Online Data Supplement

Variation in Cilia Protein Genes and Progression of Lung Disease in Cystic Fibrosis

Elizabeth Blue, Tin L. Louie, Jessica X. Chong, Scott J. Hebbring, Kathleen C. Barnes, Nicholas M. Rafaels, Michael R. Knowles, Ronald L. Gibson, Michael J. Bamshad, Mary J. Emond, and NHLBI GO Exome Sequencing Project, Lung GO

MATERIALS AND METHODS

The Discovery Cohort

Relative to the larger and unselected Canadian Consortium for Genetic Studies (CGS; N = 1,357) CF cohort (E1), the threshold for the lower tertile in this data set is near the 20th percentile, and the upper tertile threshold is approximately the 80th percentile for SaKnorm in the CGS. Note that SaKnorm is normally distributed in a random sample of persons with CF (E1), but our samples were selected to include individuals having extreme lung function phenotypes, making our distribution bimodal.

Controls were drawn from the National Heart, Lung, and Blood Institute (NHLBI) "Grand Opportunity" (GO) Exome Sequencing Project (ESP) Lung and Heart Cohorts (E2). Because the CF cases are also from the ESP collection, these control exomes have the desirable property of contemporaneous sequencing within the same umbrella project. To create a proper control set, we excluded individuals with lung disease, all but one sibling, individuals with non-European ancestry, and PC outliers. The excluded projects and their database of Genotype and Phenotype (dbGaP) accession numbers are: Cystic Fibrosis (phs000254), Acute Lung Injury (ALI; phs000334), COPD Genetic Epidemiology (COPDGene; phs000296), Lung Health Study of Chronic Obstructive Pulmonary Disease (LHS; phs000291), Pulmonary Arterial Hypertension (PAH; phs000290), and Severe Asthma Research Project (SARP; phs000422). The exome sequence data and phenotype data are available through dbGaP, with the accession numbers for each NHLBI-GO ESP project summarized on the ESP website.

PCs were estimated by smartPCA v.9003 (E3, E4) from all autosomal SNVs passing the minimum read depth filter. Outliers were removed iteratively by smartPCA to achieve a homogeneous data set (Supplemental Figures E1 and E2). We restricted association testing to autosomal or pseudo-autosomal code-altering and UTR single nucleotide variants (SNVs), excluding synonymous variants and splash bases annotated only as "intronic", in order to minimize our multiple testing penalty and focus on the subset of variants most likely to be functional. The candidate gene list was based on a published PCD candidate gene list (E5) containing 82 genes that was updated to include 12 additional genes for PCD identified since its publication: *ARMC4* (E6, E7), *CCDC39* and *CCDC40* (E8, E9), *CCDC65* (E10, E11), *CCDC103* (E12), *CCDC164* (E13), *DNAAF3* (C19orf51 (E14)), *DNAAF5/HEATR2* (E15), *DYX1C1* (E16, E17), *LRRC6* (E18, E19), *SPAG1* (E20), and *ZMYND10* (E21, E22).

Samples with small numbers can cause test statistics to provide biased p-values, as the accuracy of these p-values is based on laws of large numbers that ensure a Gaussian distribution for the appropriate statistic in large samples. The adjusted-SKAT-O approach is designed specifically to avoid this problem (E23), allowing for non-Gaussian test statistic distributions. We used adjusted-SKAT-O (v0.81) to perform by-variant association tests. Variants were tested for association if genotypes were called for at least 65 cases and 130 controls, with minor allele counts >2. Analysis parameters include method="optimal.adj", missing_cutoff=0.9, kernel = "linear", and the inclusion of PCs 1 and 2 as covariates. The upper tertile was compared to 3,148 ESP controls (excludes subjects from ALI, CF, COPDGene, LHS, PAH, and SARP studies). The lower tertile was compared to 3,269 controls (3,148 plus COPDGene “super healthy lung” controls).

The Bonferroni approach to correct for multiple testing was highly conservative, as we tested multiple highly-correlated variants within the same gene. The number of independent tests were estimated using the Genetic type I error calculator(E24), which were then used for the second multiple testing correction approach. For all the subjects together, the total estimated effective number of independent markers is 785.2 for the 1213 markers, with an effective ratio of 0.65 on all chromosomes. This gives a corrected p-value threshold of 6.37×10^{-5}.

