Complex Disorders

9/26/13

Joseph G. Hacia
hacia@hsc.usc.edu
USC Keck School of Medicine

Lecture Outline

- Trust worthy resources for medical genetics
- Brief review of basic inheritance patterns
- Genetics of common disorders
- Genome-wide association studies
- Searches for missing heritability

Trust worthy internet resources

- **G**enetics **H**ome **R**eference

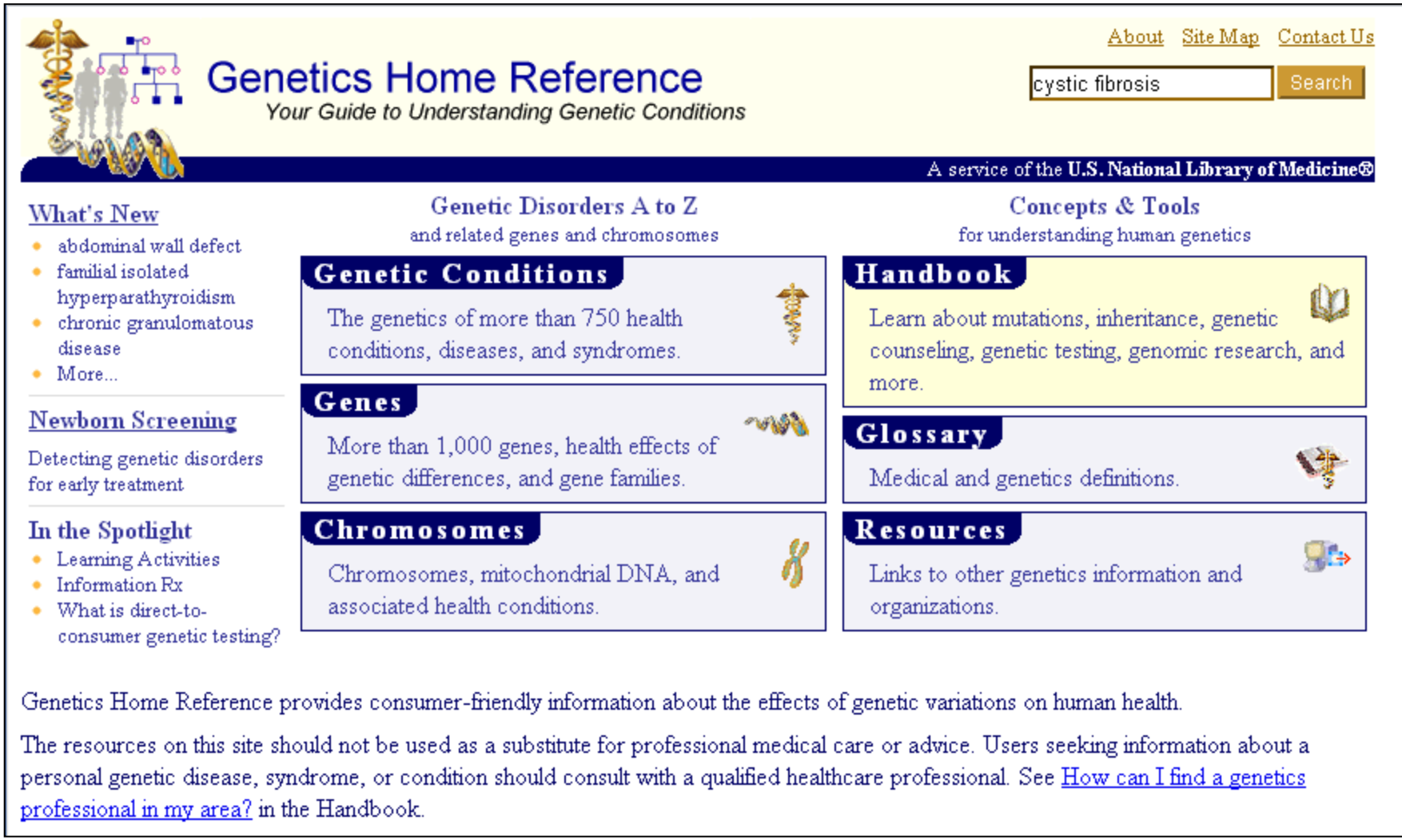
  ghr.nlm.nih.gov

- **G**ene **R**eviews

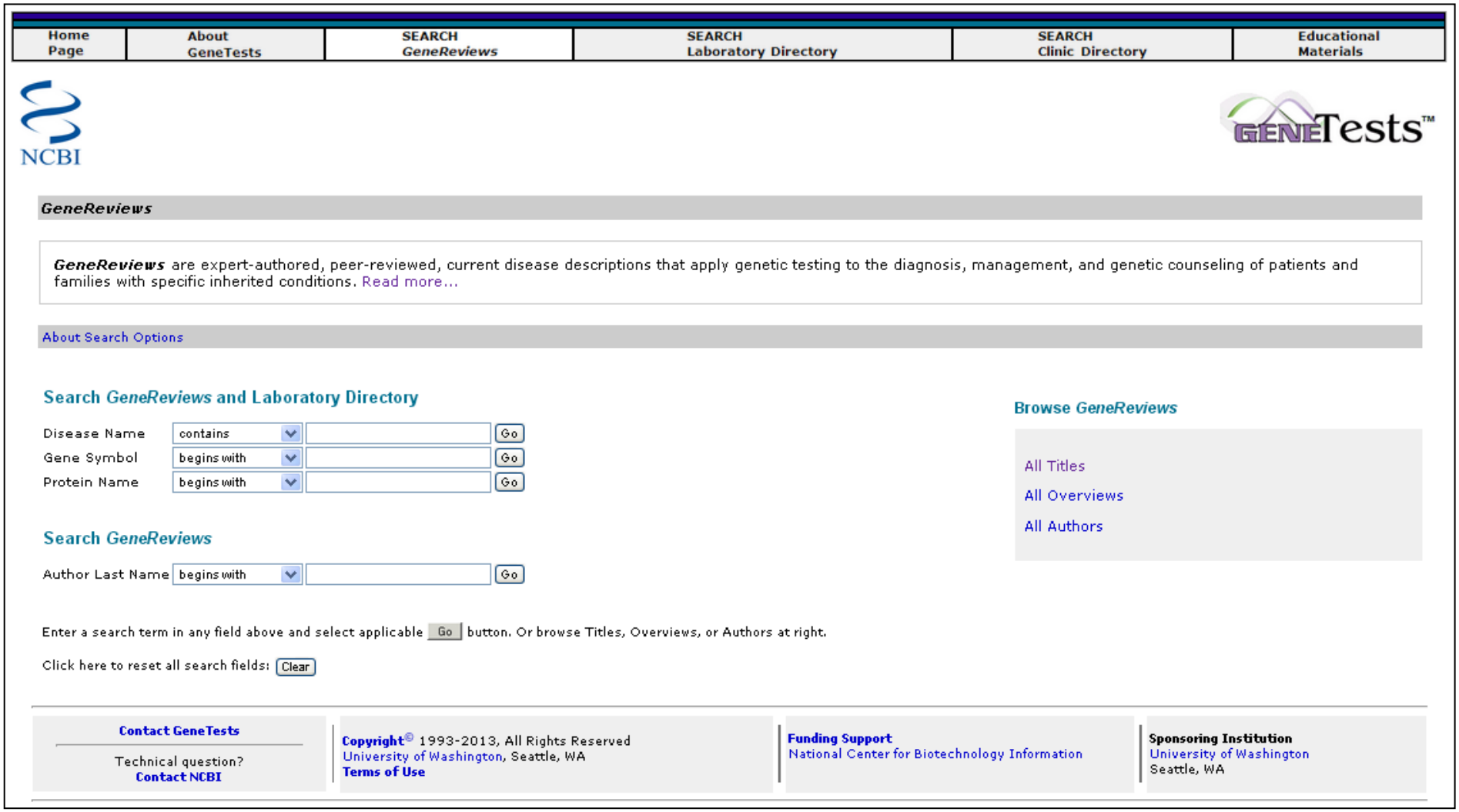
  www.ncbi.nlm.nih.gov/sites/GeneTests/review

- **O**nline **M**endelian **I**nheritance of **M**an

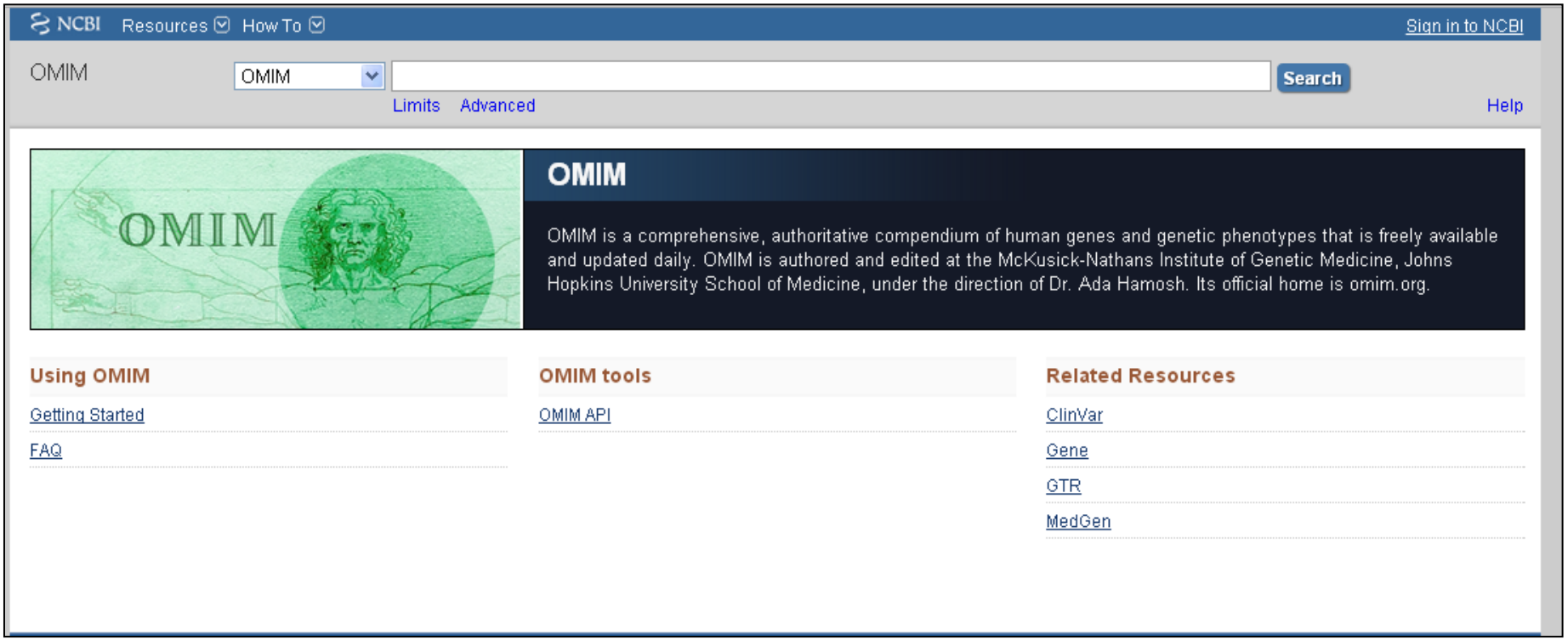
  www.ncbi.nlm.nih.gov/omim



ghr.nlm.nih.gov/



www.ncbi.nlm.nih.gov/sites/GeneTests/review



www.ncbi.nlm.nih.gov/omim

Continuum of genetic disease risks

Genetic diseases
(cystic fibrosis)

Common diseases
(diabetes, most cancers)

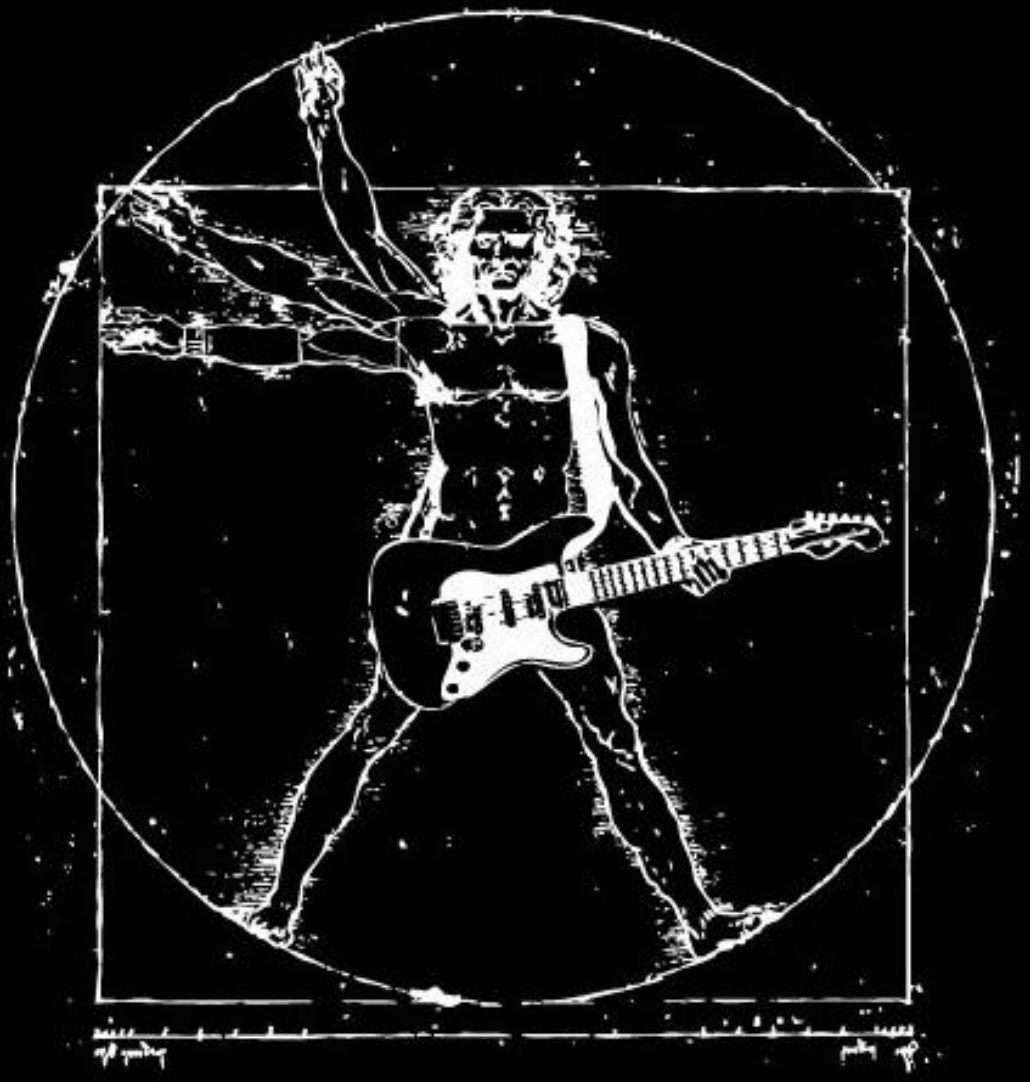
Infectious diseases
(chicken pox)

Mainly caused by genetic change

Mainly caused by genes & environment

Mainly caused by environment

Annu. Rev. Genomics Hum. Genet. 5:1–14 (2004)



Heavy Mendel

Mendelian Traits and Diseases

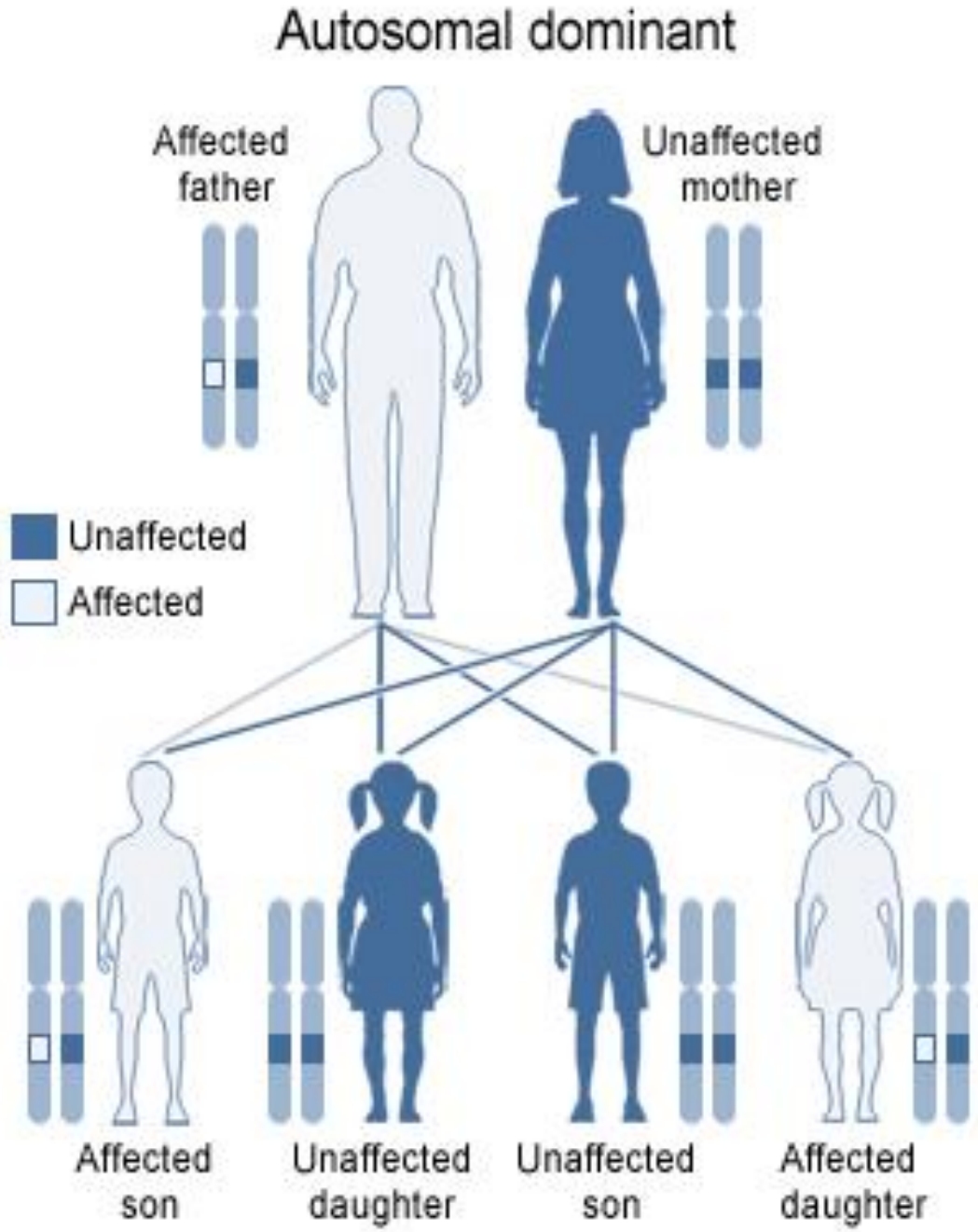
- Monogenic (single gene) traits inherited in a comparatively simple pattern
- Involve genetic variation present in the nuclear genome
- >4,000 diseases with Mendelian inheritance patterns

