



# Genetics of Stroke

Daniel Woo, M.D., M.S.

University of Cincinnati

# Objectives

- To understand the basic terms and concepts of genetics
- To understand how they have been applied to genetic discovery
- To understand the added potential in research

# What I won't Cover

- Classic Mendelian genetic risks
- ALL of the hits in stroke genetics

# Basics

- DNA → RNA → Protein
  - Each Human cell contains  $\sim 6.6 \times 10^9$  bases (A,C,G or T) or  $\sim 5$  feet of DNA
  - Adult body contains  $\sim 10$  trillion cells
  - (5 feet)(10 trillion) is enough DNA to reach from the Earth to the Sun  $\sim 90$  times
- 1-2% of our DNA codes for proteins

# Of Alleles and Polymorphisms

If everyone in the world has the same sequence at a specific locus, it is **monomorphic**.

If there are different forms, the site is **polymorphic** and the different types are called **alleles**

# Polymorphisms

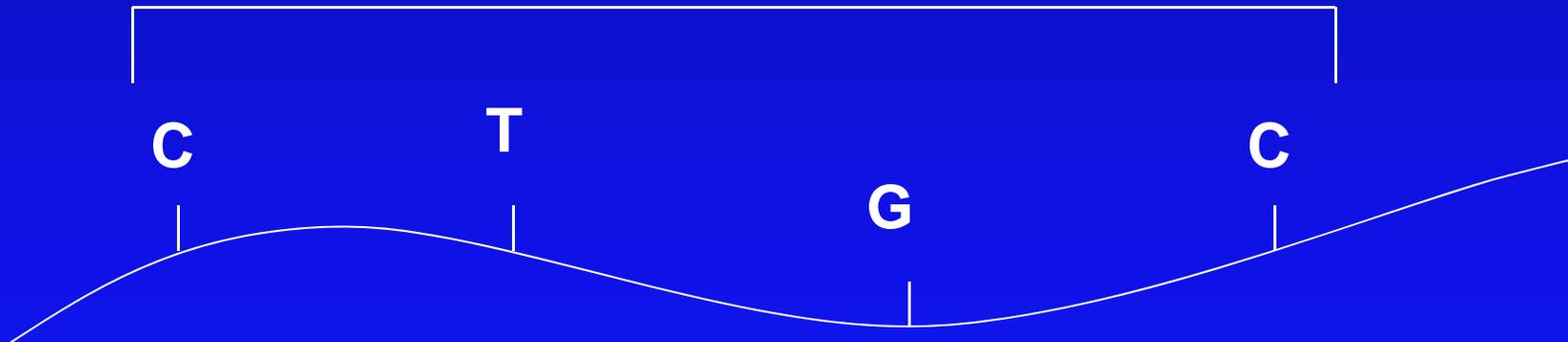
- Any variation of the genetic sequence occurring greater than 1% of the population
- Includes single nucleotide polymorphisms (SNPs), up to multinucleotide repeats like CAG repeats (microsatellites).
- Can be in the exon or intronic regions
- Can be pathologic or non-pathologic



C & T are different **Alleles** of this Single Nucleotide Polymorphism or **SNP**

