**Specific Aims**

As the most common chronic inflammatory disease of the lung, asthma will likely afflict over 10% of the U.S. population by the end of this decade.1 Scientific efforts over the last forty have identified Type 2 inflammation as the driving immunologic response in many individuals with asthma.2 These efforts also demonstrated that many patients have low or no evidence of Type 2 inflammation and that the complex pathophysiology seen among patients with asthma is due to differential contributions from genetic susceptibility and environmental exposures that include diet, air quality, geography, and the endogenous microbiome.3,4 This is especially true in individuals with severe or refractory disease (5-10% of asthmatics or 25-30 million in the U.S) and underscores the fact that the current paradigm of asthma severity, used in the clinic and for research, is inadequate because it combines patients with different inflammatory endotypes and obscures many molecular networks that contribute to disease pathobiology.5 New tools and approaches are needed to identify subgroups of disease so we can move from the traditional approach of defining severity as a continuum from mild to severe, to the model that defines distinct endotypes of disease that are driven by different patterns of dysregulated biologic pathways that will respond to a precisely determined therapeutic approach.6 Up to now, most efforts to identify asthma subgroups have used clustering algorithms of clinical variables, validating lung function and age of disease onset as key determinants of disease phenotypes.7 But, a true understanding of disease endotypes requires the examination of the molecular perturbations associated with the disease and/or clusters already established. Our translational research program is focused on such efforts through the study of a well characterized heterogeneous cohort of individuals with asthma. We have identified two endotypes of disease by: 1) the identification of a molecule that is associated with subgroups of patients with refractory asthma, Chitinase-3-Like-1(CHI3L1)/YKL-40 and 2) unsupervised cluster analysis of sputum transcriptomes that identified transcriptomic endotypes of asthma (TEA) clusters that are associated with severe disease.8-10 The first discovery was that circulating levels of CHI3L1/YKL-40 are elevated in a subgroup of patients with severe asthma, and correlates with airway remodeling, genetic polymorphisms in *CHI3L1*, and the levels of YKL-40 in the airway. The second discovery was made with a novel pathway-based, unsupervised cluster analysis of sputum gene expression that identified 3 “transcriptional endotypes” of asthma (TEA clusters): one subgroup with a history of near fatal attacks, one with a high rate of hospitalizations for asthma, and a subgroup with milder disease. However, while we have established these clinically meaningful endotypes of disease and validated them in other cohorts, a detailed understanding of the genes, pathways, and networks that drive the YKL-40 endotypes and TEA clusters has not been determined. **We hypothesize the existence of unique cell populations, transcriptomes, and functional responses that are associated with YKL-40 endotypes and TEA clusters of asthma.**  Our multidisciplinary team will determine the cellular immunophenotypes, transcriptomes, gene pathways and networks that are associated with the YKL-40 endotypes, TEA clusters and integrated endotypes that will be identified in Project 3, in a heterogeneous cohort of individuals with asthma enrolled in the NextGen study**.**  Correlations with T follicular helper (Tfh) cells, Dkk-1, and immunogenic microbiota will also be characterized (Project 2). We will pursue the following aims:

**Aim 1: Determine the single cell signatures and functional responses associated with YKL-40 endotypes of asthma from baseline and follow-up samples from the NextGen Study (Core B).** Studies will be conducted on NextGen study samples to determine: a) the immunophenotype of the cell populations that produce YKL-40 in the airway using single cell CyTOF analysis of sputum cells at baseline, b) single cell signatures and networks that are associated with YKL-40 endotypes of asthma, and c) the functional responses of the cell populations defined in Aim 1a and 1b to cytokines including IL-13 and YKL-40.

**Aim 2: Determine the single cell signatures and functional responses associated with TEA clusters of asthma from baseline and follow-up samples from the NextGen Study (Core B).** Studies will be conducted on NextGen study samples to determine the cell populations associated with TEA clusters of asthma determined by single cell CyTOF analysis of sputum cell populations at baseline, b) the single cell transcriptomic signatures and networks associated with TEA clusters and c) the functional responses of the cell populations defined in Aim 1a and 1b to cytokines including TNF, IL-13 and YKL-40.

**Aim 3: Define the functional responses associated with endotypes of asthma identified in Projects 2 and 3 compared to YKL-40 endotypes and TEA clusters.** Studies will be conducted on samples collected from subjects enrolled in the NextGen study at follow-up visits to determine the functional responses of cell populations associated with novel asthma clusters of disease identified in Project 3 by integrated modeling, b) how the cellular immunophenotypes and cellular responses differ among different endotypes compared to the integrated clusters, and c) define the relationship between YKL-40 endotypes and TEA clusters with Tfh cells, Dkk-1, and immunogenic microbiota endotypes defined in Project 2.