Extension Cohorts

Variants associated with variation in lung function among persons with CF could also be associated with outcomes in respiratory diseases with phenotypes overlapping CF, particularly if the associated variants are causal. Hence, finding an association with a related phenotype in the same direction of effect provides additional evidence of causality. We explored this line of evidence using two approaches. First, we tested for association between the candidate variants and the extreme tertiles of percent of predicted forced expiratory volume in 1 second (FEV_1) in 338 persons with COPD in the Lung Health Study. We calculated the percent predicted value of FEV_1, using the observed pre-bronchodilator values and the Global Lung Function Initiative formula (E25). We excluded 19 individuals with reactive airways, defined as a ≥12% change in FEV_1 after taking a bronchodilator, then defined lower and upper tertiles as described above. PCs were estimated as described with these subjects and the ESP controls defined above. We restricted the analysis to persons of self-reported European ancestry, consistent with their PC results. We performed by-variant adjusted-SKAT-O analyses for variants found to be

significantly associated with SaKnorm extremes in the discovery cohort, adjusting for the first two PCs as in the CF discovery analysis.

Second, we tested for association between variants found to be significantly associated with SaKnorm extremes in the CF discovery cohort and the following sub-phenotypes in an independent data set: chronic sinusitis, chronic bronchitis, and bronchiectasis. Phenotypes were defined by diagnostic ICD9 coding extracted from Marshfield Clinic's Personalized Medicine Research Project (PMRP). PMRP is a cohort of approximately 20,000 adult participants, 99% Caucasian with over 70% claiming German ancestry, and most linked to over 30 years of electronic health record data (E26). Cases were defined by those with a specific code (e.g., ICD9 473, chronic sinusitis) whereas controls were defined by those without the code (E27-31). Genotype data from 7418 subjects were generated using the Illumina Human Core Exome Chip. One variant of interest, rs17522489, was not present on the Exome Chip and was instead represented by rs12032942, a variant in strong LD in European populations (r^2=0.94). Both a Chi-squared test and a Fisher's Exact test were performed for each variant. For the Chi-squared test, sex and completeness of electronic health record data was included as covariates.

Acknowledgements

The authors thank the families for their participation, and specifically the EPIC study site investigators and research coordinators for their work: LungGO (includes EPIC Observational Study sites) includes M.J. Bamshad (Seattle Children's Hospital, University of Washington), K.C. Barnes (Johns Hopkins University), I. Abdulhamid (Children's Hospital of Michigan), F. Accurso

(University of Colorado), R. Anbar (Upstate Medical University), T. Beaty (Johns Hopkins University), A. Bigham (University of Washington), P. Black (Children's Mercy Hospital), E. Bleecker (Wake Forest University), K. Buckingham (University of Washington), A.M. Cairns (Maine Medical Center), D. Caplan (Emory University), B. Chatfield (University of Utah), A. Chidekel (A.I. Dupont Institute Medical Center), M. Cho (Brigham and Women's Hospital, Harvard Medical School), D.C. Christiani (Massachusetts General Hospital), J.D. Crapo (National Jewish Health), J. Crouch (Seattle Children's Hospital), D. Daley (University of British Columbia), A. Dang (University of North Carolina Chapel Hill), H. Dang (University of North Carolina Chapel Hill), A. De Paula (Ochsner Health System), J. DeCelie-Germana (Schneider Children's Hospital), A. Dozor (New York Medical College, Westchester Medical Center), M. Drumm (University of North Carolina Chapel Hill), M. Dyson (Cook Children's Med. Center), J. Emerson (Seattle Children's Hospital, University of Washington), M.J. Emond (University of Washington), T. Ferkol (St. Louis Children's Hospital, Washington University School of Medicine), R. Fink (Children's Medical Center of Dayton), C. Foster (Johns Hopkins University), D. Froh (University of Virginia), L. Gao (Johns Hopkins University), W. Gershan (Children's Hospital of Wisconsin), R.L. Gibson (Seattle Children's Hospital, University of Washington), E. Godwin (University of North Carolina Chapel Hill), M. Gondor (All Children's Hospital Cystic Fibrosis Center), H. Gutierrez (University of Alabama at Birmingham), N.N. Hansel (Johns Hopkins University, Johns Hopkins University School of Public Health), P.M. Hassoun (Johns Hopkins University), P. Hiatt (Texas Children's Hospital), J.E. Hokanson (University of Colorado), M. Howenstine (Indiana University, Riley Hospital for Children), L.K. Hummer (Johns Hopkins University), J. Kanga (University of Kentucky), Y. Kim (National Human Genome Research Institute), M.R. Knowles