CT would then be the '**Genotype**' at this **Locus**

## Haplotype

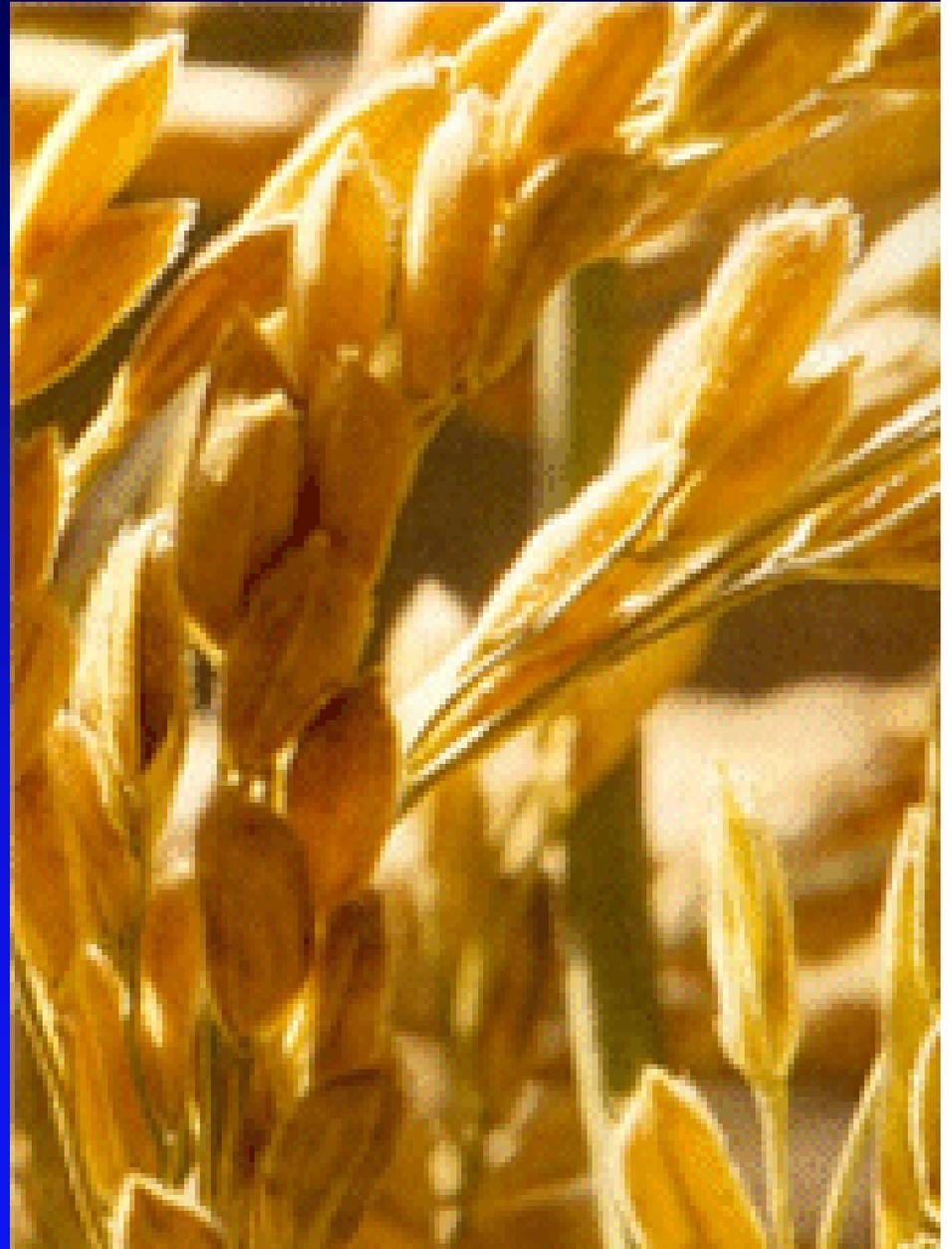


# Number of Genes

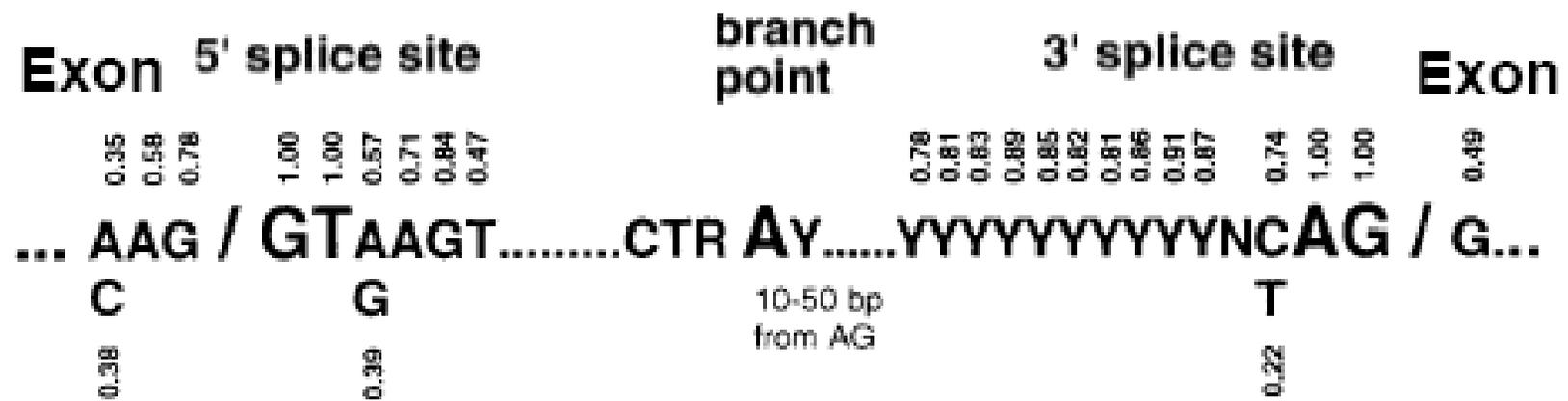
- 13,000 – Fruit Fly
- 19,000 – Roundworm
- 25,000 – Mustard Weed
- 30-40,000 – Humans

# Oryza Sativa

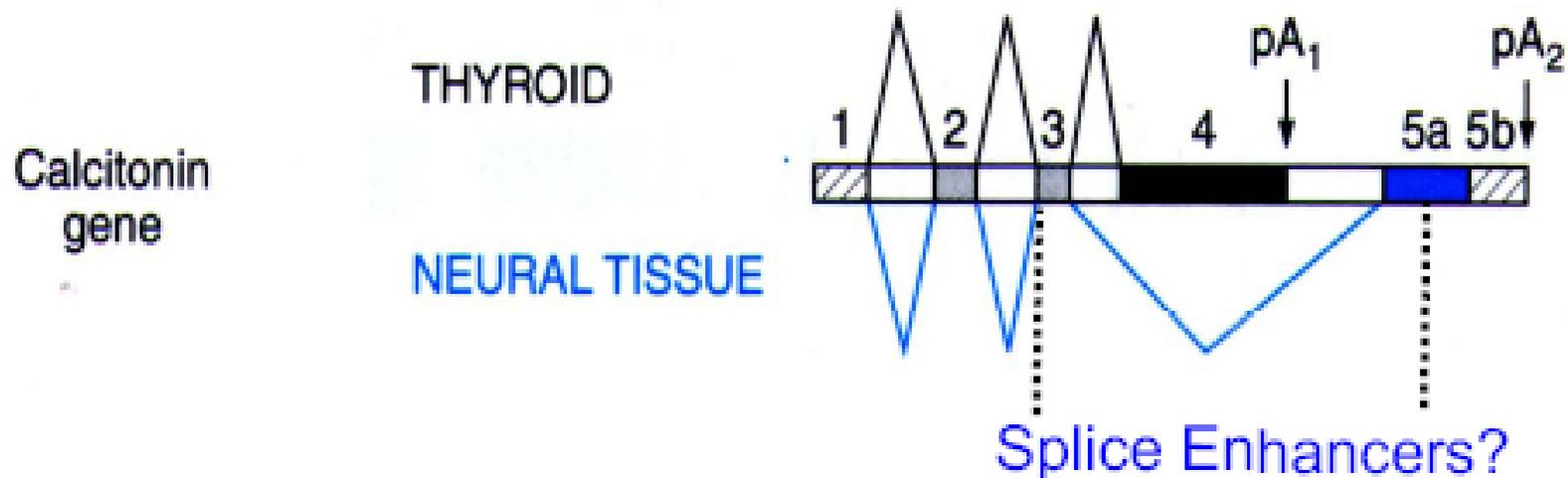
- 56,000 GENES!
- Humans gain complexity through regulation and splicing



