

RESEARCH

Open Access



# X chromosome associations with chronic obstructive pulmonary disease and related phenotypes: an X chromosome-wide association study

Lystra P. Hayden<sup>1,2</sup>, Brian D. Hobbs<sup>2,3</sup>, Robert Busch<sup>4</sup>, Michael H. Cho<sup>2,3</sup>, Ming Liu<sup>5</sup>, Camila M. Lopes-Ramos<sup>2,6</sup>, David A. Lomas<sup>7</sup>, Per Bakke<sup>8</sup>, Amund Gulsvik<sup>8</sup>, Edwin K. Silverman<sup>2,3</sup>, James D. Crapo<sup>9</sup>, Terri H. Beaty<sup>10</sup>, Nan M. Laird<sup>6</sup>, Christoph Lange<sup>6</sup> and Dawn L. DeMeo<sup>2,3\*</sup>

## Abstract

**Background** The association between genetic variants on the X chromosome to risk of COPD has not been fully explored. We hypothesize that the X chromosome harbors variants important in determining risk of COPD related phenotypes and may drive sex differences in COPD manifestations.

**Methods** Using X chromosome data from three COPD-enriched cohorts of adult smokers, we performed X chromosome specific quality control, imputation, and testing for association with COPD case–control status, lung function, and quantitative emphysema. Analyses were performed among all subjects, then stratified by sex, and subsequently combined in meta-analyses.

**Results** Among 10,193 subjects of non-Hispanic white or European ancestry, a variant near *TMSB4X*, rs5979771, reached genome-wide significance for association with lung function measured by FEV<sub>1</sub>/FVC ( $\beta$  0.020, *SE* 0.004,  $p$   $4.97 \times 10^{-08}$ ), with suggestive evidence of association with FEV<sub>1</sub> ( $\beta$  0.092, *SE* 0.018,  $p$   $3.40 \times 10^{-07}$ ). Sex-stratified analyses revealed X chromosome variants that were differentially trending in one sex, with significantly different effect sizes or directions.

**Conclusions** This investigation identified loci influencing lung function, COPD, and emphysema in a comprehensive genetic association meta-analysis of X chromosome genetic markers from multiple COPD-related datasets. Sex differences play an important role in the pathobiology of complex lung disease, including X chromosome variants that demonstrate differential effects by sex and variants that may be relevant through escape from X chromosome inactivation. Comprehensive interrogation of the X chromosome to better understand genetic control of COPD and lung function is important to further understanding of disease pathology.

**Trial registration** Genetic Epidemiology of COPD Study (COPDGene) is registered at ClinicalTrials.gov, NCT00608764 (Active since January 28, 2008). Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints Study (ECLIPSE), GlaxoSmithKline study code SCO104960, is registered at ClinicalTrials.gov, NCT00292552 (Active since

\*Correspondence:

Dawn L. DeMeo

dawn.demeo@channing.harvard.edu

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

February 16, 2006). Genetics of COPD in Norway Study (GenKOLS) holds GlaxoSmithKline study code RES11080, Genetics of Chronic Obstructive Lung Disease.

**Keywords** COPD, Lung function, Emphysema, X chromosome-wide association study, Sex differences, X chromosome inactivation

## Background

Chronic obstructive pulmonary disease (COPD) is a disease with both environmental and genetic risk factors [1, 2]. Several genome-wide association studies (GWAS) have demonstrated multiple autosomal associations [3]. However, much of the overall heritability of COPD remains unexplained [4]. One potential source of additional information is genetic variation on the X chromosome, which has been excluded in most prior COPD GWAS [5].

Inclusion of the X chromosome in association studies requires modified quality control procedures and poses statistical challenges distinct from autosomes, including the presence of only one X chromosome in males and random inactivation of one X in females [5, 6]. Standard GWAS methodology for autosomal variants relies on the fact that each locus has 2 alleles and thus each paired autosomal variant generally has three possible genotypes, 0/1/2. If these methods are applied to the X chromosome, where males are hemizygous, the X variants would be coded as 0/1 for males and 0/1/2 for females. This may limit the ability to detect associations compared to methods that consider male X chromosome genetic variants as 0/2 [7]. Underutilization of genetic information on the X chromosome has been noted along with calls for greater incorporation of allosomal data in future GWAS of complex human traits and diseases [8].

The motivation of this study is to properly include and assess the X chromosome variation on a genome-wide level to more completely understand sex differences in COPD. Women are more susceptible to severe, early-onset COPD [9]. Disease features such as emphysema and exacerbations show sex differences, as do disease manifestations of dyspnea, depression, and anxiety [10, 11]. Several GWAS have begun incorporating the X chromosome [12–14]. Variants associated with lung function on the X chromosome have been identified in previous GWAS [14–16]. There has been limited prior interrogation of the X chromosome in GWAS of COPD, and related phenotypes including emphysema, that employed comprehensive accounting for the X chromosome and included sex-stratified analysis [16, 17]. Prior studies have shown the significance of sex differences and the X chromosome in cardiovascular disease and related traits, but similar examinations have not been performed in COPD [18].

This study aims to perform an X chromosome-wide association study (XWAS) and meta-analyses of COPD datasets, including sex-stratified analyses to test for association between genetic loci on the X chromosome with COPD and related phenotypes of lung function and quantitative emphysema (emphysema) on chest computed tomography (CT) scans. To achieve this, we utilize quality control and statistical methods that specifically consider X chromosome inheritance patterns and inactivation. We hypothesize that the X chromosome harbors variants important in determining risk of COPD and related quantitative phenotypes, and that X chromosome variants may drive some sex differences in COPD manifestations.

Some results have been previously reported as an abstract [19].

## Methods

### Study participants and phenotyping

Study participants were current and former smokers in three previously described studies: Genetic Epidemiology of COPD Study (COPDGene, non-Hispanic white subset), Genetics of COPD in Norway Study (GenKOLS), and Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints Study (ECLIPSE) [1, 20, 21]. Subjects with COPD (cases) were defined by Global Initiative for Chronic Obstructive Lung Disease (GOLD) airflow limitation severity grades 2–4, on the basis of standardized post-bronchodilator spirometry with Forced Expiratory Volume in one second ( $FEV_1$ ) < 80% predicted, and  $FEV_1$  to Forced Vital Capacity (FVC) ratio < 0.7 [22]. Control subjects (controls) included smokers with normal spirometry ( $FEV_1 \geq 80\%$ ,  $FEV_1/FVC \geq 0.7$ ). Measurement of quantitative emphysema (emphysema) was defined as the natural logarithm transform of the percentage of lung voxels with density less than  $-950$  Hounsfield units on chest CT inspiratory images ( $\log - 950$ ), determined using Thirona software (<http://www.thirona.eu>) for COPDGene, and Slicer software (<http://www.slicer.org>) for ECLIPSE and GenKOLS [23].

Study subjects provided written informed consent and each study's research protocol was approved by institutional review boards at participating institutions. Study phenotypes are further defined in the supplement

(Additional file 1). Datasets from COPDGene (accession number phs000179.v6.p2) and ECLIPSE (accession number phs001252.v1.p1) are publicly available in dbGaP.