**Significance:**  In the case of asthma, we can only achieve true precision medicine if we understand the molecular mechanisms that underlie endotypes of disease. We believe that defining these mechanisms at the single cell level that are associated with YKL-40 endotypes and TEA clusters will be a significant step forward towards reaching this ambitious goal and will have implications for development of therapeutics and a deeper understanding of asthma pathogenesis.

In most individuals with asthma, symptoms are easily controlled by treatment with bronchodilators and low doses of inhaled corticosteroids. However, a significant proportion of asthmatics have persistent symptoms despite receiving high doses of inhaled corticosteroids. Compared to controlled asthmatics, these “refractory” asthmatics have medical costs that are 6-7-fold higher and require frequent hospitalization.11 Studies over the last two decades have defined several molecular targets in the Type 2 inflammatory pathway that contribute to refractory asthma, leading multiple pharmaceutical companies to develop therapeutic monoclonal antibodies against targets such as IL-5 and IL-13.12-14 Since as many as 30% of severe asthmatics are not allergic or have evidence of eosinophilia, but have significant airway obstruction, there remains a significant unmet need for therapeutics in severe asthma.5 Thus, despite being a small percentage of all asthmatics, these subgroups constitute a large number of patients (millions in the U.S.) whose disease is unlikely to be controlled with these new biologic therapies. Since current animal models of asthma only recapitulate adaptive Th2 driven disease (most commonly to ovalbumin or house dust mite antigen), at this point in time research on asthma heterogeneity can only be accomplished through the study of asthma in humans.15,16 *We have established a cohort of individuals with asthma to study asthma heterogeneity and resolve endotypes of disease. A major goal of this U19 is to define the pathobiology of asthma endotypes in humans by examining cellular phenomena in the sputum.*

Although differences in environmental exposures and compliance may explain a significant proportion of asthma heterogeneity, it is clear that the biologic networks that drive asthmatic inflammation are pathobiologically diverse across the spectrum of disease. This concept supports the growing belief that asthma is actually a collection of several different airway diseases, each driven by a different set of gene networks with similar and overlapping inflammatory, physiologic, and clinical features. Research efforts in asthma are now focused on defining asthma subgroups or phenotypes and endotypes to improve the development and targeting of therapeutics to improve healthcare outcomes, especially for the most severe patients with asthma. The best example of novel phenotypes in asthma are the SARP clusters.7,17,18 These subgroups have validated the importance of lung function and age of disease onset as critical discriminating features that underlie asthma heterogeneity. The best example of an asthma endotype is the Th2 high definition of disease that is strongly linked to exaggerated IL-13 expression.12 Importantly, while both endotypes and phenotypes reduce the heterogeneity of the population being examined, there remains significant heterogeneity within each subgroup identified. This is true of the SARP clusters, Th2 high/low defined disease, as well as the endotypes that we have established including YKL-40 endotypes, and the TEA clusters. This highlights the need for further study of the pathobiologic mechanisms that are associated within these endotypes. Only then will we know which subgroups are most generalizable, robust and relevant to pathobiology and clinical care. *To do this properly, we will need to employ a combination of methods that include, leveraging existing clusters of asthma with novel biologic measurements, examining the pathobiology of the endotypes at the single cell level, and defining new endotypes as higher resolution datasets and computational approaches become available.*

**Innovation***.* There are several innovative aspects to this proposal that warrant discussion:

***Novel models of asthma heterogeneity.*** A central theme of this U19 proposal is to develop novel, clinically and pathobiologically meaningful ways to deconstruct the complex disease of asthma and make these tools publicly accessible. The YK-40 endotypes and TEA clusters that we have identified provide two approaches that are complementary and innovative ways to define subgroups of asthma. The studies outlined herein will significantly enhance our understanding of the biologic drivers of these endotypes at an unprecedented level of resolution. Combined with our goal to define novel asthma clusters through integrated analysis of multidimensional data, these studies will move our understanding of asthma heterogeneity significantly forward.

