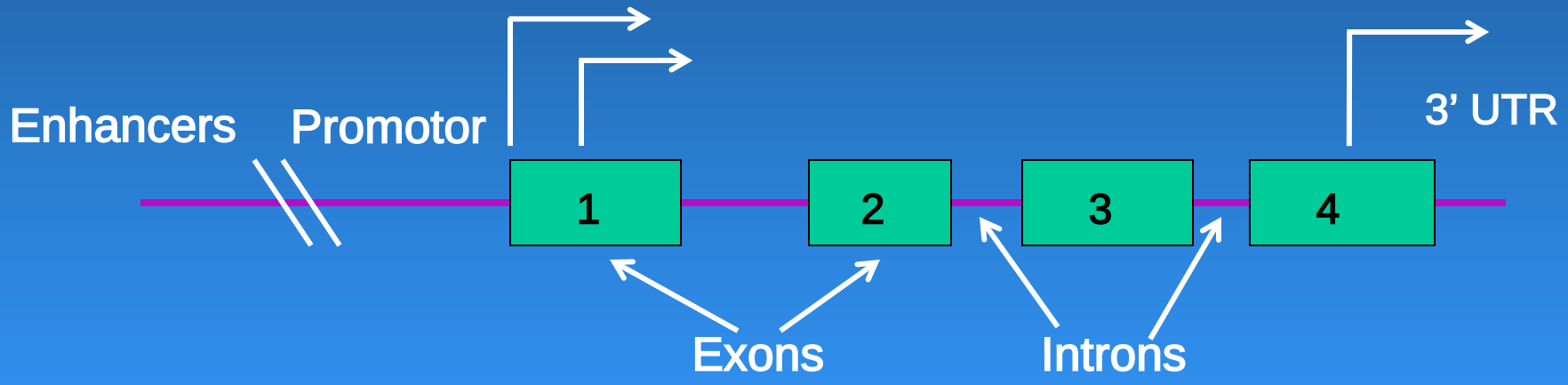
Neurodevelopment and Genetics 2, Current Insights and Future Prospects

Prof W Ted Brown, MD, PhD
University of Sydney,
New York State Institute for Basic Research
In Developmental Disabilities,

Topics

- Genetic analysis
- Genetics of Intellectual Deficiency
- Genetics of Complex Disorders
- Genetics of Autism

What is a gene?



Features of the Genome

Size	~ 3 billion bp
Genes (coding proteins)	~ 22,000
know function	~ 75%
exons / transcript	~ 9
total exons “Exome”	~ 200,000
the “Exome”	~ 50 Mb (1.5%)
highly conserved	~ 170 Mb (5%)
ncRNAs	~ 85% ??

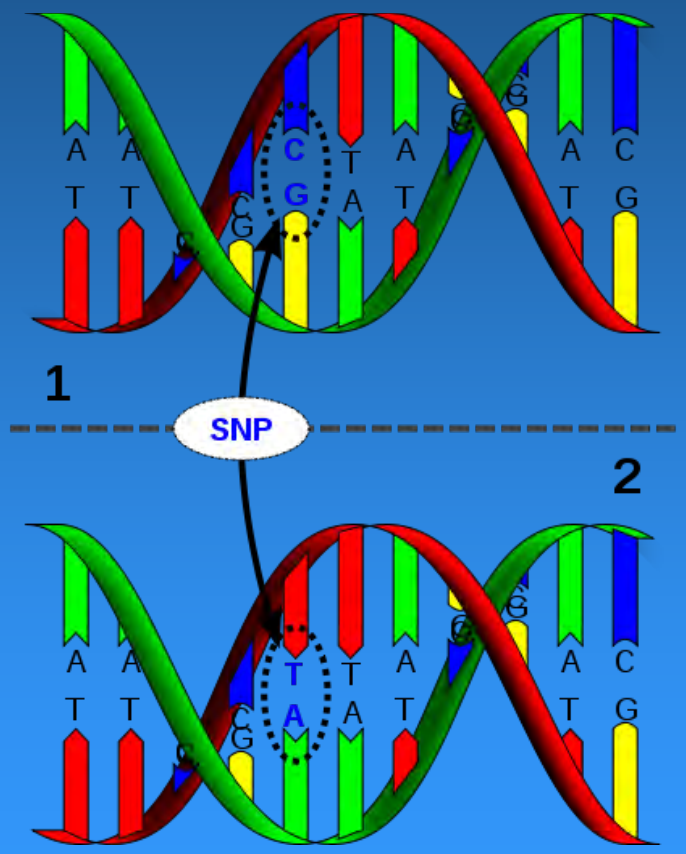
DNA Sequence: Are We All the Same?

- Humans are 99.6% identical at the sequence level
- Evolutionary perspective:
- Homo sapiens - a young species (~100,000 yrs)
- A small founding population (~10,000)
- Similarity with our relatives
- 98.5% identity with Chimpanzee
- ~ 90% identity with Mouse

Sources of Genetic Variation

- Insertion/deletions – “indels” ~ 10%
- Length polymorphisms - STRs ~ 5%
  (short tandem repeats)
- Single nucleotide polymorphisms ~ 4%
  (SNPs)
- Copy number variants (CNVs) ~ 40%

Single Nucleotide Polymorphisms (SNPs)



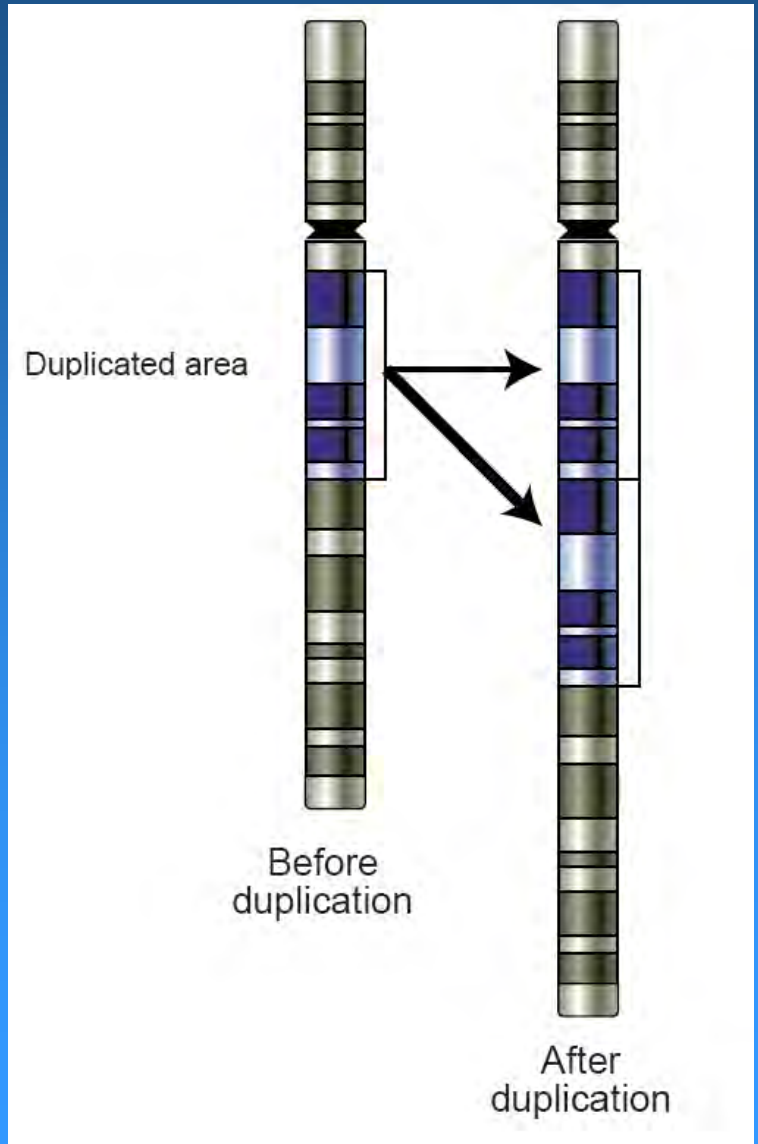
- Single base pair variants with both possibilities relatively frequent (generally >1%)

  allele 1 … G C **C** T A …

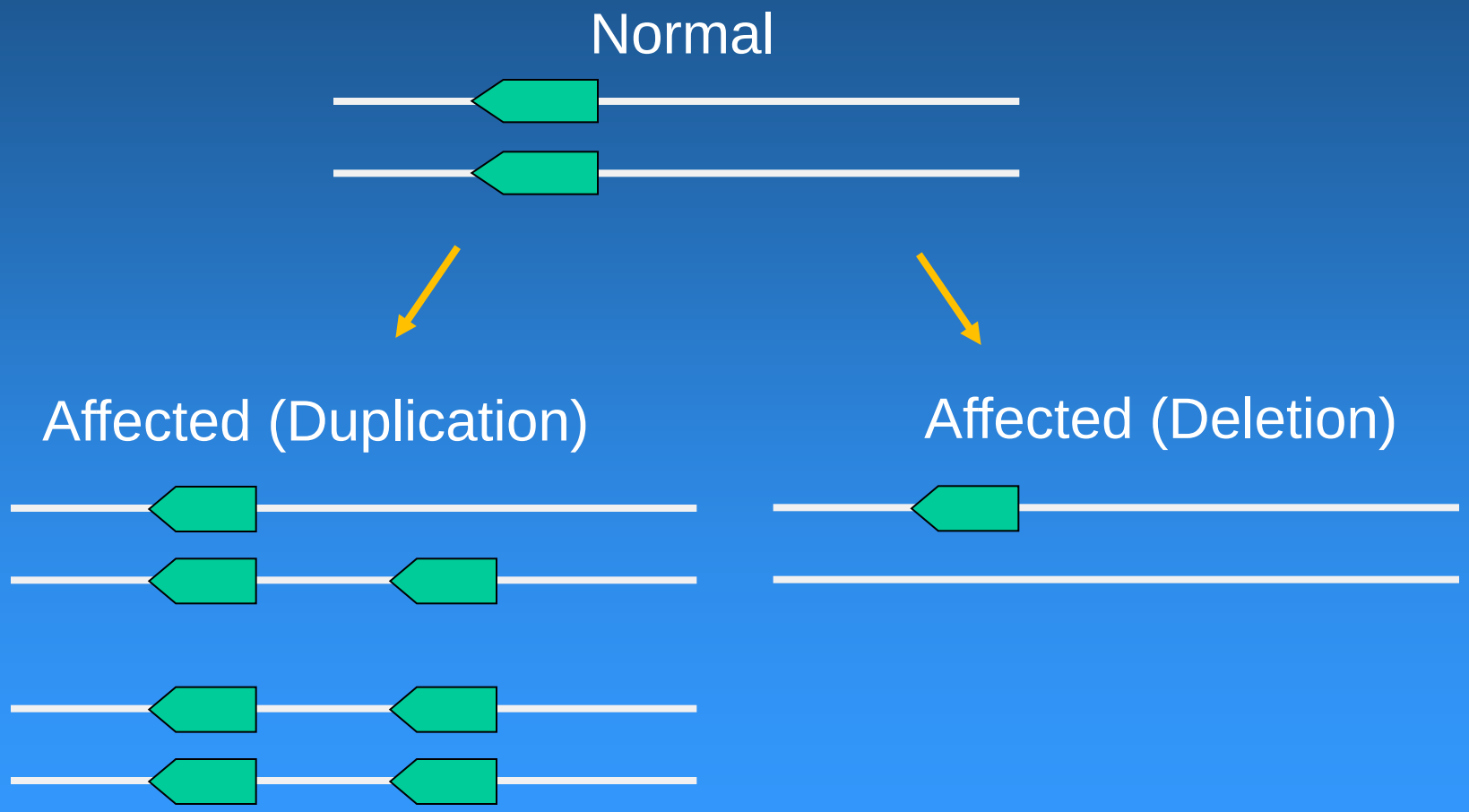
  allele 2 … G C **T** T A …

- Frequent ~ 1/1000 bp or at least 3 M per haploid genome
- More than 50 M now known
- Any two individuals differ by ~3 M
- Current SNP genotyping systems score close to 3 M across the genome.

Copy Number Variants - CNVs



Copy Number Variants - CNVs



Copy Number Variants - CNVs

Increasing appreciation for human genetic variation

- 3 – 7 Average Large (10-100 kb) CNVs / person
- 5 – 10% of persons have 1 CNV > 100 kb (~1 gene)
- 1 – 2% have 1 CNV > 1 Mb (~7 genes)
- CNVs can expose dosage sensitive genes
- Can create “fusion” genes with new functions
- Deletions can expose otherwise normal variation on the remaining allele

Non-coding RNAs (ncRNAs)

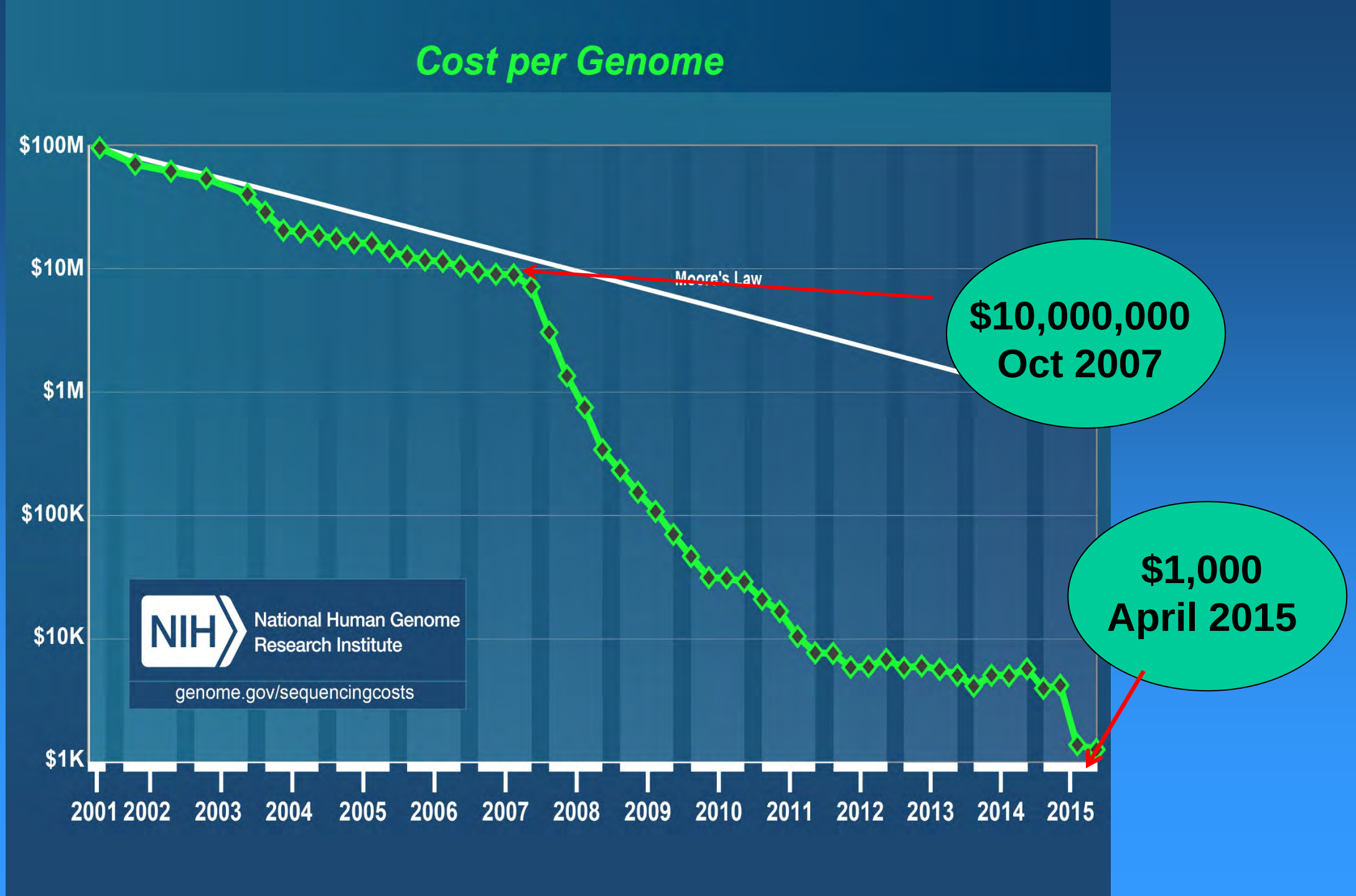
- Short ncRNAs 7-30 bps (miRNA, piRNA)
- Mid-sized ncRNAs 30-200 bps
- Long ncRNAs 200-1000+ bps

Non-coding RNAs

Esteller. Nature Reviews Genetics 2011

Name	Size	Location	Number in humans	Functions	Illustrative examples	Refs
Short ncRNAs						
miRNAs	19–24 bp	Encoded at widespread locations	>1,424	Targeting of mRNAs and many others	miR-15/16, miR-124a, miR-34b/c, miR-200	3–6
piRNAs	26–31bp	Clusters, Intragenic	23,439	Transposon repression, DNA methylation	piRNAs targeting *RASGRF1* and LINE1 and IAP elements	13–19
tiRNAs	17–18bp	Downstream of TSSs	>5,000	Regulation of transcription?	Associated with the *CAP1* gene	37
Mid-size ncRNAs						
snoRNAs	60–300 bp	Intronic	>300	rRNA modifications	U50, SNORD	20–22
PASRs	22–200 bp	5′ regions of protein-coding genes	>10,000	Unknown	Half of protein-coding genes	10
TSSa-RNAs	20–90 bp	−250 and +50 bp of TSSs	>10,000	Maintenance of transcription?	Associated with *RNF12* and *CCDC52* genes	35
PROMPTs	<200 bp	−205 bp and −5 kb of TSSs	Unknown	Activation of transcription?	Associated with *EXT1* and *RBM39* genes	36
Long ncRNAs						
lincRNAs	>200 bp	Widespread loci	>1,000	Examples include scaffold DNA–chromatin complexes	*HOTAIR*, *HOTTIP*, *lincRNA-p21*	2,28–30
T-UCRs	>200 bp	Widespread loci	>350	Regulation of miRNA and mRNA levels?	uc.283+, uc.338, uc160+	31–34
Other lncRNAs	>200 bp	Widespread loci	>3,000	Examples include X-chromosome inactivation, telomere regulation, imprinting	*XIST*, *TSIX*, TERRAs, *p15AS*, *H19*, *HYMAI*	2,23–25

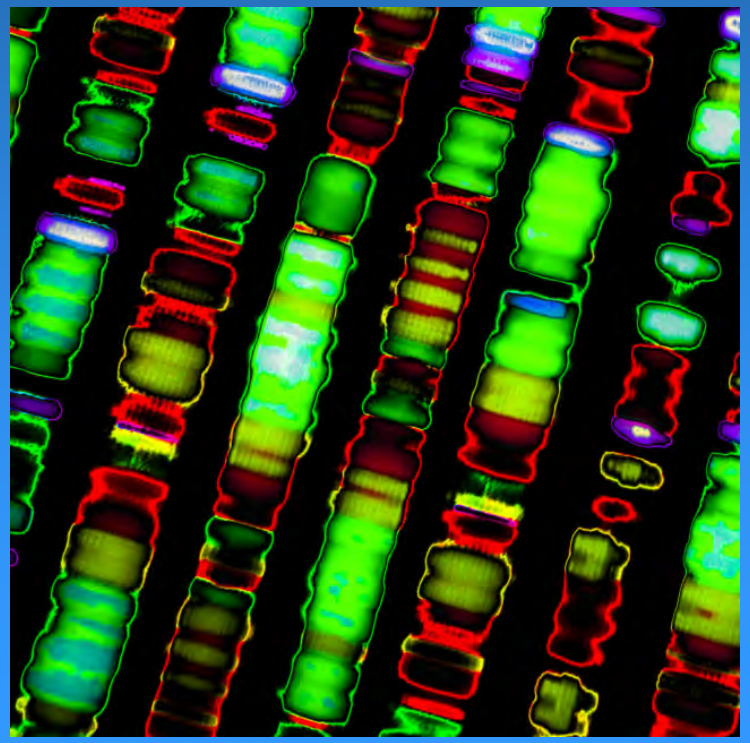
*There is not necessarily a clear delineation between classes of non-coding RNA (ncRNA); for example, X-inactivation specific transcript (*XIST*) and its antisense transcript *TSIX* could be considered as large intergenic non-coding RNAs (lincRNAs). In the 'Location' column, '−' represents the number of base pairs upstream of the transcription start site (TSS) and '+' represents the number of base pairs downstream of the TSS. *CAP1*, CAP, adenylate cyclase-associated protein 1; *CCDC52*, coiled-coil domain containing 52 (also known as *SPICE1*); *EXT1*, exostosin 1; *HOTAIR*, homeobox (HOX) transcript antisense RNA; *HOTTIP*, HOXA distal transcript antisense RNA; *HYMAI*, hydatidiform mole associated and imprinted; IAP, intracisternal A-particle; lncRNA, long non-coding RNA; miRNAs, microRNAs; piRNAs, PIWI-interacting RNAs; PASRs, promoter-associated small RNAs; PROMPTs, promoter upstream transcripts; *RASGRF1*, RAS-protein-specific guanine nucleotide-releasing factor 1; *RBM39*, RNA-binding motif protein 39; *RNF12*, ring finger protein 12 (also known as *RLIM*); snoRNAs, small nucleolar RNAs; TERRAs, telomeric repeat containing RNAs; tiRNAs, transcription initiation RNAs; TSSa-RNAs, TSS-associated RNAs; T-UCRs, transcribed ultraconserved regions.


