Finding Modifiers of Known Disease – Related Variants

Cystic Fibrosis: Model of "Monogenic" Recessive Disorder

NIH Workshop: Sequencing in Cohort Studies and Large Sample Collections

June 29, 2012

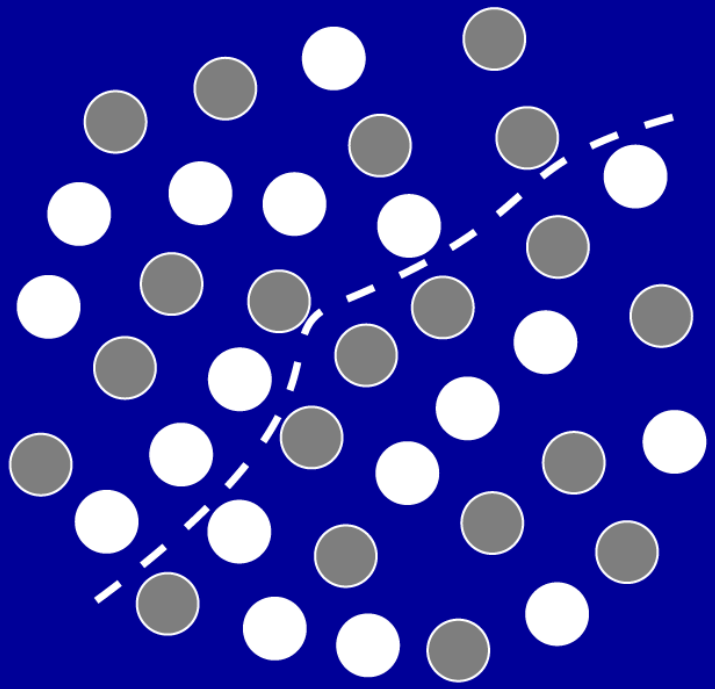
Michael R. Knowles, M.D.

The University of North Carolina at Chapel Hill

CF Gene Modifier Consortium

Genetics in Complex vs. Mendelian Genetic Disorders

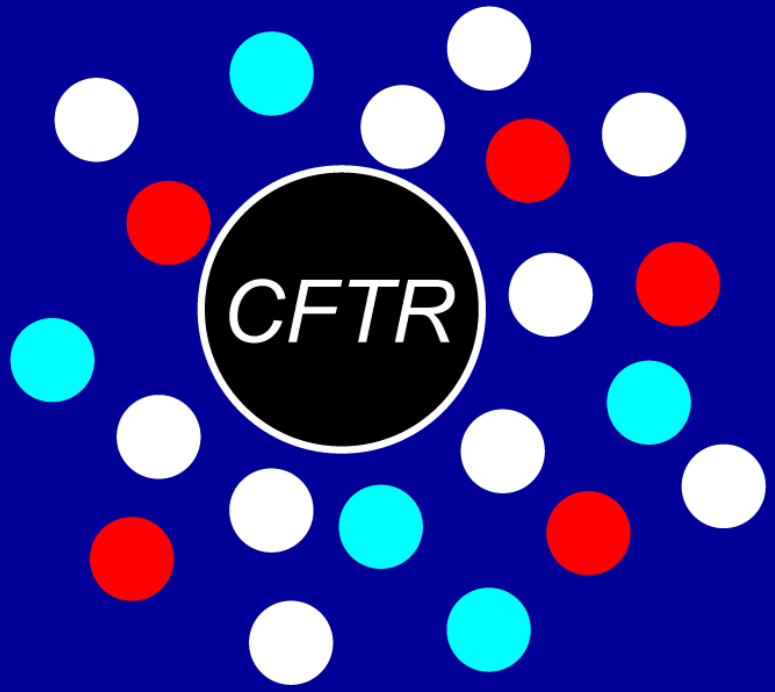
Non-Mendelian ("complex")



Genetic factors (non-Mendelian)

Environmental factors

Mendelian ("recessive")

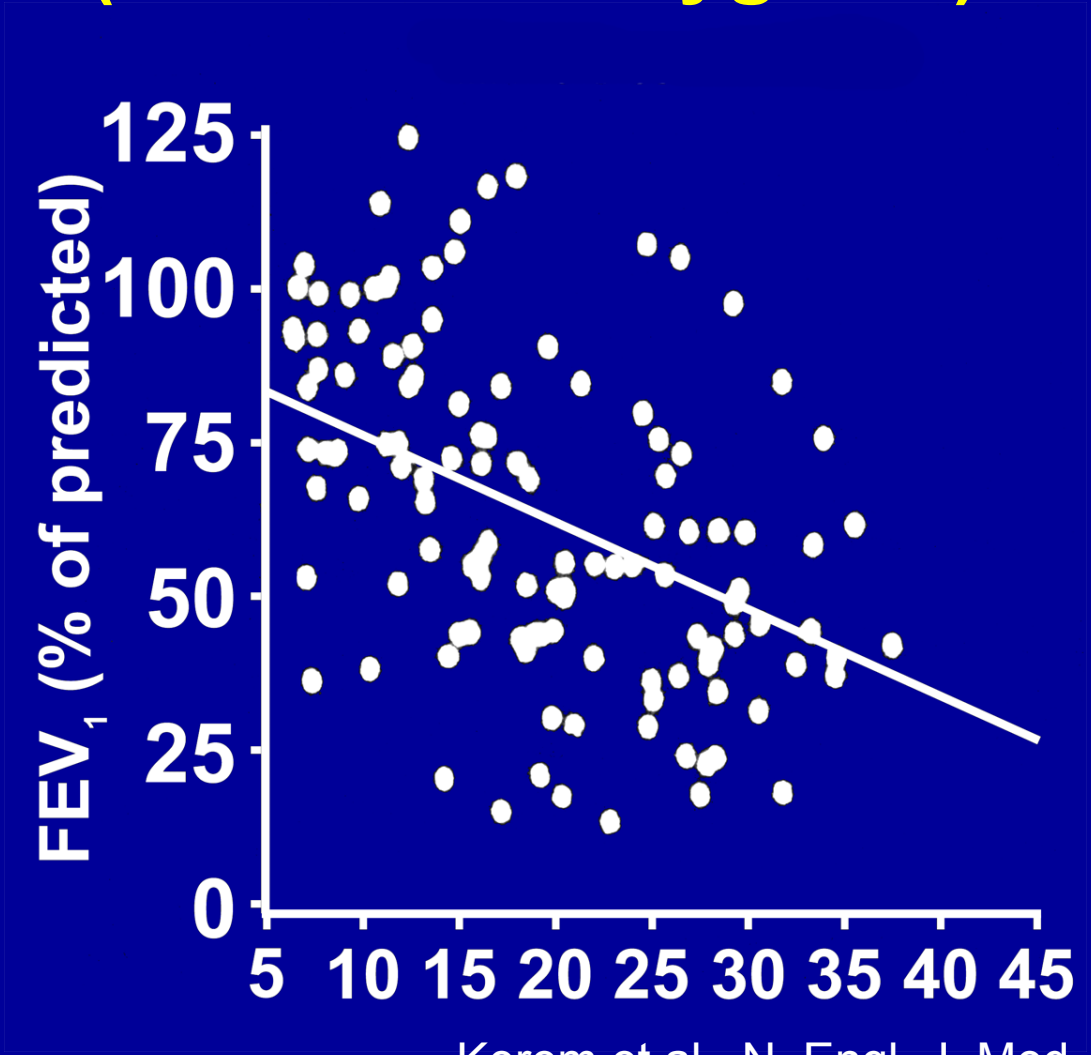


Adverse genetic factors (non-*CFTR*)

Beneficial genetic factors (non-*CFTR*)

Environmental factors

Variability of Lung Function in CF
(ΔF508 Homozygotes)



Kerem et al., N. Engl. J. Med. 1990;323:1517

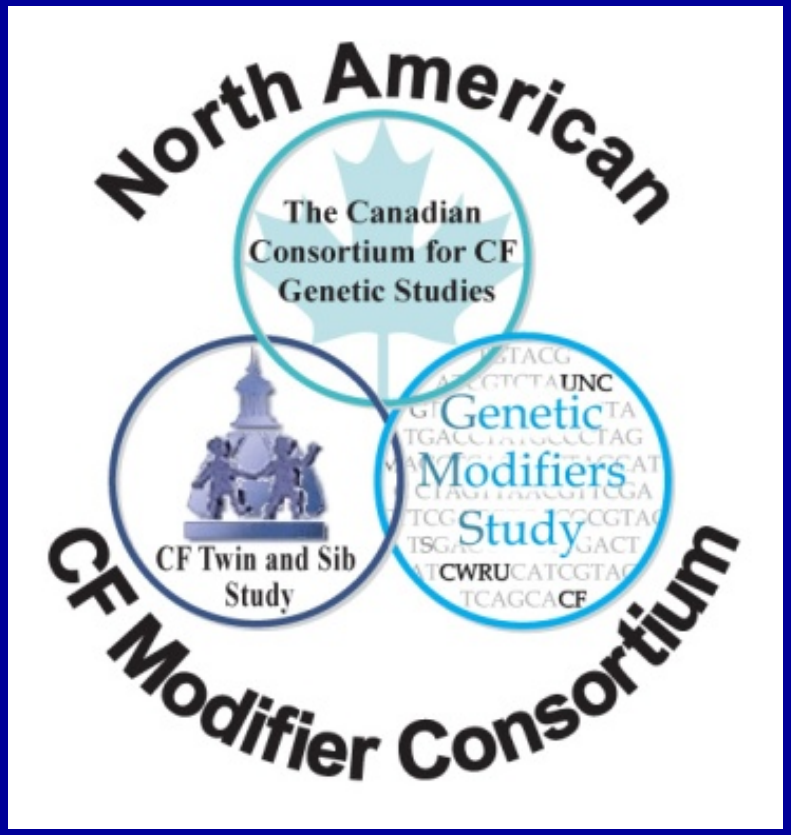
Variability of Lung Function Reflects Genetics and Environment

- European twin/sib study suggested strong (~50%) genetic contribution to CF phenotype (lung/nutrition)*
- U.S. twin/sib study indicates variability of CF lung disease primarily due to genetic factors (0.55-0.86)†

Challenge: What are the genes?

*Mekus, F. et al. 2000. Twin Res.

†Vanscoy et al. 2007. Am. J. Respir. Crit. Care Med.



