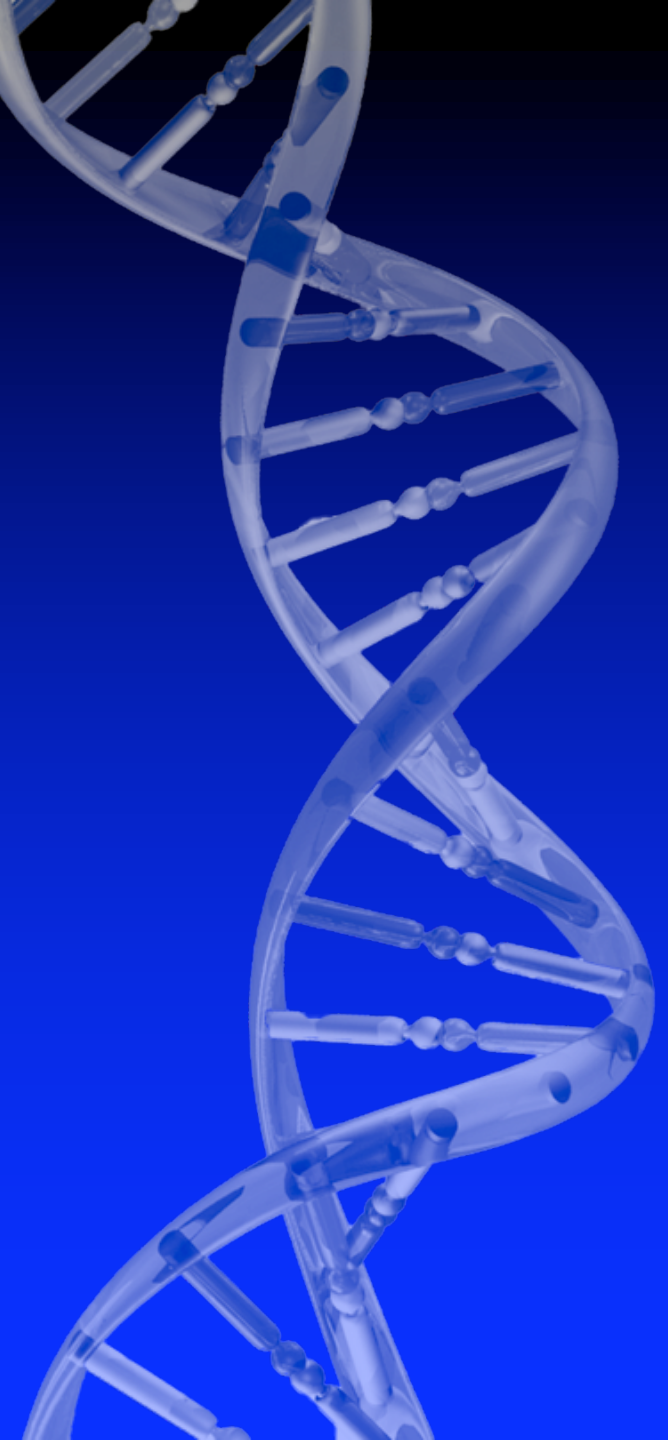

Summary


NEXT STEPS
dreamstime

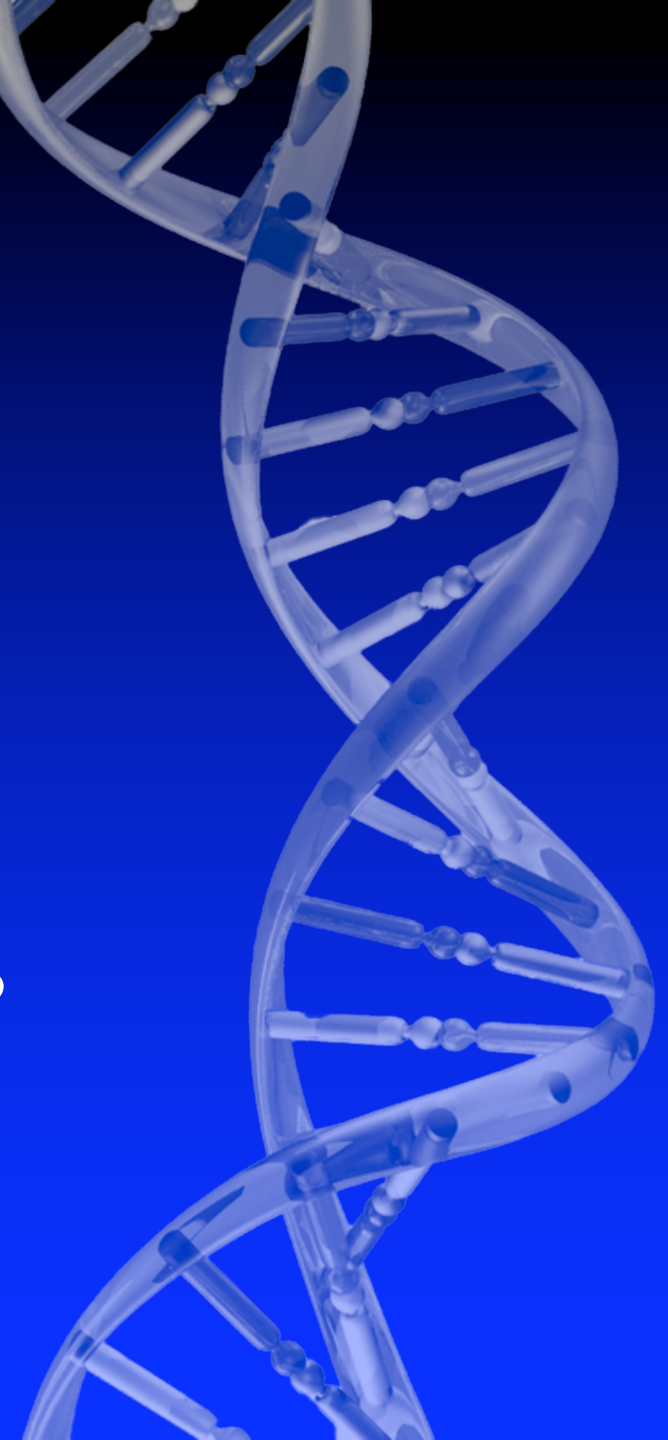
What Have We Learned?

- Using all associated or even all genotyped SNPs explains much more h2 than genome-wide significant (GWS) SNPs: height 45% vs 5%; LDL
- Rare variants have larger effects but collectively contribute small amount h2
- Few examples of dominance variance (1-3%)
- Much genetic variance captured by arrays
- ➢ Few interactions currently seen are with very large effect loci like MHC



