Genomewide association study of the age of onset of childhood asthma

Online Supplement

Erick Forno, MD MPH[1,3,4,*], Jessica Lasky-Su, ScD[1,4], Blanca Himes, PhD[1,4], Judie Howrylak, MD[1], Clare Ramsey, MD MPH[5], John Brehm, MD MPH[1,2,4], Barbara Klanderman, PhD[1], John Ziniti, BA[1], Erik Melén, MD PhD[1,6,7], Goran Pershagen, MD PhD[6], Magnus Wickman, MD PhD[6,8], Fernando Martinez, MD[9], Dave Mauger, PhD[10], Christine Sorkness, PharmD[11], Kelan Tantisira, MD[1,2,4], Benjamin A. Raby, MD MPH[1,2,4], Scott T. Weiss, MD MS[1,4], Juan C. Celedón, MD DrPH[12]

[1]Channing Laboratory and [2]Division of Pulmonary/Critical Care Medicine, Dept. of Medicine, Brigham and Women's Hospital, Boston, MA; [3]Division of Pediatric Pulmonology, Dept. of Pediatrics, University of Miami, Miami, FL; [4]Harvard Medical School, Boston, MA; and [5]Dept. of Medicine, University of Manitoba, Winnipeg, Canada; [6]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; [7]Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden; [8]Sachs Children's Hospital, Stockholm, Sweden; [9]Department of Pediatrics, University of Arizona, Tucson, AZ; [10]Penn State Hershey College of Medicine, Hershey, PA; [11]University of Wisconsin, Madison, WI; [12]Division of Pediatric Pulmonary Medicine, Allergy and Immunology, Dept. of Pediatrics, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh School of Medicine, Pittsburgh, PA.

*Corresponding author: Erick Forno, M.D., M.P.H.

Department of Pediatrics, University of Miami

1580 NW 10th Ave #125, Miami, FL 33136

Phone: 305.243.1420; Fax: 305.243.1262

Email: eforno@med.miami.edu

METHODS

Population for the GWAS (CAMP)

For the CAMP trial, asthma was defined by symptoms greater than 2 times per week, the use of an inhaled bronchodilator at least twice weekly or the use of daily medication for asthma, and airway responsiveness to $\leq$12.5 mg/ml of methacholine. Children with severe asthma or other clinically significant conditions were excluded. In CAMP, 1,041 asthmatic children were followed up for 4-6 years. Of these, 968 children and 1,518 of their parents contributed DNA samples. For this study, we restricted our analysis to 573 non-Hispanic white children (413 index children in nuclear families and 160 singletons).

Phenotypic assessment

Spirometry was conducted at baseline in CAMP, GACRS and PACT following American Thoracic Society recommendations (1). In CAMP, subsequent measurements were obtained at 4-month intervals throughout the 48 months of the study; completion rate was ~94%. Albuterol use was assessed at the same intervals in CAMP via parental report (range of the score was 0-4, where 0=none, 1=less than once a week, 2=at least once a week, 3=at least twice a week, 4=daily). Total serum IgE and peripheral eosinophil counts were assessed at the beginning of each study, and were $\log_{10}$-transformed for analysis.

Genotyping and QC in CAMP

Stringent quality-control was conducted for the genome-wide genotypic data (see **Table S1**): 6,257 markers were removed due to low clustering scores, and 1,329 markers were removed because their flanking sequences did not map to a unique position on the hg17 reference genome sequence. Further quality control was performed with PLINK v1.03 (2). The average completion rate for each marker was >99%. Monomorphic markers (n=3,790) and those with $\geq$5 Mendelian errors (n=2,445) were removed. We assessed genotype reproducibility by plating 4 subjects once on each of 14 plates. All of these replicates had >99.8% concordance. The average genotyping completion rate for each subject was 99.75%. No filtering was done based on Hardy-Weinberg equilibrium due to ascertainment of the cohort through affected probands. Thus, 547,645 (97.5%) of the 561,466 SNPs in the BeadChip passed quality control. Further, SNPs with low

minor allele frequencies (MAF) <1% and SNPs located in sex chromosomes were excluded, leaving 512,296 SNPs for analysis.