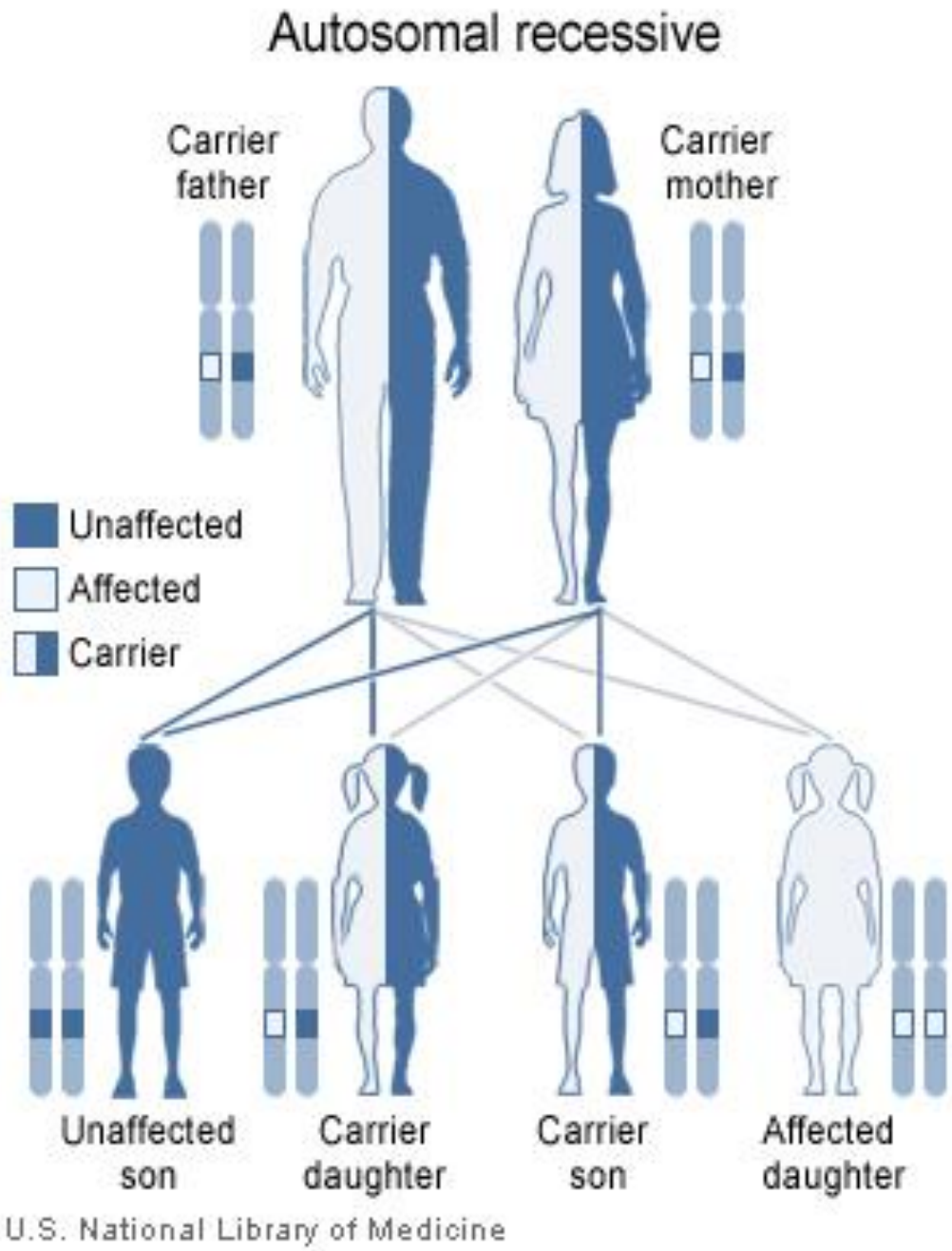
Gregor Mendel
1822 – 1884

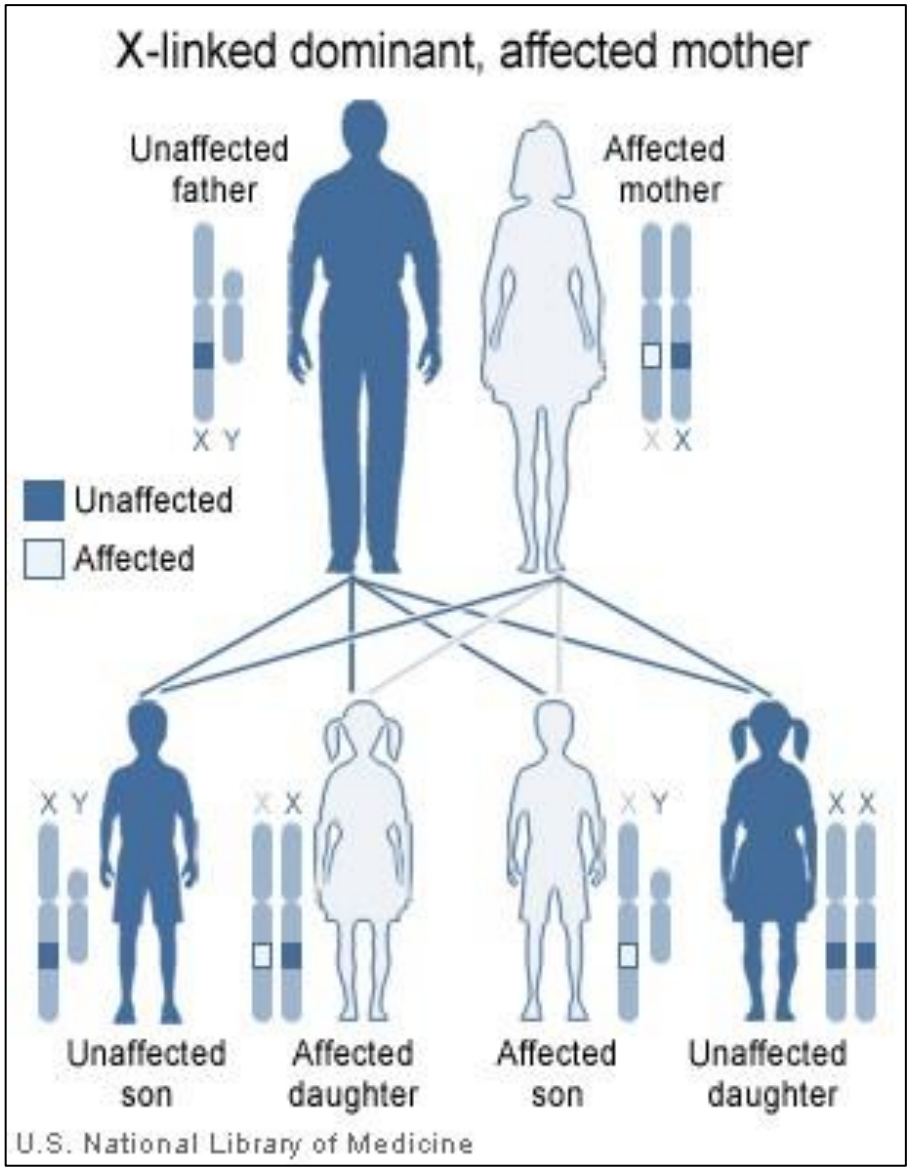

Give peas a chance

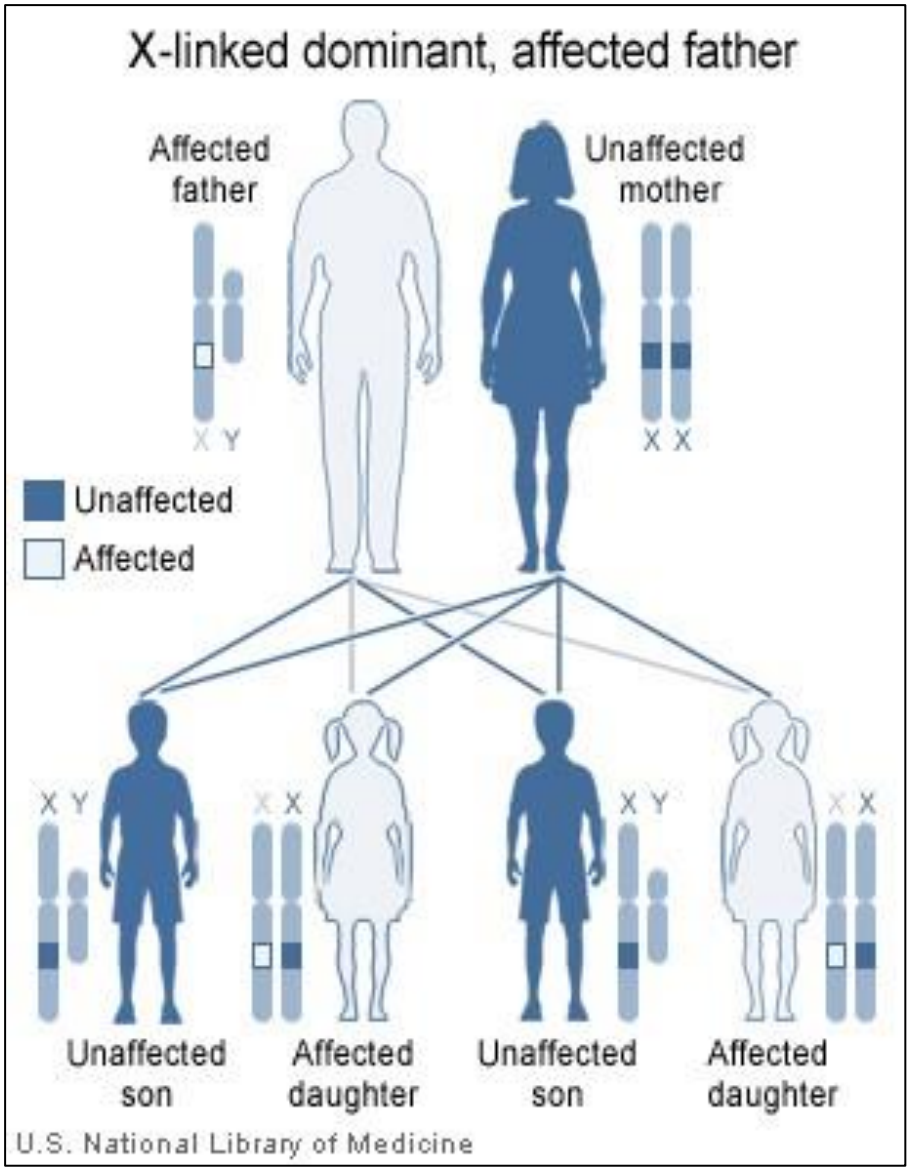
Mendelian Inheritance Patterns

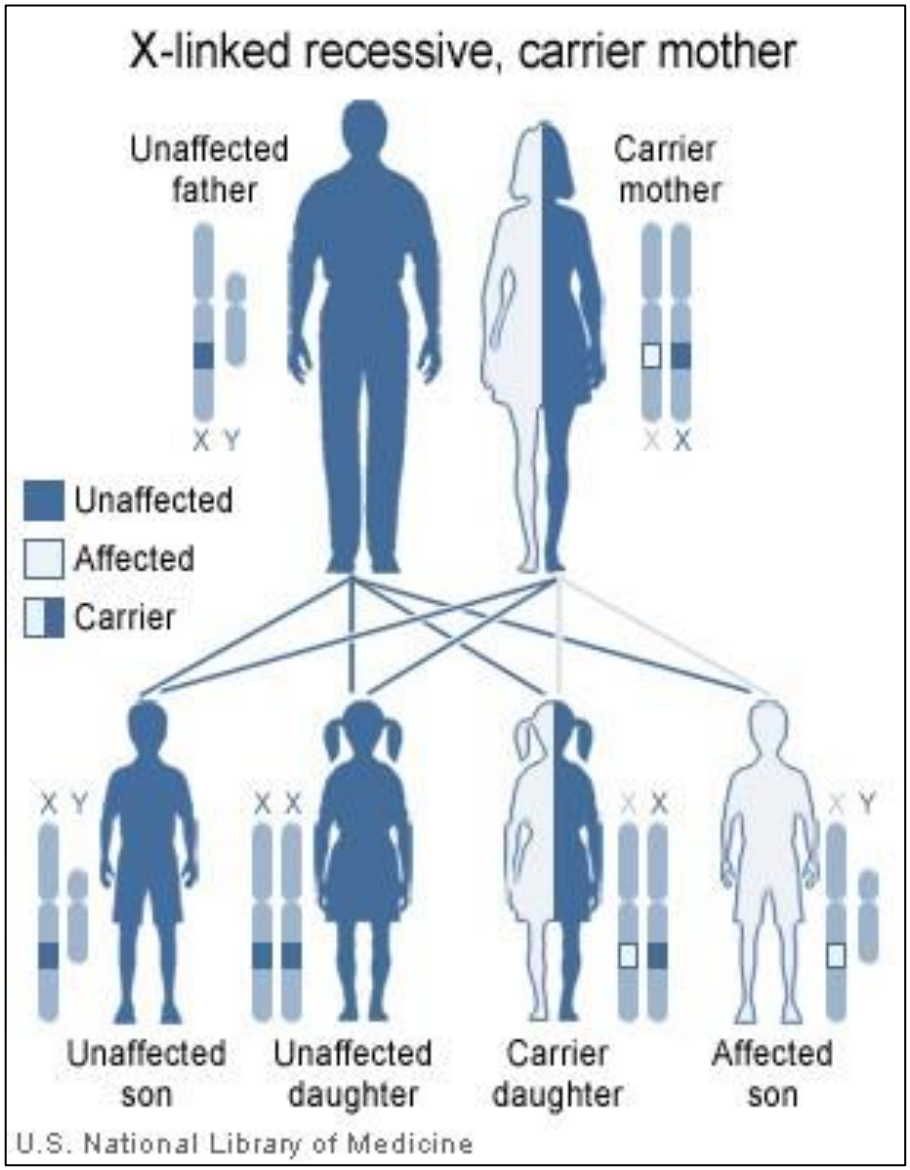
- Autosomal
- X-linked
- Y-linked (exceptionally rare)
- Dominant
- Recessive

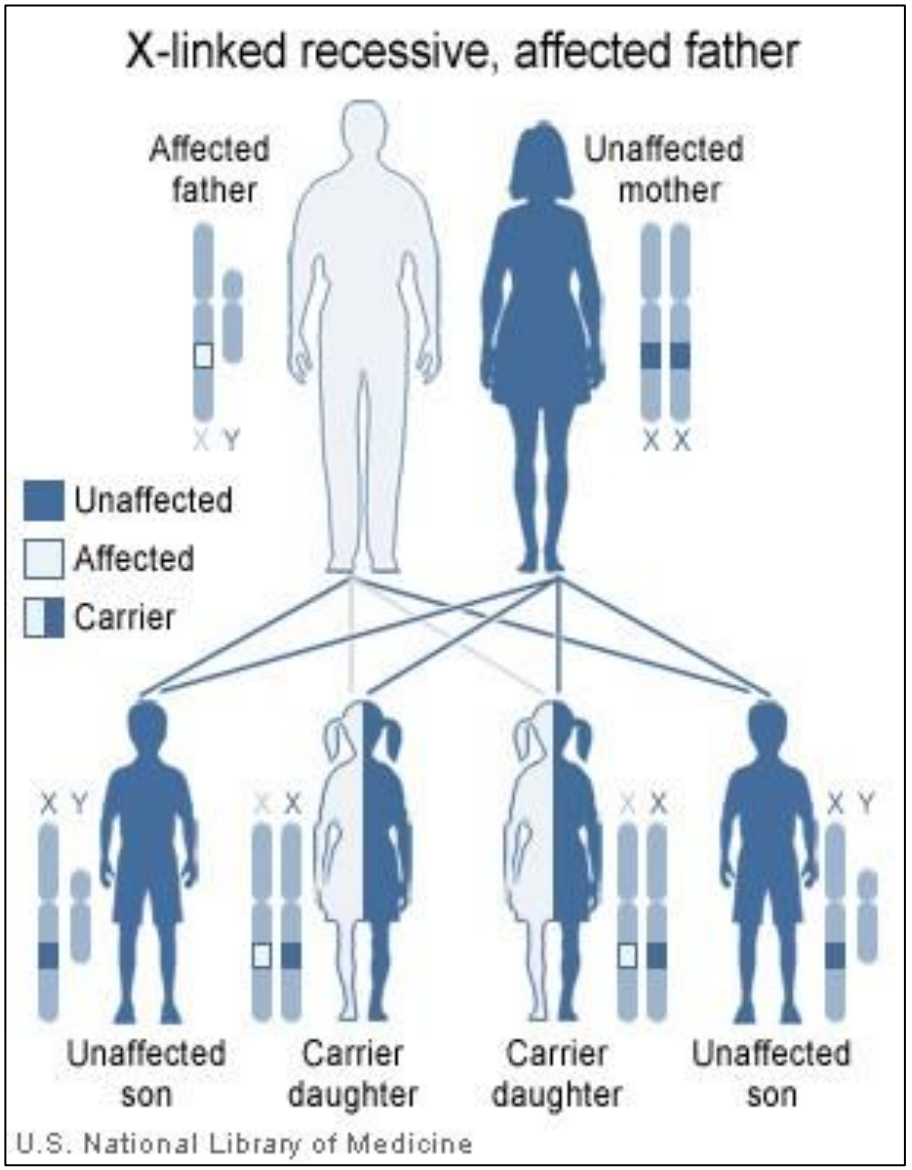

Autosomal dominant
Affected father
Unaffected mother
Unaffected
Affected
Affected son
Unaffected daughter
Unaffected son
Affected daughter
U.S. National Library of Medicine


Autosomal recessive
Carrier father
Carrier mother
Unaffected
Affected
Carrier
Unaffected son
Carrier daughter
Carrier son
Affected daughter
U.S. National Library of Medicine


X-linked dominant, affected mother
Unaffected father
Affected mother
X Y
X X
Unaffected
Affected
X Y
X X
X Y
X X
Unaffected son
Affected daughter
Affected son
Unaffected daughter
U.S. National Library of Medicine

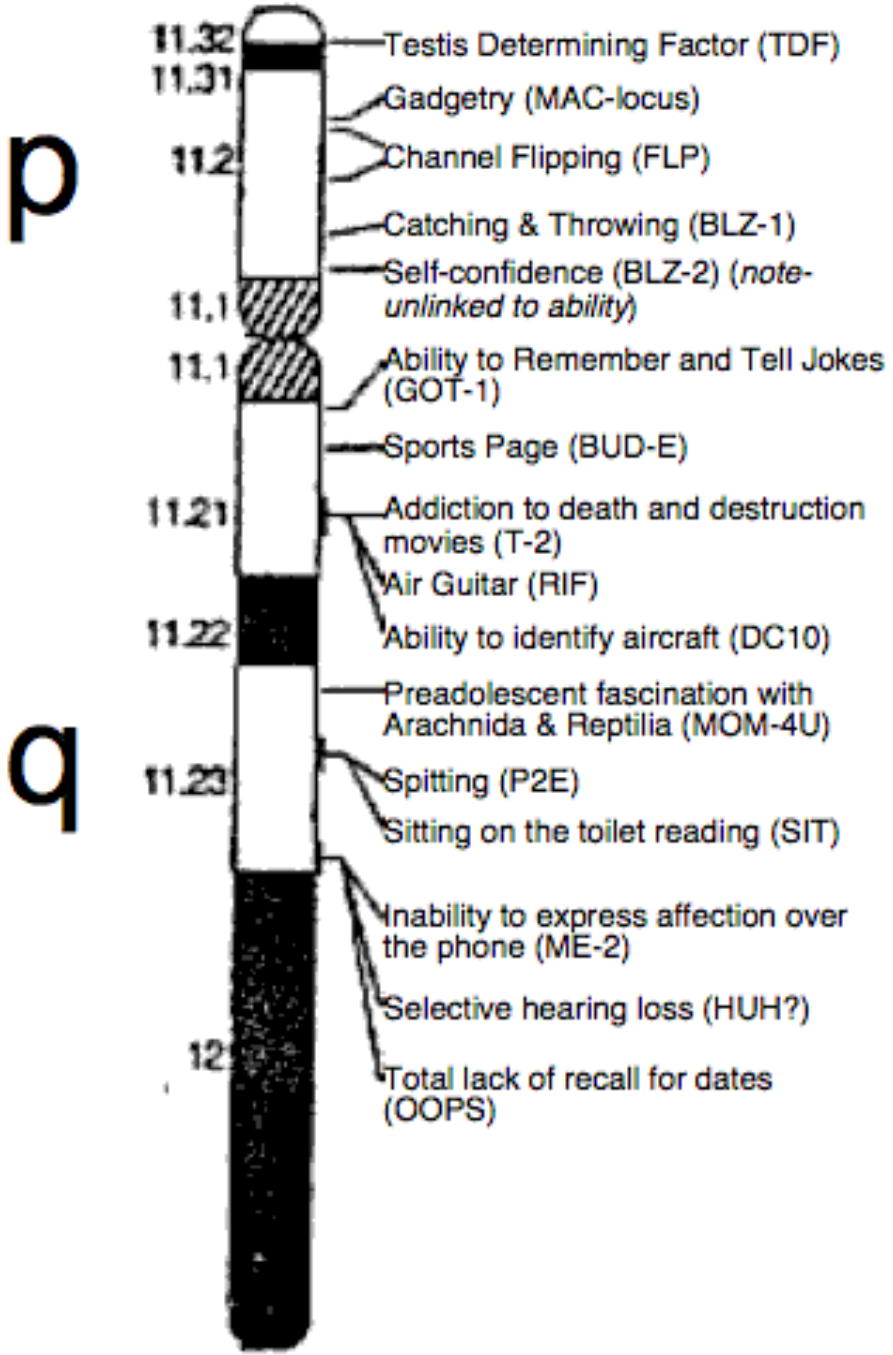

X-linked dominant, affected father
Affected father
Unaffected mother
X Y
X X
Unaffected
Affected
X Y
X X
X Y
X X
Unaffected son
Affected daughter
Unaffected son
Affected daughter
U.S. National Library of Medicine


X-linked recessive, carrier mother
Unaffected father
Carrier mother
X Y
X X
Unaffected
Affected
Carrier
X Y
X X
X X
X Y
Unaffected son
Unaffected daughter
Carrier daughter
Affected son
U.S. National Library of Medicine


X-linked recessive, affected father
Affected father
Unaffected mother
X Y
X X
Unaffected
Affected
Carrier
X Y
X X
X X
X Y
Unaffected son
Carrier daughter
Carrier daughter
Unaffected son
U.S. National Library of Medicine

Some genes on the Y chromosome:

This explains a lot!!

