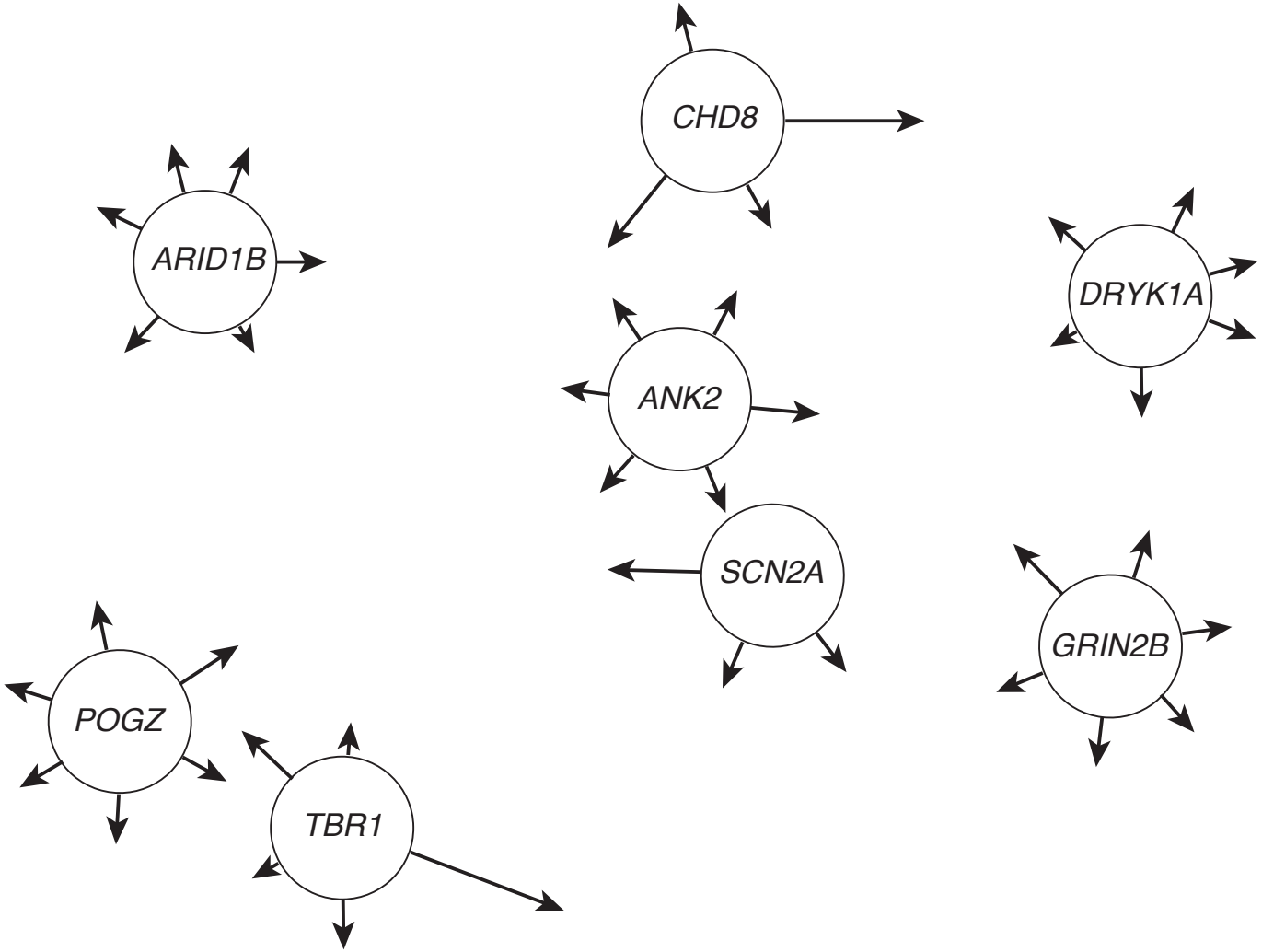


Having 1,000 genes may be a benefit, not a disadvantage

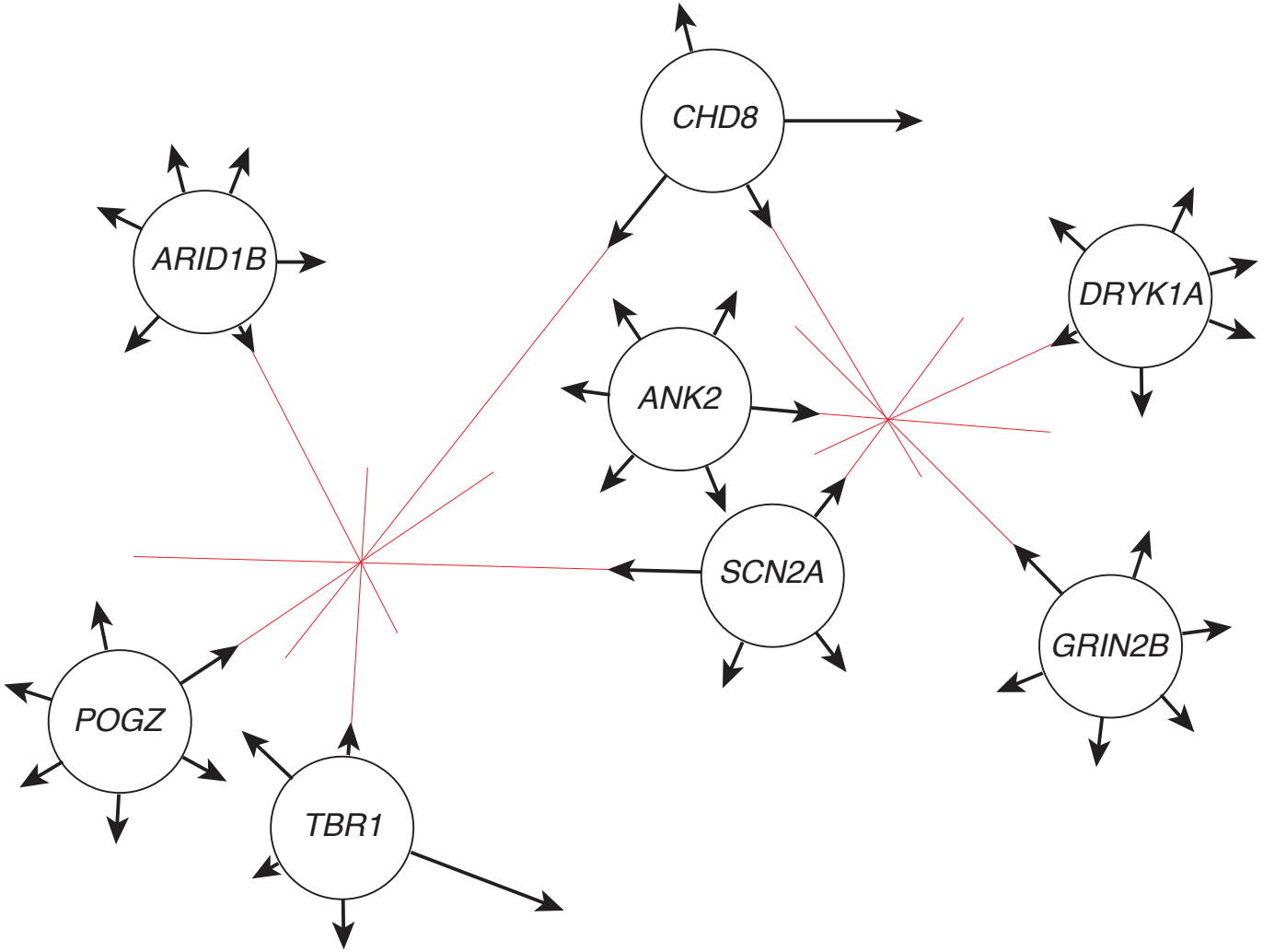


ASD genes are often highly pleiotropic (multiple functions). Only a subset of their functions will play a role in ASD pathology

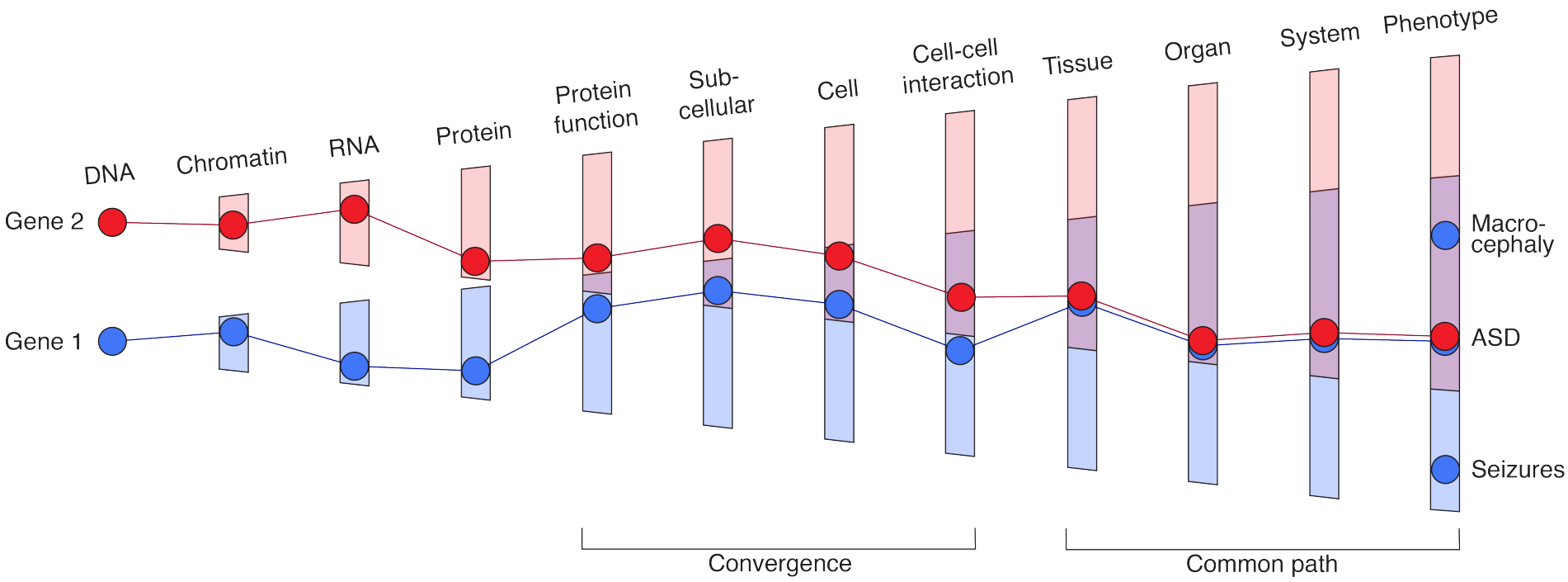
Having 1,000 genes may be a benefit, not a disadvantage