Cost per Genome
$100M
$10M
$1M
$100K
$10K
$1K
Moore's Law
$10,000,000
Oct 2007
$1,000
April 2015
NIH
National Human Genome
Research Institute
genome.gov/sequencingcosts
2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015

Illumina says it can deliver a $100 genome — soon

BY MEGHANA KESHAVAN STATNEWS.COM JANUARY 9, 2017

The **NovaSeq** system will be three times faster than the previous generation of sequencers, able to sequence 48 human genomes in the time that current technology can sequence16.

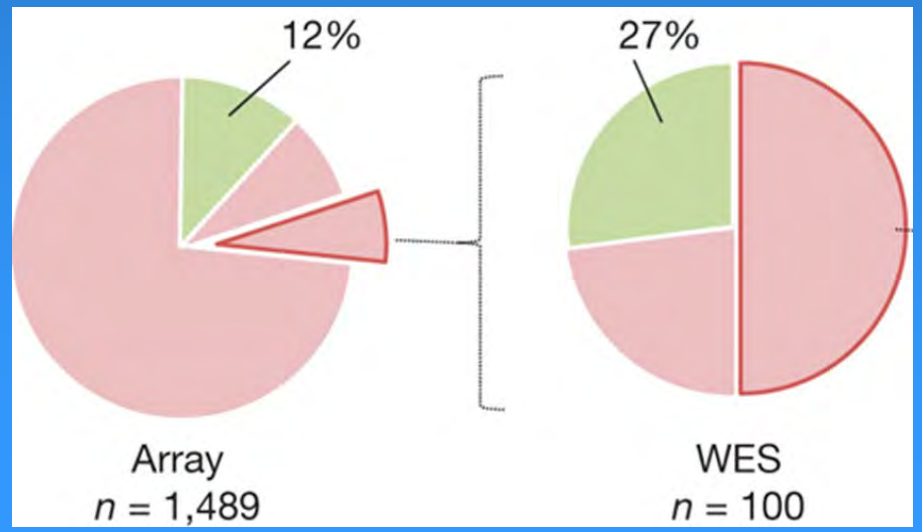


Genetics of Intellectual Disability (ID)

- About 0.5% of the population have an IQ < 50
- Can be due to non-genetic factors: infection, hypoxia
- But in developed countries, most ID is thought to have a genetic cause
- Estimated that mutations in ≥1,000 genes can cause ID
- Diagnostic evaluations: physical examination, metabolic screening, FXS and other targeted gene tests, then:
- Chromosome microarray analysis (CMA), whole exome sequencing (WES), finally whole genome sequencing (WGS)

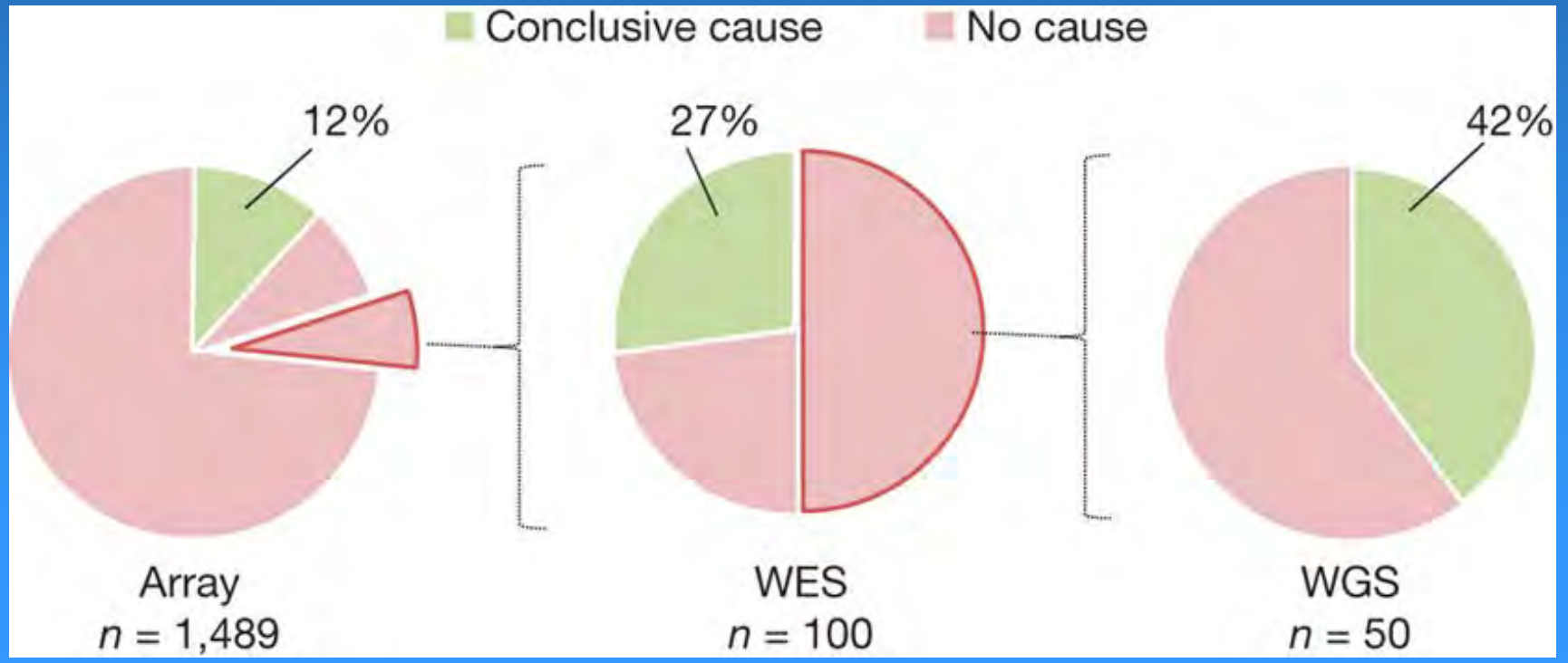
CMA followed by Whole Exome Sequncing for ID

- 1,489 patients had CMA and FX/metabolic screening, which identified 12% as positive. (de Ligt, NEJM 2012)
- 100 of remaining patients with IQ < 50 were studied by WES.
- All studies included both parents (trios).
- 27 were found to have causative mutations and 11 had potential new mutations for 38 total.
- 62 were negative.
- 36% combined yield.

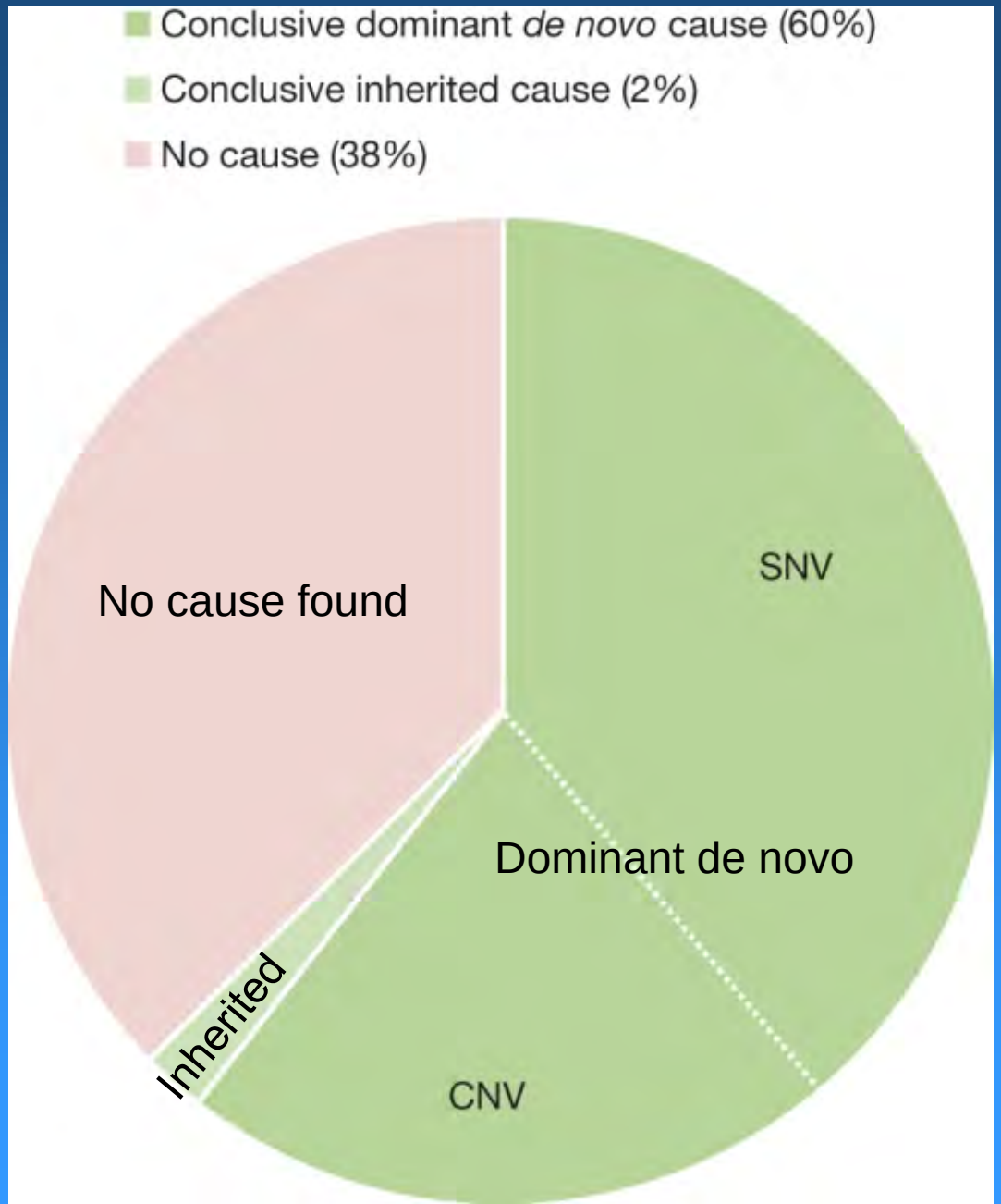


Whole Genome Sequencing for ID

- 50 patients with IQ < 50 were analyzed (Gilissen, Nature 2014)
- 21 positive diagnoses were made (42%) (candidate ID genes 8 -16%)
- Overall yield of CMA+WES+WGS calculated to be 62%.

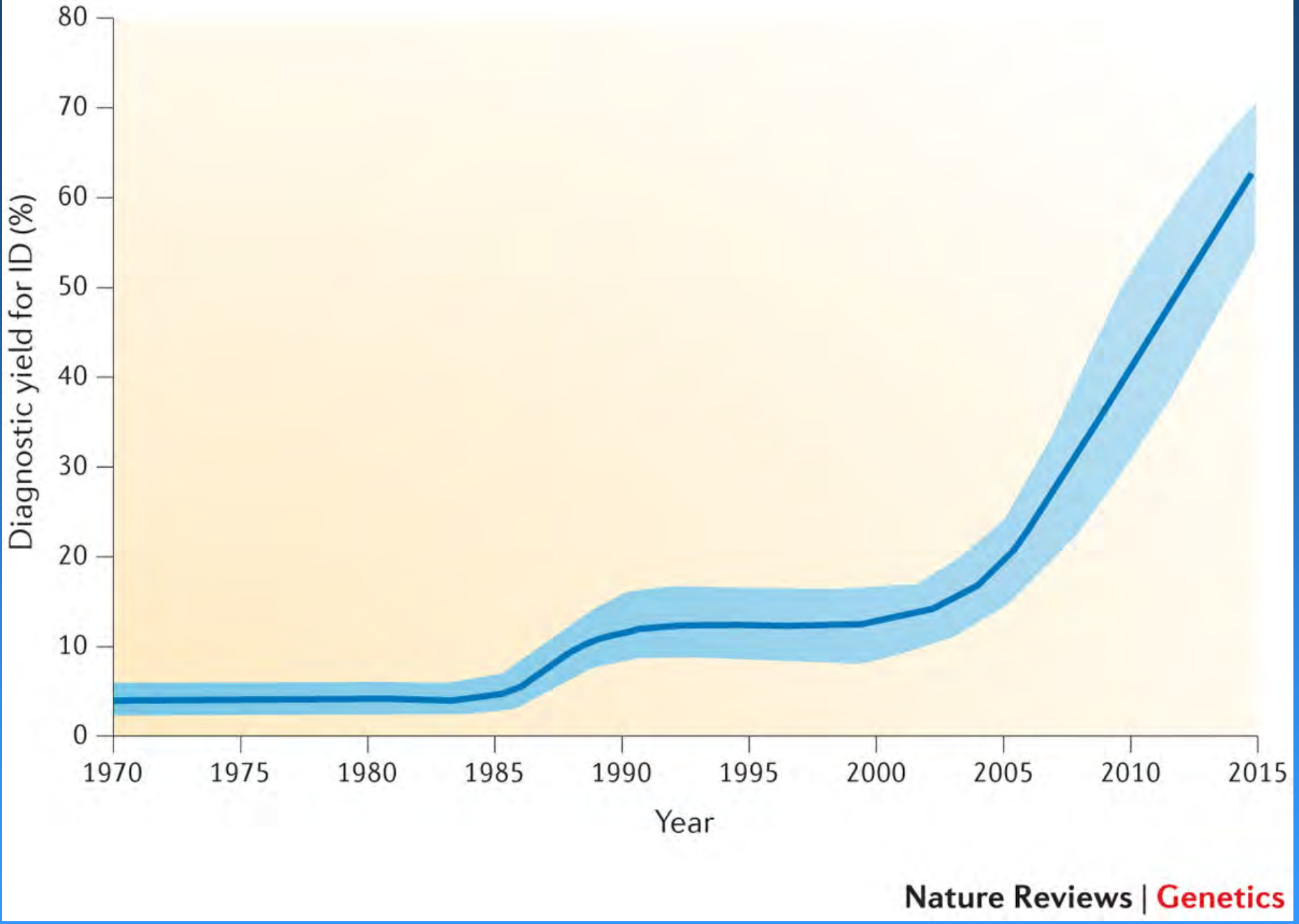


Whole Genome Sequencing for ID



Genome Sequencing identifies major causes of severe intellectual disability.

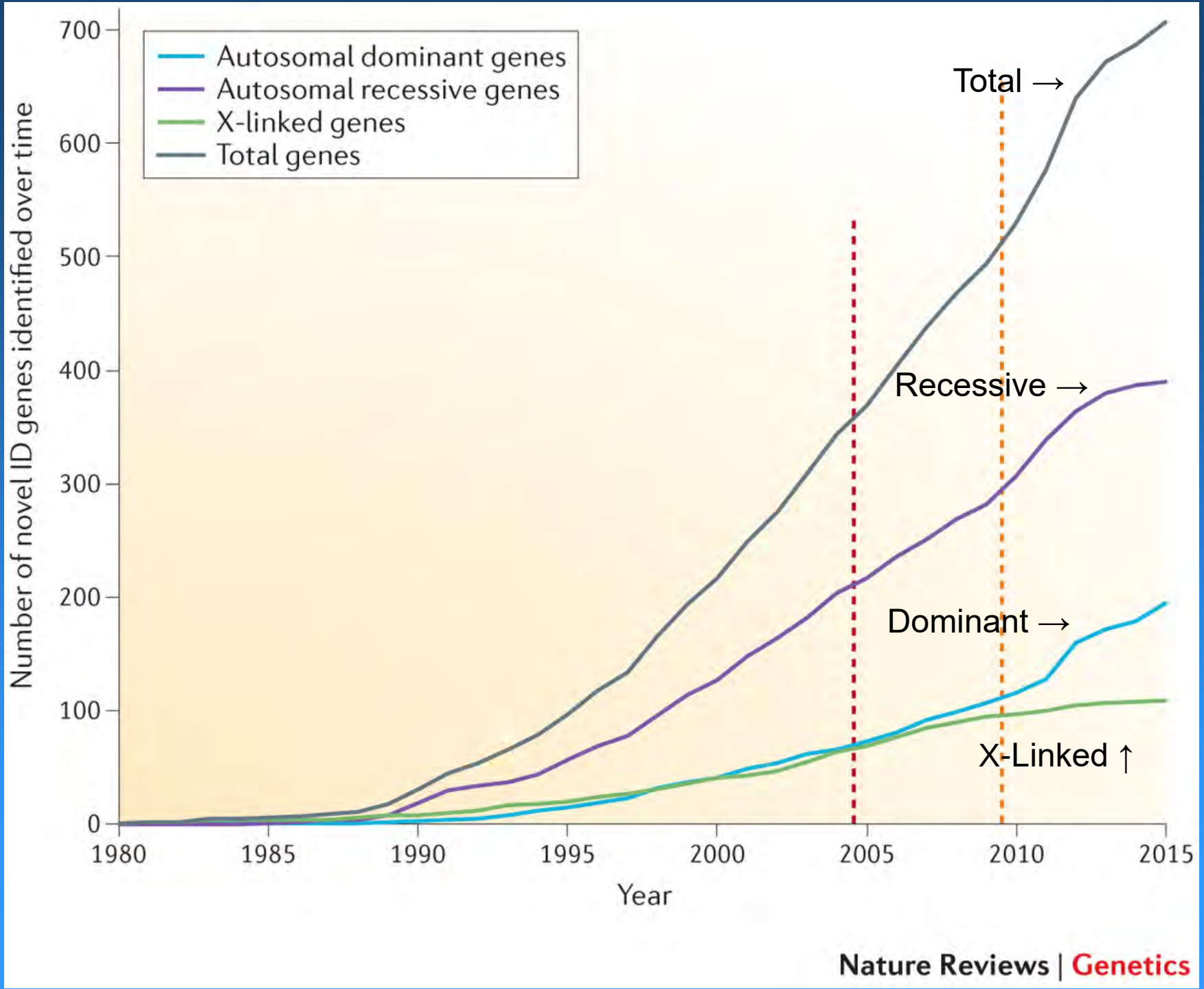
Gilissen et al.
Nature 511: 344 (2014)