# mRNA Splicing: Consensus Sequences



## Alternative Splicing: Splicing Enhancers

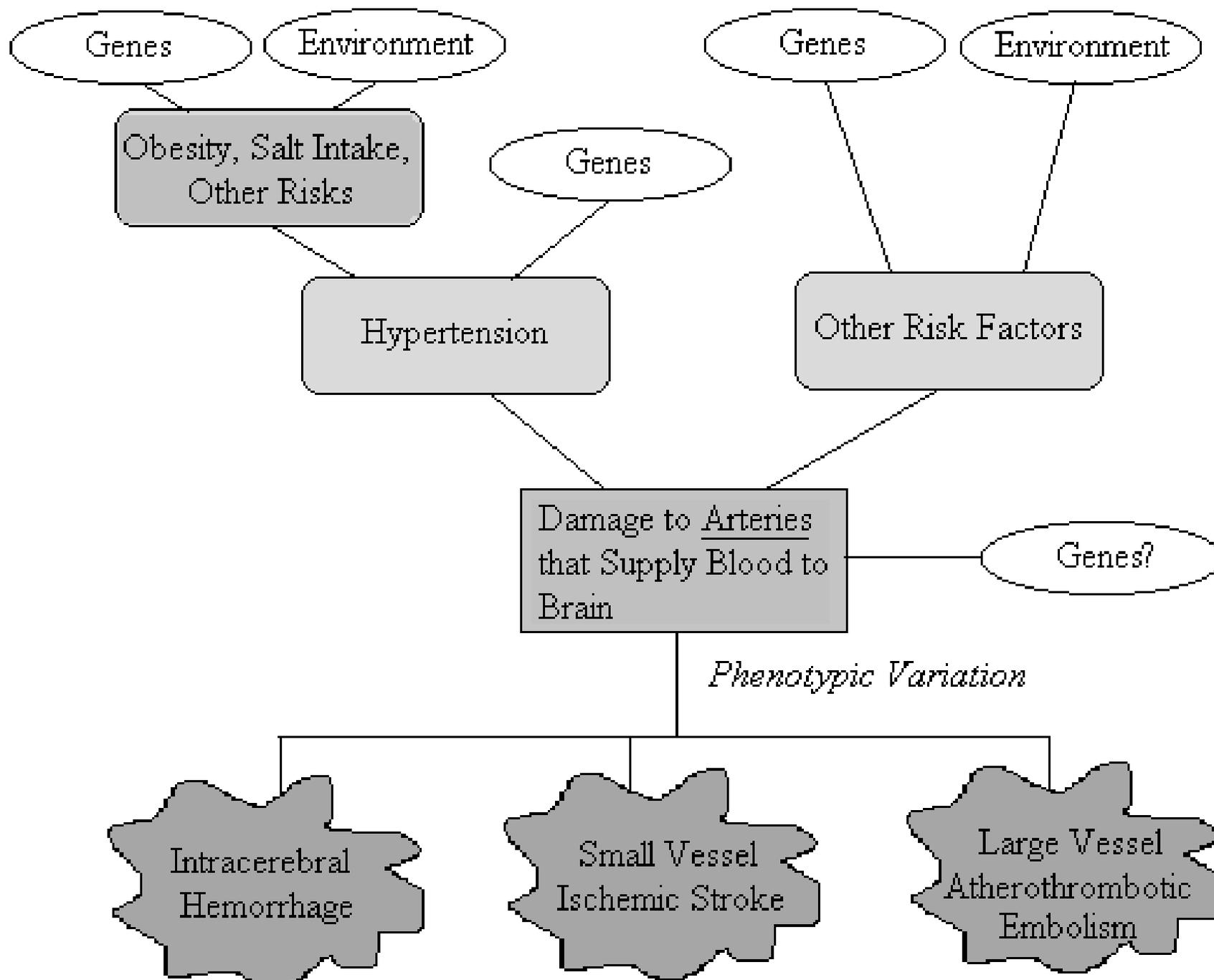


# Simple versus Complex Traits

- Classic Mendelian inheritance (Recessive/dominant, autosomal/sex linked) or one variant one disease
  - Sickle Cell
  - Fabry's
  - MELAS
- Complex –
  - Multiple variants, often with tiny impact
  - Variants that interact with one another (two hit, three hit, twenty hit)
  - Variants that affect expression of a gene
  - Variants that are hits only if an environmental exposure occurs
  - Phenocopies
  - Stroke, Hypertension, Diabetes, Cancer

# Stroke is not just one disease but many

- Each of the subtypes of stroke meet the criteria for a complex genetic disease
- In fact, within subtypes such as ICH there are likely different phenotypes



**Figure 1 - Complex Etiology of Stroke**

# Intracranial Aneurysm

- Pop'n based studies show that around 10% of first-degree relatives have a history of SAH and around 5% of second-degree relatives have history of SAH
- The ratio suggests that this phenotype is highly genetic and there could be 1 major gene with limited penetrance.

# Intracerebral Hemorrhage

- About 6% of patients with ICH have a first-degree relative with ICH and 1% have a second-degree relative with ICH.
- This works out to be modestly genetic and to have between 2 and 3 major genes with very limited penetrance

# Twin Studies

- Brass et al found that monozygotic twins had a 17% concordance rate of stroke and dizygotic twins had 4% rate.
- Bak et al found 10% vs. 5%
- These studies suggests 1 to 2 genes with variable penetrance may be related to stroke

enland

meadows

### World of the Vikings

Numbers refer to separate  
over primary places of  
ing Age Interest.

ow areas represent  
ing Land". Red lines mark  
rtes of the Vikings.



# Iceland deCODE Study

- Reported in 10/02 as having discovered the gene for stroke(!)
- 1500 patients and 3000 close relatives
- Genome scan on 476 pts + 438 relatives
- Multipoint LOD Score of 4.4(!)
- Reconfirmed region as of interest in Scotland (1063 patients vs 729 controls)

# STRK1

- On 5q12 – a 4.1 megabp region
- Enzyme expressed in vascular tissue
- More important for large vessel and cardiogenic
- RR=1.25
- PDE4D – 17 exons
- cAMP mediated arterial smooth muscle proliferation and atherogenic plaques

# PDE4D

- Phosphodiesterase 4D
- Two key processes in atherosclerosis are:
  - Inflammation
  - Intimal Smooth Muscle Proliferation
- Both processes are affected by PDE4D

# Winner's Curse

- Refers to a financial transaction in which the winner of a bidding war often has a bid far in excess of the actual value of the prize
- In genetics research, it refers to the upward bias of the estimated effect of a newly identified risk which is far in excess of what the actual value of the allele is.

# Variation in the PDE4D Gene and Ischemic Stroke Risk

## A Systematic Review and Meta-analysis on 5200 Cases and 6600 Controls

Steve Bevan, BSc, PhD; Martin Dichgans, MD; Andreas Gschwendtner, MD;  
Gregor Kuhlenbäumer, MD; E.B. Ringelstein, MD; Hugh S. Markus, DM, FRCP

Int J Clin Exp Med 2015;8(2):171  
[www.ijcem.com](http://www.ijcem.com) /ISSN:1940-590

### Original Article

### Association between PDE4D SNP 87 and ischemic stroke risk

Wenzhao Liang<sup>1\*</sup>, Dezhi Zhang<sup>2\*</sup>, J  
Zhongxin Xu<sup>1</sup>

<sup>1</sup>Department of Neurology, China-Japan  
Abdomen Ultrasound, The First Hosp  
China-Japan Union hospital of Jilin U

Received October 20, 2014; Accepted

**Abstract:** Background and purpose: [

are contentious and debatable. The present meta-analysis was undertaken to systematically summarize the possible association. Methods: Based on comprehensive search of PubMed, Embase, and CNKI databases, we identified 18 eligible articles examining the relationship between PDE4D SNP 87 and ischemic stroke risk. We evaluated the strength of relationship using odds ratios (ORs) with 95% confidence intervals (CIs). Results: In the overall analysis, PDE4D SNP 87 was not found to have effects on the risk of ischemic stroke. The null association persisted in the subgroup analyses according to ethnicity and sample size. Conclusions: Our meta-analysis suggests that PDE4D SNP 87 may not represent an independent risk factor for ischemic stroke development.

**Keywords:** PDE4D, SNP, ischemic stroke, risk

**Background and Purpose**—PDE4D was identified as the first novel gene associated with ischemic stroke risk. Replication studies have produced conflicting results, but many have been small and underpowered. Meta-analysis provides a method to combine this data and determine in a larger sample size whether the association with PDE4D can be replicated.

**Methods**—A meta-analysis of all PDE4D variants investigated in relation to ischemic stroke has been undertaken. Analysis of any variant appearing in 2 or more replication studies was included; this comprised 6 single nucleotide polymorphisms together with allele 0 of minisatellite AC008818 and the G0 haplotype. A total of 16 studies were identified, allowing examination of up to 5216 cases and 6615 controls for a single variant. Analyses were performed including all data, excluding data from the original report (providing true replication data), and for individual stroke subtypes and limited to white ethnicity.

**Results**—No individual single nucleotide polymorphism was associated with all ischemic stroke cases. Allele 0 of AC008818 and haplotype G0 carriers was associated with increased risk (relative risk, 1.12; 95% CI, 1.01 to 1.25;  $P=0.03$  and relative risk, 1.18; 95% CI, 1.05 to 1.33;  $P=0.007$ ), but these associations became nonsignificant after exclusion of the original study from the analysis (relative risk, 1.06; 95% CI, 0.94 to 1.20;  $P=0.34$  and relative risk, 1.16; 95% CI, 1.00 to 1.34;  $P=0.06$ , respectively). Analyzing only whites, the majority of cases studied, did not result in a significant association for any analysis. Few robust associations were found with individual stroke subtypes.