(University of North Carolina Chapel Hill), M. Konstan (Rainbow Babies & Children's Hospital), T. Lahiri (Vermont Children's Hospital at Fletcher Allen Health Care), N. Laird (Harvard School of Public Health), C. Lange (Harvard School of Public Health), L. Lin (Harvard Medical School), T.L. Louie (University of Washington), D. Lynch (National Jewish Health), B. Make (National Jewish Health), T.R. Martin (University of Washington, VA Puget Sound Medical Center), S.C. Mathai (Johns Hopkins University), R.A. Mathias (Johns Hopkins University), J. McNamara (Children's Hospitals and Clinics of Minnesota), S. McNamara (Seattle Children's Hospital), D. Meyers (Wake Forest University), S. Millard (DeVos Children's Butterworth Hospital, Spectrum Health Systems), P. Mogayzel (Johns Hopkins University), R. Moss (Stanford University), T. Murray (Johns Hopkins University), D. Nielson (University of California at San Francisco), B. Noyes (Cardinal Glennon Children's Hospital), W. O'Neal (University of North Carolina Chapel Hill), D. Orenstein (Children's Hospital of Pittsburgh), B. O'Sullivan (University of Massachusetts Memorial Health Care), R. Pace (University of North Carolina Chapel Hill), P. Pare (St. Paul's Hospital, University of Washington), H.W. Parker (Dartmouth-Hitchcock Medical Center, New Hampshire Cystic Fibrosis Center), M.A. Passero (Rhode Island Hospital), E. Perkett (Vanderbilt University), A. Prestridge (Children's Memorial Hospital), N.M. Rafaels (Johns Hopkins University), B. Ramsey (Seattle Children's Hospital, University of Washington), E. Regan (National Jewish Health), C. Ren (University of Rochester), G. Retsch-Bogart (University of North Carolina Chapel Hill), M. Rock (University of Wisconsin Hospital and Clinics), A. Rosen (Johns Hopkins University), M. Rosenfeld (Seattle Children's Hospital, University of Washington), I. Ruczinski (Johns Hopkins University School of Public Health), A. Sanford (University of British Columbia), D. Schaeffer (Nemours Children's Clinic), C. Sell (University of North Carolina Chapel

Hill), D. Sheehan (Children's Hospital of Buffalo), E.K. Silverman (Brigham and Women's Hospital, Harvard Medical School), D. Sin (Children's Medical Center of Dayton), T. Spencer (Children's Hospital Boston), J. Stonebraker (University of North Carolina Chapel Hill), H.K. Tabor (Seattle Children's Hospital, University of Washington), L. Varlotta (St. Christopher's Hospital for Children), C.I. Vergara (Johns Hopkins University), F. Wigley (Johns Hopkins University), R.A. Wise (Johns Hopkins University), F.A. Wright (University of North Carolina Chapel Hill), M.M. Wurfel (Harvard School of Public Health, Massachusetts General Hospital, University of Washington), R. Zanni (Monmouth Medical Center), F. Zou (University of North Carolina Chapel Hill).

Supplemental References

E1. Wright FA, Strug LJ, Doshi VK, Commander CW, Blackman SM, Sun L, Berthiaume Y, Cutler D, Cojocaru A, Collaco JM, Corey M, Dorfman R, Goddard K, Green D, Kent JW, Jr., Lange EM, Lee S, Li W, Luo J, Mayhew GM, Naughton KM, Pace RG, Pare P, Rommens JM, Sandford A, Stonebraker JR, Sun W, Taylor C, Vanscoy LL, Zou F, Blangero J, Zielenski J, O'Neal WK, Drumm ML, Durie PR, Knowles MR, Cutting GR. Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. *Nat Genet* 2011; 43: 539-546.

E2. Auer PL, Reiner AP, Wang G, Kang HM, Abecasis GR, Altshuler D, Bamshad MJ, Nickerson DA, Tracy RP, Rich SS, Project NGES, Leal SM. Guidelines for Large-Scale Sequence-Based Complex Trait Association Studies: Lessons Learned from the NHLBI Exome Sequencing Project. *Am J Hum Genet* 2016; 99: 791-801.

