

SUPPLEMENT

X Chromosome Genetic Associations with Chronic Obstructive Pulmonary Disease and Related Phenotypes

Lystra P Hayden, Brian D Hobbs, Robert Busch, Michael H Cho, Ming Liu, Camila M Lopes-Ramos, David A Lomas, Per Bakke, Amund Gulsvik, Edwin K Silverman, James D Crapo, Terri H Beaty, Nan M Laird, Christoph Lange, Dawn L DeMeo; on behalf of the COPDGene and ECLIPSE Investigators.

METHODS

Study phenotype

Subjects were current and former smokers with at least a 10 pack-year smoking history in COPDGene and ECLIPSE, and at least a 2.5 pack-year smoking history in GenKOLS. Subjects completed a modified American Thoracic Society respiratory questionnaire and standardized spirometry, as well as chest CT, and provided blood for DNA extraction and genetic analysis (1, 2).

Quality control

Study based subject-level data cleaning removed subjects with sex discrepancies including chromosomal aneuploidies, and tested for excess or deficiency of heterozygosity (inbreeding coefficients, F-test absolute value > 0.2) (3). Study based variant-level cleaning was conducted separately in male and female datasets. Variants were excluded if they had minor allele frequency $< 1\%$, variant level missingness $> 2\%$, differential missingness between sexes of $\geq 2\%$, or for a significant departure from Hardy-Weinberg equilibrium in females ($p < 1 \times 10^{-4}$). The data cleaned by sex was merged, and only variants passing quality control in both sexes for that study were retained for XWAS.

Imputation

Data from subjects and variants passing quality control were used for imputation of untyped variants based on the Haplotype Reference Consortium reference panel (version r1.1 2016) via the Michigan Imputation Server v1.0.2 (<https://imputationserver.readthedocs.io/en/latest/pipeline/>) (4). Imputed posterior genotype dosage probabilities were filtered for imputation accuracy (INFO-score) and removed for $R^2 < 0.5$ (3).

Significance and suggestive thresholds

The primary threshold for identifying genome-wide significant associations was a p-value for association $p < 5 \times 10^{-8}$ (5, 6). We examined additional suggestive associations, defined as p-values at the level of 10^{-6} or less if associations were present at that level, and at the level to 10^{-5} if no other associations were observed.

Annotation

Annotation of variants was performed using National Center for Biotechnology Information databases dbSNP/Gene, UCSC Genome Browser, Ensembl, LDlink, and LocusZoom (7-11).

RESULTS

Annotation of rs142755000

The sex-stratified analysis by Zhao et al. found that rs142755000 reached genome-wide significance for FEV₁. Zhao found it has the same direction of effect in males and females, but a notably larger effect in males, which was not seen in our current study. We annotated rs142755000 to *HMGN5* and a nearby top suggested variant we identified in *HMGN5*, rs185387095, did have significant sex-differences in this XWAS for FEV₁ with larger effect in males (β males -0.039, β females -0.029, sex-difference p 3.30 x 10⁻⁰²). The closest gene to rs142755000 is *BRWD3*, 149kb upstream. In this study we annotated rs142755000 to *HMGN5*, which is 155kb downstream (Table 2). This annotation to *HMGN5* was made due to the fact that another top suggested association in this study, rs185387095, is found in *HMGN5*, and rs142755000 is in linkage disequilibrium (LD) at R² 0.52 (11). Additionally, there are other LD variants within the same recombination hotspot that includes *HMGN5* and are in and near rs142755000 (supplement Figure 3).

DISCUSSION

Suggestive associations

In COPD there was a suggestive association in *SH3KBP1*, a gene that escapes XCI, encoding a protein that facilitates protein-protein interactions and has been implicated in cellular processes including cell adhesion, cytoskeletal rearrangement, apoptosis, endocytosis, and it has been shown to play a role in maturation of alveolar epithelial cells and surfactant production in mice (7, 12-14). *SH3KBP1* is expressed in lung tissue and exhibits sex-biased gene expression in whole blood (15, 16).

In lung function XWAS, variants in Xp11.21 were implicated including the dense region in *FOXR2*/near *RRAGB*/near *PAGE5*. *FOXR2* is a member of the FOX superfamily of genes known to interact with β -catenin and to play a role in epithelial-mesenchymal transition of cancer cells, and in non-small cell lung cancer (NSCLC). *FOXR2* expression inactivates the Wnt/ β -catenin pathway (17, 18). *RRAGB*, a gene expressed in lung tissue, is part of a large family of signal transducers, and high expression of *RRAGB* has been found to predict good survival in NSCLC (7, 16, 19). *PAGE5* is part of family of proteins expressed in some fetal tissues as well as in a variety of tumors, and it encodes a protein that may protect cells from programmed cell death (20).

Among top suggestive associations in females were a number of genes that are expressed in lung tissue and have interesting implications in disease pathobiology (16). *LINC0259* is a long noncoding RNA that mediates TGF- β signaling and has been implicated in lung cancer cell migration and invasion (16, 21). *TMEM47* encodes a highly conserved protein that is a member of the PMP22/EMP/claudin family important in cell morphology that is involved in localization of tight junction proteins and actomyosin structure (7, 22, 23). *ITM2A* has been implicated in ankylosing spondylitis where it is differentially expressed by CD4⁺ T cells and is involved in T cell

activation (24, 25). *TAB3* encodes a protein that functions in the NF- κ B signaling pathway, which plays a role in response to pro-inflammatory cytokines TNF or IL-1. It has been reported to be involved in signaling events in pathogenesis or progression of idiopathic pulmonary fibrosis, where it is thought to alter immune response, tissue repair, and fibrosis (26).

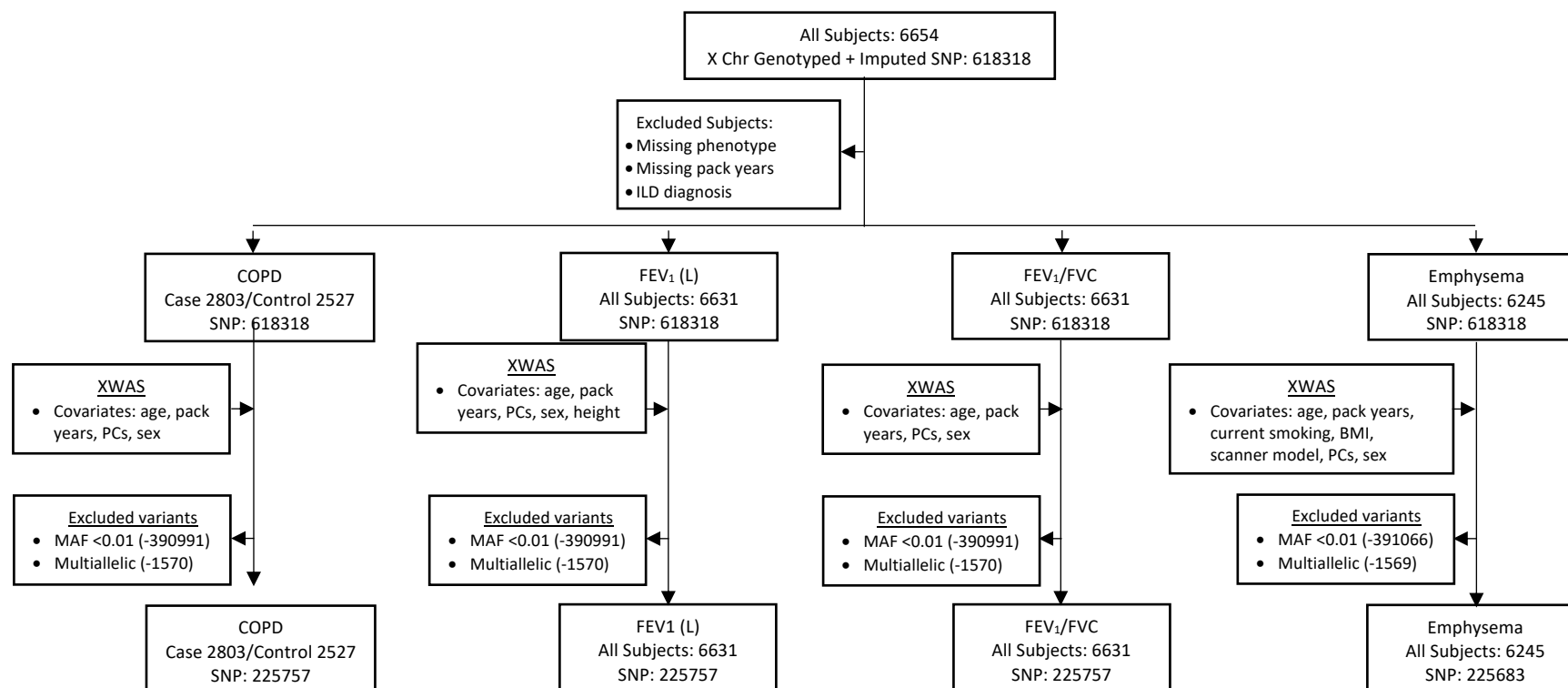
DMD, a top suggestive association for COPD for males as well as for FEV₁/FVC for females, encodes dystrophin and is the largest gene identified in humans; it has been associated with an extensive number of traits (7, 13, 14, 27). *DMD* is known to escape X chromosome inactivation, demonstrate sex-biased gene expression in lung tissue with a female bias, and have female-biased expression in numerous other tissues (13, 15, 28). Dystrophin is present in cardiac and cytoskeletal muscle and mutations in *DMD* lead to Duchenne's and Becker's muscular dystrophies, X-linked recessive disorders that manifests only in males and results from decreased dystrophin production (29). Duchenne's muscular dystrophy patients have early mortality related to respiratory muscle weakness, with decreased muscle strength leading to hypoventilation.