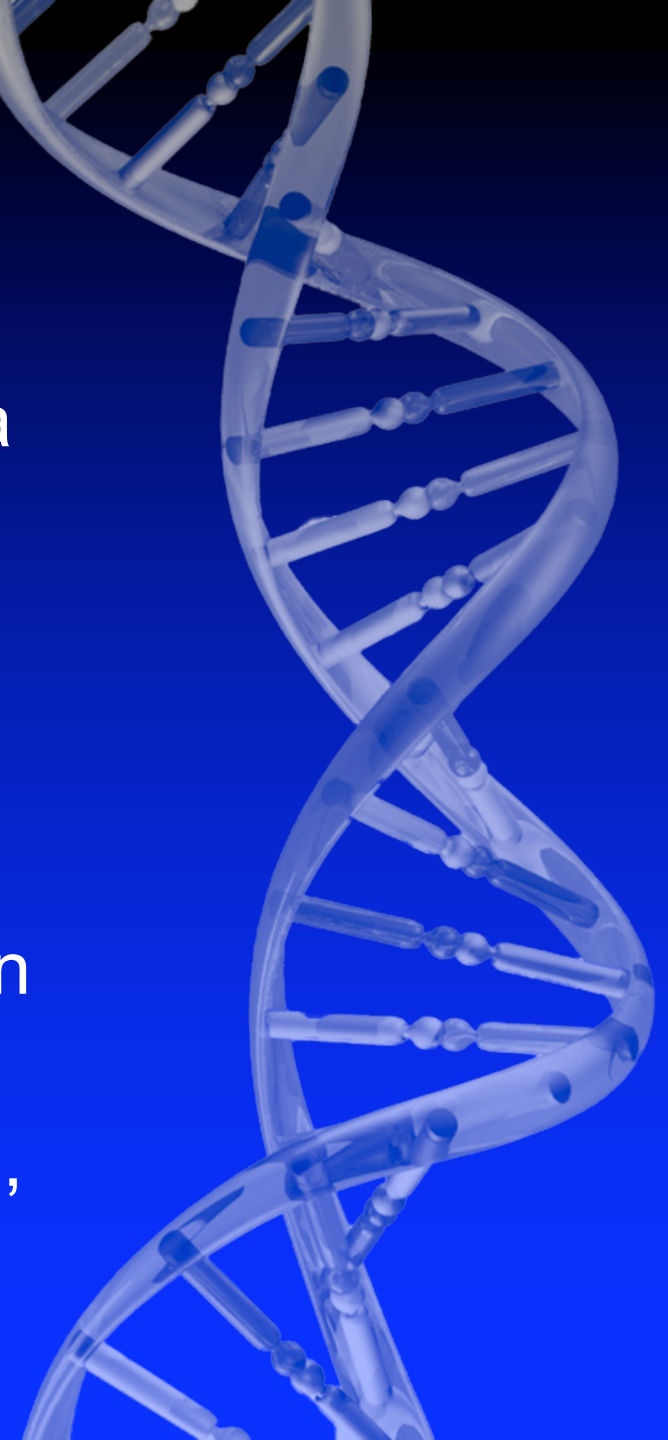
What Have We Learned?

- In Crohn's transcriptomics more predictive of disease course than genomics, probably more environmental– *others?*
- Lifetime risk has major impact on h2 estimates yet rarely know lifetime risks
- PheWAS new since 2009, shows IBD co-segregates with P disorders, long QT; protective for tongue-tied
- Beginning to explain pleiotropic surprises: *LRKK2* kinase domain variants in Crohn's and Parkinson's



What Have We Learned?

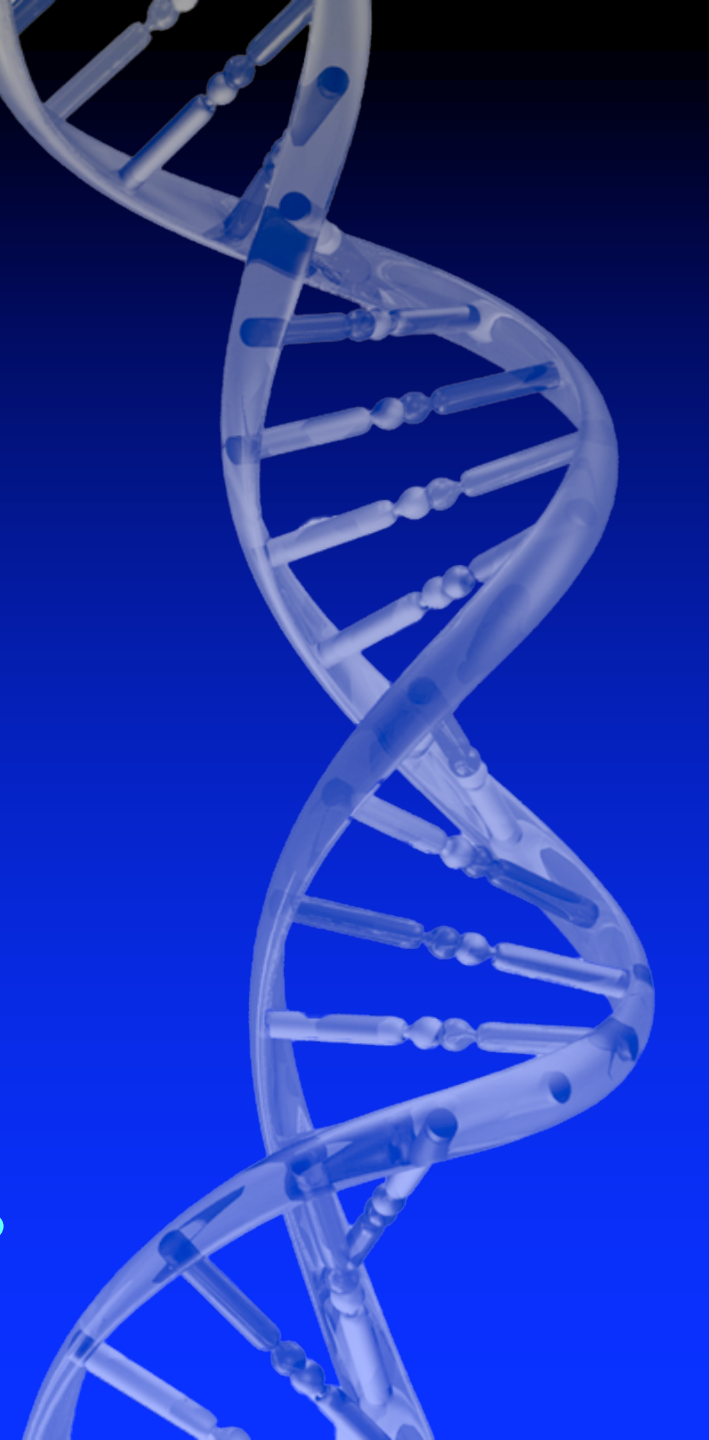
- Clinical diagnostic sequencing can work well for subset of diseases and well-defined genes, even with poor phenotypic characterization and no idea of gene; would not have expected
- Sequencing in complex diseases identifies significant number monogenics where treatment altered: Alport's, Wilson's, MODY– when to look?
- Rare that point mutation outside gene will have strong effects on gene expression because built-in redundancy– remains to be proven
- SVs across genome make very small contribution, because relatively few associations (~1000?) but at individual locus 2-4X variance of lead SNP



What Have We Learned?