**Genotyping, quality control, and imputation**

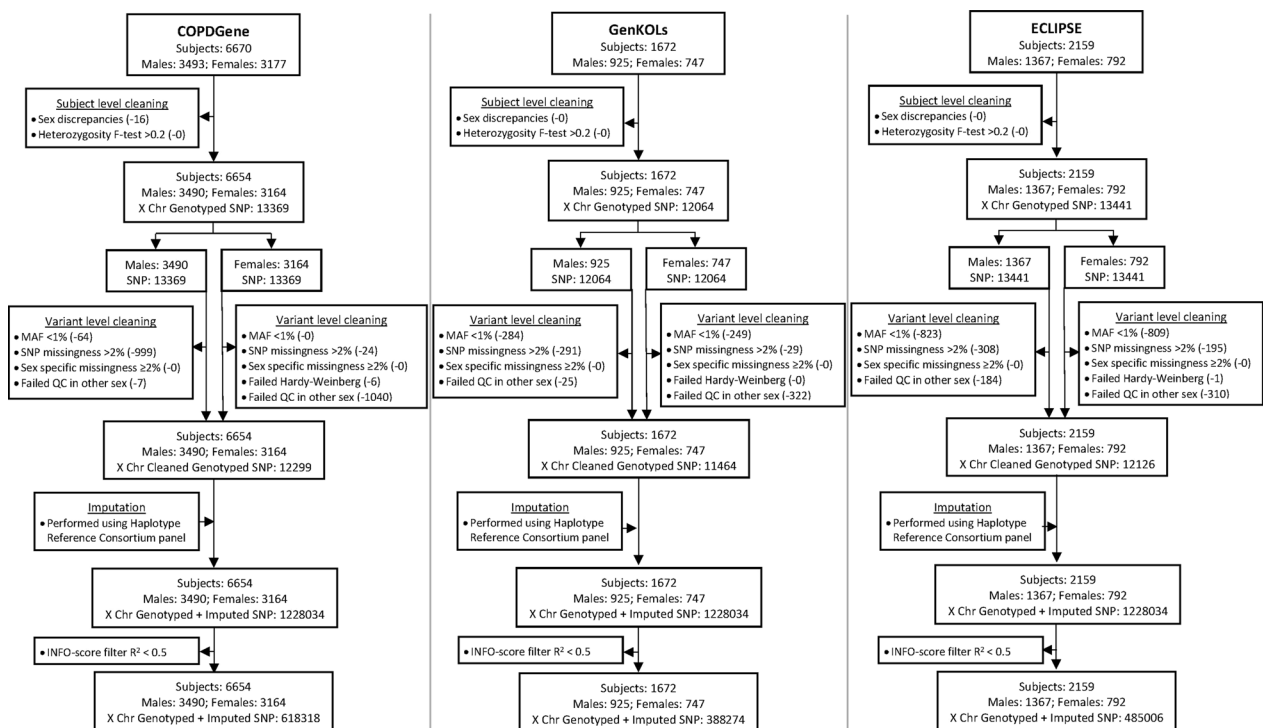
Genotyping for all studies was performed using various Illumina (San Diego, CA) platforms. Genotyping and initial quality control methods have been previously described, and included cleaning based on subject-level missingness, cryptic relatedness, and sex checks based on X and Y chromosomes [21, 24]. Principal components of genetic ancestry were generated separately, based on genotyped autosomal data in each case-control population using EIGENSOFT, as previously described. The pseudoautosomal region was excluded from the X chromosome prior to association analysis, defined for human genome build 19 (GRCh37/hg19, <https://www.ncbi.nlm.nih.gov/grc/human>) by X chromosome base pair coordinates 60,001–2,699,520 and 154,931,044–155,260,560.

Additional quality control steps for variants on the X chromosome followed published recommendations and were performed using PLINK v1.9 (Fig. 1) [5, 25]. A full description of study based subject level data cleaning as

well as imputation methods are in the supplement (Additional file 1).

**Statistical analysis**

X chromosome variants were tested in each of the three studies independently for association with four individual phenotypes: COPD case-control status, FEV<sub>1</sub> in liters, FEV<sub>1</sub>/FVC and quantitative emphysema using logistic regression for COPD status and linear regression analysis for quantitative phenotypes. XWAS were run in all subjects and additionally stratified by sex as demonstrated in Additional file 1: Figure S1a–c for COPDGene. Regression models for each phenotype were adjusted for age, pack-years of smoking, and principal components of genetic ancestry for all analyses [6, 25, 26]. For FEV<sub>1</sub> in liters height was included as a covariate. For the analyses of all subjects, sex was considered as a covariate (females 0, males 1). For the emphysema analyses, additional covariates included current smoking, body mass index, and scanner model if more than one scanner was used in that study. Variants were excluded if they had a minimum minor allele frequency of <0.01, were multiallelic, or in the COPD case/control XWAS if they were present in only one group.



**Fig. 1** Cleaning and imputation by study. Steps in subject cleaning, variant cleaning, and imputation prior to X chromosome association analysis. Numbers in parentheses represent subjects or variants removed. Abbreviations: Chr chromosome, SNP single nucleotide polymorphism, MAF minor allele frequency, QC quality control

Clayton's method for XWAS was employed, which accounts for both female X chromosome inactivation (XCI) and male X chromosome hemizygoty [6, 27]. Males carry only one copy of the X chromosome, while in female most loci are subject to XCI so that a female will have approximately half of their cells with one copy active and the remainder will have the other copy activated. Males are equivalent to homozygous females in respect to such loci. Clayton's method for analysis of the X chromosome, implemented in PLINK v2.0, performs logistic regression models with one degree of freedom encoding female loci as 0 (homozygous for reference allele), 1 (heterozygous), or 2 (homozygous for alternate allele) and male loci as 0 (no copies of alternate allele) or 2 (single copy of alternate allele). With this approach the heterozygous female genotype falls between two homozygous genotypes on the linear predictor scale, thus accounting for XCI as only 50% of cells will have a normal active allele [27].

Results were subsequently combined into fixed-effects meta-analyses using all quality-controlled variants from COPDGene, ECLIPSE and GenKOLs studies with PLINK v1.9 (Additional file 1: Figure S2). Meta-analyses were run for each of the four phenotypes, both for all subjects and separately in sex-stratified datasets. Variants were excluded if only present in one study. Testing

for sex-difference was performed comparing effect estimates among males and females among the top suggested associations in the sex-stratified meta-analysis, with sex-difference significance assessed at  $P < 0.05$  [28]. A description of methods for significance thresholds, the suggestive level of association examined, and annotation is in the supplement (Additional file 1).

### Replication

Ten X chromosome variants previously discovered in GWAS for lung function were examined for replication in the meta-analysis results for this XWAS [14, 16, 17]. One variant, rs28382751 from Shrine et al. was not evaluated as it was not present in our study.

## Results

### Subjects

A total of 10,193 subjects of non-Hispanic white or European ancestry with X chromosome data were included in the analysis. Baseline characteristics and summary statistics for each dataset among all subjects and in sex-stratified subsets are shown in Table 1. The largest study population was from COPDGene, with 6631 subjects, while GenKOLs had 1658 subjects and ECLIPSE 1904 subjects. Compared to the other two studies, GenKOLs had fewer mean pack-years of smoking history, and a

**Table 1** Characteristics of study populations

	COPDGene			GenKOLS			ECLIPSE		
	All Subjects	Males	Females	All Subjects	Males	Females	All Subjects	Males	Females
Subjects with spirometry	6631	3474 (52.39%)	3157 (47.61%)	1658	917 (55.31%)	741 (44.69%)	1904	1258 (66.07%)	646 (33.93%)
Mean age, years (SD)	62.03 (8.84)	62.38 (8.81)	61.65 (8.85)	60.69 (11.05)	61.88 (11.15)	59.21 (10.74)	63.14 (7.51)	63.62 (7.43)	62.22 (7.59)
Mean pack Year History (SD)	47.27 (26.03)	51.54 (28.06)	42.57 (22.68)	25.98 (17.41)	29.49 (18.86)	21.64 (14.28)	48.7 (27.75)	52.12 (29.66)	42.14 (22.20)
Current smoking (%)	2594 (39.12%)	1385 (39.87%)	1209 (38.30%)	731 (44.09%)	377 (41.11%)	354 (47.77%)	661 (35.46%)*	422 (34.20%)*	239 (37.94%)*
GOLD II-IV COPD (%)	2803 (42.27%)	1564 (45.02%)	1239 (39.25%)	853 (51.45%)	513 (55.94%)	340 (45.88%)	1726 (91.08%)	1155 (92.33%)	571 (88.66%)
FEV <sub>1</sub> , % predicted (SD)	73.56 (25.90)	72.41 (26.49)	74.83 (25.18)	72.18 (26.17)	70.06 (26.19)	74.80 (25.92)	53.13 (23.23)	51.33 (22.74)	56.64 (23.78)
FEV <sub>1</sub> , L (SD)	2.22 (0.95)	2.51 (1.03)	1.88 (0.72)	2.38 (1.10)	2.66 (1.19)	2.04 (0.87)	1.51 (0.78)	1.63 (0.83)	1.28 (0.63)
FEV <sub>1</sub> /FVC (SD)	64.09 (16.61)	62.59 (16.93)	65.74 (16.09)	64.84 (16.92)	63.15 (17.24)	66.94 (16.28)	47.75 (14.90)	46.57 (14.77)	50.04 (14.89)
Subjects with chest CT data	6245	3267	2978	827	482	345	1502	975	527
Emphysema, log -950 (SD)	0.94 (1.71)	1.26 (1.54)	0.60 (1.82)	0.56 (1.80)	0.91 (1.66)	0.07 (1.87)	2.42 (1.11)	2.50 (1.05)	2.29 (1.21)

Comparison of characteristics of subjects from each the three study populations among all study participant and by sex. \*Data is missing for some subjects. Abbreviations: SD standard deviation; GOLD Global Initiative for Chronic Obstructive Lung Disease; COPD chronic obstructive pulmonary disease; FEV<sub>1</sub> Forced expiratory volume in one second; L liters; FVC forced vital capacity; CT computed tomography

higher proportion of current smokers. Compared to the other two studies, ECLIPSE, which enrolled a majority of subjects with COPD and comparatively few control smokers, had fewer female subjects and more severe disease. Among males and females in each study group, female subjects were slightly younger, had a shorter pack-year smoking history, and had less severe disease, characterized by higher lung function, less emphysema (especially among females in ECLIPSE), and a lower proportion of subjects with COPD.