**Conclusion**—No genetic variant examined in PDE4D showed a robust and reproducible association to ischemic stroke. Any association that may exist is likely to be weak and potentially restricted to specific populations. (*Stroke*. 2008;39:1966-1971.)

**Key Words:** genetics ■ meta-analysis ■ PDE4D

# Replication, replication, replication

- Thus far, the best check against a false positive result has been replication in an independent sample set
- Need to find the same allele in the same direction and of relatively the same effect size
- Technically speaking, a much larger sample size still trumps replication as a check against false positives if the studies do not have an inherent bias.

# Linkage Disequilibrium

- All humans are related if we go back far enough
- Ie. Two 'unrelated' people from the UK would likely share common ancestors not more than 22 generations ago or about 500 years.
- Because of the multiple recombinations, only those alleles VERY close to the disease allele will remain shared.

# What's LD?

- Linkage disequilibrium
- Deck of cards out of the pack, all of the suits are in order.
- If you shuffle it once, some disorder occurs but cards that are near one another are still likely to be near one another and are much more likely than cards that are far away.

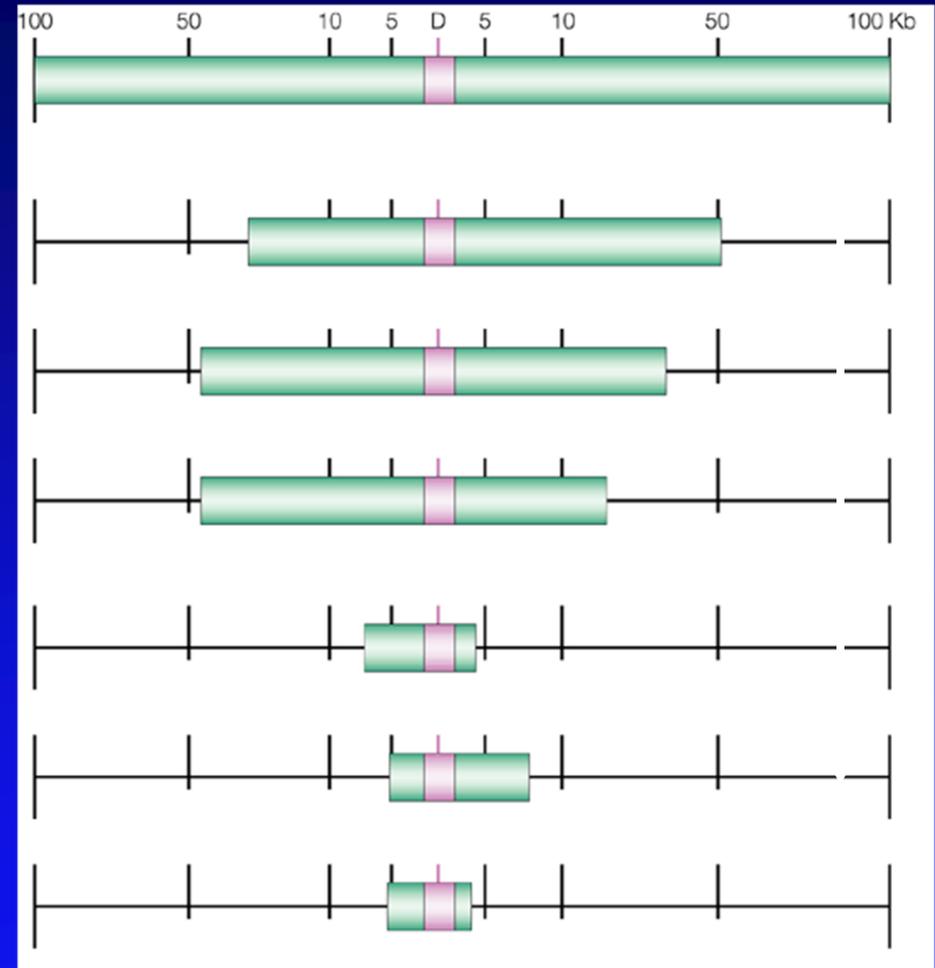


# Let's imagine a really big deck of cards

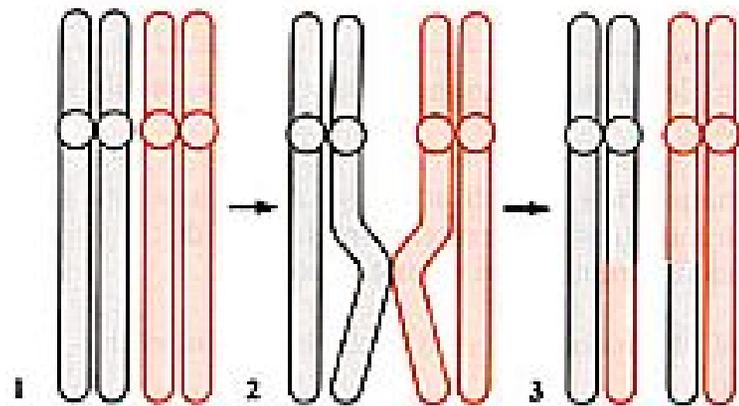
- Say...  $\sim 6.6 \times 10^9$  cards. Each card represents one of four suits (or DNA residues).
- As you shuffle the deck, parts of the DNA that are closer to one another are more likely to stay together than parts of the DNA that are far away from each other.

# LD

- a function of the age of the mutations, distance, number of haplotypes in ancestral population (founder effects)
- look for polymorphisms that are enriched in affected individuals >> general population
- works best in isolated populations--small number of haplotypes (founder effect)



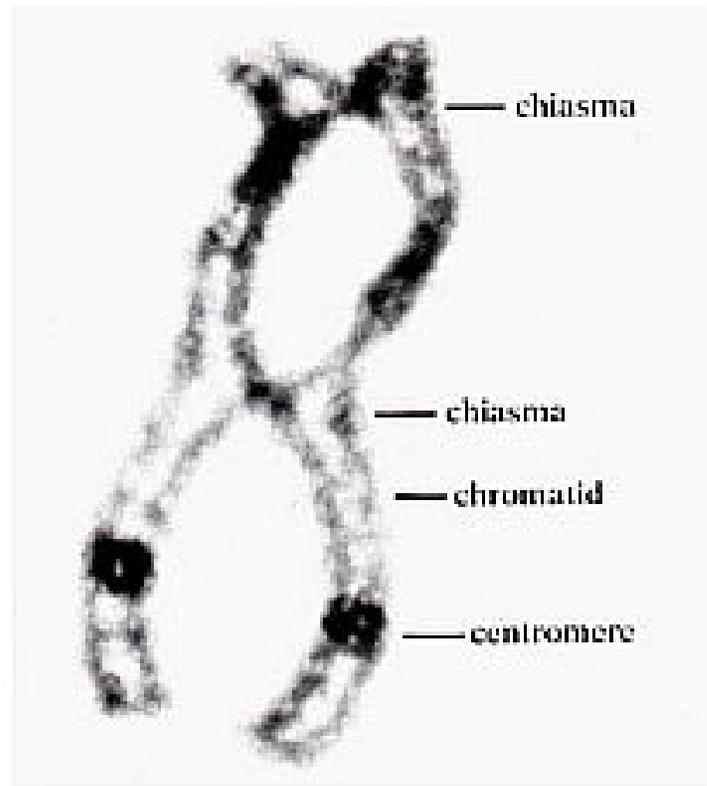
--haplotype blocks (green) will decrease in size over time  
--haplotypes provide a record of the entire recombinational history of the region



