SOX5 is a candidate gene for COPD susceptibility and is necessary for lung development

Craig P. Hersh, Edwin K. Silverman, Jody Gascon, Soumyaroop Bhattacharya, Barbara J. Klanderman, Augusto A. Litonjua, Veronique Lefebvre, David Sparrow, John J. Reilly, David A. Lomas, Thomas J. Mariani

Online data supplement

Methods

Study Subjects

The National Emphysema Treatment Trial (NETT) was a randomized clinical trial of lung volume reduction surgery in patients with emphysema and severe airflow obstruction (1). NETT subjects had $FEV_1 \leq 45\%$ predicted, hyperinflation on pulmonary function tests, and emphysema on chest CT scan. Subjects in the NETT Genetics Ancillary Study were recruited from 16 of 17 NETT clinical centers and provided a blood sample for genetic analysis. Since NETT did not enroll subjects without COPD, we compared NETT COPD cases to control smokers from the Normative Aging Study (NAS), a longitudinal study of aging in initially healthy men conducted by the Boston Veterans Administration (2). In this analysis, we included NAS subjects with a smoking history of at least 10 pack-years with normal spirometry at the last study visit (FEV_1>80% predicted, FEV_1/FVC>90% predicted). From the full set of 389 NETT and 472 NAS subjects (Table 1), 386 NETT and 424 NAS subjects with adequate DNA samples were included in the case-control analysis of the fine mapping Illumina SNP panel.

The Boston Early-Onset COPD (BEOCOPD) Study is a family-based study (extended pedigrees) of COPD genetics; subject enrollment and phenotype determination have been described previously (3). Probands with a diagnosis of COPD had FEV_1<40% predicted at an age<53 years, without severe α1-antitrypsin deficiency All first-degree relatives and older second-degree relatives were invited to participate.

The International COPD Genetics Network (ICGN) was a multi-center family-based study (predominantly sib-pairs) conducted in the U.S. and Europe (4, 5). Probands were age 45-65 with FEV_1<60% predicted, FEV_1/VC<90% predicted, a smoking history of at least 5 pack-years, and at least one sibling with at least a 5 pack-year smoking history. Computerized image analysis of chest CT scans was used to determine the percent emphysema at -950HU and the square root wall area of a 10mm airway.

The distributions of post-bronchodilator FEV_1 (% predicted) of the NETT subjects, BEOCOPD probands and ICGN probands are shown in Figure E1 in the online supplement.

Human studies were approved by institutional review boards at Partners Healthcare and other participating centers. Subjects provided written informed consent.

SNP Selection and Genotyping

Genomewide linkage analysis in the Boston Early-Onset COPD Study identified a region on chromosome 12p with linkage to post-bronchodilator FEV_1, in smokers only, that nearly reached criteria for genomewide significance (LOD [logarithm of the odds of linkage] score = 3.26 at 36cM)(6). This region also showed suggestive evidence of linkage to moderate airflow obstruction (FEV_1<60% predicted with FEV_1/FVC<90% predicted) as a qualitative trait (7). The 1.5 LOD-drop interval (akin to a 95% confidence interval) ranged from 10.2 – 25.8 Mb on the human genome map (NCBI build 36). Based on data from CEPH Caucasians in the International HapMap Project (8), we used a linkage disequilibrium tagging algorithm (r^2>0.8, minor allele frequency 10%) implemented in Tagger (9) to select a set of 1534 SNPs to tag common genetic variants across the region. SNPs were genotyped in the NETT-NAS case-control study using custom-designed Illumina (San Diego, CA) GoldenGate assays. In NETT-NAS, we had previously genotyped a panel of 195 intergenic SNPs throughout the genome, excluding regions linked to COPD and found no evidence of population stratification (10). To follow-up significant results from NETT-NAS, selected SNPs were genotyped in the BEOCOPD study and the ICGN using Sequenom (San Diego, CA) or TaqMan (Applied Biosystems [ABI], Foster City, CA) assays.