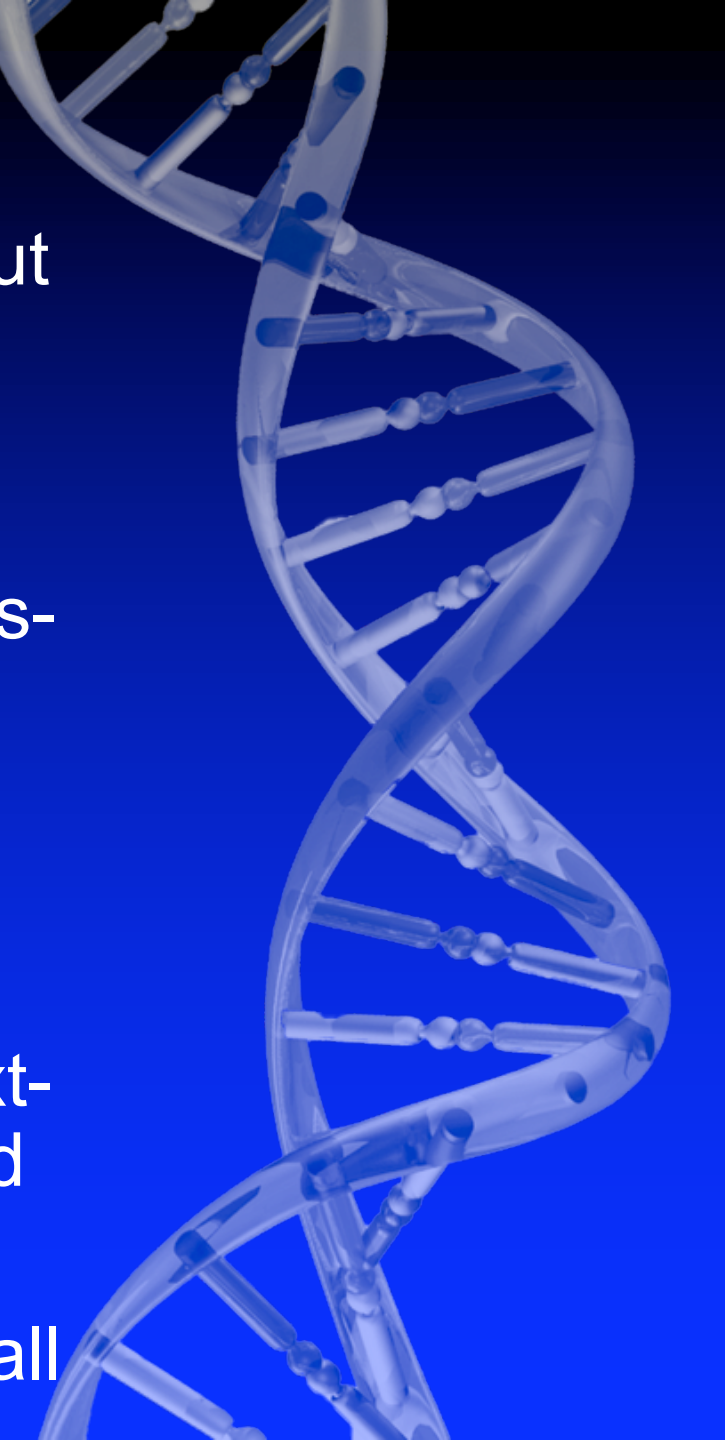
- Acquired mutations may contribute to h2—tendency to mutate and object of clonal selection are inherited?
- > 8K mosaic segmental mutations at least 1% fraction in 150K UKBB ppts; cluster in genomic hotspots like fragile sites
- Value of widely accessible datasets on vast numbers of people; imputation and IBD more powerful as datasets expand

➢ Most phenotypic variance due to regulatory variation in genes expressed in “right” tissues but *without direct roles in disease– how do we know?*



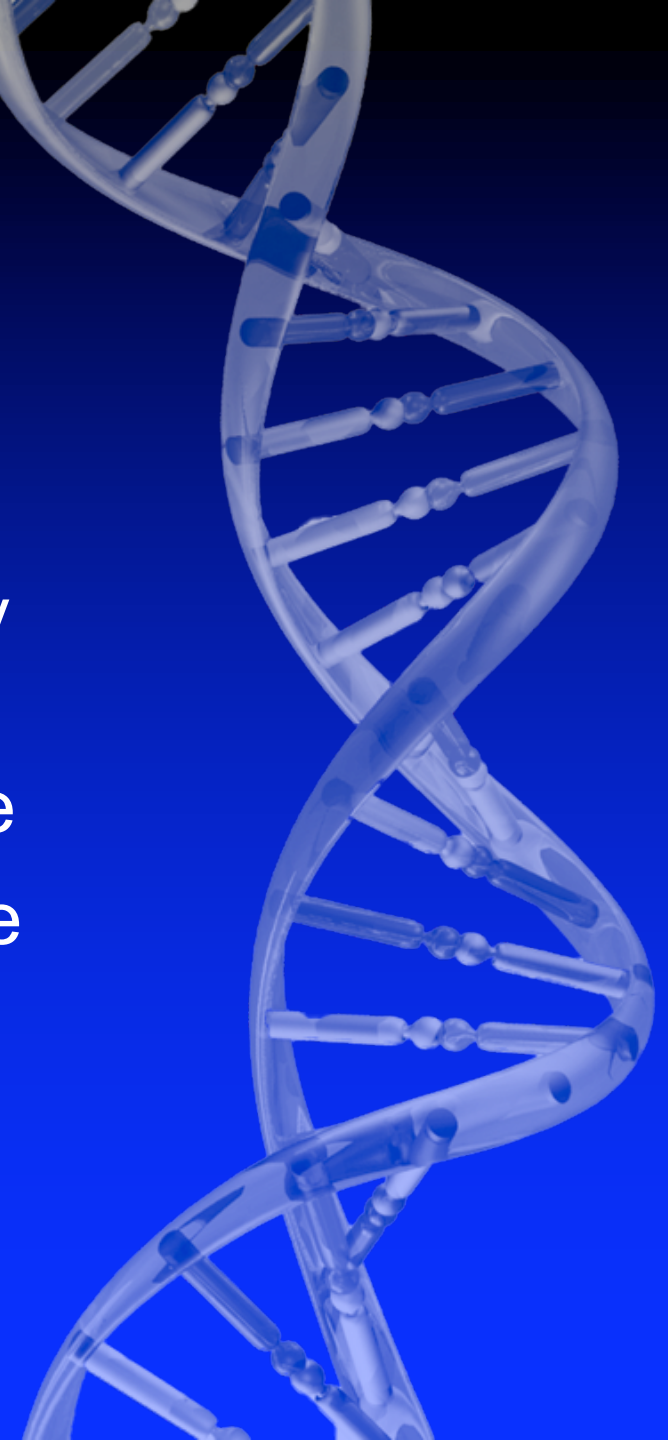
What Have We Learned?

- Peripheral genes outnumber core genes 100:1 but effects very small, may explain why huge fraction of genome contributes to single trait- model to be studied
- Genetics of gene expression: large catalogs of cis-eQTLs, diverse contexts, variants, phenotypes
- Rare variants drive extreme expression levels, in aggregate may explain large proportion h2 of expression
- ➢ Genetic variation in sexual dimorphism as context-dependent effect; in flies massive gene X sex and gXe interactions
- Much interaction is antagonistic, may explain small effect sizes– *could this be similar in humans?*



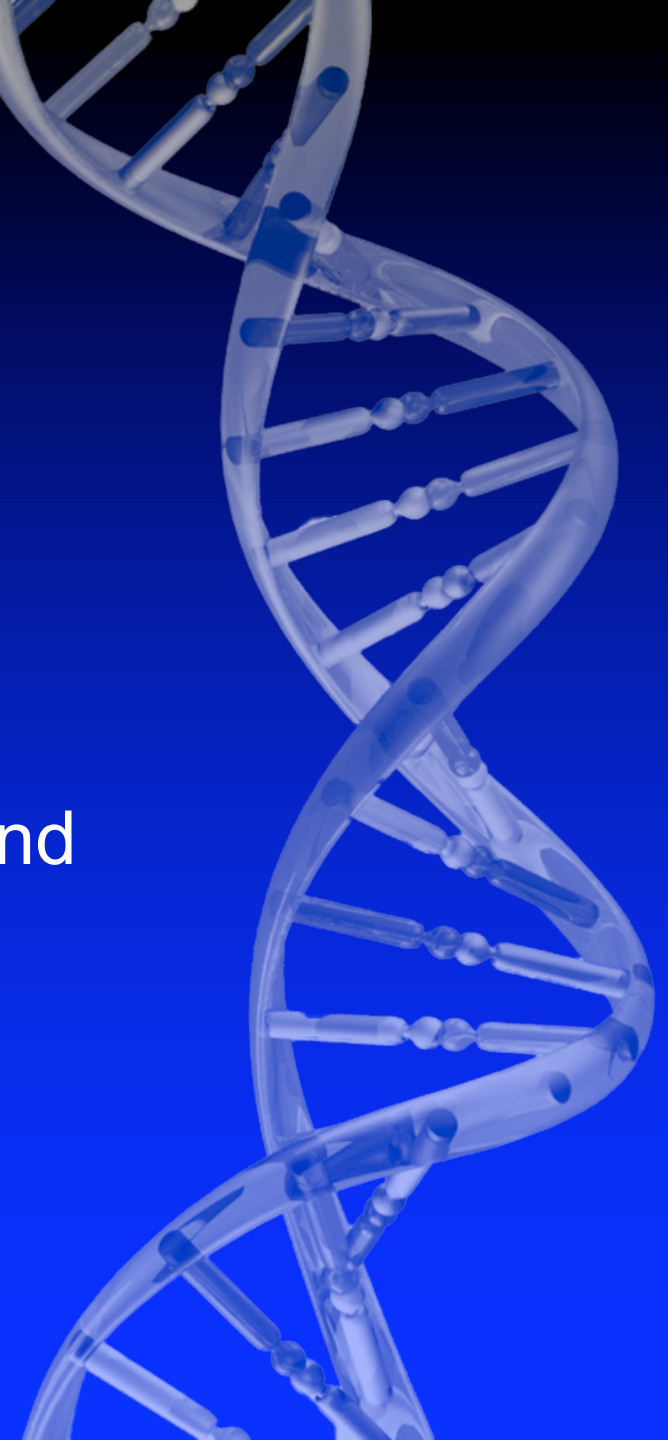
What Have We Learned?