***Single cell study of airway cells from asthmatics.*** In addition to the novel endotypes of disease which the Yale U19 team has described to date, our multidisciplinary team is pioneering the use of single cell technologies on sputum derived inflammatory cells including CyTOF and single cell RNA-seq (technical aspects and preliminary data are described in Core C and computational approaches are described in Project 3). This provides an opportunity to move beyond Gaussian based statistics to a transformative approach that can precisely determine cellular lineage and immunophenotypes of cell populations from a large population of cells that are contributing to disease. We will also characterize cell populations in the airway using single cell transcriptomes as opposed to our current approach of staining cells with a Wright stain, or immunophenotyping by FACS. These are truly innovative and transformative concepts that will move the field of asthma heterogeneity forward.

**Research Strategy**

***Aim 1: Determine the single cell signatures and functional responses associated with YKL-40 endotypes of asthma from baseline and follow-up samples from the NextGen Study (Core B).*** *Studies will be conducted on NextGen study samples to determine: a) the immunophenotype of the cell populations that produce YKL-40 in the airway using single cell CyTOF analysis of sputum cells at baseline, b) single cell signatures and networks that are associated with YKL-40 endotypes of asthma, and c) the functional responses of the cell populations defined in Aim 1a and 1b to cytokines including IL-13 and YKL-40.*

**Rationale.** The driving goal of this project is to characterize the cell populations and transcriptomic signatures in the airway that are associated with two endotypes of severe asthma that we have established. We will also determine the functional responses of cells in the airway that are associated with these endotypes after stimulation with cytokines that are associated with asthmatic airway inflammation. We will use this information to identify the cellular populations and expression signatures that are linked to each endotype, ultimately defining molecular pathways associated with CHI3L1/YKL-40 and the TEA clusters. We will also interrogate the functional cellular responses associated with the integrated endotypes that will be identified in project 3. We will begin these studies in aim 1, by defining the cell populations and cellular responses that are associated with CHI3L1/YKL-40, a chitinase-like-molecule that we discovered is elevated in blood and airway of a subgroup of severe asthmatics and a critical mediator of inflammation and remodeling responses in asthma.

***Preliminary studies.*** Chitinases are a family of evolutionarily conserved hydrolases that belong to the 18-glycosyl hydrolase (GH 18) gene family.19 These proteins were characterized by their ability to cleave chitin, the second most common naturally occurring polysaccharide (cellulose is the most common). Chitinases have been studied extensively in lower life forms where they function to control chitin homeostasis and degrade chitin in the surrounding environment. Mammals do not make chitin, but have been shown to have genes for enzymatically active chitinases and several chitinase-like proteins (CLP) that bind to chitin, but do not catalyze it.20 Evidence in animals and humans suggests that both chitinases and chitinase-like proteins are potent regulators of inflammatory responses. In the airway, chitinases and CLP are juxtaposed between the environment and the host by controlling exposure to chitin inhaled from the environment by degrading it and modulating innate and adaptive immune responses to it.19 The mammalian chitinases that have been studied and implicated in the pathogenesis of asthma are the true chitinases, chitotriosidase (CHIT1) and acidic mammalian chitinase (AMCase), and the CLP, Chitinase-3-Like-1 (CHI3L1)/YKL-40 in (Breast Regression Protein 39 (BRP-39) in mice). In humans, YKL-40 regulates homeostasis of many organs and tissues. It is produced by many different cell types including monocytes, macrophages, neutrophils, bronchial epithelial cells, chondrocytes and synovial cells, indicating its widespread importance.19 In addition, YKL-40 controls cell proliferation, survival and stimulates human lung fibroblasts and airway smooth muscle cells to produce cytokines. Recently, co-immunoprecipitation studies identified IL-13 receptor alpha 2 (IL-13R2) as a receptor for YKL-40. YKL-40 binds to IL-13R2 in a complex with IL-13 that leads to activation of ERK, mitogen-activated protein kinase (MAPK), protein kinase /AKT, followed by Wnt/-catenin signaling. This cascade of events regulates oxidant injury, apoptosis, pyroptosis, inflammasome activation, antibacterial responses, melanoma metastasis, and TGF-production *in vivo*.21,22 *Our research in human asthma demonstrates a strong link between YKL-40 and severe asthma and more recently specific endotypes of asthma (outlined below).*