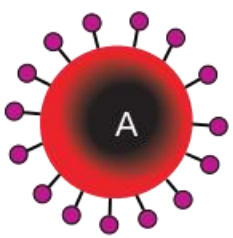
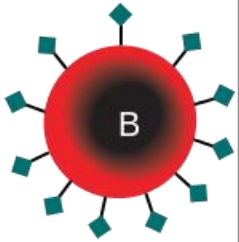
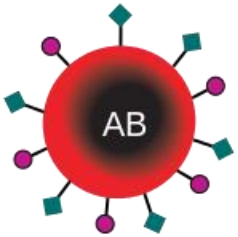
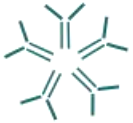
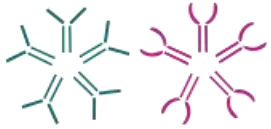


Brief look at special cases

Codominant trait

http://imgace.com/pic/tag/child-has-horizontal-and-vertical-striped-shirt/

Inheritance of ABO Blood Group System

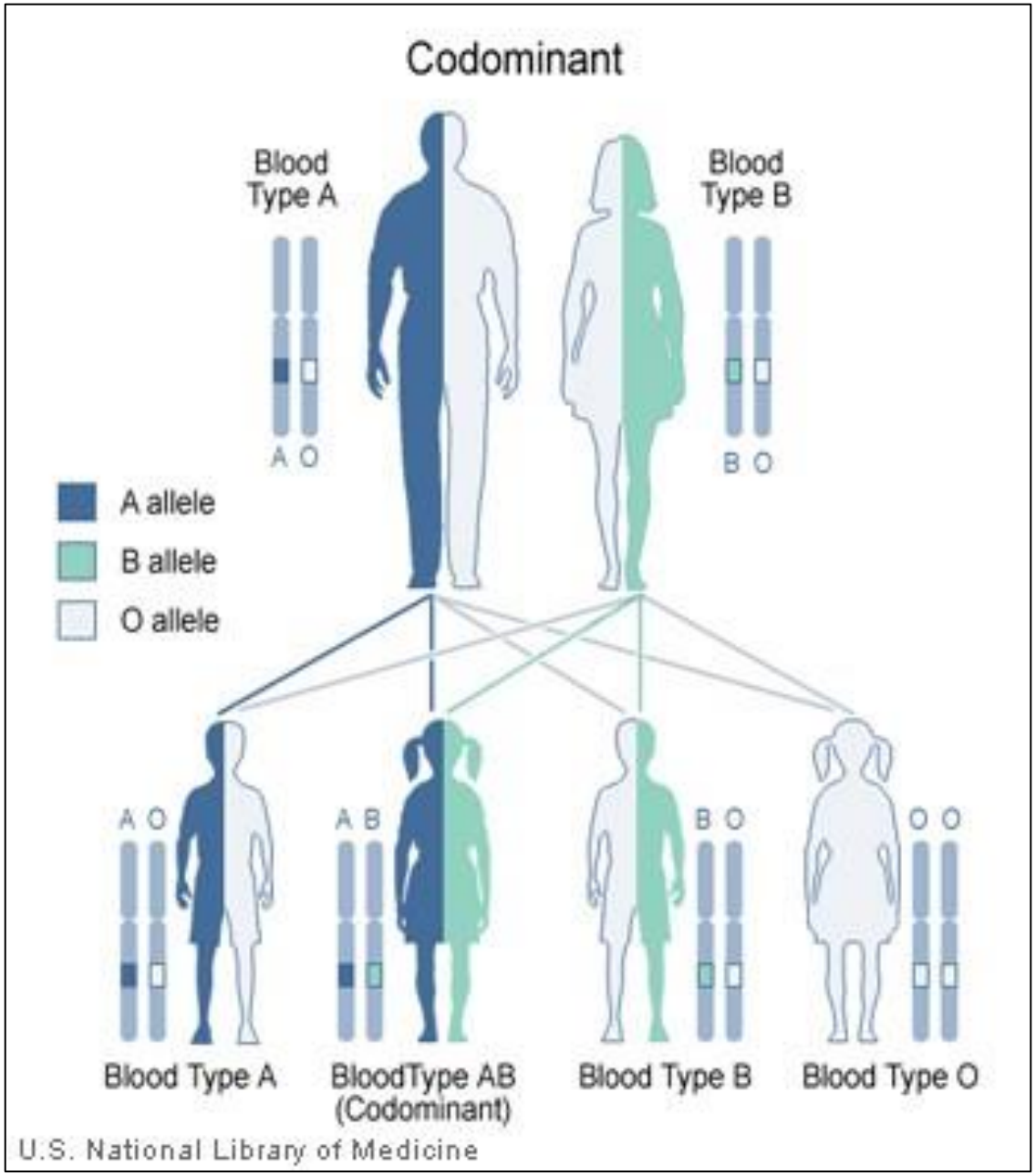
	Group A	Group B	Group AB	Group O
Red blood cell type				
Antibodies in Plasma	 Anti-B	 Anti-A	None	 Anti-A and Anti-B
Antigens in Red Blood Cell	 A antigen	 B antigen	 A and B antigens	None

en.wikipedia.org/wiki/ABO_blood_group_system

- *ABO* gene encodes a glycosyltransferase enzyme
- “A” allele: modifies the H antigen with D-galactose
- “B” allele: modifies the H antigen with N-acetylgalactosamine
- “O” allele: no activity and thus H antigen is unmodified
- A and B alleles are codominant; O allele is recessive

Scenario demonstrating codominant inheritance of A and B alleles

http://imgace.com/pic/tag/child-has-horizontal-and-vertical-striped-shirt/

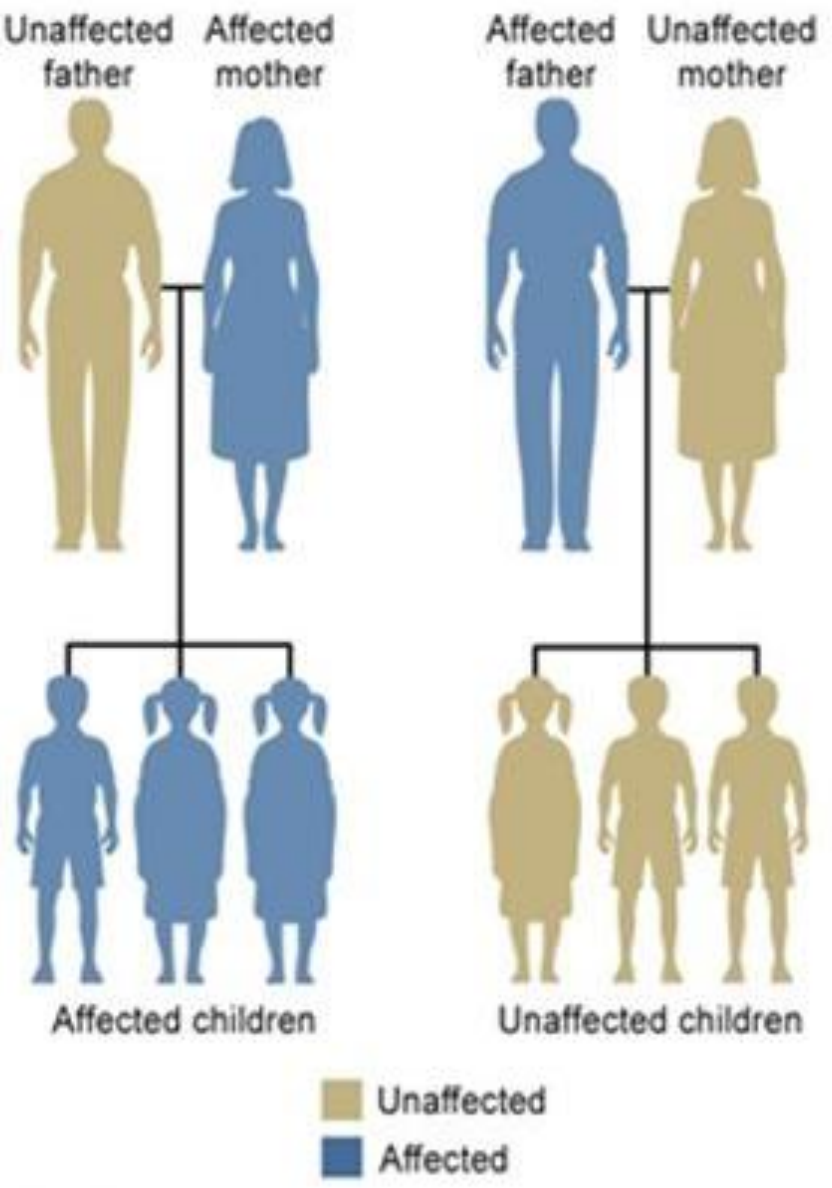


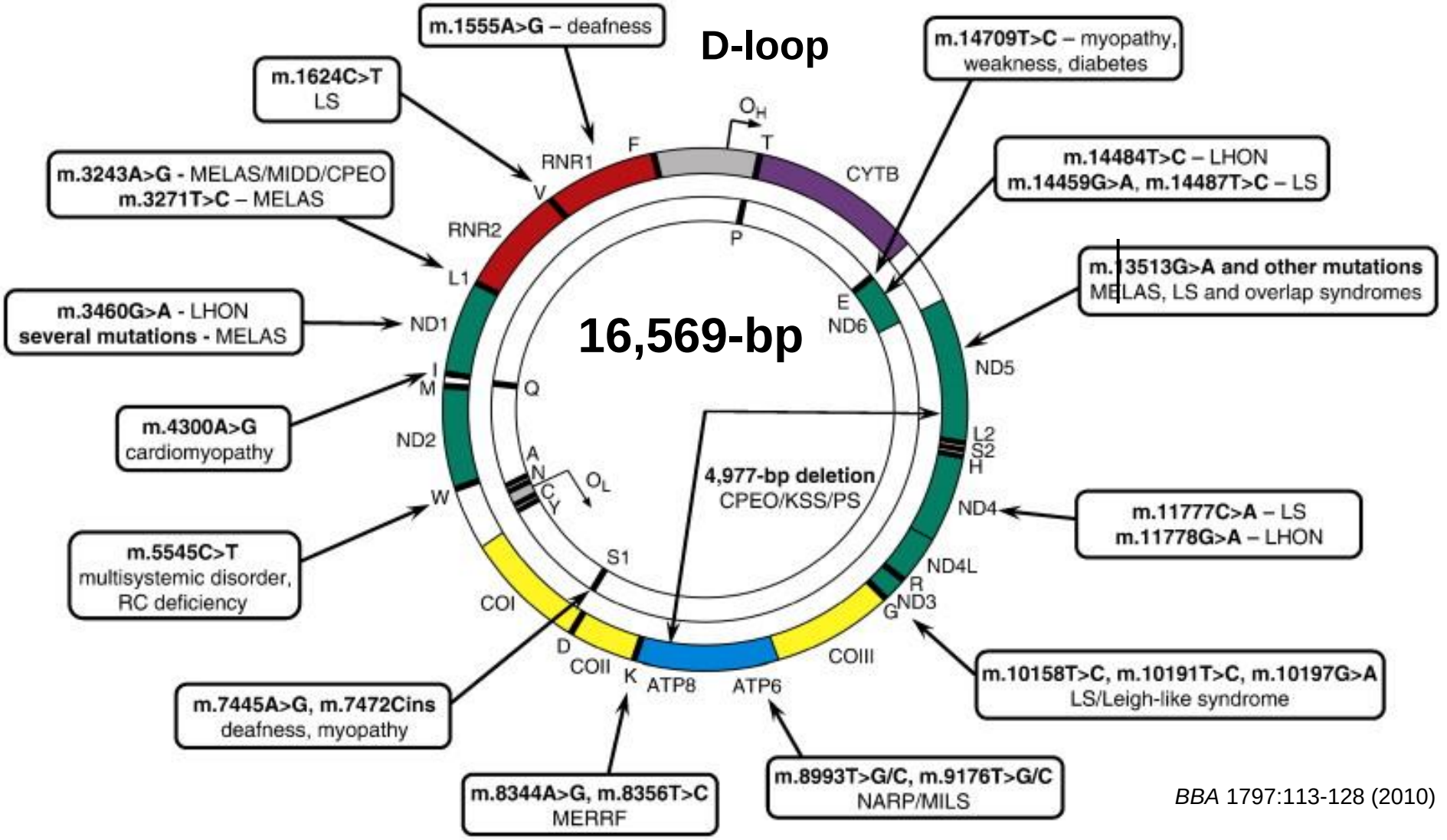
Example with heterozygous parents

Mitochondrial Disorders

- 44 diseases listed by the United Mitochondrial Disease Foundation <www.umdf.org>
- Caused by mutations in mtDNA or in genomic DNA
- Overall prevalence of ~1/8,800 individuals
- Maternal inheritance of mtDNA mutations
- Most of mitochondrial disorders with gDNA mutations are autosomal recessive
- Estimating relevance is complicated by disease complexity and difficulty in diagnosis

Mitochondrial Inheritance (Non-Mendelian)

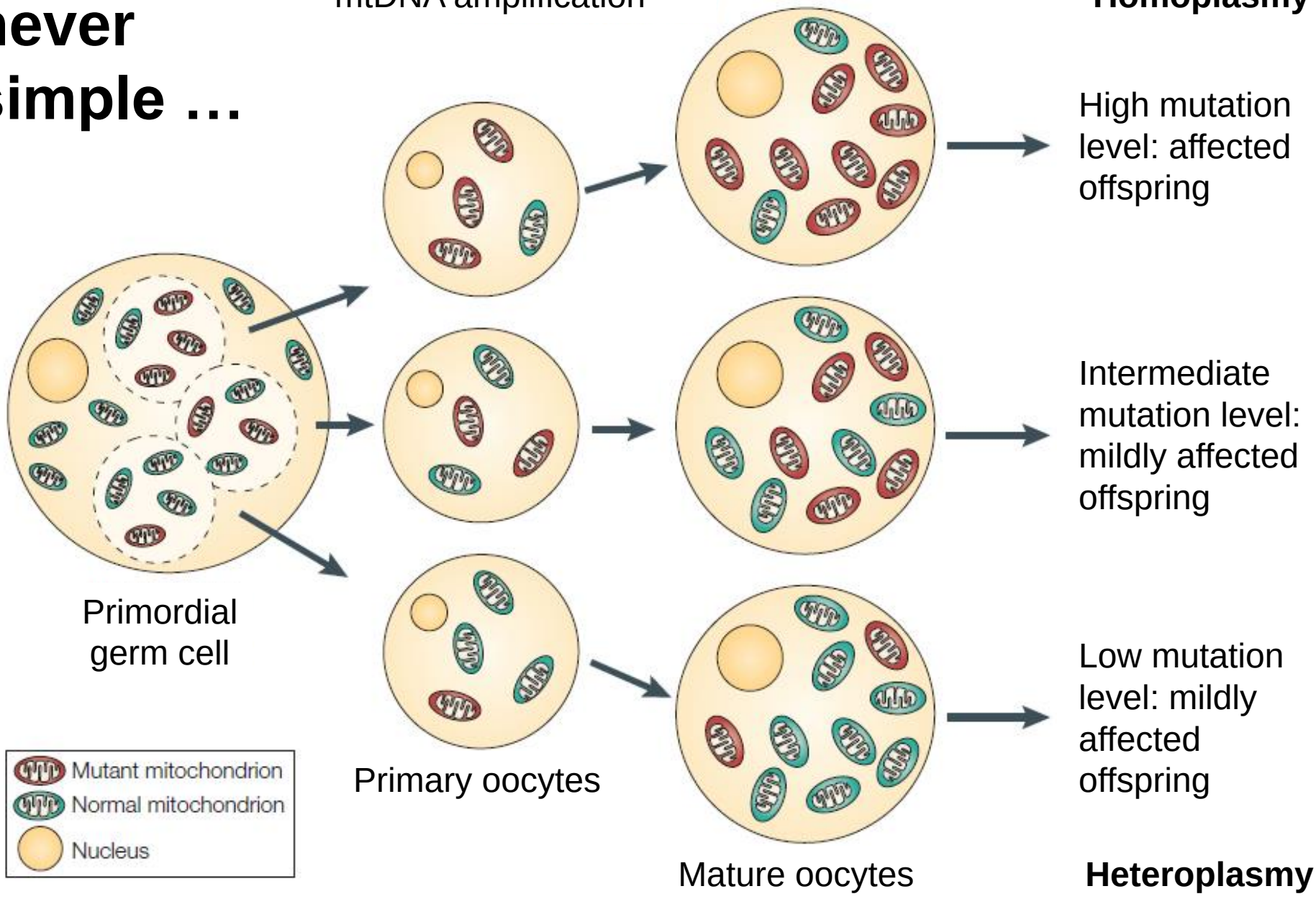




BBA 1797:113-128 (2010)

- 37 genes: 13 protein coding (all respiratory complex)
- 22 transfer RNAs, 2 ribosomal RNAs
- D-loop contains origins of replication and transcription

Life is never simple …



Nature Reviews Genet. 6: 389-402 (2005)

Mitochondrial Heteroplasmy

- Varies among cells
- Mutation-dependent
- Tissue-dependent
- Threshold effect for each tissue
  - Influences disease severity
  - Inlfuences age of onset

Heteroplasmy is common in MELAS

- Mitochondrial Myopathy, Encephalopathy, Lactic Acidosis and Stroke-Like episodes
- Prevalence still being determined
- 80% patients have A3243G mutation in tRNA-leu^{UUR}
- Many other reported phenotypes including diabetes and deafness occur due to this mutation

Homoplasmy in Leber Hereditary Optic Neuropathy (LHON)

- Likely the most common mitochondrial disorder
  - 1/30,000 – 1/50,000 in Europe
- Heteroplasmy in only 10-15% of patients
- Three common mutations in mtDNA complex I genes
- Rapid visual loss in second and third decades
  - Males are 4× more likely to go blind than females
- Low penetrance: ~50% of males and ~90% of females with LHON-causing mutations do not develop disease

Continuum of genetic disease risks

Genetic diseases
(cystic fibrosis)

Common diseases
(diabetes, most cancers)

Infectious diseases
(chicken pox)

Mainly caused by genetic change

Mainly caused by genes & environment

Mainly caused by environment

Annu. Rev. Genomics Hum. Genet. 5:1–14 (2004)

Complex traits and disorders

- Non-Mendelian inheritance patterns
- Familial aggregation (clustering in families), but no clearly defined pattern of transmission
- Complex disorders are present at higher population frequency than ‘Mendelian’ disorders

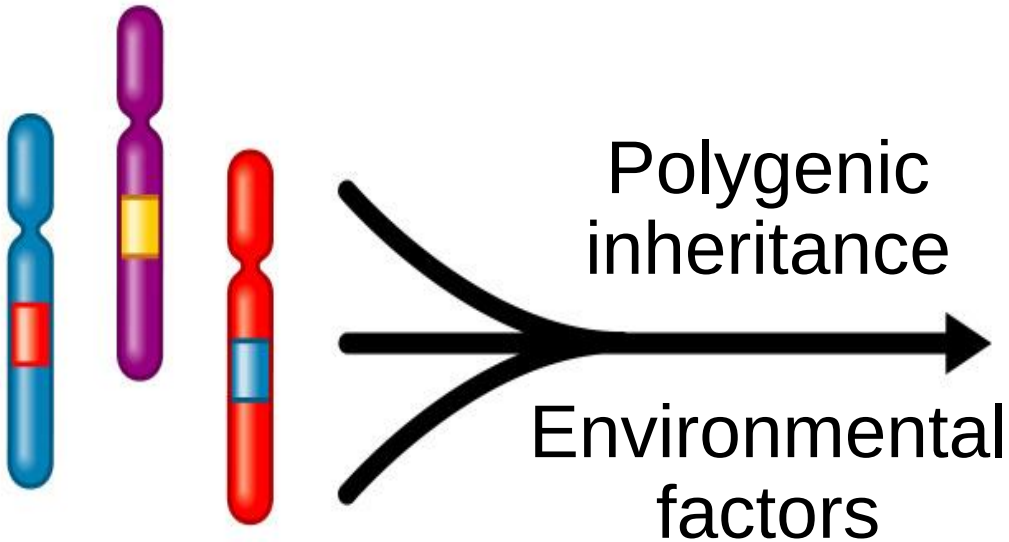
Frequency of Different Types of Genetic Diseases