- Had expected SNPs in cancer pathways to affect multiple cancer types but now 90% of SNPs or even loci in cancer not seen in another cancer
- Polygenic risk scores can separate 10-fold differences in risk, will soon be important clinically
- True gXe rare, as is eXe; partly due to need for large studies and accurate classification exposure
- Even if no true interaction, absolute risk difference of non-genetic RF at high genetic risk much greater
- Risks seem to multiply without synergism, interactions unlikely to improve prediction– good news for risk prediction algorithms



What Have We Learned?

- Family studies valuable for:
  - causal *de novo* mutations: false negatives 4%
  - detection of shared genomic segments
- Heterogeneity in Mendelian conditions extensive, both allelic and locus, “multi-Mendels” 3-4%
- Two locus models can explain incomplete penetrance: *SMAD6* and *BMP2*; also *TBX6* null and hypomorphic alleles
- Benefit of adding 50K non-Europeans to GIANT: reduces credible set sizes, increases post prob
- Integrated analysis with PC across diverse population more effective than stratified



What Have We Learned?

- Controlling for global ancestry does not remove effect local ancestry, use chromosomal segment as unit of analysis
- Admixture mapping methods evolved now to point of segmental analysis (2009 question)
- Specific populations starting to reveal strong effect alleles, largely founder effects (2009 question)
- X and mito still under-studied (Y?)

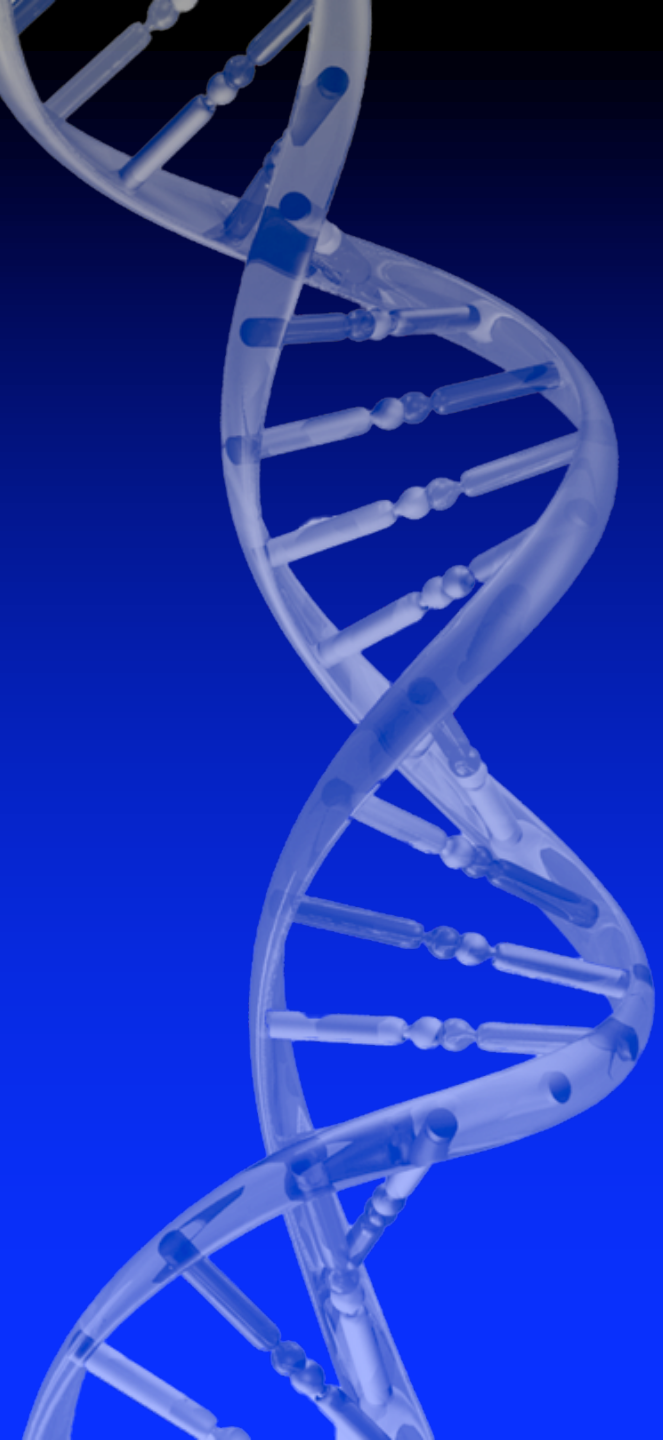
➢ Can have pervasive epistasis and additive models still fit– “mundane finding”

➢ h2 and effect size estimates can be biased upwards

Where Do We Go Next?

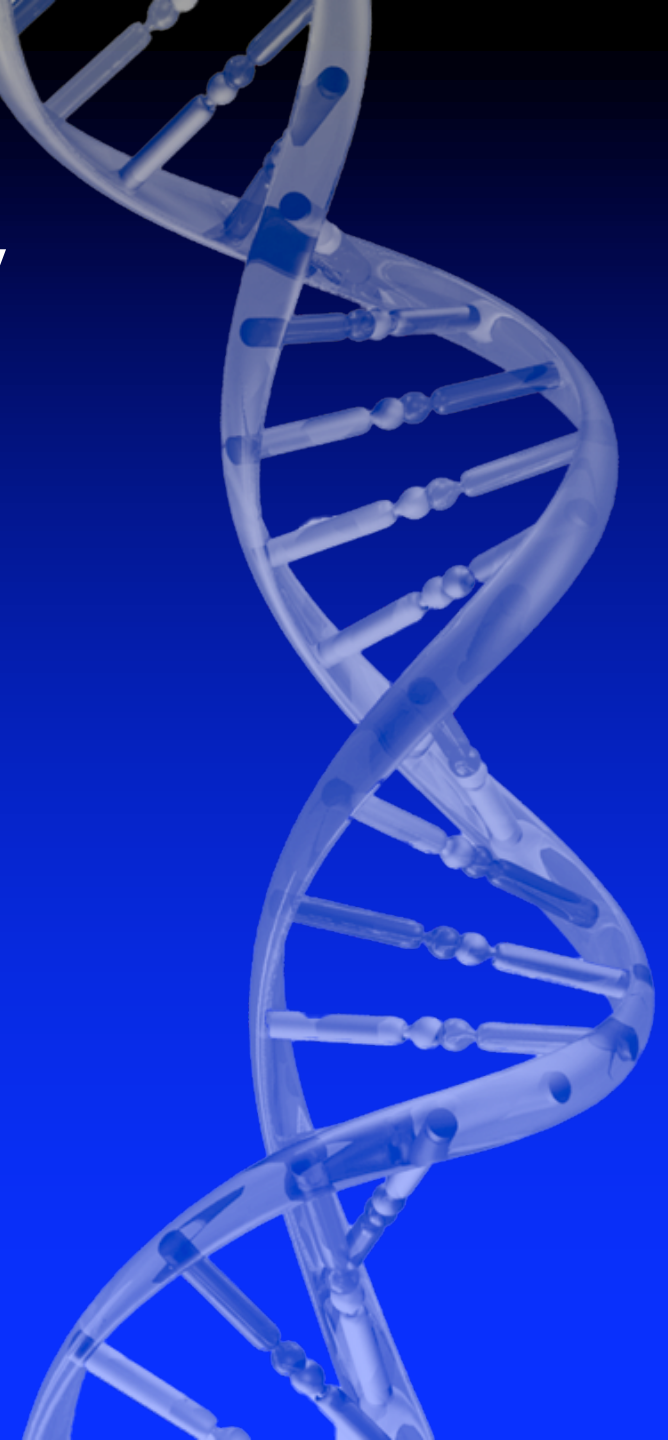
Where Do We Go Next?

