



Weiss - The Genetic Epidemiology of Asthma in Costa Rica and the Childhood Asthma Management Program (CRA_CAMP)

Updated 06/27/2018

[Introductory slides from the June 4, 2015 Steering Committee/EAP meeting \(requires log-in\).](#) [1]

[Introductory slides from the July 13, 2017 All-hands meeting \(requires log-in\).](#) [2]

The Genetic Epidemiology of Asthma in Costa Rica (CRA) Cohort Parent Grant, R37 HL066289

From February 2001 to August 2008, questionnaires were sent to the parents of 16,912 children (ages 6-14 years) enrolled in 140 Costa Rican schools; 9,180 (54.3%) questionnaires were returned. Children were eligible for the study if they had asthma (physician-diagnosed asthma and ≥ 2 respiratory symptoms or asthma attacks in the prior year) and a high probability of having ≥ 6 great-grandparents born in the Central Valley of Costa Rica (as determined by the study genealogist on the basis of the paternal and maternal last names of each of the child's parents). Of the 9,180 children screened, 3,113 (33.9%) had asthma. By the close of the recruitment and enrollment of the CRA Study in 2011, samples from 4,245 individuals have been collected for the main study and related subsequent studies. Children, parents and pedigree relatives gave blood samples for DNA extraction. All probands/children completed a protocol including questionnaires, spirometry, methacholine challenge testing (if their FEV₁ was $\geq 65\%$ of predicted), allergy skin testing, and collection of blood (for plasma, DNA and RNA extraction, and measurement of serum total and allergen-specific IgE) and house dust (for measurement of dust mite/cockroach allergens) samples.

The parent grant, R37 HL066289 was initially funded as an R01 and then was successfully renewed with a 1st percentile score and was then converted to an R37 MERIT Award that has subsequently been renewed. The grant is currently in its 13th year and has two years remaining in its current segment; thus, it is eligible for this administrative supplement.

The unique aspects of this population are that the asthma prevalence rates in Costa Rica (CR) are among the highest in the world and the Central Valley of Costa Rica, the primary recruitment/enrollment area, is a relative genetic isolate where we have had the ability to ascertain relatives in large extended pedigrees and phenotype these subjects for asthma and related traits. A total of 671 subjects in 8 large pedigrees were ascertained and phenotyped, and we have collected an additional 1053 trios. Initial efforts were focused on linkage studies and then we subsequently moved to genetic association studies, at which time we added the trios (N=720) in the Childhood Asthma Management Program (CAMP) to the grant as a replication population for our results, as these two populations have exactly the same study design (e.g., trios) and identical protocols for phenotyping subjects. We have approximately 100 phenotypes relevant to heart, lung, blood and sleep disorders, including asthma, obesity, height, COPD and blood and lipid disorders via metabolomics. It is highly unlikely that any application has the ability to test private mutations in extended pedigrees, as well as generalize from inbred to outbred populations to the extent

that we do in this proposal.

We have been productive in publishing our work. A total of 114 linkage and association papers have been written using these two study populations. Some of the most important findings from the initial 13 years of the study are the following: we have (1) identified *MMP12* as an important gene for asthma, COPD and lung function decline; (2) identified allele specific chromatin remodeling via an insulator on chromosome 17 that co-regulates the *ZBP2/GSDMB/ORMDL3* locus; (3) identified, through linkage and fine-mapping, *PRKCA* as a novel gene for asthma and obesity; (4) replicated GWAS results for *PDE4D* from CAMP; (5) identified a novel gene for obesity, *ROBO1*; and (6) identified vitamin D deficiency as an important risk factor for increased asthma severity. In the third five-year cycle, we added GWAS and integrative genomics, including whole blood gene expression and methylation arrays on 384 probands. The current aims of the grant are to perform GWAS and association analysis on Costa Rica and replicate them in CAMP and other Hispanic populations. This aim is directly related with this administrative supplement. The GWAS analyses are just now being completed and the gene expression, microRNA and methylation arrays are just being run. This means that the aims of the parent grant are directly aligned with the aims of the proposed administrative supplement. Therefore, with two years remaining on the parent grant, there is ample time to perform the whole genome sequencing proposed and report the results from the project. As described in the significance section of the administrative supplement