REFERENCES

1. COPDGene. COPDGene, Phase 1 Study Documents [Available from: <http://www.copdgene.org/phase-1-study-documents>].
2. Ferris BG. Epidemiology Standardization Project (American Thoracic Society). *Am Rev Respir Dis*. 1978;118(6 Pt 2):1-120.
3. König IR, Loley C, Erdmann J, Ziegler A. How to include chromosome X in your genome-wide association study. *Genetic epidemiology*. 2014;38(2):97-103.
4. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics*. 2016;48(10):1279-83.
5. Hoggart CJ, Clark TG, De Iorio M, Whittaker JC, Balding DJ. Genome-wide significance for dense SNP and resequencing data. *Genetic epidemiology*. 2008;32(2):179-85.
6. Chu X, Shen M, Xie F, Miao XJ, Shou WH, Liu L, et al. An X chromosome-wide association analysis identifies variants in GPR174 as a risk factor for Graves' disease. *J Med Genet*. 2013;50(7):479-85.
7. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic acids research*. 2016;44(D1):D7-19.
8. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-7.
9. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006.
10. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122.
11. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31(21):3555-7.
12. Coulombe P, Paliouras GN, Clayton A, Hussainkhel A, Fuller M, Jovanovic V, et al. Endothelial Sash1 Is Required for Lung Maturation through Nitric Oxide Signaling. *Cell Rep*. 2019;27(6):1769-80 e4.
13. Zhang Y, Castillo-Morales A, Jiang M, Zhu Y, Hu L, Urrutia AO, et al. Genes that escape X-inactivation in humans have high intraspecific variability in expression, are associated with mental impairment but are not slow evolving. *Mol Biol Evol*. 2013;30(12):2588-601.
14. Cotton AM, Ge B, Light N, Adoue V, Pastinen T, Brown CJ. Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biol*. 2013;14(11):R122.
15. Oliva M, Munoz-Aguirre M, Kim-Hellmuth S, Wucher V, Gewirtz ADH, Cotter DJ, et al. The impact of sex on gene expression across human tissues. *Science*. 2020;369(6509).
16. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nature genetics*. 2013;45(6):580-5.
17. Zheng X, Lin J, Wu H, Mo Z, Lian Y, Wang P, et al. Forkhead box (FOX) G1 promotes hepatocellular carcinoma epithelial-Mesenchymal transition by activating Wnt signal through forming T-cell factor-4/Beta-catenin/FOXG1 complex. *J Exp Clin Cancer Res*. 2019;38(1):475.
18. Wang XH, Cui YX, Wang ZM, Liu J. Down-regulation of FOXR2 inhibits non-small cell lung cancer cell proliferation and invasion through the Wnt/beta-catenin signaling pathway. *Biochem Biophys Res Commun*. 2018;500(2):229-35.

19. Xie H, Xie C. A Six-Gene Signature Predicts Survival of Adenocarcinoma Type of Non-Small-Cell Lung Cancer Patients: A Comprehensive Study Based on Integrated Analysis and Weighted Gene Coexpression Network. *Biomed Res Int.* 2019;2019:4250613.
20. The Health Consequences of Smoking-50 Years of Progress: A Report of the Surgeon General. *Reports of the Surgeon General.* Atlanta (GA)2014.
21. Hao Y, Yang X, Zhang D, Luo J, Chen R. Long noncoding RNA LINC01186, regulated by TGF-beta/SMAD3, inhibits migration and invasion through Epithelial-Mesenchymal-Transition in lung cancer. *Gene.* 2017;608:1-12.
22. VanBelzen DJ, Malik AS, Henthorn PS, Kornegay JN, Stedman HH. Mechanism of Deletion Removing All Dystrophin Exons in a Canine Model for DMD Implicates Concerted Evolution of X Chromosome Pseudogenes. *Mol Ther Methods Clin Dev.* 2017;4:62-71.
23. Dong Y, Simske JS. Vertebrate Claudin/PMP22/EMP22/MP20 family protein TMEM47 regulates epithelial cell junction maturation and morphogenesis. *Dev Dyn.* 2016;245(6):653-66.
24. Lee YH, Song GG. Meta-analysis of differentially expressed genes in ankylosing spondylitis. *Genet Mol Res.* 2015;14(2):5161-70.
25. Tukiainen T, Pirinen M, Sarin AP, Ladenvall C, Kettunen J, Lehtimäki T, et al. Chromosome X-wide association study identifies Loci for fasting insulin and height and evidence for incomplete dosage compensation. *PLoS Genet.* 2014;10(2):e1004127.
26. Liu B, Jiang T, Hu X, Liu Z, Zhao L, Liu H, et al. Downregulation of microRNA30a in bronchoalveolar lavage fluid from idiopathic pulmonary fibrosis patients. *Mol Med Rep.* 2018;18(6):5799-806.
27. Mitchell JA, McCray AT. The Genetics Home Reference: a new NLM consumer health resource. *AMIA Annu Symp Proc.* 2003:936.
28. Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R, et al. Landscape of X chromosome inactivation across human tissues. *Nature.* 2017;550(7675):244-8.
29. Mosqueira M, Zeiger U, Forderer M, Brinkmeier H, Fink RH. Cardiac and Respiratory Dysfunction in Duchenne Muscular Dystrophy and the Role of Second Messengers. *Med Res Rev.* 2013.

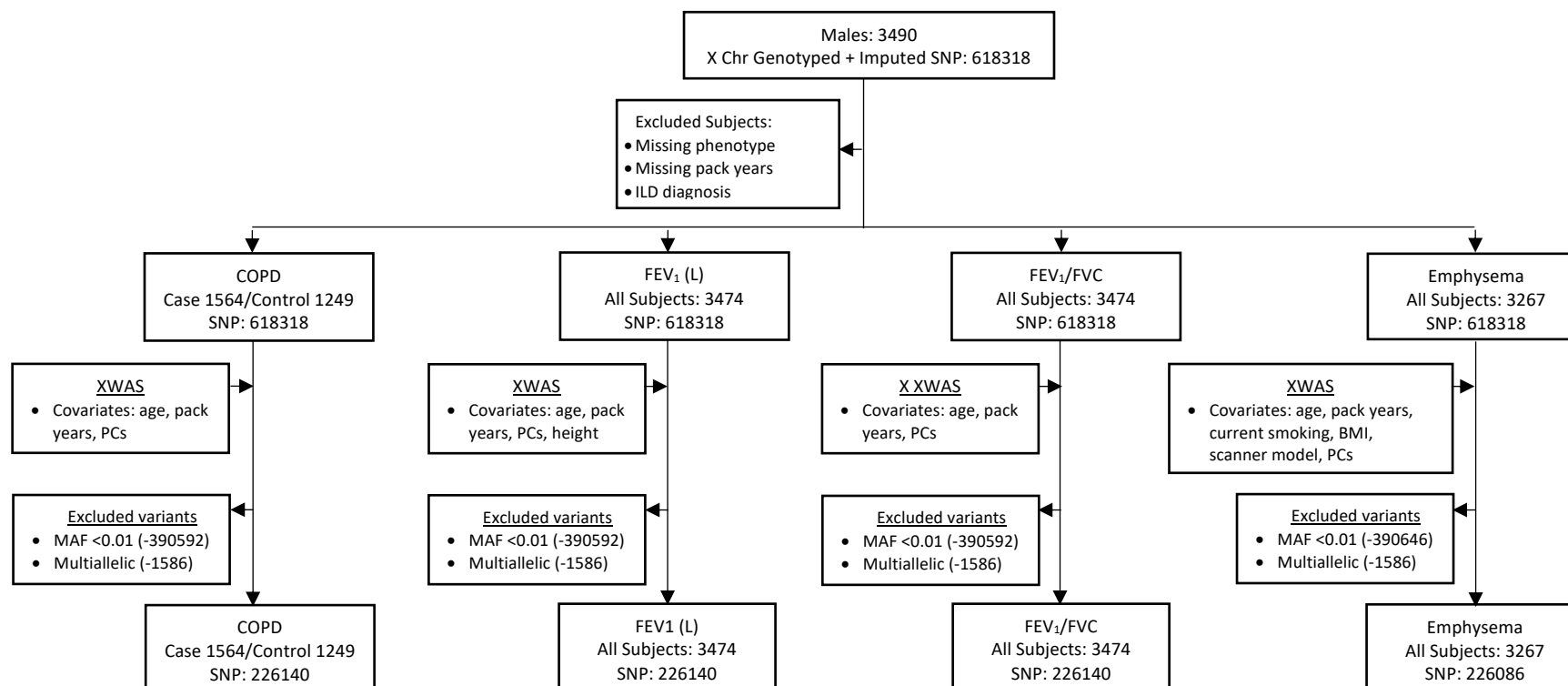
Supplement Figure 1 a: COPDGene XWAS in Stratified Populations All Subjects



Supplement Figure 1 a Steps in XWAS for COPDGene all subjects. Numbers in parentheses represent variants removed.