Genotyping in replication cohorts

SNPs selected for replication were subsequently genotyped in GACRS using Sequenom MassArray genotyping with iPlex chemistry (12 SNPs) and Taqman (Applied Biosystems) assays (2 SNPs). Completion rate was 96% and concordance was 99.9%. In BAMSE, genotyping was performed using the Illumina Human610 quad array. All samples had genotyping success rate >95%. Genotyping in PACT was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA). Completion rate was >96.5% for all included subjects, with the average completion rate >99%. When markers from Illumina were not available in Affymetrix, we used imputed data (see below) and also performed an exploratory analysis using selected available SNPs with the highest LD possible (highest r^2 for CEU trios from the HapMap).

Genotypic data imputation in PACT

Two of the selected SNPs were not genotyped in PACT (rs4658627 and rs7927044). For rs4658726, imputation was performed on the June 2010 release of the 1000 Genome Project (1KGP) data was performed using the Markov Chain Haplotyping software (MaCH)(3). For rs7927044, imputation was performed on data from the HapMap project (Phase 2 Release 22) using the same software. The ratio of the empirically observed dosage variance to the expected (binomial) dosage variance for imputed SNPs utilized was >0.5 for both imputed SNPs, indicating good quality of imputation. For imputed SNPs, dosage data was used to compute association statistics.

Genotypic data imputation in CAMP

Imputation of all SNPs available in the June 2010 release of the 1KGP data that were not genotyped or failed QC was performed with the Markov Chain Haplotyping software (MaCH)(3). The ratio of the empirically observed dosage variance to the expected (binomial) dosage variance for imputed SNPs utilized was greater than 0.5, indicating good quality of imputation.

Gene expression profiling

CD4+ lymphocytes were isolated from peripheral blood samples collected from 299 subjects from four clinical centers (Baltimore, Boston, Denver, St. Louis) participating in the Childhood Asthma Management Program (CAMP) Continuation Study, part 2 (CAMPCS/2); CAMPCS/2 was the second of two 4-years observational follow-up studies of CAMP participants carried out upon completion of the original CAMP study. Blood samples for this analysis were obtained during a routine CAMPCS/2 clinical visit between May 2004 and July 2007. CD4+ T cells were isolated from the collected mononuclear cell layer using anti-CD4+ microbeads by column separation (Miltenyi Biotec, Auburn, CA) (4). Total RNA was extracted using the RNeasy Mini Protocol (QIAGEN, Valencia, CA) (5). Expression profiles were generated with the Illumina HumanRef8 v2 BeadChip arrays (Illumina, San Diego, CA) and arrays were read using the Illumina BeadArray scanner and analyzed using BeadStudio (version 3.1.7) without background correction. Raw expression intensities were processed using the lumi package (6) of Bioconductor with background adjustment with RMA convolution (7) and $\log_2$ transformation of each array. The combined samples were quantile normalized. The normalized microarray data are available through the GeneExpression Omnibus of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/geo/, accesion number GSE22324). The expression data for C1orf100 (assayed on the HumanRef8 v2 array by probe ID ILMN_8320, with sequence TAGCCACAGTTTCGCTGAATCCTCGACCGCTTAATTCACTGCCAGAGCTC) was assessed for association with FEV_1 (percent predicted), FEV_1/FVC, and albuterol use by repeated measures analysis in SAS v9.2, using the MIXED procedure assuming a fixed-effects covariance structure, and adjusted for all covariates in the main analyses.