## Schematic model of crossing-over

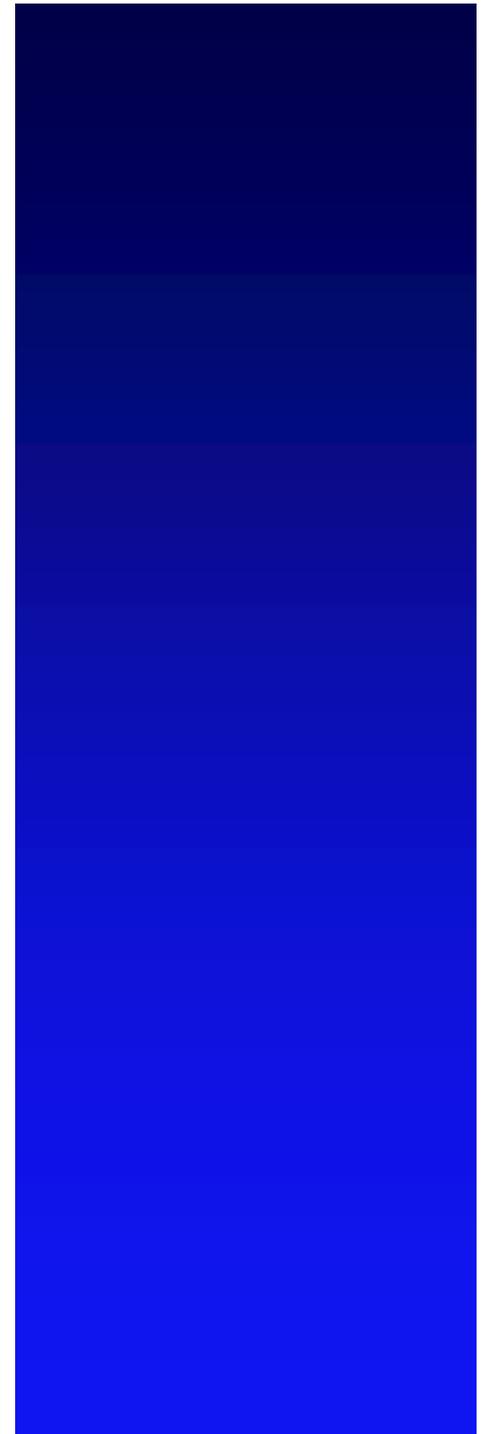
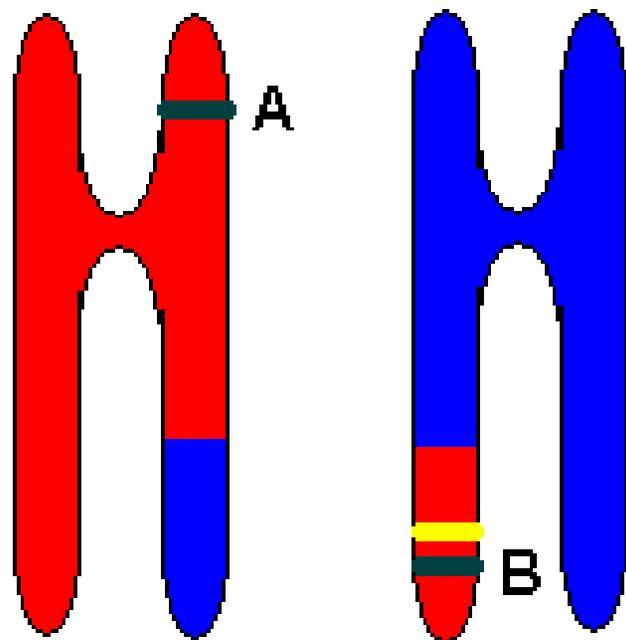
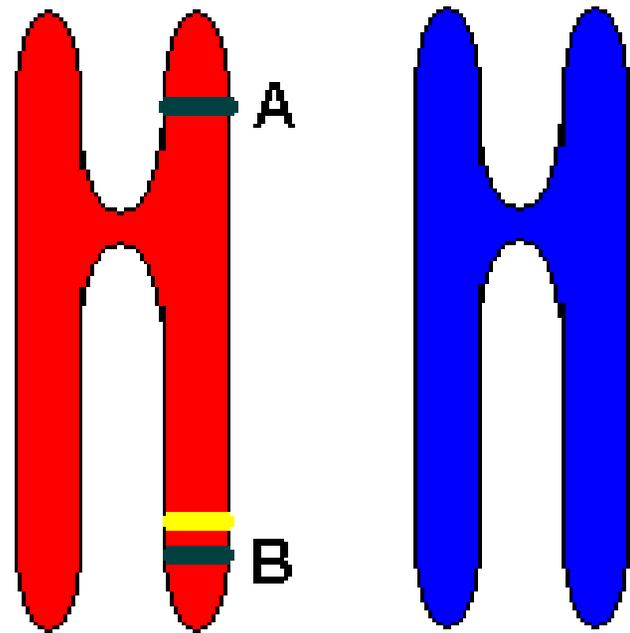
- (1) Two homologous double-stranded chromosomes lie side by side during synapsis.
- (2) Chromatids break at corresponding points and exchange parts.
- (3) The fragments rejoin; each of the four chromatids is now different from the others.

*from* Elements of Biological Science  
Third Edition 1983  
W.W.Norton & Co.  
New York. p. 403



## Photomicrograph showing two chiasmata between homologous chromosomes

Each chromosome is clearly recognizable as double-stranded. The centromeres, too, are clearly visible. Crossing-over is taking place at two points - the two chiasmata. These chromosomes are in late prophase I.