Subject level cleaning, variant level cleaning, and imputation of genotyped data was done individually for each study (Fig. 1). The resultant variants were included in the XWAS, with 618318 genotyped and imputed variants in COPDGene, 388274 in GenKOLs, and 485006 in ECLIPSE. XWAS were performed separately for each phenotype among all subjects and then in sex-stratified analyses (Additional file 1: Figure S1a–c demonstrates the XWAS analyses for COPDGene). This included association studies for each of the four phenotypes assessed separately in each of the three study strata: all subjects, males, and females. Results were combined using fixed-effects meta-analyses, one for each of the four phenotypes (Additional file 1: Figure S2), with 223295–224268 variants included in each meta-analysis depending on phenotype and population.

### XWAS

Top suggestive associations from the all-subjects and sex-stratified XWAS meta-analyses are presented in Table 2 and Additional file 2: Table S1. Figure 2 and Additional file 1: Figure S3 provide locus plot visualizations of selected top association results. Quantile–Quantile and Manhattan plots for the meta-analyses are shown in Additional file 1: Figure S4. The top association in rs5979771, a variant closest to *TMSB4X*, achieved genome-wide significance for association with FEV<sub>1</sub>/FVC among all subjects (Beta ( $\beta$ ) 0.020, standard error (SE) 0.004,  $p 4.97 \times 10^{-08}$ ). The same variant was also the top suggestive variant to show evidence of association with FEV<sub>1</sub> among all subjects. No other variants reached genome-wide significance, but suggestive associations approaching genome-wide significance were identified. Power calculations can be found in Additional file 1.

### Sex differences

Testing for sex difference compared effect estimates among the top suggested associations in sex-stratified meta-analysis for males and females (Table 2). Testing was run in 32 variants, which included 28 unique variants annotated to 25 unique genes. There were significant sex differences identified among 20 of the variants, which implicated 17 unique genes based on annotation

to the closest gene. These variants were found across all four phenotypes, including 7 variants with larger effect in males and 13 variants with larger effect in females. There were two genes, *DMD* and *POU3F4*, both implicated by more than one variant among the top suggested associations, where the sex-effect was different dependent on the variant.

### ACE2

Recent attention has come to the X chromosome gene *ACE2* and its role in SARS-CoV-2 susceptibility. SARS-CoV-2 severity has been shown to vary by sex, and there is increased risk for severe disease in those with chronic respiratory conditions including COPD [29, 30]. Meta-analysis results for variants in *ACE2*, located at Xp22.2 (base pair region 15579156–15620192), were examined and top associations with each phenotype and sex strata are in Additional file 3: Table S2. None of the *ACE2* variants approached genome-wide significance.

### Replication

Examination of ten X chromosome variants previously discovered to have genome-wide significant associations with lung function in prior studies included one variant, rs142755000, that was among the top suggested associations in this meta-analysis XWAS for FEV<sub>1</sub> ( $\beta - 0.134$ , SE 0.029,  $p 4.30 \times 10^{-06}$ ) and FEV<sub>1</sub>/FVC ( $\beta - 0.028$ , SE 0.006,  $p 3.36 \times 10^{-06}$ ), with the effect allele A having a frequency of 0.03 (Additional file 4: Table S3). This was comparable to results of Zhao et al. who found, among 5768 subjects in a COPD-enriched population of White race including COPDGene, that rs142755000 was a genome-wide association for FEV<sub>1</sub> ( $\beta - 0.18$ , SE 0.03,  $p 3.58 \times 10^{-08}$ ), also having effect allele of A with a frequency of 0.03, and a similar appearance to the locus zoom plot [17]. Description of annotation for rs142755000 is in the supplement (Additional file 1).

### Discussion

This meta-analysis is the first comprehensive examination of markers on the X chromosome specifically testing for association with COPD and COPD-related phenotypes of pulmonary function and emphysema in three case–control populations. Using stringent data quality control and statistical methods specific to the X chromosome, we identified a genome-wide significant association with FEV<sub>1</sub>/FVC ratio for one variant near *TMSB4X*. There were additional suggestive associations with COPD-related phenotypes in analyses of all subjects and in sex-stratified XWAS. These findings emphasize the importance of properly including X chromosome variants in association analyses. They also support the important role sex differences play in the pathobiology

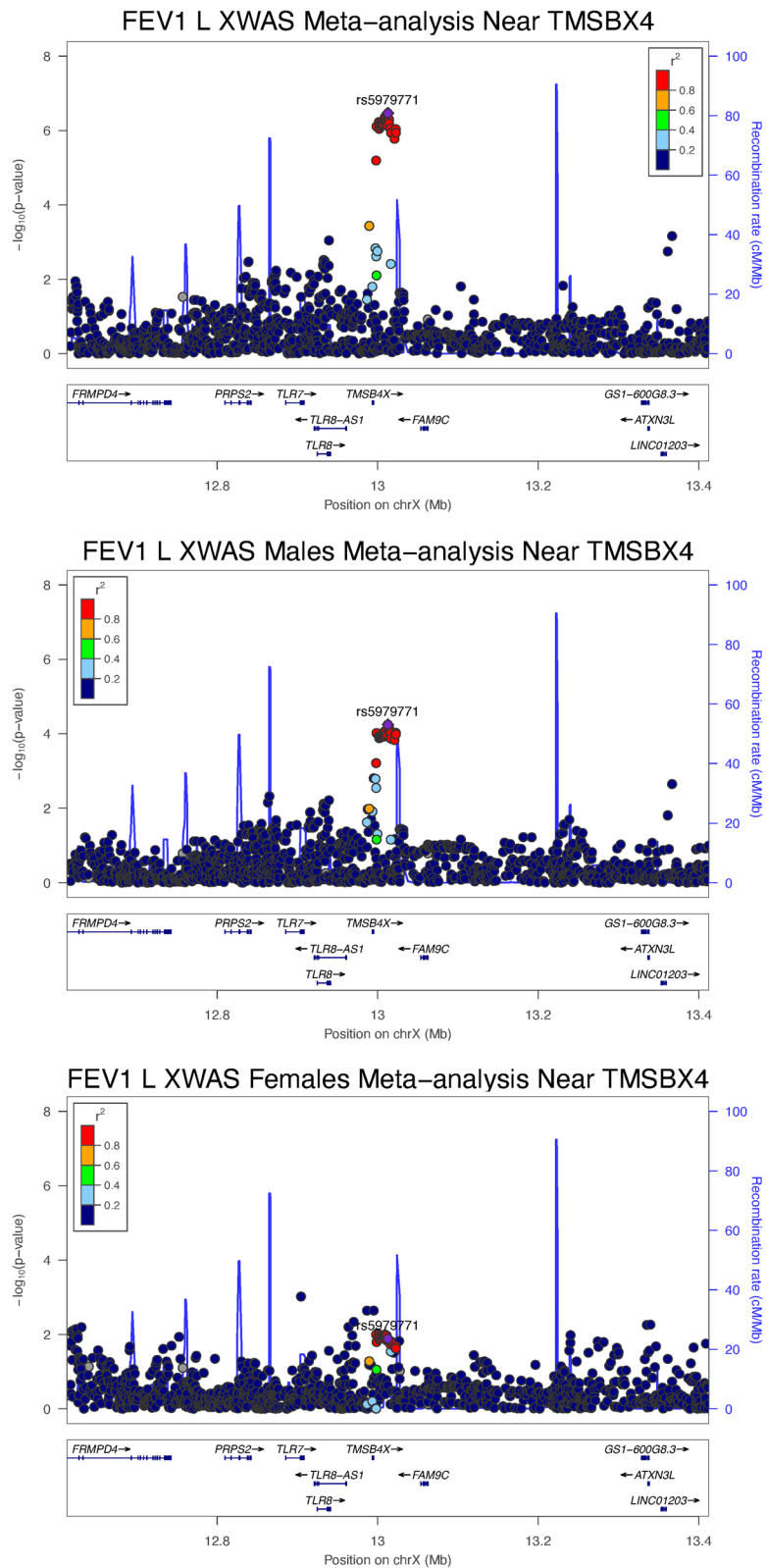
**Table 2** XWAS Meta-analysis Top Suggested Associations in COPD Related Phenotypes