The importance of CHI3L1/YKL-40 in asthma pathogenesis was recognized from studies in transgenic (TG) mice that overexpress IL-13 in the airway.23 These IL-13 TG animals develop intense peri-bronchial eosinophilic inflammation, airway remodeling and airway hyperresponsiveness, with histopathology reminiscent of the lungs from autopsies of severe asthmatics.24 Microarray analysis of whole lung lysates from IL-13 TG compared to wild-type mice revealed marked elevations of mRNA encoding the murine homologue of YKL-40, BRP-39. These studies demonstrated that IL-13 stimulates expression of BRP-39 mRNA and protein in the murine lung and suggested this CLP might contribute to allergic airway inflammation. Additional studies in knockout mice that lack BRP-39 demonstrated that eosinophilic inflammation and Th2 cytokine induction was decreased by > 85% after standard OVA sensitization and challenge compared to WT mice.25 To gain insights into the mechanisms by which YKL-40/BRP-39 contribute to Th2 inflammation, we characterized the recovery and survival of T cells in OVA sensitized and challenged WT and BRP-39 knockout mice and demonstrated increased percentages of cells undergoing cell death with accumulation of apoptotic cells compared to WT mice.15,25,26 *These studies demonstrate that CHI3L1/YKL-40/BRP-39 is an important modulator of Th2 inflammation and regulates apoptosis. This supports our hypothesis that YKL-40 is associated with an asthma endotype and that individuals with asthma that have elevated levels of YKL-40 have increased cell survival airway inflammation compared to individuals with lower levels of YKL-40 expression..*

***YKL-40 is associated with severe asthma in humans.*10,20,27**Based on the animal studies described above, we hypothesized that YKL-40, would be increased in asthmatics compared to normal individuals. Although asthma is traditionally considered an organ-specific disease, we believed that YKL-40 levels in the blood could discriminate asthma severity. To test this, we performed a cross-sectional analysis of serum samples from our cohort of asthmatic subjects from the Yale Center for Asthma and Airway Disease (YCAAD). YKL-40 was readily appreciated in the serum of normal volunteers and was significantly higher in the serum of asthmatics (P=0.02). YKL-40 increased with asthma severity, with the highest levels being observed in refractory asthmatics, compared to moderate and mild asthmatics (P for trend = 0.003, Figure 1). Circulating YKL-40 levels also correlated with asthma severity in the two other cohorts examined from the University of Wisconsin and the University of Paris (data not shown).

Figure 1. YKL-40 is increased in the blood in asthma.



We then performed immunohistochemistry for CHI3L1/YKL-40 on bronchial biopsies from the University of Paris cohort (Figure 2). The numbers of YKL-40-positive staining cells in asthmatics was significantly increased over control subjects that exhibited rare YKL-40 expressing cells (Figure 2, panel A). As shown inFigure 2, YKL-40 staining was seen in subepithelial cells from the majority of asthmatics (panels B-E). In severe asthmatics, the number of YKL-40 staining subepithelial cells was increased, and staining of the bronchial epithelium was also evident (Figure 2, panels D and E). In BAL, cytospin preparations showed that YKL-40 was found in the cytoplasm of macrophages and neutrophils (Figure 2, panel F). Importantly, in asthmatics, YKL-40 staining cells in the airway wall correlated with asthma severity and serum YKL-40 levels (r = 0.55, p < 0.001) (Figures 3). Serum YKL-40 levels also correlated inversely with FEV1 in all three cohorts (Yale, r = -0.22, P = 0.01; Wisconsin, r = -0.33, P = 0.009 and Paris, r = -0.21, P = 0.005; data not shown). There was a significant correlation between sub-basement membrane (SBM) thickness and the serum YKL-40 levels in this population (r = 0.51, P = 0.003, Figure 4). In addition to these seminal observations, in more recent studies we have demonstrated that YKL-40 levels in the sputum correlate with asthma severity and that YKL-40 levels rise significantly in bronchoalveolar lavage fluid of atopic asthmatics after segmental antigen challenge (Figure 5).9,28

Figure 3. YKL-40 levels in the airway correlate with asthma severity.

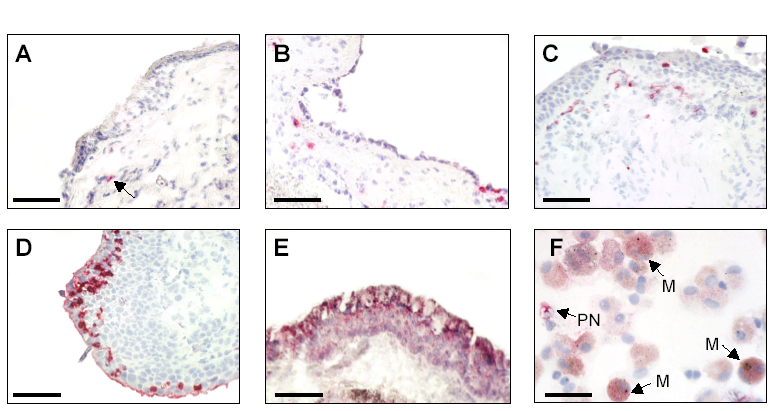


Figure 2. Immunohistochemical analysis of YKL-40 in the airway and BAL in asthma.