Genetic Modifiers Study (GMS)
(Extremes of phenotype)

- University of North Carolina
  - Mike Knowles, Wanda O'Neal
  - Fred Wright
- Case-Western
  - Mitch Drumm

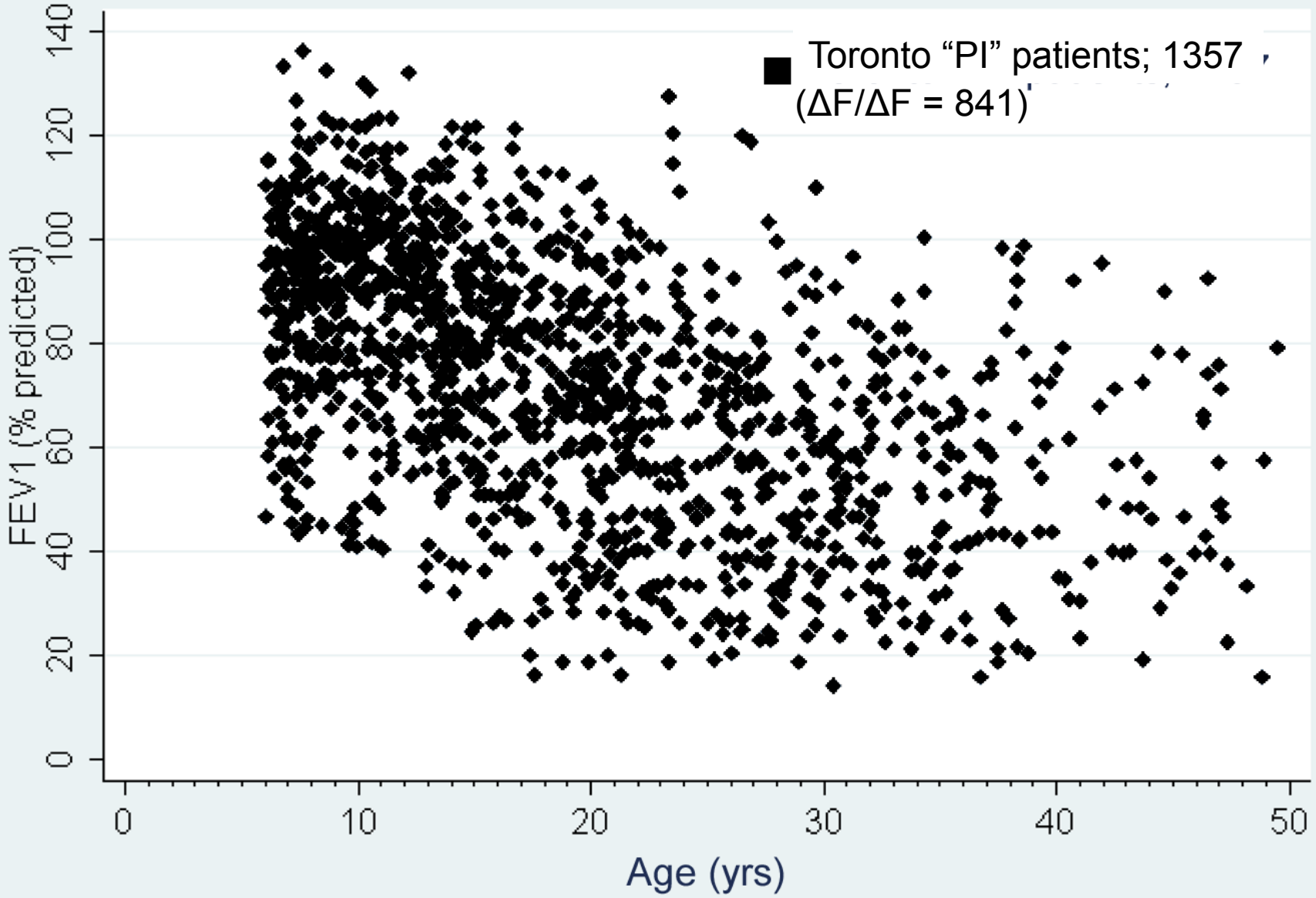
Canadian Consortium for CF Genetic Studies (CGS)

- Hospital for Sickkids
  - Peter Durie, Johanna Rommens

CF Twin and Sibling Study (TSS)

- Johns Hopkins
  - Garry Cutting
  - Scott Blackman

Challenge of Lung Disease Phenotype
FEV1 (% Pred) vs. Age



Quantitative Lung Phenotype*

- Key for joint analyses (quantitative lung phenotype)
- Used multiple measures of FEV_1 (3 years), and referenced to Kulich percentiles (age, sex, height adjusted), then survival corrected and z-transformed
- Validated; strong correlation ($r^2 > 0.90$) with Schluchter cross-sectional, longitudinal and survival models†
- Validated; robust genetic influence ($h^2 = 0.51$)‡

*Taylor, C., et al., Pediatr. Pulmonol. 2011

†Schluchter, M. et al., Am. J. Respir. Crit. Care Med. 2006

‡Wright, F., et al., Nat. Genet. 2011

Overview of Strategy

North American Consortium for CF Modifier Studies

US CF Sibling and Twin (Hopkins) (973)	Canadian Consortium (Toronto) (1,357)	Extremes of Phenotype (UNC/CWRU) (1,137 ΔF/ΔF)

(3,467 patients)
(600,000 SNPs/CNVs)

Supported by the NIH, Genome Canada and US CFF

- Replicate each other's findings
- Increase power for rare disease traits
- Maximize modifier discovery

Genome Wide Association Study (GWAS)

Genome Quebec Platform

Analyses: Key Concepts

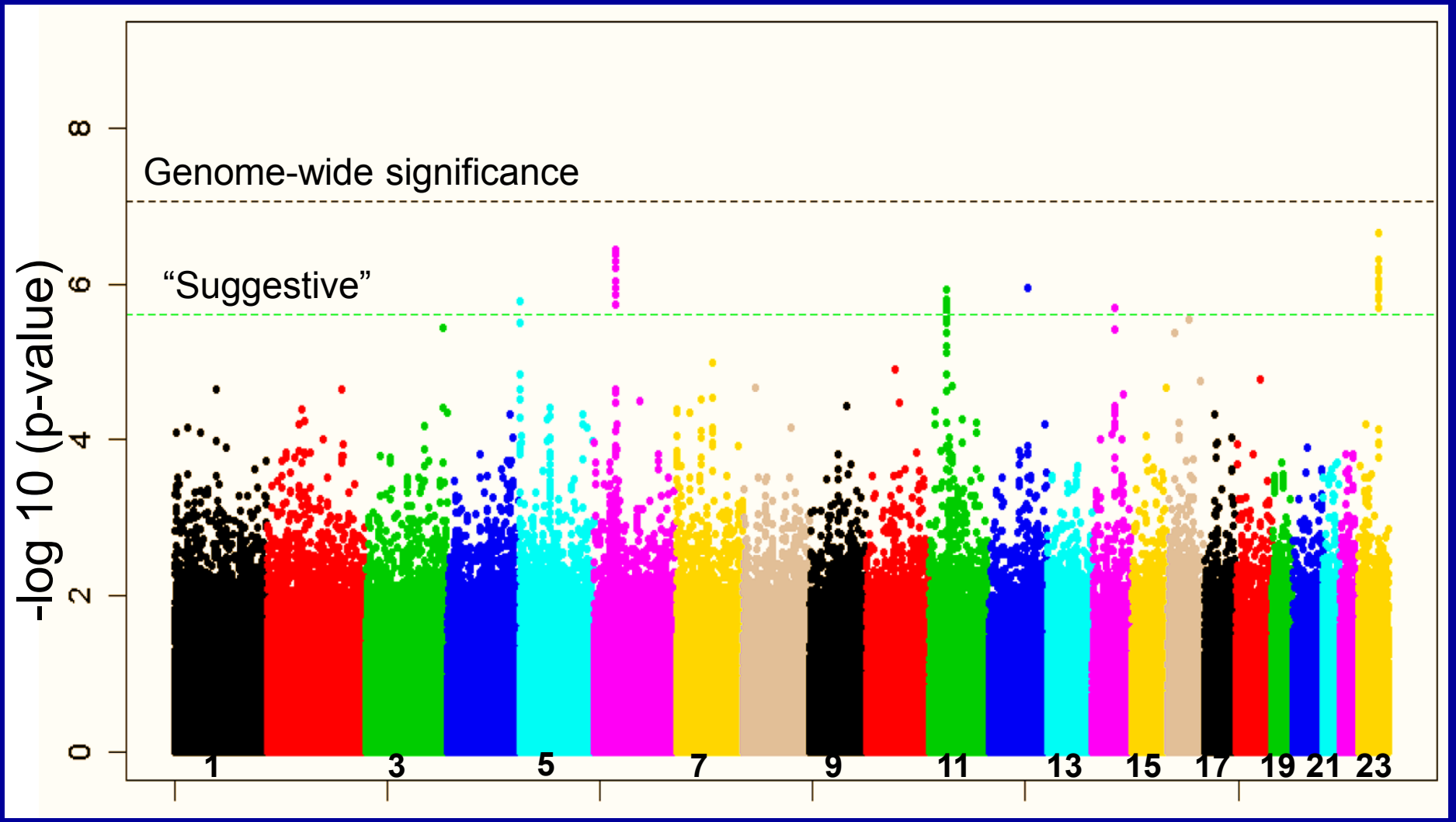
Three Different Study Designs and Populations

1. Twin/Sibs (Hopkins) – Family-based linkage analysis (plus association)
   - Tests inheritance of variants (SNPs) in con- and discordant sibs
   - Linkage has more power to identify less common (rare) variants
2. Canadian CF patients (Toronto) – Association Analysis
   - Representative sample (70%); mixed CFTR ("PI") genotypes
   - More power to identify common (>5%) variants
3. Extremes-of-Phenotype (UNC/CWRU) – Association Analysis
   - More power in ΔF/ΔF extremes of phenotype ("severe" and "mild")
   - More power to identify common (>5%) variants

Approach

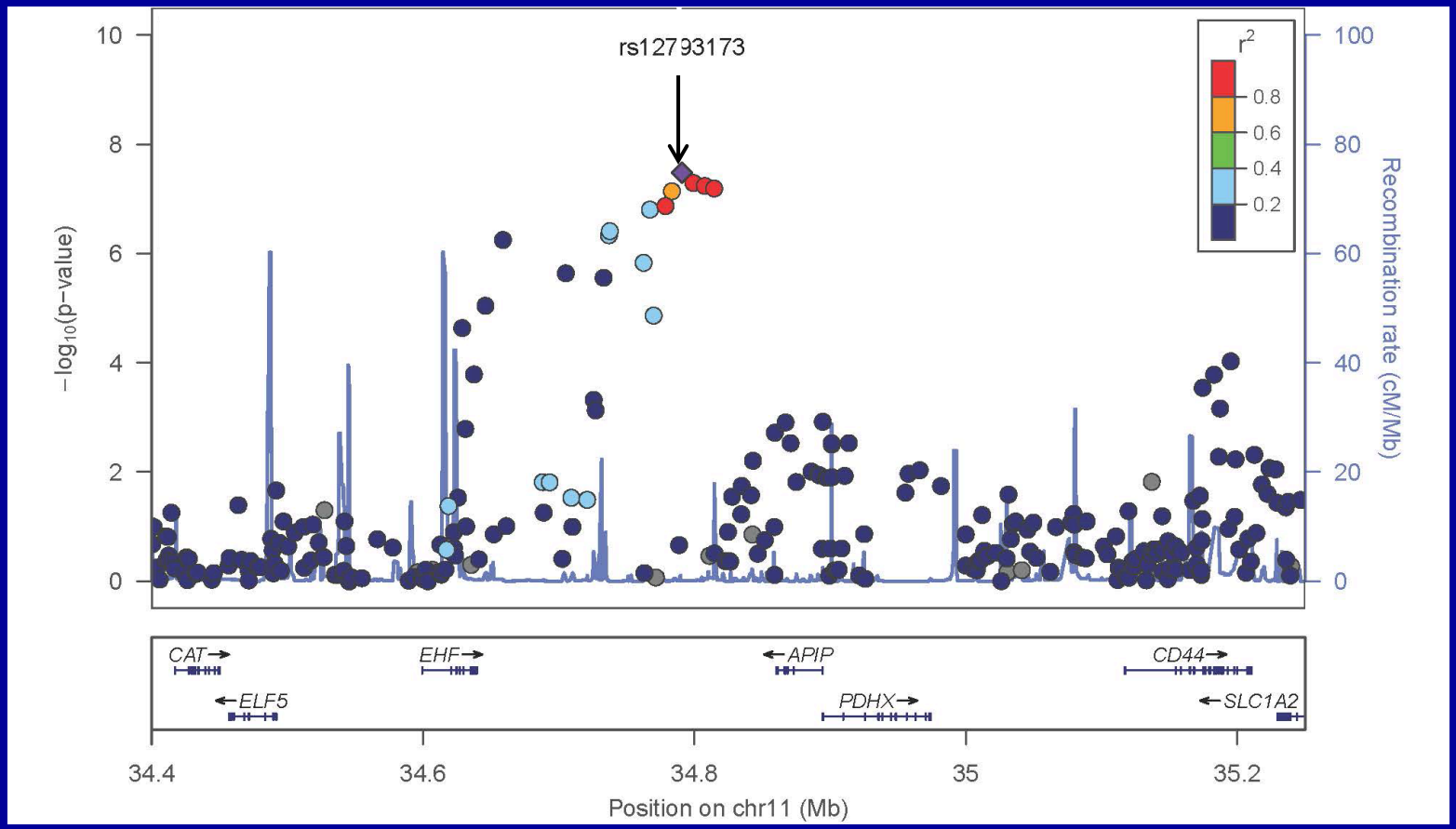
1. Perform "silo" analysis of each study (could use different phenotypes)
2. Perform joint analyses; requires common lung phenotype
   - Linkage (Hopkins) + Association (UNC/CWRU + Toronto)
   - "Meta-Analysis"; control for population stratification (PCs)