Having 1,000 genes may be a benefit, not a disadvantage



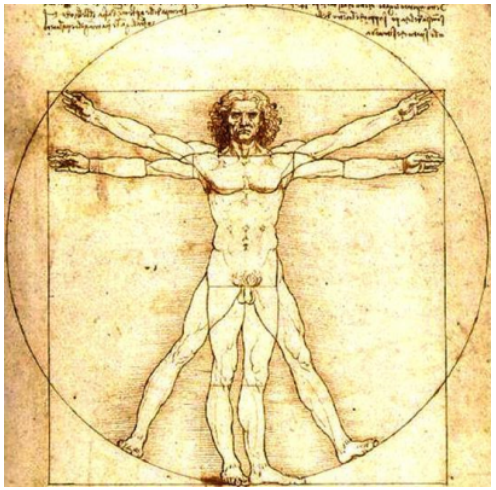
Convergence between variants can be quantified to estimate causation



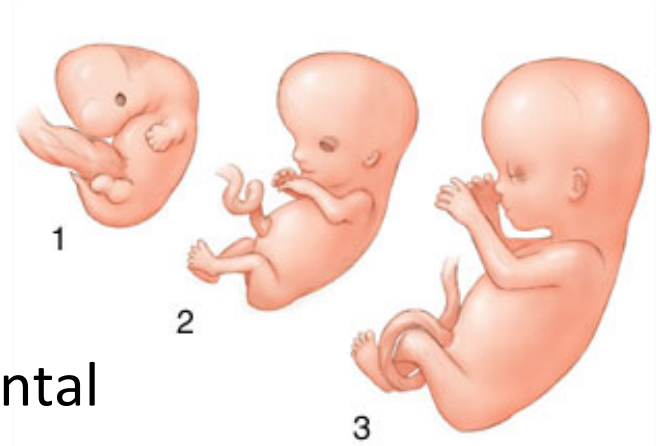
Sanders SJ, Curr Opin Genet Dev, 2015

Understanding neuropsychiatric disorders may require complex resources