Finally, to further understand the patients with high levels of circulating YKL-40, a post-hoc analysis correlating serum YKL-40 levels and asthma characteristics in the Yale cohort demonstrated that YKL-40 levels correlated positively with the number of corticosteroid tapers in the last year, the dose of oral corticosteroids and the frequency of rescue inhaler use, and negatively with the percent predicted FEV1. YKL-40 was not associated with history of atopy or IgE level. Multivariable analysis of the data was undertaken to determine if the correlation between YKL-40 and asthma severity persisted after adjustments for confounders, including age, race, gender, history of atopy, BMI, and levels of serum IgE. In accord with our initial observations, this analysis demonstrated that asthma severity was associated with YKL-40 levels (adjusted P for trend = 0.02, data not shown) after adjustment for these factors. *These findings indicate that circulating YKL-40 is elevated in severe asthma, independent of IgE, atopy, and eosinophilic inflammation, and negatively correlates with lung function and airway remodeling. Although YKL-40 was initially identified downstream of IL-13, these studies show that in human asthma, YKL-40 is not specific to Th2 inflammatory responses, and is most strongly linked to severe phenotypes of disease. These data support our contention that YKL-40 is a marker of a specific endotype of asthma.*

***CHI3L1/YKL-40 polymorphisms and asthma severity.*** We have also identified genetic polymorphisms in *CHI3L1* that suggest this molecule defines an endotype of asthma, in particular, severe asthma.9,29 We conducted studies to determine if there are genetic polymorphisms in *CHI3L1* that are associated with asthma and asthma severity. In collaboration with Dr. Carole Ober at the University of Chicago, three asthma populations were examined; a well characterized founder population of European descent--the Hutterites (a related 13-generation, 1,623-person pedigree), 3 outbred Caucasian populations from the Childhood Origins of ASThma (COAST) birth cohort study, and 2 outbred asthma case-control populations from Freiberg Germany, and Chicago. A genome-wide association study was performed using the ~ 500,000 affymetrix SNP chip for YKL-40 levels. Four SNP with locus P values < 10-7, meeting the criteria for genome-wide significance (P < 0.05) were identified. All 4 SNP were in the *CHI3L1* gene, and one SNP was in the promoter (rs4950928) that is known to influence *CHI3L1* transcription. In all 3 populations, several of these SNP were associated with asthma prevalence, lung function, and bronchial hyperresponsiveness. Since the rs4950928 -131C/G SNP had been shown to interact with MYC/MAX and increase YKL-40 levels, it was considered the functional polymorphism. A second population-basedstudy of 6,514 Caucasian individuals also found an association between the rs4950928 SNP and asthma, however, this study did not examine serum YKL-40 levels.30 *These studies indicate that genetic variation in the promoter region of CHI3L1 are associated with asthma, but severe asthma was not examined.*

Figure 4. Serum YKL-40 levels correlate with airway remodeling (sub-basement membrane thickness). HcGP-39=YKL-40

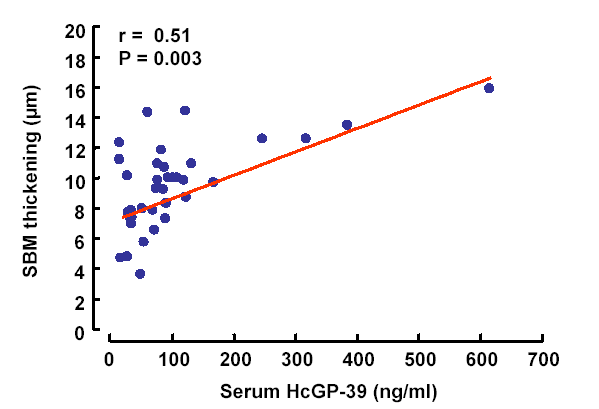


Figure 5. Meso Scale-measured YKL-40 levels in BAL before and after segmental antigen challenge (P<0.001).