Type (Disorders Due To …)	Incidence at birth (per 1000)	Prevalence at Age 25 Years (per 1000)	Population Prevalence (per 1000)
Chromosome abnormalities	6	1.8	3.8
Single-gene mutations	10	3.6	20
Multifactorial inheritance	□50	□50	□600

Adapted from Thompson and Thompson, Genetics in Medicine 7th Edition, Table 8-1

Complex traits and disorders

- Can be polygenic: many loci, each with small effect, but no environmental factors (e.g. eye color)
- Can be multifactorial: dependent on a combination of genetic and environmental factors (most of the time)



Multiple genes

Polygenic inheritance

Environmental factors

Alzheimer Disease
Arthritis
Autism
Orofacial Clefting
Schizophrenia
Diabetes

Relative risk ratios (λr) of some common diseases

Disease	Relation	λr
Schizophrenia	Sibling	12
Autism	Sibling	150
Manic-depressive disorder	Sibling	7
Type 1 diabetes mellitus	Sibling	35
Crohn's disease	Sibling	25
Multiple sclerosis	Sibling	24

$$\frac{\text{Prevalence of disease in relatives of affected person}}{\text{Prevalence of disease in general population}}$$

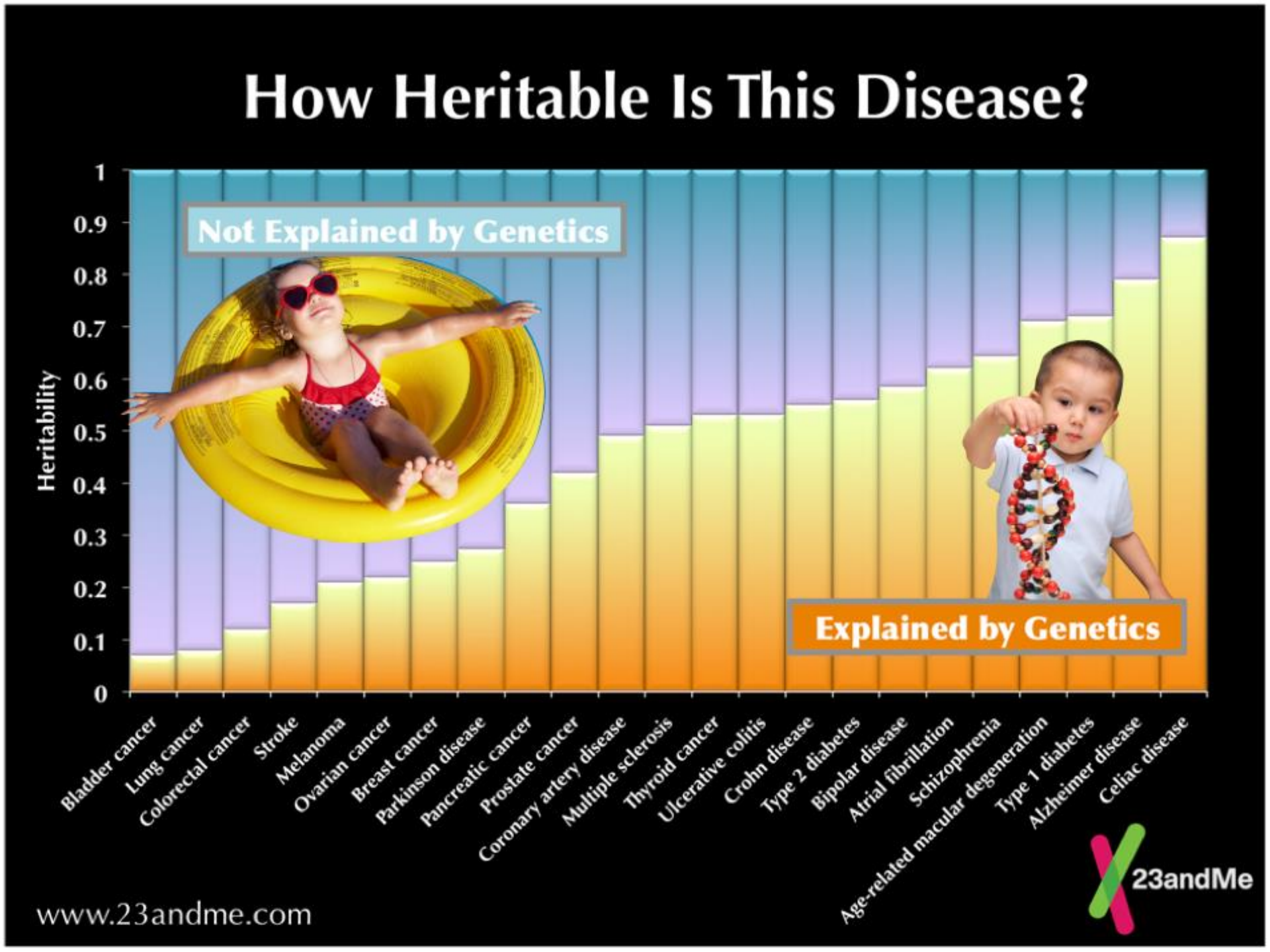
Heritability

- Fraction of total phenotypic variance of a quantitative trait that is caused by genes
- Measures the extent to which different alleles at various loci are responsible for the variability in that trait seen across a population
- Heritability (h^2) can be estimated based on twin studies
  - Ranges from 0 (no genetic basis) to 1 (complete genetic basis)

$$\frac{\text{Variance in DZ pairs} - \text{Variance in MZ pairs}}{\text{Variance in DZ pairs}}$$

DZ = dizygotic (fraternal twins) ; MZ = monozygotic (identical twins)

How Heritable Is This Disease?



www.23andme.com

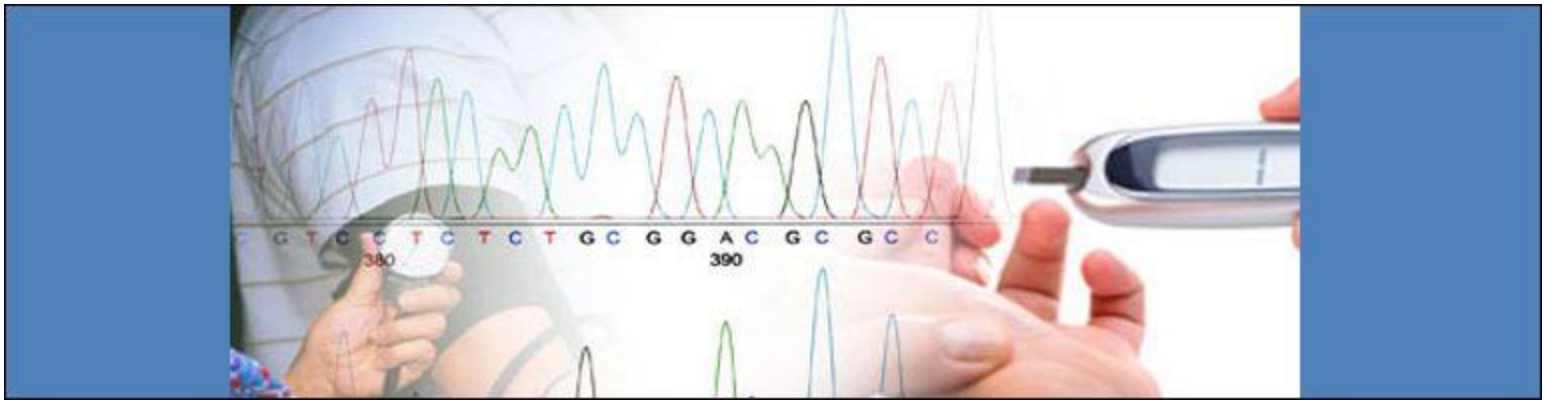
23andMe

Common Disease – Common Variant Hypothesis

Common, interacting disease alleles underlie most common diseases, perhaps in association with environmental factors

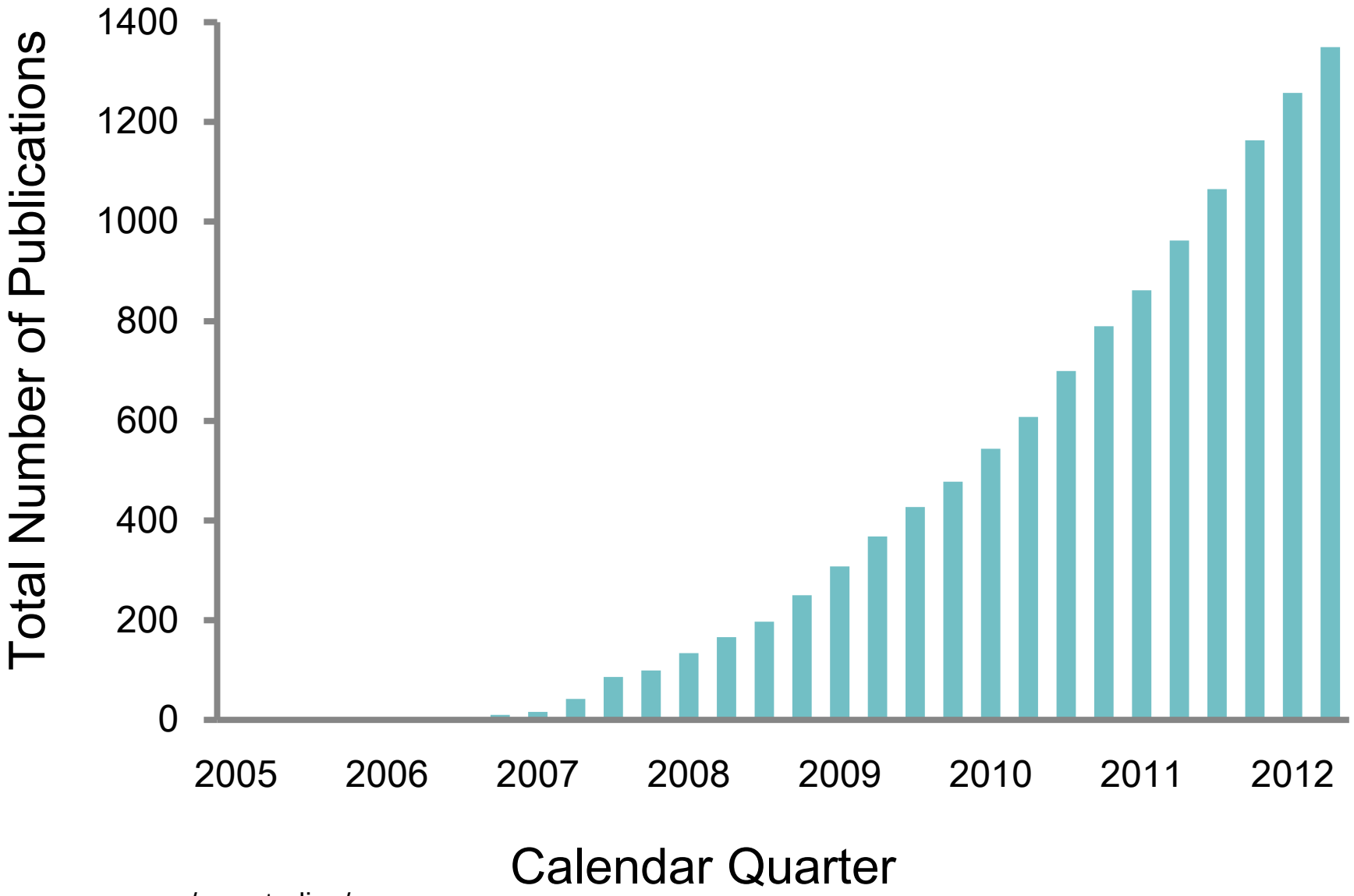
Genome Wide Association Studies (GWAS)

- Test a significant portion of common genetic variation in human populations for association with a disease or variation in a quantitative trait
- Find disease/quantitative trait-related variants without a prior hypothesis of gene function



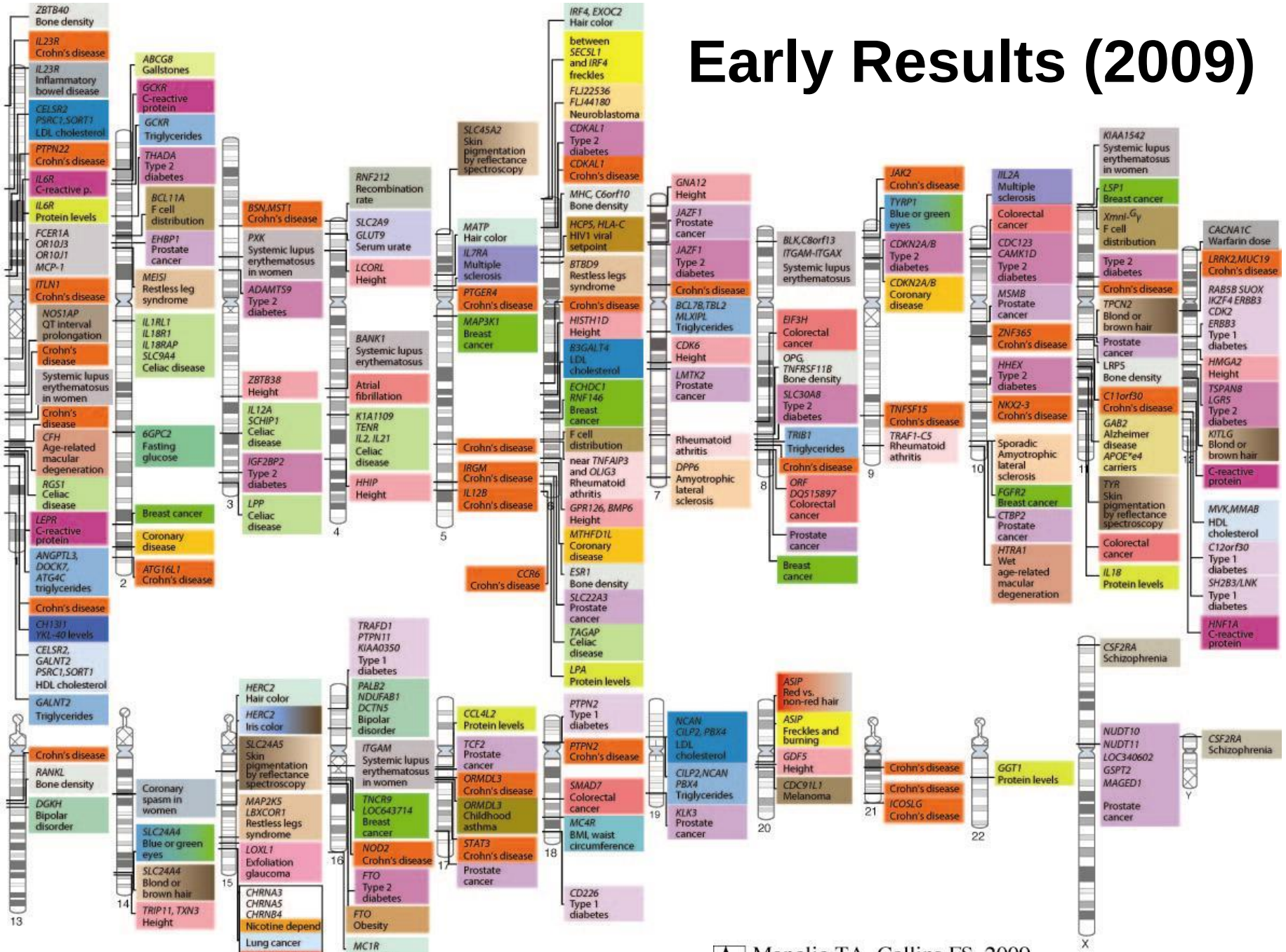
gwas.nih.gov

Published GWA Reports, 2005 – 6/2012



www.genome.gov/gwastudies/

Early Results (2009)



Manolio TA, Collins FS. 2009.
Annu. Rev. Med. 60:443–56

genome.gov
National Human Genome Research Institute
National Institutes of Health

Google™ Search SEARCH

Research Funding | Research at NHGRI | Health | Education | Issues in Genetics | Newsroom | Careers & Training | About | For You

Home › Research Funding › Research Funding Divisions › Division of Genomic Medicine › **GWAS Catalog**

Share Print

Division of Genomic Medicine

A Catalog of Published Genome-Wide Association Studies

Division Staff | Funding Opportunities | Genomic Medicine Activities | GWAS Catalog | Meetings & Workshops | Potential Sample Collections for Sequencing | Programs | Publications | Trans-NIH Sequencing Inventory

Additional information has been added to the HTML catalog columns below. For a description of column headings for the HTML catalog, go to: Catalog Heading Descriptions PDF new

Potential etiologic and functional implications of genome-wide association loci for human diseases and traits PDF
Click here to read our recent *Proceedings of the Academy of Sciences (PNAS)* article on catalog methods and analysis.

View the Interactive Diagram new **View the Full Catalog** **Download the Catalog** **Search the Catalog**

Published Genome-Wide Associations

The genome-wide association study (GWAS) publications listed here include only those attempting to assay at least 100,000 single nucleotide polymorphisms (SNPs) in the initial stage. Publications are organized from most to least recent date of publication, indexing from online publication if available. Studies focusing only on candidate genes are excluded from this catalog. Studies are identified through weekly PubMed literature searches, daily NIH-distributed compilations of news and media reports, and occasional comparisons with an existing database of GWAS literature (HuGE Navigator).

SNP-trait associations listed here are limited to those with p-values < 1.0 x 10-5 (see full methods for additional details). Multipliers of powers of 10 in p-values are rounded to the nearest single digit; odds ratios and allele frequencies are rounded to two decimals. Standard errors are converted to 95 percent confidence intervals where applicable. Allele frequencies, p-values, and odds ratios derived from the largest sample size, typically a combined analysis (initial plus replication studies), are recorded below if reported; otherwise statistics from the initial study sample are recorded. For quantitative traits, information on % variance explained, SD increment, or unit difference is reported where available. Odds ratios < 1 in the original paper are converted to OR > 1 for the alternate allele. Where results from multiple genetic models

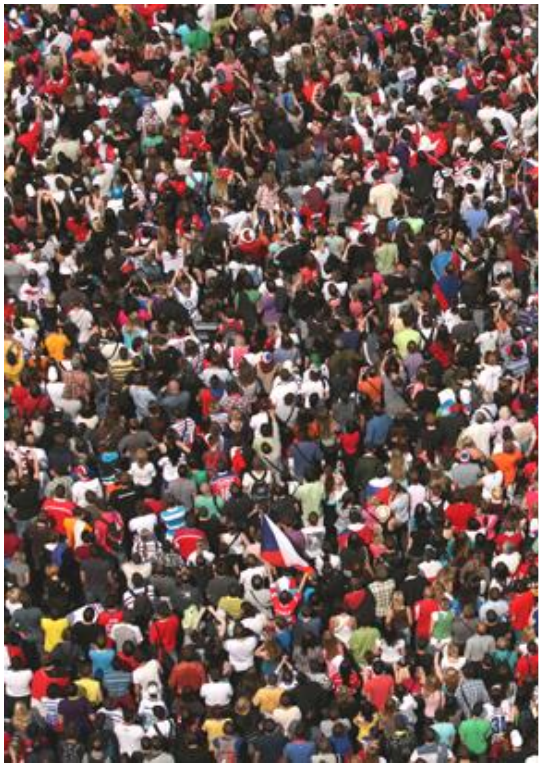
www.genome.gov/GWAStudies/

Fundamentals of GWAS

- Study Design
  - Acquire DNA from appropriate individuals with disease or trait of interest and matching controls
- Genotyping
  - Determine allele frequencies in each group
- Statistical Analysis
  - Identify alleles found more frequently than expected by chance in the affected relative to control group
- Validation
  - Confirm findings in other case-control studies

Study Design

Strategies to maximize the size of genetic effects



Cases

- More extreme phenotypes
- Other relative has phenotype
- Younger age-of-disease onset

Controls

- Less extreme phenotypes
- No family history of phenotype
- Matched ancestry, age, sex, etc.