Abbreviations: *XWAS* X chromosome association study; *Chr* chromosome; *SNP* single nucleotide polymorphism; *ILD* interstitial lung disease; *COPD* chronic obstructive pulmonary disease; *FEV₁* Forced expiratory volume in one second; *L* liters; *FVC* forced vital capacity; *PCs* principal components of genetic ancestry; *MAF* minor allele frequency.

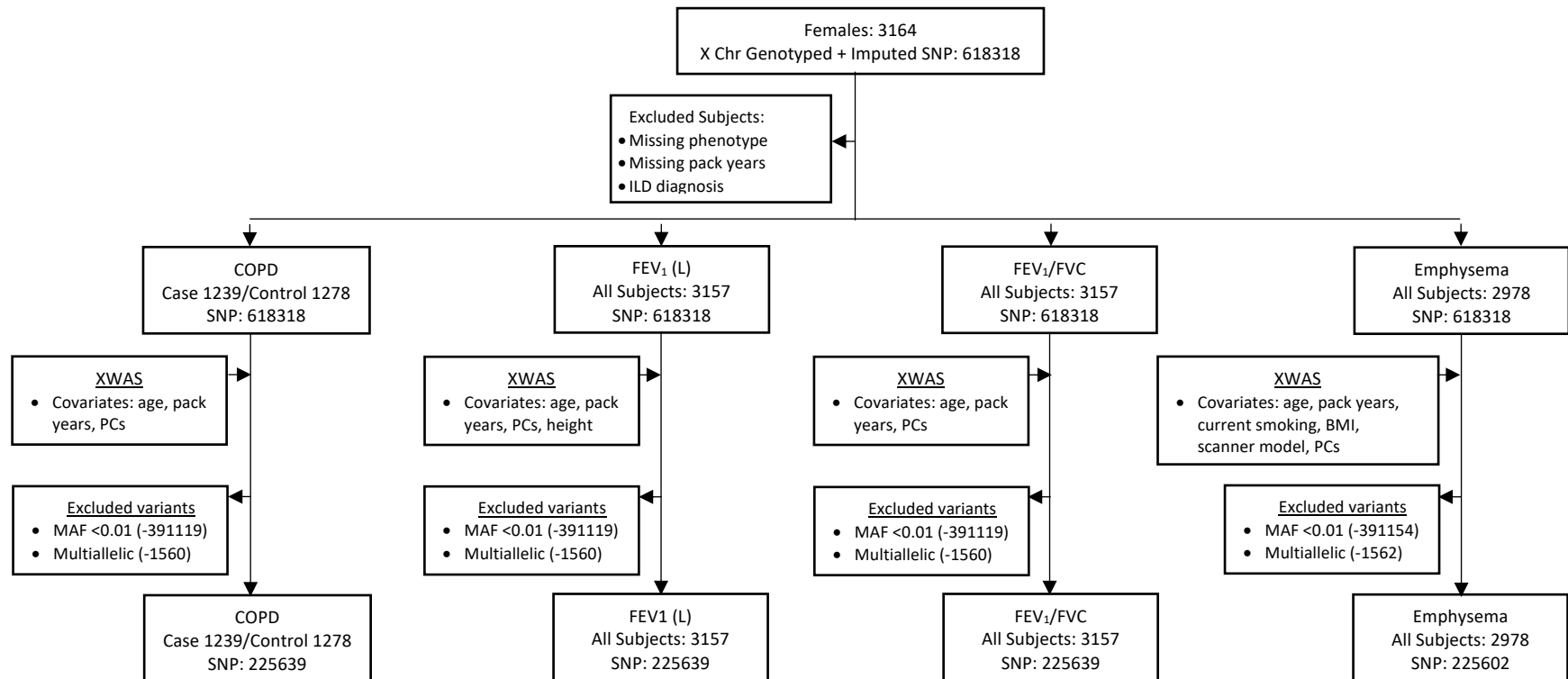
Supplement Figure 1 b: COPDGene XWAS in Stratified Populations Males



Supplement Figure 1 b Steps in XWAS for COPDGene male subjects. Numbers in parentheses represent variants removed.

Abbreviations: *XWAS* X chromosome association study; *Chr* chromosome; *SNP* single nucleotide polymorphism; *ILD* interstitial lung disease; *COPD* chronic obstructive pulmonary disease; *FEV₁* Forced expiratory volume in one second; *L* liters; *FVC* forced vital capacity; *PCs* principal components of genetic ancestry; *MAF* minor allele frequency.

Supplement Figure 1 c: COPDGene XWAS in Stratified Populations Females



Supplement Figure 1 c Steps in XWAS for COPDGene female subjects. Numbers in parentheses represent variants removed.

Abbreviations: *XWAS* X chromosome association study; *Chr* chromosome; *SNP* single nucleotide polymorphism; *ILD* interstitial lung disease; *COPD* chronic obstructive pulmonary disease; *FEV₁* Forced expiratory volume in one second; *L* liters; *FVC* forced vital capacity; *PCs* principal components of genetic ancestry; *MAF* minor allele frequency.

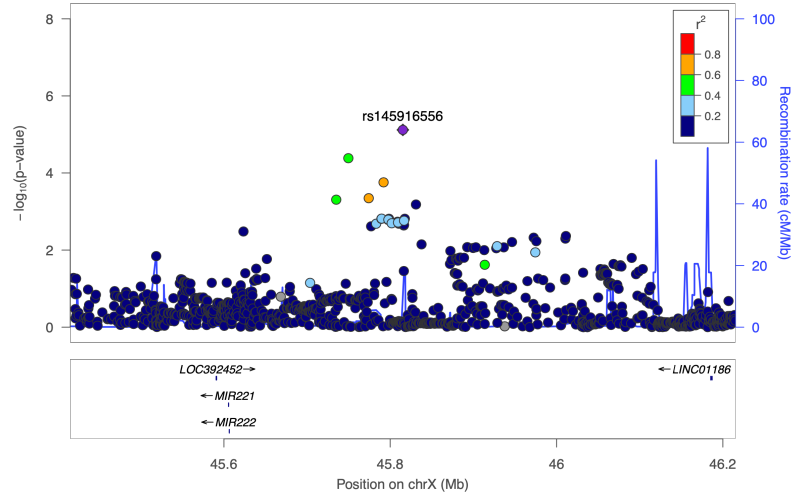
Supplement Figure 2: Meta-analysis XWAS in Stratified Population

ALL SUBJECTS										MALES										FEMALES																				
Subjects					Variants					Subjects					Variants					Subjects					Variants															
COPD																																								
Case	COPDGene	GenKOLS	ECLIPSE	TOTAL	COPDGene	GenKOLS	ECLIPSE													COPDGene	GenKOLS	ECLIPSE				COPDGene	GenKOLS	ECLIPSE												
	2803	853	1726	5382	225757	225861	222412													226140	226202	222038				225639	225766	221661												
Control	2527	805	169	3501	↓ ← plink meta-analysis					1249					404	96	1749	↓ ← plink meta-analysis					1278					401	73	1752	↓ ← plink meta-analysis									
	5330	1658	1895	8883	238339					2813					917	1251	4981	241757					2517					741	644	3902	239431									
					238337															241753										239429										
					↓ → Only in one study (-14383)															↓ → Only in one study (-17724)										↓ → Only in one study (-16134)										
					223954															224029										223295										
FEV1 (L)																																								
Case	COPDGene	GenKOLS	ECLIPSE	TOTAL	COPDGene	GenKOLS	ECLIPSE													COPDGene	GenKOLS	ECLIPSE				COPDGene	GenKOLS	ECLIPSE												
	6631	1658	1904	10193	225757	225861	222499													226140	226202	223607				225639	225766	222193												
Control					238344										3474					917	1258	5649	241852					3157					741	646	4544	239502				
					↓ → Multiallelic (-2)															↓ → Multiallelic (-4)										↓ → Multiallelic (-2)										
					238342															241848										239500										
					↓ → Only in one study (-14346)															↓ → Only in one study (-17580)										↓ → Only in one study (-16138)										
					223996															224268										223362										
FEV1/FVC																																								
Case	COPDGene	GenKOLS	ECLIPSE	TOTAL	COPDGene	GenKOLS	ECLIPSE													COPDGene	GenKOLS	ECLIPSE				COPDGene	GenKOLS	ECLIPSE												
	6631	1658	1904	10193	225757	225861	222498													226140	226202	223607				225639	225766	222181												
Control					238339										3474					917	1258	5649	241852					3157					741	646	4544	239493				
					↓ → Multiallelic (-2)															↓ → Multiallelic (-4)										↓ → Multiallelic (-2)										
					238337															241848										239491										
					↓ → Only in one study (-14338)															↓ → Only in one study (-17580)										↓ → Only in one study (-16130)										
					223999															224268										223361										
Emphysema																																								
Case	COPDGene	GenKOLS	ECLIPSE	TOTAL	COPDGene	GenKOLS	ECLIPSE													COPDGene	GenKOLS	ECLIPSE				COPDGene	GenKOLS	ECLIPSE												
	6245	827	1502	8574	225683	226354	222574													226086	226800	223680				225602	226634	222026												
Control					239714										3267					482	975	4724	244397					2978					345	527	3850	241321				
					↓ → Multiallelic (-4)															↓ → Multiallelic (-2)										↓ → Multiallelic (-2)										
					239710															244395										241319										
					↓ → Only in one study (-15764)															↓ → Only in one study (-20288)										↓ → Only in one study (-17935)										
					223946															224107										223384										