Diagnostic Yield Increasing

Genetic studies in intellectual disability and related disorders

Vissers et al. Nature Reviews Genetics 17: 9-18 (2016)

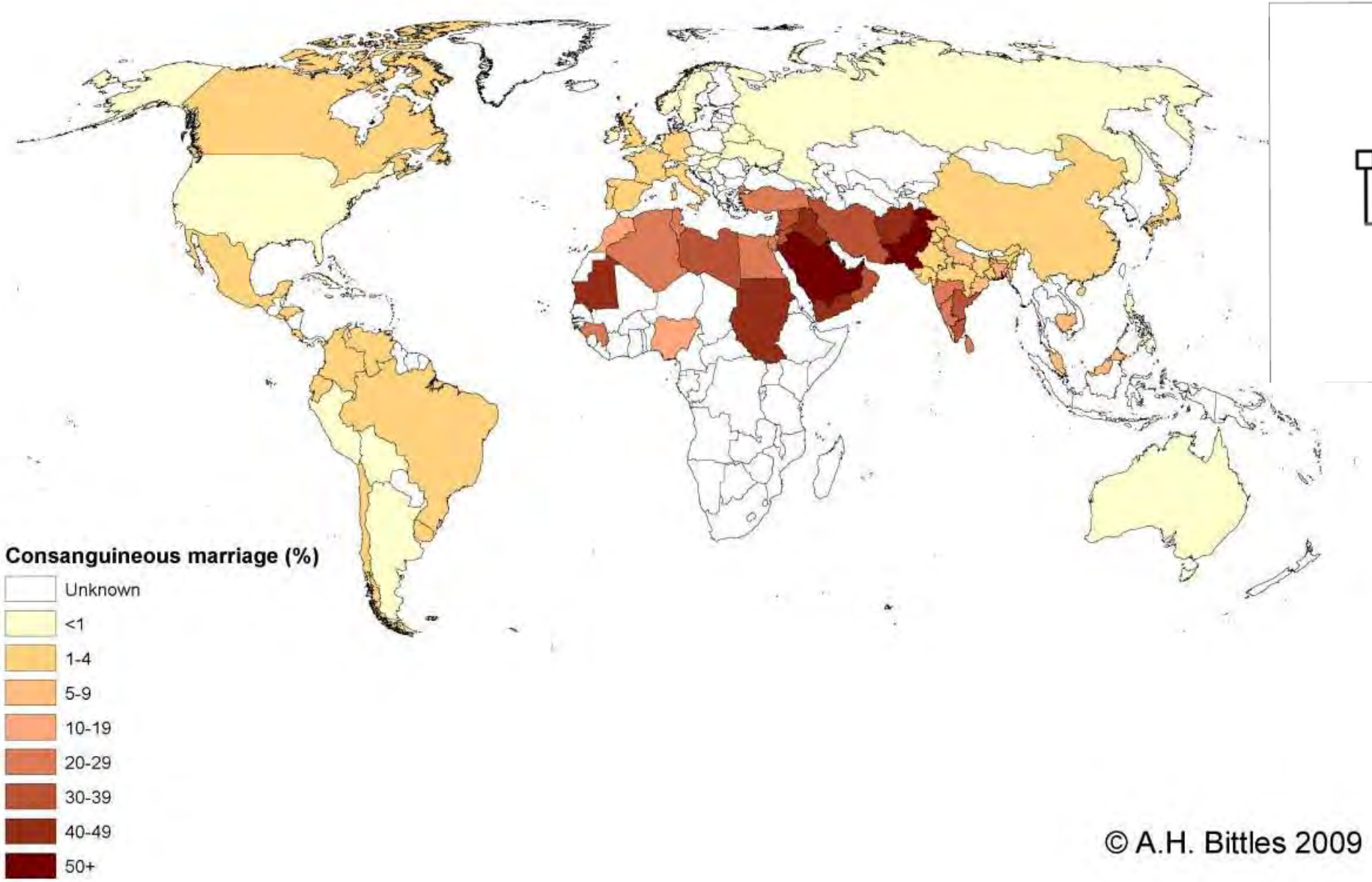


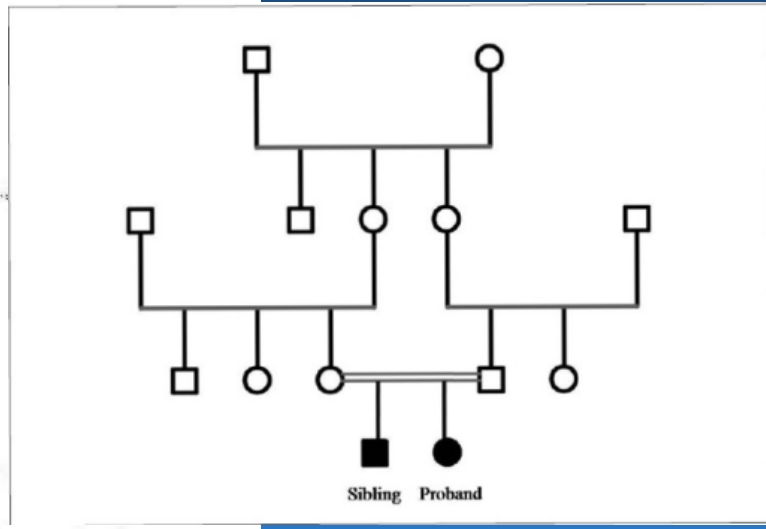
700+ ID Genes identified

Genetic studies in intellectual disability and related disorders.

Vissers et al.
Nature Reviews Genetics
17: 9-18 (2016)

Consanguinity Belt





Genetics of Recessive Intellectual Disabilities

- About 100 X-linked recessive genes have been identified.
- About 400 autosomal recessive genes identified.
- Homozygosity mapping in consanguineous families has been the strategy of choice for identifying recessive genes.
- Next Generation Sequencing now becoming increasingly used.
- Ropers analyzed 136 Iranian consanguineous families with ID, finding 50 new genes. Harripaul added 26 new genes from 192 Pakistani families.

Najmabadi & Ropers et al. (2011) Deep sequencing reveals 50 novel genes for recessive cognitive disorders.

Harripaul et al. (2017) Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families.

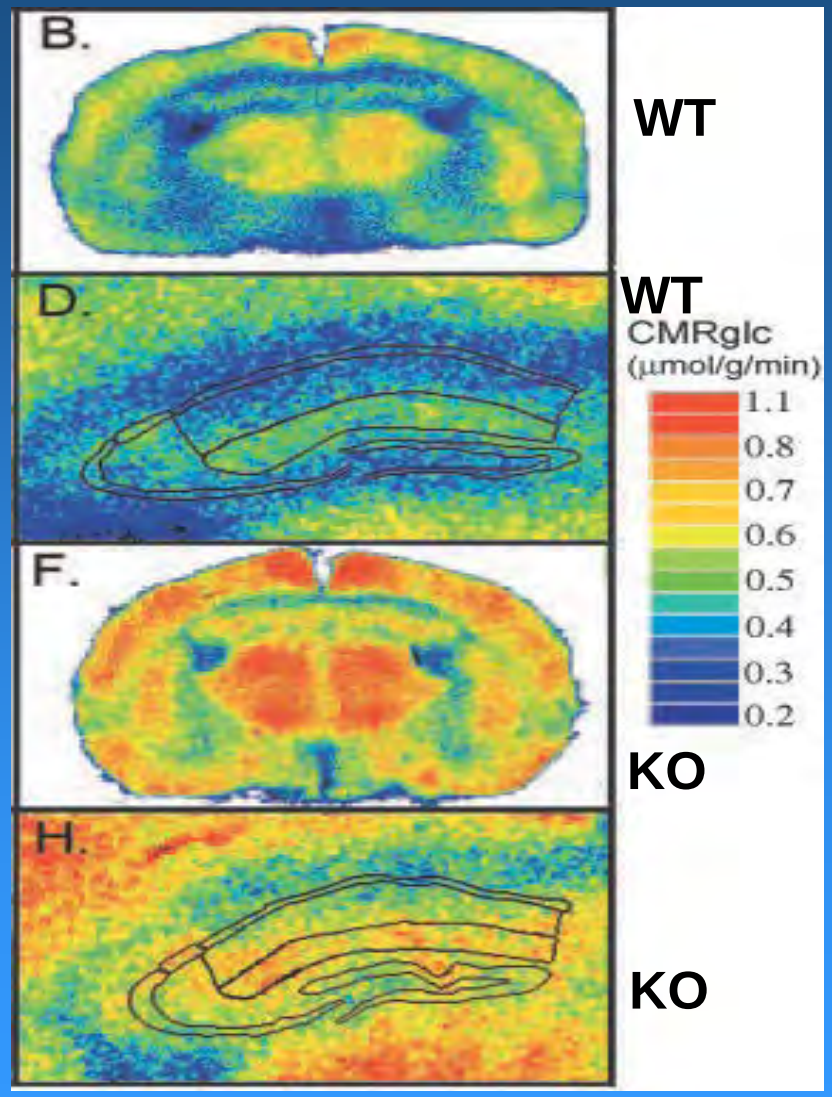
Some Genetic Disorders associated with Autism

• Smith-Magenis (del 17p11.2)	~90%
• San Filippo syndrome	~90%
• Phelan-McDerm (del 22q13.3)	~75%
• Fragile X syndrome	~50%
• Dup 15q11-15 syndrome	~50%
• Angelman syndrome	~40%
• Tuberous Sclerosis	~25%
• Prader-Willi syndrome	~25%
• VCF/ DiGeorge (del 22q11)	~25%
• Down syndrome	~10%

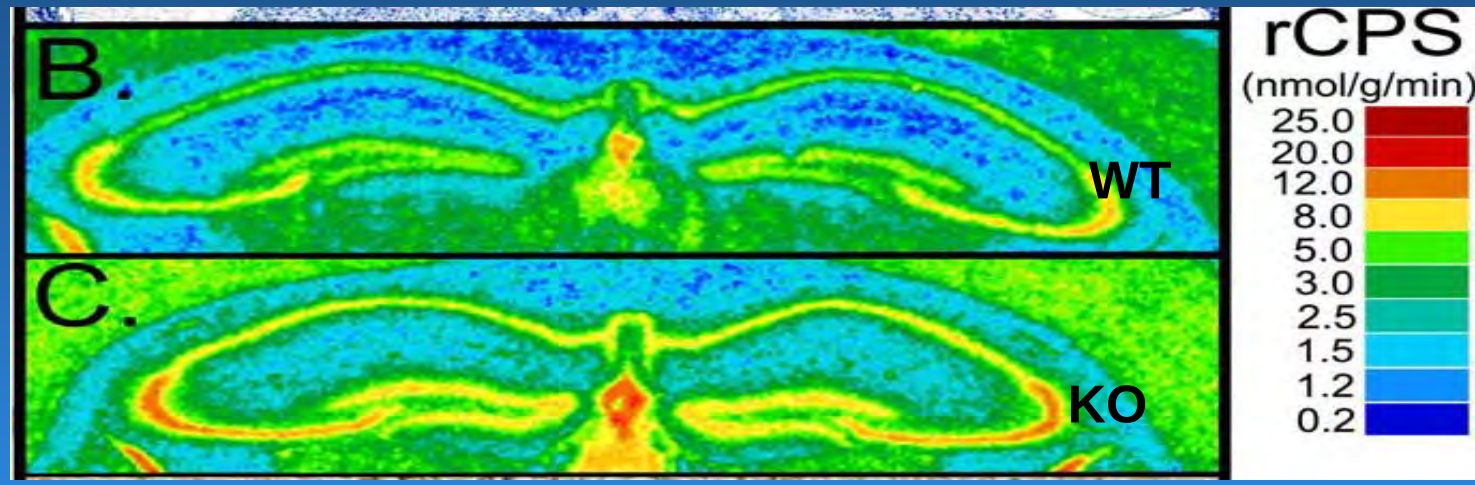
The Fragile X Syndrome

- The most common cause of inherited intellectual disability.
- This common intellectual disability syndrome is due to the silencing of the *FMR1* gene on the X chromosome.
- The *FMR1* protein (FMRP) is an RNA binding protein.
- It modulates the expression of ~5% of expressed brain proteins.

FMR1 KOs have Increased Regional Cerebral and Hypocampal Protein Synthesis



Qin M, PNAS 2002

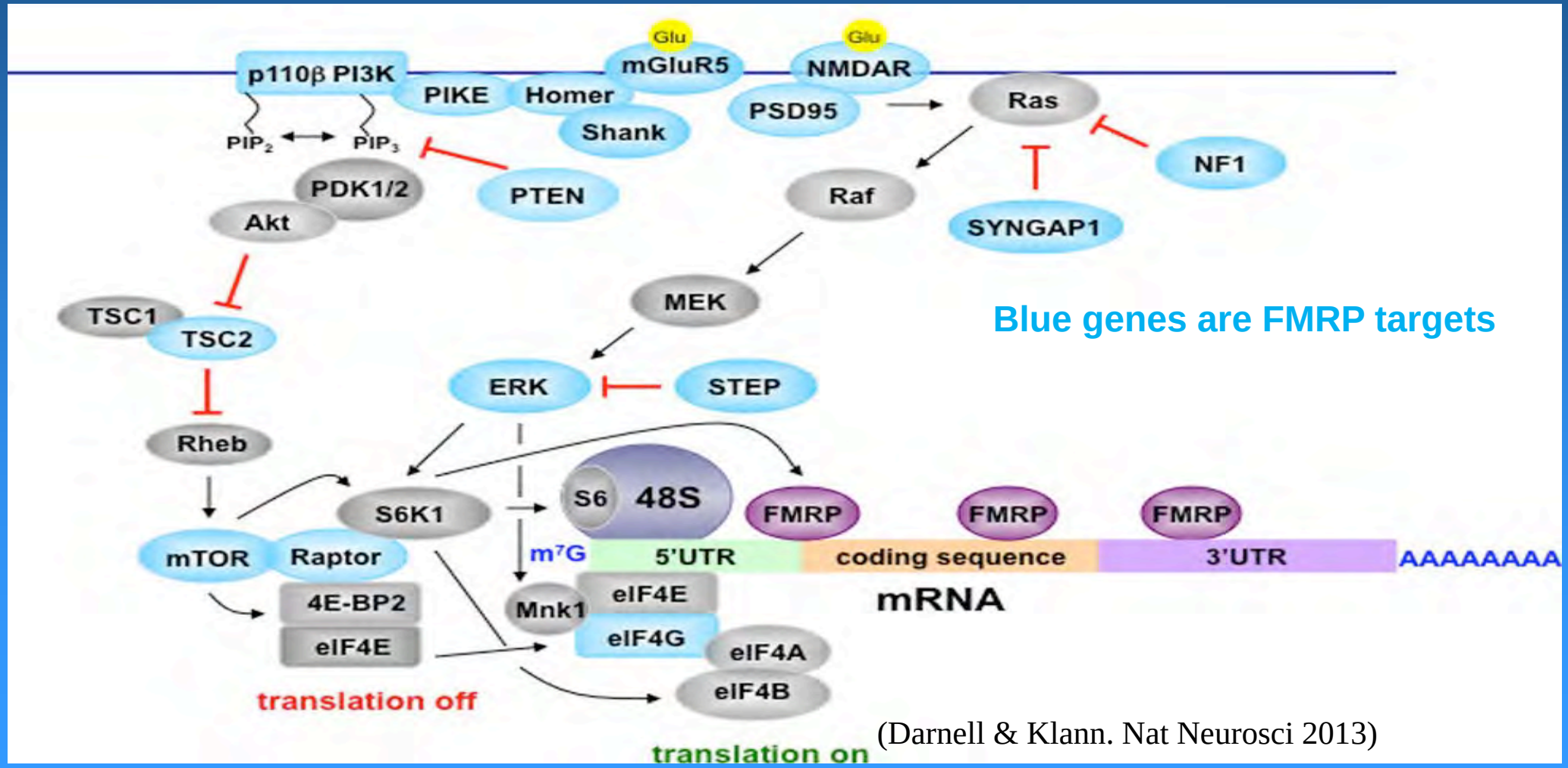


Qin M et al., J NS 2005

FMRP is an mRNA transport protein that regulates the translation of many genes at the synapse

- Neuroligins
- Neurorexins
- SHANK3
- PTEN
- TSC2
- NF1
- Many others leading to down regulation of PTEN and up regulation of mTOR

FMRP is an mRNA transport protein that regulates the translation of many genes at the synapse (Darnell 2011)



(Darnell & Klann. Nat Neurosci 2013)

842 Targets of FMRP were identified by a sensitive assay (Darnell, Cell 2011).

- The overlap with the SFARI database of candidate autism genes was highly significant (~1/3), including
- Neuroligans and Neurorexins
- SHANK3
- PTEN
- TSC2
- NF1
- Many others involved with synapses

SFARI GENE DATABASE

- SFARI Gene is an evolving genetic database for the autism research community. Gene.SFARI.org
- It is focused on genes implicated in autism susceptibility.
- The SFARI Gene web portal integrates different kinds of genetic data generated by research studies.

SFARI Gene - Autism Data

https://gene.sfari.org/autdb/Welcome.do

Quick Links

- SFARI.org
- SFARI Gene
- Human Gene
- Animal Models
- PIN
- CNV
- Gene Scoring

TOOLS

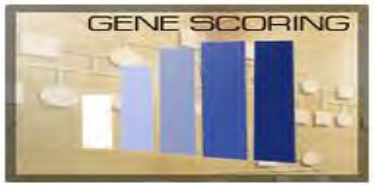
- Advanced Search
- User guide
- Submit New Gene
- Workspace
- Pending Updates

SFARI GENE Home

A Modular Database for Autism Research

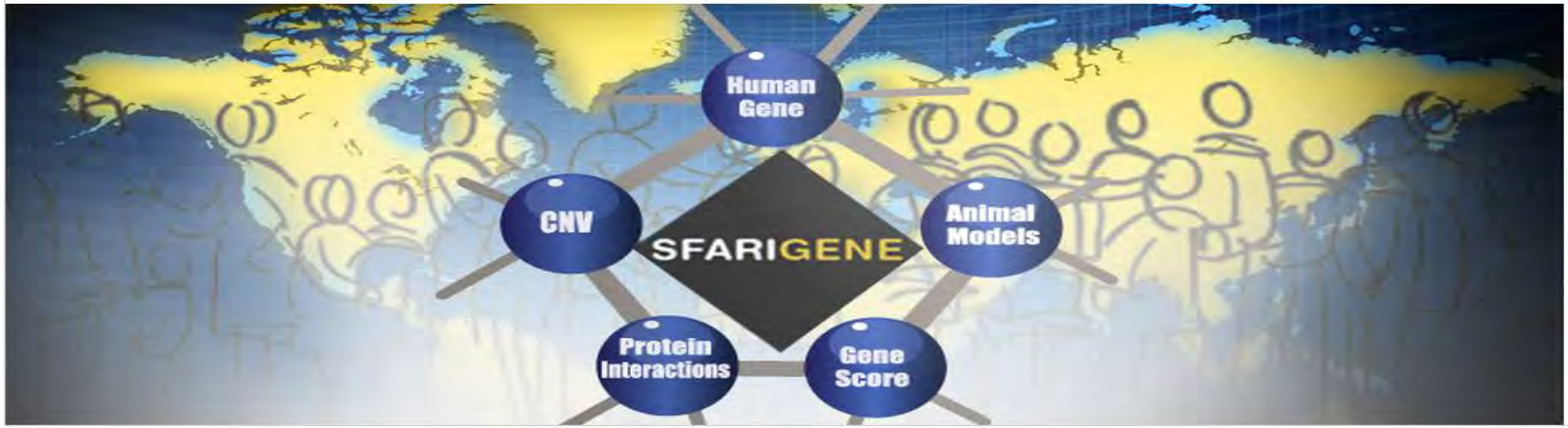
GENE.SFARI.org

ABOUT | NEWS | TIMELINE | HELP



Latest News → articipation of the autism research community. May 15, 2012

SFARI Gene is an integrated resource for the autism research community. It is a publicly available, curated, web-based, searchable database for autism research. This resource is built on information extracted from the studies on molecular genetics and biology of Autism Spectrum Disorders (ASD). The genetic information includes data from linkage and association studies, cytogenetic abnormalities, and specific mutations associated with ASD.[Read More]



GENE.SFARI.org

SFARI Gene - Gene Scoring Module - Windows Internet Explorer provided by New York State OPWDD

https://gene.sfari.org/autdb/GS_Home.do

File Edit View Favorites Tools Help

Favorites CUNY SUNY SFS Sign-in Suggested Sites Line Breaks OPWDD-FB Prism

OPWDD Homepage | Vegetarian Guide -- Recipes,... | PubMed Central, Figure 1: B... | SFARI Gene - Gene Scori...