Manhattan Plot: Association Evidence in GMS + CGS "All" (PI) Patients



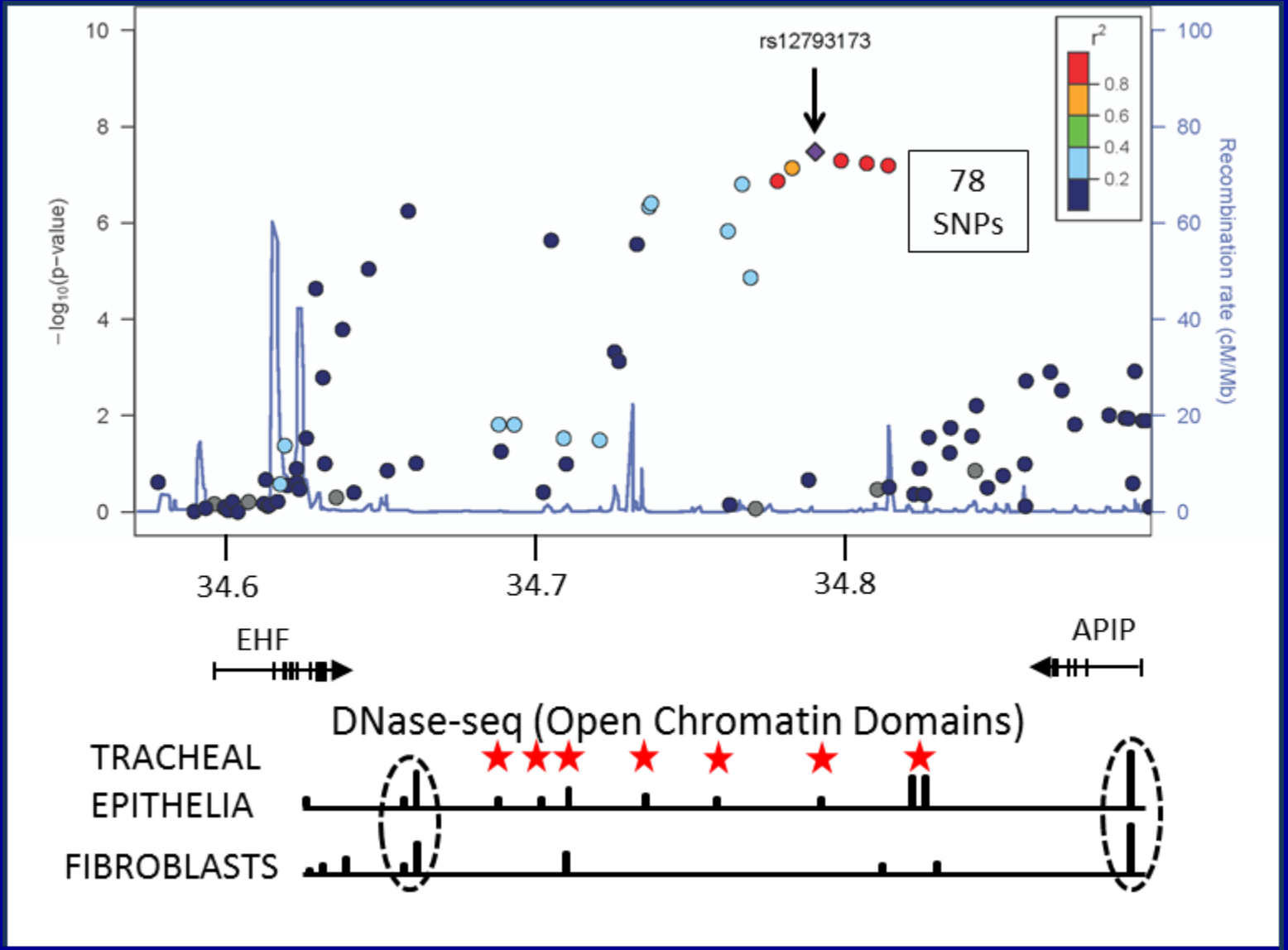
Colors denote Chromosomes 1-23 (571,000 SNPS)

Association Evidence in GMS+CGS F508del/F508del in the Chromosome 11 p13 EHF/APIP Region*

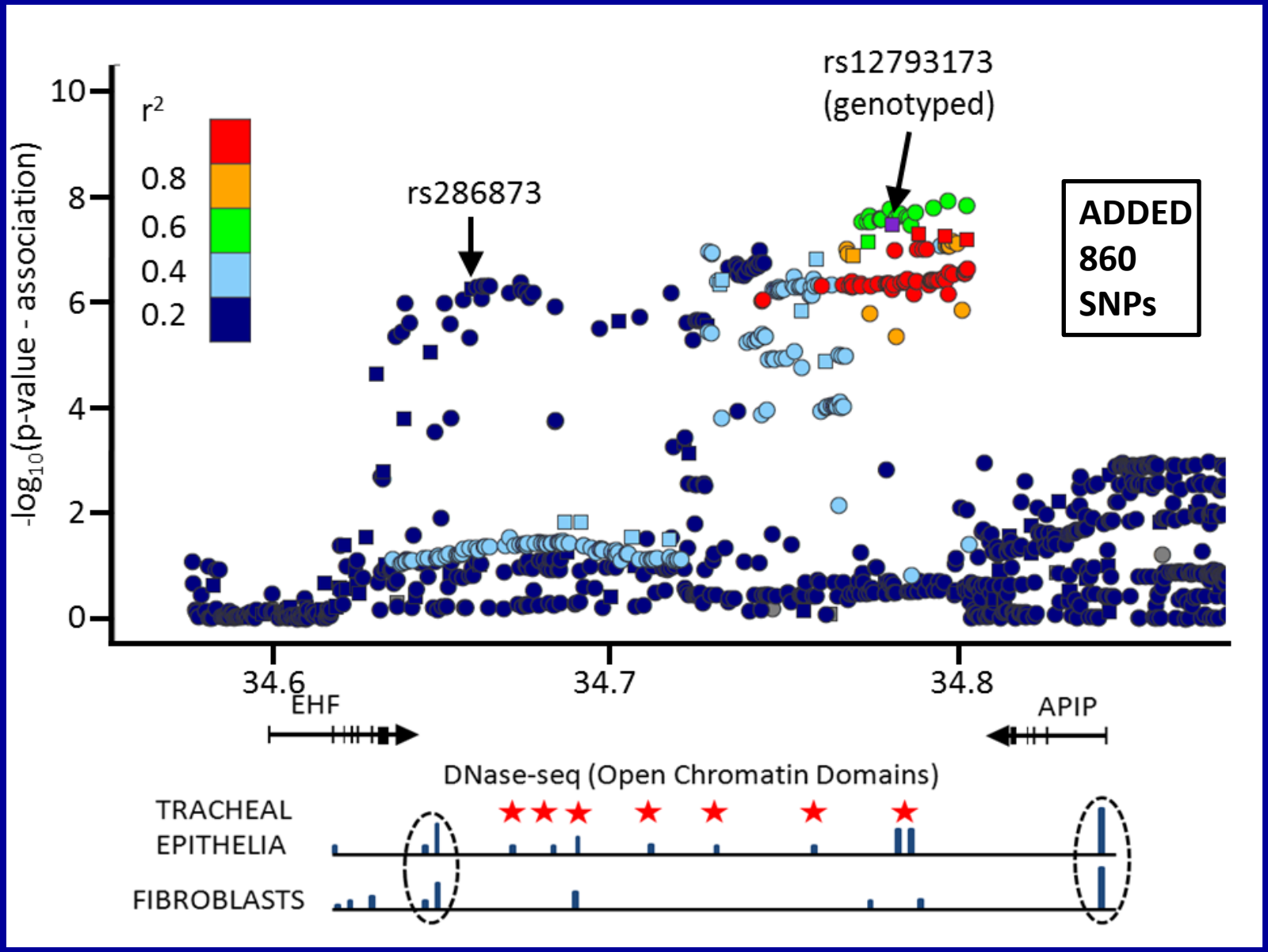


*Replicated in Hopkins TSS; Wright, F., et al., Nat. Genet. 2011

Association Evidence in GMS+CGS F508del/F508del in the Chromosome 11 p13 EHF/APIP Region



Chr11p13: Imputed SNPs (1000G)

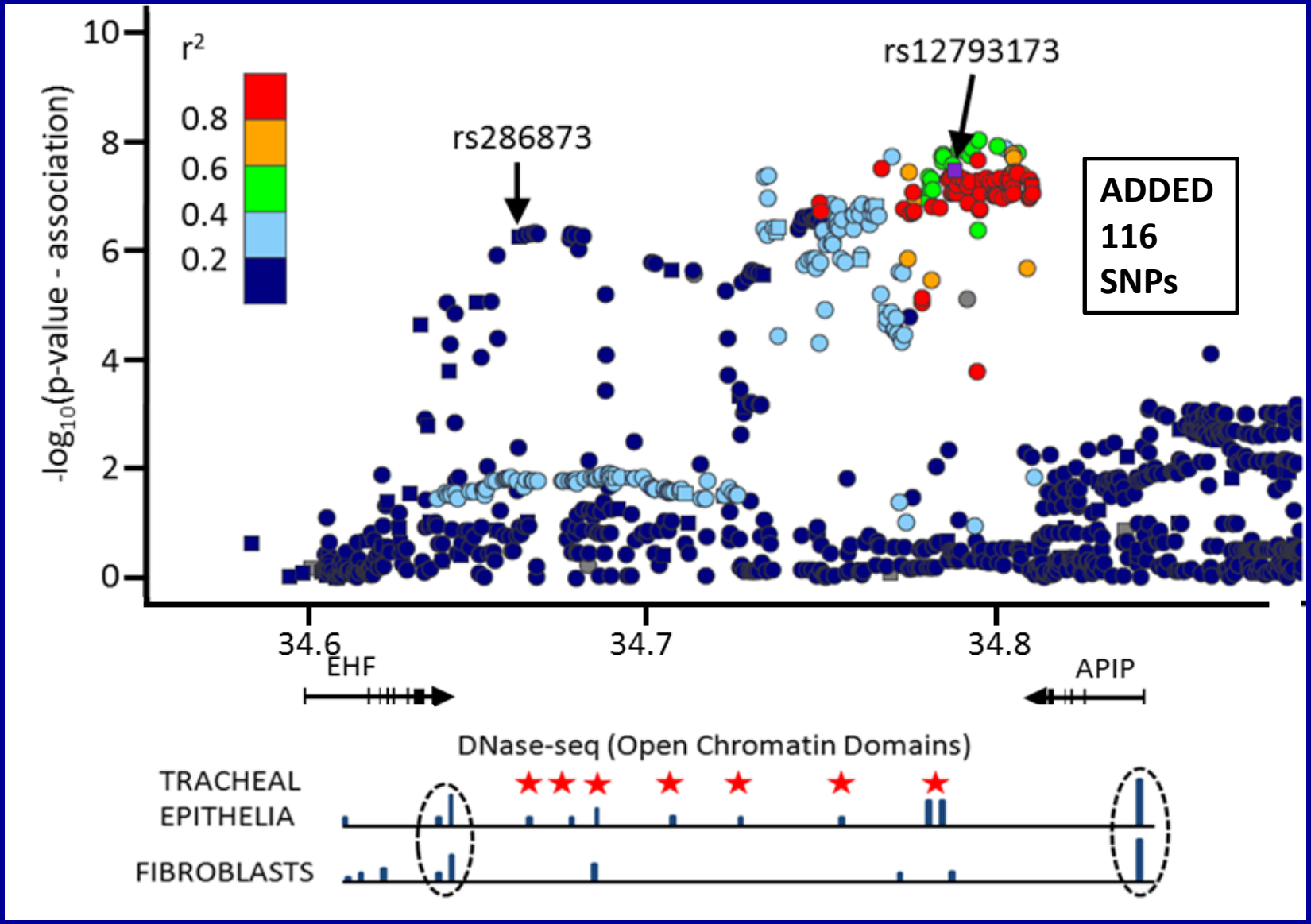


Selection of Patients for RS&G Sequence Across Chr11p13 EHF/APIP Region

Stratified 370 patients by:

- Gender (female/male)
- Extremes of extremes (severe/mild)
- Key genotypes (risk/non-risk)

Chr11p13: Imputed SNPs (RS&G Sequence 370 Patients)

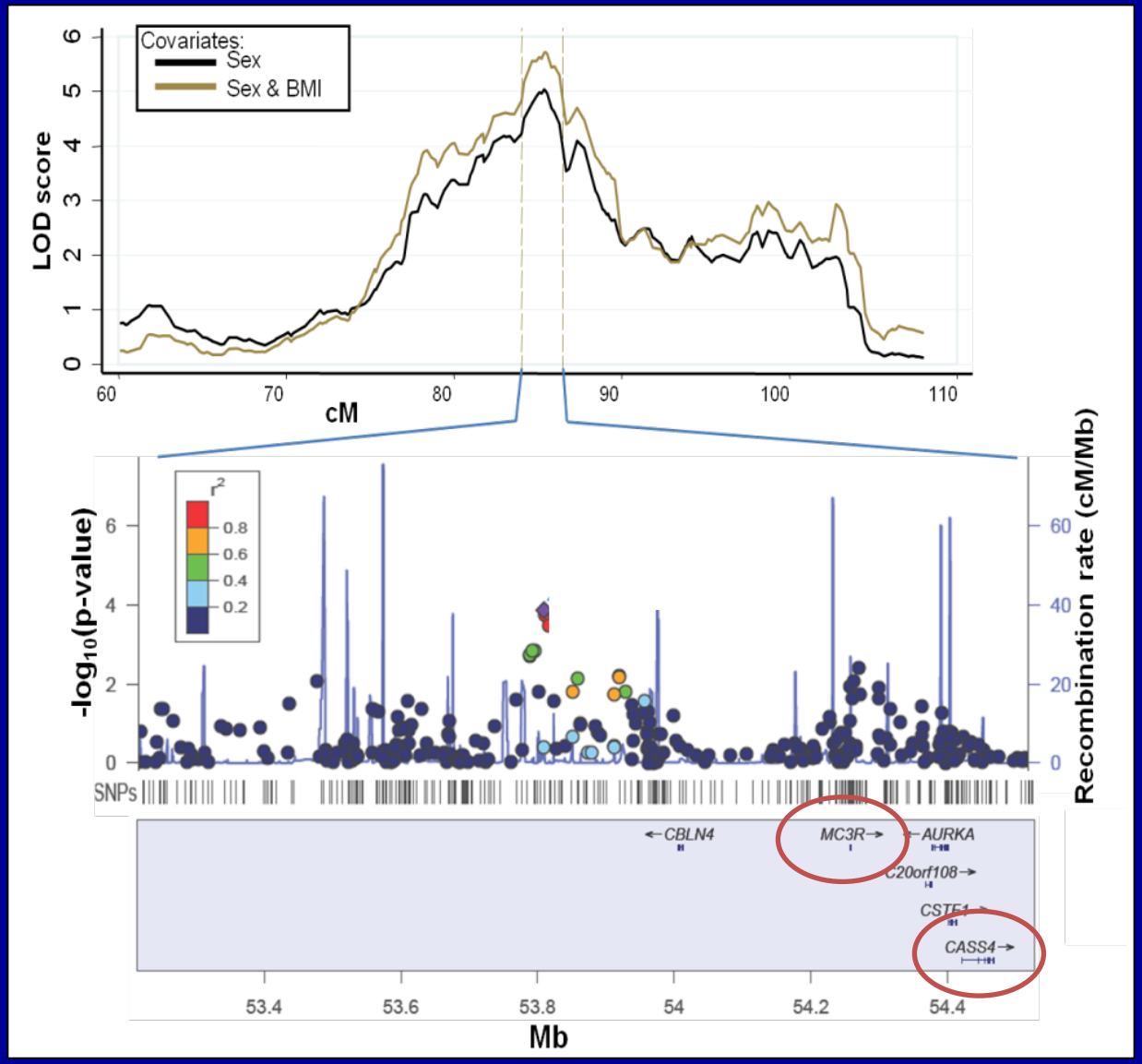


Studies to Link Variation at Chr11p13 to CF Lung Disease Severity*

1. Identify regulatory elements and cis-interaction with promotors of genes in region (chr. conformation capture; 3C).
2. Determine function of regulatory elements to control expression of genes in region (enhancers; repressors; insulators), and how SNPs may alter their properties.
3. Identify targets in human airway cells (CHiPseq), and define mechanisms of cellular pathobiology reflecting altered regulation/function of genes in region (knockdown;CHiPseq; RNAseq).

*Collaboration: Ann Harris, Ph.D.; Northwestern

Linkage at Chr20q13 for TSS, and Regional Analysis of QTL in GMS/CGS



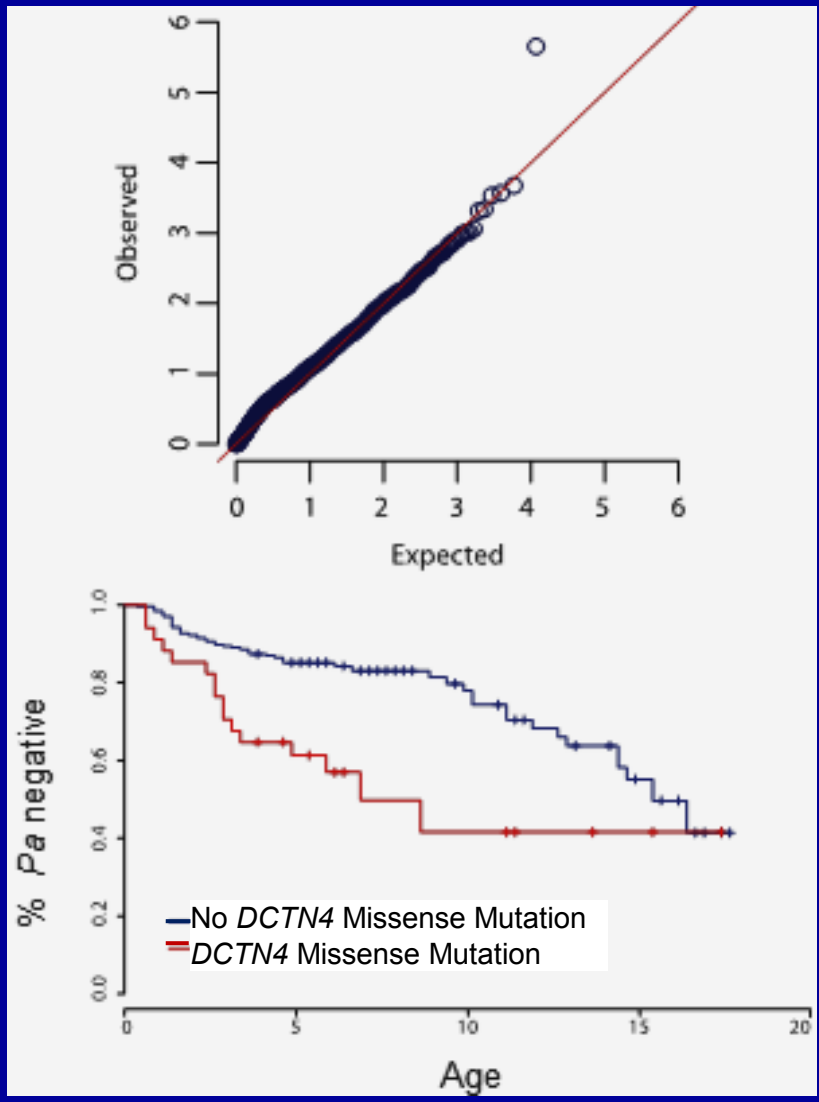
Wright, F., et al., Nat Genet. 2011

Exome Sequence: *DCTN4* is Modifier of CF

- early *Pseudomonas* Infection Control (EPIC) and UNC/CWRU cohorts
- phenotype: time to chronic *Pa* infection

- discard sites with observed MAF > 0.125 (cases and controls combined – use high threshold to allow for enrichment and small sample size)
- collapse by gene: count any variant at any ns site in a gene; allow one count per individual (present/absent; counts independent then)
- 11,542 genes testable
- n = 91 exomes
- apply two sample test of Morris and Zeginni

- validation
- Sanger sequence *DCTN4* in 696 EPIC cases
- time to chronic *Pa*
- Cox regression analysis stratified on enrollment age
- p < 0.004; HR 1.9 (1.2-2.9)



Emond et al. *Nature Genetics* (*in press*)

Modifiers of Other non-Lung Phenotypes in CF (GWAS)

1. CF-Related Diabetes Mellitus: 5-25%, depending on age and criteria* (S. Blackman, Hopkins; Consortium)
2. CF Liver Disease: 5% develop “severe” liver disease, with portal hypertension† (J. Stonebraker, UNC/CWRU; Consortium, plus International)
3. Meconium Ileus: 20% have MI (surgery or medical treatment)‡ (Strug, L., Toronto; Consortium)

The infrequency and/or complexity of these phenotypes requires collaborative interactions. GWAS analyses are ongoing. Will be important to add more patients.

*Blackman, S. et al., Abstr. #161 (pg. 267, Pediatr. Pulmonol. Suppl. 32, 2009)
†Bartlett, J. et al., JAMA 2009
‡Dorfman, R. et al., Hum. Genet. 2009
‡Henderson, L.B., et al., PLoS Genet., 2012
‡Sun, L. et. al., Nat. Gen. 2012

Conclusions

1. Sequencing cohorts aiding discovery of modifiers in CF (exomes) and defining mechanisms of variants in non-coding regions to modify lung disease.

2. More opportunities soon available
   a) Expect more regions to be defined for lung disease in GWAS2 (~3,500 additional CF patients)
   b) GWAS data available in other non-lung heritable phenotypes in CF (sweat CI; meconium ileus CF-related diabetes; CF liver disease)

Conclusions

1. Sequencing cohorts aiding discovery of modifiers in CF (exomes) and defining mechanisms of variants in non-coding regions to modify lung disease.
2. More opportunities soon available
   a) Expect more regions to be defined for lung disease in GWAS2 (~3,500 additional CF patients)
   b) GWAS data available in other non-lung heritable phenotypes in CF (sweat Cl; meconium ileus CF-related diabetes; CF liver disease)
3. Limitation: Calls for indels and CNVs

Acknowledgements

Participating sites, including Consortium and international sites.