Statistical Genetics Analysis

In the NETT-NAS case-control study, SNPs were analyzed using logistic regression under an additive genetic model, without covariate adjustment, using PLINK version 1.0.7 (11). Secondary

analyses were adjusted for age and pack-years of smoking. There was no adjustment for sex, since all NAS controls were male. In the BEOCOPD and ICGN family-based studies, SNPs were analyzed for association with COPD status under an additive model, without covariate adjustment, using the extended pedigree family-based association test implemented in Golden Helix (Bozeman, MT) PBAT version 6.4.3 (12). In the BEOCOPD study, SNPs were tested under the null hypothesis of linkage but no association. Fisher's method was used to combine p-values across studies (13). In secondary analyses in all studies, measured FEV_1 was modeled as a continuous trait, adjusted for age, sex, height, and pack-years of smoking.

DNA Sequencing

In 23 probands from the BEOCOPD Study and 1 CEPH control subject, we sequenced the 14 exons and corresponding intron-exon boundaries in SOX5, as well as 10 highly conserved regions in the 3' end of the gene, due to the 3' location of rs11046966. Conserved regions were selected using PhastCons scores (14), implemented in the UCSC Genome Browser (15). We used dye-labeled dideoxy sequencing reactions and an ABI 3730 DNA sequencing machine. Sequence tracings were analyzed with Phred, Phrap, and Consed (16, 17), and polymorphisms were identified using PolyPhred (18) and by manual review. SNPs found on more than one chromosome (MAF>2%) during sequencing were genotyped in NETT-NAS and BEOCOPD using Sequenom or TaqMan assays.

Sox5 Null Mouse

All animal experiments were conducted in accordance with the University of Rochester Animal Care and Use Policy and following an approved animal studies protocol. Mice harboring an allele containing a mutated form of the Sox5 gene (19) in a mixed C57BL/6x129/SvEv background were obtained from Dr. V. Lefebvre (Lerner Research Institute, Cleveland, OH). These mice were bred to

generate Sox5 deficient ($Sox5^{-/-}$), heterozygous ($Sox5^{+/-}$) and wild type offspring. Timed matings were performed and offspring were harvested for analysis at 16.5, 17.5 or 18.5 days of embryonic gestation (E) by euthanization of the pregnant dam. Genotypes were defined using embryonic tissues as previously described (19). Embryos were dissected to isolate the entire thoracic cavity or individual lung lobes and fixed in buffered formalin for histological analysis or snap frozen in liquid nitrogen for molecular analysis. Formalin-fixed tissues were embedded in paraffin and lung morphology was assessed in 5 um sections by H&E staining. RNA was isolated from E17.5 lung tissue using the Trizol method (Invitrogen, Carlsbad, CA). Crude RNA preparations were re-purified and rendered DNA-free using the Stratagene Absolutely RNA kit (Santa Clara, CA), according to the manufacturer's protocol. DNA-free RNA was reverse transcribed into cDNA using using oligo dT primers and MultiScribe reverse transcriptase (Applied Biosystems, Foster City, CA), and subjected to quantitative Real-Time PCR (qPCR) as previously described (20). Relative steady-state mRNA expression levels for Sox5 (Forward 5'- ATC AAC GGA GAG ATT TAC GAG GA -3'; Reverse 5'- CCG CAA TGT GGT TTT CGC T -3') and fibronectin (Forward 5'- TGG ATA GCA CCC AGT GTT CAG -3'; Reverse 5'- CCT GTC TTC TCT TTC GGG TTC A -3') were defined using cyclophillin A (PPIA; Forward 5'- GGT GGT GAC TTT ACA CGC CA -3'; Reverse 5'- TCT CCG TAG ATG GAC CTG CC -3') as a constitutively expressed housekeeping gene using the ddCt method.