Phenotype	SNP	EA/Other	EAF	Studies	All Subjects			Males			Females			Sex Difference P value	Annotated gene	Locus
					OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value			
COPD	rs138704174	A/G	0.03	3	1.58	(1.31,1.91)	3.55E-06	1.47	(1.16,1.87)	1.54E-03	1.78	(1.29,2.45)	4.32E-04	3.48E-01	SH3KBP1 <sup>†</sup>	Xp22.12
	rs5928132	T/C	0.39	3	0.89	(0.84,0.95)	2.19E-04	0.85	(0.79,0.91)	9.03E-06	0.99	(0.90,1.10)	8.97E-01	1.30E-02	DMD <sup>‡</sup>	Xp21.1
	rs145916556	A/G	0.10	3	0.90	(0.81,1.00)	4.52E-02	1.02	(0.91,1.15)	7.04E-01	0.65	(0.53,0.78)	7.61E-06	4.84E-05	LINC02595	Xp11.3
	rs150086151	T/A	0.01	3	1.55	(1.14,2.12)	5.45E-03	0.87	(0.60,1.25)	4.52E-01	3.83	(2.20,6.66)	2.65E-06	1.18E-05	LOC112268302 <sup>‡</sup>	Xq24
FEV <sub>1</sub> , L	rs5979771	C/T	0.05	3	0.092	0.018	3.40E-07	0.095	0.023	5.70E-05	0.070	0.028	1.30E-02	4.91E-01	TMSB4X	Xp22.2
	rs5927319	G/A	0.25	3	-0.024	0.009	1.07E-02	-0.012	0.012	3.33E-01	-0.063	0.014	9.84E-06	6.43E-03	TMEM47	Xp21.1
	rs5960415	T/C	0.04	3	-0.101	0.022	6.55E-06	-0.099	0.029	5.60E-04	-0.100	0.037	6.58E-03	9.80E-01	PAGE5	Xp11.21
	rs2375465	T/C	0.03	3	-0.105	0.023	7.73E-06	-0.105	0.030	4.79E-04	-0.102	0.039	9.69E-03	9.44E-01	FOXR2 <sup>§</sup>	Xp11.21
	rs139674496	C/T	0.03	3	-0.105	0.023	5.59E-06	-0.102	0.030	5.81E-04	-0.114	0.039	3.40E-03	8.13E-01	RRAGB	Xp11.21
	rs5965560	G/A	0.01	2	0.154	0.036	1.93E-05	0.244	0.049	8.52E-07	-0.047	0.055	3.99E-01	8.64E-05	OPHN1 <sup>‡</sup>	Xq12
	rs14275000	A/G	0.03	3	-0.134	0.029	4.30E-06	-0.149	0.040	1.91E-04	-0.105	0.041	1.17E-02	4.40E-01	HMGNS	Xq21.1
	rs185387095	T/A	0.01	2	-0.187	0.044	2.84E-05	-0.311	0.067	4.29E-06	-0.118	0.060	4.79E-02	3.30E-02	HMGNS <sup>‡</sup>	Xq21.1
	rs145089551	A/G	0.01	2	-0.164	0.042	8.62E-05	-0.310	0.063	1.16E-06	-0.052	0.054	3.42E-01	2.01E-03	SH3BGRL	Xq21.1
	rs6616818	C/G	0.48	3	0.038	0.008	3.85E-06	0.048	0.011	9.66E-06	0.007	0.012	5.93E-01	1.17E-02	POU3F4	Xq21.1
	rs185667977	G/T	0.01	2	0.287	0.059	1.66E-06	0.307	0.071	1.76E-05	-	-	-	-	FMRI <sup>‡</sup>	Xq27.3
	rs5979771	C/T	0.05	3	0.020	0.004	4.97E-08	0.021	0.004	2.57E-06	0.019	0.007	6.58E-03	8.57E-01	TMSB4X	Xp22.2
	rs12015039	A/G	0.06	3	-0.011	0.004	2.01E-03	-0.006	0.004	1.89E-01	-0.028	0.007	5.93E-05	6.09E-03	DMD <sup>‡</sup>	Xp21.1
rs138641025	A/G	0.02	3	-0.020	0.006	9.89E-04	-0.010	0.007	1.75E-01	-0.043	0.011	4.74E-05	9.34E-03	TMEM47	Xp21.1	
rs190159509	T/C	0.05	3	-0.002	0.005	6.50E-01	0.010	0.006	7.82E-02	-0.040	0.010	5.07E-05	1.13E-05	ITM2A	Xq21.1	
rs14275000	A/G	0.03	3	-0.028	0.006	3.36E-06	-0.028	0.007	1.68E-04	-0.028	0.011	7.07E-03	9.81E-01	HMGNS	Xq21.1	
rs185387095	T/A	0.01	2	-0.039	0.009	2.82E-05	-0.060	0.012	2.15E-06	-0.029	0.015	5.51E-02	1.15E-01	HMGNS <sup>‡</sup>	Xq21.1	
rs145089551	A/G	0.01	2	-0.033	0.009	1.50E-04	-0.060	0.012	4.95E-07	-0.013	0.014	3.45E-01	9.55E-03	SH3BGRL	Xq21.1	
rs73230918	A/G	0.05	3	-0.010	0.004	1.90E-02	-0.002	0.005	6.21E-01	-0.031	0.008	8.53E-05	1.82E-03	POU3F4	Xq21.1	
rs5988060	A/G	0.36	3	-0.005	0.002	1.37E-02	-0.001	0.002	8.02E-01	-0.016	0.004	1.65E-05	3.89E-04	HTR2C	Xq23	
rs185903771	A/G	0.04	3	-0.015	0.005	1.22E-03	-0.025	0.005	5.29E-06	0.010	0.008	2.28E-01	4.36E-04	ARHGAP36	Xq26.1	

**Table 2** (continued)

				Beta	SE	P value	Beta	SE	P value	Beta	SE	P value			
Emphysema	rs147204746	G/A	0.01	2	-0.511	0.120	2.36E-05	0.124	4.93E-05	-0.565	0.318	7.48E-02	8.63E-01	<b>FRMPD4<sup>†</sup></b>	<b>Xp22.2</b>
	rs184018661	G/A	0.02	3	-0.130	0.057	2.37E-02	0.060	6.27E-01	-0.485	0.114	2.45E-05	6.53E-05	<b>TAB3</b>	<b>Xp21.2</b>
	rs5905883	T/A	0.33	3	0.072	0.018	9.57E-05	0.020	1.41E-04	0.061	0.040	1.26E-01	7.55E-01	<i>DUSP21</i>	Xp11.3
	rs12845098	G/T	0.17	3	0.022	0.024	3.74E-01	0.026	9.91E-02	0.207	0.052	9.28E-05	2.15E-05	<i>TBX22</i>	Xq21.1
	rs184737280	T/C	0.02	3	-0.216	0.056	1.50E-04	0.062	1.00E-01	-0.461	0.114	6.20E-05	5.66E-03	<i>TCEAL1</i>	Xq22.2
	rs187046350	A/G	0.02	3	-0.090	0.083	2.82E-01	0.088	1.58E-01	-0.768	0.183	3.38E-05	1.23E-05	<i>GUCY2F</i>	Xq22.3
	rs2748634	A/G	0.43	3	0.052	0.016	1.12E-03	0.017	2.61E-06	-0.028	0.033	4.07E-01	3.66E-03	<b>SLITRK2</b>	<b>Xq27.3</b>
	rs149254153	G/T	0.02	3	-0.124	0.067	6.15E-02	0.068	4.93E-01	-0.612	0.147	3.61E-05	4.78E-05	<i>FMRI-AS1</i>	Xq27.3

Summary of the top suggested associations in the XWAS meta-analysis in COPD as well as related phenotypes of spirometric lung function and emphysema measured by quantitative chest CT. Included are the top variants per annotated gene in at least one population strata for a given phenotype with P value  $\leq 10^{-6}$  if present, otherwise  $\leq 10^{-5}$  or less (Note:  $\leq 10^{-5}$  was used for XWAS of all subjects emphysema, females FEV<sub>1</sub>, FEV<sub>1</sub>/FVC, and emphysema). Bold rows are top suggested association for that phenotype and population strata. \* P < 0.05 for significant sex-difference. † The variant is in the gene and intronic. ‡ The variant is in the gene and a coding variant. Abbreviations: XWAS X-chromosome association study; COPD chronic obstructive pulmonary disease; SNP single nucleotide polymorphism; EA effect allele; EAF effect allele frequency; OR odds ratio; CI confidence interval; FEV<sub>1</sub> Forced expiratory volume in one second; L liters; FVC forced vital capacity



**Fig. 2** Meta-analysis locus plots. Plots for rs5979771, the genome-wide association for FEV<sub>1</sub> near TMSB4 X in the XWAS meta-analysis among all subjects and in sex-stratified populations. Abbreviations: COPD chronic obstructive pulmonary disease, XWAS X chromosome association study, FEV<sub>1</sub> forced expiratory volume in one second, L liters, FVC forced vital capacity



of complex lung disease, including factors such as escape from XCI as seen with *TMSB4X*, and sex differences in effect size, as seen among 20 of the top suggested variants in stratified analysis.