The studies detailed above clearly demonstrated a strong link of the promoter SNP in *CHI3L1* to asthma prevalence and lung function, however, the populations examined were of mild severity. Therefore, we pursued studies to determine if genetic polymorphisms in *CHI3L1* are associated with asthma severity and airway levels of CHI3L1/YKL-40.9,31,32 We genotyped multiple SNP spanning the *CHI3L1* promoter and coding regions and correlated these SNP with asthma severity (based on the SARP clustering algorithm) and the levels of CHI3L1/YKL-40 in the serum and sputum in 261 YCAAD and 919 SARP subjects.17 Association and haplotype analyses were conducted to identify effects on airflow obstruction, serum and sputum YKL-40 levels. Fifteen SNPs in *CHI3L1* were associated with FEV1 and serum YKL-40 levels. Importantly, rs12141494, residing in intron 6, was most significantly associated with serum YKL-40 levels and post-bronchodilator FEV1 in both the YCAAD and SARP cohorts, in contrast to the promoter SNP, rs4950928, which did not correlate. In asthmatics, rs12141494 was the only SNP associated with both serum YKL-40 levels and asthma severity, an effect that is independent of the previously defined promoter SNP, rs4950928, in both study populations. Haplotype analysis in all European ancestry individuals demonstrated that the combination of the G allele at rs12141494 and the minor allele (G) at rs4950928 were associated with lower YKL-40 levels and higher FEV1 percent predicted. Subsequent analysis of sputum supernatants demonstrated that individuals with asthma and the risk allele A at rs12141494 are associated with higher levels of YKL-40 in the airway (P = <0.05, Figure 6). Therefore, genetic abnormalities in *CHI3L1* are also associated with asthma severity, airway remodeling, and airway levels of YKL-40. This suggests that mutations in the *CHI3L1* coding region affect YKL-40 expression in the airway and contribute to airway remodeling in asthma severity. *These discoveries further support our hypothesis that YKL-40 is a marker of a disease endotype characterized by the activation of specific cell populations and pathways that contribute to severe features of disease.*

Figure 6: Coding region *CHI3L1* SNP associated with sputum YKL-40 levels.



Since we have determined that YKL-40 in the circulation and genetic polymorphisms correlate with asthma severity, we hypothesized that combining YKL-40 levels in the circulation with clinical characteristics in a clustering algorithm would identify pathobiologically and clinically meaningful CHI3L1/YKL-40 endotypes. For this analysis, Dr. Gomez, Co-lead of this project, conducted an unsupervised cluster analysis of 156 individuals with asthma in YCAAD to identify the YKL-40 endotypes. Then, recursive partitioning was used to develop a classifier for cross-validation of the YKL-40 clusters in 167 individuals from the Severe Asthma Research Program (SARP). We also analyzed total sputum cell transcriptome measured by Affymetrix ST 1.0 Gene microarrays to identify pathways that are associated with the resulting endotype groups. Clustering analysis revealed that 4 clusters provided the lowest connectivity in the YCAAD cohort. Two of the groups were characterized by higher sputum and serum YKL-40 levels compared to the other clusters. YKL-40 endotype group 3 (E3) was associated with severe airflow obstruction, early life onset of disease, low FeNO levels, and a history of near fatal asthma attacks, while YKL-40 endotype 4 (E4) was characterized by the highest BMI and significantly higher serum YKL-40 levels, and preserved lung function (Figure 7 and data not shown). Evaluation of the YKL-40 endotype groups in the SARP cohort showed similar differences in age of onset, lung function, and YKL-40 level among groups E3 and E4. Analysis of the E3 transcriptome showed down-regulation of apoptotic pathways consistent with findings in mouse models of asthma outlined above.25 *These studies demonstrate the utilization of YKL-40 levels and clinical features of asthma in a clustering analysis that identified two YKL-40 endotypes of disease characterized by severe persistent airflow obstruction, near-fatal asthma and obesity. These clusters have distinct transcriptomic profiles in the airway. We will use the YKL-40 endotype algorithm in subgroups patients and determine cellular immunotypes and transcriptomes associated with YKL-40 endotype.*

Figure 8: YKL-40 Endotypes are replicated in SARP. P < 0.05 for bars with asterisks



Figure 7: Differences in lung function among YKL-40 Endotypes (P < 0.05 for bars with asterisks)



***Summary****:*

***A. CHI3L1/YKL-40 drives pathobiology in mouse models of asthma.***  (1) YKL-40 is expressed in an exaggerated manner in Th2 inflammation in asthma models. 2) In YKL-40-deficient mice the biologic effects of IL-13 and ovalbumin-driven inflammation are significantly inhibited.

***B. Human studies show that YKL-40 is important to the pathogenesis of asthma.***(1) YKL-40 is expressed in exaggerated quantities in the serum of asthma and correlates positively with asthma severity and airway remodeling, but inversely with lung function in multiple asthma populations. (2) YKL-40 levels are associated with polymorphisms in *CHI3L1* that correlate with asthma prevalence, severity, bronchial hyperresponsiveness, and inversely with lung function. These findings have been validated in multiple studies indicating that YKL-40 is important to the pathogenesis of asthma.