Members of Consortium:

Hopkins: G. Cutting, S. Blackman, J. Blangero (SFBR), L. Bremer, M. Collaco, V. Doshi, D. Green, A. Hamosh, L. Henderson, T. Lai, K. McDougal, R. McWilliams, K. Haughton, L. Vanscoy

Toronto: P. Durie, J. Rommens, L. Strug, M. Corey, A. Cojocaru, R. Dorfman, L. Li, P. Pare, A. Sandford, L. Sun, C. Taylor,

UNC/CWRU: M. Knowles, M. Drumm, F. Wright, W. O'Neal, J. Dunn, E. Lange, G. Mayhew, R. Pace, J. Stonebraker, F. Zou

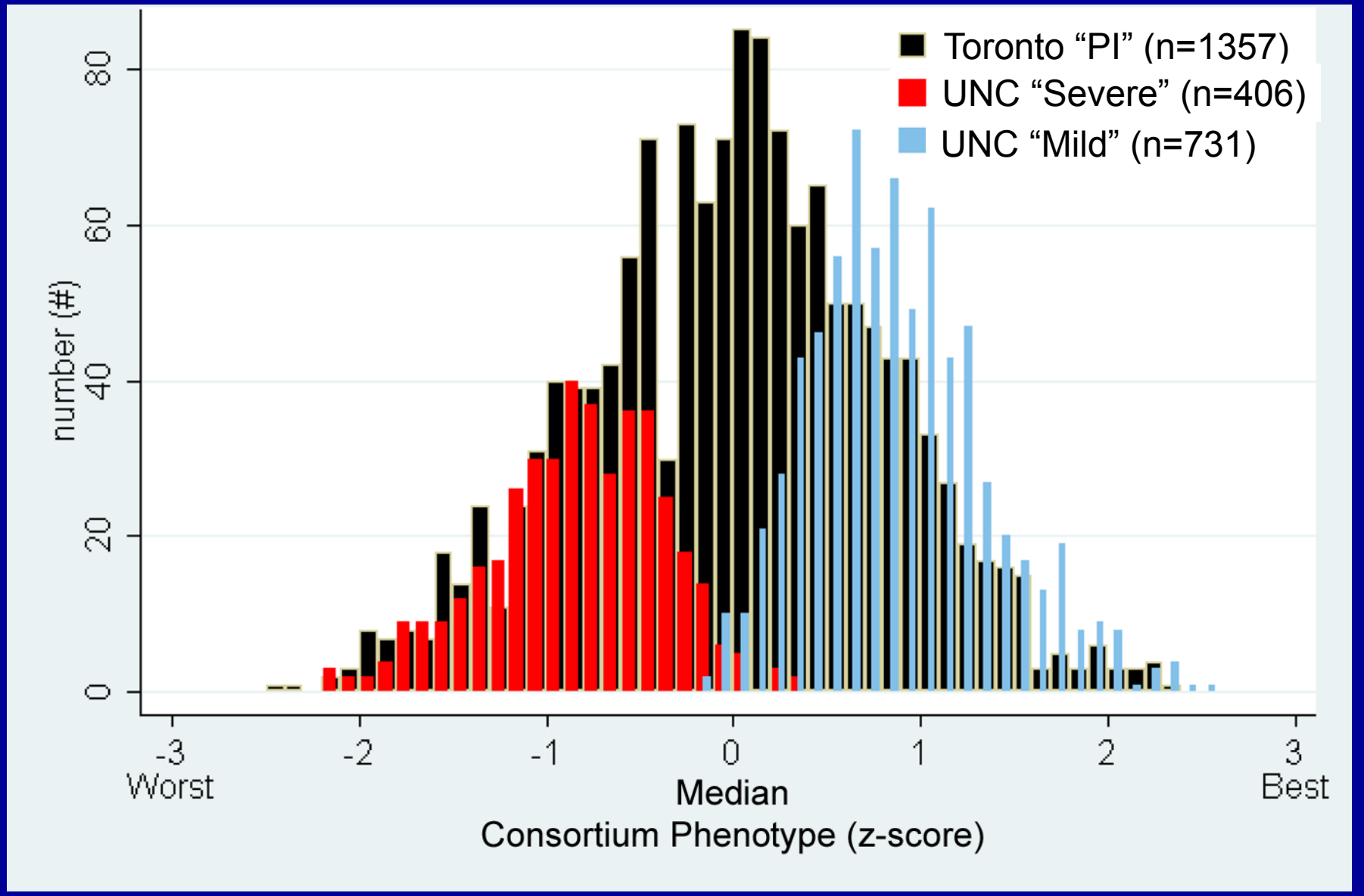
U.S. CF Foundation (C. Penland, Ph.D.)

RS&G Service (D. Nickerson): Northwest Genomics Center; Univ. of Washington, Department of Genome Sciences, NHLBI Contract Number HHSN268201100037C

U.S. CFF (Cutting 07PO; DRUMM 06PO; Knowles 07PO; Consortium GWAS grant); NIH (HL68890; HL68927; DK066368); RS&G (N01-HV-48194); Genome Canada (OGI-3-05); and Canadian CFF.

cfmod@med.unc.edu

Quantitative Lung Phenotype: Distribution



Predicted Effect of Modifier Loci: EHF-APIP region

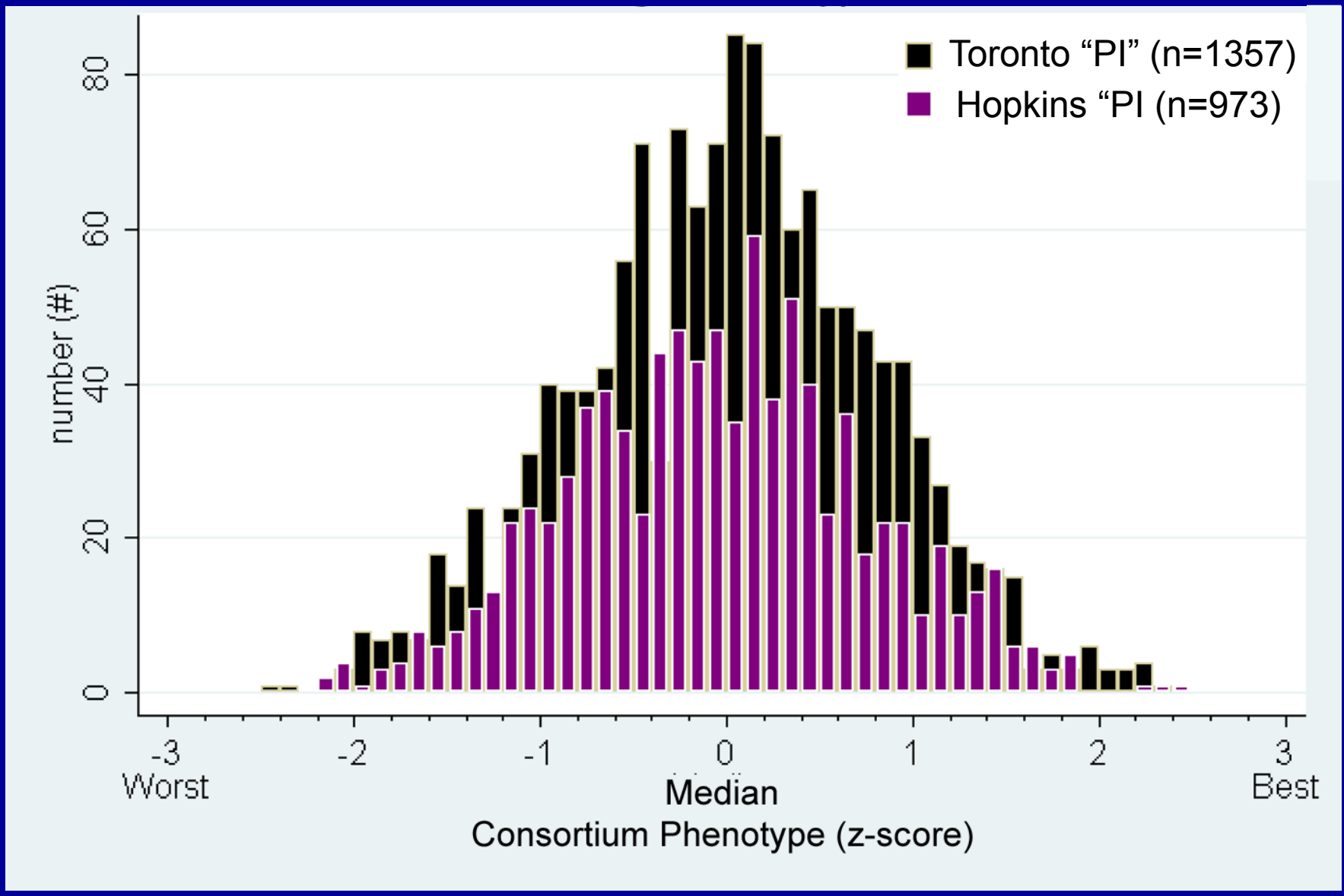
Per allele:

- change of 0.2 units of the quantitative lung disease phenotype
- equivalent to 5.1 ± 1.9% change in the FEV_1 (% Pred.), when calculated as recommended by the U.S. Cystic Fibrosis Foundation
- or, change of 254 ± 86mL in subjects older than 18 years.

What is Needed?

- Sequencing of many patients, high depth, select regions, low cost
  - Increase number of individual sequencing tags per run
- High quality variants files
  - 1000 genomes, Complete Genomics
- Improved filtering for loss-of-function variants
  - MacArthur DG et al Science 2012

Quantitative Lung Phenotype: Distribution



APIP/EHF

APiP (Apaf-1-interacting protein) – inhibits mitochondrial apoptosis

a) Binds to Apaf-1 (result – inhibition of caspase-9)
b) During hypoxia, inhibits caspase-9 (through AKT&ERK1/2)

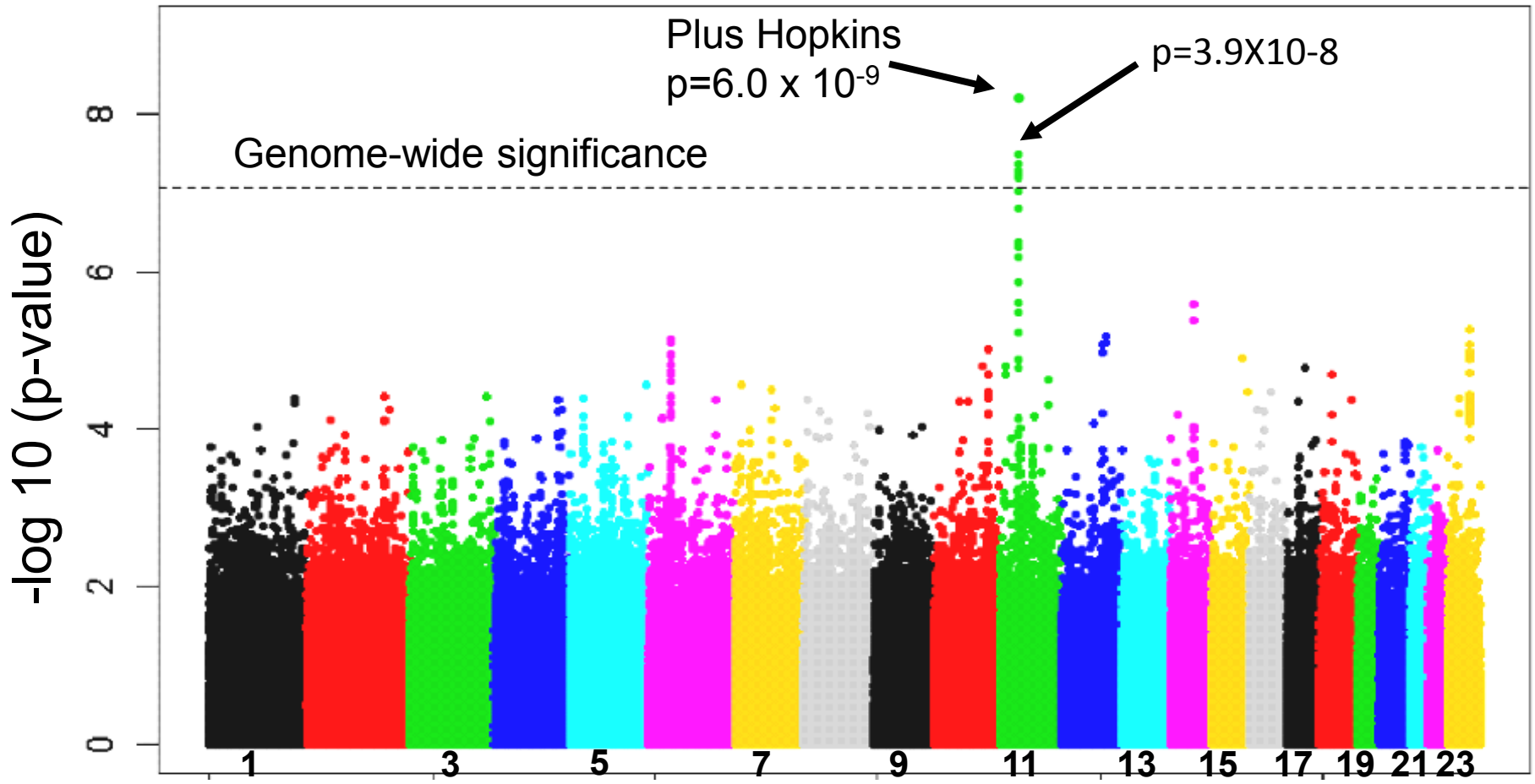
EHF (ESE-3) – Epithelial-Specific Ets transcription factor

a) 3 ESE human genes: organs with differentiated epithelia
b) In lung, regulates differentiation/inflammation under stress (MAP kinases)
c) Transcriptional repressor of ETS/AP-1-responsive genes*
d) Activity induced in airway cells by IL-1b & TNFα*

These genes may play a role in the degree of CF airway inflammation.