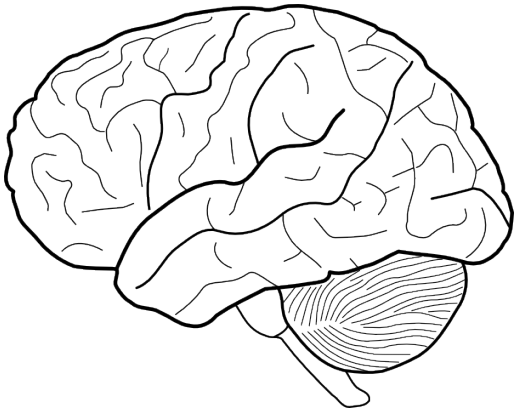
Agnostic; driven by best data



Human context

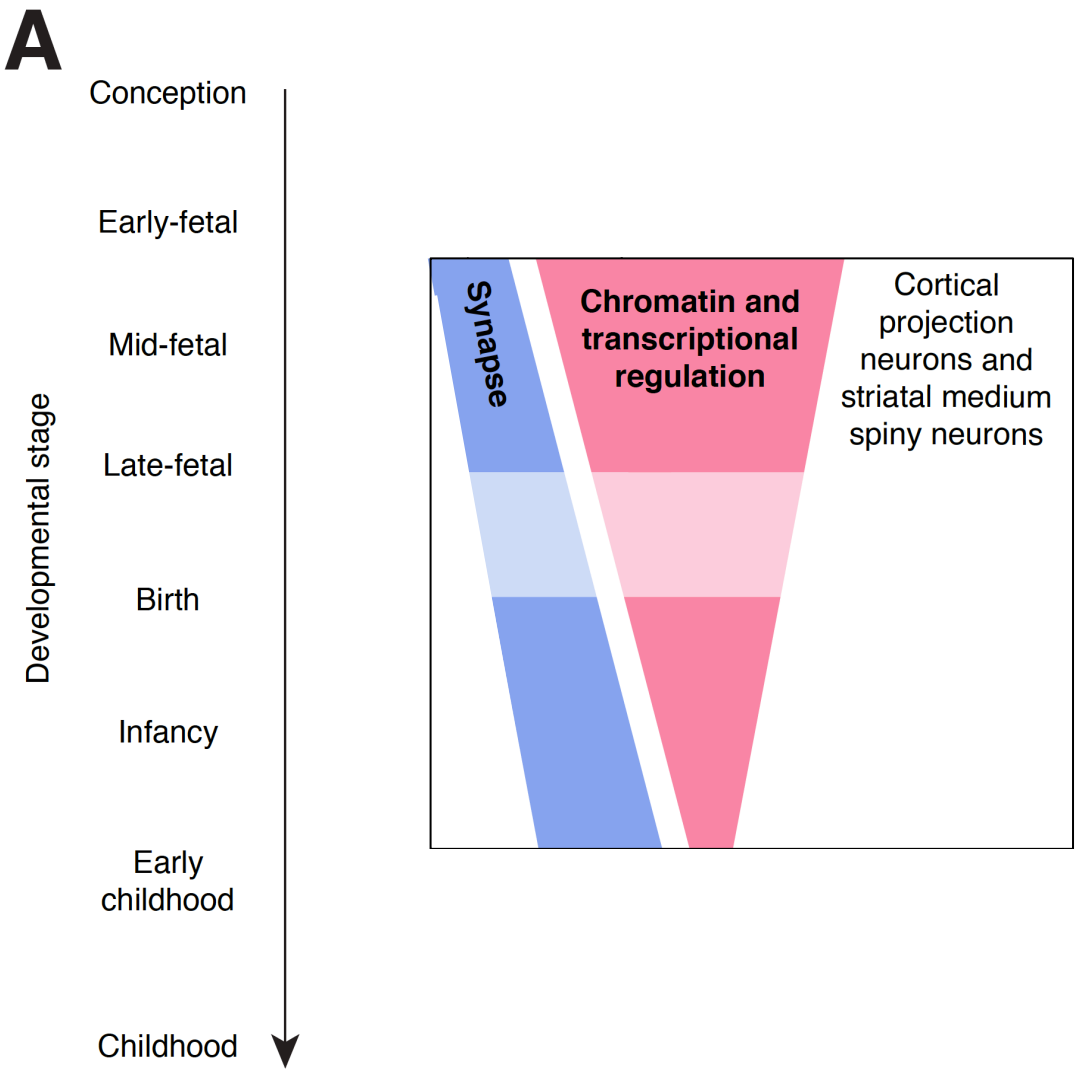


Developmental context



Neural context

The relationship between chromatin and synaptic genes is a key quesiton



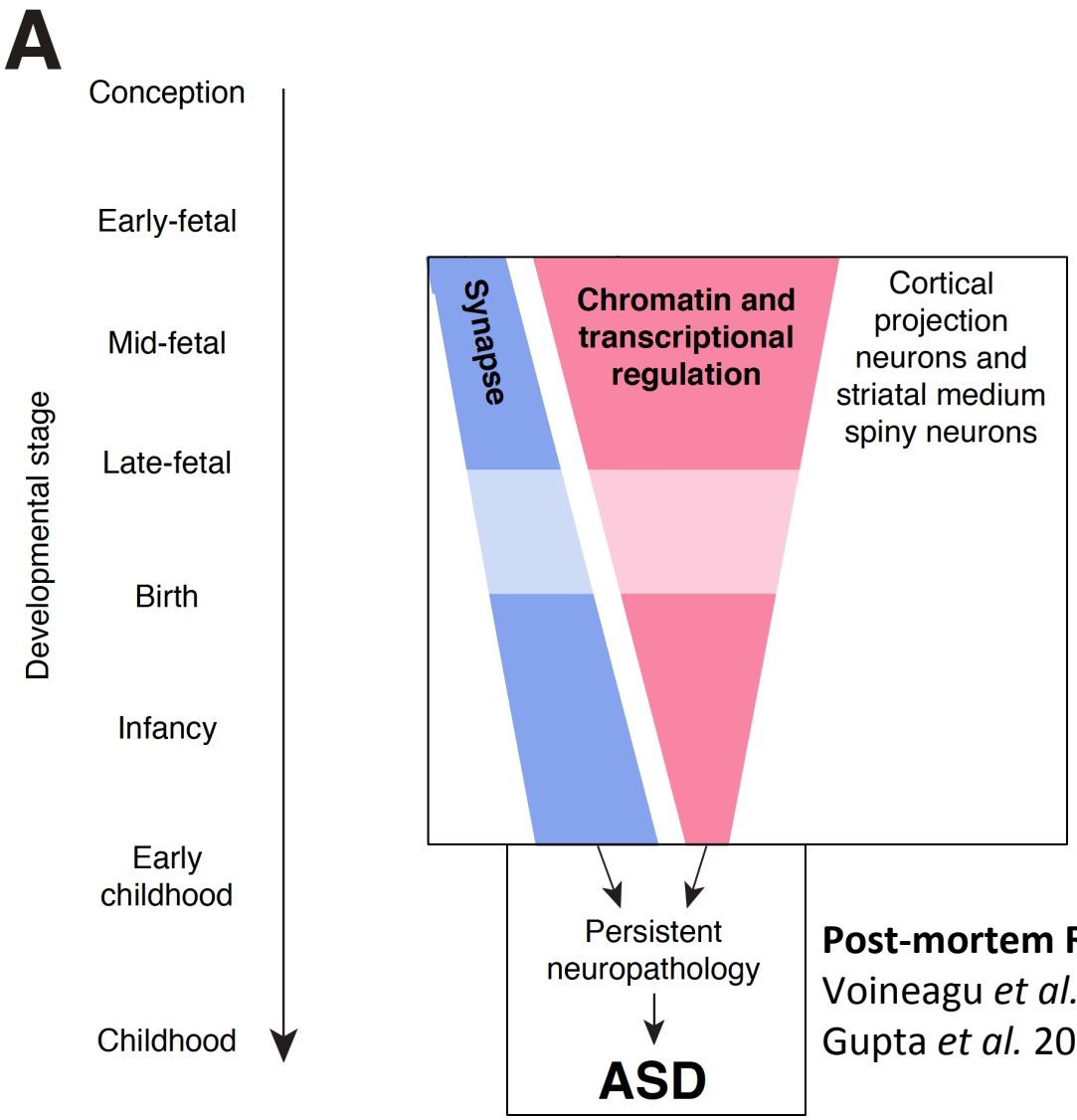
Co-expression
Willsey *et al.* 2013
Parikshak *et al.* 2013