Genetic risk factors for COPD have only been partially determined [2]. The X chromosome may contribute additional genetic risk that could explain some of the missing heritability for this complex disease not accounted for by autosomal GWAS results [3, 8]. COPD is a heterogeneous disease that also shows sex specific pathobiology [10]. Females have been suggested to have earlier onset of disease, and greater lung function decline per daily number of cigarettes smoked, while males may have more emphysema by quantitative imaging [9, 11]. An autosomal variant previously studied by Hardin et al. found sex-specific associations with COPD [4]. A larger GWAS by Sakornsakolpat et al. where 82 autosomal COPD variants were identified did not show sex-specific effects; this question may be better answered by examining the X chromosome [3].

The X chromosome has been included in prior lung function GWAS, although not routinely, and in that setting has been included in COPD investigations, though there has not previously been a direct XWAS of COPD and related phenotypes [14–17]. A lung function GWAS including the X chromosome by Soler Artigas et al. identified one variant at genome-wide significance that does not replicate in this study [14]. Wyss and colleagues included the X chromosome in their assessment of lung function and did not find any significant associations [15]. A large GWAS by Shrine et al. looked at a genetic risk score for lung function, including examination of the X chromosome with identification of five variants at genome-wide significance in the UK Biobank; however, these variants did not replicate in SpiroMeta and no variants met the study threshold for inclusion in their genetic risk score [16]. None of these variants replicated in our current XWAS. Most recently, Zhao and colleagues examined pulmonary function and COPD in whole genome sequencing data, identifying four variant signals on the X chromosome for FVC and one for FEV<sub>1</sub>, rs142755000, which was in a COPD-enriched stratum [17]. This FEV<sub>1</sub> variant near *HMGNS* was among the top suggested associations in our current meta-analysis XWAS for FEV<sub>1</sub> and FEV<sub>1</sub>/FVC.

In this analysis we demonstrated the importance of proper inclusion of the X chromosome as it harbors variants with association and suggestive association with COPD and COPD-related phenotypes. A variant near *TMSB4X* reached the genome-wide significance threshold for association with lung function in all subjects for FEV<sub>1</sub>/FVC and showed a suggestive association with FEV<sub>1</sub>. In the emphysema XWAS, the top

suggestive association in all subjects was a variant in *FRMPD4*, located 332 kb upstream of the *TMSB4X* variant in the Xp22.2 locus; both *FRMPD4* and *TMSB4X* escape XCI [31]. *TMSB4X* encodes an actin sequestering protein, thymosin  $\beta$  4, that plays a role in regulation of actin polymerization and is important in organization of the cytoskeleton, as well as being involved in cell proliferation, migration, and differentiation [32]. *TMSB4X* is expressed equivalently in both male and female lung tissue but has been observed to have male-biased expression in skin, adipose, and kidney [33–35]. A comprehensive analysis of transcriptome sequencing data in COPD lung tissue demonstrated decreased expression in COPD compared to normal tissue based on both RNA-seq and quantitative real-time PCR [36]. Thymosin  $\beta$  4 and one of its methionine oxidation products, thymosin  $\beta$  4 sulfoxide, has been found at increased levels in bronchioalveolar lavage fluid of smokers [37]. In smokers, methionine oxidation plays a role in  $\alpha$ (1)-antitrypsin inactivation and pathologic lung remodeling [37, 38]. Thymosin  $\beta$  4 is thought to limit inflammation through autophagy and has been found to have a protective effect in interstitial lung diseases including scleroderma, bleomycin-induced lung damage, and reperfusion-induced acute lung injury [37, 39–41]. A variant near *TMSB4X* has been associated with risk of childhood onset asthma in the UK Biobank; though it is not found in linkage disequilibrium with the variant identified in this study the direction of effect is the same [42, 43].

There were additional suggestive associations found among the all-subject XWAS. In this lung function XWAS, variants from the Xq21.1 locus including in *HMGNS* and near *SH3BGRL* were implicated, which are both expressed in lung tissue [33]. A variant in *HMGNS*, rs185387095, was a top suggested association in this study, and a nearby variant in rs142755000 in linkage disequilibrium was also suggested; rs142755000 was identified in association with FEV<sub>1</sub> in prior whole genome sequence analysis of COPD-enriched White race populations including COPDGene [17]. *HMGNS* modulates cellular transcription and, in a murine model, mutations in *HMGNS* are associated with a lung function phenotype on pulmonary function tests as well as an emphysema-like phenotype [44].

The sex-stratified analysis of COPD-related phenotypes provides additional revealing information and suggests functional associations based on size and direction of effect not identified in XWAS of all subjects. There was a larger effect in females found in sex-stratified testing among top suggested variants: two for COPD, one for FEV<sub>1</sub>, five for FEV<sub>1</sub>/FVC and five for emphysema. This included different directions of effect among one variant for lung function near *ITM2A* and four variants for emphysema near *TAB3*,

*TBX22*, *GUCY2F*, and *FMRI-AS1*. There was a larger effect in males found in sex-stratified testing among top suggested variants: one for COPD, five for lung function, and one for emphysema. This included different directions of effect among two variants for lung function, one in *OPHNI* and one near *ARHGAP36*, as well as one variant for emphysema near *SLITRK2*. Further discussion of suggestive associations can be found in the Additional file 1).

One source of potential disease pathology for some of these variants is escape from XCI [45, 46]. Mammalian females carry two copies of chromosome X, and the majority of X-linked human genes are subject to XCI, where one of the two copies is silenced. However, at least 15–23% of genes escape XCI to some extent, and thus both X chromosome genes are expressed, with a variable continuum of expression likely due to epigenetic effects [31, 34, 47]. XCI is not random, with the majority of escape genes found on the short arm of chromosome X, a region enriched for male-biased and female-biased genes [34, 35, 48]. This is likely related to the evolutionary history of the sex chromosomes where the short arm of the X chromosome is a recent addition to an ancestral chromosome [48]. The impact of XCI is complex and variable across individuals and tissue types, and without functional studies it is not possible to know if the closest annotated genes are causal [34, 35]. *TMSB4X*, the closest gene to the top variant implicated in this analysis, is known to escape XCI, though there is not a significant sex difference in the current analysis. Evidence of escape from XCI has been reported for other genes implicated by suggested associations in or near *DMD*, *HTR2C*, and *GUCY2F*, which demonstrate a significant sex difference in our analysis, as well as *SH3KBP1* and *FRMPD4*, which do not have significant sex differences in the current analysis [31, 32, 34, 35, 49, 50]. Future investigation of gene expression and methylation patterns together with genetic variation may reveal pathologic relevance for these genes that escape XCI despite nominal evidence for sex-specific genetic association.

Escape from XCI results in sex biases in gene expression and has implications for the role of the X chromosome in human diseases, including intellectual disability, autoimmune disease, and cancer [34, 35, 45, 46, 48, 51]. It has been suggested that escape from XCI is a mechanism for respiratory disease pathology in COPD and that escape from XCI influences lung tissue transcription [52]. *POU3F4* was the only suggested variant in the current analysis that is a transcription factor. In our published sex-specific network analysis of gene expression data in normal lung tissue, *POU3F4* demonstrated sexed biased targeting of 15 X chromosome genes, 14 of which escape XCI, including two important XCI regulators, *XIST* and *JPX* (Additional file 1: Figure S5) [53, 54]. This points to

sex-biased genetic variants as one mechanism implicated in sex-biased transcriptional targeting.

We implemented X chromosome-specific genetic association testing for our analyses, using parameters conforming to Clayton's method for analysis to account for XCI in females and male X chromosome hemizyosity [6, 27]. Hickey and colleagues compared Clayton's method to autosomal methods and alternate X chromosome methods and found this approach provided the best power to detect an association [7, 27]. This method has been used in several GWAS that identified significant association between X chromosome markers and complex disease traits [14, 55, 56].