E3. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006; 2: e190.

E4. 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: 904-909.

E5. Berg JS, Evans JP, Leigh MW, Omran H, Bizon C, Mane K, Knowles MR, Weck KE, Zariwala MA. Next generation massively parallel sequencing of targeted exomes to identify genetic mutations in primary ciliary dyskinesia: implications for application to clinical testing. *Genet Med* 2011; 13: 218-229.

E6. Onoufriadis A, Shoemark A, Munye MM, James CT, Schmidts M, Patel M, Rosser EM, Bacchelli C, Beales PL, Scambler PJ, Hart SL, Danke-Roelse JE, Sloper JJ, Hull S, Hogg C, Emes RD, Pals G, Moore AT, Chung EM, Uk10K, Mitchison HM. Combined exome and whole-genome sequencing identifies mutations in ARMC4 as a cause of primary ciliary dyskinesia with defects in the outer dynein arm. *J Med Genet* 2014; 51: 61-67.

E7. Hjeij R, Lindstrand A, Francis R, Zariwala MA, Liu X, Li Y, Damerla R, Dougherty GW, Abouhamed M, Olbrich H, Loges NT, Pennekamp P, Davis EE, Carvalho CM, Pehlivan D, Werner C, Raidt J, Kohler G, Haffner K, Reyes-Mugica M, Lupski JR, Leigh MW, Rosenfeld M, Morgan LC, Knowles MR, Lo CW, Katsanis N, Omran H. ARMC4 mutations cause primary ciliary dyskinesia with randomization of left/right body asymmetry. *Am J Hum Genet* 2013; 93: 357-367.

E8. Antony D, Becker-Heck A, Zariwala MA, Schmidts M, Onoufriadis A, Forouhan M, Wilson R, Taylor-Cox T, Dewar A, Jackson C, Goggin P, Loges NT, Olbrich H, Jaspers M, Jorissen M, Leigh MW, Wolf WE, Daniels ML, Noone PG, Ferkol TW, Sagel SD, Rosenfeld M, Rutman A, Dixit A, O'Callaghan C, Lucas JS, Hogg C, Scambler PJ, Emes RD, Uk10k, Chung EM, Shoemark A, Knowles MR, Omran H, Mitchison HM. Mutations in CCDC39 and CCDC40 are the major cause of primary ciliary dyskinesia with axonemal disorganization and absent inner dynein arms. *Hum Mutat* 2013; 34: 462-472.

E9. Blanchon S, Legendre M, Copin B, Duquesnoy P, Montantin G, Kott E, Dastot F, Jeanson L, Cachanado M, Rousseau A, Papon JF, Beydon N, Brouard J, Crestani B, Deschildre A, Desir J, Dollfus H, Leheup B, Tamalet A, Thumerelle C, Vojtek AM, Escalier D, Coste A, de Blic J, Clement A, Escudier E, Amselem S. Delineation of CCDC39/CCDC40 mutation spectrum and associated phenotypes in primary ciliary dyskinesia. *J Med Genet* 2012; 49: 410-416.

E10. Austin-Tse C, Halbritter J, Zariwala MA, Gilberti RM, Gee HY, Hellman N, Pathak N, Liu Y, Panizzi JR, Patel-King RS, Tritschler D, Bower R, O'Toole E, Porath JD, Hurd TW, Chaki M, Diaz KA, Kohl S, Lovric S, Hwang DY, Braun DA, Schueler M, Airik R, Otto EA, Leigh MW, Noone PG, Carson JL, Davis SD, Pittman JE, Ferkol TW, Atkinson JJ, Olivier KN, Sagel SD, Dell SD, Rosenfeld M, Milla CE, Loges NT, Omran H, Porter ME, King SM, Knowles MR, Drummond IA, Hildebrandt F. Zebrafish Ciliopathy Screen Plus Human Mutational Analysis Identifies C21orf59 and CCDC65 Defects as Causing Primary Ciliary Dyskinesia. *Am J Hum Genet* 2013; 93: 672-686.

E11. Horani A, Brody SL, Ferkol TW, Shoseyov D, Wasserman MG, Ta-shma A, Wilson KS, Bayly PV, Amirav I, Cohen-Cymberknoh M, Dutcher SK, Elpeleg O, Kerem E. CCDC65 mutation causes primary ciliary dyskinesia with normal ultrastructure and hyperkinetic cilia. *PLoS One* 2013; 8: e72299.

