Online Data Supplement

Plasma Soluble Receptor for Advanced Glycation Endproducts in Idiopathic Pulmonary Fibrosis

Ani Manichaikul, Li Sun, Alain C. Borczuk, Suna Onengut-Gumuscu, Emily A. Farber, Susan K. Mathai, Weiming Zhang, Ganesh Raghu, Joel D. Kaufman, Karen D. Hinckley-Stukovsky, Steven M. Kawut, Sanja Jelic, Wen Liu, Tasha E. Fingerlin, David A. Schwartz, Jessica L. Sell, Stephen S. Rich, R. Graham Barr, David J. Lederer

SUPPLEMENTARY METHODS

We prospectively enrolled adults with ILD presenting to the interstitial lung disease and lung transplant programs at Columbia University Medical Center between 2007 and 2011. For the current analysis, we selected a convenience sample with available plasma and/or DNA for the following studies.

MESA Study Design: MESA is a longitudinal study of subclinical cardiovascular disease and risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease (1). Between 2000 and 2002, MESA recruited 6,814 men and women 45 to 84 years of age from Forsyth County, North Carolina; New York City; Baltimore; St. Paul, Minnesota; Chicago; and Los Angeles. Exclusion criteria were clinical cardiovascular disease, weight exceeding 136 kg (300 lb.), pregnancy, and impediment to long-term participation.
The MESA Family Study recruited 1,595 African American and Hispanic participants, generally siblings of MESA participants, using the same inclusion and exclusion criteria as MESA except that clinical cardiovascular disease was permitted. The MESA Air Pollution Study recruited an additional 257 participants from Los Angeles and Riverside County, CA, and Rockland County, NY, using the same criteria as MESA, except that participants were ages 50 to 89 who lived in the area $\geq$ 50% of the year and had no plans to move in the next five years (2).

sRAGE measurement: We measured plasma levels of soluble RAGE using commercially available ELISA kits (R&D Systems) in 291 participants the Columbia ILD Study. We excluded 6 with unreliable RAGE measurement and 5 with missing covariate data, leaving 280 for analysis. We further measured sRAGE levels in 21 healthy controls for comparison.

ILD case genotyping: Plasma and buffy coat DNA were obtained at the time of a baseline study visit and stored at -80^{o}C until analysis. There were 364 participants (representing Whites, African Americans, Hispanics and Asians) who had DNA available. DNA was genotyped using the Axiom Biobank chip. Genotyping of rs2070600 was carried out on the Affymetrix Axiom Biobank chip. Genotypes were called according to the APT Best practices workflow as recommended by the manufacturer (Affymetrix, Santa Clara, CA). SNPs were filtered on call rate > 95%, and the SNP with lower call rate was removed for all pairs of duplicates based on RSID. We further applied a filter of HWE *P*-value > 1E-5 as calculated for samples pooled across race/ethnic groups, as well as in the stratified sample of Whites. A total of 553,114 SNPs remained after QC, including rs2070600. Samples were filtered for call rate > 97%, heterozygosity > 18% and sex mismatch. These high-throughput genotypes were used to carry out relationship inference in KING (3), and one individual was removed from each pair of first degree relatives to construct a subset of unrelated individuals. After applying these individual level filters for genotype QC, we retained data for 349 participants.

Genotyping in MESA: The SNP rs2070600 was genotyped on the HumanExome BeadChip v.1.0 (Illumina) with genotypes called through a multi-cohort CHARGE effort at the University of Texas Health Science Center at Houston (4, 5). In MESA, rs2070600 passed variant-level quality control filters including missing > 5%, and Hardy-Weinberg Equilirbium (HWE) *P*-values < 1E-5 in each of the four MESA race/ethnic groups. Sample level quality control (QC) included checking concordance to GWAS data and excluding those individuals missing >5% genotypes, population clustering outliers, individuals with high inbreeding coefficients or heterozygote rates, individuals with gender mismatches, and one individual from duplicate pairs. IMPUTE version

2.2.2 was used to perform imputation for the MESA SHARe participants using the cosmopolitan 1,000 Genomes Phase 1 v3 March 2012 reference set.