Many thousands of DNA samples are genotyped

Genotyping

In theory, 3 SNPs can appear in 8 different patterns called ‘haplotypes’

ACCGTAACGCGCTAGCACATGCTACCGTC

ACCGTAACGCGCTAGCACATGCTGCCGTC

ACCGTAACGCGCTAGCCCATGCTACCGTC

ACCGTAACGCGCTAGCCCATGCTGCCGTC

ACCGTAATGCGCTAGCACATGCTACCGTC

ACCGTAATGCGCTAGCACATGCTGCCGTC

ACCGTAATGCGCTAGCCCATGCTACCGTC

ACCGTAATGCGCTAGCCCATGCTGCCGTC

In practice, not all ‘haplotypes’ are present in a given population

ACCGTAACGCGCTAGCACATGCTACCGTC

ACCGTAACGCGCTAGCACATGCTGCCGTC

ACCGTAACGCGCTAGCCCATGCTACCGTC

ACCGTAACGCGCTAGCCCATGCTGCCGTC

ACCGTAATGCGCTAGCACATGCTACCGTC

ACCGTAATGCGCTAGCACATGCTGCCGTC

ACCGTAATGCGCTAGCCCATGCTACCGTC

ACCGTAATGCGCTAGCCCATGCTGCCGTC

These 3 variants are in linkage disequilibrium

ACCGTAACGCGCTAGCACATGCTACCGTC

ACCGTAACGCGCTAGCACATGCTGCCGTC

ACCGTAACGCGCTAGCCCATGCTACCGTC

ACCGTAACGCGCTAGCCCATGCTGCCGTC

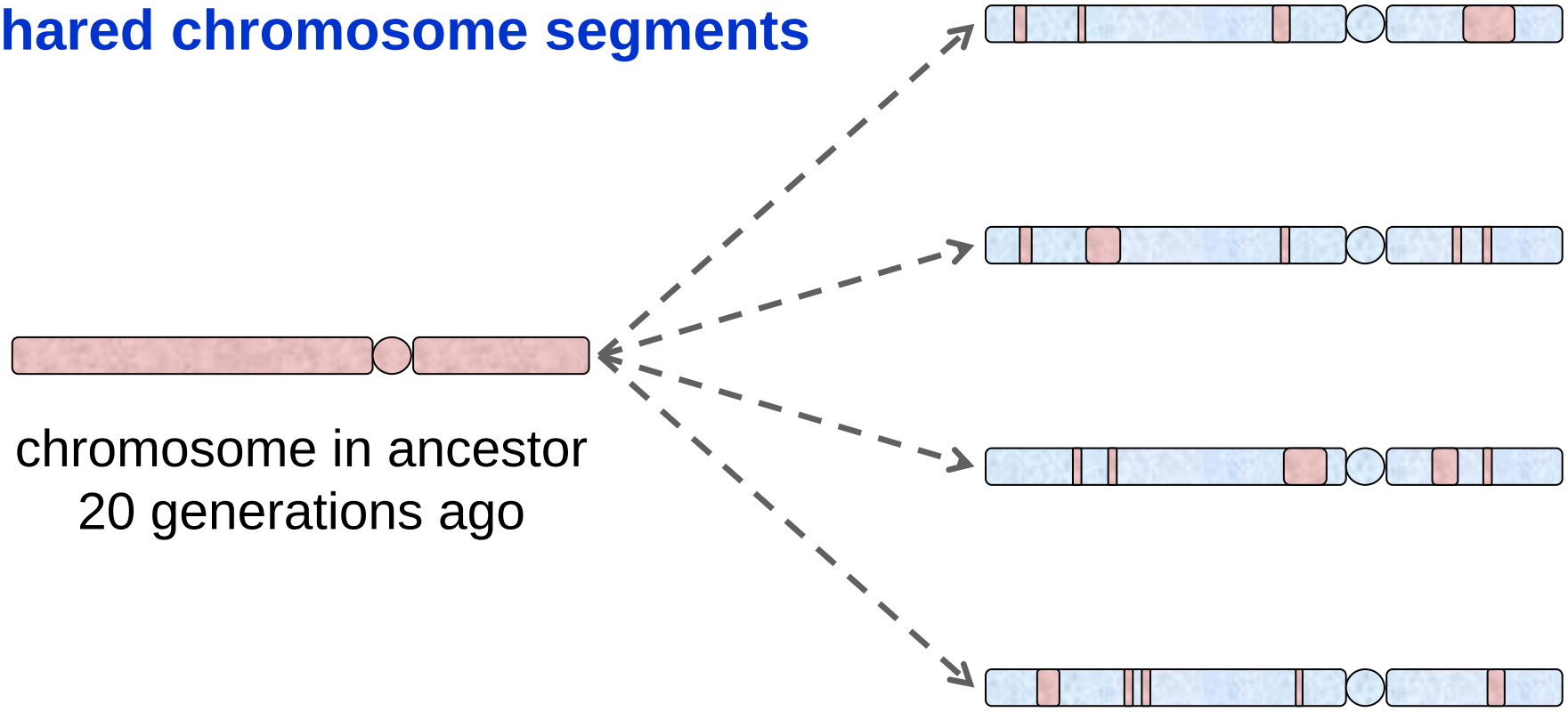
ACCGTAATGCGCTAGCACATGCTACCGTC

ACCGTAATGCGCTAGCACATGCTGCCGTC

ACCGTAATGCGCTAGCCCATGCTACCGTC

ACCGTAATGCGCTAGCCCATGCTGCCGTC

Effect of recombination on shared chromosome segments



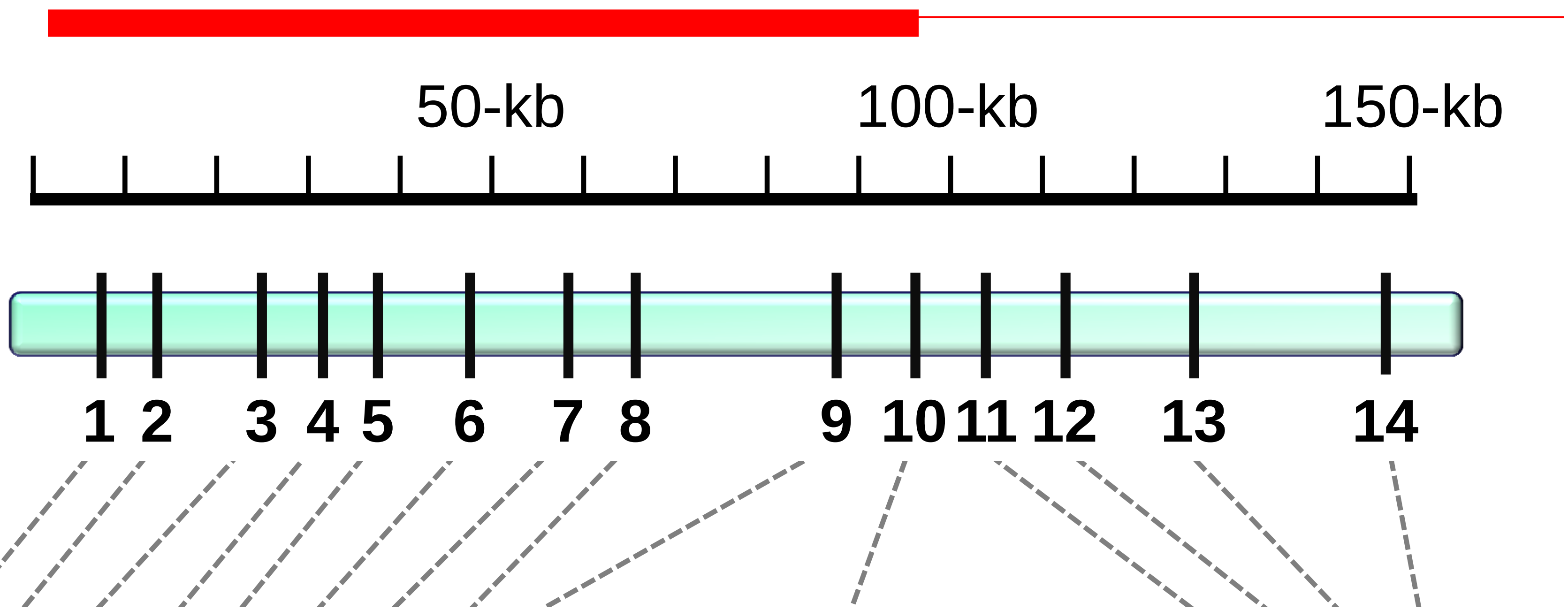
chromosome in ancestor
20 generations ago

same chromosome in four
present day descendents

20 meioses with 1-2 random crossovers in each arm per meiosis

Ancestral segments shared by a significant number of descendents in this situation is typically 5 – 15-kb

Linkage Disequilibrium (LD) Blocks



G	T	C	C	T	C	C	G	A	40%
A	C	T	T	G	T	T	T	T	30%
A	C	C	C	T	C	T	G	A	11%
G	T	C	C	T	C	T	G	A	9%
G	T	T	C	T	C	T	G	A	8%
									98%

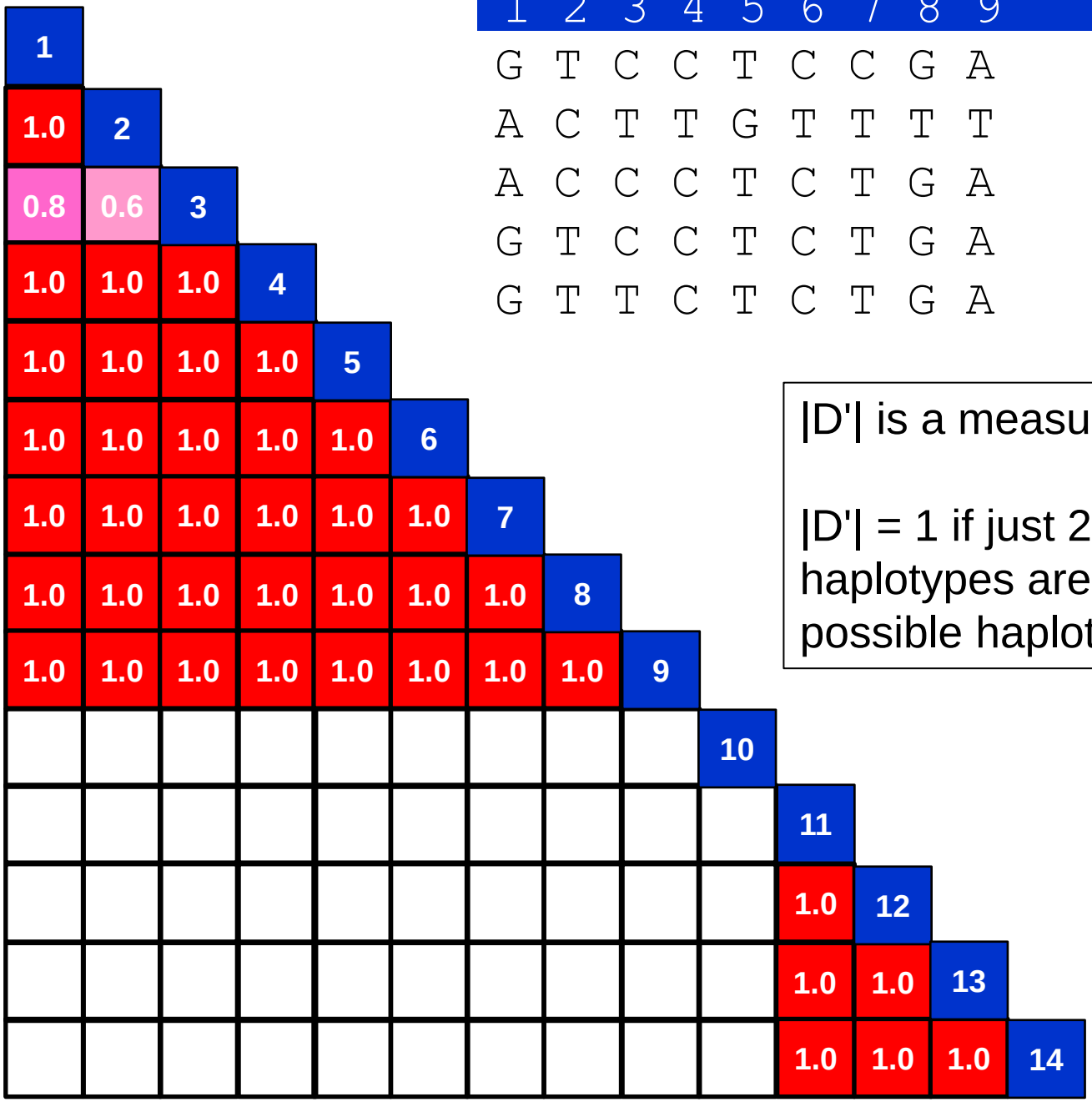
G	40%
A	60%
	100%

C	G	T	A	42%
T	G	T	A	31%
C	T	C	T	26%
				100%

CLUSTER_1

CLUSTER_2

1	2	3	4	5	6	7	8	9	10	11	12	13	14
G	T	C	C	T	C	C	G	A	G	C	G	T	A
A	C	T	T	G	T	T	T	T	A	T	G	T	A
A	C	C	C	T	C	T	G	A		C	T	C	T
G	T	C	C	T	C	T	G	A					
G	T	T	C	T	C	T	G	A					



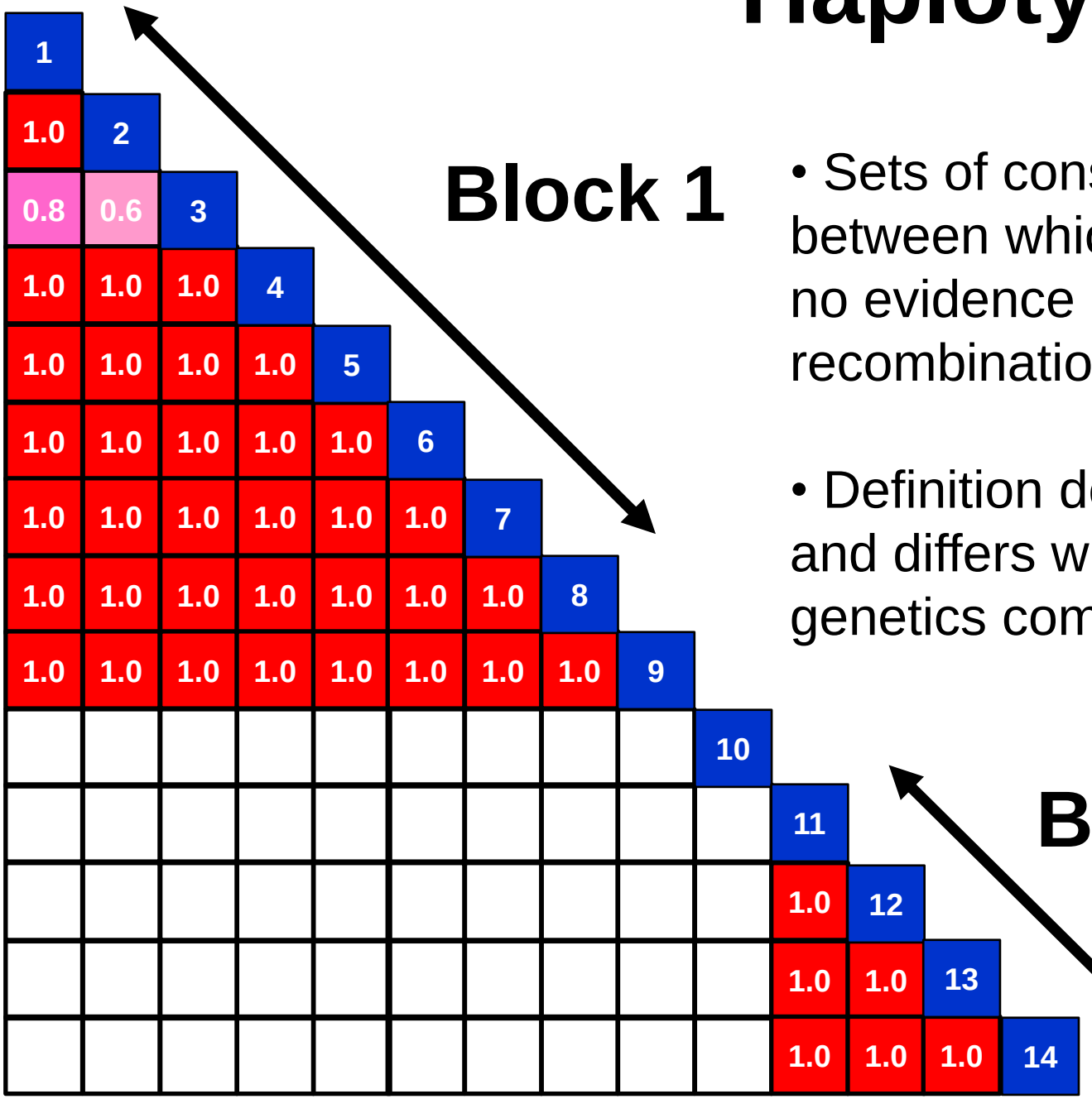
|D'| is a measure of LD

|D'| = 1 if just 2 or 3 of the possible haplotypes are present, and <1 if all 4 possible haplotypes are present

|D'| value Color Scale

1.0	0.8	0.6	>0.06

Haplotype Blocks



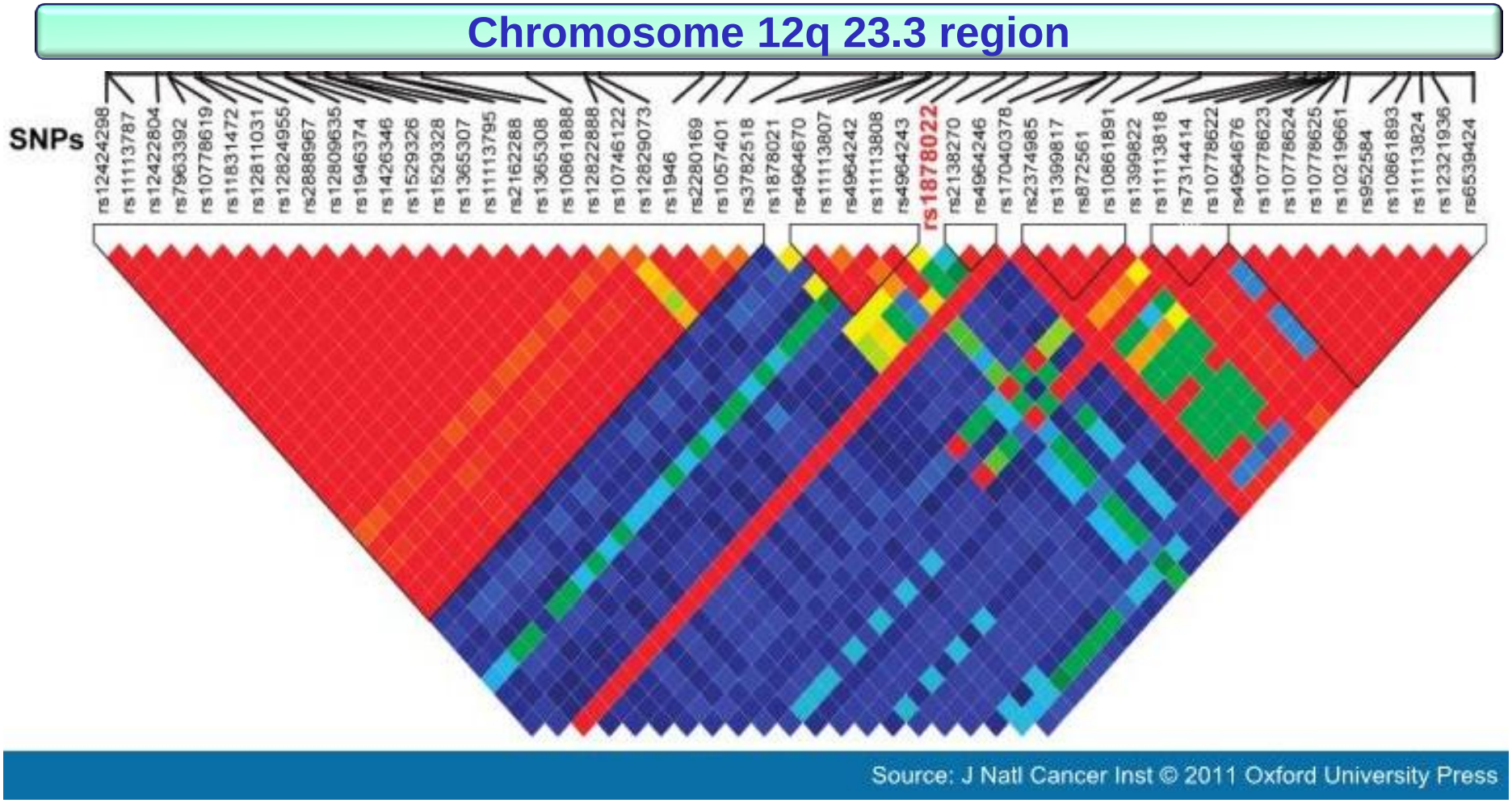
Block 1

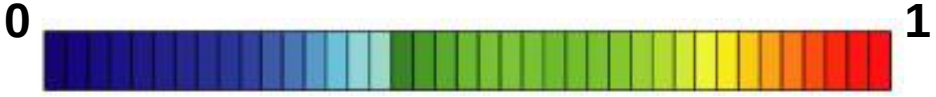
Block 2

- Sets of consecutive sites between which there is little or no evidence of historical recombination.