- PIN
- CNV
- Gene Scoring

TOOLS

- Advanced Search
- User guide
- Submit New Gene
- Workspace
- Pending Updates

Community - Wide Annotation

To encourage participation in our effort to assess the evidence for autism risk genes, we have implemented a process for community-based gene scoring. Tow short tutorials will guide users through the site layout (Tutorial 1) and the scoring procress (Tutorial 2). To contibute your own gene scores, click here.

Category S	FMR1	CACNA1C	Syndromic
Category 1			High Confidence
Category 2	NRXN1	MET	Strong Candidate
Category 3	AUTS2	FOXP2	Suggestive Evidence
Category 4	DPP6	DLGAP2	Minimal Evidence
Category 5	DAB1	DIAPH3	Hypothesized
Category 6	APC	TPH2	Not Supported

Click on category name to obtain a list of scored genes.

Internet | Protected Mode: On 110%

GENE.SFARI.org

- Syndromic Genes – 97
- Genes with High Confidence – 23
- Genes are Strong Candidates – 42
- Genes with Suggestive Evidence – 149
- Genes with Minimal Evidence – 256
- Genes Hypothesized not Tested – 135

- Total Number of Autism Genes – 702

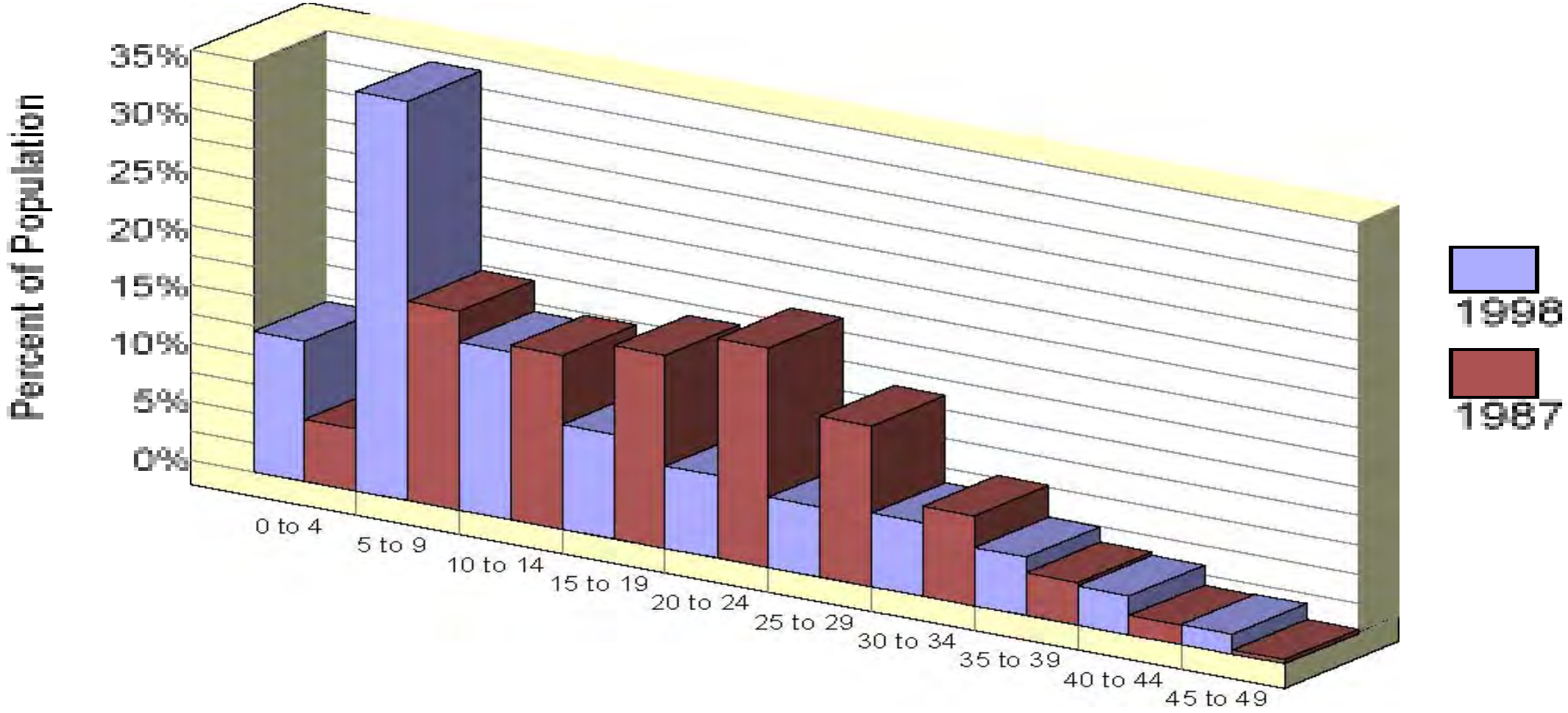
Epidemiology of Autism

- Rates of autism are increasing
- The reasons are unclear
  - Better diagnosis?
  - Broader diagnosis?
  - Better programs?
  - True increase?

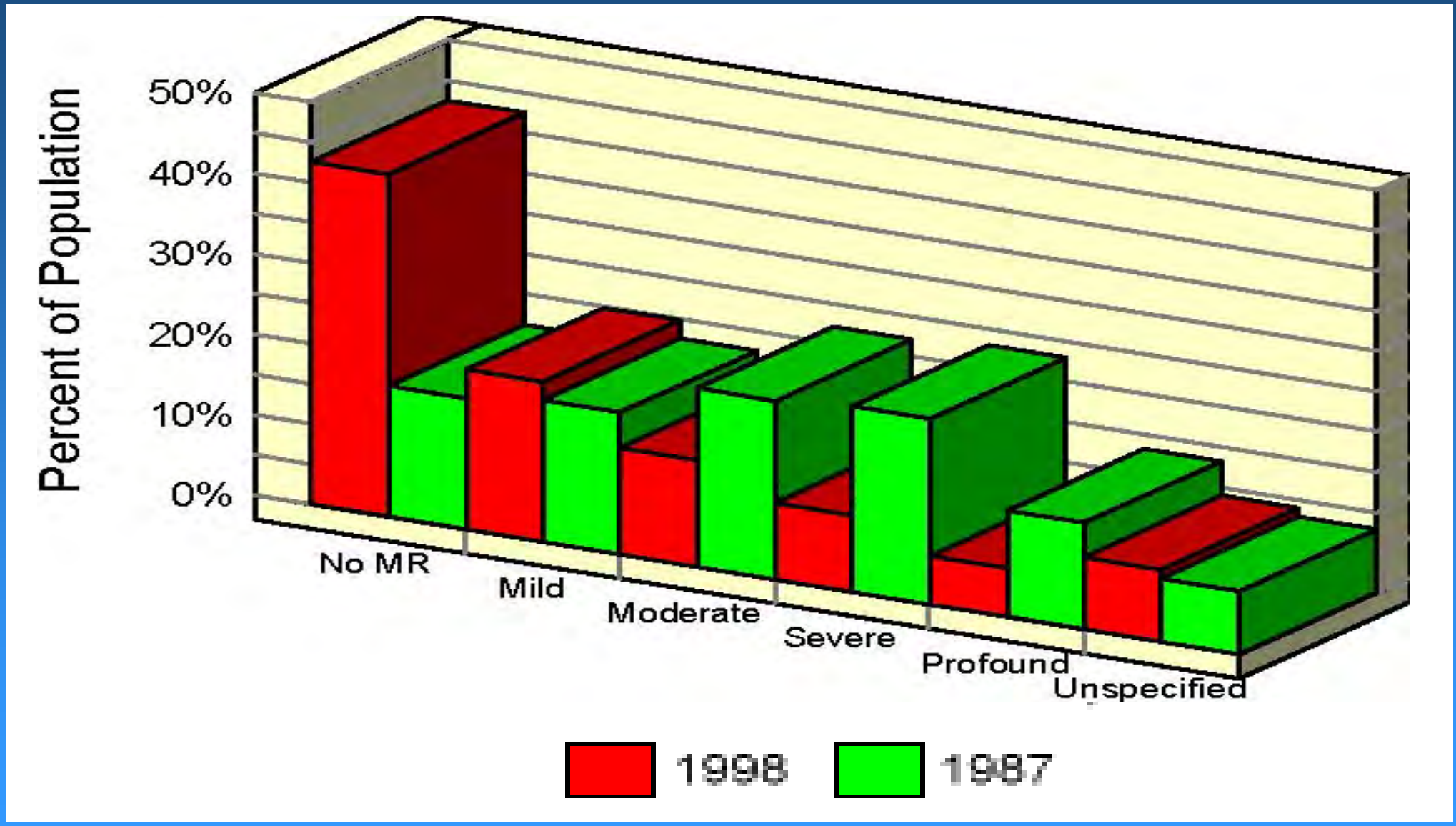
How Common is Autism?

- The general observation is of a dramatic increase.
- The standard figure up to about 1980 was 1 in 2000.
- Since then, a consistent increase has been seen.
- CDC (2007) → 6.6/1000 or 1 in 150
- CDC (2009) → 9.0/1000 or 1 in 110
- CDC (2012) → 11.3/1000 or 1 in 88
- CDC (2014, 16) → 14.6/1000 or **1 in 68**
- Sex ratio 4:1 → about **1 in 42 boys, 1 in 189 girls**

Autistic Age Distribution:
1987 *vs.* 1998

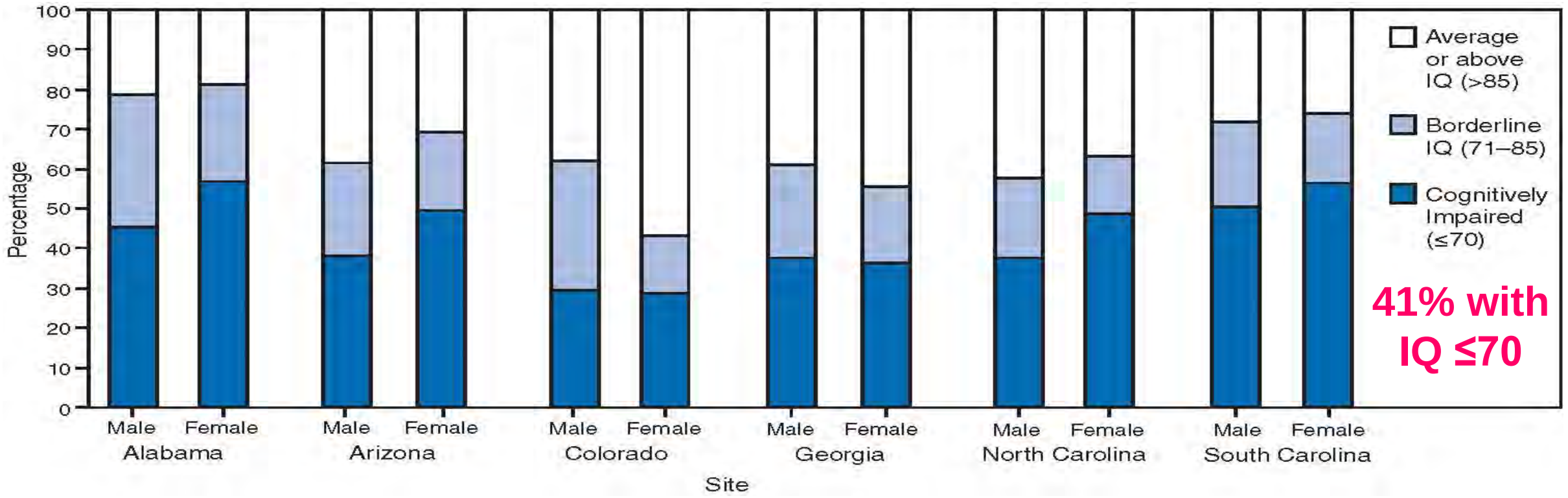


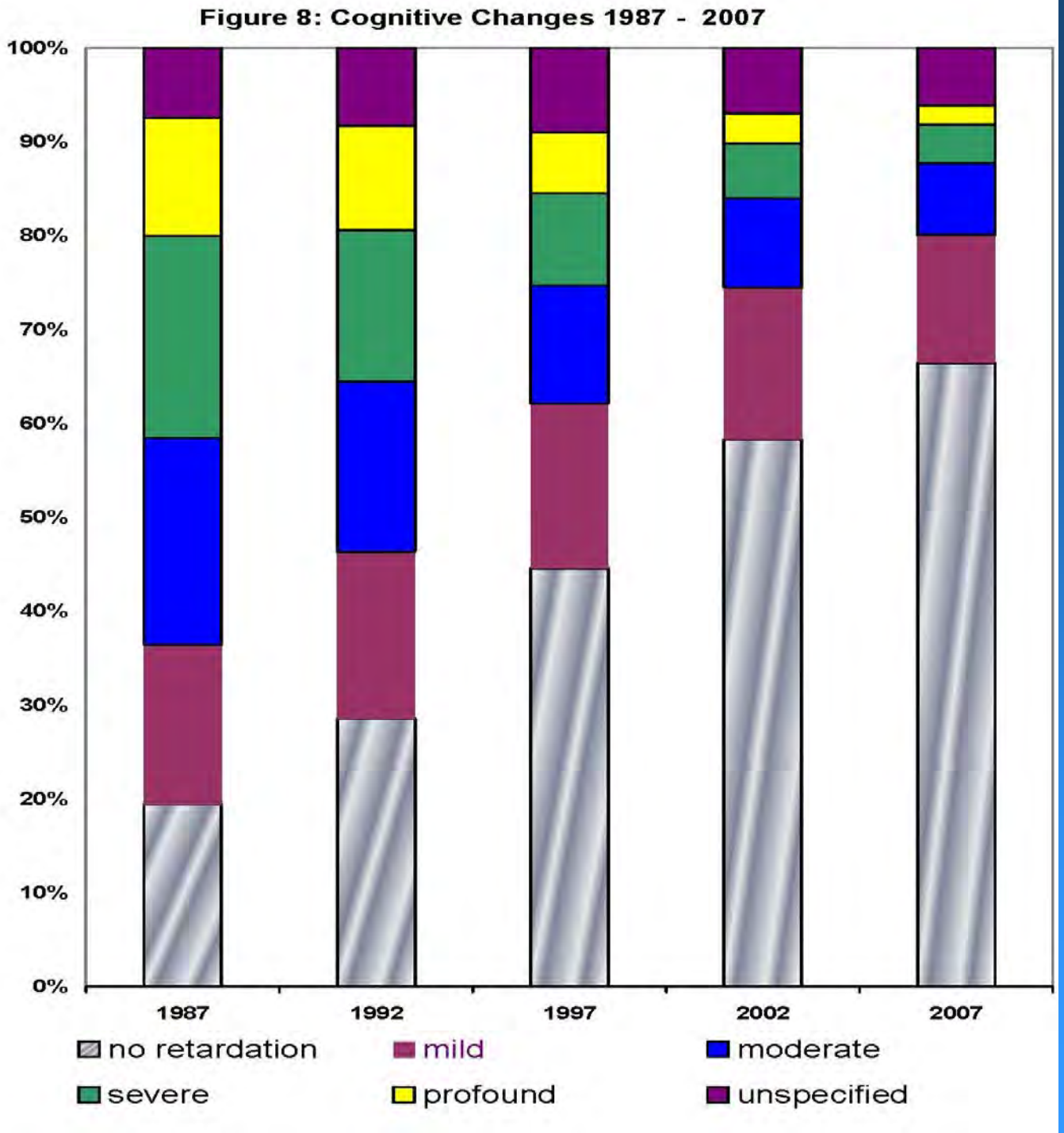
Mental Ability and Autism



Mental Ability and Autism

CDC study of 2,756 eight year olds with ASD





Mental Ability and Autism in the California System 1987 to 2007

Baby Siblings Risk

- Ozonoff studied recurrence risk for autism in 664 baby sibs born in ASD families
- Overall **18.7%** had ASD.
- Among males **26%**
- Among females **09%**
- Relative Risk of males : females **2.8**
- Among multiplex families, RR was **2.2**
- The recurrence rate of ASD was higher than previously assumed

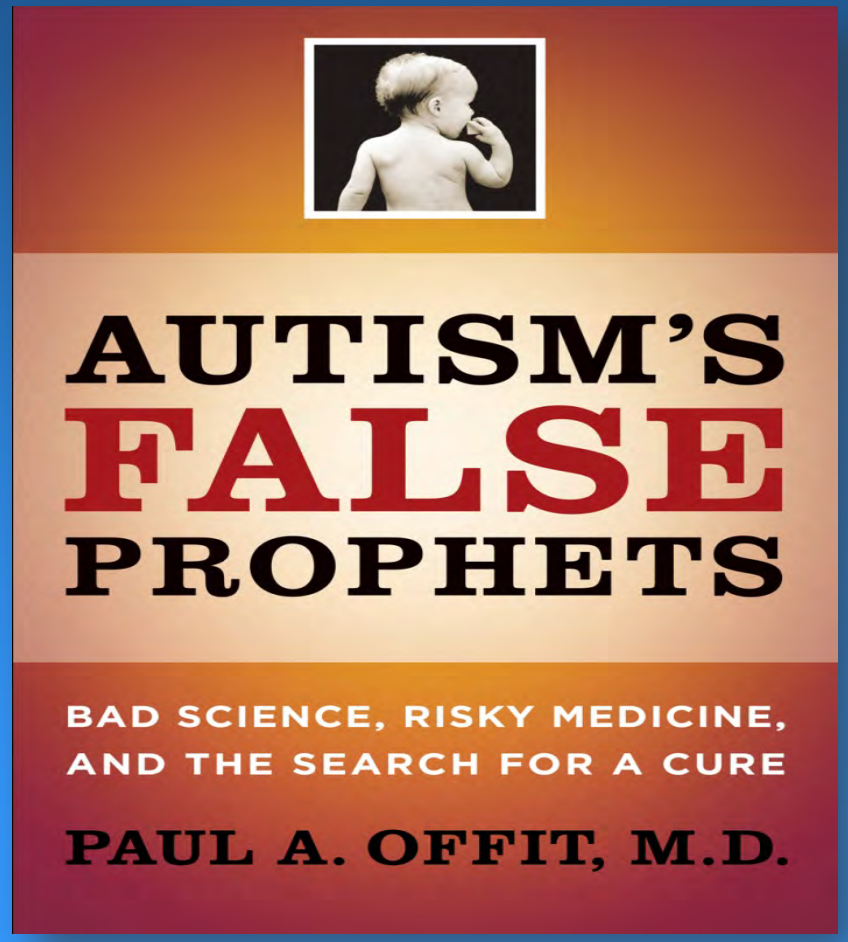
(Ozonoff et al. Recurrence Risk for Autism Spectrum Disorders:
A Baby Siblings Research Consortium Study. Pediatrics 128: e488, 2011)

Environment and Autism

- Several non-genetic factors have been associated with autism:
- Rubella (German Measles) Outbreak in the 1970s, 7% autistic
- Thalidomide - Swedish registry, 5% were autistic
- Valproic Acid – an anticonvulsant
- Fetal Alcohol Syndrome
- Terbutaline – used to suppress labor

Environment and Autism

Vaccines do not cause autism



Fascinating and readable book.