Statistical methods for phenotypic variables

To assess whether SNPs associated with asthma onset were also associated with indicators of asthma severity, we conducted longitudinal analyses of FEV_1 and albuterol use score in CAMP adjusted for age and height at randomization, gender, race, ETS, study center, total IgE, and peripheral eosinophil count. Analyses were done by mixed-effects regression models. Children in the budesonide arm were excluded to avoid confounding by treatment with inhaled corticosteroids, which had an effect on lung function in CAMP. Residual maximum likelihood

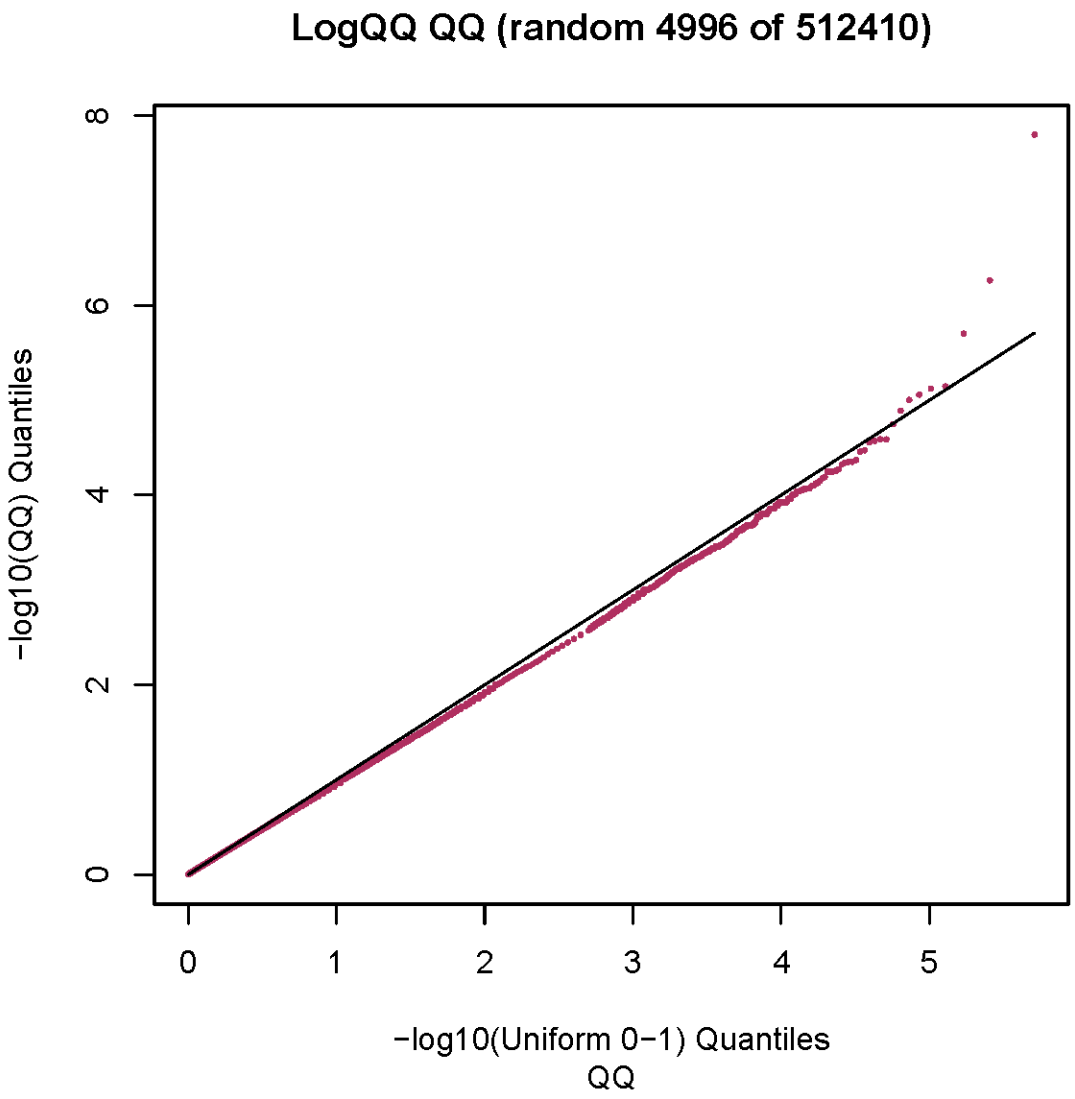
estimation with spatial-exponential covariance structure was used. Fixed-effects test statistics were adjusted using the "sandwich" error estimator. Longitudinal P-values reported are from χ^2 tests with *n-1* degrees of freedom, where *n* is the number of measurements for each outcome. Analysis was done using SAS v9.2 (SAS Institute, Cary, NC).