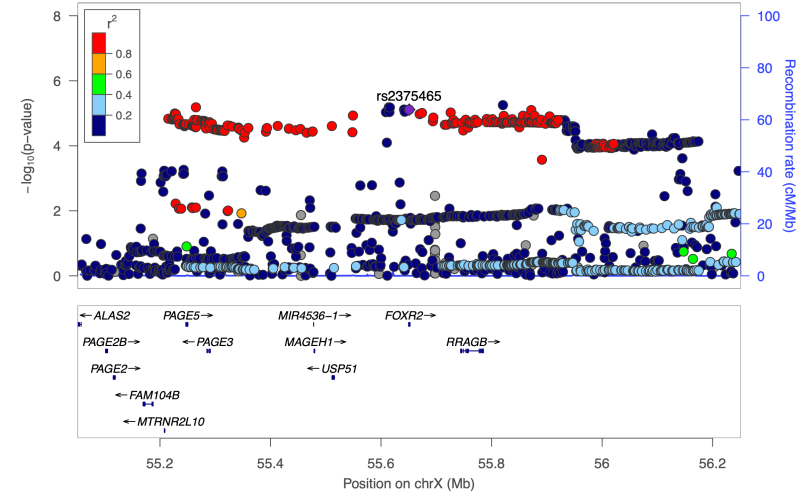
Supplement Figure 2 Steps in XWAS meta-analysis. Numbers in parentheses represent variants removed. Abbreviations: XWAS X chromosome association study; COPD chronic obstructive pulmonary disease; FEV₁ Forced expiratory volume in one second; L liters; FVC forced vital capacity.

Supplement Figure 3: Meta-analysis Locus Plots

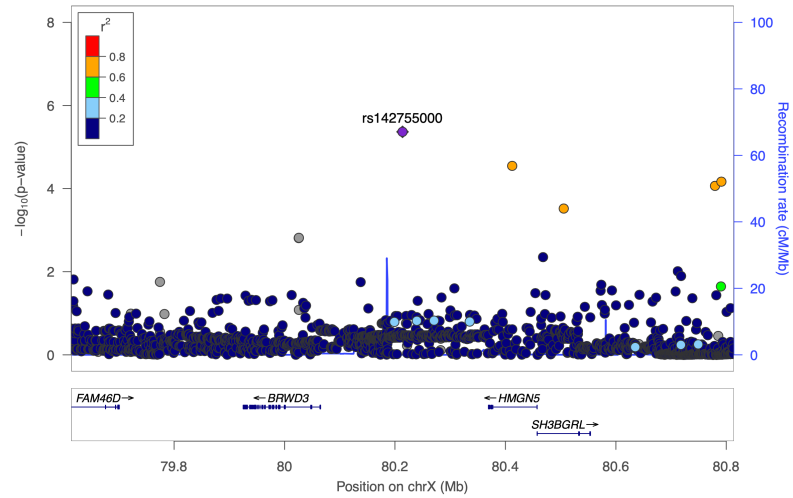
COPD FEMALES XWAS Meta-analysis Near LINC02595



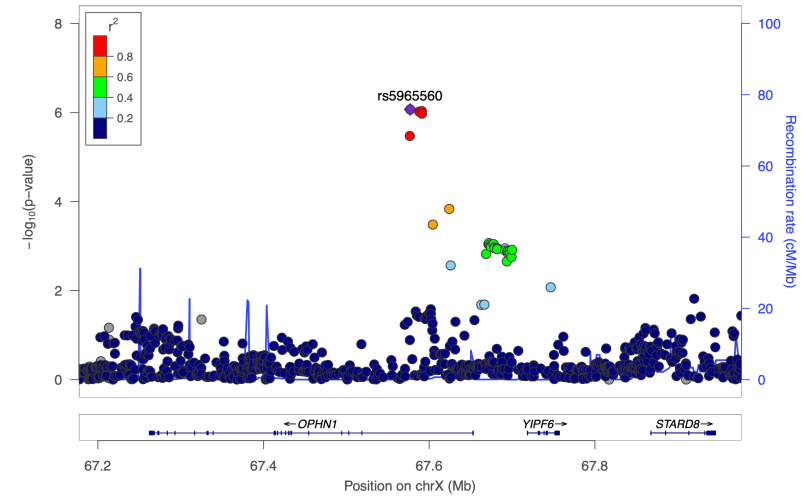
FEV1 L XWAS Meta-analysis Xp11.21 Locus



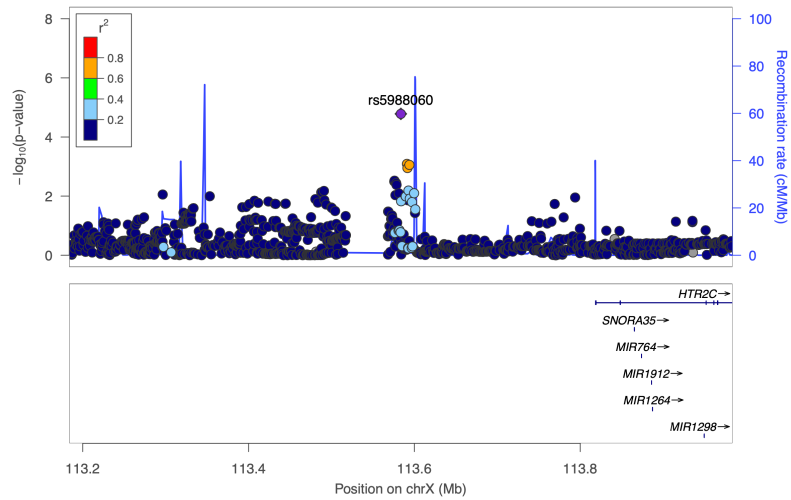
FEV1 L XWAS Meta-analysis Xq21.1 Locus



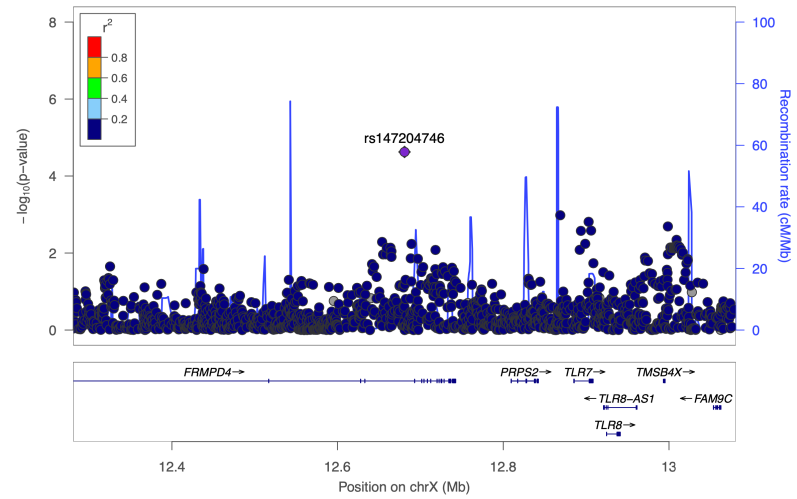
FEV1 L MALES XWAS Meta-analysis In OPHN1