References

E1. Fishman A, Martinez F, Naunheim K, Piantadosi S, Wise R, Ries A, Weinmann G, Wood DE. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003;348:2059-2073.

E2. Bell B, Rose CL, Damon H. The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. *Aging Hum Dev* 1972;3:5-17.

E3. Silverman EK, Chapman HA, Drazen JM, Weiss ST, Rosner B, Campbell EJ, O'Donnell WJ, Reilly JJ, Ginns L, Mentzer S, et al. Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis. *Am J Respir Crit Care Med* 1998;157:1770-1778.

E4. Patel BD, Coxson HO, Pillai SG, Agusti AG, Calverley PM, Donner CF, Make BJ, Muller NL, Rennard SI, Vestbo J, et al. Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2008;178:500-505.

E5. Zhu G, Warren L, Aponte J, Gulsvik A, Bakke P, Anderson WH, Lomas DA, Silverman EK, Pillai SG. The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. *Am J Respir Crit Care Med* 2007;176:167-173.

E6. DeMeo DL, Celedon JC, Lange C, Reilly JJ, Chapman HA, Sylvia JS, Speizer FE, Weiss ST, Silverman EK. Genome-wide linkage of forced mid-expiratory flow in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2004;170:1294-1301.

E7. Silverman EK, Mosley JD, Palmer LJ, Barth M, Senter JM, Brown A, Drazen JM, Kwiatkowski DJ, Chapman HA, Campbell EJ, et al. Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. *Hum Mol Genet* 2002;11:623-632.

E8. International HapMap Consortium. The International HapMap Project. *Nature* 2003;426:789-796.

E9. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet* 2005;37:1217-1223.

E10. Hersh CP, Hansel NN, Barnes KC, Lomas DA, Pillai SG, Coxson HO, Mathias RA, Rafaels NM, Wise RA, Connett JE, et al. Transforming growth factor-beta receptor-3 is associated with pulmonary emphysema. *Am J Respir Cell Mol Biol* 2009;41:324-331.

E11. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-575.

E12. Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM. PBAT: tools for family-based association studies. *Am J Hum Genet* 2004;74:367-369.

E13. Fisher RA. Statistical Methods for Research Workers. London: Oliver and Boyd, Ltd.; 1925.

E14. Felsenstein J, Churchill GA. A Hidden Markov Model approach to variation among sites in rate of evolution. *Mol Biol Evol* 1996;13:93-104.

E15. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res* 2002;12:996-1006.

E16. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998;8:175-185.

E17. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res* 1998;8:195-202.

E18. Nickerson DA, Tobe VO, Taylor SL. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 1997;25:2745-2751.

E19. Smits P, Li P, Mandel J, Zhang Z, Deng JM, Behringer RR, de Crombrugghe B, Lefebvre V. The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. *Dev Cell* 2001;1:277-290.
E20. Srisuma S, Bhattacharya S, Simon DM, Solleti SK, Tyagi S, Starcher B, Mariani TJ. Fibroblast growth factor receptors control epithelial-mesenchymal interactions necessary for alveolar elastogenesis. *Am J Respir Crit Care Med* 2010;181:838-850.

Table E1: Single nucleotide polymorphisms identified by sequencing of SOX5 exons and conserved regions in 24 subjects.