123 **Table E1. Summary of the QC and cleaning procedures in the CAMP GWAS**

Attribute	***Count (% of 561,466 markers on array)***
Low Illumina QC score	6,257 (1.1%)
Flank sequences do not map to hg17	1,329 (0.2%)
Monomorphic	3,790 (0.7%)
Parent-offspring inconsistencies >4	2,445 (0.4%)
Total number of failed markers	13,821 (2.5%)
Total number of passed markers	547,645 (97.5%)
Autosomes	534,290 (95.2%)
Sex-linked	13,229 (2.4%)
Mitochondrial genome	126 (0.02%)
Autosomal markers with MAF < 1%	21,994 (3.9%)
Total autosomal markers used in the analysis	**512,296 (91.2%)**

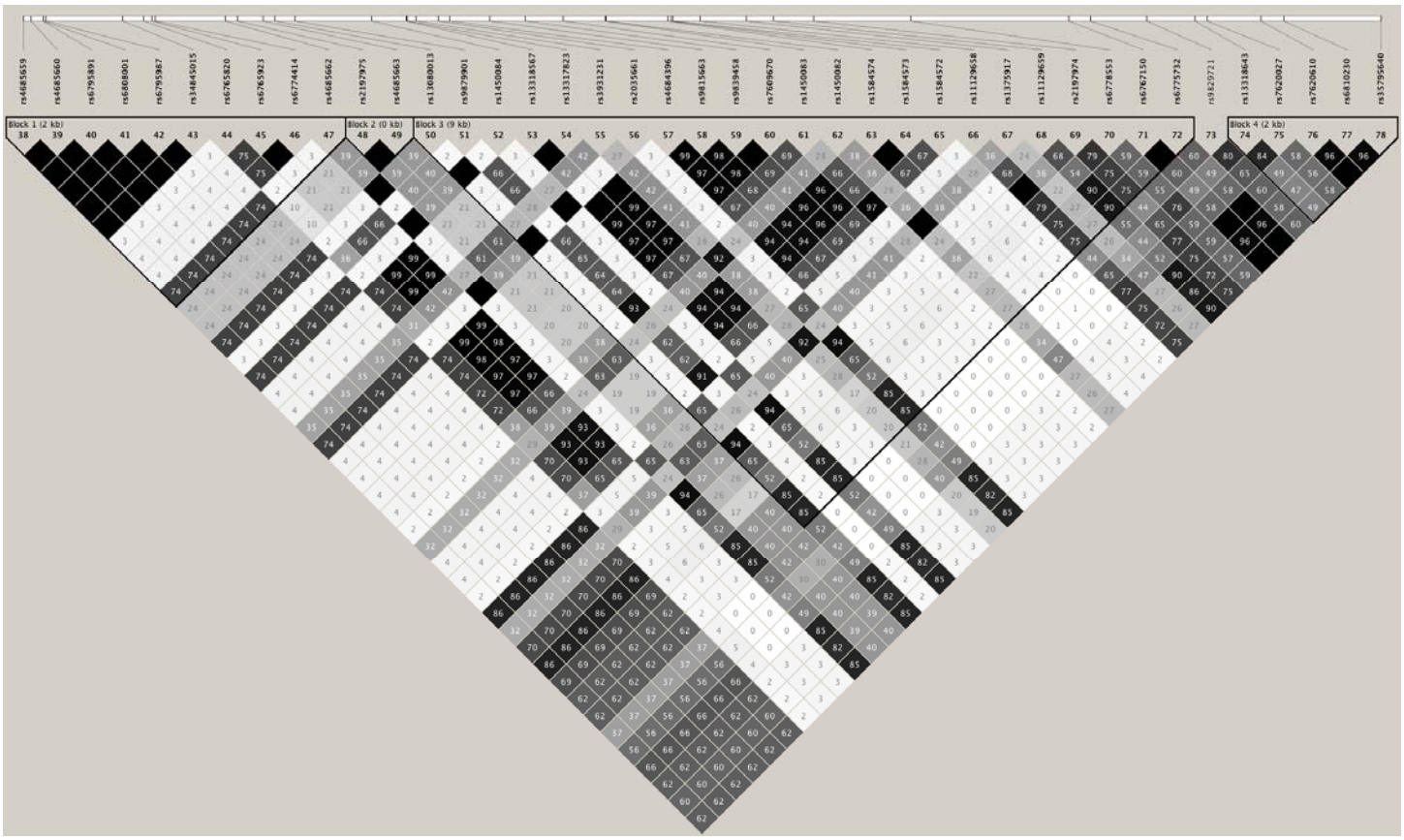
124

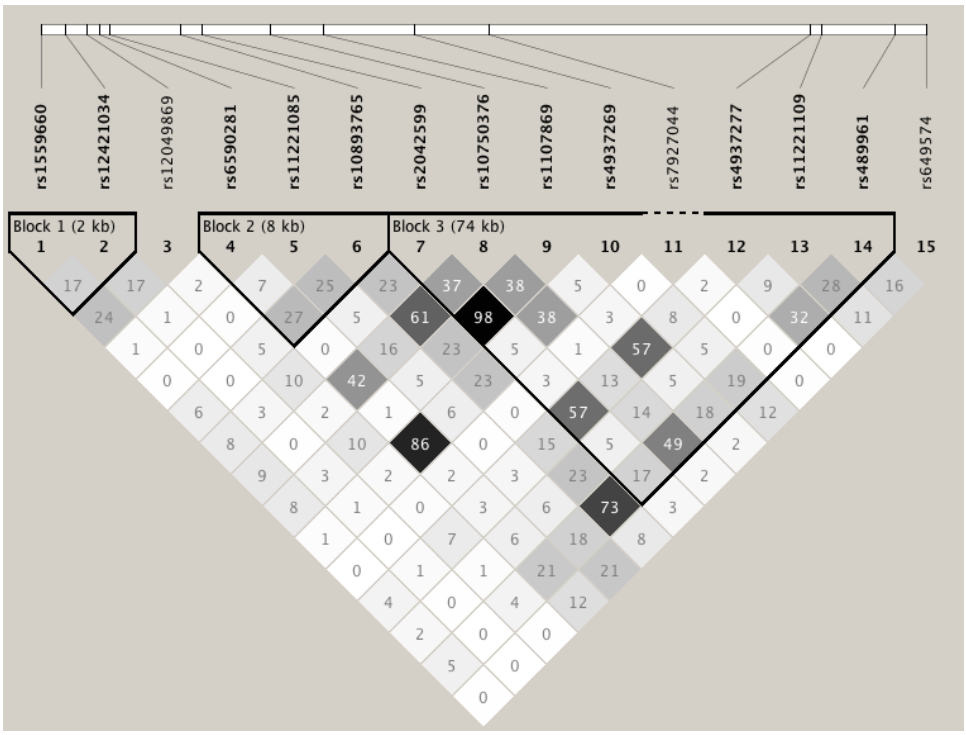
Figure E1. QQ plot for CAMP GWAS