- Estimate genetic variation using large (>50K) WGS samples
- Estimate variance due to non-SNP variation
- X chromosome!!
- Large numbers of families in studies needed to dissect within vs between family effects
- Genetics of disease progression/severity
- Need analysis paradigm that's true burden analysis, not collapsing of point mutations
- Systematic analysis of biobank genetic data
- Domain-specific sequence annotation
- Need to know more about transreg networks– how they behave when perturbed



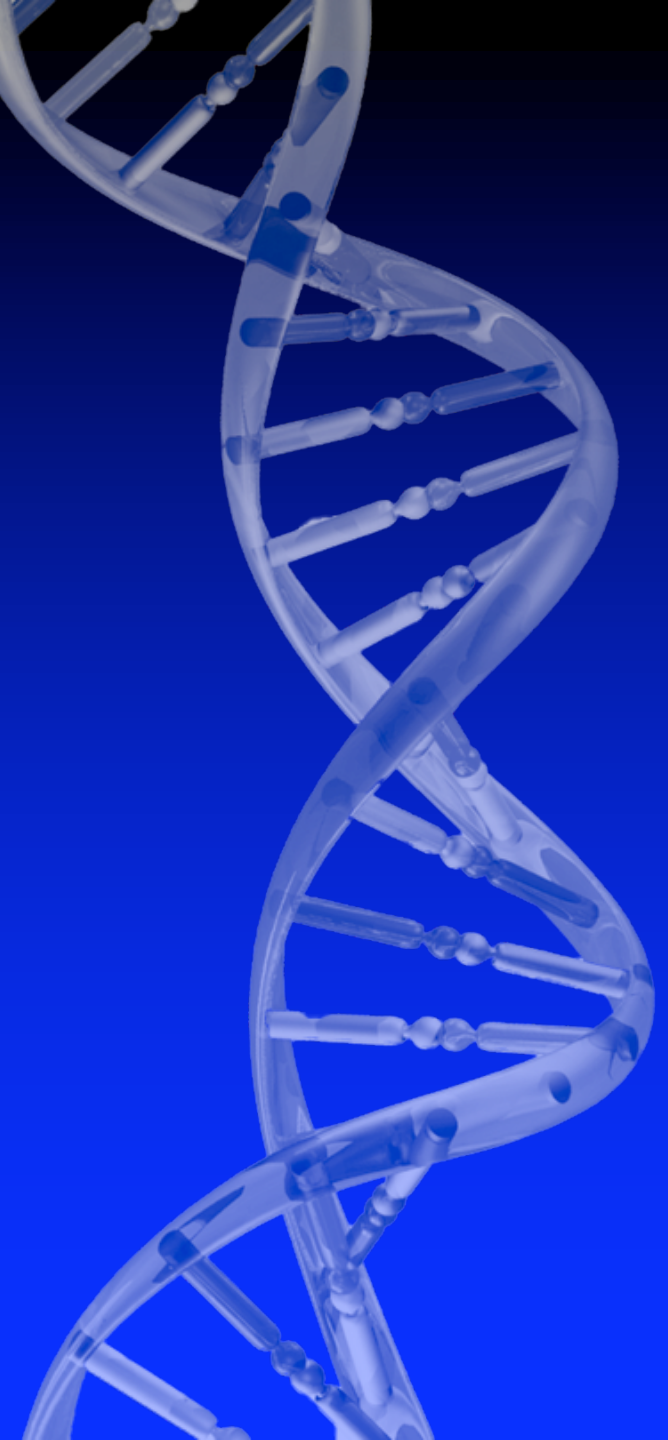
Where Do We Go Next?

- VNTRs and clonal expansion of mosaic sites may be whole new frontier
- Scaling up expression studies overall:
  - Larger sample sizes
  - Single cell analysis - Human Cell Atlas, etc
  - Integrated analyses connecting epigenetic and expression data and GWAS
- Study expression during development and “de-development” (in cancer)
- Phenotype risk scores to find hidden Mendelians in population, characterize phenotypic variation associated with genes not looked at yet



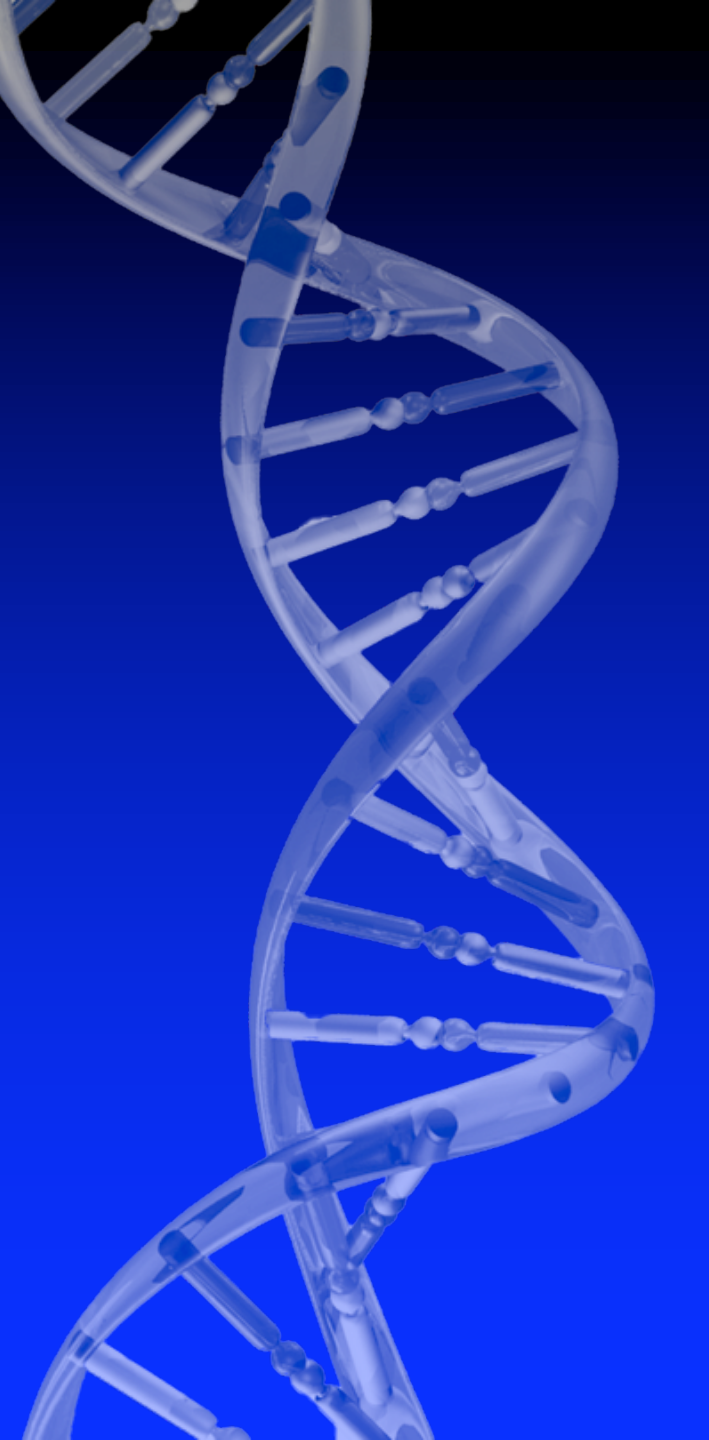
Where Do We Go Next?