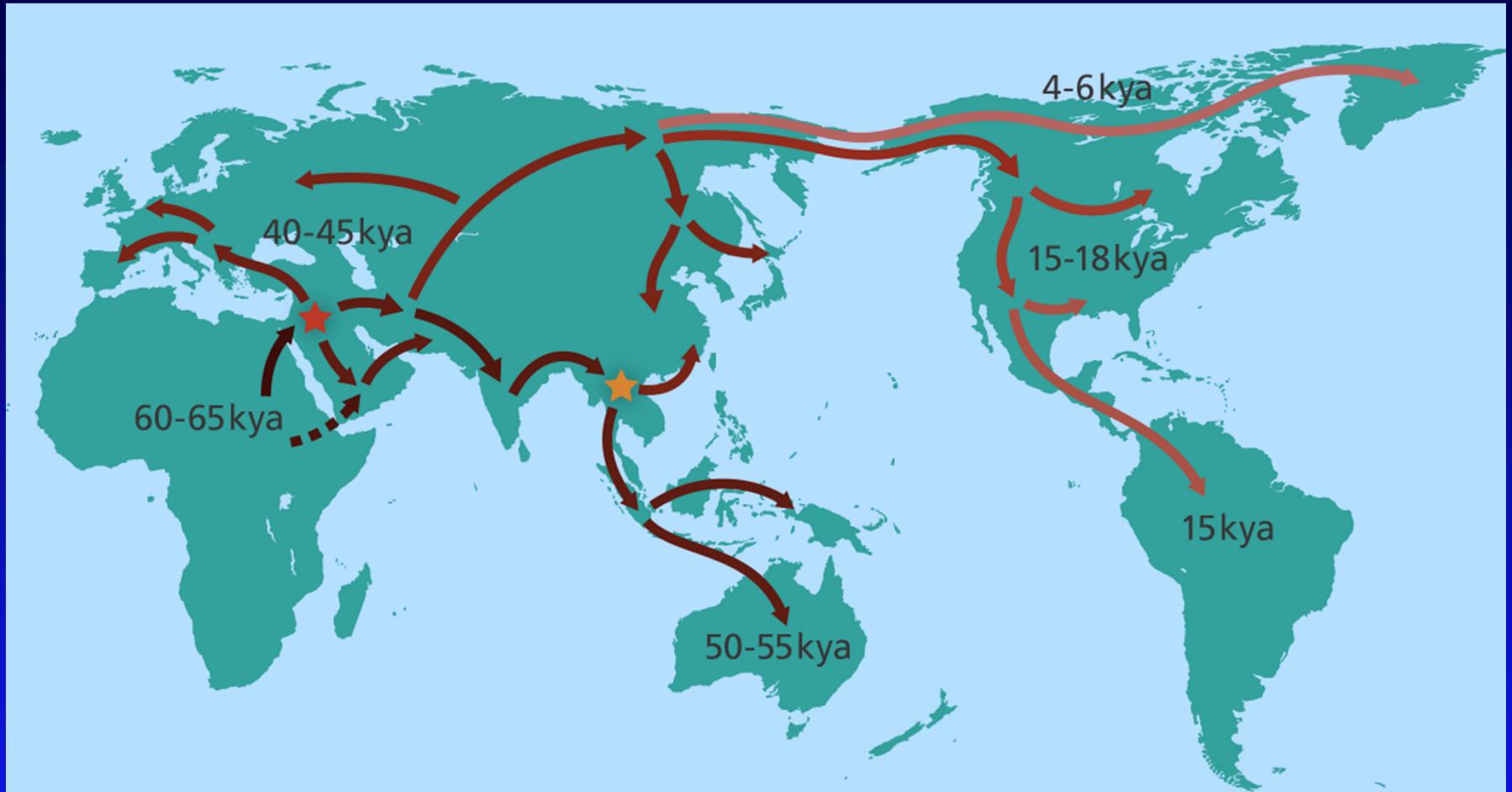


# Genomic Coverage

- Thus, we don't have to get "The Stroke Gene", we only have to get a common marker that is so close to the stroke gene, that it will be inherited with it
- Thus, when we speak of genomic coverage, we mean how much of the DNA are we covering with the markers whose position we know of.

# LD by Race

- From a genetic standpoint, Caucasians or European whites are VERY new. Very few shuffles of the deck
- But those of African Descent have had several tens of thousands of years more generations of shuffling than whites
  - Thus, their LD region is VERY small
- So not only are minorities...well, a minority of the US population, they also have very small LD blocks



■■■■ alternative route

★ possible location of admixture with Neanderthals

kya 1,000 years ago

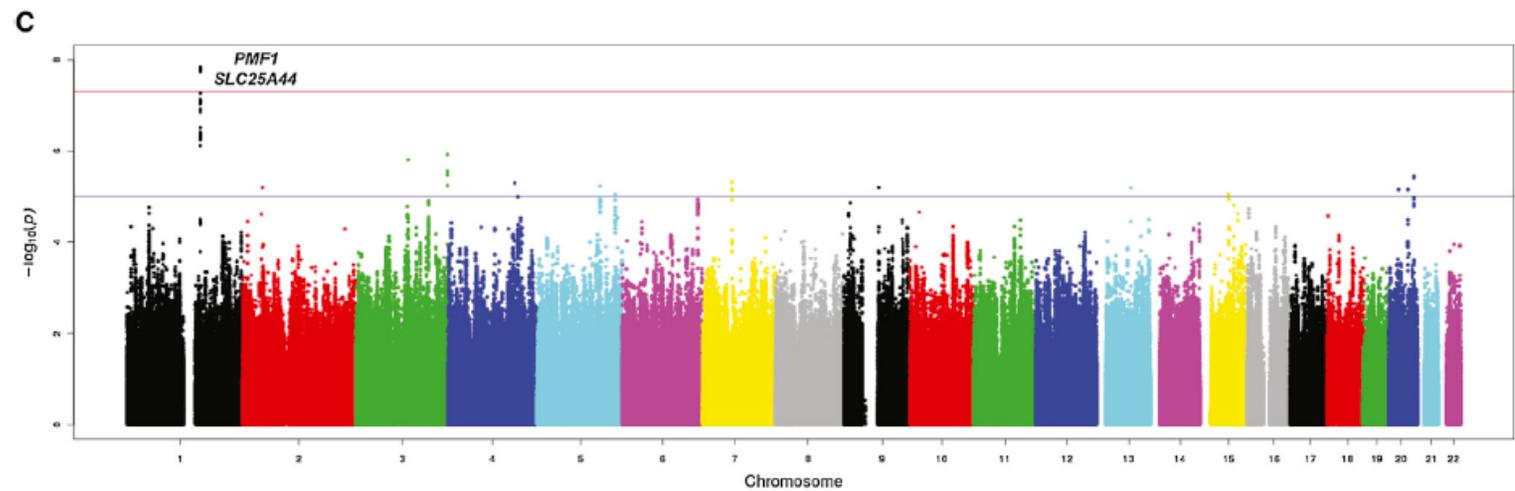
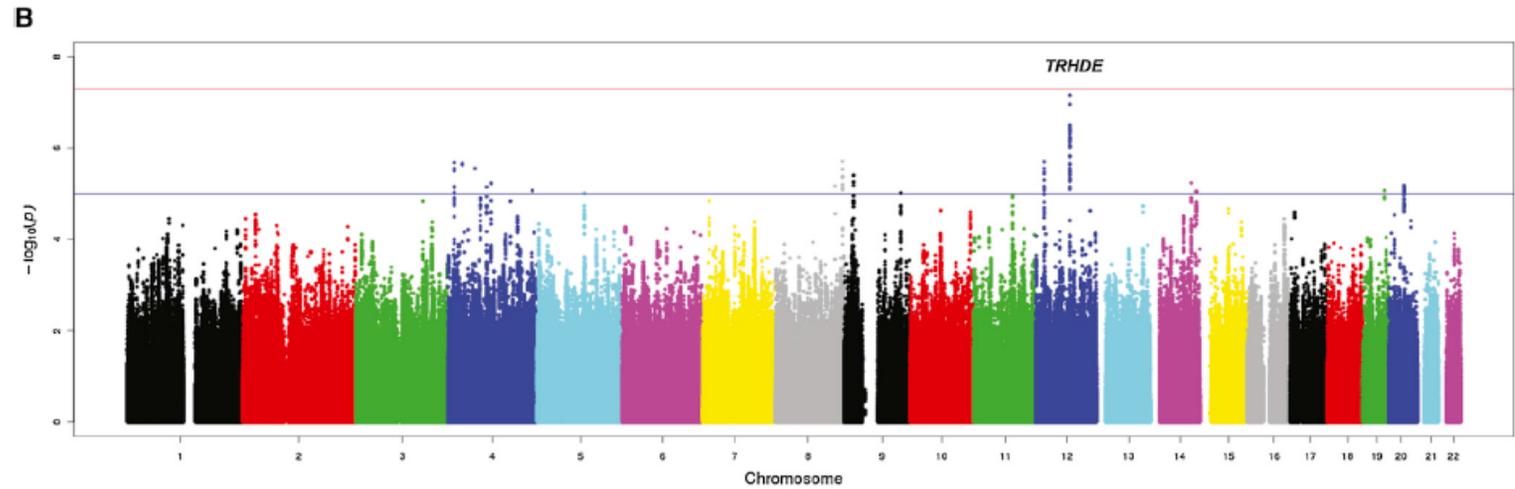
★ possible location of admixture with Denisovans