Cell type specific genes
Xu *et al.* 2014

NetBag+ systems analysis
Chang *et al.* 2015

Sanders SJ, Curr Opin Genet Dev, in press

The relationship between chromatin and synaptic genes is a key quesiton



Co-expression
Willsey *et al.* 2013
Parikshak *et al.* 2013

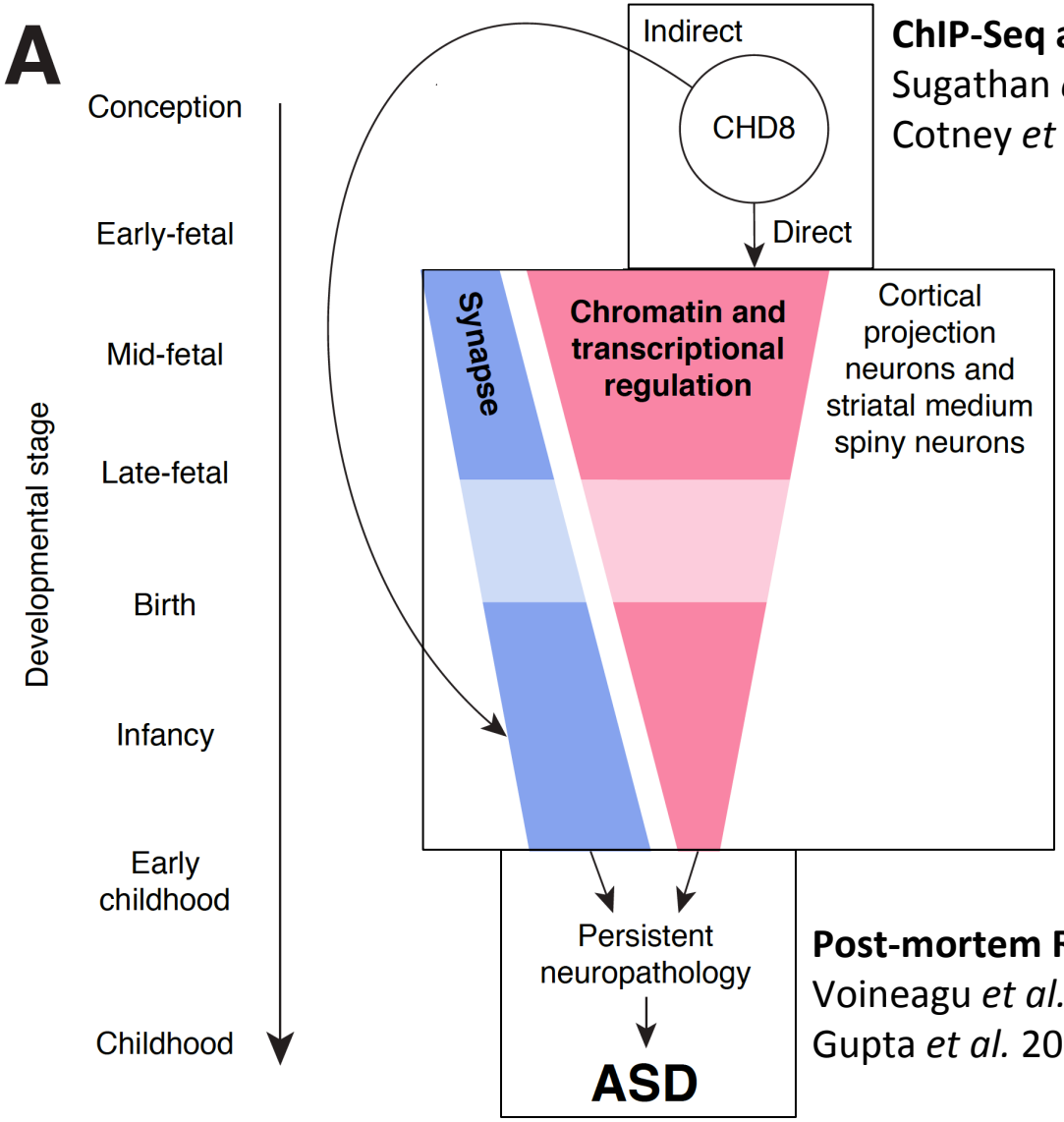
Cell type specific genes
Xu *et al.* 2014