"Autism's False Prophets" traces the histories of the MMR-autism and thimerosal-autism controversies, and discusses the science in clear, layman's language.

Autism Prevalence Following Prenatal Exposure to Hurricanes and Tropical Storms in Louisiana

- Hurricanes and tropical storms serve as natural experiments for investigating whether autism is associated with exposure to stressful events during sensitive periods of gestation.
- Weather service data identified severe storms in Louisiana from 1980 to 1995 and parishes hit by storm centers during this period.
- Autism prevalences in different cohorts were calculated, together with corresponding census data on all births in Louisiana.
- Prevalence increased in dose-response fashion with severity of prenatal storm exposure, especially for cohorts exposed near the middle or end of gestation ($p < 0.001$).
- Results provide further evidence that factors disrupting development during sensitive gestational periods may contribute to autism.

Kinney et al., J Autism Dev Disord 38:481–488 (2008)

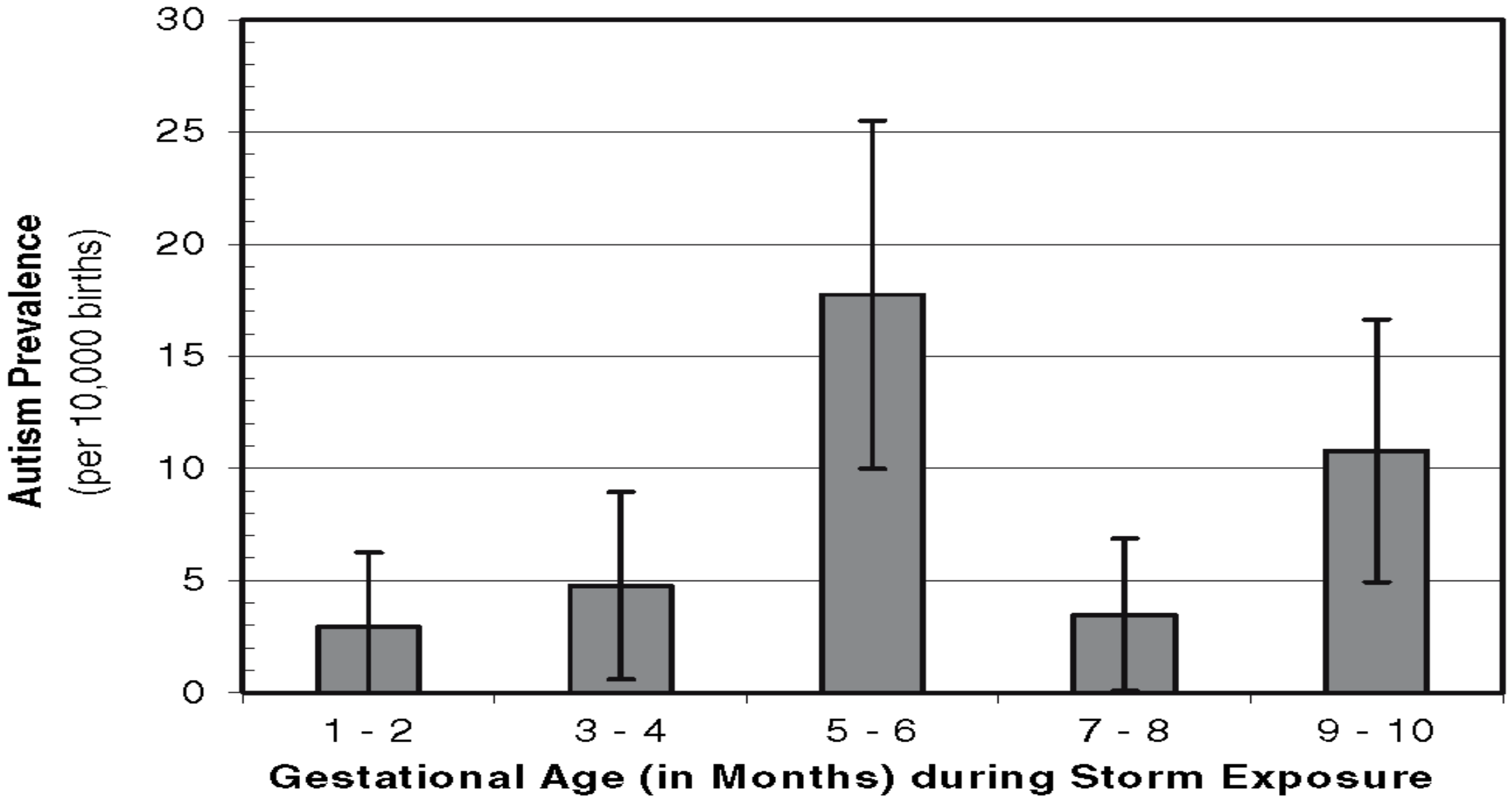


Fig. 1 Prevalence of Autistic Disorder (*AD*) among children born in Orleans parish, by gestational age at time of storm exposure.*Note*: Error brackets represent the 95% confidence limits on prevalence

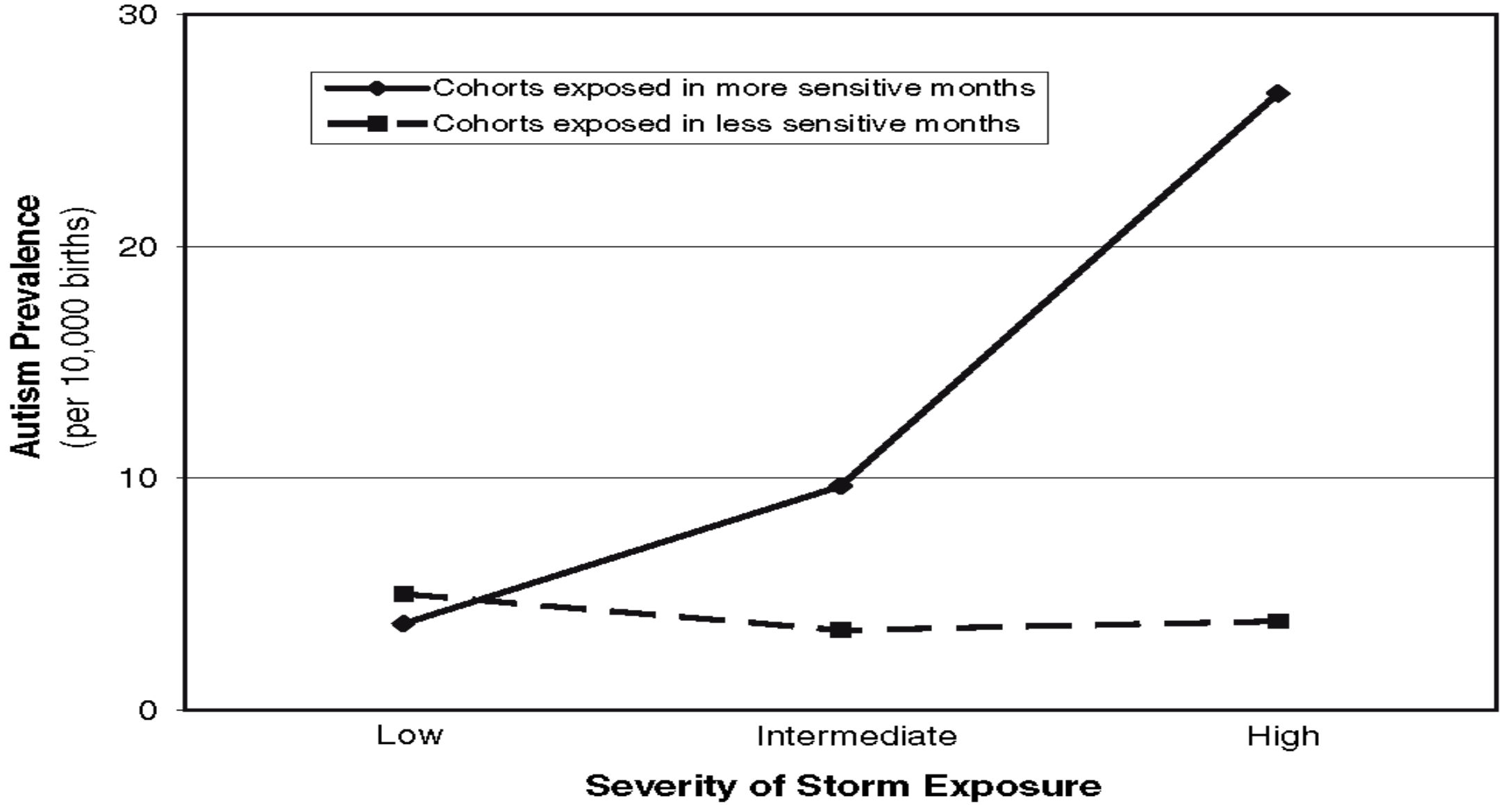


Fig. 2 The relation of AD prevalence to severity of storm exposure depends upon the gestational period when exposure occurs

Fever Improves Autism Symptoms

- Children with autism appear to improve when they have a fever reported Andrew Zimmerman, MD, of Johns Hopkins.
- Fever was associated with less hyperactivity, less irritability, and improved communication in a study of children with ASD.
- The improvement in communication and socialization in the study suggests that fever directly affects brain function.

Curran et al. Pediatrics Dec 07

Genetics of Complex Disorders

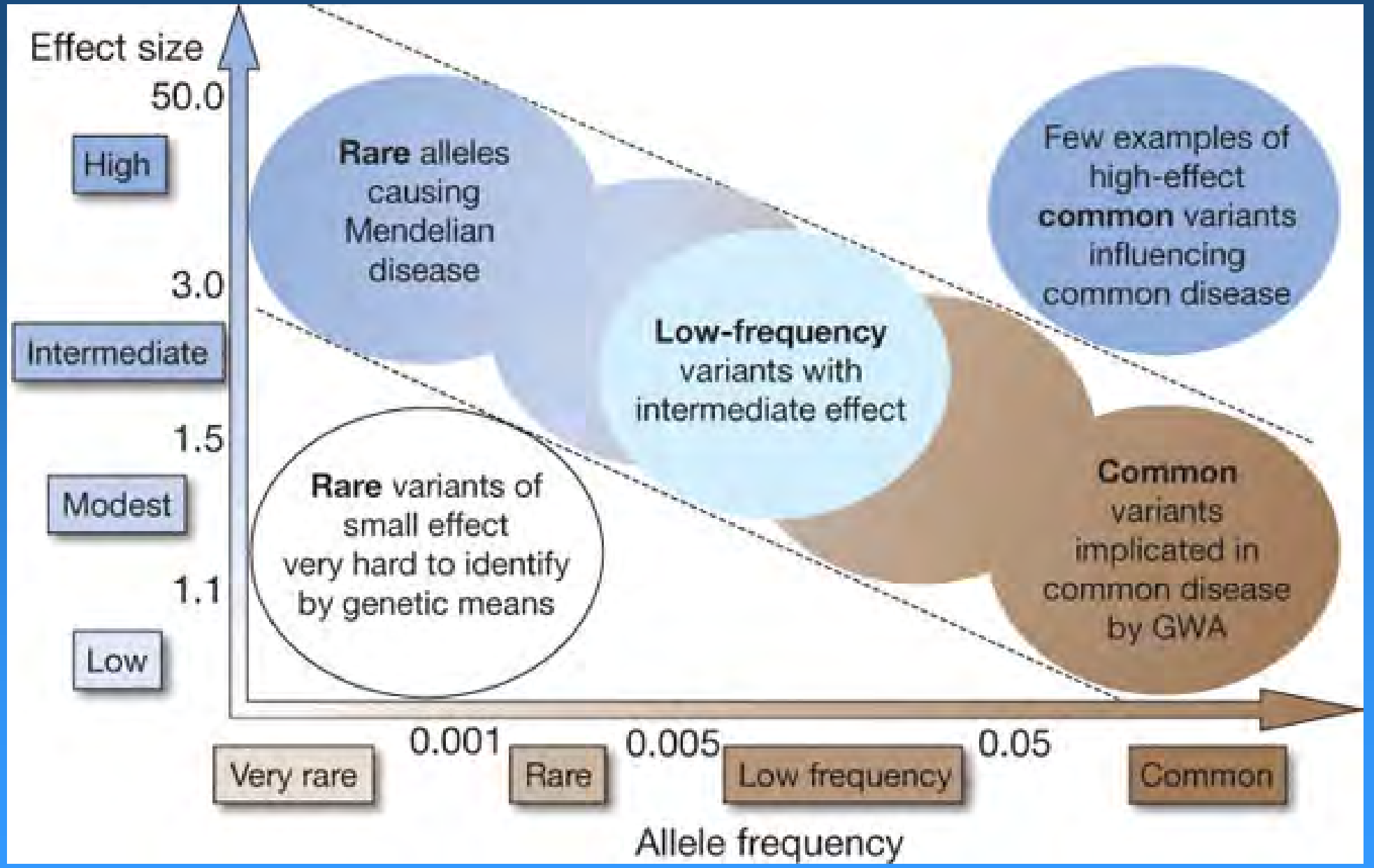
- Polygenetic and multifactorial conditions
- Height, Weight, Hypertension, Diabetes …
- Autism, Schizophrenia, Bipolar Disorder, Depression…
- Heritability: How much variation in a phenotypic trait in a population is due to variation among individuals in that population.
- Often use identical vs. fraternal twin studies.

Heritability for Several Neuropsychiatric Disorders

Disorder	Heritability estimate
Autism	50-55 %
Schizophrenia	80-84 %
Bipolar Disorder	60-70 %
Panic Disorder	50-60 %
Substance Dependence	30-50 %
Major Depression	28-40 %

Sandin (2014)

Merikangas & Risch (2003)


Effect size
50.0
High
3.0
Intermediate
1.5
Modest
1.1
Low
Rare alleles causing Mendelian disease
Few examples of high-effect common variants influencing common disease
Low-frequency variants with intermediate effect
Rare variants of small effect very hard to identify by genetic means
Common variants implicated in common disease by GWA
0.001
0.005
0.05
Very rare
Rare
Low frequency
Common
Allele frequency

Where did the heritability go?

The missing heritability problem: individual genes cannot explain the heritability of traits

The case of the missing heritability

Vol 461|8 October 2009|doi:10.1038/nature08494 nature

REVIEWS

Finding the missing heritability of complex diseases

Teri A. Manolio[1], Francis S. Collins[2], Nancy J. Cox[3], David B. Goldstein[4], Lucia A. Hindorff[5], David J. Hunter[6], Mark I. McCarthy[7], Erin M. Ramos[5], Lon R. Cardon[8], Aravinda Chakravarti[9], Judy H. Cho[10], Alan E. Guttmacher[1], Augustine Kong[11], Leonid Kruglyak[12], Elaine Mardis[13], Charles N. Rotimi[14], Montgomery Slatkin[15], David Valle[9], Alice S. Whittemore[16], Michael Boehnke[17], Andrew G. Clark[18], Evan E. Eichler[19], Greg Gibson[20], Jonathan L. Haines[21], Trudy F. C. Mackay[22], Steven A. McCarroll[23] & Peter M. Visscher[24]

Genome-wide association studies have identified hundreds of genetic variants associated with complex human diseases and traits, and have provided valuable insights into their genetic architecture. Most variants identified so far confer relatively small increments in risk, and explain only a small proportion of familial clustering, leading many to question how the remaining, 'missing' heritability can be explained. Here we examine potential sources of missing heritability and propose research strategies, including and extending beyond current genome-wide association approaches, to illuminate the genetics of complex diseases and enhance its potential to enable effective disease prevention or treatment.

How to explain this problem?

Rare Variants, rare CNVs, epigenetics or.. epistatic effects?

Table 1 | Estimates of heritability and number of loci for several complex traits

Disease	Number of loci	Proportion of heritability explained
Age-related macular degeneration[72]	5	50%
Crohn's disease[21]	32	20%
Systemic lupus erythematosus[73]	6	15%
Type 2 diabetes[74]	18	6%
HDL cholesterol[75]	7	5.2%
Height[15]	40	5%
Early onset myocardial infarction[76]	9	2.8%
Fasting glucose[77]	4	1.5%

* Residual is after adjustment for age, gender, diabetes.

Is the heritability really missing or are we looking at the wrong place?

How to explain missing heritability?
*Rare Variants, rare CNVs, epigenetics or.. **epistatic effects**?*

Table 1 | Estimates of heritability and number of loci for several complex traits

Disease	Number of loci	Proportion of heritability explained
Age-related macular degeneration[72]	5	50%
Crohn's disease[21]	32	20%
Systemic lupus erythematosus[73]	6	15%
Type 2 diabetes[74]	18	6%
HDL cholesterol[75]	7	[illegible]
Height[15]	40	5%
Early onset myocardial infarction[76]	9	[illegible]
Fasting glucose[77]	4	1.5%

* Residual is after adjustment for age, gender, diabetes.