- Specific test of core vs. peripheral regulation– condition on set of core genes
- Major need to build African and other non-European populations; African chip will help
- Leverage subpopulation differences like Dominican and Puerto Rican
- More explicit modeling with infectious agents
- Push for available data

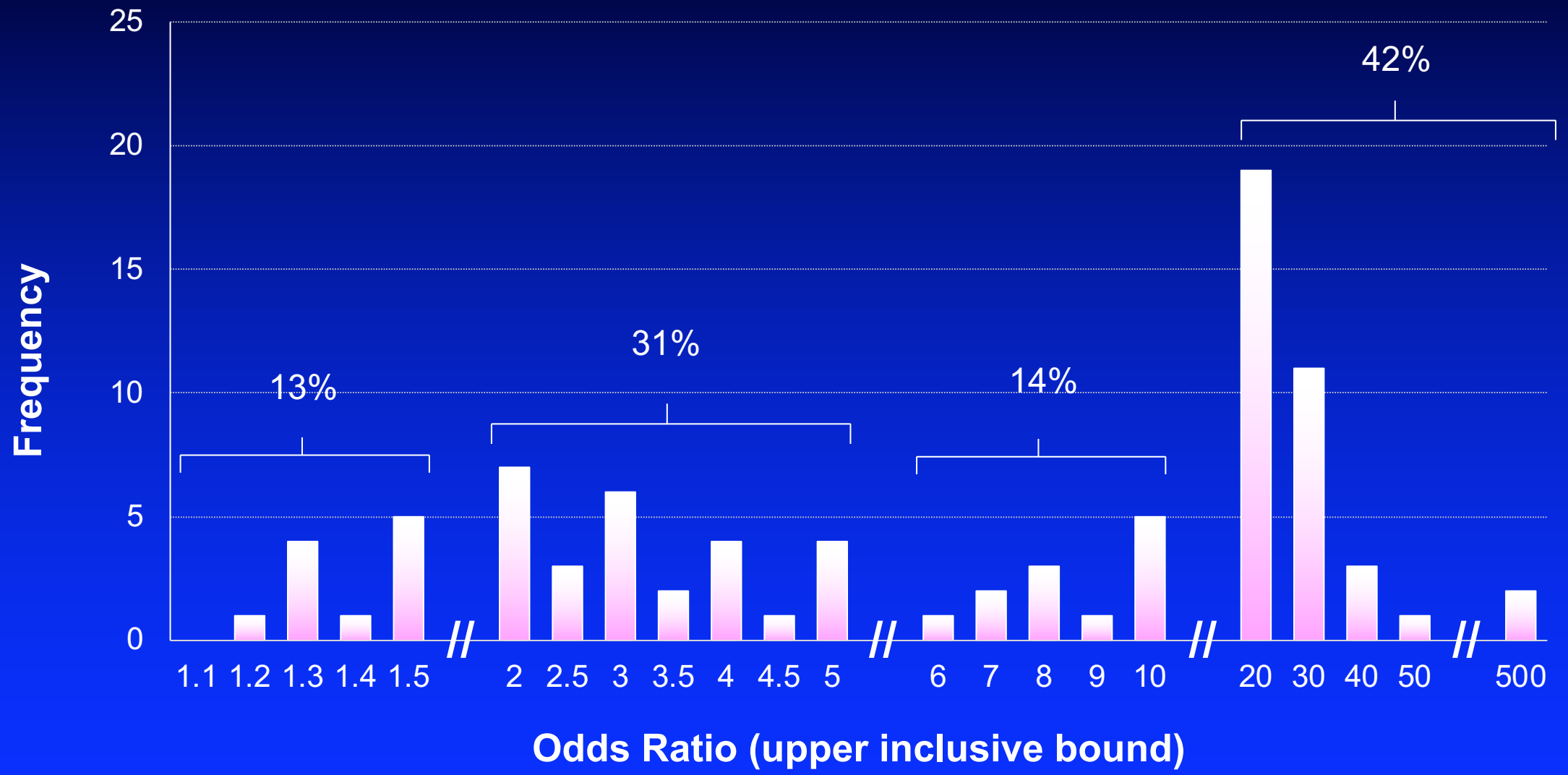


Open Questions

- What modifies large effect mutations?
- What explains widespread signals throughout the genome?
- How much h2 driven by expression?
- How much h2 driven by epigenetics?
- Why do we see such strong gXe and epistatic interactions in animal and plant literature and not in humans?



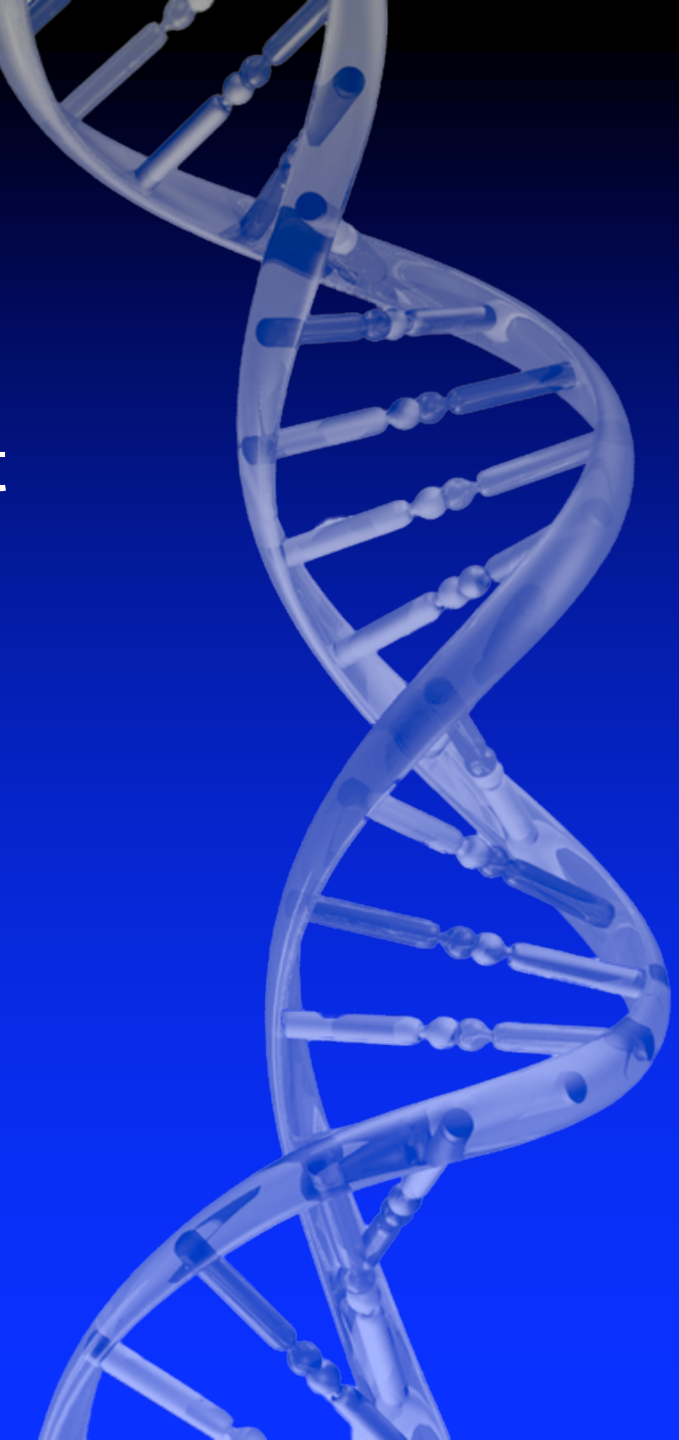
Histogram of Odds Ratios, MAF ≤ 0.01



NHGRI-EBI Catalog, 86 discrete trait OR for MAF < 0.01, 4/26/18

What is the Value?