NetBag+ systems analysis
Chang *et al.* 2015

Post-mortem RNA-Seq of ASD brain
Voineagu *et al.* 2011
Gupta *et al.* 2014

Sanders SJ, Curr Opin Genet Dev, in press

The relationship between chromatin and synaptic genes is a key quesiton



ChIP-Seq and RNA-Seq
Sugathan *et al.* 2014
Cotney *et al.* 2015

Co-expression
Willsey *et al.* 2013
Parikshak *et al.* 2013

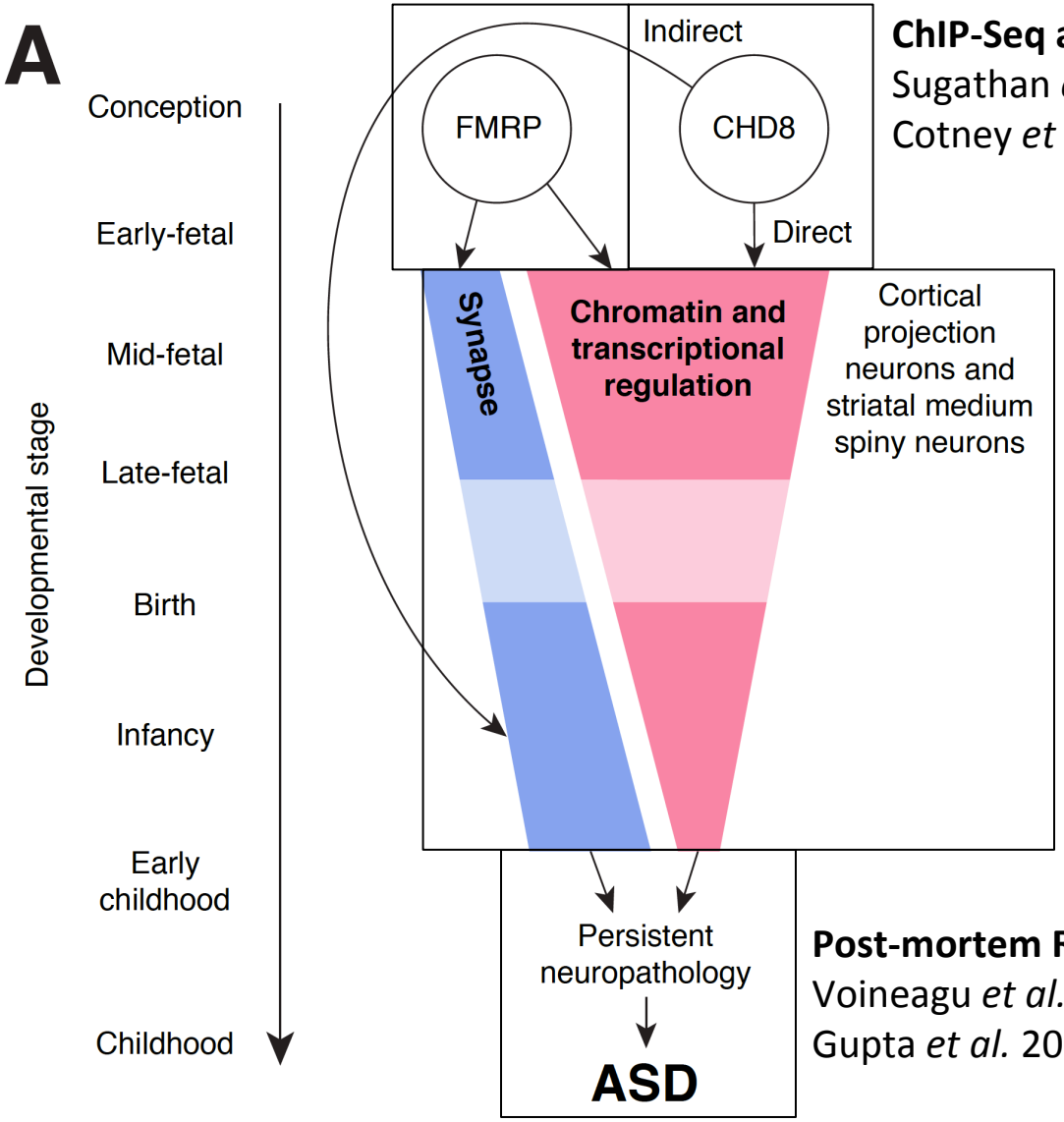
Cell type specific genes
Xu *et al.* 2014