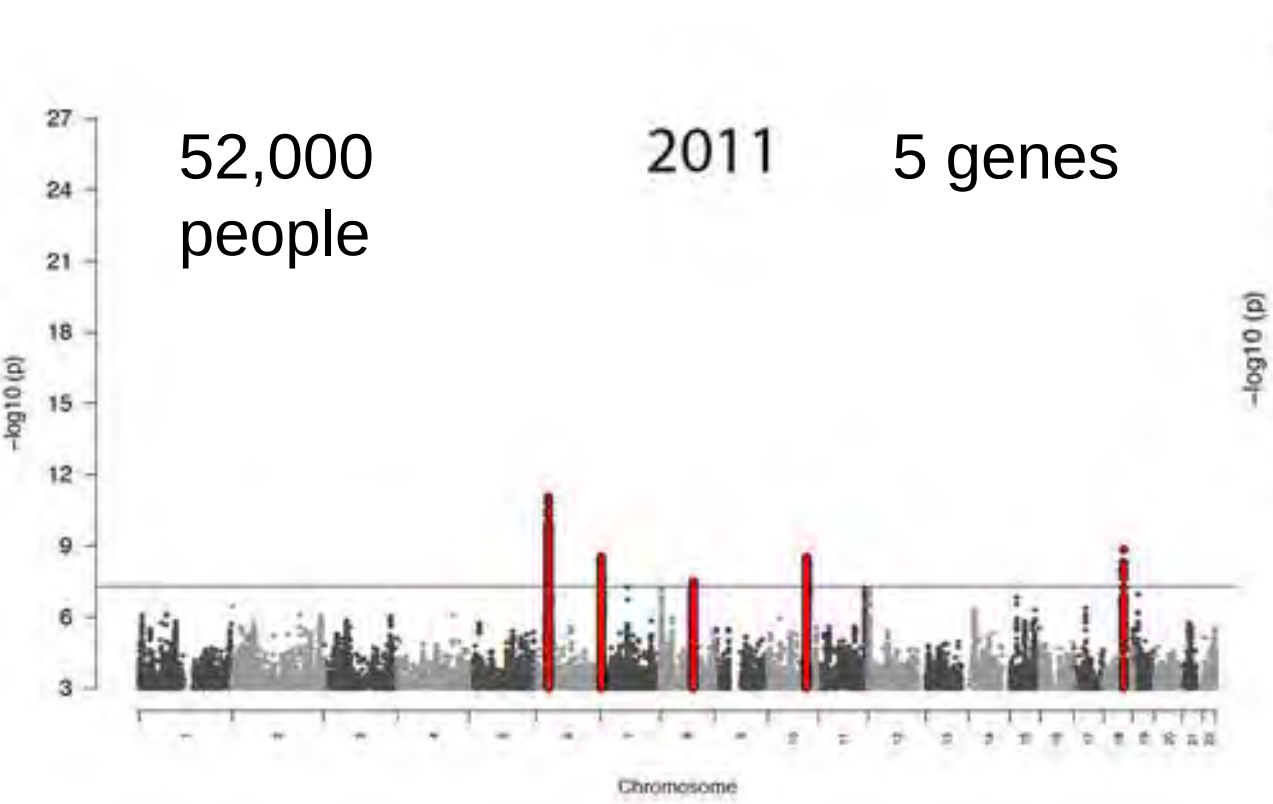
The case of the missing heritability

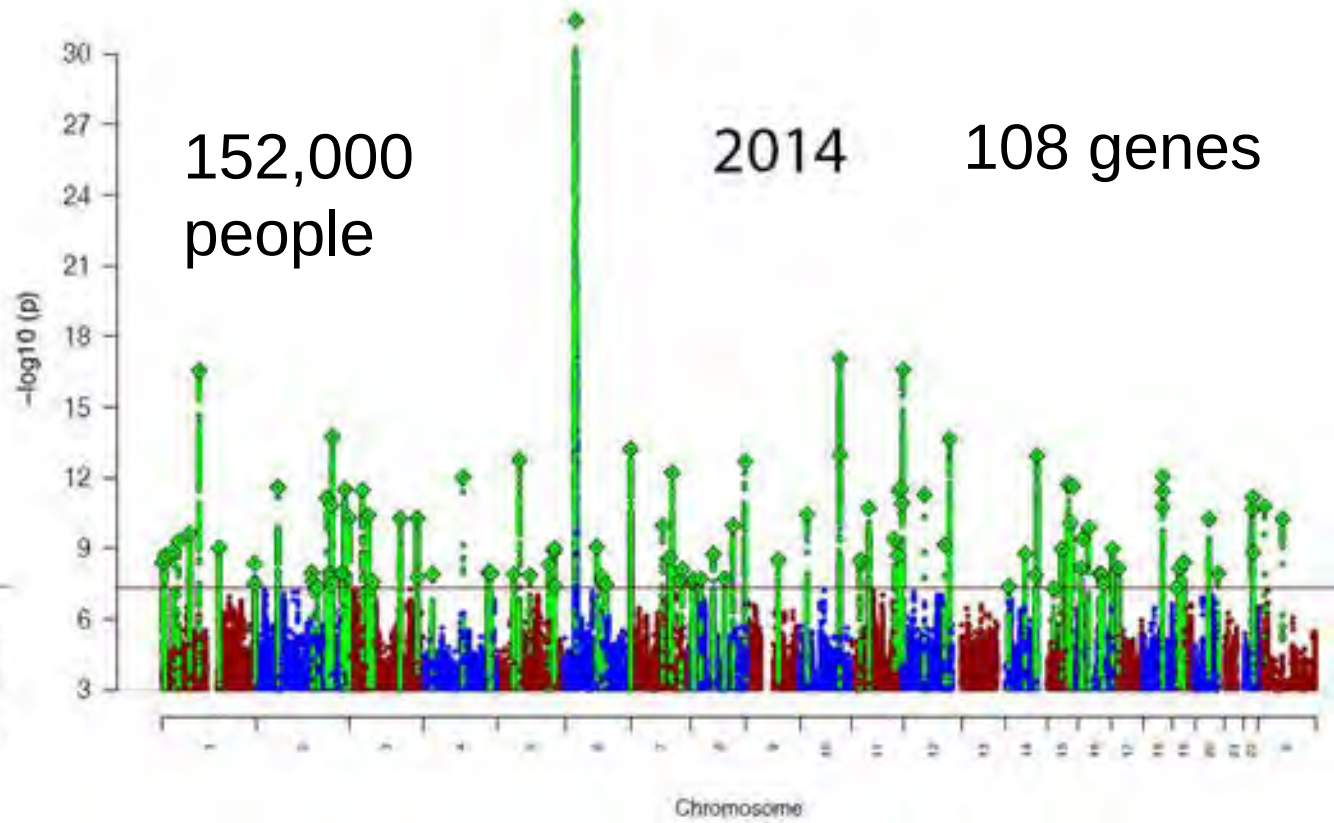
genetics

Common SNPs explain a large proportion of the heritability for human height

At the end, most of the heritability was there…

Manhattan plot showing schizophrenia associations.

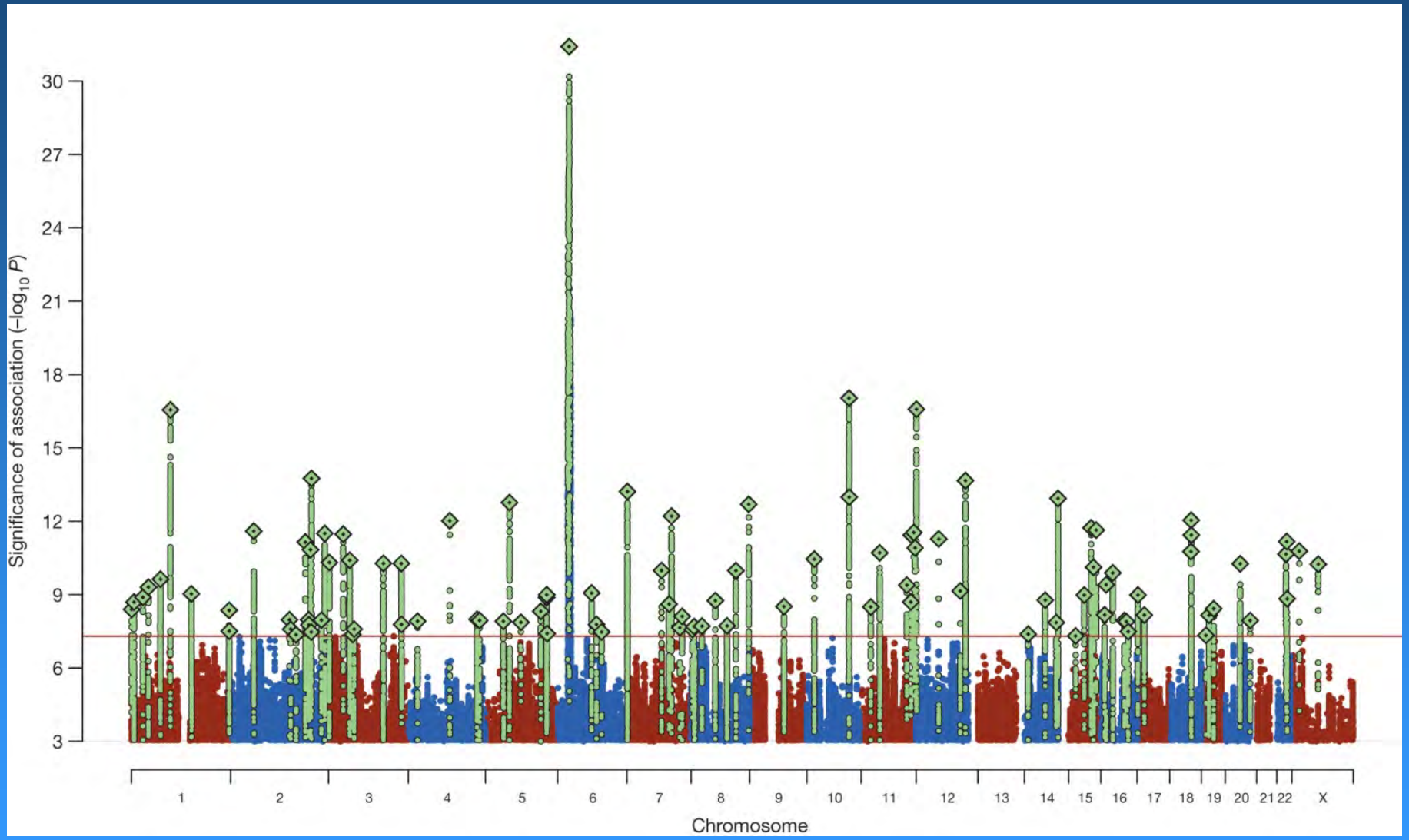




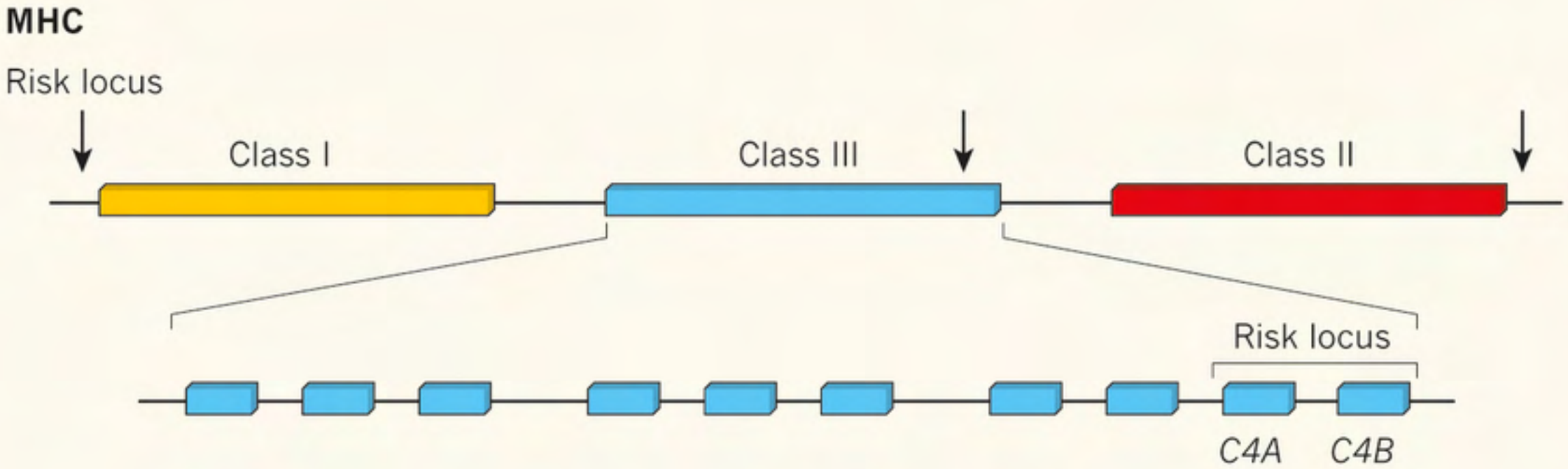
Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics* 43: 969 (2011)

Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511: 421 (2014)

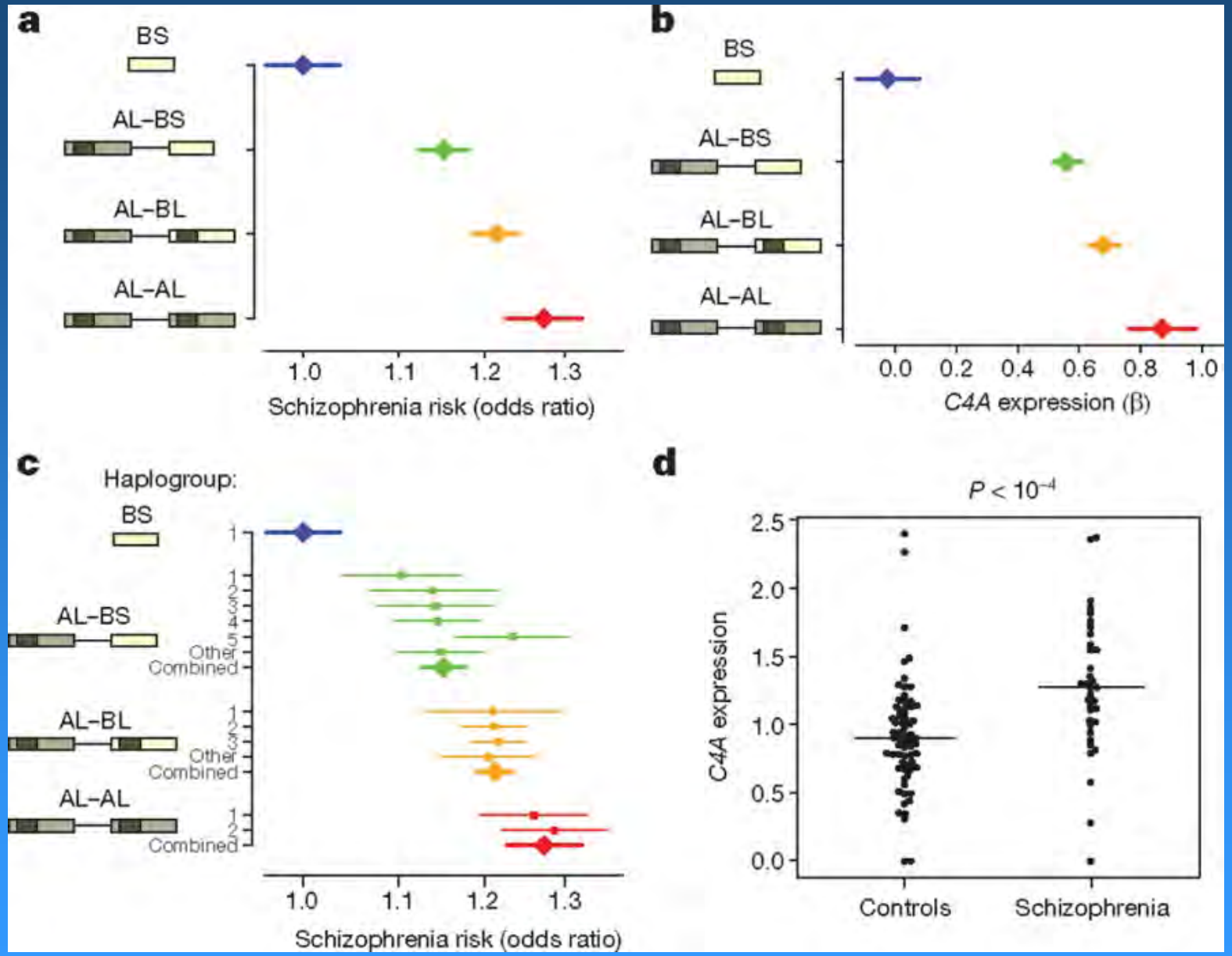
Manhattan plot showing schizophrenia associations.



Ripke *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511: 421 (2014)

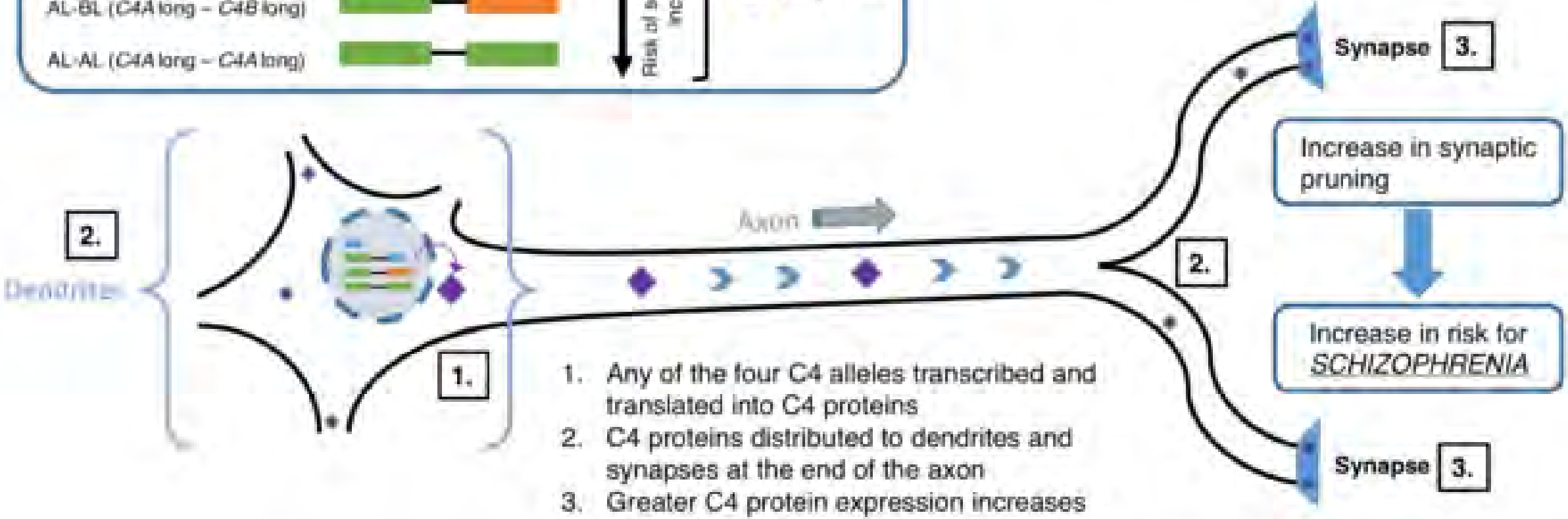

MHC
Risk locus
Class I
Class III
Class II
Risk locus
C4A
C4B

C4 structures, *C4A* expression, and schizophrenia risk



Sekar *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature 530:177* (2016)

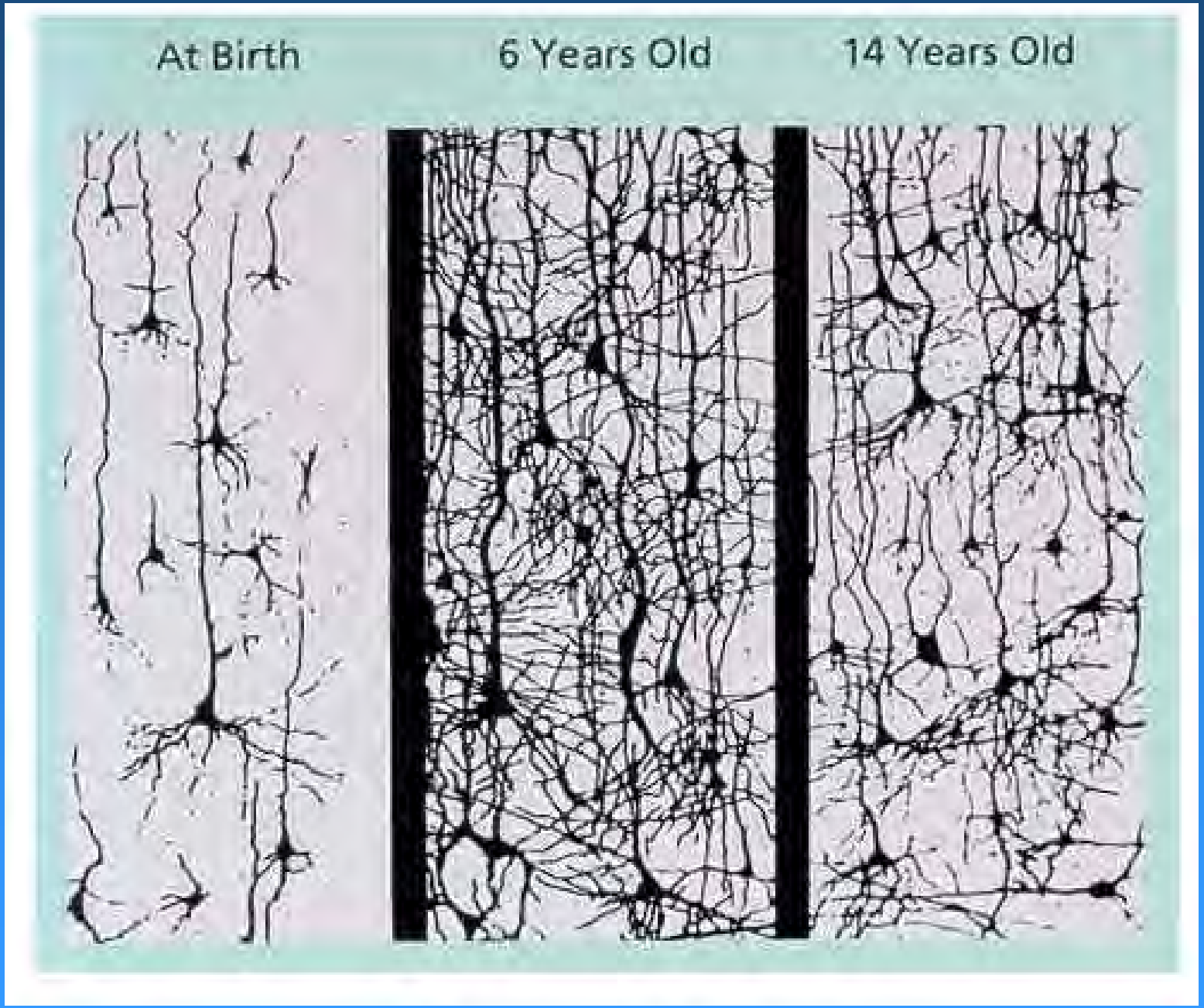
Four common structural variants of the C4 gene (*C4A/C4B* loci):