FEV1/FVC FEMALES XWAS Meta-analysis Near HTR2C

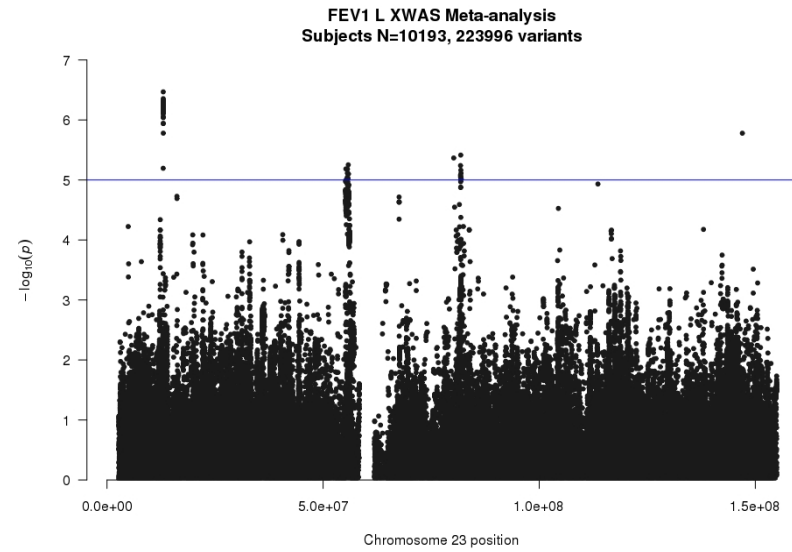
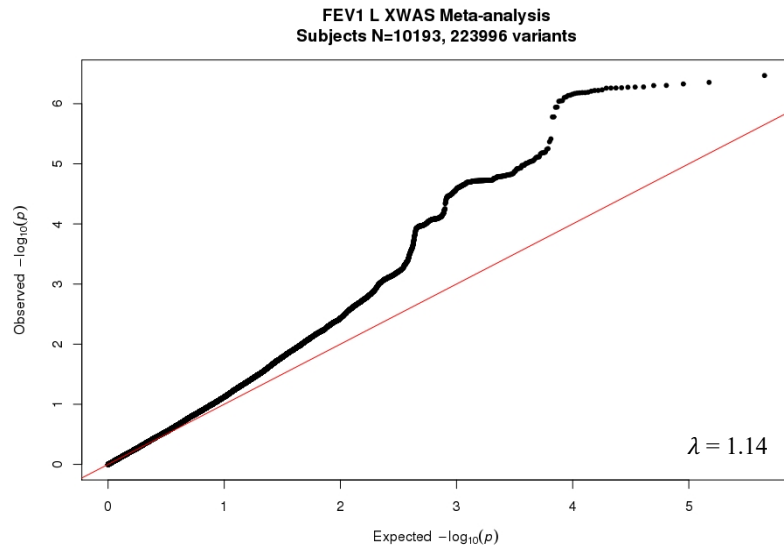
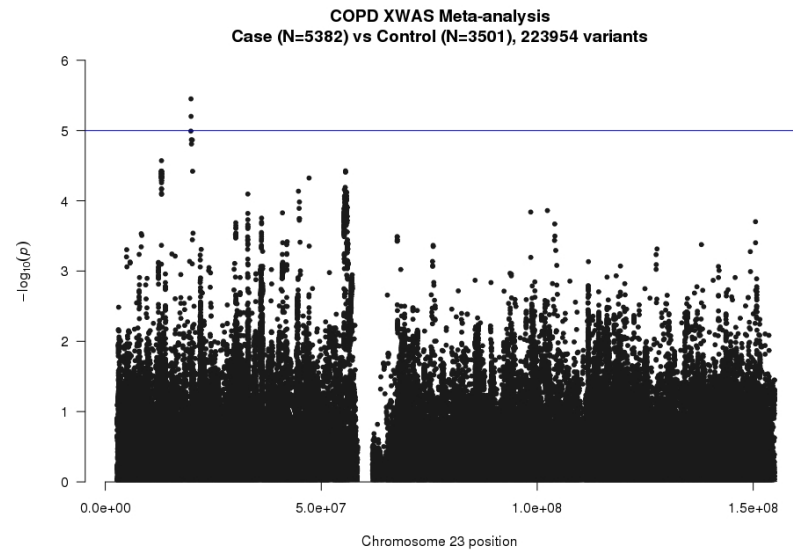
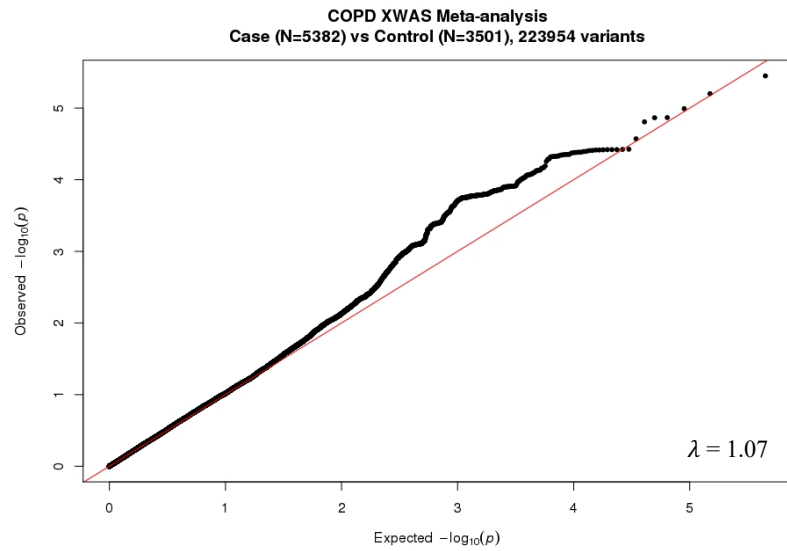


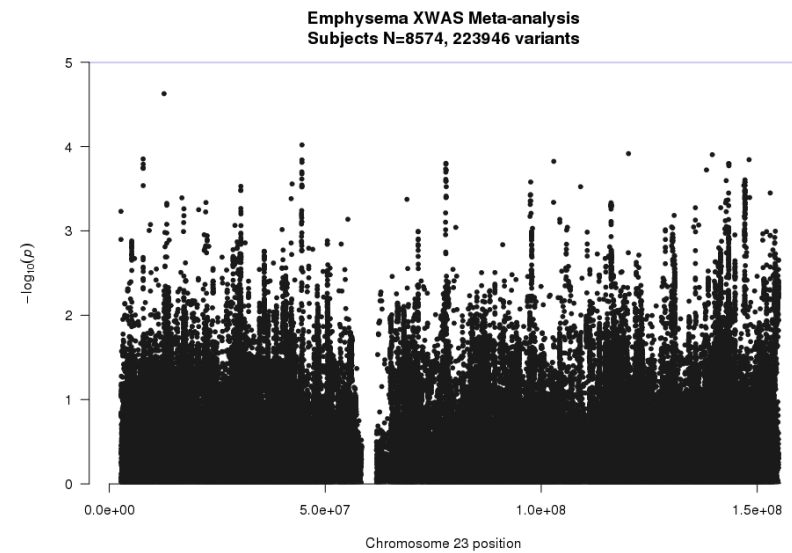
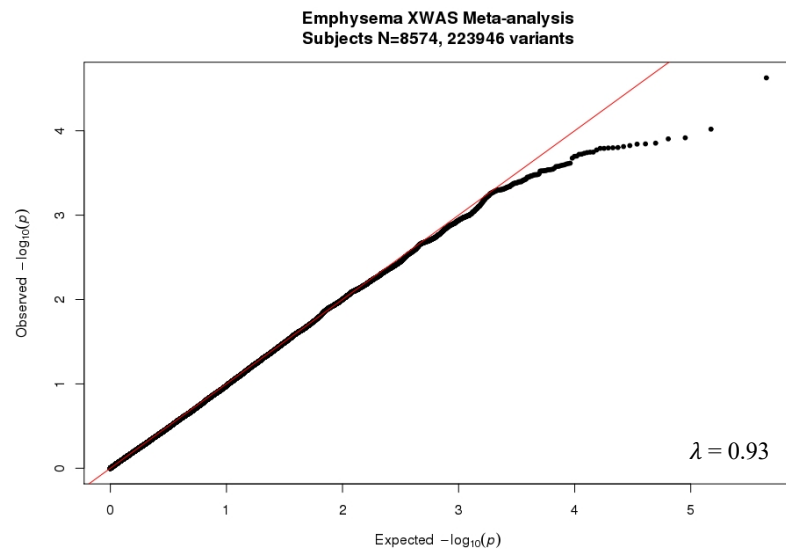
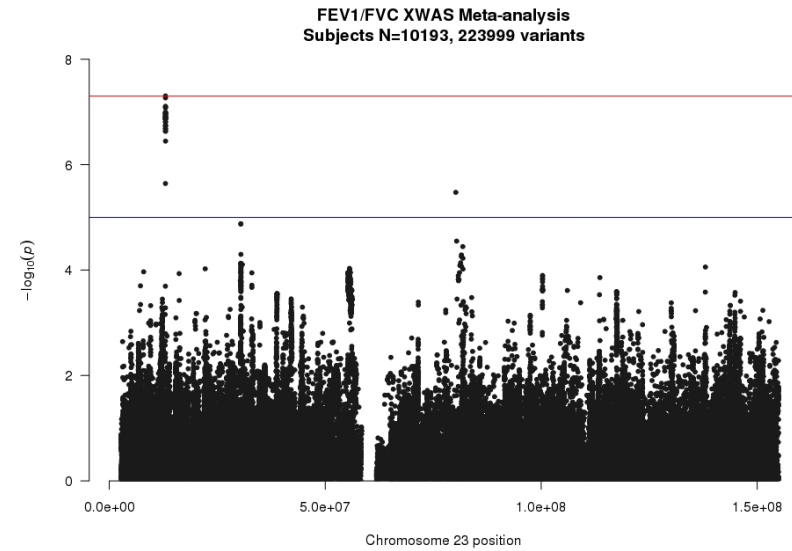
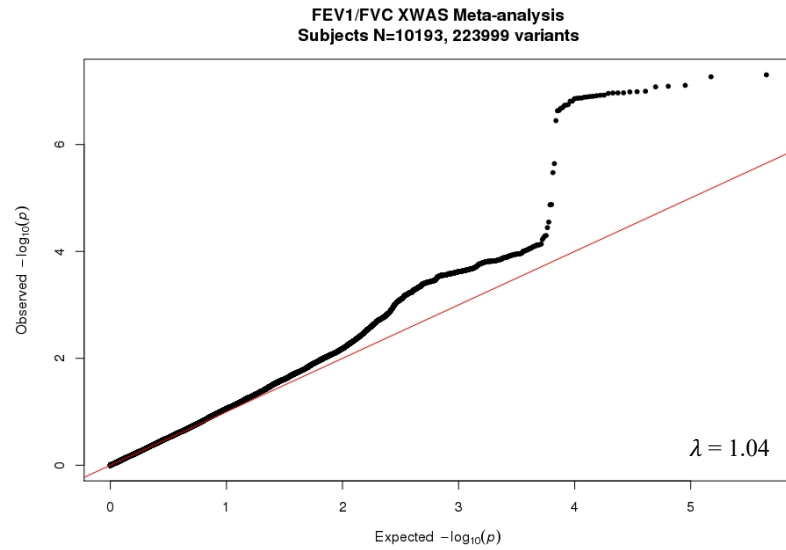
Emphysema XWAS Meta-analysis In FRMPD4