E12. Panizzi JR, Becker-Heck A, Castleman VH, Al-Mutairi DA, Liu Y, Loges NT, Pathak N, Austin-Tse C, Sheridan E, Schmidts M, Olbrich H, Werner C, Haffner K, Hellman N, Chodhari R, Gupta A, Kramer-Zucker A, Olale F, Burdine RD, Schier AF, O'Callaghan C, Chung EM, Reinhardt R, Mitchison HM, King SM, Omran H, Drummond IA. CCDC103 mutations cause primary ciliary dyskinesia by disrupting assembly of ciliary dynein arms. *Nat Genet* 2012; 44: 714-719.

E13. Wirschell M, Olbrich H, Werner C, Tritschler D, Bower R, Sale WS, Loges NT, Pennekamp P, Lindberg S, Stenram U, Carlen B, Horak E, Kohler G, Nurnberg P, Nurnberg G, Porter ME, Omran H. The

nexin-dynein regulatory complex subunit DRC1 is essential for motile cilia function in algae and humans. *Nat Genet* 2013; 45: 262-268.

E14. Mitchison HM, Schmidts M, Loges NT, Freshour J, Dritsoula A, Hirst RA, O'Callaghan C, Blau H, Al Dabbagh M, Olbrich H, Beales PL, Yagi T, Mussaffi H, Chung EM, Omran H, Mitchell DR. Mutations in axonemal dynein assembly factor DNAAF3 cause primary ciliary dyskinesia. *Nat Genet* 2012; 44: 381-389, S381-382.

E15. Horani A, Druley TE, Zariwala MA, Patel AC, Levinson BT, Van Arendonk LG, Thornton KC, Giacalone JC, Albee AJ, Wilson KS, Turner EH, Nickerson DA, Shendure J, Bayly PV, Leigh MW, Knowles MR, Brody SL, Dutcher SK, Ferkol TW. Whole-exome capture and sequencing identifies HEATR2 mutation as a cause of primary ciliary dyskinesia. *Am J Hum Genet* 2012; 91: 685-693.

E16. Tarkar A, Loges NT, Slagle CE, Francis R, Dougherty GW, Tamayo JV, Shook B, Cantino M, Schwartz D, Jahnke C, Olbrich H, Werner C, Raidt J, Pennekamp P, Abouhamed M, Hjeij R, Kohler G, Griese M, Li Y, Lemke K, Klena N, Liu X, Gabriel G, Tobita K, Jaspers M, Morgan LC, Shapiro AJ, Letteboer SJ, Mans DA, Carson JL, Leigh MW, Wolf WE, Chen S, Lucas JS, Onoufriadis A, Plagnol V, Schmidts M, Boldt K, Uk10K, Roepman R, Zariwala MA, Lo CW, Mitchison HM, Knowles MR, Burdine RD, Loturco JJ, Omran H. DYX1C1 is required for axonemal dynein assembly and ciliary motility. *Nat Genet* 2013; 45: 995-1003.

E17. Casey JP, McGettigan PA, Healy F, Hogg C, Reynolds A, Kennedy BN, Ennis S, Slattery D, Lynch SA. Unexpected genetic heterogeneity for primary ciliary dyskinesia in the Irish Traveller population. *Eur J Hum Genet* 2015; 23: 210-217.

E18. Kott E, Duquesnoy P, Copin B, Legendre M, Dastot-Le Moal F, Montantin G, Jeanson L, Tamalet A, Papon JF, Siffroi JP, Rives N, Mitchell V, de Blic J, Coste A, Clement A, Escalier D, Toure A, Escudier E, Amselem S. Loss-of-function mutations in LRRC6, a gene essential for proper

axonemal assembly of inner and outer dynein arms, cause primary ciliary dyskinesia. *Am J Hum Genet* 2012; 91: 958-964.

E19. Horani A, Ferkol TW, Shoseyov D, Wasserman MG, Oren YS, Kerem B, Amirav I, Cohen-Cymberknoh M, Dutcher SK, Brody SL, Elpeleg O, Kerem E. LRRC6 mutation causes primary ciliary dyskinesia with dynein arm defects. *PLoS One* 2013; 8: e59436.