- Essential for prediction: 45% variance, 40% prediction for height
- So much of genome contributes to variation, impt contribution
- Value fo r understanding evolution
- Neil Risch's 15 alleles for autism-- revolutionary



10 years of discovery (not being smug)

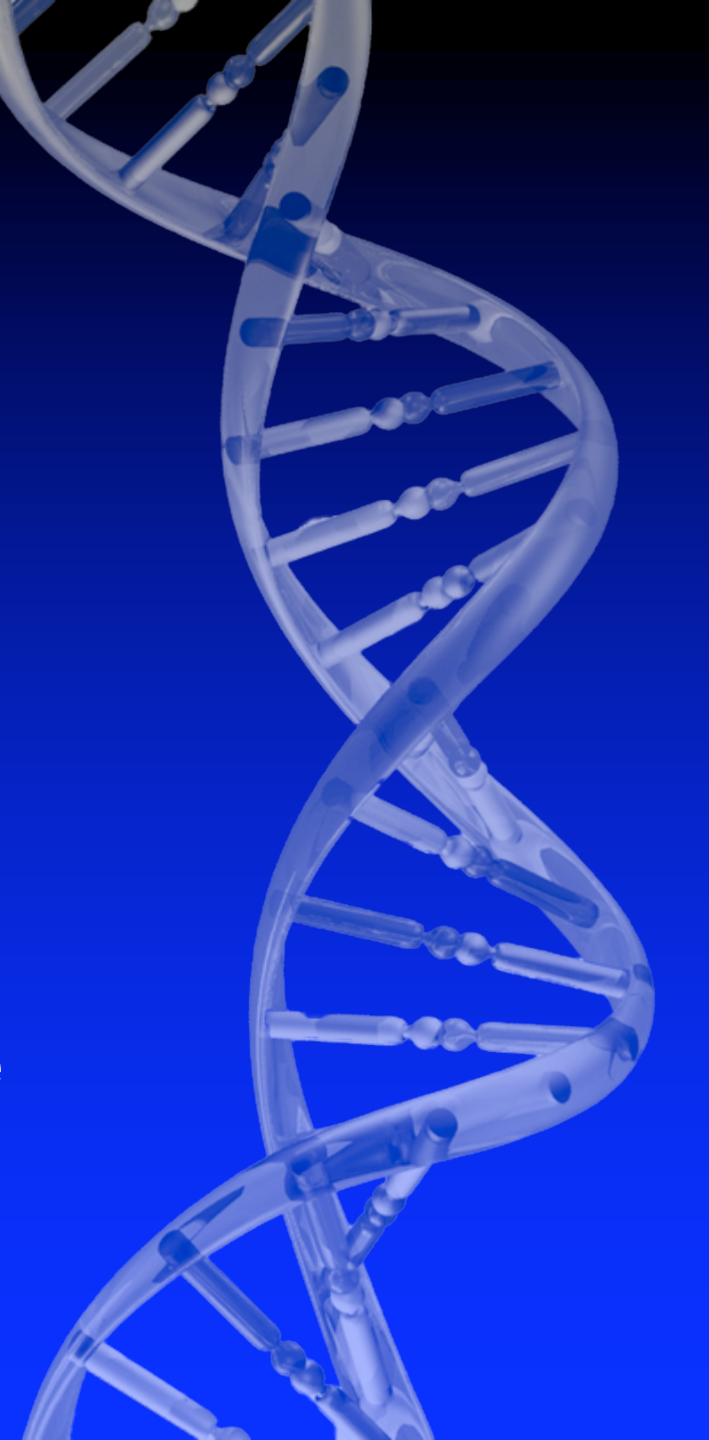
- Enormous progress (e.g. contrast with decades before)
  - GWAS as an experimental design not questioned anymore
  - Common vs rare variant debate largely disappeared
  - Resolution of genes and (some) gene variants
  - New questions, new discoveries, new knowledge
- Technology & data-driven hypothesis generating science
- Powerful data resources
  - GWAS summary statistics & GWAS Catalogue
  - GTEx
  - Epigenetic Roadmap; ENCODE
  - UK Biobank

Broad summary of workshop

- Core: Quantification of genetic architecture of complex traits
  - from SNP arrays to WGS (including *de novo*), within and between populations
  - substantial proportion of h^2 now captured from known variants
  - (nearly) all traits are polygenic: many genes & gene variants contribute to genetic variation
- Front-end: How does natural selection shape trait-specific architecture?
  - trait-fitness relationships
- Back-end: How does polygenicity work biologically?
  - coding changes, gene regulation
  - gene expression networks (core vs peripheral genes)

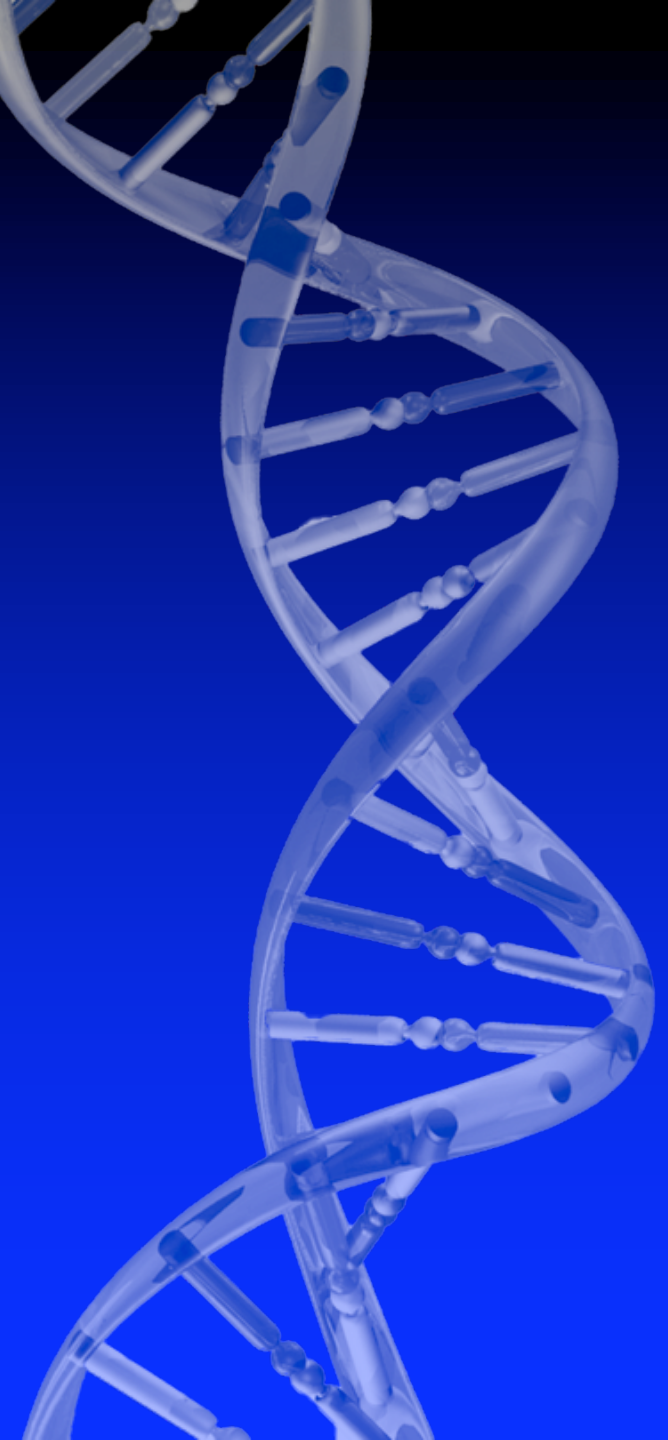
Deleted Learned

- Structurally unstable loci more challenging
- Improve inferences about trait-specific evolutionary forces to allow better predictions
- Quantify additional effect of rare variants: 57%
- Dichotomy between germline (heritable, predictable) and acquired mutations (capricious, random) not as firm as previously thought
- Sex differences also seen with induced mutations
- Understanding how evolution shapes architecture helps explain missing heritability
- Common variants explain 1/3 – 2/3 (based on method); *shouldn't this differ across traits?*