Genetic association analysis of ILD cases and race/ethnicity-matched MESA controls: Before matching the ILD cases with controls the eligible MESA controls, we further used KING (3) to verify there were no unexpected first degree relatives across these two sets of samples. To limit the presence of ILD among the selected controls, we restricted selection of controls to those MESA participants free of self-reported chronic lung diseases other than asthma. PCs of ancestry were computed within race/ethnic groups using up to 53,243 SNPs common to the Affymetrix Axiom Biobank and the Affymetrix 6.0 SNP arrays used to genotype ILD cases and MESA participants, respectively, and those participants identified as outliers in PCs analysis were excluded. Each ILD case was then matched to three eligible MESA controls of the same race/ethnicity. After restricting the ILD cases to those with self-reported White, African American, Hispanic or Asian race/ethnicity for concordance with MESA (6 participants removed), pruning an additional 27 cases identified as outliers in race/ethnic-specific principal components of ancestry, there were 316 ILD cases remaining for genetic analyses.

Analysis of ILD cases vs. race/ethnicity-matched MESA controls was performed by logistic regression with adjustment for sex and race/ethnic-specific PCs among Whites, representing the largest group of cases in the Columbia ILD study. We further performed pooled analysis of ILD cases and controls across all race/ethnic groups using logistic regression with adjustment for sex and PCs of ancestry computed for the pooled set of cases and controls across race/ethnic groups. We used Cochran's Q test as implemented in the package R/metafor

(6) to examine evidence of heterogeneity with respect to diagnosis of IPF versus non-IPF forms of ILD.

Gene and protein expression analyses: We measured mRNA expression of *AGER* along with the housekeeping gene GAPDH in OCT-embedded fresh frozen lung tissue obtained from 15 adults with IPF and a histologic usual interstitial pneumonia pattern and 15 adults without lung disease stored in the Columbia University Pathology Tissue Bank. RNA was isolated using the RNeasy Plus Mini Kit (Qiagen) and reverse transcribed to cDNA using the High Capacity RNA to cDNA Kit (LifeTechnologies). Quantitative PCR was performed using the StepOnePlus real-time PCR system (LifeTechnologies) in triplicate for 40 cycles. Mean fold changes in gene expression between IPF cases and controls were calculated using the ddCT method. Statistical significance was determined using Wilcoxon rank sum tests. We also performed immunohistochemical staining for RAGE using standard methods in IPF and control lung tissue.

Table E1: Characteristics of participants in with sRAGE measured and with available *AGER* genotyping

Characteristic	sRAGE participants	Genetic Participants
Age, years	59.6 ± 10.9	57.3 ± 11.8
Female	43%	43%
Race/ethnicity		
White	76%	75%
Black	11%	16%
Hispanic	6%	5%
Asian	6%	5%
Diagnosis		
Idiopathic Pulmonary Fibrosis	37%	34%
Other ILD	63%	66%
Forced vital capacity, % predicted	57.0 ± 18.9	55.4 ± 17.9
Diffusing capacity for carbon monoxide, % predicted	33.5 ± 11.3	32.5 ± 11.7
Body mass index, kg/m^2	28.5 ± 5.4	27.6 ± 5.4

Data are mean ± standard deviation and percentage.

Table E2: Analysis of rs2070600 in ILD cases vs. MESA controls.

Group	EAF*		N	N cases	Beta	SE	Conditional Odds ratio (95% CI)	*P*-value
	Case	Control						
All race/ethnic groups (pooled)								
IPF	0.042	0.028	432	108	0.412	0.407	1.51 (0.68, 3.36)	0.31
Other forms of ILD only	0.048	0.031	672	168	0.541	0.331	1.71 (0.90, 3.29)	0.10
Pooled	0.044	0.031	1264	316	0.450	0.240	1.57 (0.98, 2.51)	0.06
White only								
IPF	0.049	0.029	368	92	0.397	0.419	1.49 (0.65, 3.38)	0.34
Other forms of ILD only	0.065	0.037	432	108	0.779	0.363	2.18 (1.07, 4.43)	0.03
Pooled	0.057	0.035	912	228	0.574	0.257	1.78 (1.07, 2.94)	0.03

*EAF = effect allele frequency

Results for rs2070600 shown based on an additive genetic model for the effect allele A (vs. other allele G). Analyses are based on logistic regression of race/ethnicity-matched cases and controls with adjustment for sex and PCs of ancestry.