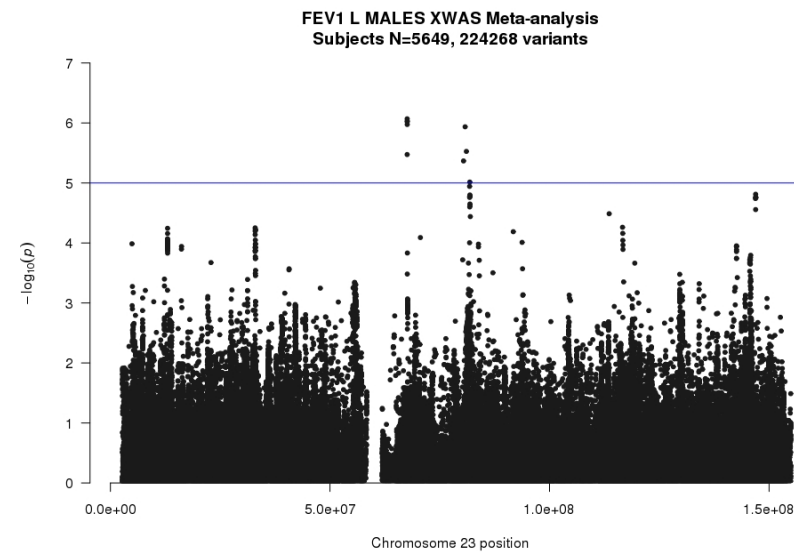
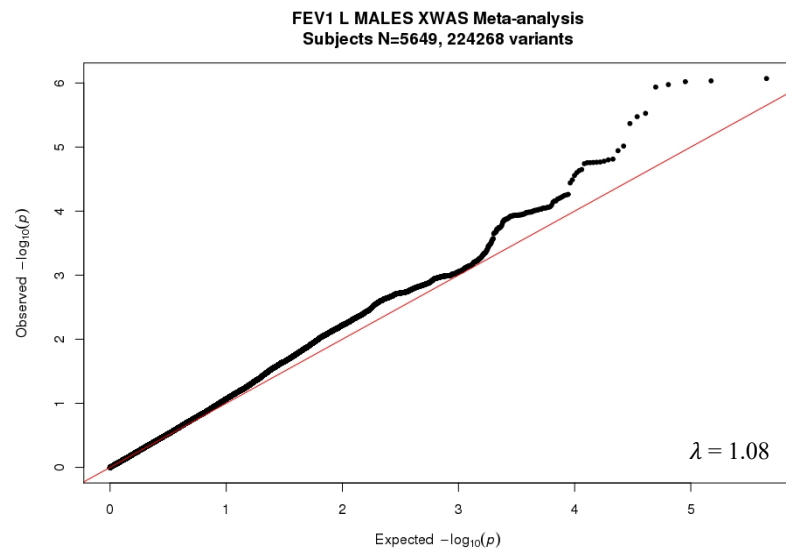
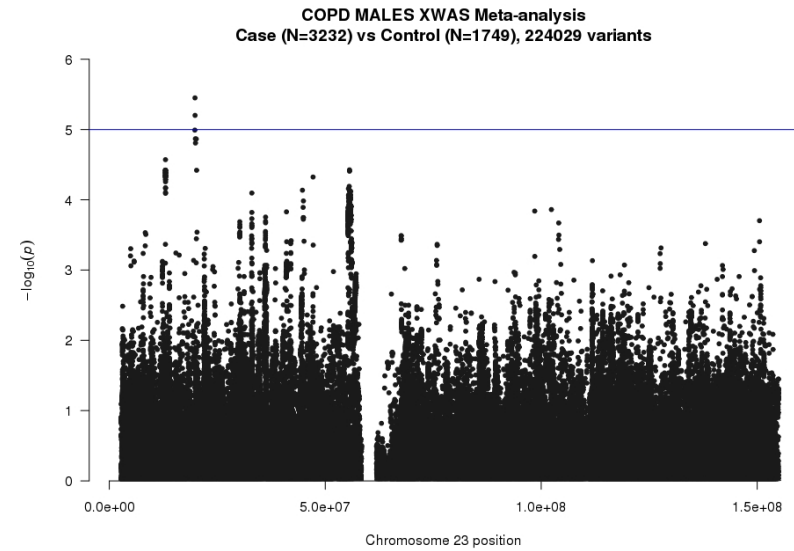
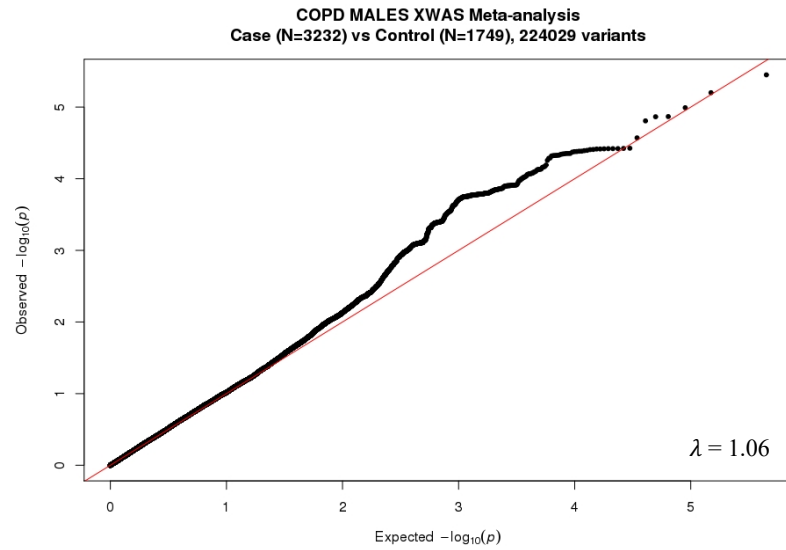


Supplement Figure 3 Locus plots for additional top suggested associations in meta-analysis. Abbreviations: *COPD* chronic obstructive pulmonary disease; *XWAS* X chromosome association study; *FEV₁* Forced expiratory volume in one second; *L* liters; *FVC* forced vital capacity.