NetBag+ systems analysis
Chang *et al.* 2015

Post-mortem RNA-Seq of ASD brain
Voineagu *et al.* 2011
Gupta *et al.* 2014

Sanders SJ, Curr Opin Genet Dev, in press

The relationship between chromatin and synaptic genes is a key quesiton



ChIP-Seq and RNA-Seq
Sugathan *et al.* 2014
Cotney *et al.* 2015

FMRP mRNA targets
Darnell *et al.* 2011
Iossifov *et al.* 2012
Dong *et al.* 2014
Iossifov *et al.* 2014

Co-expression
Willsey *et al.* 2013
Parikshak *et al.* 2013

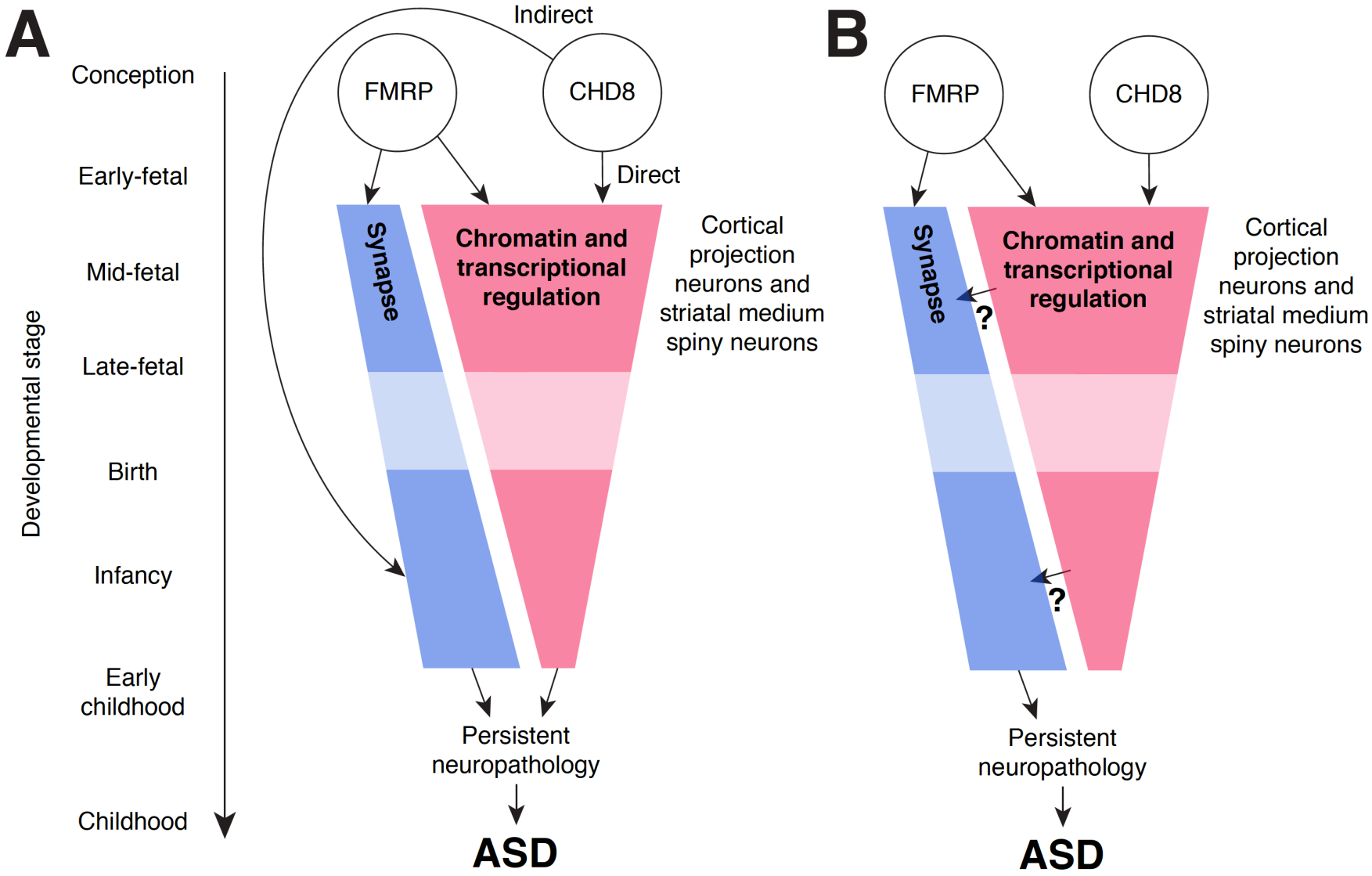
Cell type specific genes
Xu *et al.* 2014