SPECIFIC AIMS Sequencing supplement to R37 HL066289

We focus on asthma because of its public health significance. Asthma affects 26 million U.S. children and adults,¹ [3] remains a major cause of morbidity (one-half million hospitalizations a year), and is the most common cause of school and work days lost. Asthma-related costs are estimated to be over \$12.7 billion annually.¹ [3] The three primary goals of this project are to: (1) identify common and rare genetic variants that determine asthma and its associated phenotypes through whole genome sequencing (WGS); (2) perform novel family based association analysis of our WGS data to identify novel genes for asthma; and (3) integrate epigenomic and transcriptomic data with our WGS data and determine the epistatic interactions present using systems genomics approaches. Identification of the molecular determinants of asthma remains an important priority in translational science. Genome wide association studies (GWAS) have been successful in this regard, identifying at least 10 novel susceptibility genes for asthma.² [4],³ [5] However, as with most complex traits, the variants identified by GWAS explain only a fraction of the estimated heritability of this disorder. **Herein, we propose a novel family-based study design and state-of-the-art genome sequencing techniques to map a set of sequence variants for asthma and its associated phenotypes and assess the interrelationships of the identified genes and variants using systems genomics methods. We are sequencing 1198 subjects: 291 trios in CR 873 subjects and 325/671 or 49% of the extended pedigrees in CR.**

Specific Aim 1: We hypothesize that coding and regulatory variants for asthma and its associated phenotypes are identifiable through whole genome sequencing. In **Subaim 1a**, we will generate genome wide DNA sequencing data (WGS) and imputed WGS in 1,053 childhood asthma trios, 1053 from the Costa Rica Asthma (CRA) study and 694 from the Childhood Asthma Management Program (CAMP) that are part of R37 HL066289. In **Subaim 1b**, with the help of the Data Coordinating Center, we will QC these sequence data and relate this QC information to other NHLBI consortium members' data. In **Subaim 1c**, we will deposit the cleaned and QC'ed WGS data into dbGaP with the phenotypic information.

Specific Aim 2: We hypothesize that family based association analysis will identify novel genes for asthma. In **Subaim 2a**, we will perform standard family-based association analyses as well as the novel two-stage screening algorithm we have previously described in our 1,747 trios. In **Subaim 2b**, we

will apply standard and novel approaches to rare variant analysis of trio's data and in **Subaim 2c**, the WGS data will be combined with existing exome sequencing and genotyping of extended pedigrees in the Costa Rica asthma population and in a rare variant meta-analysis. Finally, in **Subaim 2d**, we will utilize novel methodology to create and analyze haplotypes in family data to identify all possible genes for childhood asthma using the largest family-based data set for asthma in the world.

Specific Aim 3: We hypothesize that multidimensional interaction networks (epistasis) underlie the genomic complexity of asthma and we will elucidate those interactions using our WGS, methylation and transcriptomic data and a molecular interaction network (interactome) approach. In Subaim 3a, we will use a molecular interaction network (interactome) to understand the genomic loci associated with asthma and its associated phenotypes. We will evaluate: (i) whether genes from WGS (**from Specific Aim 2, our transcriptomic and epigenetic data, and the literature**) are significantly connected via protein-protein interactions and (ii) we will use network-based separation to analyze the closeness of different data types in molecular interaction networks. In **Subaim 3b**, we will identify a *disease module*, a connected sub-network that can be mechanistically linked to the asthma phenotype. We will apply the Disease Module Detection method (DIAMOnD) that exploits the structural properties of the interactome to identify the disease module for asthma and its associated phenotypes. In **Subaim 3c**, we will assess commonality and dissimilarity of the pathways enriched in asthma and its associated phenotype modules. In **Subaim 3d**, we will prioritize genes with rare, deleterious variants that demonstrate an association with asthma in **Specific Aim 2**. In **Subaim 3e**, we will use an “edgetics” strategy to interpret the genotype-phenotype relationship. We will analyze the effect of the rare variants from **Subaim 3d** on the network rewiring based on the edgetic changes.

Source URL (modified on 06/27/2018 - 7:10pm):https://topmed.nhlbi.nih.gov/group/cra_camp

Links

[1] https://topmed.nhlbi.nih.gov/system/files/meetings/Weiss_WGS-CR_06042015FINAL%20%5BCompatibility%20Mode%5D.pdf [2] <https://topmed.nhlbi.nih.gov/system/files/meetings/SuCAMPTOPMed%20presentation07132017.pdf> [3] https://topmed.nhlbi.nih.gov/group/cra_camp#_ENREF_1 [4] https://topmed.nhlbi.nih.gov/group/cra_camp#_ENREF_2 [5] https://topmed.nhlbi.nih.gov/group/cra_camp#_ENREF_3