Prior COPD GWAS have excluded genomic information from the X chromosome in their analyses [3, 16, 24]. Although individual reasons for excluding these data are not known, many studies exclude X chromosome genetic variants because of the statistical difficulties inherent to testing for association with the sex chromosomes. We implemented X-specific approaches to both quality control and statistical analysis to improve both accuracy and power to detect an association. Larger studies of X chromosome markers will be required to reliably break the threshold of genome-wide significance on the X for COPD and related phenotypes, but we feel that our results represent an important step in determining associations involved in smoking-related obstructive lung disease.

Our study has several limitations. Since the study cohorts are limited to smokers, we were not able to assess risk variants or effect modification among non-smokers and our results may not generalize to non-smokers. Clayton's method provides a statistical approach to random XCI, but some areas of the X undergo non-random inactivation [27]. Future studies including allele-specific methylation will be helpful to directly investigate the effect of XCI more thoroughly. Variants with minor allele frequency of 1% or greater were included in this analysis, though the allele frequency of some of the suggestive associations was near 1%, which increases the likelihood that these suggestive associations are driven by a small number of subjects; low frequency variants should be interpreted with caution. We were unable to replicate all the previously discovered genome-wide significant associations with lung function, which may be related to differences in phenotypes examined (five variants were for FVC only, which was not examined in the current study), study design, power, and patient populations including the inclusion of current and former smokers only [14, 16, 17]. The top hits described in this study were not replicated in another population, which would be an interesting future direction. Despite these limitations, we did

identify several intriguing biologically plausible genes that could play a potential role in COPD; additional investigation of allosomal genetic variation and lung disease is imperative.

## Conclusions

In summary, these results represent a comprehensive analysis of markers on the X chromosome for association with COPD and related phenotypes including emphysema in three COPD case–control cohorts. We identified one genome-wide significant variant and several promising associations between markers on the X chromosome that may contribute to sex differences in COPD. Among 33 top suggested variants there were 20 unique variants that had significant sex differences in stratified analysis, seven with larger effect in males and thirteen with larger effect in females. Among the 25 top genes implicated with suggestive associations in this study, at least six have evidence for escape from XCI, including the top genome-wide association near *TMSB4X*. XCI may be an important contributor to disease pathology, though the impact of XCI is complex and variable. Detecting variants associated with complex traits is inherently difficult and requires large sample sizes for males and females to address the statistical complexities of studying the X chromosome. Genetic association studies in human lung disease should routinely consider systematic interrogation of X chromosome variants as these may reveal new genes for sex-specific diagnostic and therapeutic approaches to COPD.

## Abbreviations

β	Beta
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
COPDGene	The Genetic Epidemiology of COPD Study
CT	Computed tomography
EA	Effect allele
EAF	Effect allele frequency
ECLIPSE	Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints Study
Emphysema	Quantitative emphysema
FEV <sub>1</sub>	Forced expiratory volume in one second
FVC	Forced vital capacity
GenKOLS	Genetics of COPD in Norway Study
GOLD	Global Initiative for Chronic Obstructive Lung Disease
GWAS	Genome-wide association studies
L	Liters
OR	Odds ratio
QC	Quality control
SD	Standard deviation
SE	Standard error
SNP	Single nucleotide polymorphism
XCI	X chromosome inactivation
XWAS	X-chromosome-wide association study

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12931-023-02337-1>.

**Additional file 1: Text and Figures S1–S5.** Supplementary text for the methods (study phenotype, quality control, imputation, significance and suggestive thresholds, annotation), results (power to detect associations, annotation of rs142755000), and discussion (suggestive associations) as well as related references. **Figure S1.** a: COPDGene XWAS in Stratified Populations All Subjects. b: COPDGene XWAS in Stratified Populations Males. c: COPDGene XWAS in Stratified Populations Females. **Figure S2.** Meta-analysis XWAS in Stratified Population. **Figure S3.** Meta-analysis Locus Plots of Top Suggested Associations in COPD Related Phenotypes. **Figure S4.** Meta-analysis Quantile–Quantile and Manhattan Plots. **Figure S5.** Sex differential edge weights connecting the transcription factor *POU3F4* in GTEx Lung Tissue.

**Additional file 2: Table S1.** XWAS Meta-analysis All Top Suggested Associations At  $10^{-6}$  Or Less If Present, Otherwise Top Suggested Associations At  $10^{-5}$  Or Less

**Additional file 3: Table S2.** XWAS Meta-analysis Top ACE2 Xp22.2 Variants

**Additional file 4: Table S3.** Replication Examination Of Associations From Other Studies In This XWAS Meta-analysis

## Acknowledgements

**Collaborating investigators for the COPDGene® Study:** Ann Arbor VA: Jeffrey L. Curtis, MD; Perry G. Pernicano, MD *Baylor College of Medicine, Houston, TX*: Nicola Hanania, MD, MS; Mustafa Atik, MD; Aladin Boriek, PhD; Kalpatha Guntupalli, MD; Elizabeth Guy, MD; Amit Parulekar, MD; *Brigham and Women's Hospital, Boston, MA*: Dawn L. DeMeo, MD, MPH; Craig Hersh, MD, MPH; Francine L. Jacobson, MD, MPH; George Washko, MD *Columbia University, New York, NY*: R. Graham Barr, MD, DrPH; John Austin, MD; Belinda D'Souza, MD; Byron Thomashow, MD *Duke University Medical Center, Durham, NC*: Neil MacIntyre, Jr., MD; H. Page McAdams, MD; Lacey Washington, MD *HealthPartners Research Institute, Minneapolis, MN*: Charlene McEvoy, MD, MPH; Joseph Tashjian, MD *Johns Hopkins University, Baltimore, MD*: Robert Wise, MD; Robert Brown, MD; Nadia N. Hansel, MD, MPH; Karen Horton, MD; Allison Lambert, MD, MHS; Nirupama Putcha, MD, MHS *Lundquist Institute for Biomedical Innovation at Harbor UCLA Medical Center, Torrance, CA*: Richard Casaburi, PhD, MD; Alessandra Adami, PhD; Matthew Budoff, MD; Hans Fischer, MD; Janos Porszasz, MD, PhD; Harry Rossiter, PhD; William Stringer, MD *Michael E. DeBakey VAMC, Houston, TX*: Amir Sharafkhaneh, MD, PhD; Charlie Lan, DO *Minneapolis VA*: Christine Wendt, MD; Brian Bell, MD; Ken M. Kunisaki, MD, MPH; *Joseph School of Medicine, Atlanta, GA*: Eric L. Flenaugh, MD; Hirut Gebrekristos, PhD; Mario Ponce, MD; Silanath Terpenning, MD; Gloria Westney, MD, MS *National Jewish Health, Denver, CO*: Russell Bowler, MD, PhD; David A. Lynch, MD *Reliant Medical Group, Worcester, MA*: Richard Rosiello, MD; David Pace, MD *Temple University, Philadelphia, PA*: Gerard Criner, MD; David Ciccolella, MD; Francis Cordova, MD; Chandra Dass, MD; Gilbert D'Alonzo, DO; Parag Desai, MD; Michael Jacobs, PharmD; Steven Kelsen, MD, PhD; Victor Kim, MD; A. James Mamary, MD; Nathaniel Marchetti, DO; Aditi Satti, MD; Kartik Shenoy, MD; Robert M. Steiner, MD; Alex Swift, MD; Irene Swift, MD; Maria Elena Vega-Sanchez, MD *University of Alabama, Birmingham, AL*: Mark Dransfield, MD; William Bailey, MD; Surya P. Bhatt, MD; Anand Iyer, MD; Hrudaya Nath, MD; J. Michael Wells, MD *University of California, San Diego, CA*: Douglas Conrad, MD; Xavier Soler, MD, PhD; Andrew Yen, MD *University of Iowa, Iowa City, IA*: Alejandro P. Comellas, MD; Karin F. Hoth, PhD; John Newell, Jr., MD; Brad Thompson, MD *University of Michigan, Ann Arbor, MI*: MeiLan K. Han, MD MS; Ella Kazerooni, MD MS; Wassim Labaki, MD MS; Craig Galban, PhD; Dharshan Vummidi, MD *University of Minnesota, Minneapolis, MN*: Joanne Billings, MD; Abbie Begnaud, MD; Tadashi Allen, MD *University of Pittsburgh, Pittsburgh, PA*: Frank Sciurba, MD; Jessica Bon, MD; Divay Chandra, MD, MSc; Joel Weissfeld, MD, MPH *University of Texas Health, San Antonio, TX*: Antonio Anzueto, MD; Sandra Adams, MD; Diego Maselli-Caceres, MD; Mario E. Ruiz, MD; Harjinder Singh.