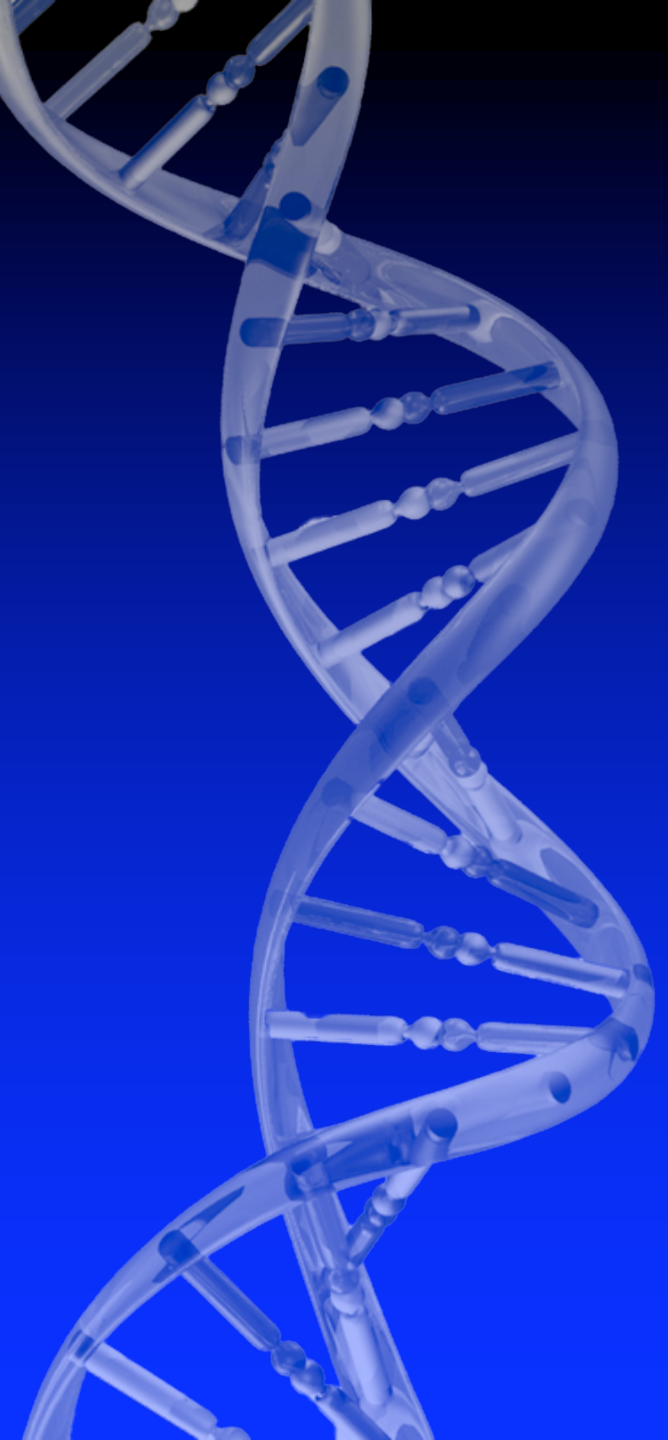
Deleted Learned

- Common variants explain 1/3 – 2/3 (based on method); *shouldn't this differ across traits?*
- Multiple methods for assessing h2-SNPs
- Binary traits are hard:
  - Difference between SNP-h2 and pedigree-h2 greater for discrete traits than quantitative
  - Methodologic assumptions violated
  - Need better data to quantitate: age of onset
- GWAS of recurring mutations, low freq causal variants (0.05-0.5%) with high ORs (19-700)



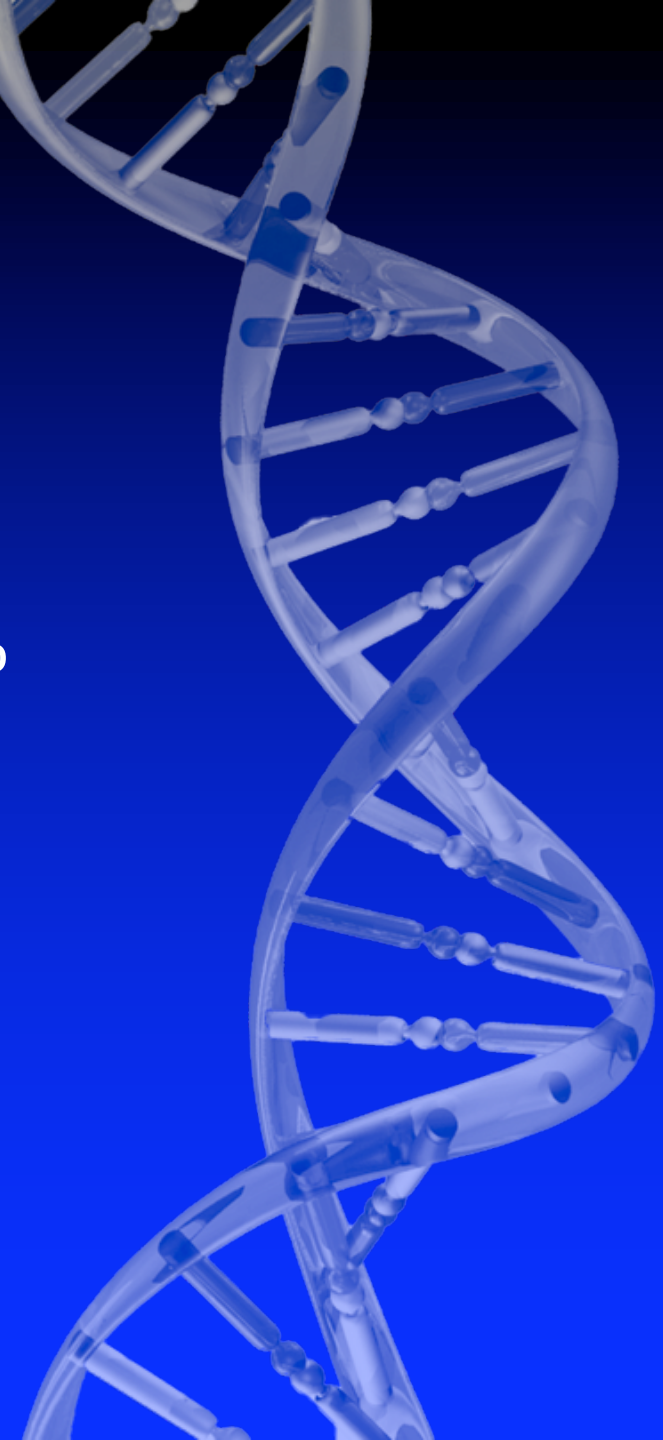
Deleted Learned

- Gene expression contributes sizeable but not majority fraction to trait h2
- Perhaps developmental order is key
- Increasing evidence of polygenic risk score-by-environment interactions: upper levels BMI in UKBB—*is this an epidemiologist's interaction?*
- Family studies valuable for:
  - causal *de novo* mutations: false negatives 4%
  - detection of shared genomic segments
- PAGE finds 150 variants with MAF>0.05 not seen in other databases; 40% increase in ClinVar conflicted variants that can be adjudicated



Deleted Learned

- Most additive variance explained by markers undifferentiated by ancestry
- Can use genetics-first approach to get into treatment early: two major infections < age 50
- GWS loci explain 20% variance in LDL-C vs. 80% genome-wide
- Common diallelic SVs now part of large datasets, routinely imputed



What Have We Learned?