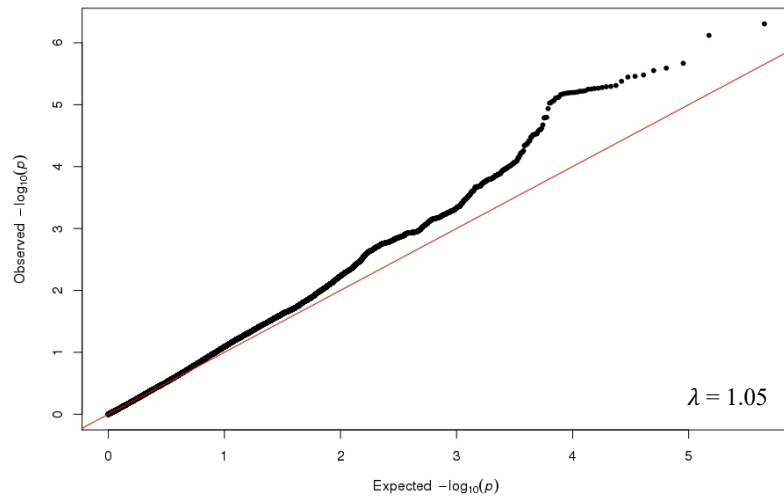
Supplement Figure 4: Meta-analysis Quantile-Quantile and Manhattan Plots



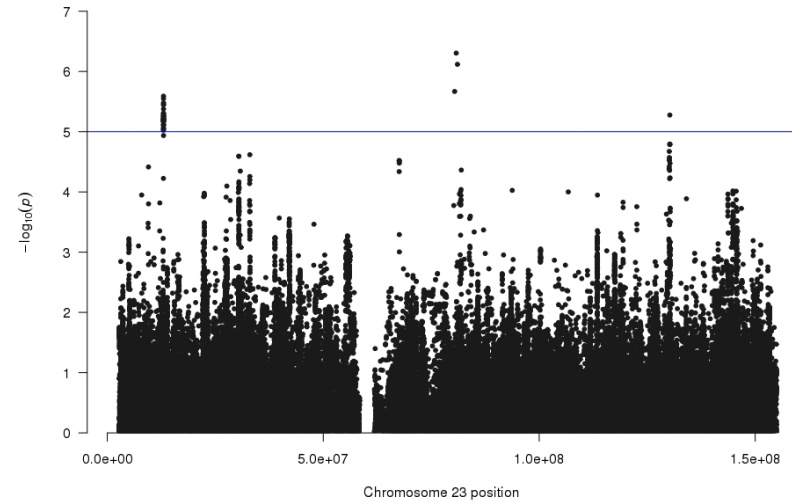




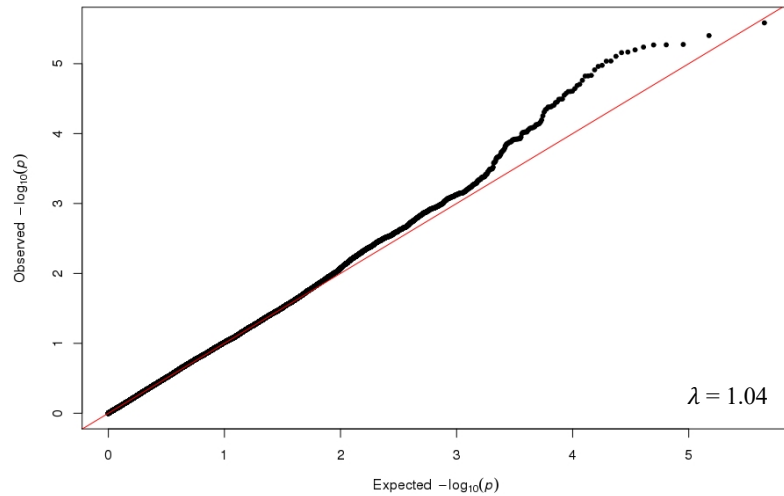
FEV1/FVC MALES XWAS Meta-analysis
Subjects N=5649, 224268 variants



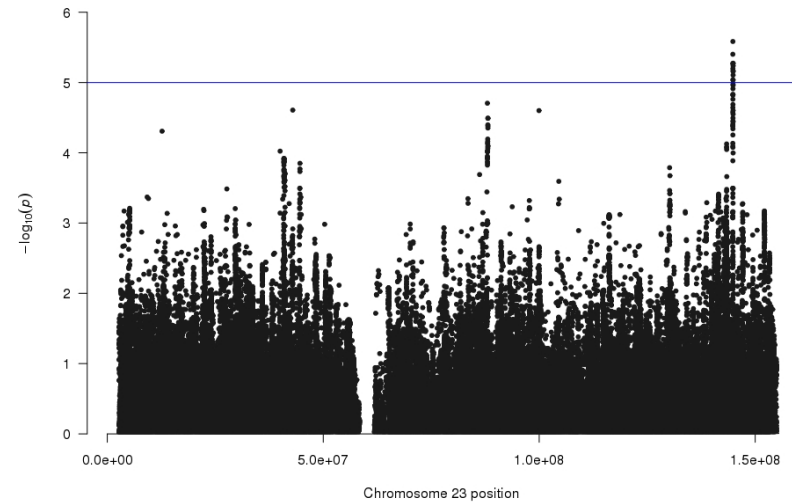
FEV1/FVC MALES XWAS Meta-analysis
Subjects N=5649, 224268 variants



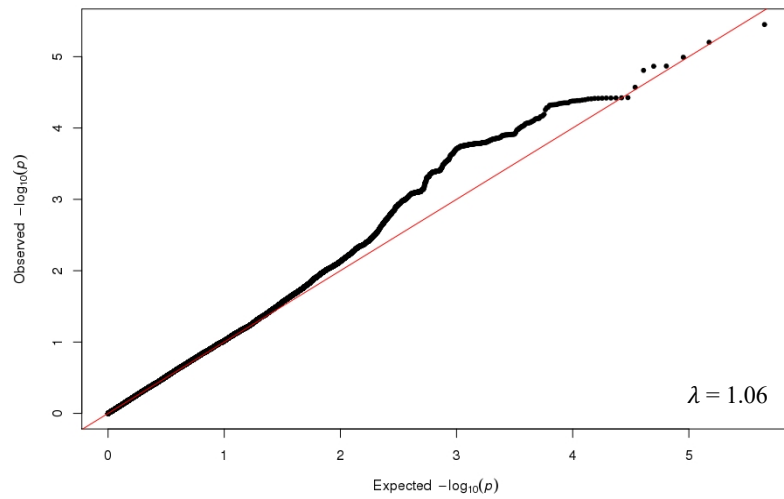
Emphysema MALES XWAS Meta-analysis
Subjects N=4724, 224107 variants



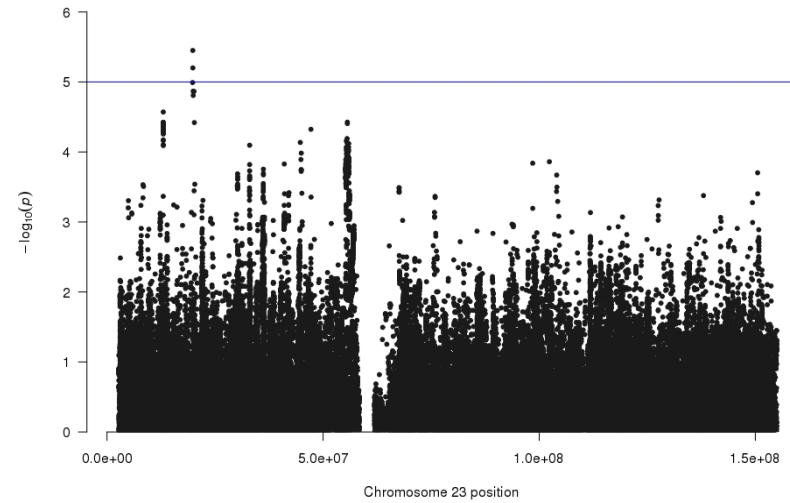
Emphysema MALES XWAS Meta-analysis
Subjects N=4724, 224107 variants



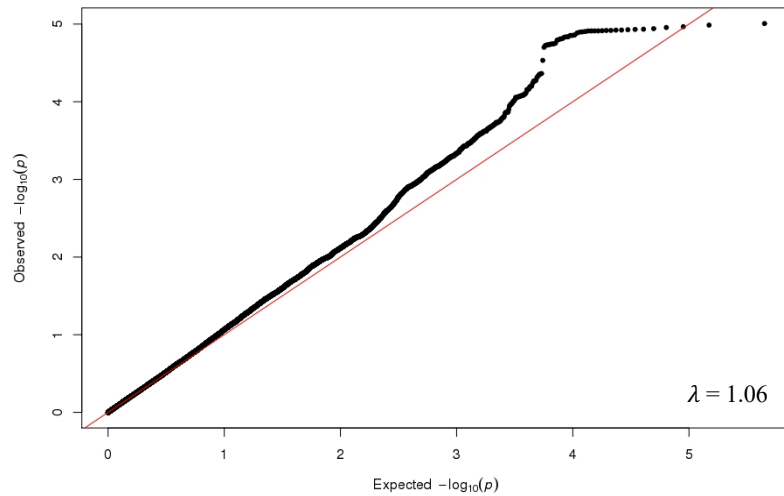
COPD FEMALES XWAS Meta-analysis
Case (N=2150) vs Control (N=1752), 223295 variants