- Definition depends on context and differs within the population genetics community

Haptopype blocks in GWAS of Chemotherapy Responses





|D'| value color scale

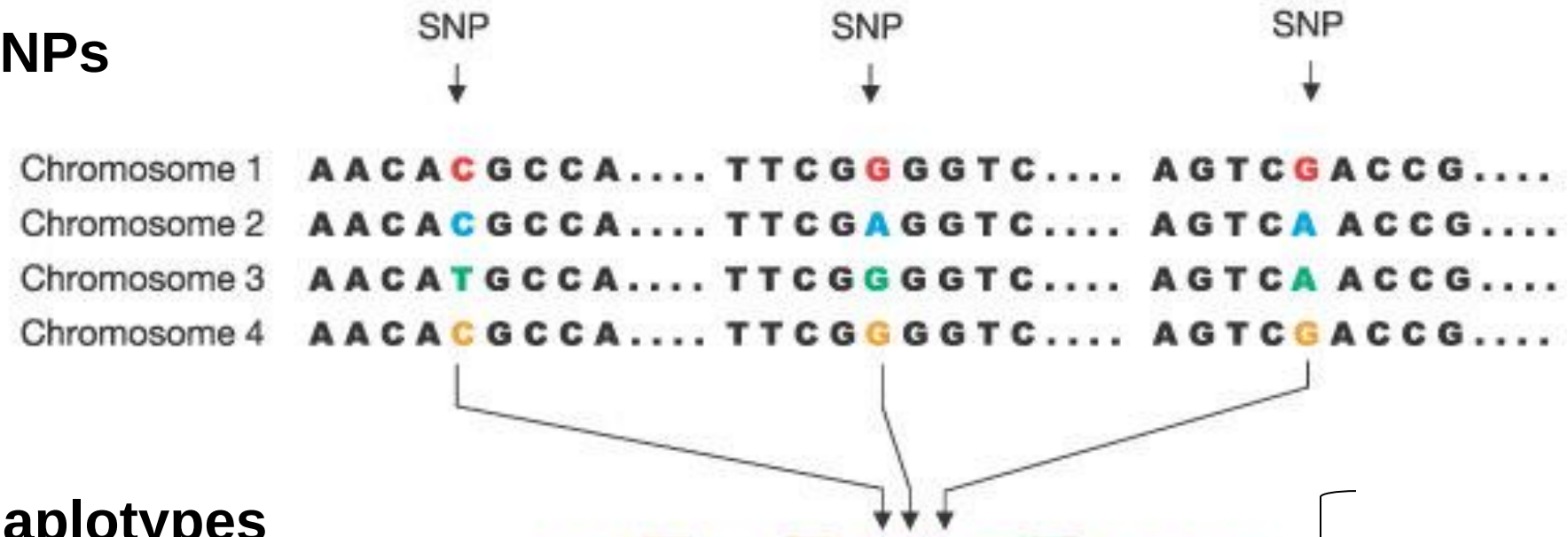
J Natl Cancer Inst. 103:817-825 (2011)

Haplotype blocks in human populations

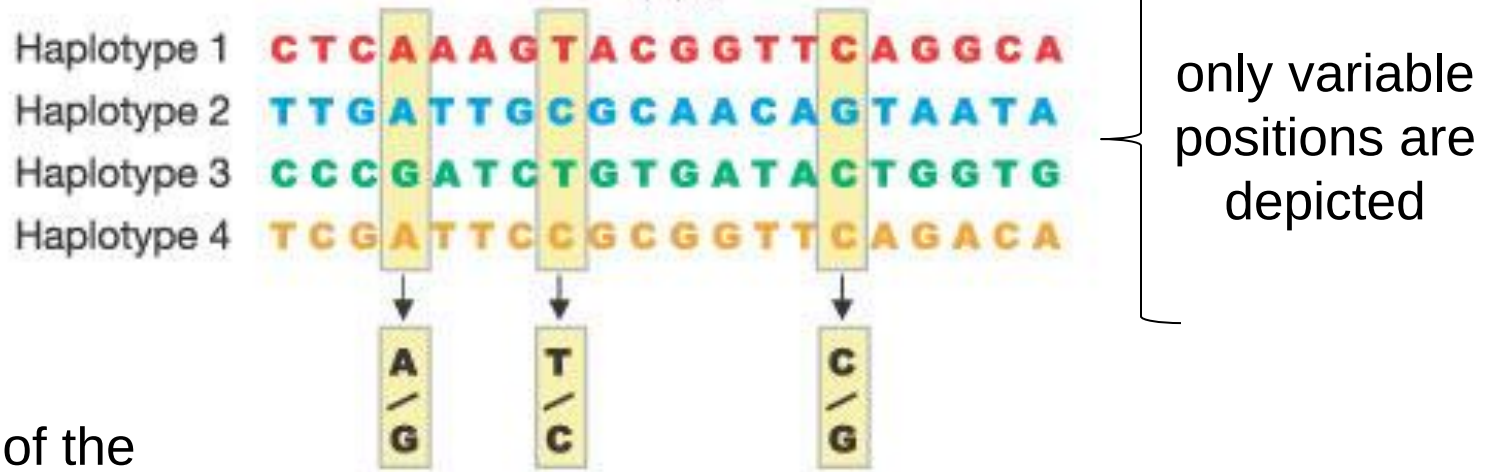
Parameter	YRI (Nigerian)	CEU (N & W Europe)	CHB & JPT (Chinese & Japanese)
Average # SNPs per block	30.3	70.1	54.4
Average length per block (kb)	7.3	16.3	13.2
% of genome spanned by block	67	87	81
Average # haplotypes per block	5.6	4.7	4.0
% of chromosomes accounted for by these haplotypes	94	93	95

Nature 437: 1299-1320 (2005)
http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/README.populations

SNPs



Haplotypes



Tag SNPs

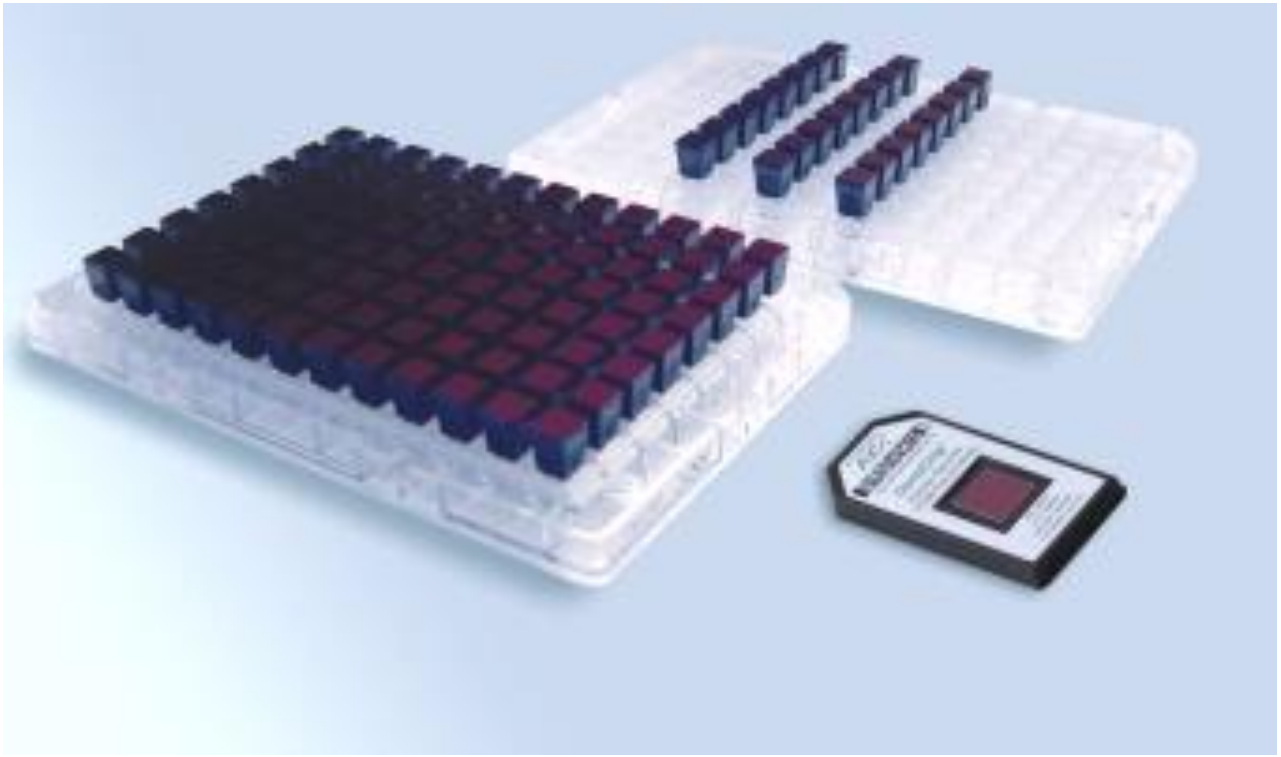
capture the bulk of the information content

Practical application:

Tag SNPs greatly reduce the genotyping workload!

Nature 426, 789-796 (2003)

Commonly Used Genotyping Platforms

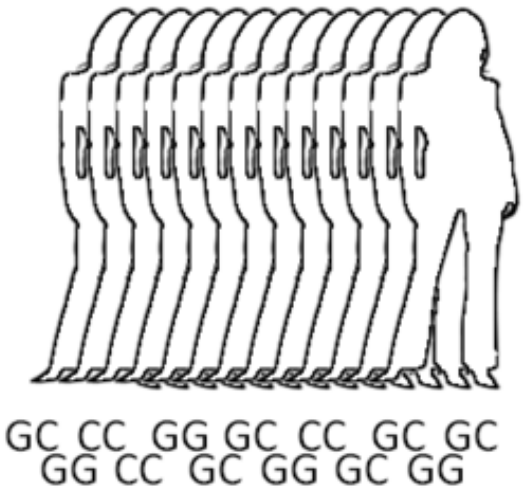


www.atlas-biolabs.de/snp_genotyping/affymetrix_snp_genotyping_service

Genotype several million human SNPs

www.illumina.com/products/humanomni5-quad_beadchip_kit.ilmn

Statistical Analysis



SNP1	*SNP2*	*SNP …*
Cases Count of G: 2104 of 4000 Frequency of G: 52.6%	**Cases** Count of G: 1648 of 4000 Frequency of G: 41.2%	*Repeat for all SNPs*
Controls Count of G: 2676 of 6000 Frequency of G: 44.6%	**Controls** Count of G: 2532 of 6000 Frequency of G: 42.2%	
P-value: $5.0 \cdot 10^{-15}$	**P-value:** 0.33	

en.wikipedia.org/wiki/File:Method_example_for_GWA_study_designs.png

Pearson's chi-square test is often used to assess departure from the null hypothesis that case and controls have the same the distribution of genotype counts. Cold Spring Harbor Protoc; 2012; doi:10.1101/pdb.top068163

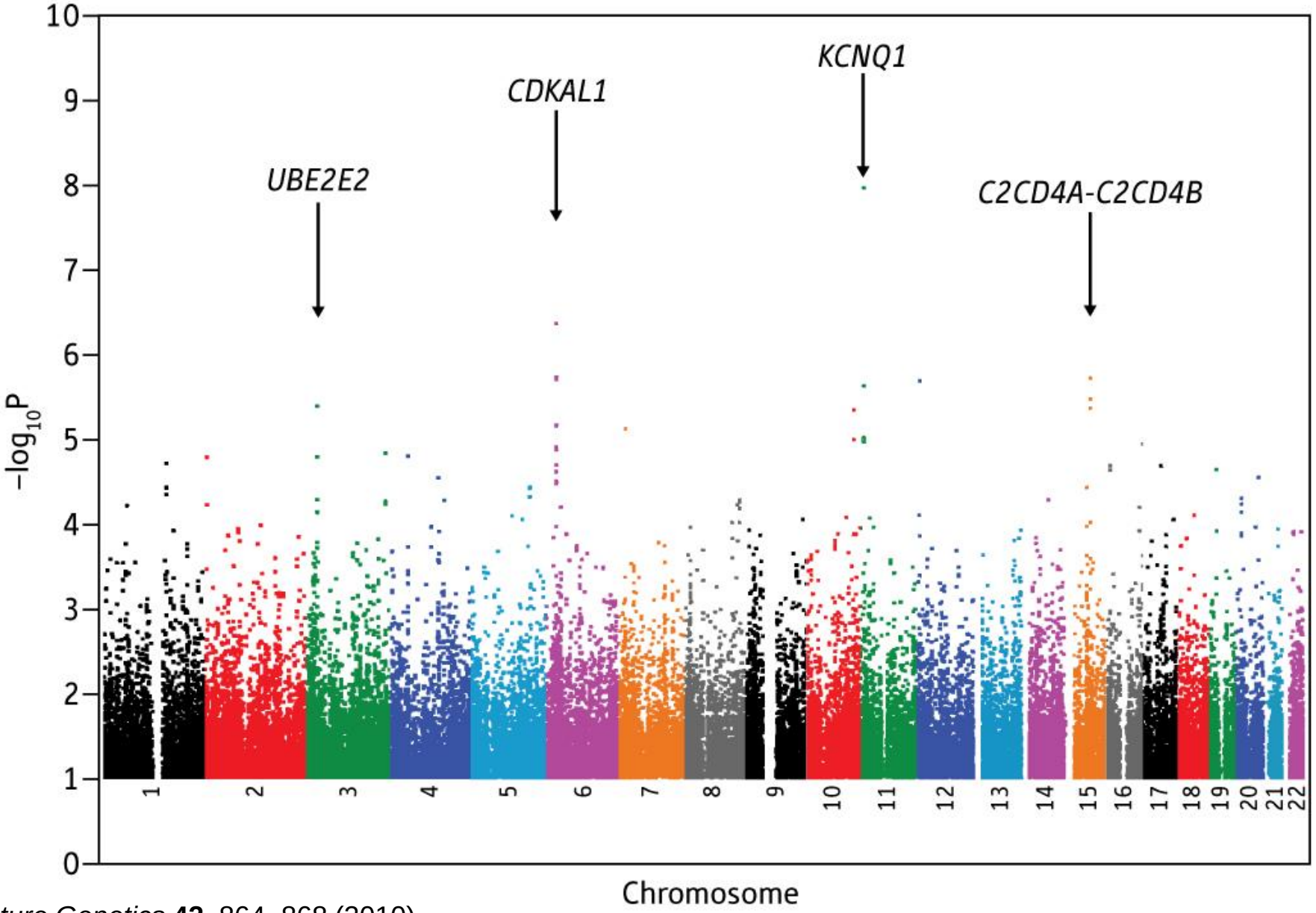
Multiple hypothesis testing

- Millions of independent hypotheses tested
- Type I error: incorrect rejection of null hypothesis
  - Without true effect, 5% of results will be significant at $p = 0.05$ level
- Correction factors that minimize Type I errors
  - Bonferroni correction: divide *p* by the # of tests
  - More permissive correction methods exist

Manhattan Plots

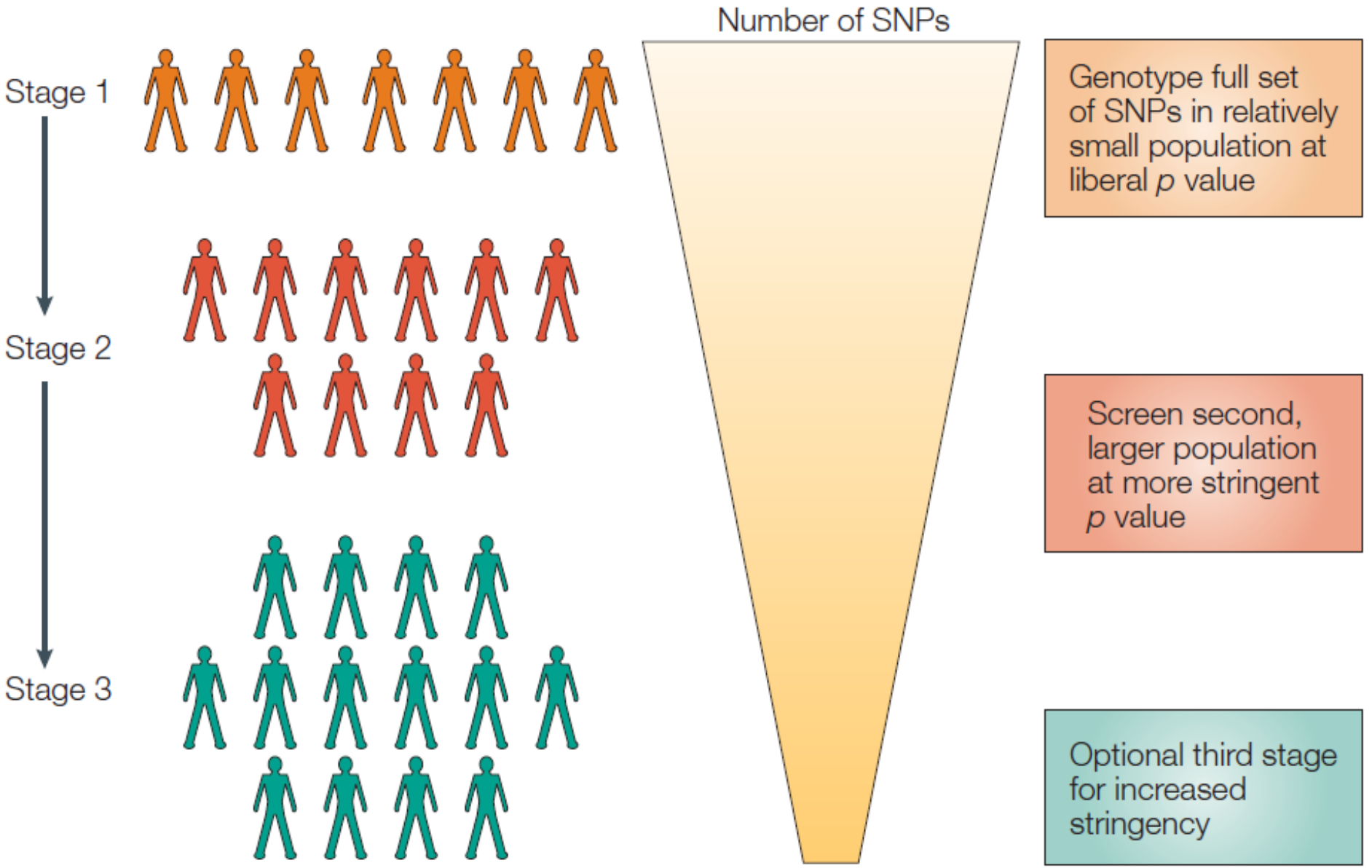
commons.wikimedia.org/wiki/File:Manhattan_Skyline.jpg

GWAS for T2DM Risk Alleles in the Japanese Population



Nature Genetics **42,** 864–868 (2010)

Validation and Study Replication



Nat Rev Genet. 6:95-108 (2005)

Magnitude of the Effect

Disease Odds Ratio (OR)

$$\frac{\dfrac{\text{carriers with disease}}{\text{carriers without disease}}}{\dfrac{\text{noncarriers with disease}}{\text{noncarriers without disease}}}$$

Disease Odds Ratio (OR)

$$\frac{(\text{carriers with disease})\,(\text{noncarriers without disease})}{(\text{carriers without disease})\,(\text{noncarriers with disease})}$$

Commonly used when summarizing the results of GWAS

Example of Disease Odds Ratio Calculation

Allele	Cases	Controls	Total
Present	2100	750	2850
Absent	1250	1400	2650
Total	3350	2150	5500

Disease OR for allele

$$\frac{\text{carriers with disease}}{\text{carriers without disease}} \Big/ \frac{\text{noncarriers with disease}}{\text{noncarriers without disease}}$$

Example of Disease Odds Ratio Calculation

Allele	Cases	Controls	Total
Present	2100	750	2850
Absent	1250	1400	2650
Total	3350	2150	5500

Disease OR for allele

$$\frac{\textbf{2100}}{\text{carriers without disease}} \Big/ \frac{\text{noncarriers with disease}}{\text{noncarriers without disease}}$$

Example of Disease Odds Ratio Calculation

Allele	Cases	Controls	Total
Present	2100	750	2850
Absent	1250	1400	2650
Total	3350	2150	5500

Disease OR for allele

$$\left(\frac{\textbf{2100}}{\textbf{750}}\right) \Big/ \left(\frac{\text{noncarriers with disease}}{\text{noncarriers without disease}}\right)$$

Example of Disease Odds Ratio Calculation

Allele	Cases	Controls	Total
Present	2100	750	2850
Absent	1250	1400	2650
Total	3350	2150	5500

Disease OR for allele

$$\frac{\text{2100}}{\text{750}} \Big/ \frac{\text{1250}}{\text{noncarriers without disease}}$$

Example of Disease Odds Ratio Calculation

Allele	Cases	Controls	Total
Present	2100	750	2850
Absent	1250	1400	2650
Total	3350	2150	5500