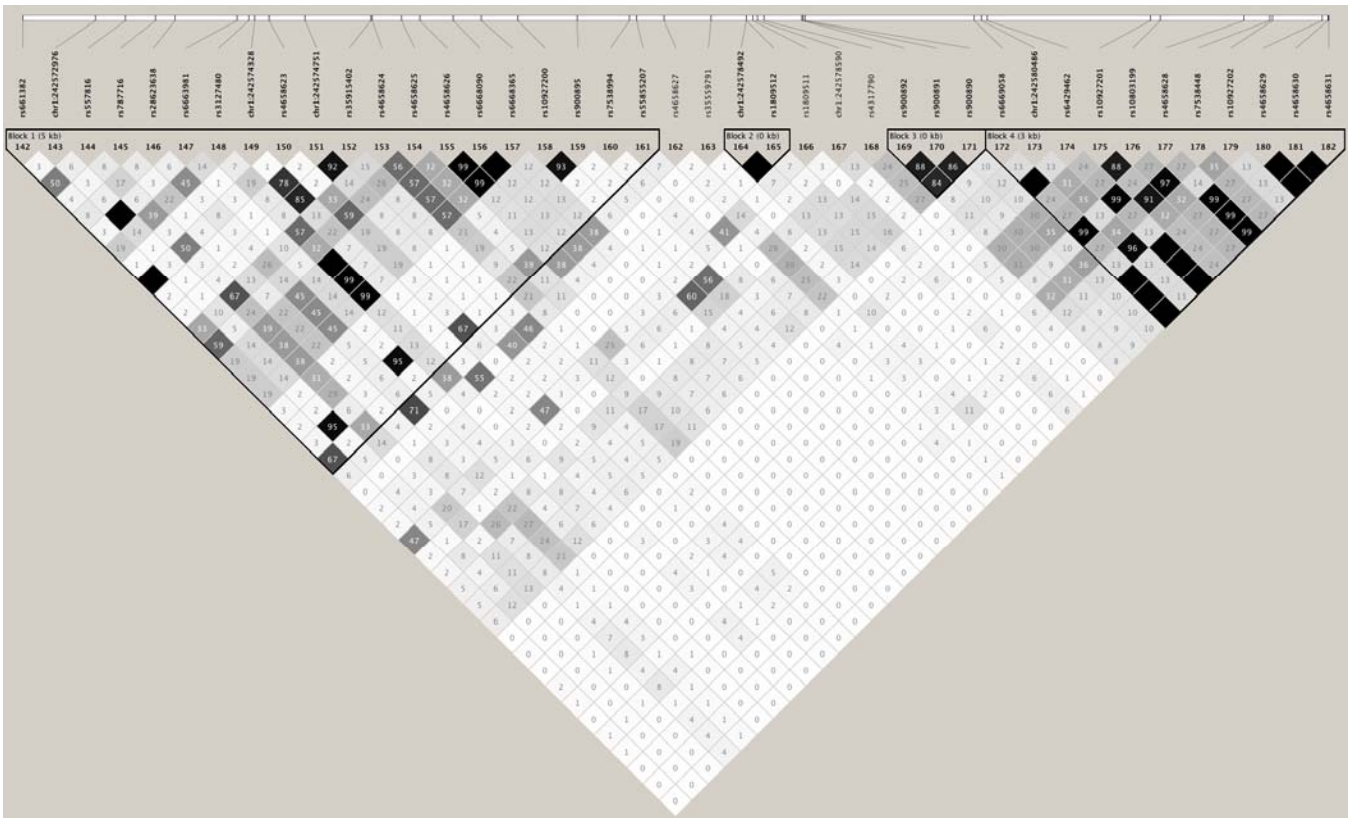


Legend: Probability quantile-quantile plot for the CAMP GWAS

Figure E2. LD plots for SNPs rs9815663 (top), rs7927044 (middle), and rs4658627 (bottom)







Legend: Figures show the LD plots for the three top SNPs from the CAMP GWAS. Plots were obtained using imputed genotypic data in CAMP generated from the 1000 Genomes Project.

REFERENCES

(1) American Thoracic Society. Standardization of Spirometry, 1994 Update. Am J Respir Crit Care Med 1995 Sep;152(3):1107-36.

(2) Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses Am J Hum Genet 2007 Sep;81(3):559-575.

(3) Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease Nat Genet 2008 Feb;40(2):161-169.

(4) Jonuleit H, Schmitt E, Schuler G, Knop J, Enk AH. Induction of interleukin 10-producing, nonproliferating CD4(+) T cells with regulatory properties by repetitive stimulation with allogeneic immature human dendritic cells J Exp Med 2000 Nov 6;192(9):1213-1222.

(5) Gonzalez P, Zigler JS,Jr, Epstein DL, Borras T. Identification and isolation of differentially expressed genes from very small tissue samples BioTechniques 1999 May;26(5):884-6, 888-92.

(6) Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray Bioinformatics 2008 Jul 1;24(13):1547-1548.

(7) Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data Biostatistics 2003 Apr;4(2):249-264.