### Author contributions

LPH, RB, EKS, THB, NL, CL and DLD have made substantial contributions to the conception or design of the work. DAL, PB, AG, EKS, JDC and DLD contributed to the acquisition of data. LPH, BDH, RB, MHC, ML, CMLR and DLD contributed to analysis or interpretation of the data. LPH, RB and DLD have drafted the work or substantively revised it. LPH prepared Tables 1, 2, Figures S1, S2, Tables S1–S3, Figures S1–S4. CMLR prepared Fig. S5. LPH, BDH, RB, MHC, ML, CMLR, DAS, PB, AG, EKS, JDC, THB, NL, CL, and DLD approved the submitted version and agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature. All authors read and approved the final manuscript.

### Funding

This project was supported by National Institutes of Health (NIH)/National Heart, Lung, and Blood Institute (NHLBI) grants including K23 HL136851 (LPH), K08 HL136928 (BDH), R01 HL089438 (DLD), P01 HL105339 (DLD), P01 HL114501 (EKS), U01 HL089856 (EKS), U01 HL089897 (JDC), R01 HL113264 (MHC and EKS), NIH/National Cancer Institute (NCI) R35 CA220523 (CMLR), and NIH T32HL007427 (CMLR). In addition, this work was supported by a CHEST Foundation 2016 grant in chronic obstructive pulmonary disease (DLD) and from the American Lung Association, LCD-821824 (CMLR). COPDGene is also supported by the COPD Foundation through contributions made to an Industry Advisory Board that has included AstraZeneca, Bayer Pharmaceuticals, Boehringer-Ingelheim, Genentech, GlaxoSmithKline, Novartis, Pfizer, and Sunovion. The Norway GenKOLS study (GSK code RES11080) and the ECLIPSE study (GSK code SCO104960) were funded by GlaxoSmithKline. Neither the NIH nor the Industry Advisory Board had a role in the study design, data collection, data analysis, interpretation of the data, writing of the report, or the decision to submit the paper for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, NHLBI, or NCI. This article reflects the views of the author and should not be construed to represent FDA's views or policies.

### Availability of data and materials

The datasets generated and/or analyzed during the current study are publicly available in dbGaP for COPDGene (accession number phs000179.v6.p2, [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000179.v6.p2](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000179.v6.p2)) and ECLIPSE (accession number phs001252.v1.p1, [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001252.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001252.v1.p1)).

### Declarations

#### Ethics approval and consent to participate

COPDGene obtained approval from the Institutional Review Board at each of the twenty-one clinical sites and all participants provided written informed consent for their medical data to be used prior to taking part in the study. ECLIPSE, GlaxoSmithKline study code SCO104960, was approved by the relevant ethics and review boards at the participating centers and all participants provided written informed consent. GenKOLS, GlaxoSmithKline study code RES11080, was approved by the Regional Committee for Medical Research Ethics (REK Vest), the Norwegian Data Inspectorate and the Norwegian Department of Health, and a written informed consent was obtained from all participants.

COPDGene Clinical Center and IRB protocol numbers:

Ann Arbor VA, Ann Arbor, MI (PCC 2008–110732).  
 Baylor College of Medicine, Houston, TX (H-22209).  
 Brigham and Women's Hospital, Boston, MA (2007-P-000554/2; BWH).  
 Columbia University, New York, NY (IRB-AAAC9324).  
 Duke University Medical Center, Durham, NC (Pro00004464).  
 Health Partners Research Foundation, Minneapolis, MN (07–127).  
 Johns Hopkins University, Baltimore, MD (NA\_00011524).  
 Los Angeles Biomedical Research Institute at Harbor UCLA Medical Center, Los Angeles, CA (12756–01).  
 Michael E. DeBakey VAMC, Houston, TX (H-22202).  
 Minneapolis VA, Minneapolis, MN (4128-A).  
 Morehouse School of Medicine, Atlanta, GA (07–1029).  
 National Jewish Health, Denver, CO (HS-1883a).

Reliant Medical Group, Worcester, MA (1143).  
 Temple University, Philadelphia, PA (11369).  
 University of Alabama, Birmingham, AL (FO70712014).  
 University of California, San Diego, CA (70876).  
 University of Iowa, Iowa City, IA (200710717).  
 University of Michigan, Ann Arbor, MI (HUM00014973).  
 University of Minnesota, Minneapolis, MN (0801M24949).  
 University of Pittsburgh, Pittsburgh, PA (PRO07120059).  
 University of Texas Health Science Center at San Antonio, San Antonio, TX (HSC20070644H).  
 ECLIPSE GSK Investigational Sites:  
 Phoenix, Arizona, United States, 85006.  
 Rancho Mirage, California, United States, 92270.  
 Torrance, California, United States, 90502.  
 Denver, Colorado, United States, 80206.  
 Hartford, Connecticut, United States, 06105.  
 New Haven, Connecticut, United States, 06519.  
 Miami, Florida, United States, 33136.  
 Boston, Massachusetts, United States, 02115.  
 Boston, Massachusetts, United States, 02135.  
 Baltimore, Maryland, United States, 21224.  
 Rochester, Minnesota, United States, 55905.  
 St. Charles, Missouri, United States, 63301.  
 Omaha, Nebraska, United States, 68131.  
 Omaha, Nebraska, United States, 68198.  
 Lebanon, New Hampshire, United States, 03756.  
 Pittsburgh, Pennsylvania, United States, 15213.  
 Providence, Rhode Island, United States, 02903.  
 Houston, Texas, United States, 77030.  
 San Antonio, Texas, United States, 78229.  
 Richmond, Virginia, United States, 23225.  
 Pleven, Bulgaria, 5800.  
 Sofia, Bulgaria, 1000.  
 Vancouver, British Columbia, Canada, V5Z 1M9.  
 Vancouver, British Columbia, Canada, V6Z 1Y6.  
 Halifax, Nova Scotia, Canada, B3H 3A7.  
 Hamilton, Ontario, Canada, L8N 3Z5.  
 Kingston, Ontario, Canada, K7L 2V7.  
 Montreal, Quebec, Canada, H2X 2P4.  
 Sainte-Foy, Quebec, Canada, G1V 4G5.  
 Praha 8, Czech Republic, 182 00.  
 Hvidovre, Denmark, 2650.  
 Horn, Netherlands, 6085 NM.  
 Wellington, New Zealand, 6035.  
 Bergen, Norway, N-5021.  
 Golnik, Slovenia, 4204.  
 Palma de Mallorca, Spain, 07014.  
 Donetsk, Ukraine, 83003.  
 Kiev, Ukraine, 03680.  
 Kiev, Ukraine, 3680.  
 Edinburgh, Midlothian, United Kingdom, EH16 4SA.  
 Cambridge, United Kingdom, CB2 2XY.  
 Liverpool, United Kingdom, L9 7AL.  
 London, United Kingdom, NW3 2QG.  
 Manchester, United Kingdom, M23 9LT.

#### Consent for publication

Not applicable.

#### Competing interests

MHC has received Grant support from GSK and Bayer and speaking and consulting fees from AstraZeneca, Genentech and Illumina. DAL is supported by the NIHR UCLH Biomedical Research Centre and is an NIHR Senior Investigator, received grant support from GlaxoSmithKline, and received honoraria from Grifols and GlaxoSmithKline. PB has received payment or honoraria from AstraZeneca, Novartis, and Chiesi. In the past three years, EKS received grant support from GlaxoSmithKline and Bayer. DLD has received grant support from Bayer and honoraria from Novartis. Authors BDH, RB, ML, CMLR, AG, JDC, THB, NL, and CL have nothing additional to disclose.

**Author details**

<sup>1</sup>Division of Pulmonary Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. <sup>2</sup>Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Ave, Boston, MA 02115, USA. <sup>3</sup>Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>4</sup>Division of Pulmonology, Allergy, and Critical Care, U.S. Food and Drug Administration, Silver Spring, MD, USA. <sup>5</sup>Bioinformatics and Computational Biology Program, Worcester Polytechnic Institute, Worcester, MA, USA. <sup>6</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>7</sup>UCL Respiratory, University College London, London, UK. <sup>8</sup>Department of Clinical Science, University of Bergen, Bergen, Norway. <sup>9</sup>Division of Pulmonary Sciences and Critical Care Medicine, National Jewish Health, Denver, CO, USA. <sup>10</sup>Johns Hopkins School of Public Health, Baltimore, MD, USA.