# Genome Wide Association Studies

- By evaluating roughly one million markers across the genome, we can evaluate which markers appear to be inherited along with 'stroke gene'
- Bonferroni correction for one million markers =  $0.05/1 \text{ million} = 0.05 \times 10^{-8}$  otherwise known as 'genome wide significant'.

# The Manhattan Plot

• Top 50 p-values



# So now you've learned

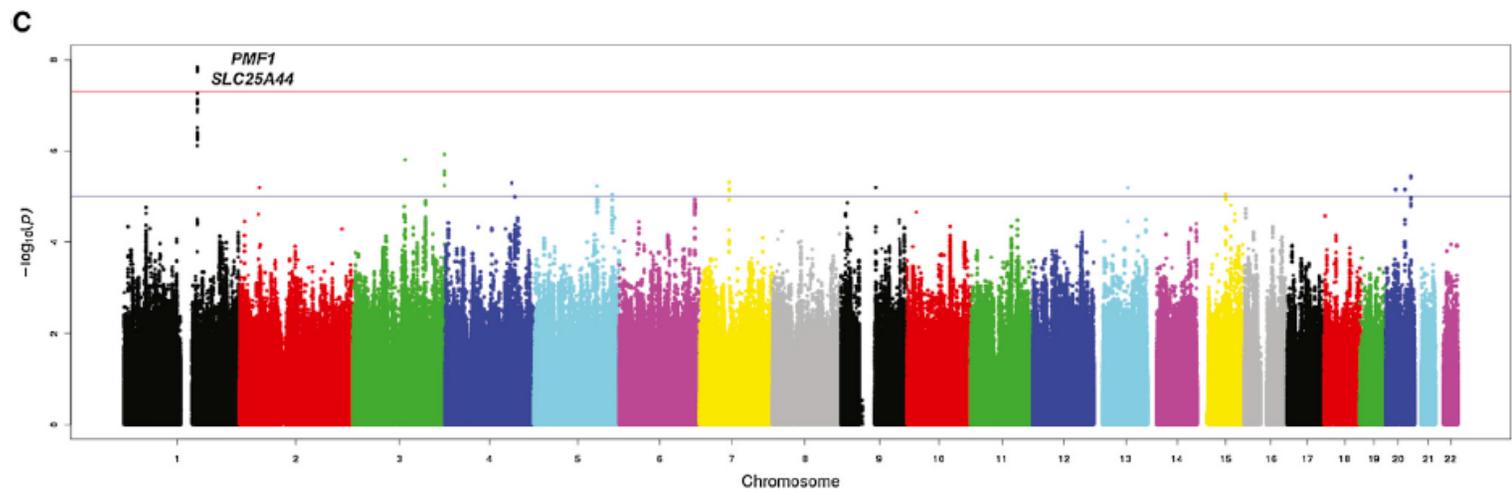
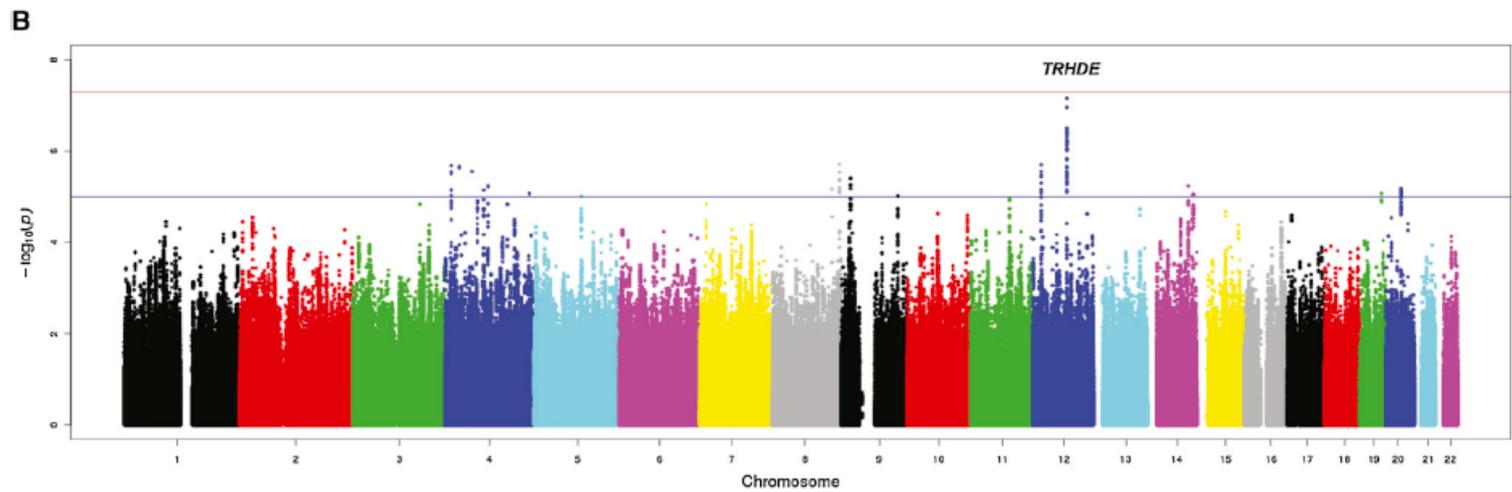
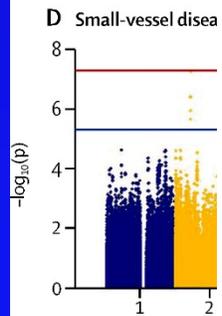
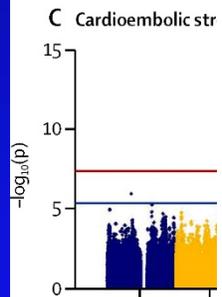
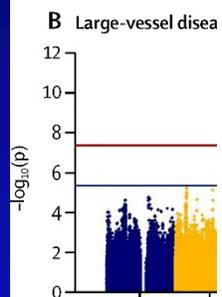
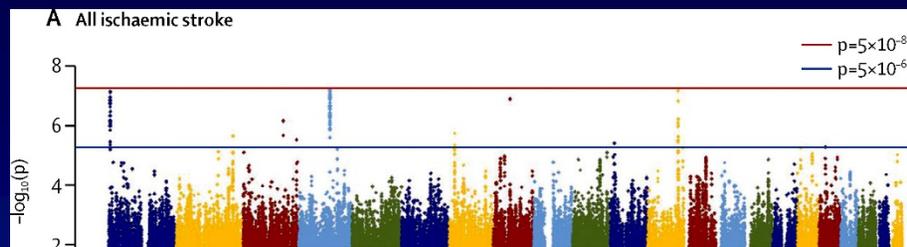
- A polymorphism is any variation in the genome
- The alleles are the different versions of the polymorphism
- A genotype are the different alleles at a single location
- We gain complexity not through more genes but through more variation of genes (splice site variation, variation in expression, gene-environment interactions)
- A simple trait is one gene/one disease while a complex trait may have multiple genes, multiple gene-environment interactions, variation by regulation making a tiny increase in risk over time
- Linkage disequilibrium is the phenomenon that two places on the chromosome will be inherited together more often than places far away from each other