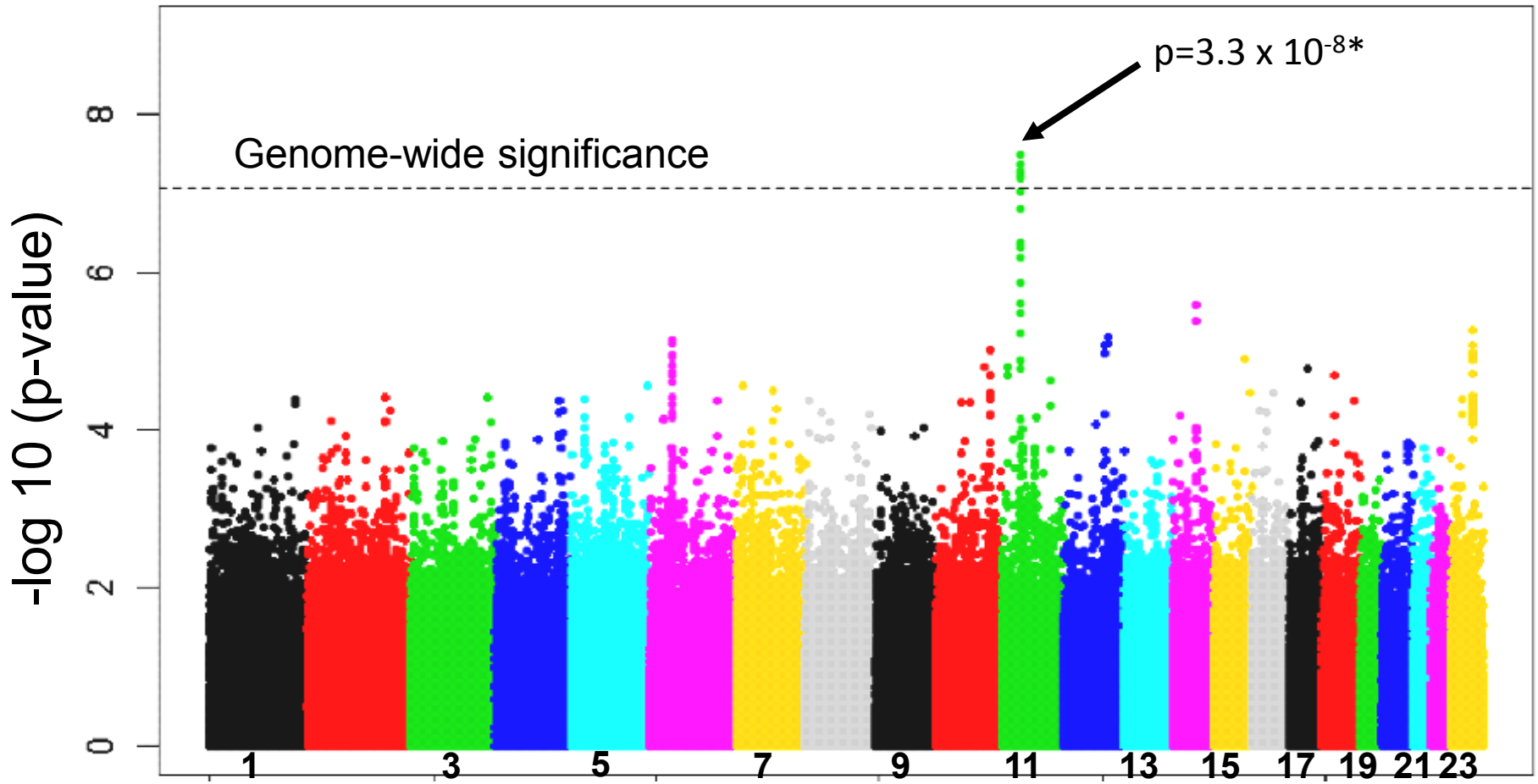
*Silverman Eric et al., 2002, AJRCMB; Wu, Jing et al., 2008, Cell Res.

Combined Association Analysis UNC/CWRU (ΔF/ΔF) + Toronto (ΔF/ΔF)



Colors denote Chromosomes 1-23 (571,000 SNPS)

Association Analysis
GMS (ΔF/ΔF) + CGS (ΔF/ΔF) (n=2,198)

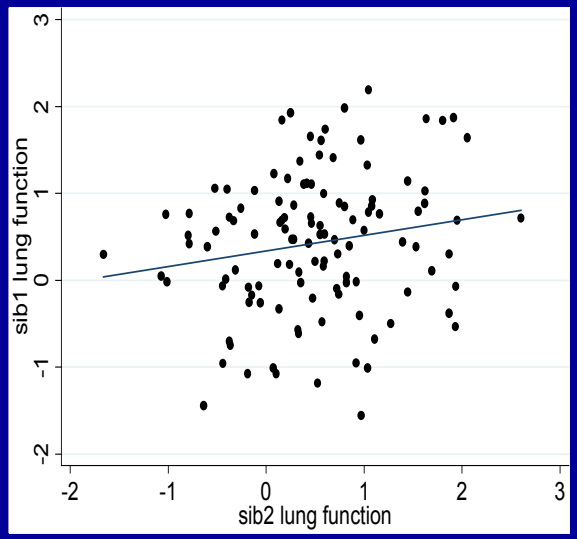


*Joint analysis; GMS + CGS + TSS; p = 1 × 10^{-9}

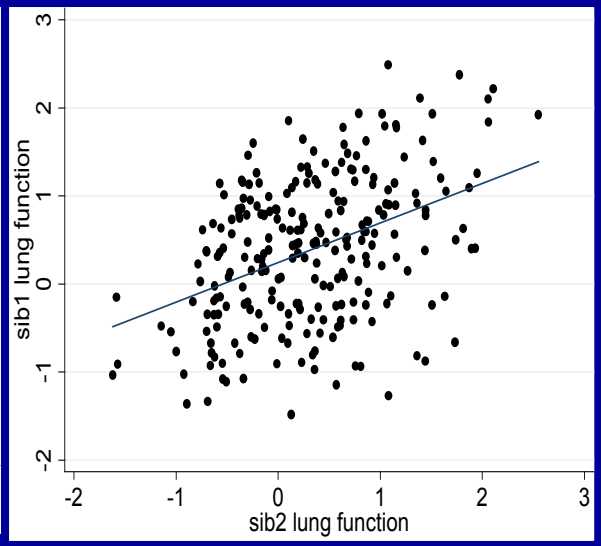
Colors denote Chromosomes 1-23 (571,000 SNPS)