NetBag+ systems analysis
Chang *et al.* 2015

Post-mortem RNA-Seq of ASD brain
Voineagu *et al.* 2011
Gupta *et al.* 2014

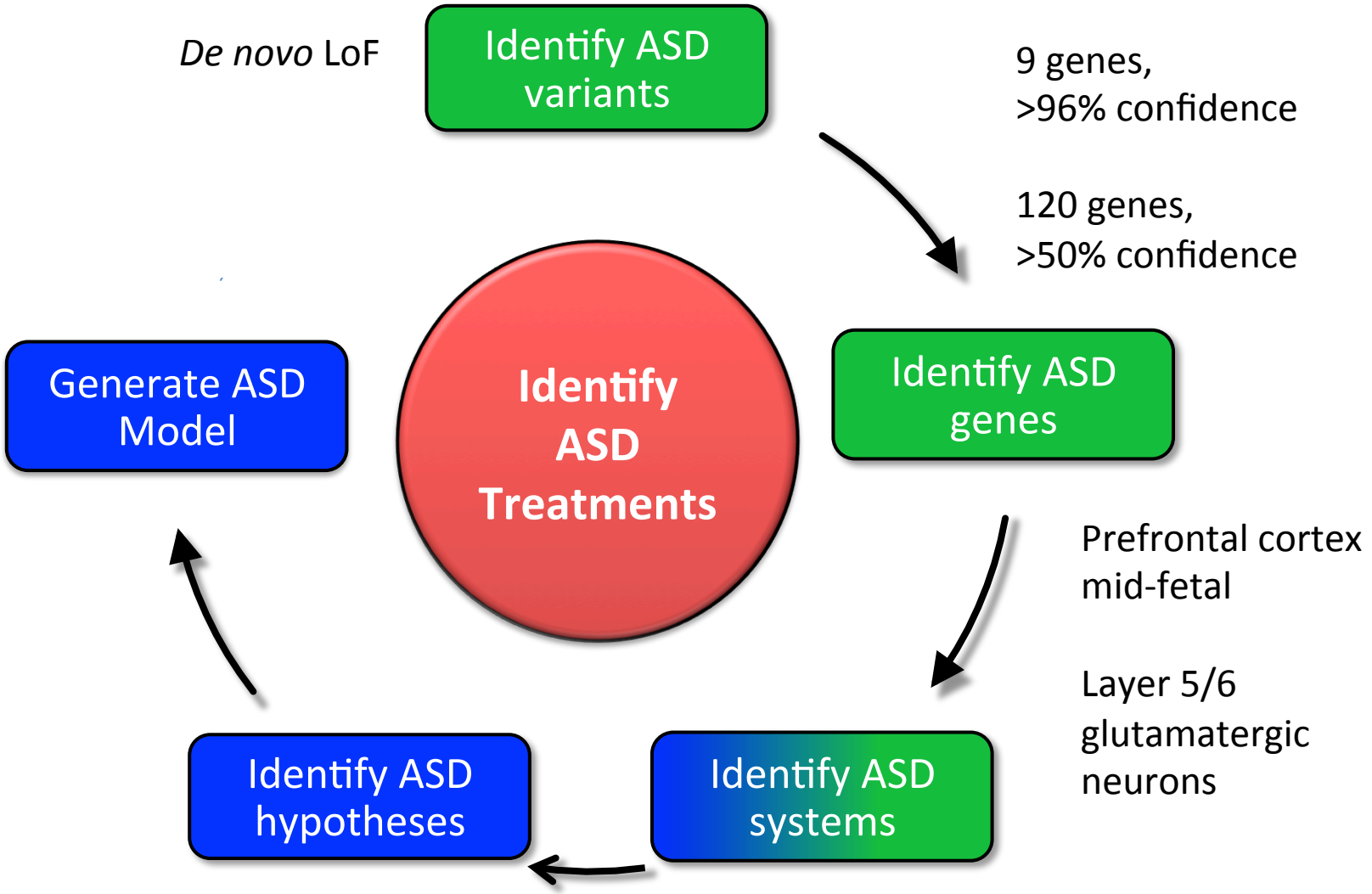
Sanders SJ, Curr Opin Genet Dev, in press

The relationship between chromatin and synaptic genes is a key question



Sanders SJ, Curr Opin Genet Dev, in press

De novo SNVs can be used to identify ASD genes with high confidence



Key points

- Be skeptical about gene associations
  - Is a gene still considered “important” with genomic data
- Most genetic loci are far from 100% penetrant
- Many genetic loci are highly pleiotropic
  - How can we distinguish causal relationships?