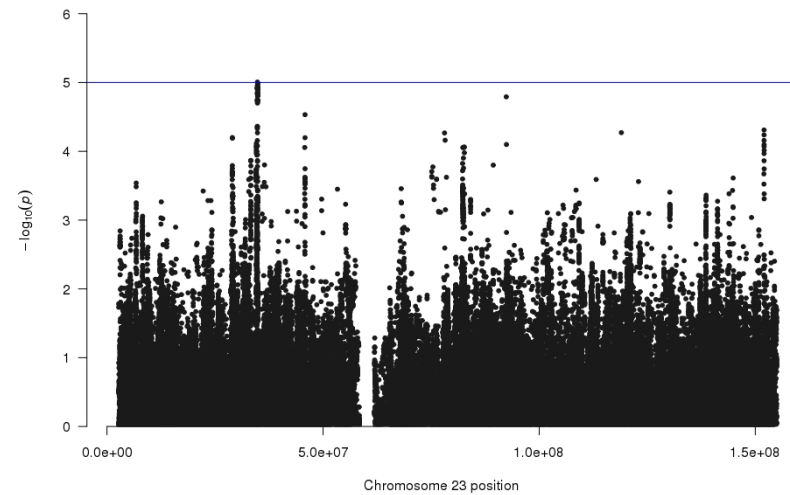
COPD FEMALES XWAS Meta-analysis
Case (N=2150) vs Control (N=1752), 223295 variants

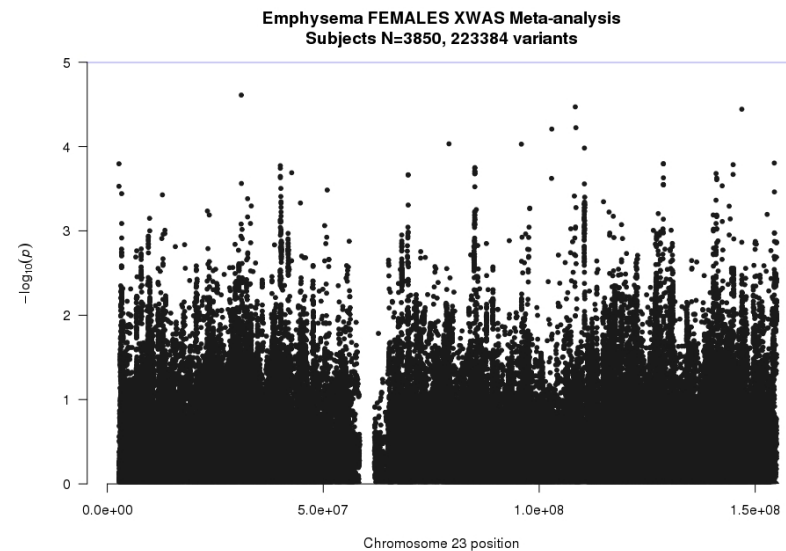
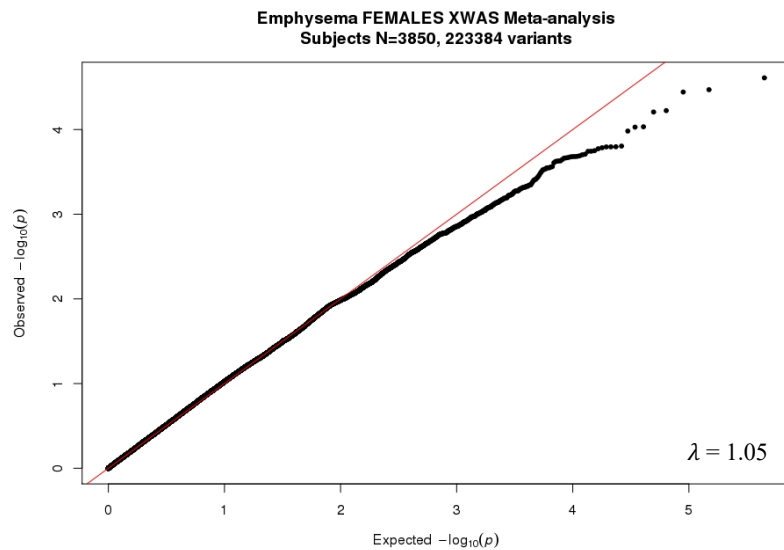
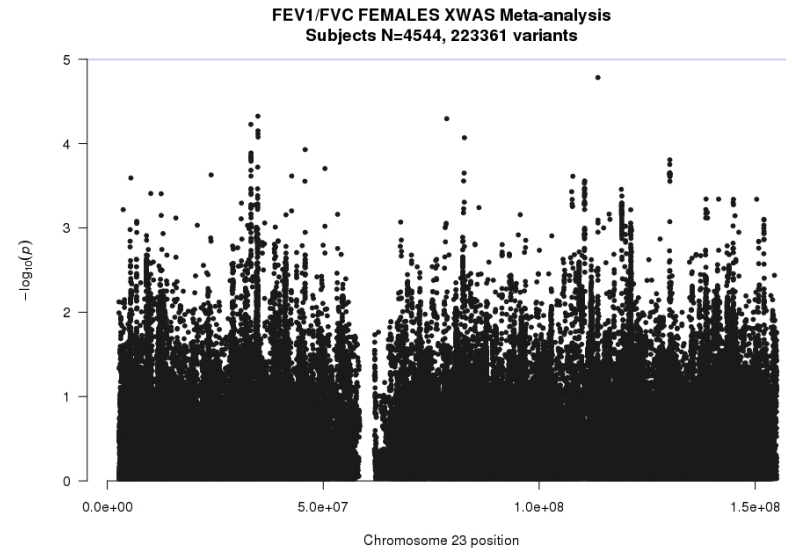
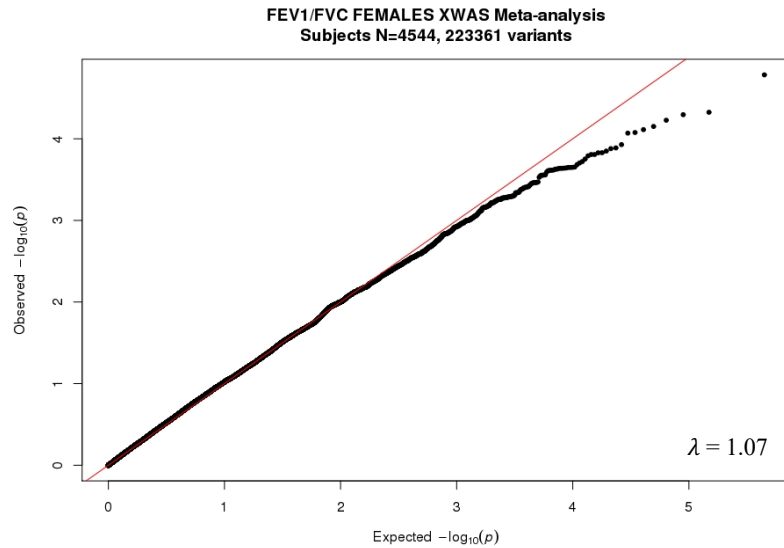


FEV1 L FEMALES XWAS Meta-analysis
Subjects N=4544, 223362 variants



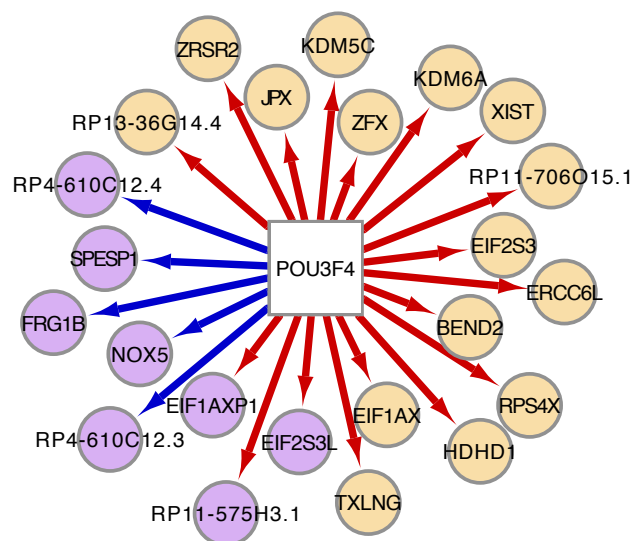
FEV1 L FEMALES XWAS Meta-analysis
Subjects N=4544, 223362 variants





Supplement Figure 4 QQ and Manhattan plots for the twelve XWAS meta-analyses. Abbreviations: *COPD* chronic obstructive pulmonary disease; *XWAS* X chromosome association study; *FEV₁* Forced expiratory volume in one second; *L* liters; *FVC* forced vital capacity.

Supplement Figure 5: Sex differential edge weights connecting the transcription factor *POU3F4* in GTEx Lung Tissue



Supplement Figure 5 Sex differential edge weights connecting the transcription factor *POU3F4* to 23 target genes (FDR<0.05). Includes 18 female-biased edges (red), 5 male-biased edges (blue). Yellow nodes represent target genes in the X chromosome and purple notes in the autosome.