Table E3. Spearman correlation coefficients between plasma sRAGE and age, body mass index, and measures of disease severity

Variable	IPF		other ILDs	
	Spearman's rho	p-value	Spearman's rho	p-value
Age	0.22	0.02	0.11	0.13
BMI	-0.12	0.22	-0.16	0.04
FVC%	0.46	<0.001	0.27	<0.001
DLCO%	0.19	0.13	0.24	0.01
6MWD	0.36	0.02	0.17	0.09
GAP score	-0.23	0.07	-0.04	0.64

Figure E1. Boxplots of serum sRAGE levels in Columbia ILD participants with IPF stratified by (A) smoking status (Wilcoxon rank-sum p-value = 0.30), (B) presence of gastroesophageal reflux disease (GERD; p = 0.81), (C) IPF vs. combined pulmonary fibrosis and emphysema (CPFE; p = 0.36), and (D) use of immunosuppressive or immunomodulatory therapy (IS; p < 0.001). Panels A, B, and D exclude participants with CPFE.

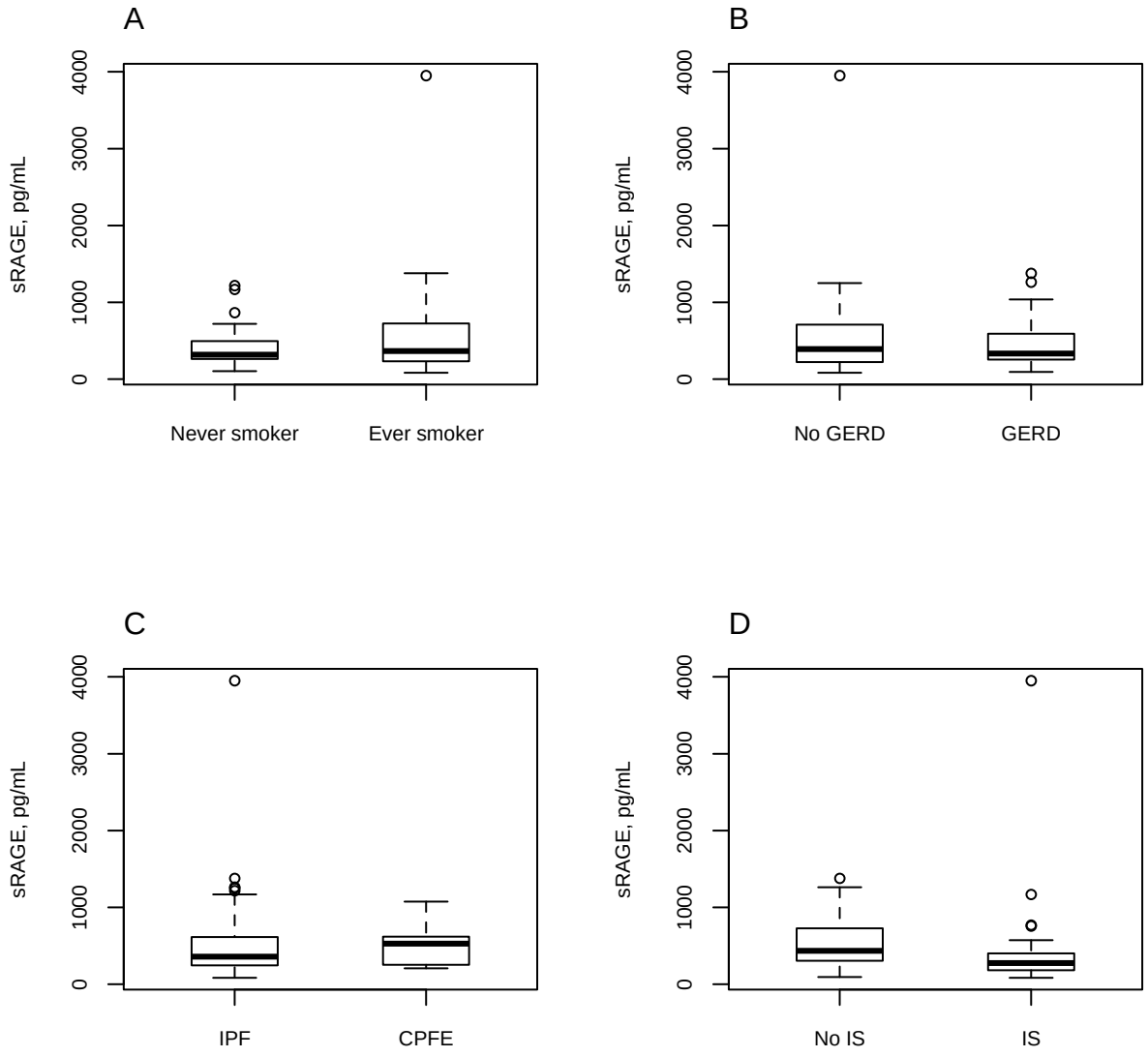
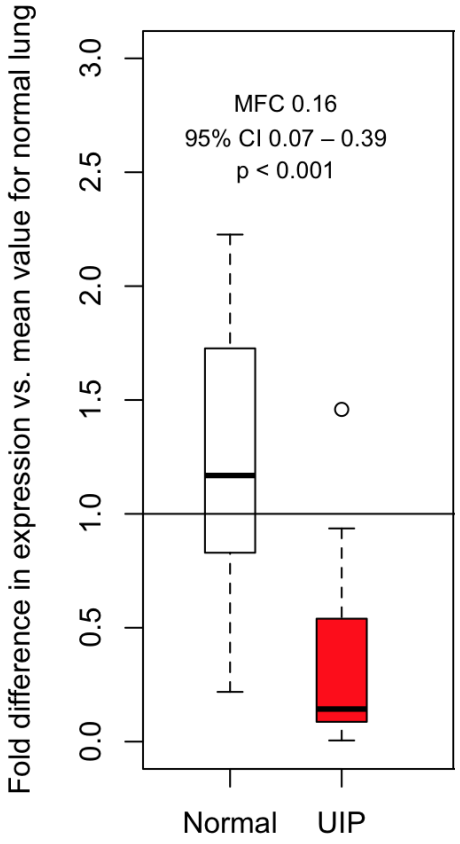