Position*	Major allele	Minor allele	Minor Allele Frequency	Known SNP (dbSNP)	Role
chr12:23994847	A	G	30%	rs7297174	5' UTR isoform b
chr12:23994746	A	G	39%	rs10505917	5' UTR isoform b
chr12:23939971	T	A	2%		intron
chr12:23890155	G	A	13%	rs7980561	intron
chr12:23854813	G	A	8%		intron
chr12:23854722	G	A	41%	rs7308759	intron
chr12:23854639	A	G	6%		intron
chr12:23785314	A	G	2%		intron
chr12:23785020	C	T	22%	rs11047075	intron
chr12:23779069	G	A	2%	rs16926690	intron
chr12:23747053	G	A	2%		intron
chr12:23717275	T	C	26%	rs10771018	intron
chr12:23716930	T	C	2%		intron
chr12:23716674	T	C	2%	rs11047037	intron
chr12:23716564	G	A	10%	rs7970223	intron
chr12:23684931	C	G	2%		intron
chr12:23666594	T	A	10%		intron
chr12:23666083	G	C	13%	rs10842193	intron
chr12:23648512	A	G	45%	rs1479452	intron
chr12:23645476	A	G	10%	rs10771006	intron
chr12:23645194	A	C	6%	rs11608710	intron
chr12:23645024	A	C	4%		intron
chr12:23609175	A	G	2%		intron
chr12:23605876	G	C	10%	rs17381981	intron
chr12:23605703	A	G	39%	rs10505893	intron
chr12:23587383	A	G	45%	rs7485662	intron
chr12:23580568	G	A	15%	rs4636755	intron
chr12:23578621	C	T	2%		exon, synonymous
chr12:23577108	T	C	15%		3' UTR
chr12:23576923	T	C	15%		3' UTR

*Chromosomal locations based on NCBI build 36

Table E2: Results from genetic association analyses of single nucleotide polymorphism rs11046966.
nt=not tested

A) National Emphysema Treatment Trial-Normative Aging Study

	Unadjusted models		**Adjusted models**		
Phenotype	**OR**	**p-value**	**Covariates**	**OR or β**	**p-value**
Case-control status	1.48	0.00060	age, pack-years	1.49	0.0017
FEV_1 post-bronchodilator*	nt		age, sex, pack-years, height	-0.0006	0.97
CT emphysema (-950 HU)*	nt		age, sex, pack-years, weight	-0.005	0.59

*Continuous traits were analyzed in NETT cases only

B) Boston Early-Onset COPD Study

	Unadjusted models	**Adjusted models**	
Phenotype	**p-value**	**Covariates**	**p-value**
COPD status	1.45e-05	age, sex, pack-years	7.0e-5
Severe COPD (GOLD 3-4)	2.1e-5	age, sex, pack-years	1.2e-4
FEV_1 post-bronchodilator	nt	age, sex, pack-years, height	0.0043
Current and former smokers only			
COPD status	1.4e-5	age, sex, pack-years	3.9e-5
FEV_1 post-bronchodilator	nt	age, sex, pack-years, height	2.5e-4

C) International COPD Genetics Network

	Unadjusted models	**Adjusted models**	
Phenotype	**p-value**	**Covariates**	**p-value**
COPD status	0.68	age, sex, pack-years	0.63
Severe COPD (GOLD 3-4)	0.54	age, sex, pack-years	0.94
FEV_1 post-bronchodilator	nt	age, sex, pack-years, height	0.79
CT emphysema (-950 HU)	nt	age, sex, pack-years, current smoking, weight	0.16
CT airway thickening	nt	age, sex, pack-years, current smoking, weight	0.37

Table E3: Single nucleotide polymorphisms (SNPs) in linkage disequilibrium with SOX5 SNP rs11046966, based on SNP data in Caucasians from phase 3 of the International HapMap project (8). SNPs with r^2>0.66 are shown.

SNP	Chrom 12 location*	r^2 with rs11046966	SOX5 role
rs11046966	23568959	--	3'
rs10505891	23569189	0.744	3'
rs11046968	23572538	1	3'
rs12227654	23580022	0.705	intron
rs17466501	23590913	0.744	intron
rs7137550	23604665	0.667	intron
rs7969423	23612911	0.667	intron
rs12229765	23621122	0.667	intron

*Chromosomal locations based on NCBI build 36

Figure E1: Distributions of post-bronchodilator values for FEV_1 % predicted in: (A) National Emphysema Treatment Trial (NETT) COPD cases, (B) Boston Early-Onset COPD Study (BEOCOPD) probands, and (C) International COPD Genetics Network (ICGN) probands.