E20. Knowles MR, Ostrowski LE, Loges NT, Hurd T, Leigh MW, Huang L, Wolf WE, Carson JL, Hazucha MJ, Yin W, Davis SD, Dell SD, Ferkol TW, Sagel SD, Olivier KN, Jahnke C, Olbrich H, Werner C, Raidt J, Wallmeier J, Pennekamp P, Dougherty GW, Hjeij R, Gee HY, Otto EA, Halbritter J, Chaki M, Diaz KA, Braun DA, Porath JD, Schueler M, Baktai G, Griese M, Turner EH, Lewis AP, Bamshad MJ, Nickerson DA, Hildebrandt F, Shendure J, Omran H, Zariwala MA. Mutations in SPAG1 cause primary ciliary dyskinesia associated with defective outer and inner dynein arms. *Am J Hum Genet* 2013; 93: 711-720.

E21. Zariwala MA, Gee HY, Kurkowiak M, Al-Mutairi DA, Leigh MW, Hurd TW, Hjeij R, Dell SD, Chaki M, Dougherty GW, Adan M, Spear PC, Esteve-Rudd J, Loges NT, Rosenfeld M, Diaz KA, Olbrich H, Wolf WE, Sheridan E, Batten TF, Halbritter J, Porath JD, Kohl S, Lovric S, Hwang DY, Pittman JE, Burns KA, Ferkol TW, Sagel SD, Olivier KN, Morgan LC, Werner C, Raidt J, Pennekamp P, Sun Z, Zhou W, Airik R, Natarajan S, Allen SJ, Amirav I, Wieczorek D, Landwehr K, Nielsen K, Schwerk N, Sertic J, Kohler G, Washburn J, Levy S, Fan S, Koerner-Rettberg C, Amselem S, Williams DS, Mitchell BJ, Drummond IA, Otto EA, Omran H, Knowles MR, Hildebrandt F. ZMYND10 is mutated in primary ciliary dyskinesia and interacts with LRRC6. *Am J Hum Genet* 2013; 93: 336-345.

E22. Moore DJ, Onoufriadis A, Shoemark A, Simpson MA, zur Lage PI, de Castro SC, Bartoloni L, Gallone G, Petridi S, Woollard WJ, Antony D, Schmidts M, Didonna T, Makrythanasis P, Bevillard J, Mongan NP, Djakow J, Pals G, Lucas JS, Marthin JK, Nielsen KG, Santoni F, Guipponi M, Hogg C, Antonarakis SE, Emes RD, Chung EM, Greene ND, Blouin JL, Jarman AP, Mitchison HM.

Mutations in ZMYND10, a gene essential for proper axonemal assembly of inner and outer dynein arms in humans and flies, cause primary ciliary dyskinesia. *Am J Hum Genet* 2013; 93: 346-356.

E23. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Team NGESP-ELP, Christiani DC, Wurfel MM, Lin X. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* 2012; 91: 224-237.

E24. Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet* 2012; 131: 747-756.

E25. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, Enright PL, Hankinson JL, Ip MS, Zheng J, Stocks J, Initiative ERSGLF. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012; 40: 1324-1343.

E26. McCarty CA, Wilke RA, Giampietro PF, Wesbrook SD, Caldwell MD. Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a large population-based biobank. *Personalized Medicine* 2005; 2: 49-79.

E27. Liu J, Ye Z, Mayer JG, Hoch BA, Green C, Rolak L, Cold C, Khor SS, Zheng X, Miyagawa T, Tokunaga K, Brilliant MH, Hebbring SJ. Phenome-wide association study maps new diseases to the human major histocompatibility complex region. *J Med Genet* 2016; 53: 681-689.

E28. Ye Z, Mayer J, Ivacic L, Zhou Z, He M, Schrodi SJ, Page D, Brilliant MH, Hebbring SJ. Phenome-wide association studies (PheWASs) for functional variants. *Eur J Hum Genet* 2015; 23: 523-529.

E29. Hebbring SJ, Schrodi SJ, Ye Z, Zhou Z, Page D, Brilliant MH. A PheWAS approach in studying HLA-DRB1*1501. *Genes Immun* 2013; 14: 187-191.

E30. Hebbring SJ. The challenges, advantages and future of phenome-wide association studies. *Immunology* 2014; 141: 157-165.

E31. Hebbring SJ, Rastegar-Mojarad M, Ye Z, Mayer J, Jacobson C, Lin S. Application of clinical text data for phenome-wide association studies (PheWASs). *Bioinformatics* 2015; 31: 1981-1987.