Disease OR for allele

$$\left(\frac{2100}{750}\right) \Big/ \left(\frac{1250}{1400}\right)$$

Example of Disease Odds Ratio Calculation

Allele	Cases	Controls	Total
Present	2100	750	2850
Absent	1250	1400	2650
Total	3350	2150	5500

Disease OR for allele: 3.14

Early GWAS Success Story

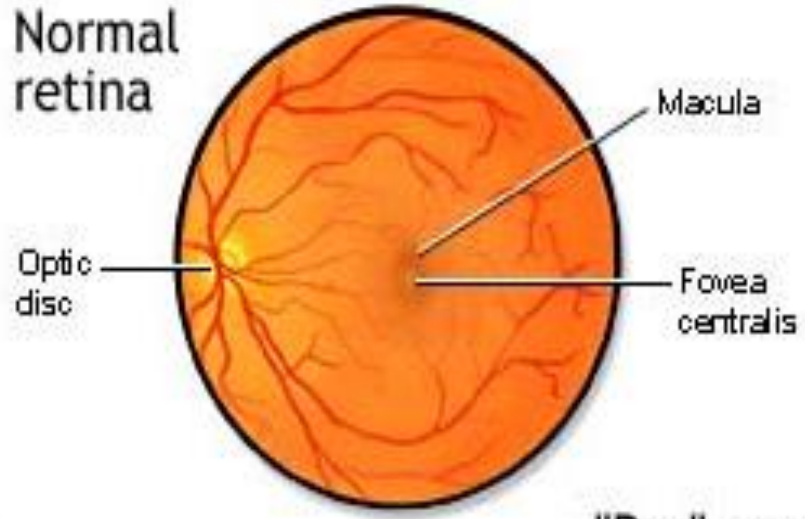
Age-Related Macular Degeneration

Age-Related Macular Degeneration (AMD)

- Leading cause of vision loss for >55 year olds in US
  - >10 million affected individuals in the US
  - ~15% disease risk: ages 70-79
  - ~30% disease risk: age >80

- Affects the macula, a region near the center of the retina where visual perception is most acute

http://www.nei.nih.gov/health/maculardegen/armd_facts.asp



Normal Vision

"Wet" macular degeneration

Abnormal blood vessels grow under the macula (10 - 15% of cases)

"Dry" macular degeneration

drusen

Light sensitive cells die (85 - 90% of cases)

Aggregate of lipids, glyconjugates, proteins, and complement proteins

AMD

http://www.nei.nih.gov/health/maculardegen/armd_facts.asp

Age-Related Macular Degeneration (AMD)

- ~70% of risk can be inferred based on SNP genotypes and environmental factors (smoking)
- *CFH* Y402H, *ARMS2* A69S, *C3* R102G alleles are associated with increased risk of developing AMD
- *CFB* R32Q is a protective allele
- Genes highlight the role of inflammation in disease

http://www.nei.nih.gov/health/maculardegen/armd_facts.asp

What usually is found …

Type II Diabetes

Diabetes in the United States

- Syndrome of impaired carbohydrate, fat, and protein metabolism

- 25.8 million children and adults (8.3% of US population)
  - 18.8 million and 7.0 million cases, respectively
  - 79 million prediabetic cases

- Leading cause of kidney failure, non-traumatic lower-limb amputations, blindness, heart attack, and stroke among US adults

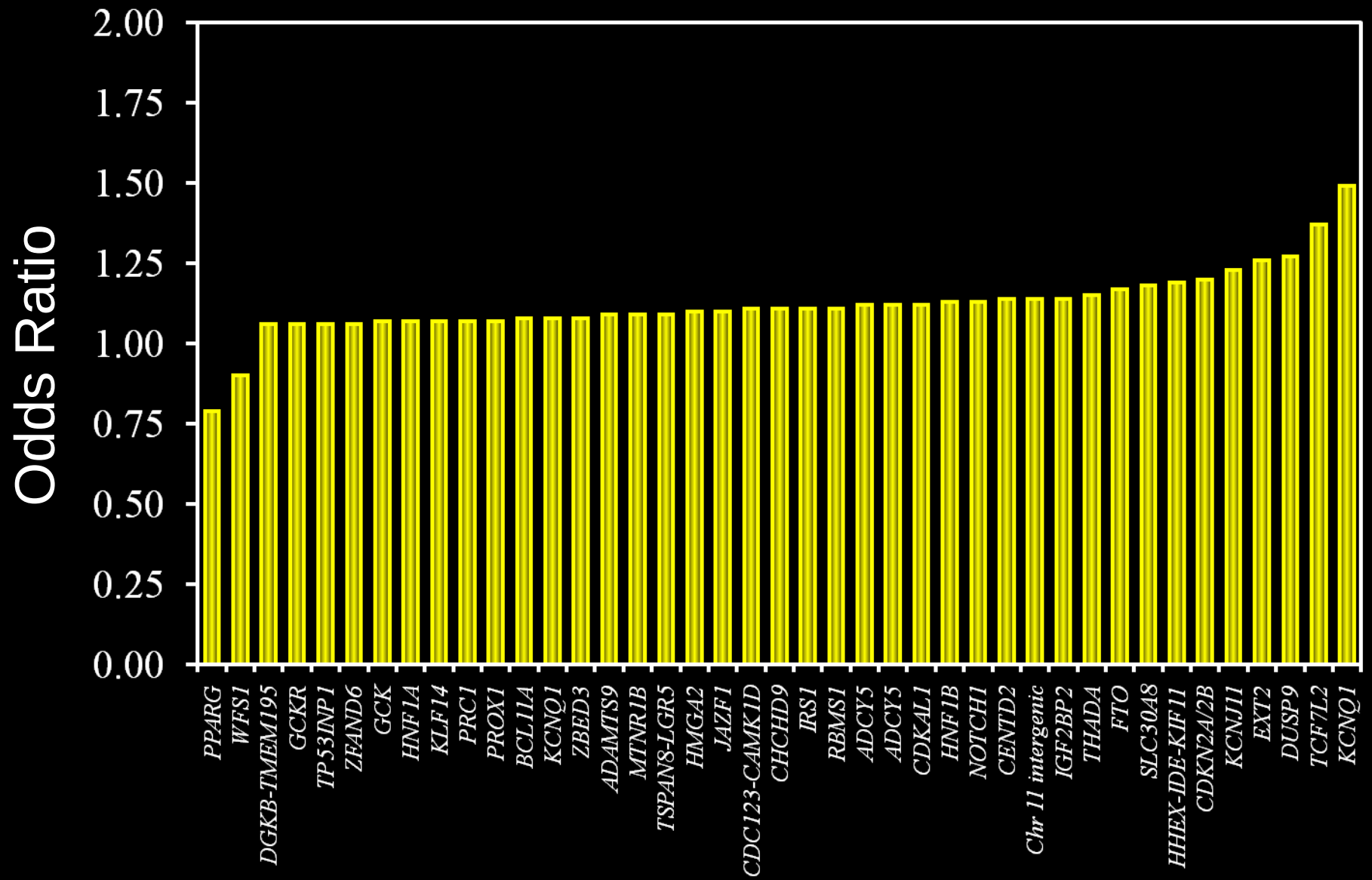
http://www.cdc.gov/diabetes/pubs/pdf/ndfs_2011.pdf

Type 2 Diabetes Mellitus (T2DM)

- Most common type of diabetes (>80% of cases)
- Decreased sensitivity of target tissues to insulin
  - Pancreatic hormone that plays a critical role in energy metabolism
- Complex risk factors
  - increased age, obesity, family history, impaired glucose metabolism, physical inactivity, and ethnic origin

http://www.nei.nih.gov/health/maculardegen/armd_facts.asp

Effect Sizes of T2DM Susceptibility Loci



from Dr. Richard Watanabe (USC)

Results of T2DM GWAS

- To date, only 10% of calculated genetic risks identified
  - Studies in different populations are still ongoing
- Predictive power using genes alone vs. clinical parameters (age, body mass index, etc.) are about the same
- Loci highlight disease mechanisms
  - Abnormal insulin processing and secretion

Lessons from GWAS

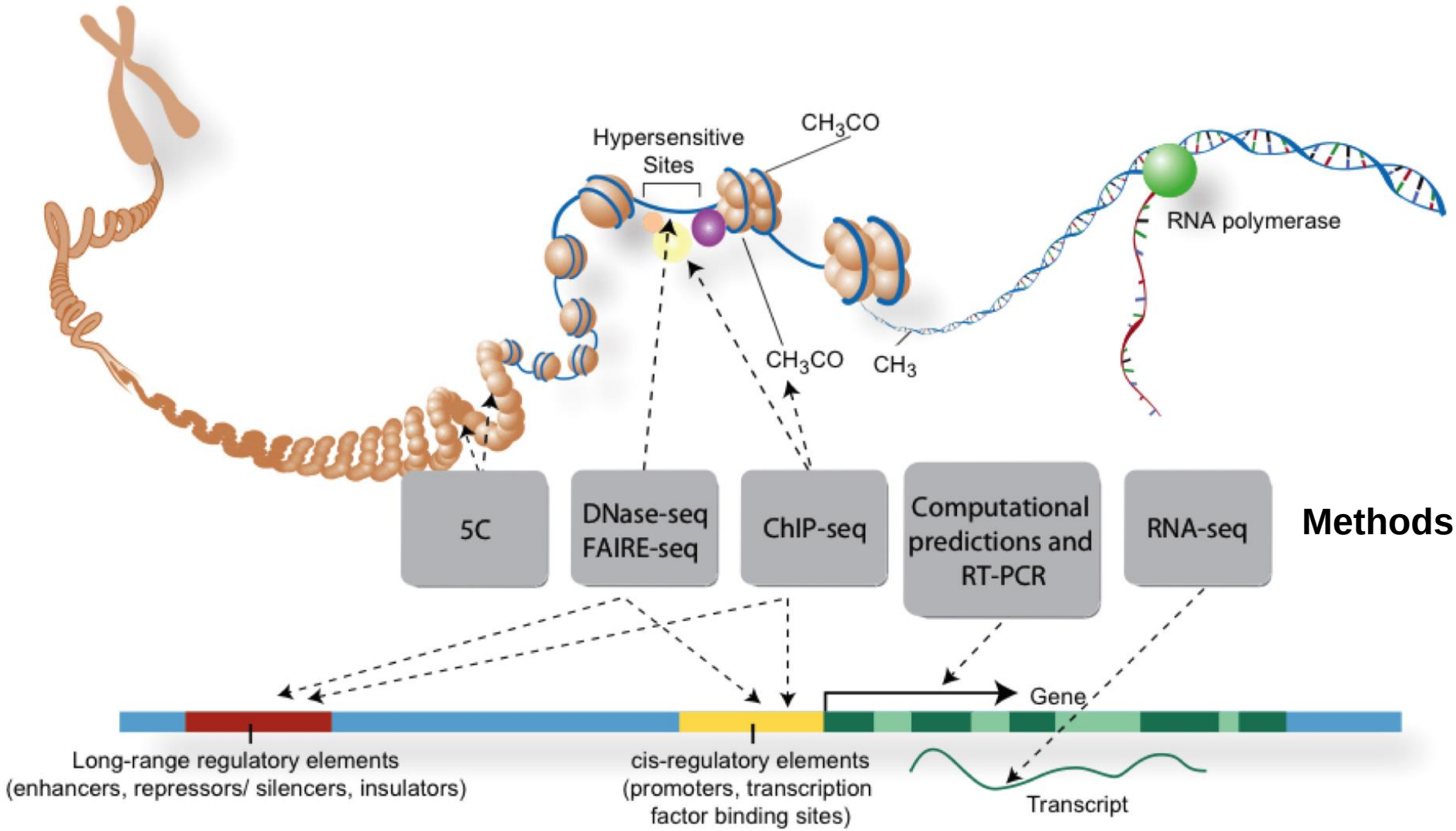
- Most common variants have modest effects on risk
  – Less than 1.5-fold odds ratio (OR)

- For most common diseases/traits: identified SNPs only account for <5-10% of overall risk
  – Not useful clinically

- Useful for understanding pathophysiology
  – New biologic pathways and drug targets

- >80% of GWAS variants are noncoding

Annotating the human genome



http://genome-mirror.duhs.duke.edu/images/EncodeNatureGraphic.png

The case of the missing heritability

Nature 456: 18-21 (2008)

The usual suspects

- Rare variants
- *De novo* mutations: germline and/or somatic
- Copy number variants (CNVs)
- Gene × Gene interactions (epistasis)
- Gene × Environment interactions
- Mitochondrial contribution
- Epigenetic influence
- Combinations of all and/or a subset of the above

Rare variants circa late 2012

ARTICLE

doi:10.1038/nature11632

An integrated map of genetic variation from 1,092 human genomes

The 1000 Genomes Project Consortium*

Nature 491: 56 – 65 (November, 2012)

De novo mutations in human genetic disease

Joris A. Veltman and Han G. Brunner

NATURE REVIEWS | GENETICS **13**: 565-575 (2012)

Rate of *de novo* mutations and the importance of father's age to disease risk

Augustine Kong[1], Michael L. Frigge[1], Gisli Masson[1], Soren Besenbacher[1,2], Patrick Sulem[1], Gisli Magnusson[1], Sigurjon A. Gudjonsson[1], Asgeir Sigurdsson[1], Aslaug Jonasdottir[1], Adalbjorg Jonasdottir[1], Wendy S. W. Wong[3], Gunnar Sigurdsson[1], G. Bragi Walters[1], Stacy Steinberg[1], Hannes Helgason[1], Gudmar Thorleifsson[1], Daniel F. Gudbjartsson[1], Agnar Helgason[1,4], Olafur Th. Magnusson[1], Unnur Thorsteinsdottir[1,5] & Kari Stefansson[1,5]

Nature **488**: 471-475 (2012)

De Novo Mutations

- Sequence variant present for the first time in a family member as a result of a mutation in a germ cell of one of the parents or in the fertilized egg itself
- ~70 *de novo* single nucleotide variants per diploid genome
- ~3 *de novo* insertion or deletion (1-50 bp) per diploid genome
- ~1 *de novo* mutation per exome

Nature Reviews Genetics 13, 565 - 575 (2012)

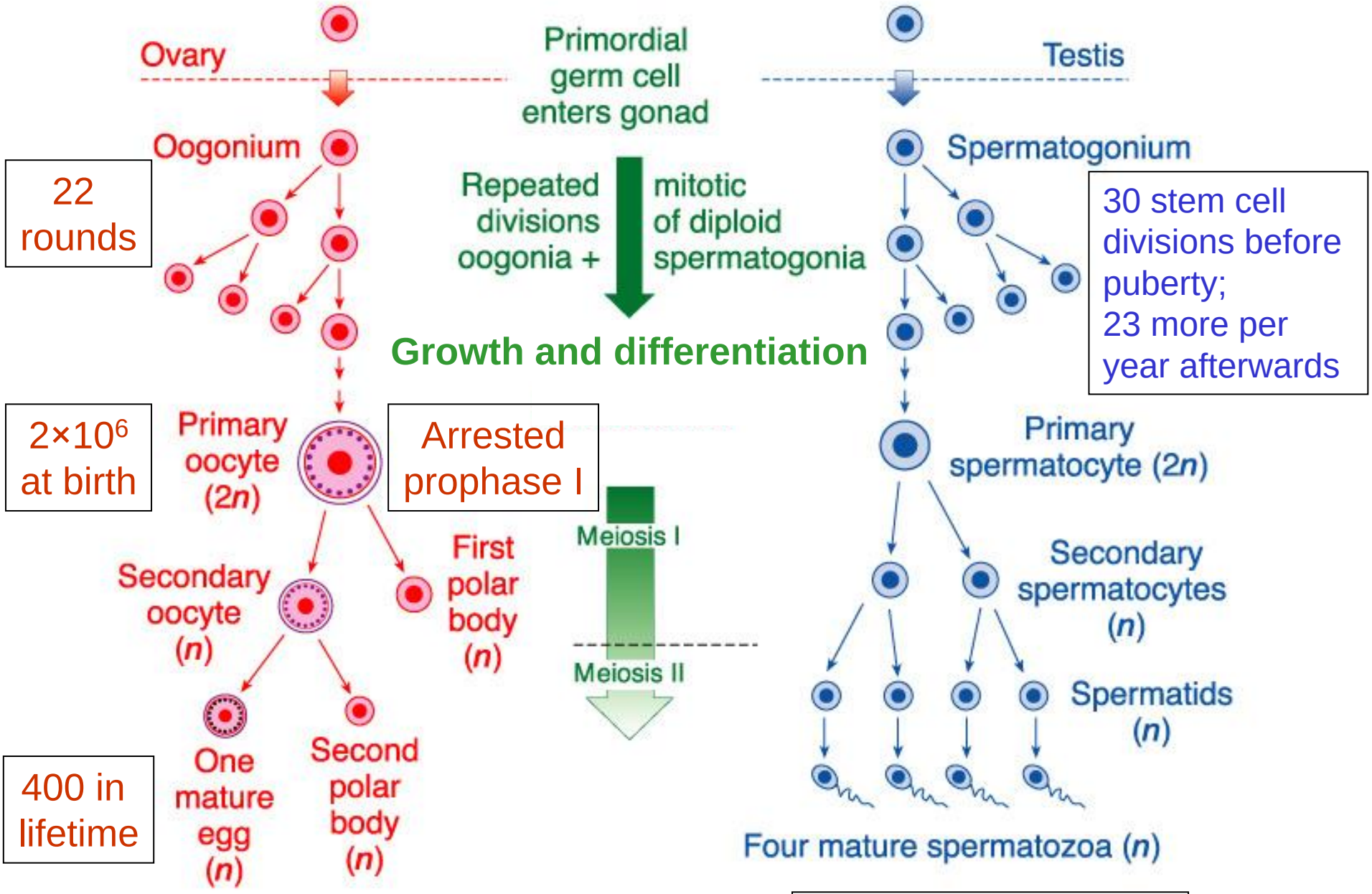


Figure 2-9 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Parent of Origin Effects

- 78 Icelandic parent-offspring trios were sequenced
- Average of 55.4 paternal and 14.2 maternal mutations
- Variation in *de novo* mutation rates driven by paternal contribution
- Number of *de novo* mutations increases with paternal age
  - Two additional *de novo* mutations per year
  - Paternal mutations double about every 16.5 years
- Factors other than father's age did not contribute substantially to the mutation rate diversity in this study

Nature 488, 471 - 475 (2012)

Autism Spectrum Disorders (ASD)

- Affects 1/100 - 1/150 children in the United States
- Shared features include: impaired social relationships, impaired language & communication, repetitive behaviors, narrow range of interests
- Approximately 75% have lifelong disability requiring social and educational support
- Genetic cause is unknown in >70% of cases

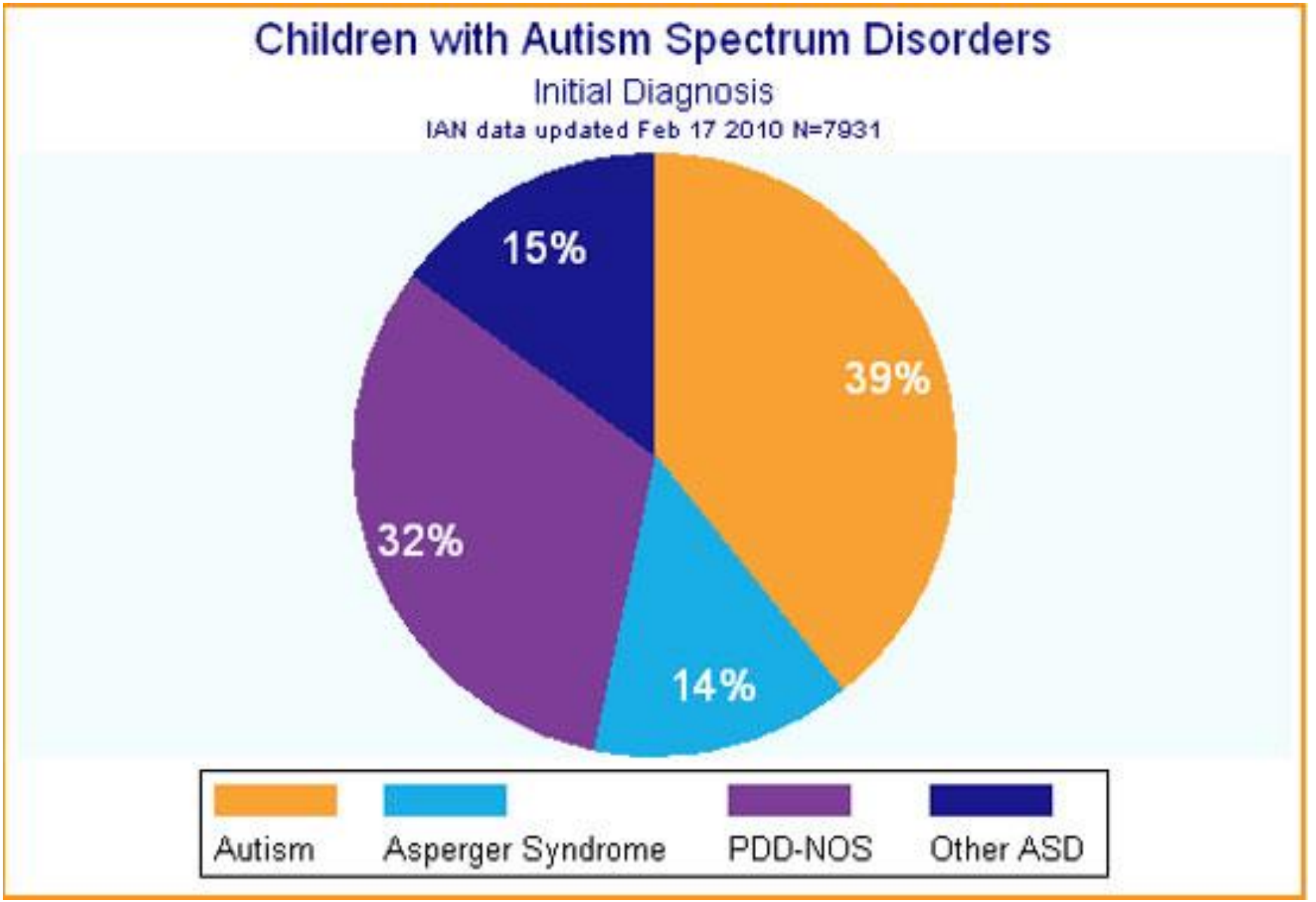
Identified Prevalence of Autism Spectrum Disorders

ADDM Network 2000-2008

Combining Data from All Sites

Surveillance Year	Birth Year	Number of ADDM Sites Reporting	Prevalence per 1,000 Children (Range)	This is about 1 in X children...
2000	1992	6	6.7 (4.5-9.9)	1 in 150
2002	1994	14	6.6 (3.3-10.6)	1 in 150
2004	1996	8	8.0 (4.6-9.8)	1 in 125
2006	1998	11	9.0 (4.2-12.1)	1 in 110
2008	2000	14	11.3 (4.8-21.2)	1 in 88

http://www.cdc.gov/features/countingautism/



www.mybraintest.org/wp-content/uploads/2012/07/autism-spectrum-disorders.jpg

Rett Syndrome (X-linked dominant)

- Caused by *MECP2* (methylCpG-binding protein 2) mutations
- Prevalence of ~1/12,000 female births
- Males with severe (null) *MECP2* mutations are not live born
- Progressive neurodevelopment disease, onset at 6-18 months

Human Molecular Genetics, 2012, Vol. 21, No. 21 **4781–4792**
doi:10.1093/hmg/dds301
Advance Access published on July 26, 2012

Individual common variants exert weak effects on the risk for autism spectrum disorders

"Despite genotyping over a million SNPs covering the genome, no single SNP shows significant association with ASD or selected phenotypes at a genome-wide level."