***C. YKL-40 is associated with a disease endotype of asthma in humans.*** (1) Clustering analysis that includes YKL-40 and SARP clustering characteristics identifies 4 subgroups of patients and 2 YKL-40 endotype clusters with elevated levels of YKL-40 in the blood and airway (YKL-40 E3 an E4). YKL-40 E3 is associated with early life onset of disease and airway remodeling, while YKL-40 E4 is associated with higher BMI, adult onset disease, and near normal lung function, compared to the other subgroups. (2) YKL-40 binds to IL-13R2 and activates AKT, MAPK, and Wnt/ catenin.

**Approach.** We will conduct studies using single cell technologies to determine the single cell signatures and functional responses associated with YKL-40 endotypes. Baseline sputum cell samples collected in the NextGen study will be evaluated by CyTOF by the Precision Profiling Core C. Immunophenotyping will be conducted using a customized panel of markers developed in collaboration with Dr. Montgomery, PI of the Precision Profiling Core C to identify M1/M2 macrophages, dendritic cells, NK, Th1, Th2, Th17, Treg, CD8, B cells, Tfh cells as well as YKL-40 pathway signaling molecules including YKL-40, MAPK, AKT, Wnt/Catenin, IL-13R2, and CRTH2, IL-8. SPADE will be used to identify subpopulations of YKL-40 producing cells within each cell lineage.33 The levels of the YKL-40 producing cell will be compared by the different YKL-40 endotypic markers that we have outlined above including: a) blood YKL-40 level, sputum YKL-40 level, *CHI3L1* haplotype, and YKL-40 endotype group. Each subject’s YKL-40 endotype group will be determined using the recursive partitioning algorithm developed by Dr. Gomez as outlined above (See biosketch). Some enrolled subjects will have had *CHI3L1* genotyping done in the past, so when needed *CHI3L1* haplotype will be determined by sequenom as previously described.9 These studies will compare the differences in YKL-40 producing cell populations among the different CHI3L1/YKL-40 endotype markers. As outlined in Core C, CyTOF can be done with as little as 10,000 cells per well, so we know there will be an adequate number of usable samples for these studies (~75% or patients enrolled into the YCAAD phenotyping protocol or 150 subjects).33 Sputum and serum YKL-40 levels will be determined in the Chupp lab using a Meso Scale immunoassay (Meso Scale Discovery, Rockville, MD). Meso Scale technology uses electrochemoluminescence detection and has enhanced sensitivity and reduced background signals compared to standard colorimetric or fluorescent ELISA kits we have used to date.9,29,34,35 One of the important features of this assay compared to standard ELISA is a significantly increased dynamic range and level of detection for sputum samples. We have determined that YKL-40 was below the level of detection in 70% of BAL fluid samples by standard ELISA, and was 100% detectable by Meso Scale that was sensitive to less than 1.0 pg/ml with a dynamic range of 4 logs (Figure 5). Sputum inflammatory cytokine levels will also be correlated with YKL-40 endotypic markers as described in the Precision Profiling Core C. CyTOF computational analyses using Density Resampled Estimate of Mutual Information DREMI and density Rescaled Visualization DREVI developed by Dr. Smita Krishnaswamy, Co-Investigator on Project 3 (See biosketch).36 These computational tools will be particularly relevant to sputum cell stimulation studies conducted on samples collected at longitudinal visits. These studies will define the cellular immunophenotype of cells producing YKL-40 and how these cellular phenotypes differ amongst different YKL-40 subgroups including genetic haplotype, YKL-40 endotype, and serum and sputum YKL-40 level.

In addition to using CyTOF to characterize the populations that produce YKL-40, we will also determine the lineage of cells that are associated with *CHI3L1* gene expression and YKL-40 synthesis. For these analyses, single cell RNA-seq transcription will be conducted at the Keck sequencing facility under the guidance of Dr. Shrikant Mane who has been part of the YCAAD research team since 2008. Details of single cell sample processing and preliminary data showing single cell transcriptomic profiling of the sputum are described in the Precision Profiling Core C. Cellular transcriptomic signatures will be correlated with each subject’s YKL-40 classifiers including: YKL-40 endotype, YKL-40 protein levels in the sputum and blood, and *CHI3L1* haplotype. *CHI3L1* gene expression will also be correlated with the YKL-40 classifiers. For these analyses, we will utilize the computational pipelines and tools that will be developed by Project 3, aim 1. These studies will determine how CHI3L1/YKL-40 expression evolves with the shifts in inflammatory populations.37