Predicted Effect of Modifier Loci: Chromosome 20 linkage region

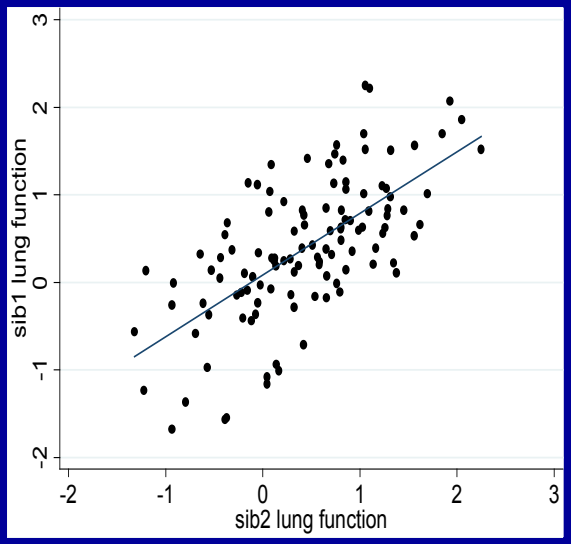
IBD0



IBD1

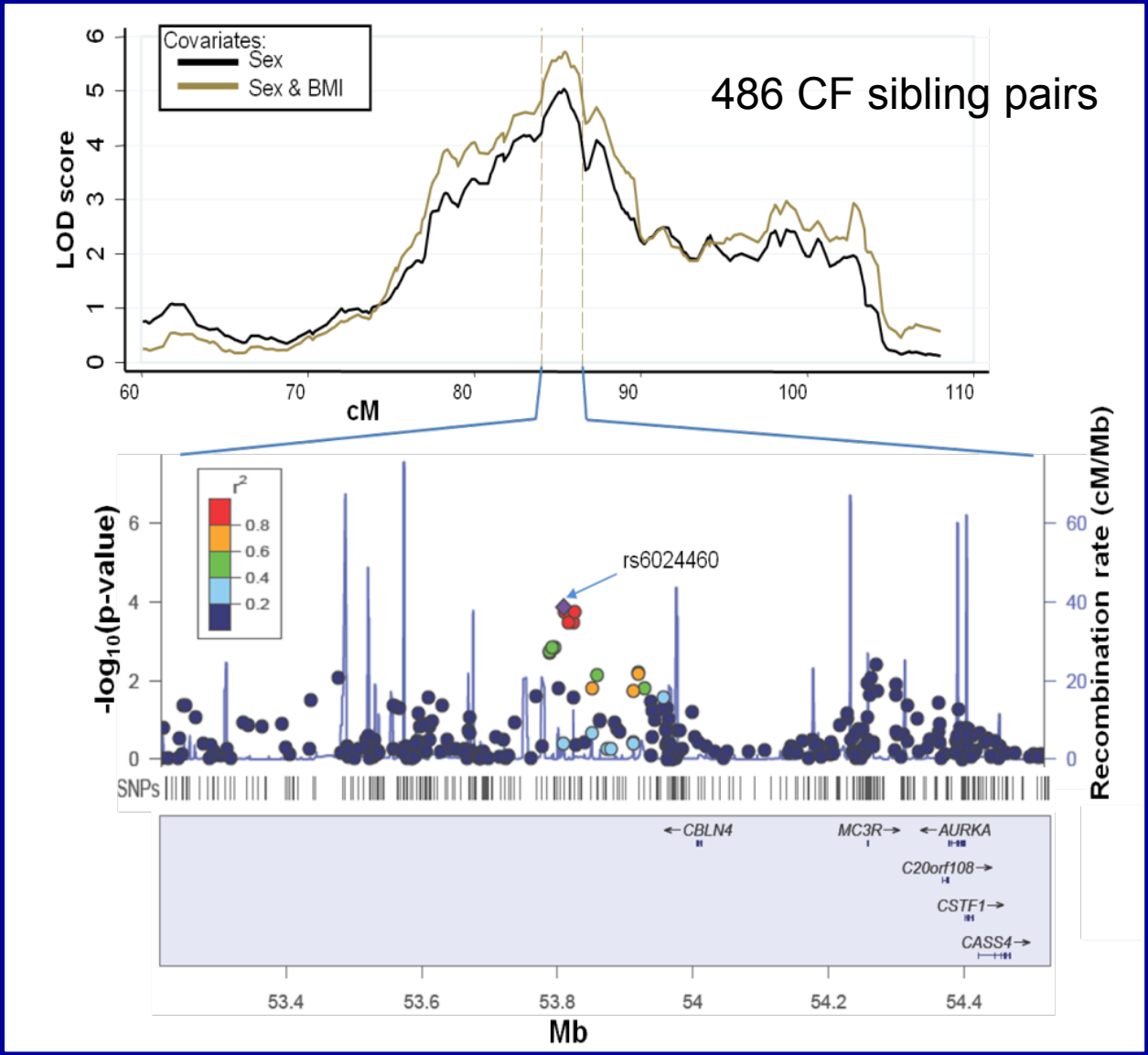


IBD2



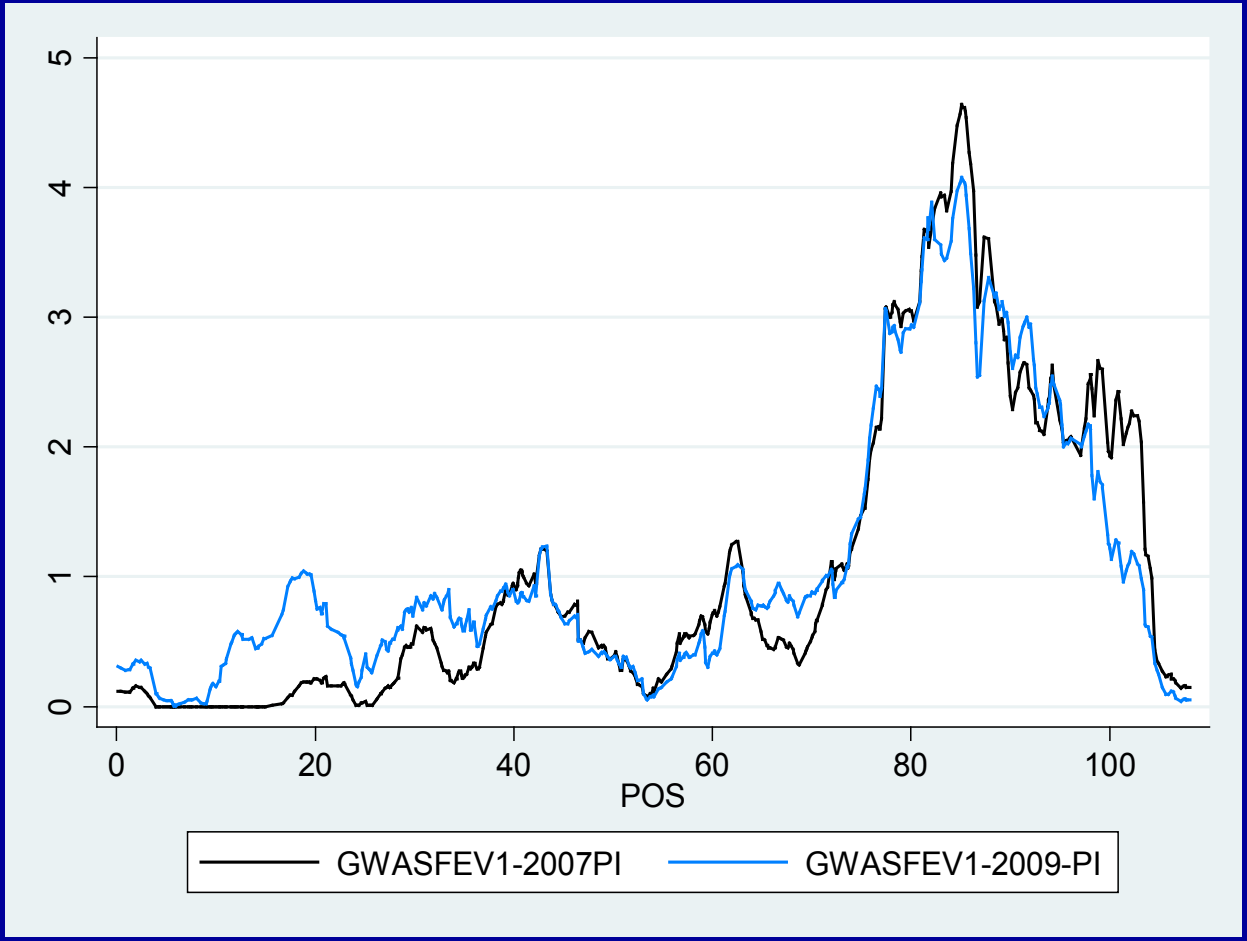
Correlation by IBD status: IBD0: r=0.1762; IBD1: r= 0.4136 and IBD2: r= 0.6542.
The contribution of the linked region to variation was estimated by subtracting the correlation in siblings who are IBD 0 from the correlation in sibling who are IBD 2 (0.6542 – 0.1762 = **0.478).**
Similar estimates were obtained using variance components methods in Merlin (**0.499**) and SOLAR (**0.496**).

Chromosome 20q13 Harbors a Modifier of Lung Function in CF (LOD 5.03)



Wright F, et al, Nat Genet 2011

Linkage Signal Remains Robust as Patients Age (2007 to 2009; 402 pairs)

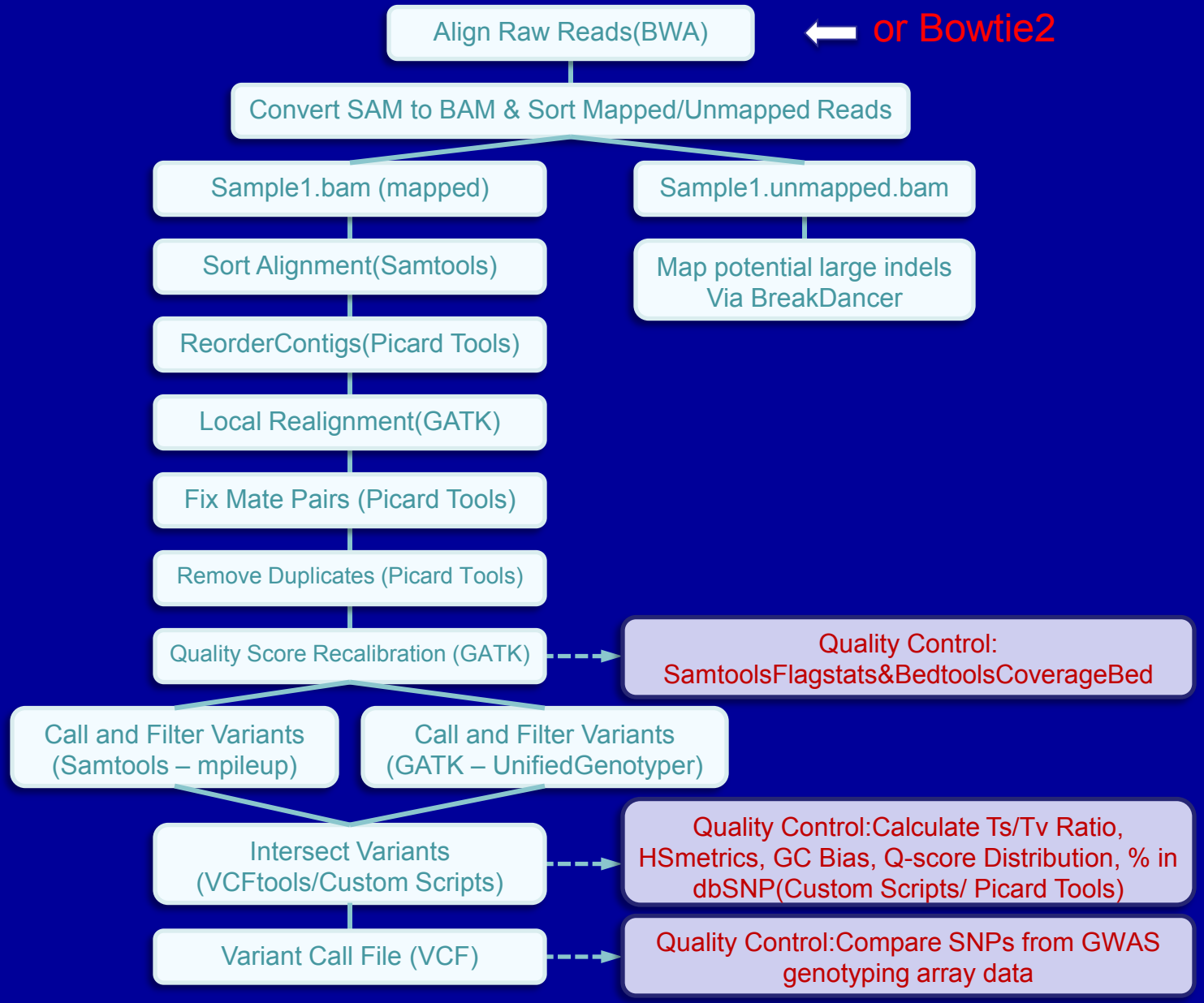


2% have completely non-overlapping data with 486 siblings in original study, 46% have a 12 month or less overlap, and 83% have an 18 month or less overlap

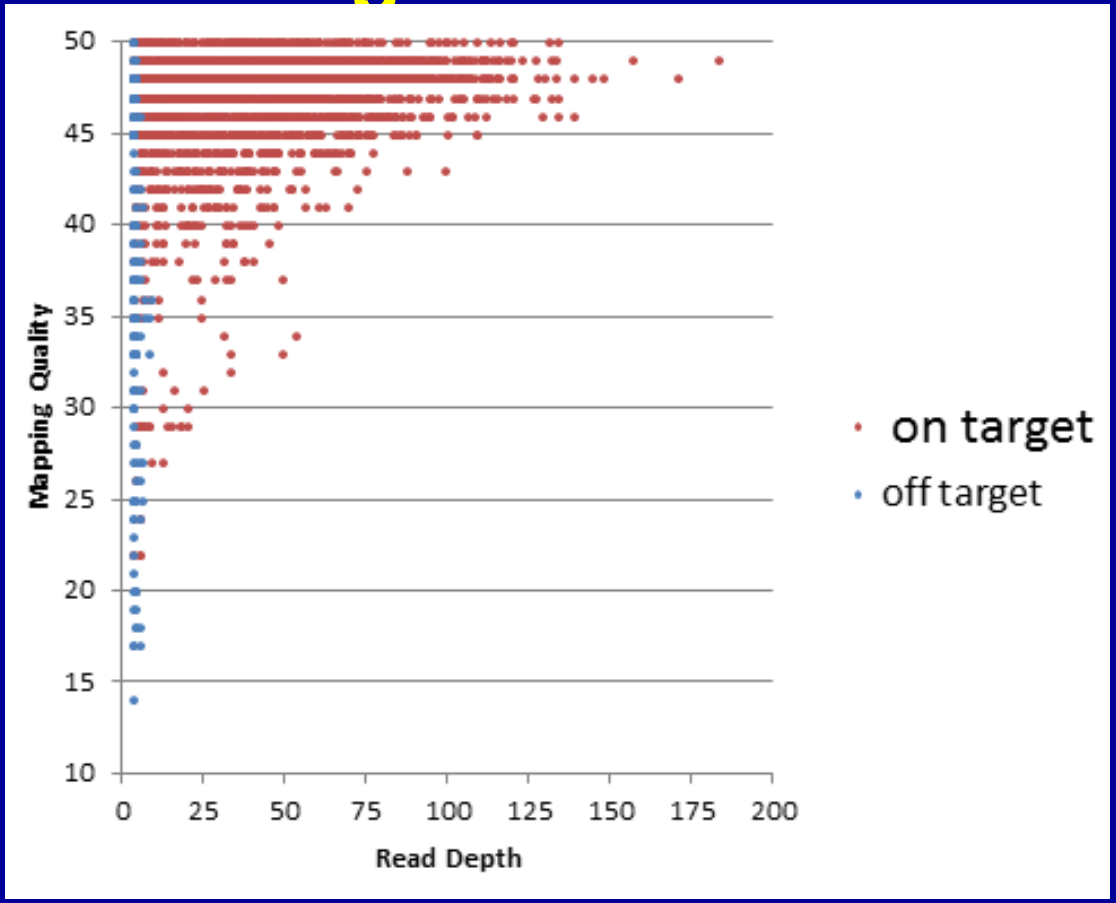
Pilot Sequencing

- Linkage region on chr 20 (1.3Mb)
- Genes (exons/introns/50kb 5'and 3')
  - Cse1L
  - MSRA
  - SLC26A9
- 4 families; 2 CF siblings each with parents; Agilent array and HiSeq Next gen sequencing (Harry Cuppens)
- Analysis and pipelines (Vecchio/Park/Yandell/Moore-VAAST; CIDR; Salzberg group)