Figure E1: Principal components 1 (X-axis) and 2 (Y-axis) for the discovery sample lower SaKnorm tertile (red) and controls (blue).

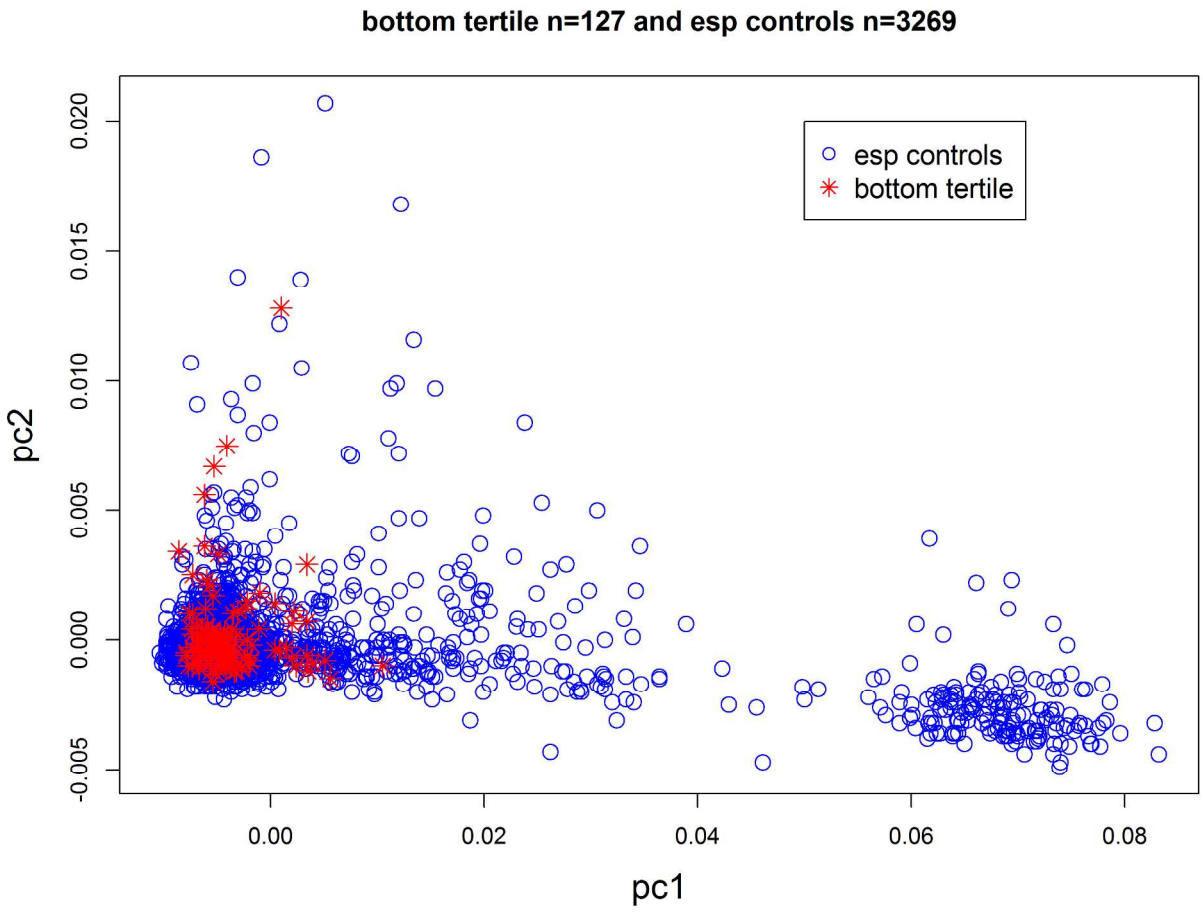

Figure E2: Principal components 1 (X-axis) and 2 (Y-axis) for the discovery sample upper SaKnorm tertile (red) and controls (blue).