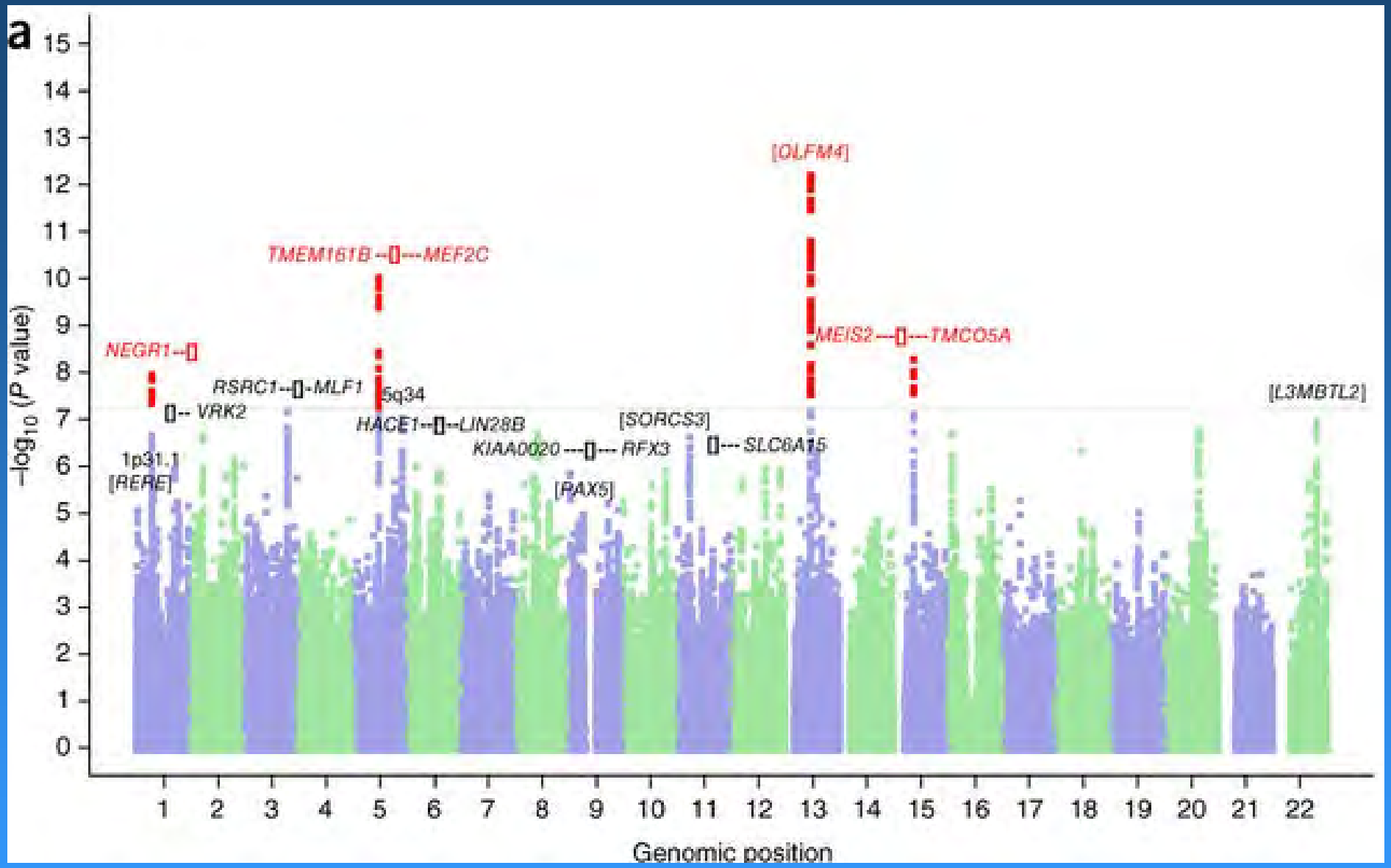


1. Any of the four C4 alleles transcribed and translated into C4 proteins
2. C4 proteins distributed to dendrites and synapses at the end of the axon
3. Greater C4 protein expression increases synaptic pruning within the neuronal circuit

Sekar *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature 530:177* (2016)

The difference in neuron density from 6 years to 14 years is a result of synaptic pruning.





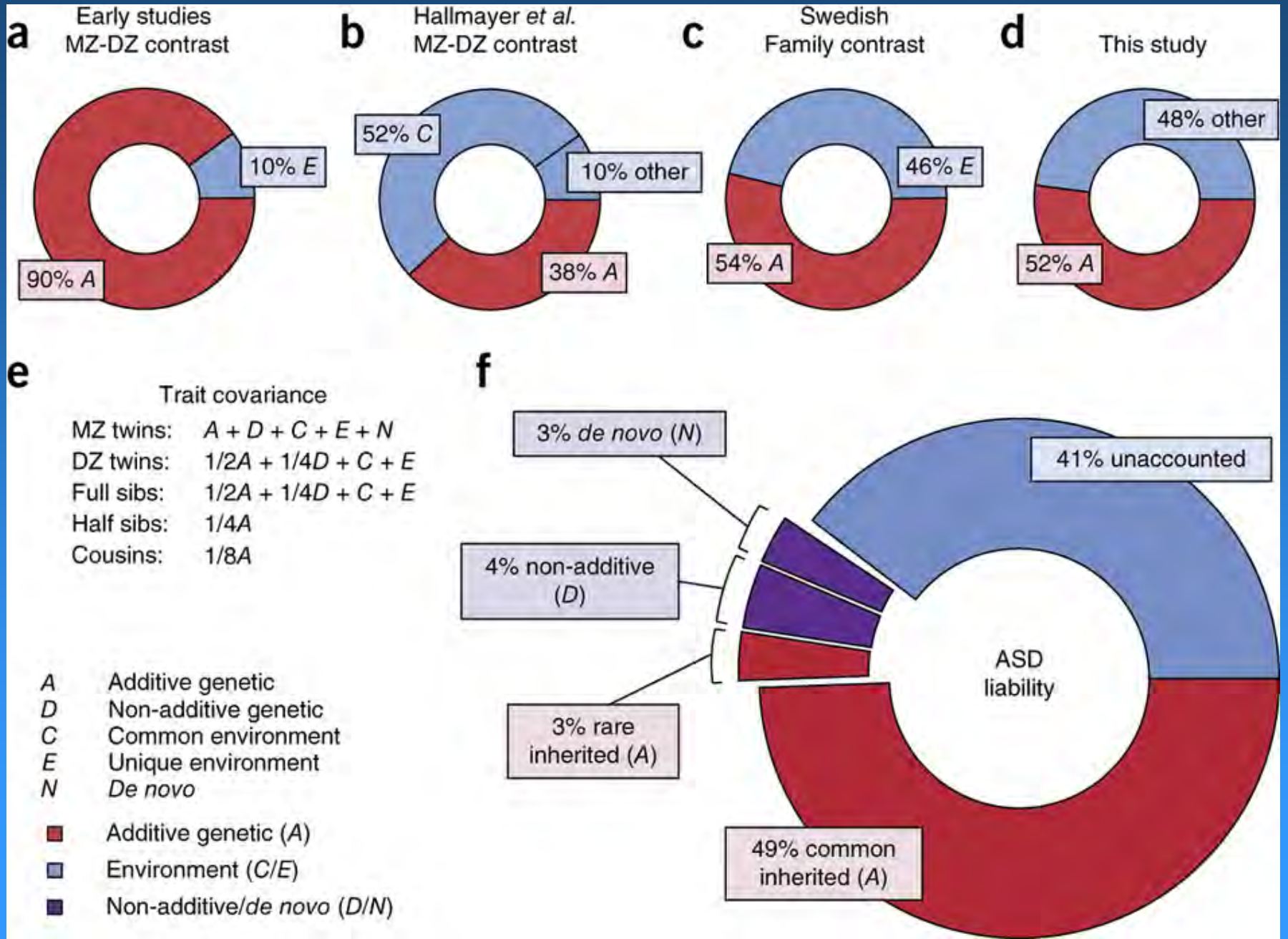
Discovery-phase meta-analysis of 23andMe self-report ascertainment of major depression (75,607 cases and 231,747 controls) **Hyde et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nature Genetics (2016)**

Genetic Risk for Autism

- Many risk-associated genes identified from rare variation.
- Yet common variation has substantial impact.
- How much effect on heritability?
- Recent new methods were applied to a Swedish twin study.
- 2.6 M sibling pairs, 37,600 twins, 14,500 ASD cases.
- Heritability was 52%, with most due to common variation.

Gaugler et al. Most genetic risk for autism resides with common variation.
Nature Genetics 46:881 (2016)

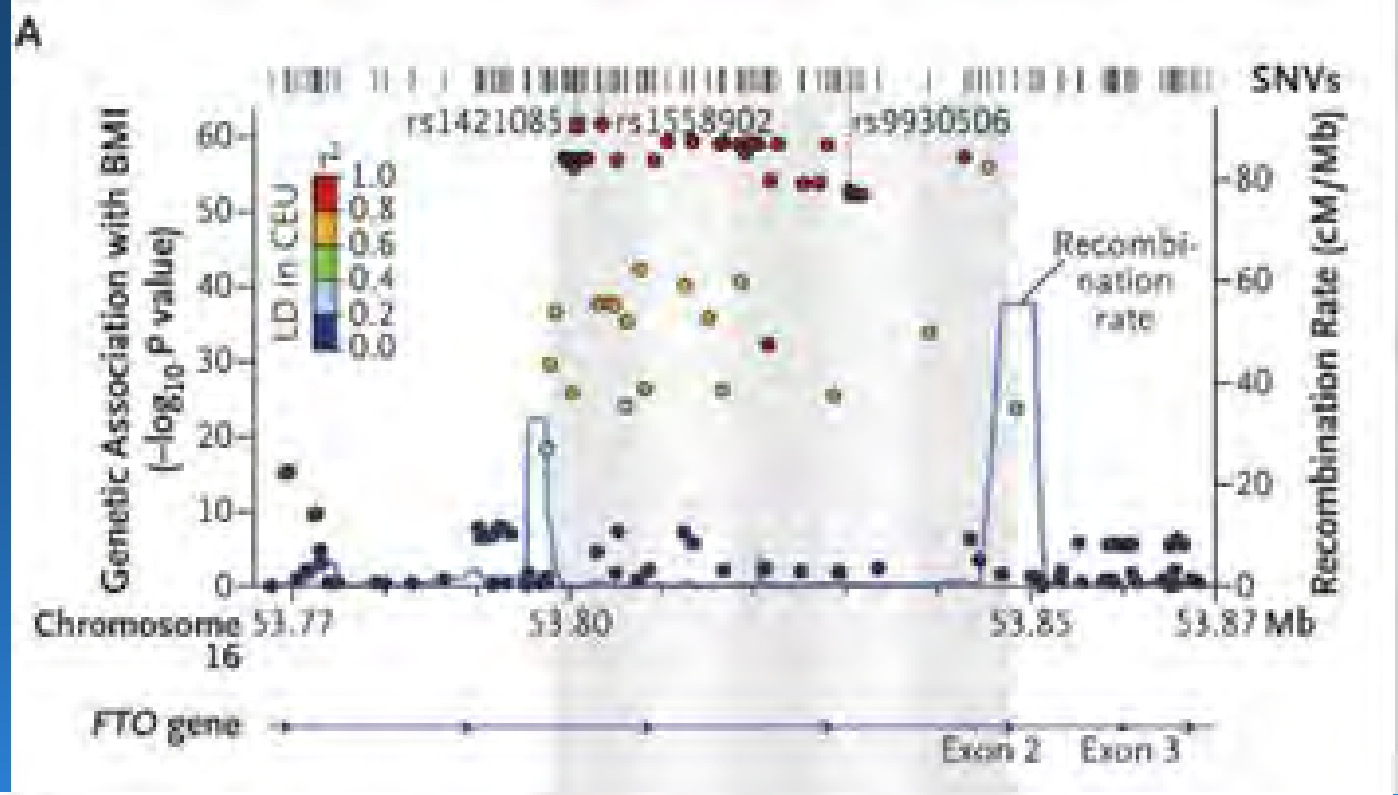
Most genetic risk for autism resides with common variation. Nature Genetics 46, 881–885 (2014)

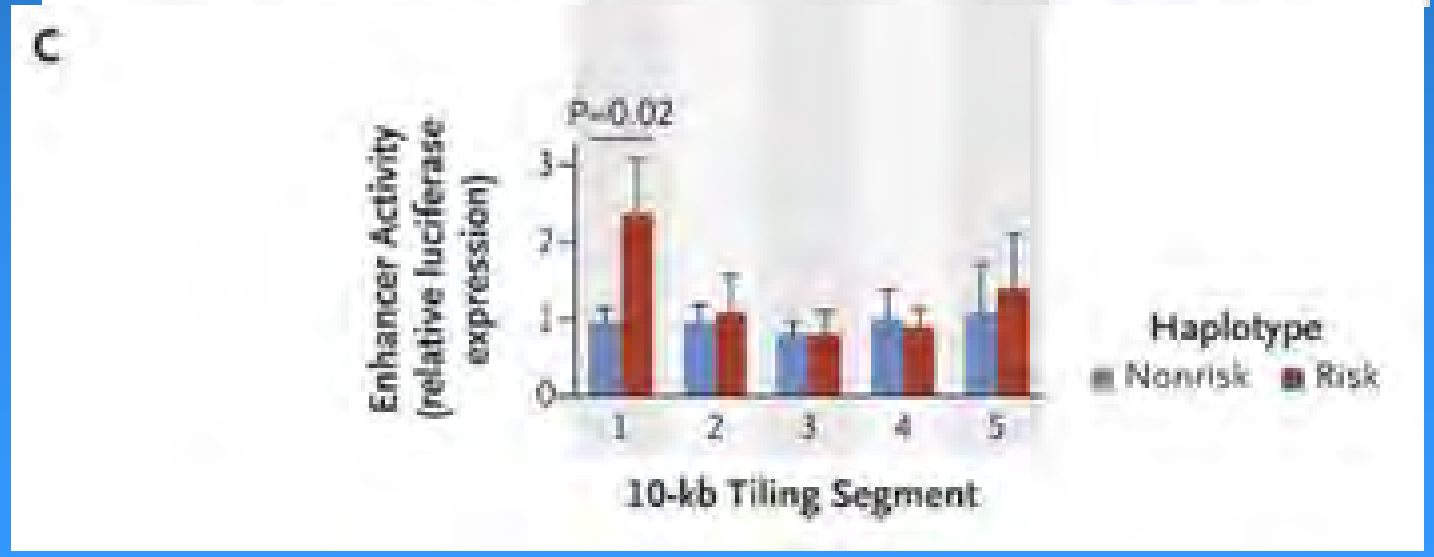


GWAS of Autism

- Total cases in 4 larges studies less than 10,000.
- Candidate SNPs have not been replicated.
- Sample sizes too small??
- A large-scale project in ASD is currently underway.
- Psychiatric Genomics Consortium (www.med.unc.edu/pgc).
- 900,000 cases with initial focus on autism, attention-deficit hyperactivity disorder, bipolar disorder, major depressive disorder, and schizophrenia.