Received: 21 April 2022 Accepted: 18 January 2023

Published online: 01 February 2023

**References**

- Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010;7(1):32–43.
- Moll M, Sakornsakolpat P, Shrine N, Hobbs BD, DeMeo DL, John C, et al. Chronic obstructive pulmonary disease and related phenotypes: polygenic risk scores in population-based and case-control cohorts. *Lancet Respir Med*. 2020;8(7):696–708.
- Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet*. 2019;51(3):494–505.
- Hardin M, Cho MH, Sharma S, Glass K, Castaldi PJ, McDonald ML, et al. Sex-based genetic association study identifies CELSR1 as a possible chronic obstructive pulmonary disease risk locus among women. *Am J Respir Cell Mol Biol*. 2017;56(3):332–41.
- Konig IR, Loley C, Erdmann J, Ziegler A. How to include chromosome X in your genome-wide association study. *Genet Epidemiol*. 2014;38(2):97–103.
- Clayton D. Testing for association on the X chromosome. *Biostatistics*. 2008;9(4):593–600.
- Hickey PF, Bahlo M. X chromosome association testing in genome wide association studies. *Genet Epidemiol*. 2011;35(7):664–70.
- Wise AL, Gyi L, Manolio TA. eXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am J Hum Genet*. 2013;92(5):643–7.
- Foreman MG, Zhang L, Murphy J, Hansel NN, Make B, Hokanson JE, et al. Early-onset chronic obstructive pulmonary disease is associated with female sex, maternal factors, and African American race in the COPDGene Study. *Am J Respir Crit Care Med*. 2011;184(4):414–20.
- Aryal S, Diaz-Guzman E, Mannino DM. Influence of sex on chronic obstructive pulmonary disease risk and treatment outcomes. *Int J Chron Obstruct Pulmon Dis*. 2014;9:1145–54.
- Hardin M, Foreman M, Dransfield MT, Hansel N, Han MK, Cho MH, et al. Sex-specific features of emphysema among current and former smokers with COPD. *Eur Respir J*. 2016;47(1):104–12.
- Marques CR, Costa GN, da Silva TM, Oliveira P, Cruz AA, Alcantara-Neves NM, et al. Suggestive association between variants in IL1RAPL and asthma symptoms in Latin American children. *Eur J Hum Genet*. 2017;25(4):439–45.
- Chang D, Gao F, Slavney A, Ma L, Waldman YY, Sams AJ, et al. Accounting for eXentricities: analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. *PLoS ONE*. 2014;9(12):e113684.
- Soler Artigas M, Wain LV, Miller S, Kheirallah AK, Huffman JE, Ntalla I, et al. Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat Commun*. 2015;6:8658.
- Wyss AB, Sofer T, Lee MK, Terzikhan N, Nguyen JN, Lahousse L, et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat Commun*. 2018;9(1):2976.
- Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019;51(3):481–93.
- Zhao X, Qiao D, Yang C, Kasela S, Kim W, Ma Y, et al. Whole genome sequence analysis of pulmonary function and COPD in 19,996 multi-ethnic participants. *Nat Commun*. 2020;11(1):5182.
- Tcheandjieu C, Zhu X, Hilliard AT, Clarke SL, Napolioni V, Ma S, et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat Med*. 2022;28(8):1679–92.
- Hayden LP, Busch R, Hobbs BD, Cho MH, Lomas DA, Bakke P, et al. X chromosome genetic associations in COPD. *Am J Respir Crit Care Med*. 2019;199:A4868.
- Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, et al. Evaluation of COPD longitudinally to identify predictive surrogate endpoints (ECLIPSE). *Eur Respir J*. 2008;31(4):869–73.
- Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet*. 2009;5(3):e1000421.
- Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary. *Am J Respir Crit Care Med*. 2017;195(5):557–82.
- Nakano Y, Muro S, Sakai H, Hirai T, Chin K, Tsukino M, et al. Computed tomographic measurements of airway dimensions and emphysema in smokers. Correlation with lung function. *Am J Respir Crit Care Med*. 2000;162(3 Pt 1):1102–8.
- Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet*. 2010;42(3):200–2.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–9.
- Wang J, Yu R, Shete S. X-chromosome genetic association test accounting for X-inactivation, skewed X-inactivation, and escape from X-inactivation. *Genet Epidemiol*. 2014;38(6):483–93.
- Randall JC, Winkler TW, Kutalik Z, Berndt SI, Jackson AU, Monda KL, et al. Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet*. 2013;9(6):e1003500.
- Takahashi T, Ellingson MK, Wong P, Israelow B, Lucas C, Klein J, et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature*. 2020;139:788.
- Leung JM, Niikura M, Yang CWT, Sin DD. COVID-19 and COPD. *Eur Respir J*. 2020;56(2):2002108.
- 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.
- NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2016;44(D1):D7–19.
- Consortium GT. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45(6):580–5.
- 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.
- 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).
- Kim WJ, Lim JH, Lee JS, Lee SD, Kim JH, Oh YM. Comprehensive analysis of transcriptome sequencing data in the lung tissues of COPD subjects. *Int J Genomics*. 2015;2015:206937.
- De Santis M, Inzitari R, Bosello SL, Peluso G, Fanali C, Iavarone F, et al. beta-Thymosins and interstitial lung disease: study of a scleroderma cohort with a one-year follow-up. *Respir Res*. 2011;12:22.
- Taggart C, Cervantes-Laurean D, Kim G, McElvaney NG, Wehr N, Moss J, et al. Oxidation of either methionine 351 or methionine 358 in alpha

- 1-antitrypsin causes loss of anti-neutrophil elastase activity. *J Biol Chem.* 2000;275(35):27258–65.
39. Conte E, Iemmolo M, Fruciano M, Fagone E, Gili E, Genovese T, et al. Effects of thymosin beta4 and its N-terminal fragment Ac-SDKP on TGF-beta-treated human lung fibroblasts and in the mouse model of bleomycin-induced lung fibrosis. *Expert Opin Biol Ther.* 2015;15(Suppl 1):S211–21.
  40. Yaman OM, Guner I, Guntas G, Sonmez OF, Tanriverdi G, Cakiris A, et al. Protective effect of thymosin beta4 against abdominal aortic ischemia-reperfusion-induced acute lung injury in rats. *Medicina (Kaunas).* 2019;55(5):187.
  41. Renga G, Oikonomou V, Stincardini C, Pariano M, Borghi M, Costantini C, et al. Thymosin beta4 limits inflammation through autophagy. *Expert Opin Biol Ther.* 2018;18(sup1):171–5.
  42. Ferreira MAR, Mathur R, Vonk JM, Szwajda A, Brumpton B, Granell R, et al. Genetic architectures of childhood- and adult-onset asthma are partly distinct. *Am J Hum Genet.* 2019;104(4):665–84.
  43. 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.
  44. Kugler JE, Horsch M, Huang D, Furusawa T, Rochman M, Garrett L, et al. High mobility group N proteins modulate the fidelity of the cellular transcriptional profile in a tissue- and variant-specific manner. *J Biol Chem.* 2013;288(23):16690–703.
  45. Laffont S, Guery JC. Deconstructing the sex bias in allergy and autoimmunity: from sex hormones and beyond. *Adv Immunol.* 2019;142:35–64.
  46. Balaton BP, Brown CJ. Escape artists of the X chromosome. *Trends Genet.* 2016;32(6):348–59.
  47. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature.* 2005;434(7031):400–4.
  48. Disteché CM. Escapees on the X chromosome. *Proc Natl Acad Sci USA.* 1999;96(25):14180–2.
  49. 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.
  50. Wainer Katsir K, Linial M. Human genes escaping X-inactivation revealed by single cell expression data. *BMC Genomics.* 2019;20(1):201.
  51. Posnyck BJ, Brown CJ. Escape from X-chromosome inactivation: an evolutionary perspective. *Front Cell Dev Biol.* 2019;7:241.
  52. Choi S, Lee S, Qiao D, Hardin M, Cho MH, Silverman EK, et al. FARVATX: family-based rare variant association test for X-linked genes. *Genet Epidemiol.* 2016;40(6):475–85.
  53. Lopes-Ramos CM, Chen CY, Kuijjer ML, Paulson JN, Sonawane AR, Fagny M, et al. Sex differences in gene expression and regulatory networks across 29 human tissues. *Cell Rep.* 2020;31(12): 107795.
  54. Naciri I, Lin B, Webb CH, Jiang S, Carmona S, Liu W, et al. Linking chromosomal silencing with xist expression from autosomal integrated transgenes. *Front Cell Dev Biol.* 2021;9: 693154.
  55. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet.* 2009;41(6):703–7.
  56. Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet.* 2012;44(7):770–6.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