Variant Detection Pipeline (Part 1)

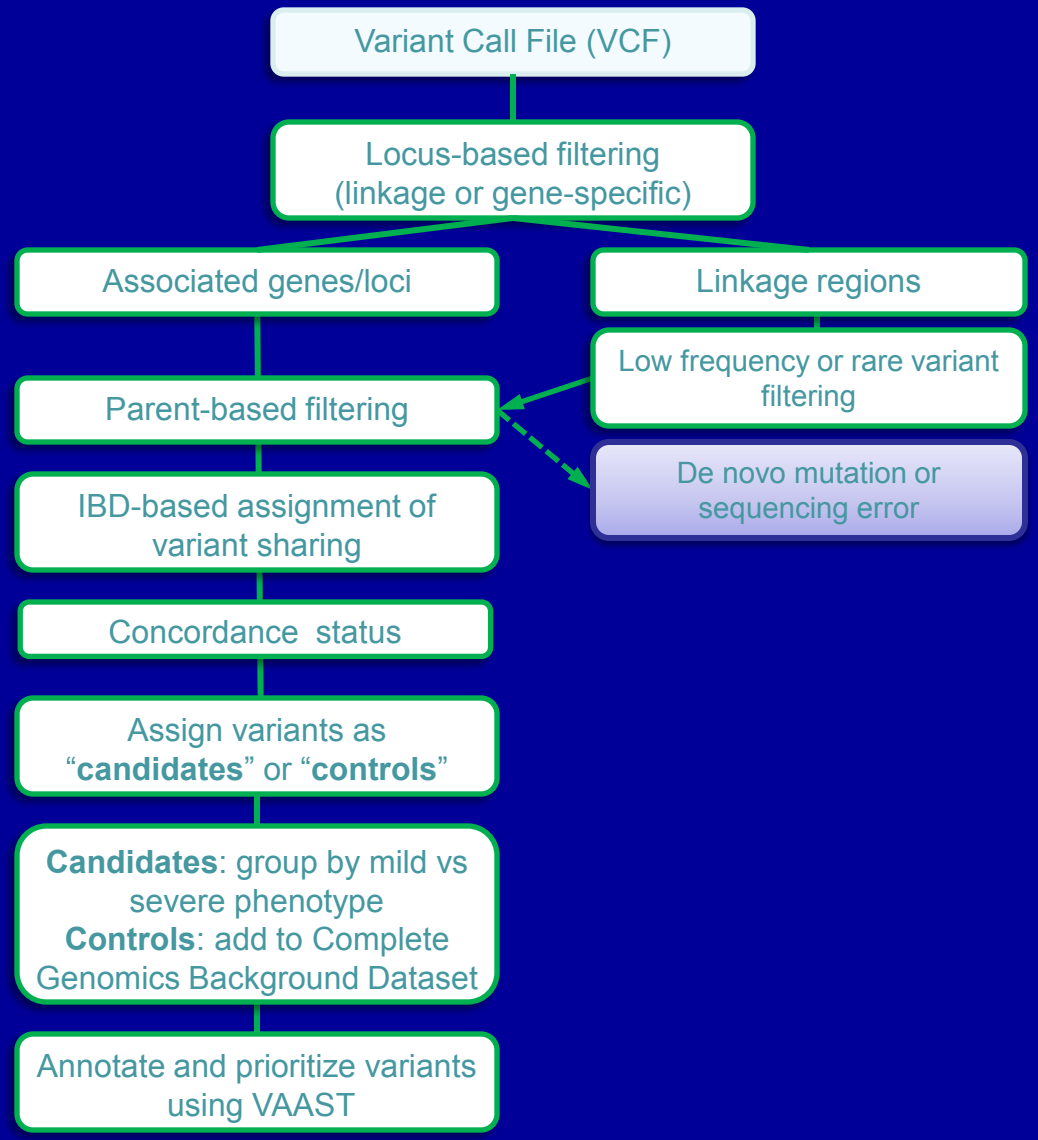


Comparison of Mean Depth and Mapping Quality of On-Target vs. Off-Target Variants for 1.6Mb Region within Chr20q13.2



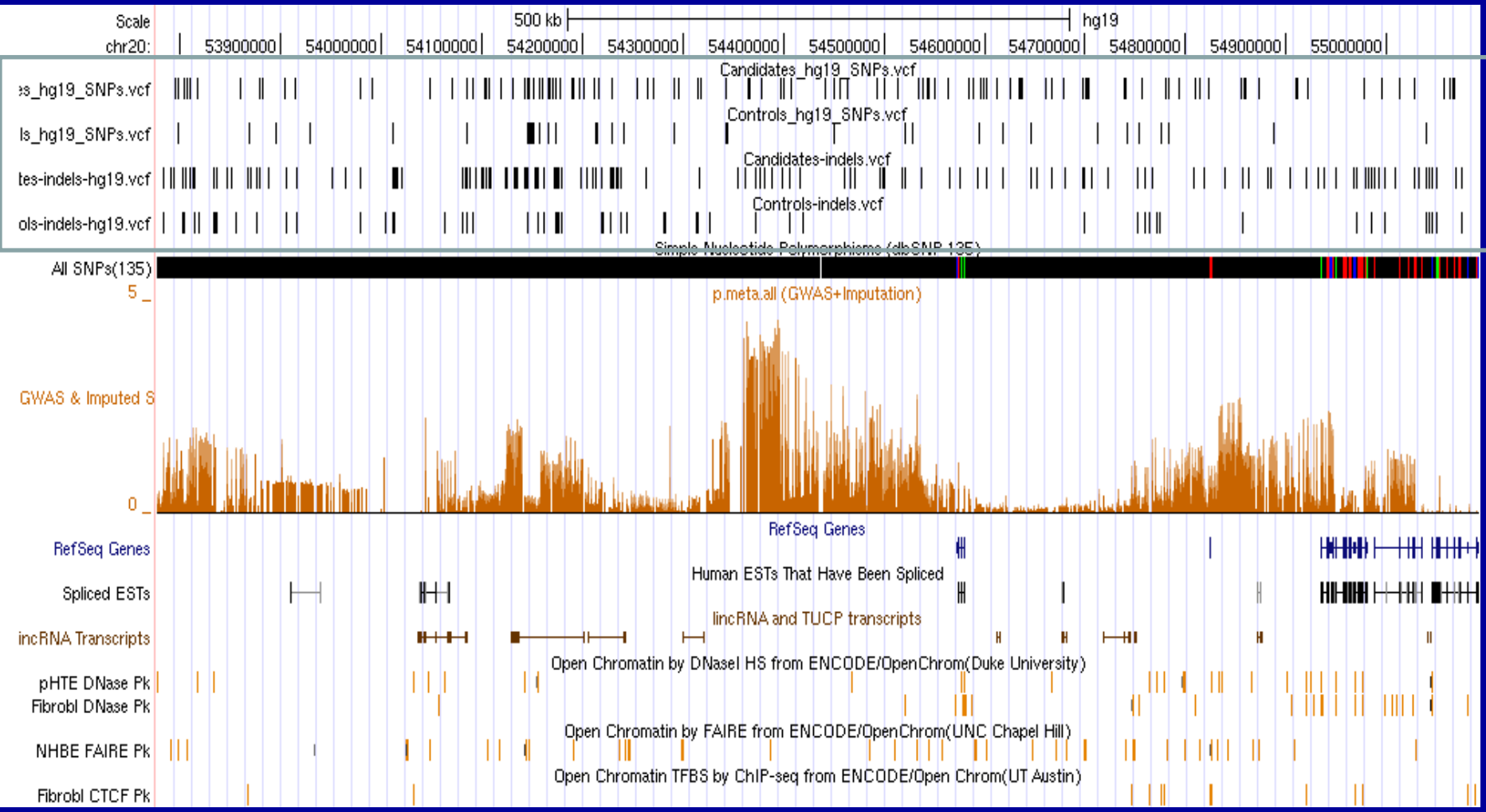
The mean depth of on-target variants found on chr20 was significantly higher (34.9x and 17.9x) compared to that of off-target variants (3.5x and 3.7x) and mean mapping quality of on-target variants was higher (46.8 and 47.1) than that of off-target variants (36.6 and 37.1) .

Variant Detection Pipeline (Part 2)



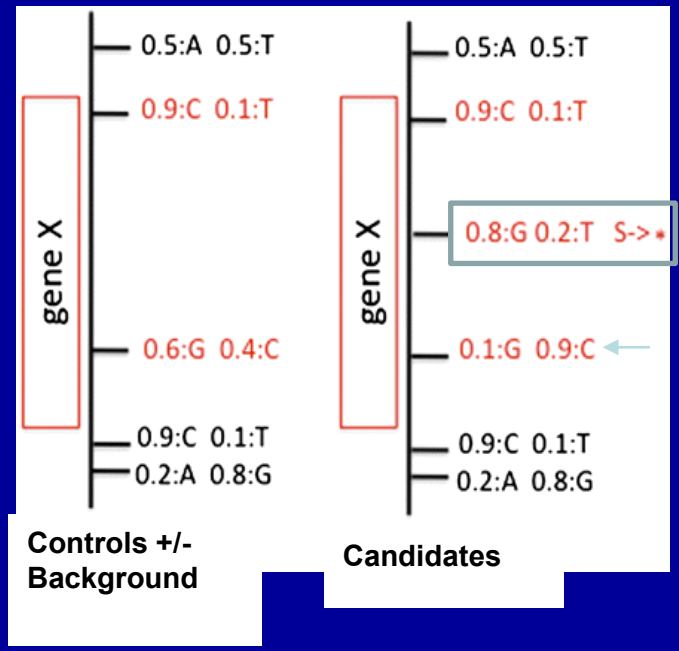
Assigning candidate and control variants		
	Concordant	**Discordant**
Shared variants	Candidates	Controls
Non-shared variants	Controls	Candidates

Variant Mapping to Chr20q13.2 Region on Santa Cruz Browser



Finding modifier variants

Variant Annotation Analysis and Search Tool (VAAST) Yandell et al Genomic Res 20



Variants encompassed by a feature are scored to give a composite likelihood for the observed genotypes at that feature under a healthy and disease model by comparing variant frequencies in the cases (candidate modifiers) compared to control (background).

Features tested will include genes, eQTLs from respiratory epithelial cells, transcription elements found in respiratory tissues and putative functional elements (conserved sequences, DNAse I sites etc.). **Sliding windows will be used to test for aggregation of variants that occur outside of the aforementioned 'functional' features.**