OBESITY ENHANCER

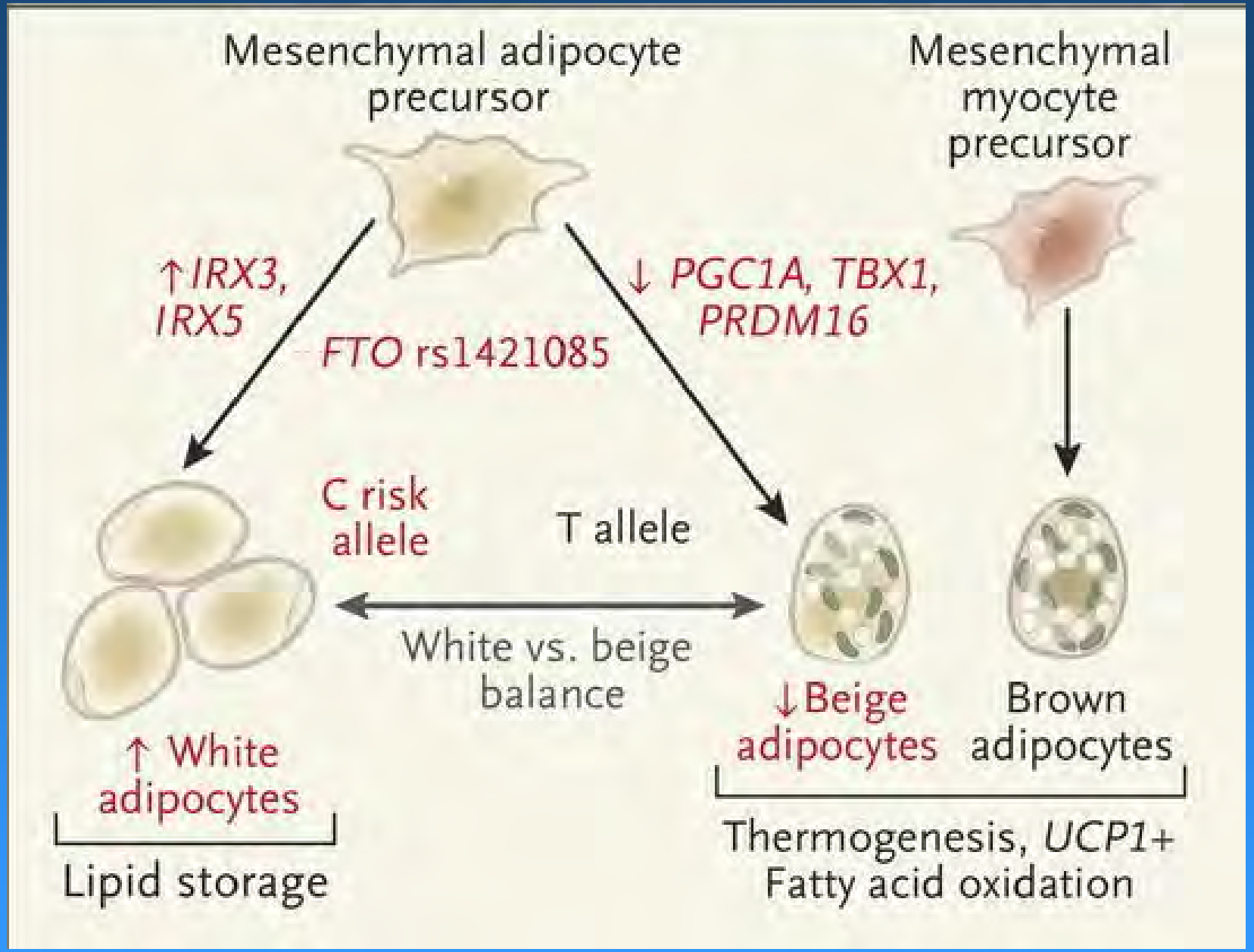




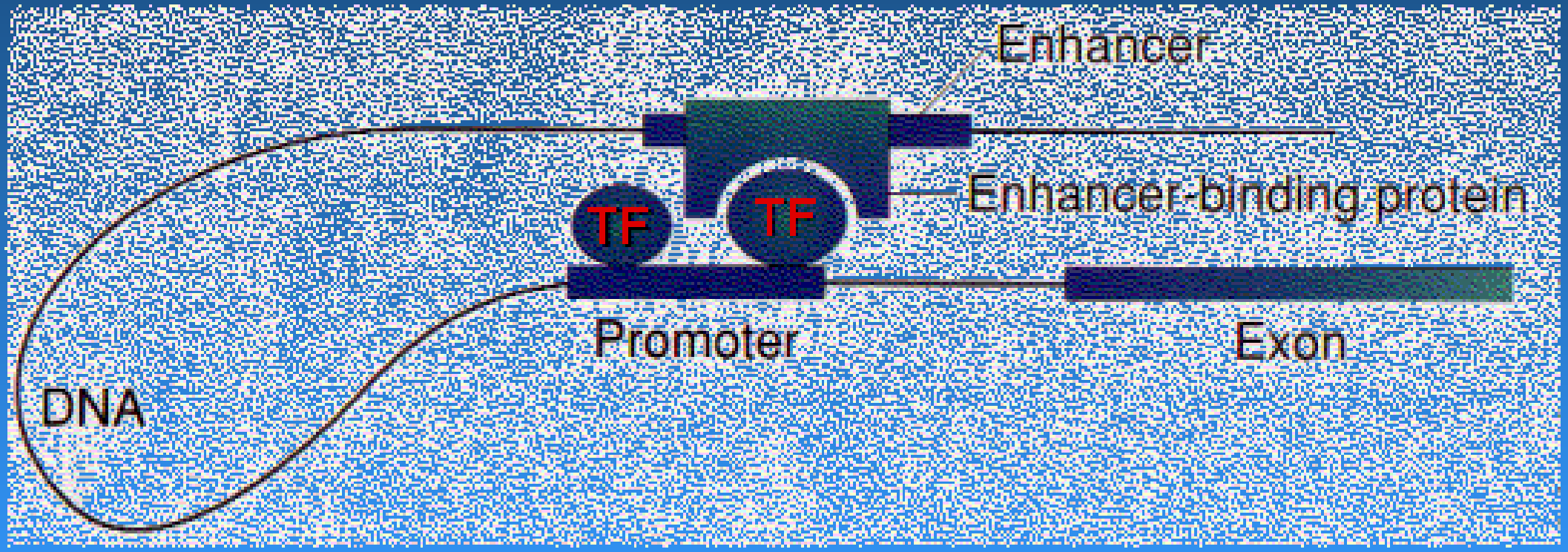
About 90% of SNVs that are associated with human traits and diseases map to non-coding genetic regions.

- The strongest linkage with obesity is to a noncoding Intron 1 region in FTO gene.
- Kellis (2015) found this intronic region contained an enhancer controlling two genes (IRX3 and IRX5) that control adipocyte energy storage and energy expenditure.
- Located 1.2 Mb away

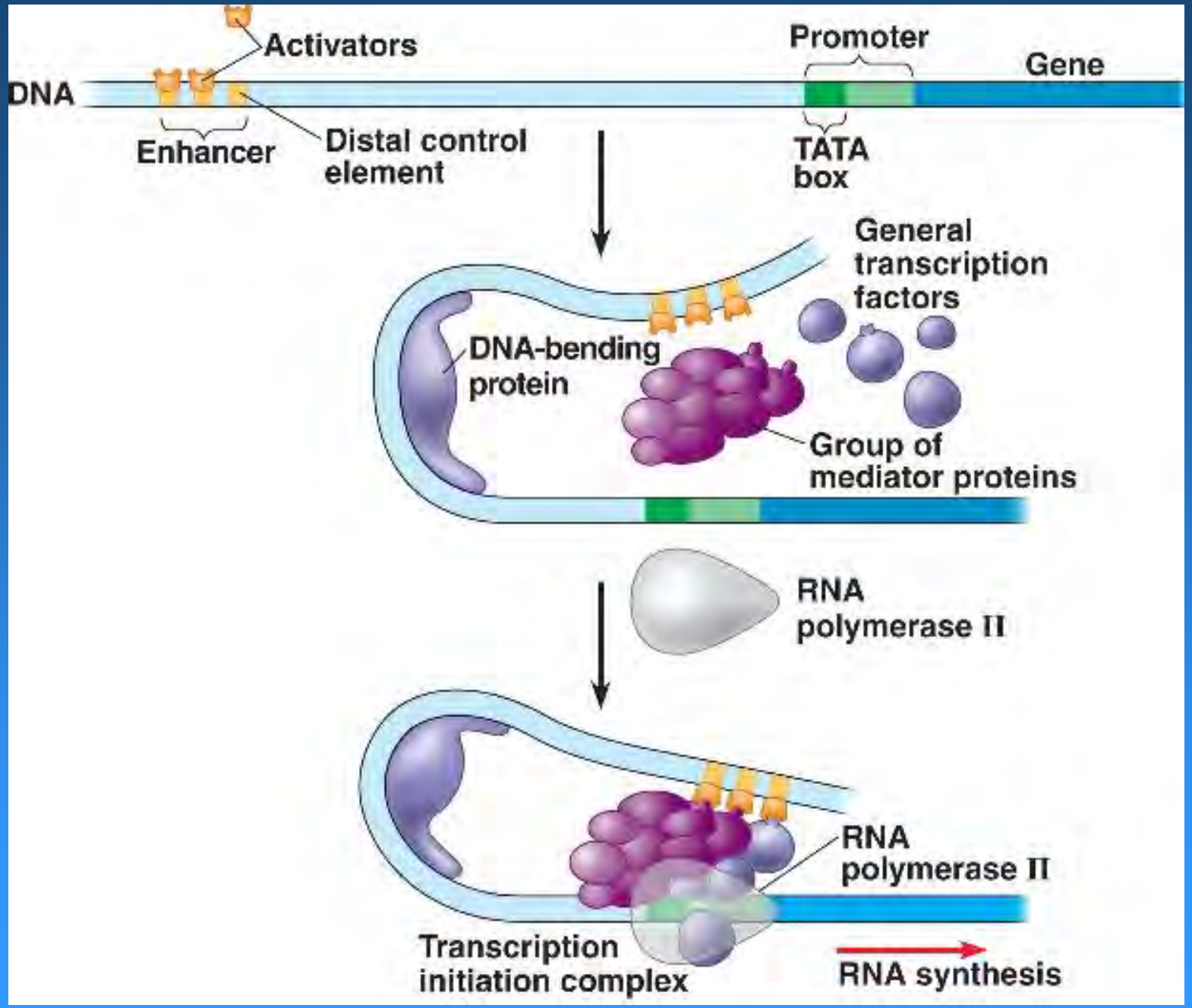
***FTO* Obesity Variant Circuitry and Adipocyte Browning in Humans. NEJM 373:895 (2015)**


Mesenchymal adipocyte precursor
Mesenchymal myocyte precursor
↑IRX3, IRX5
FTO rs1421085
↓ PGC1A, TBX1, PRDM16
C risk allele
T allele
White vs. beige balance
↑ White adipocytes
↓Beige adipocytes
Brown adipocytes
Lipid storage
Thermogenesis, UCP1+
Fatty acid oxidation

An **enhancer** is a short (50-1500 bp) region of DNA that can be bound by proteins (activators) to increase the likelihood that transcription of a particular gene will occur.



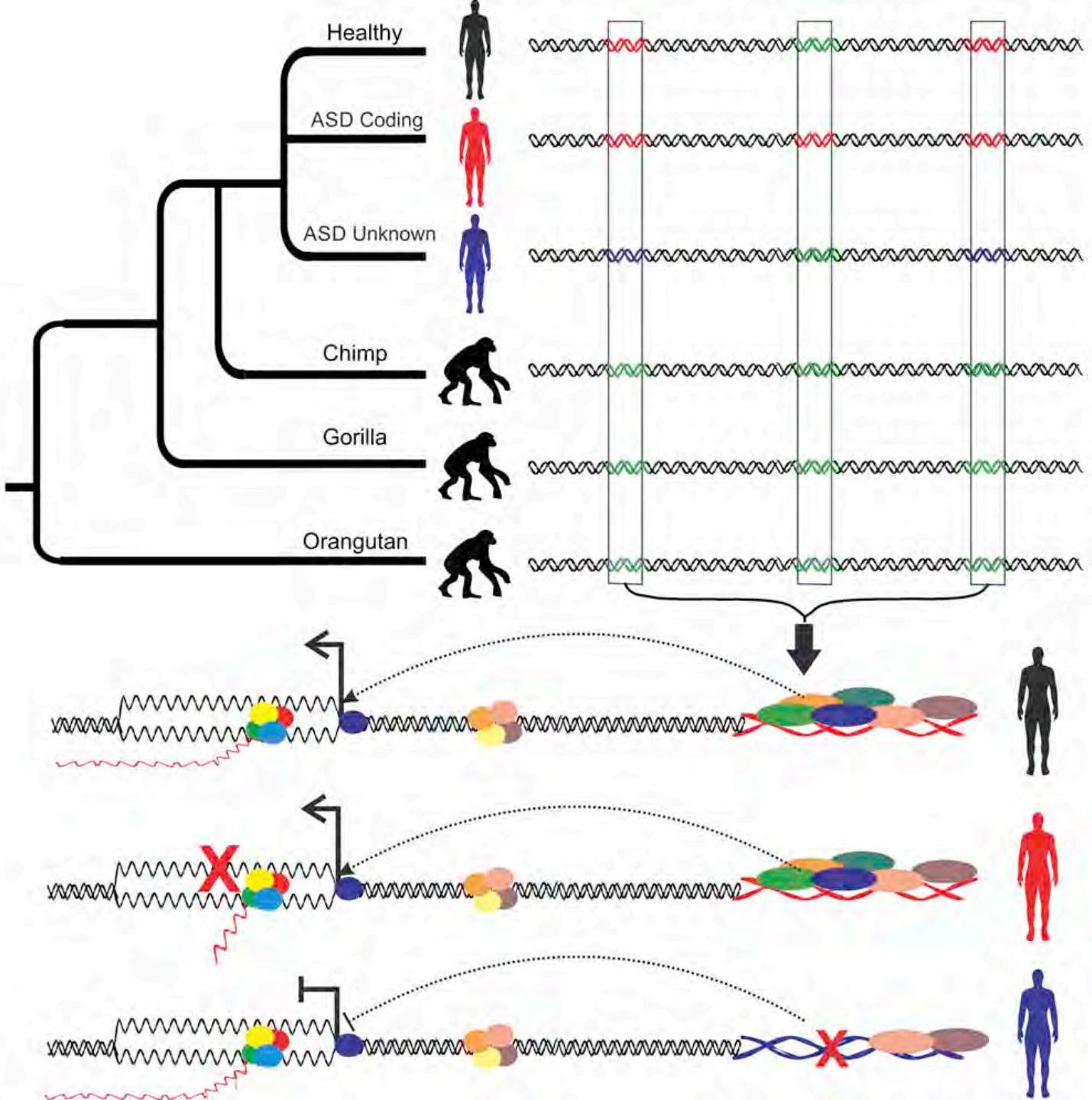
Transcription Factor (TF)


Activators
Promoter
Gene
DNA
Enhancer
Distal control element
TATA box
General transcription factors
DNA-bending protein
Group of mediator proteins
RNA polymerase II
RNA polymerase II
Transcription initiation complex
RNA synthesis

Human Accelerated Regions (HARs)

- 1.5% of genome is coding sequences (the Exome) .
- 5% of genome is highly conserved but non-coding.
- Gene regulatory regions highly conserved in mammals, but showing divergence in humans are HARs.
- What differs between humans and chimps? ~3000 HARs.
- 5% of HARs are lncRNAs, but ~95% are enhancers.
- They appear to be involved with cognition, spoken language, and fine motor skills.

Pollard. Decoding Human Accelerated Regions. The-Scientist, Aug 1, 2016



Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior

(Doan et al., 2016, Cell 167, 341–354)

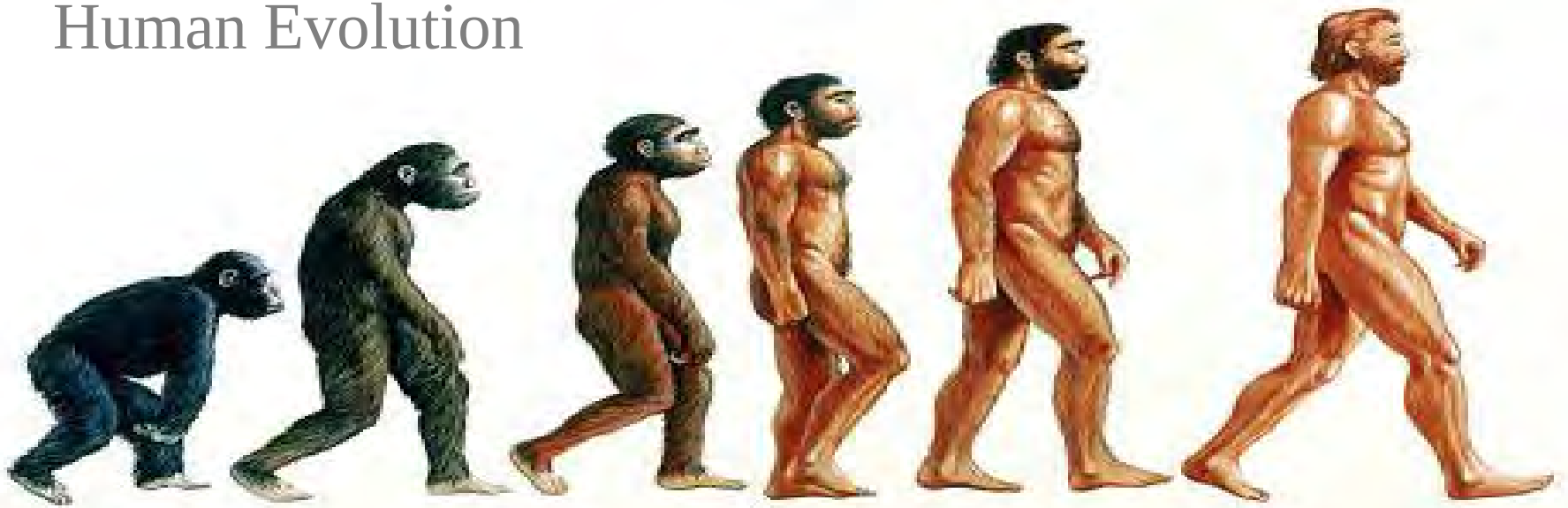
Human accelerated regions exhibit regulatory activity during neural development.

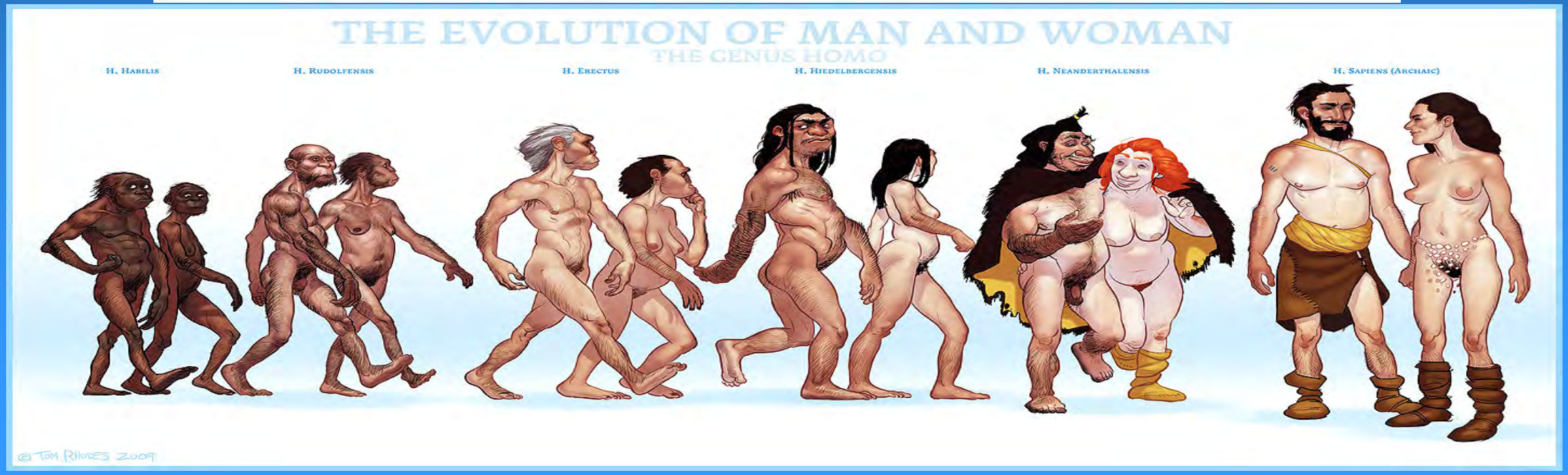
De novo CNVs impacting HARs are enriched in individuals with ASD.

Biallelic HAR mutations underlie up to 5% of consanguineous ASD cases.

Regulatory mutations reveal novel genetic architecture of ASD.

Human Evolution





Summary: Genetics of Intellectual Disabilities

- Costs for sequencing has been dropping dramatically.
- Using CMA, followed by WES, followed by WGS accelerating.
- More than 700 of estimated 1,000+ genes for ID found.
- Next several years should allow identification of most ID genes.
- GWAS of large samples yielding missing heritability.
- 5% of genome is highly conserved but non-coding.
- Although 5% of HARs are lncRNAs, but **95%** are enhancers.
- Enhancer mutations are being found relating to ID.

NEW YORK STATE INSTITUTE FOR BASIC RESEARCH in DEVELOPMENTAL DISABILITIES

GENETICS

SARAH NOLIN

CARL DOBKIN

NAN ZHONG

ANNE GLICKSMAN

XIAOHUA DING

WEINA JU

MILEN VELINOV

ED JENKINS

COLLABORATORS AT IBR

IRA COHEN

VICKI SUDHALTER

GIUSEPPE LAFAUCI

JERZY WEGIEL

PANKAJ MEHTA

GUANG WEN

RICHARD KASCSAK

XIAOHONG LI

ABHA CHAUHAN

JANU FRACKOWIAK

NAR RAMAKRISHNA

ABDESLE EL IDRISSI

MOHAMMED JUNAID

MAZHAR MALIK

ERIC LONDON