Figure E2: *AGER* mRNA expression in whole lung tissue from 15 cases with IPF (red) and 15 controls (white) using quantitative real-time PCR normalized to expression of the reference gene *GADPH*. Results are presented as boxplots of the fold difference in expression in each case or control normalized to the mean expression value among controls. Within each boxplot, the thick horizontal line represents the median fold difference, the ends of the boxplots are placed at the 25th and 75th percentiles (interquartile range), the whiskers extend to 1.5 x the interquartile range, and an outlier is represented by an open circle. Mean fold change (MFC) was calculated using the ΔΔCt method. The *P*-value is from a Wilcoxon rank sum test. The horizontal line at a value of 1 is the mean normalized expression value among controls.



SUPPLEMENTAL REFERENCES

1. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR, Jr., Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol* 2002; 156: 871-881.

2. Kaufman JD, Adar SD, Allen RW, Barr RG, Budoff MJ, Burke GL, Casillas AM, Cohen MA, Curl CL, Daviglus ML, Diez Roux AV, Jacobs DR, Jr., Kronmal RA, Larson TV, Liu SL, Lumley T, Navas-Acien A, O'Leary DH, Rotter JI, Sampson PD, Sheppard L, Siscovick DS, Stein JH, Szpiro AA, Tracy RP. Prospective study of particulate air pollution exposures, subclinical atherosclerosis, and clinical cardiovascular disease: The Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Am J Epidemiol* 2012; 176: 825-837.

3. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010; 26: 2867-2873.

4. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, Uitterlinden AG, Harris TB, Witteman JC, Boerwinkle E. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* 2009; 2: 73-80.

5. Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD, Hansen M, Borecki IB, Cupples LA, Fornage M, Gudnason V, Harris TB, Kathiresan S, Kraaij R, Launer LJ, Levy D, Liu Y, Mosley T, Peloso GM, Psaty BM, Rich SS, Rivadeneira F, Siscovick DS, Smith AV, Uitterlinden A, van Duijn CM, Wilson JG, O'Donnell CJ, Rotter JI, Boerwinkle E. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PloS one* 2013; 8: e68095.

6. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw* 2010; 36: 1-48.