" … it is reasonable to conclude that common variants affect the risk for ASD but their individual effects are modest."

LETTER

Nature 485: 246-252 (2012)

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]

Multiplex Targeted Sequencing Identifies Recurrently Mutated Genes in Autism Spectrum Disorders

Brian J. O'Roak,[1] Laura Vives,[1] Wenqing Fu,[1] Jarrett D. Egertson,[1] Ian B. Stanaway,[1] Ian G. Phelps,[2,3] Gemma Carvill,[2,3] Akash Kumar,[1] Choli Lee,[1] Katy Ankenman,[4] Jeff Munson,[4] Joseph B. Hiatt,[1] Emily H. Turner,[1] Roie Levy,[1] Diana R. O'Day,[2] Niklas Krumm,[1] Bradley P. Coe,[1] Beth K. Martin,[1] Elhanan Borenstein,[1,5,6] Deborah A. Nickerson,[1] Heather C. Mefford,[2,3] Dan Doherty,[2,3] Joshua M. Akey,[1] Raphael Bernier,[4] Evan E. Eichler,[1,7]* Jay Shendure[1]*

Science 338:1619-1622 (2012)

Rare variants and *de novo* mutations

- *Nature* paper: 677 individual exomes from 209 families

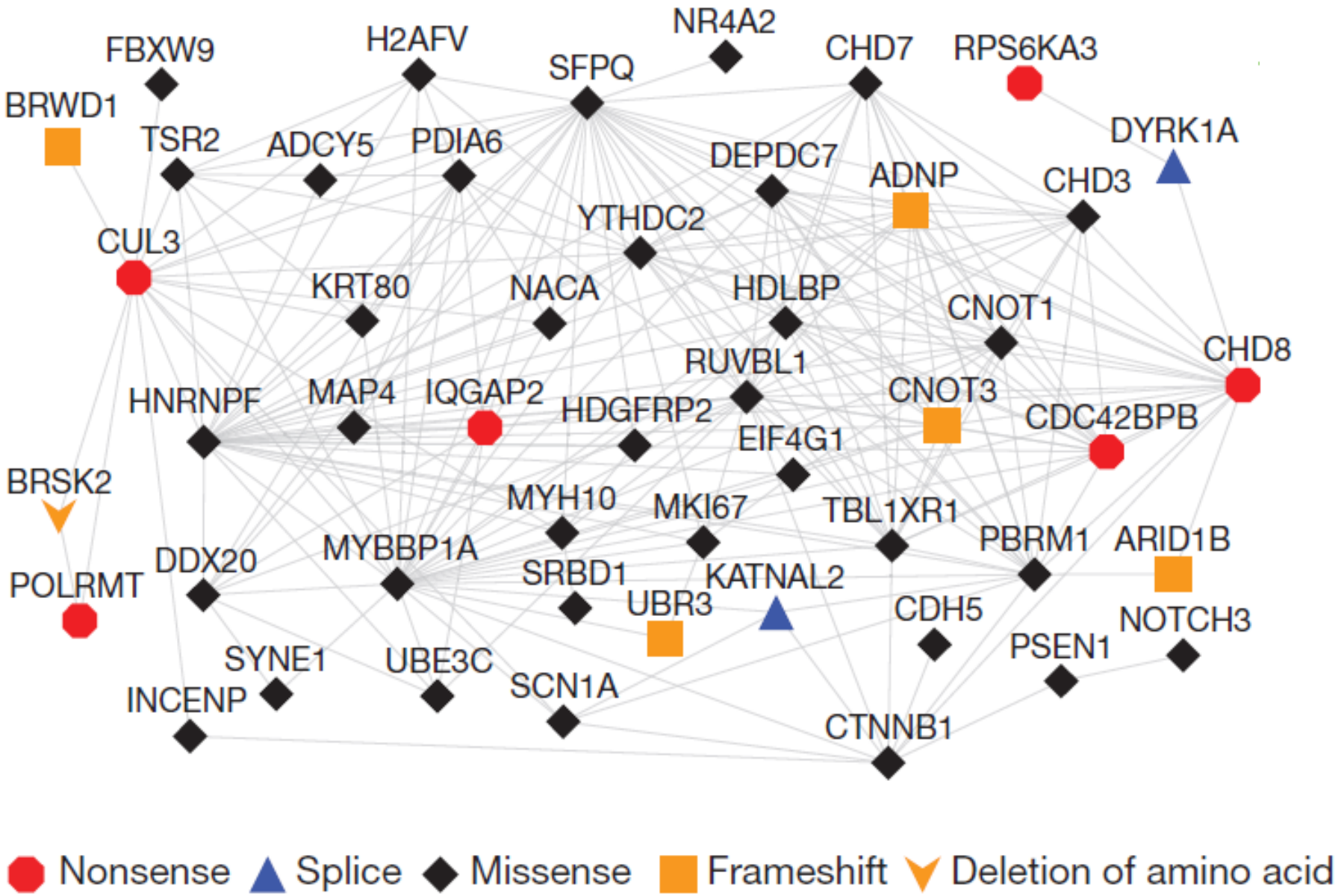
- About 40% of severe/disruptive *de novo* mutations map to β-catenin/chromatin remodeling protein network

- *Science* paper: Followed up on these results by sequencing 44 candidate genes in 2446 patients

- Mutations in 6 genes (*CHD8*, *DYRK1A*, *GRIN2B*, *TBR1*, *PTEN*, &*TBL1XR1*) may contribute to 1% of sporadic ASDs

- Consistent with oligogenic model where *de novo* mutations and rare variants contribute to genetic risk

Gene Network Analysis

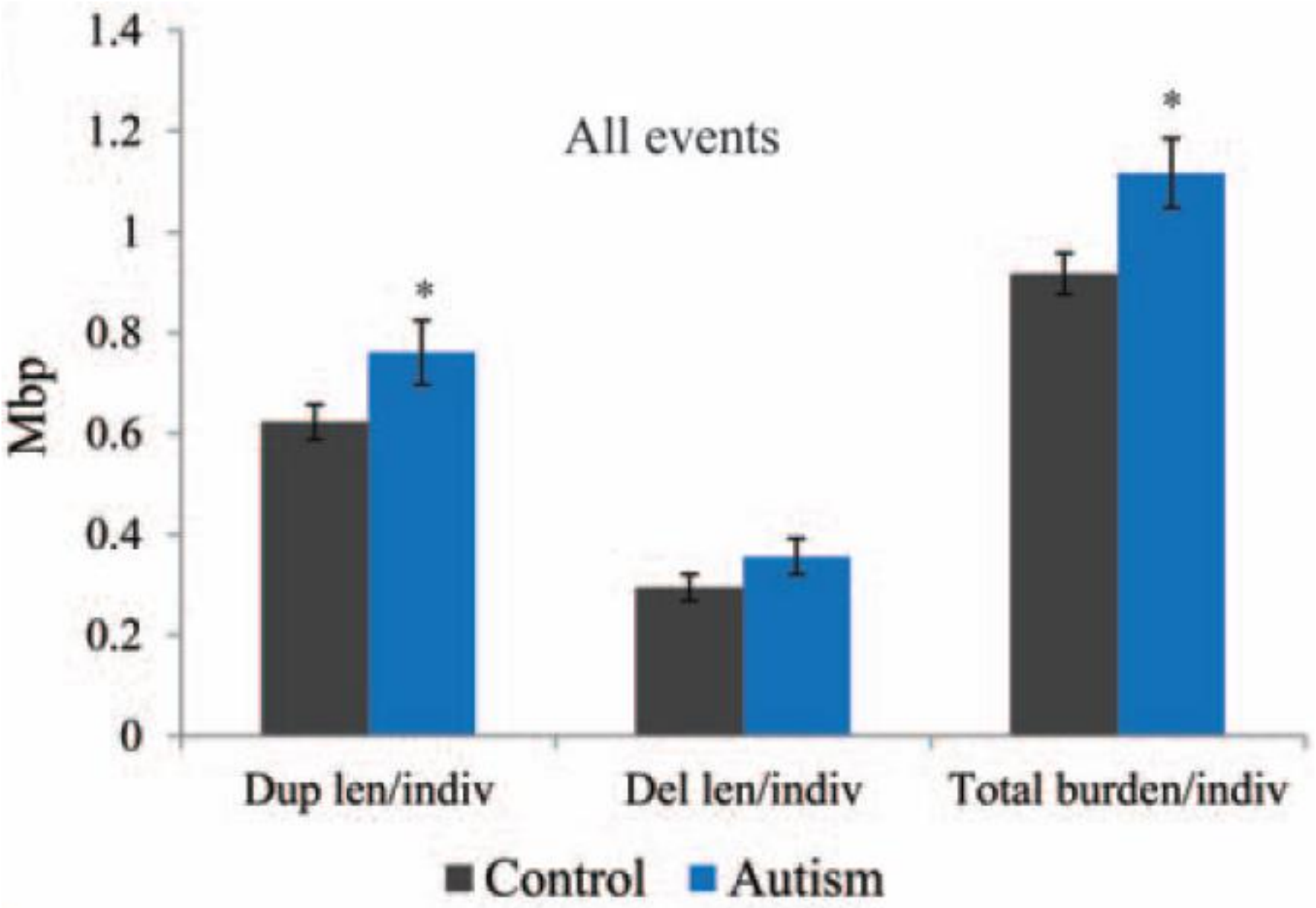


Global increases in both common and rare copy number load associated with autism

Santhosh Girirajan[1,2,3,*], Rebecca L. Johnson[4,5], Flora Tassone[6,8], Jorune Balciuniene[4,5,10], Neerja Katiyar[2], Keolu Fox[1], Carl Baker[1], Abhinaya Srikanth[2], Kian Hui Yeoh[2], Su Jen Khoo[2], Therese B. Nauth[4,5], Robin Hansen[6,7], Marylyn Ritchie[2], Irva Hertz-Picciotto[6], Evan E. Eichler[1], Isaac N. Pessah[6,9] and Scott B. Selleck[2,4,5,*]

Human Molecular Genetics Advanced Access April 19, 2013

- CNV data from 274 cases and 242 controls
- 120 genomic hotspots for >50-kb events
- Entire genome for variants >300-kb using arrayCGH


All events
Mbp
1.4
1.2
1
0.8
0.6
0.4
0.2
0
*
*
Dup len/indiv
Del len/indiv
Total burden/indiv
Control
Autism

Take home messages

- CNV load, predominantly duplications, predisposes to autism

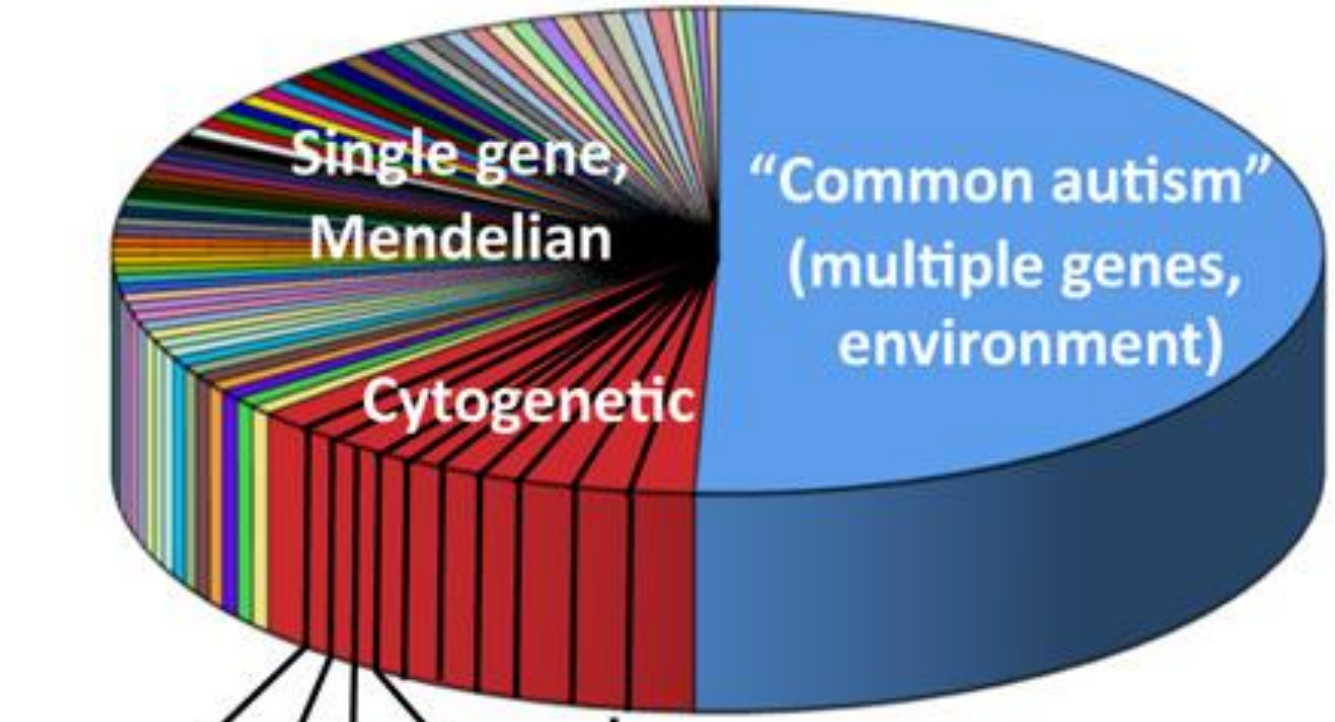
- Genomic regions with a significant role in autism are associated with substantial clinical heterogeneity
  - seizure disorder, schizophrenia, and developmental delay

- Few autism-specific genes, but a collection of genes affecting phenotypes associated with autism

- Autism-associated genetic variants discovered thus far only begin to account for the estimated heritability

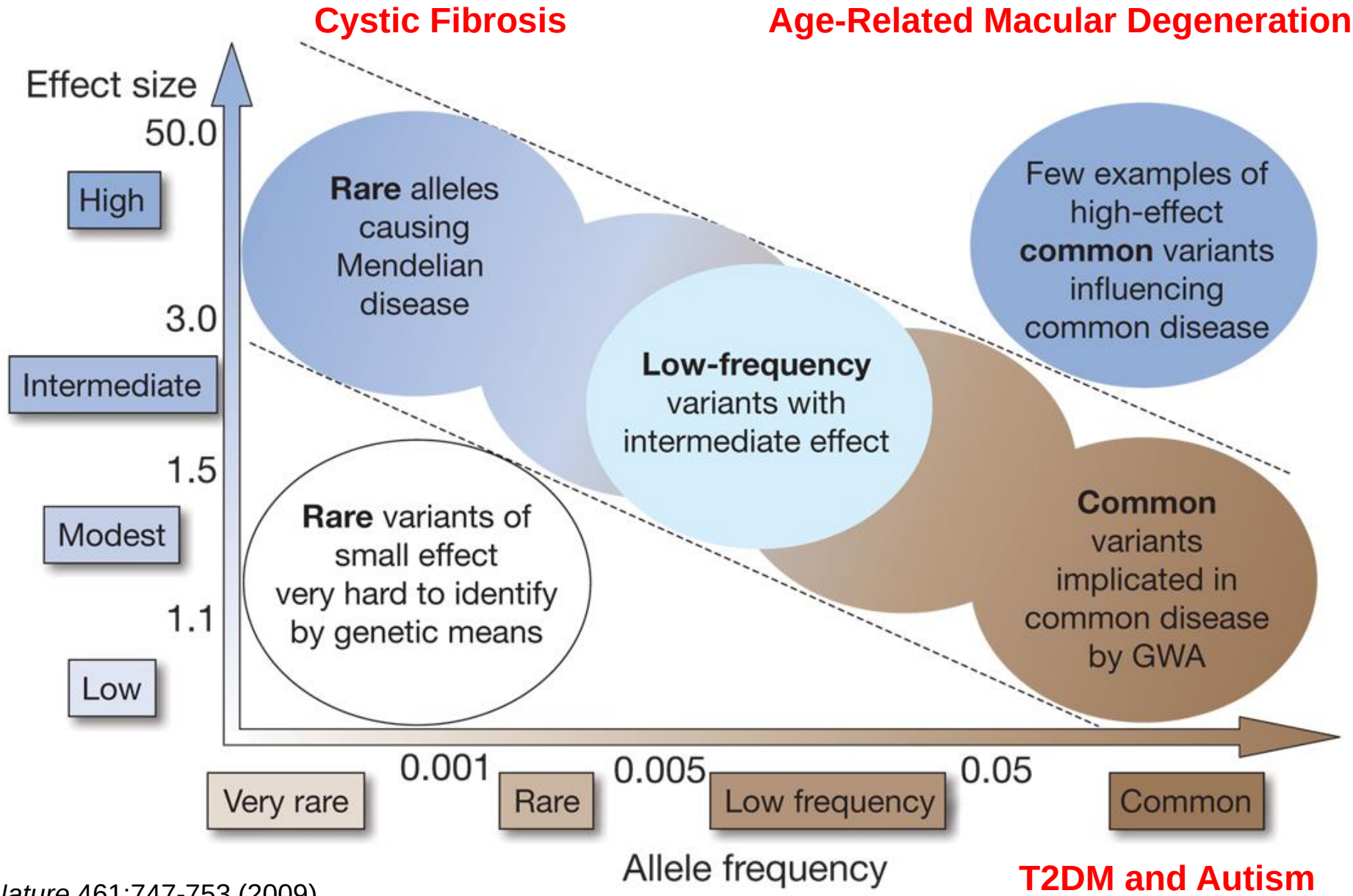
- Anything that increases genomic instability could contribute to the genesis of these disorders

Causes of Autism



genetics.emory.edu/egl/about/index.php/1583

Risk allele frequency and genetic effect sizes



Nature 461:747-753 (2009)