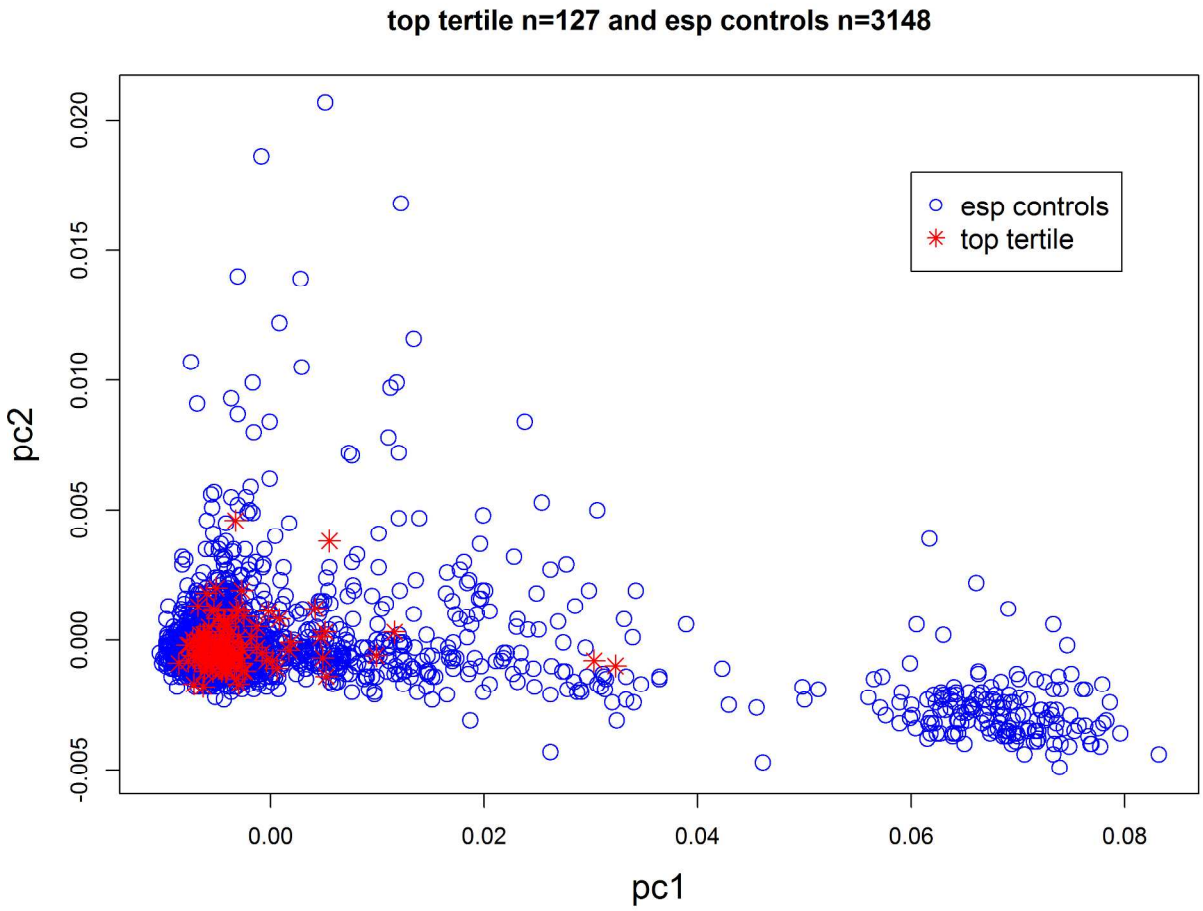

Figure E3: Q-Q plot for adjusted SKAT-O analysis of lower SaKnorm tertile in discovery cohort.

The red line is the observed =expectation line.

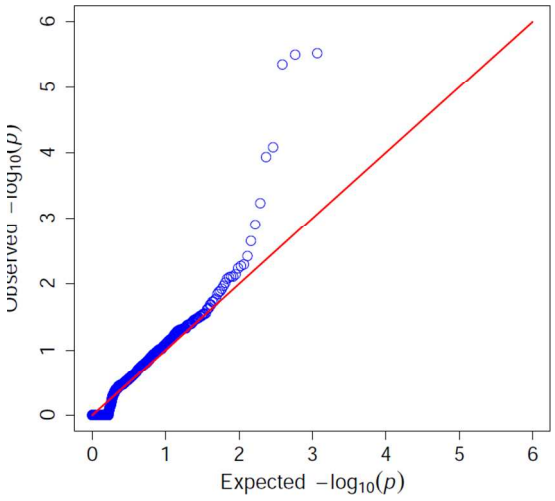

Figure E4: Q-Q plot for adjusted SKAT-O analysis of upper SaKnorm tertile in discovery cohort.

The red line is the observed =expectation line.