# So now you've learned

- Genome wide significance for a million polymorphisms is  $p < 5E-8$
- Winner's curse is the person who wins has likely overvalued the effect size (that's how they won!); replication is king
- Manhattan Plots are the log of the p-value (turns it positive and logarithmic scale)

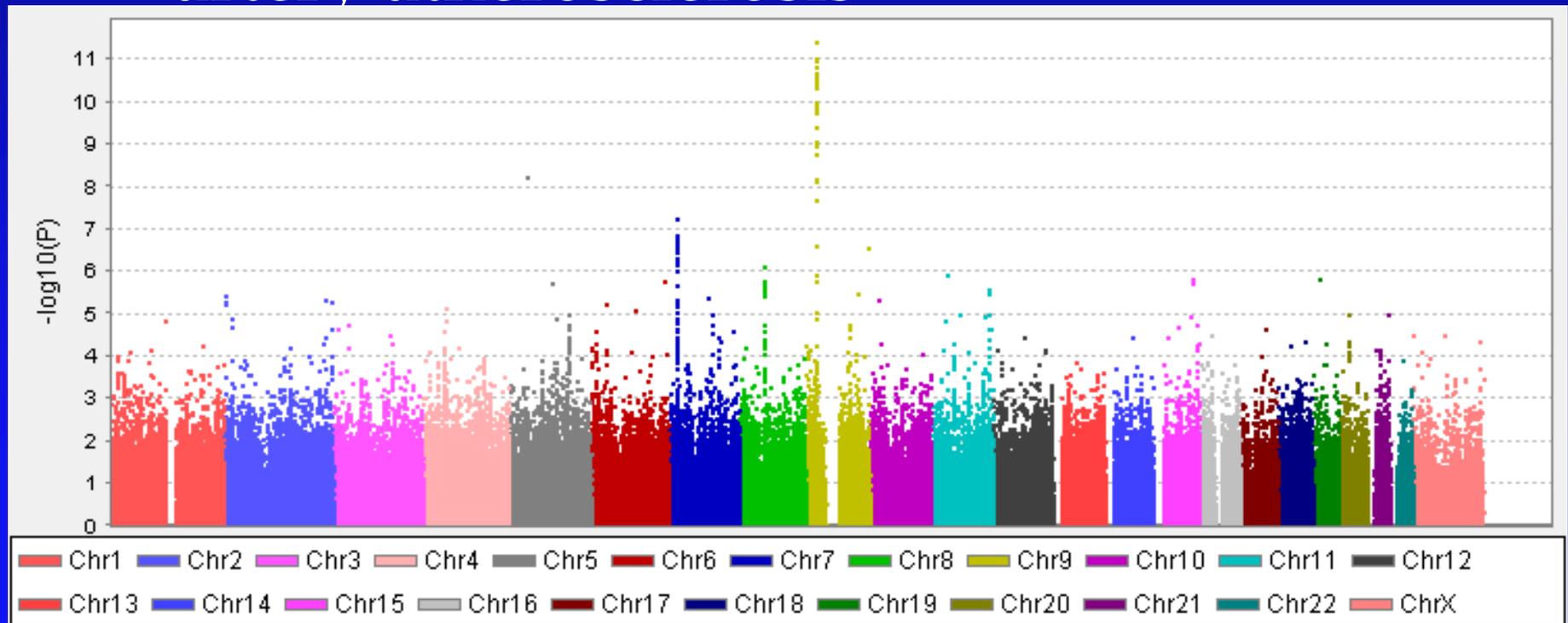
# International Stroke Genetics Consortium

- Multiple investigators from around the world; clinicians, geneticists, statisticians, basic scientists
- 2 workshops per year
- Membership is free
- Power of sample size, networking, collaboration
- Central repository of data, some samples and tissue



# New Phenotypes?

- HDAC-9 – Most consistently replicated genome wide association for large artery atherosclerosis



# New Phenotypes?

- CHR-1 locus
  - Associated with Deep ICH
  - Also genome wide significant for white matter disease
  - Also associated with microbleeds
  - Also associated with small vessel ischemic stroke
- Small vessel disease as a new phenotype with variation in expression?

# Numerous others not covered

- Dozens of genome significant associations with:
  - Intracranial aneurysm
  - Intracerebral hemorrhage
  - Ischemic stroke
    - Small vessel specific
    - Large artery specific
    - Cardioembolic

# Participation into Genetics

- Many pharma trials include a blood sample for research purposes which can include genetic analysis
- GINA – Genetic Information Non-discrimination Act (2008) makes it illegal to discriminate on someone based on genetic information including insurance companies
- Even if no direct benefit, may help future generations, identify variation that affects response to treatment

# Consent Rates

- Framingham >95% consent rates (Levy et al 2010)
- NHANES; 90.1% of eligible patients to include into a publicly available database (McQuillan et al 2006)
- Pediatric Genetics; 90% consent rates (Papaz et al 2012)
- Acceptance is growing over time

# Pharmacogenomics

- CYP2C19 and Clopidogrel
  - Genotype based prediction of poor to ultra rapid metabolizers
  - 21-26% of population are unresponsive to clopidogrel
- Statins – large variability of response
- Aspirin – COX1 and 2 variation and P2Y1 and P2Y12 variation on response to aspirin?
- tPA and ApoE – Less responsiveness after controlling for stroke severity in patients with ApoE4 alleles

# Pharmacogenomics of blood pressure response?

- Which agent is best for lowering blood pressure may be determined through a blood test in the future
- Racial variation in response to beta-blockers or thiazides
- Side effects from ACEi and myopathy in statins predictable through genetics

# The future?

- Genetic research is revolutionizing our understanding of the phenotype of stroke – atherosclerosis, small vessel disease, vascular response, neurovascular unit, inflammation
- New targets for treatment to reduce risk, improve recovery, responsiveness to treatment
- An understanding of the terminology and capabilities can enhance your research!

**Questions?**