To determine the cellular responses to YKL-40, for aim 1c, mixed cell cultures from the sputum will be stimulated *in vitro* with YKL-40, TNF, and Th1/Th2 inflammatory cytokines including IFN-and IL-13 as and analyzed by CyTOF.38 To demonstrate the feasibility of these experiments, mixed culture sputum cells were stimulated with LPS (Figure 10). Using CyTOF, we detected distinct cell subsets by SPADE analysis, where each circular node is a similar population, node color indicates intensity, and adjacent nodes are most similar (Figure 10). A CyTOF antibody panel was designed to identify multiple cell lineages and production of cytokines. The results show isolation of sufficient cells from induced sputum and the ability to detect both the variation between subjects and stimulation-induced production of the pro-inflammatory cytokine TNF in both PMN and monocyte/macrophages from sputum following treatment with LPS. As with all the Projects, we have included in the CyTOF panel surface and intracellular markers of YKL-40 pathway activation and function to best define the baseline level of YKL-40 pathway activation in sputum cells. For stimulation experiments, customized CyTOF panels will be developed in conjunction with Dr. Montgomery on results of baseline analysis. These panels will be used on samples collected from NextGen study at Nextgen longitudinal visits.

***Statistical considerations.*** Data analysis will be conducted by YCAAD’s computational biologist, Dr. Xiting Yan, Director of Precision Pulmonary Medicine Biostatistics (see biosketch and Core B) and in collaboration with the Gerstein/Krishnaswamy team (Project 3, see biosketch).39,40 For all aims, standard parametric and non-parametric statistical approaches will be used for cytokine data analysis depending on the distribution of the data. YKL-40 protein levels are usually not normally distributed.9,41 For the multidimensional data produced by CyTOF, FlowJo is not adequate, and more advanced analysis platforms are required such as SPADE, ViSNE, or Citrus. These clustering programs have been designed for CyTOF datasets to visually and quantitatively gauge the phenotypic diversity between cell types and donors (see preliminary data generated on a sputum sample in Project 3). These algorithms are in routine use in Core C. Dr. Krishnaswamy has a particular interest in conditional density based analysis of cell signaling in single-cell data and has developed conditional-Density Resampled Estimate of Mutual Information (DREMI) to quantify the strengths the influence that a protein X (in this case YKL-40) has on protein Y and conditional-density Rescaled Visualization (DREVI) to visualize and characterize the edge response function underlying protein interactions (see Krishnaswamy et al, Science 2014).36 These analytical platforms are outlined inmore detail in Project 3, aim 2. For single cell RNAseq data, analytical tools developed in Project 3 will be utilized to identify sputum cell lineages that are associated with *CHI3L1* expression and YKL-40 levels.42-45

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sample Size | SD | Table 1: Power for gene expression  For genome-wide FDR | | | | |
|  |  | **0.05** | **0.1** | **0.15** | **0.2** | **0.3** |
| 50 | 0.5 | 100 | 100 | 100 | 100 | 100 |
|  | 0.75 | 99.99 | 100 | 100 | 100 | 100 |
|  | 1.00 | 98.42 | 99.13 | 99.41 | 99.56 | 99.72 |
|  | 1.25 | 87.46 | 91.56 | 93.54 | 94.76 | 96.21 |
|  | 1.50 | 68.49 | 76.10 | 80.25 | 83.02 | 86.62 |
| 75 | 0.5 | 100 | 100 | 100 | 100 | 100 |
|  | 0.75 | 100 | 100 | 100 | 100 | 100 |
|  | 1.00 | 99.95 | 99.98 | 99.99 | 99.99 | 100 |
|  | 1.25 | 97.97 | 98.86 | 99.22 | 99.41 | 99.63 |
|  | 1.50 | 89.08 | 92.77 | 94.52 | 95.58 | 96.84 |
| 100 | 0.5 | 100 | 100 | 100 | 100 | 100 |
|  | 0.75 | 100 | 100 | 100 | 100 | 100 |
|  | 1.00 | 100 | 100 | 100 | 100 | 100 |
|  | 1.25 | 99.75 | 99.88 | 99.92 | 99.95 | 99.97 |
|  | 1.50 | 98.17 | 98.17 | 98.72 | 99.02 | 99.36 |

***Power calculation.*** We calculated power for bulk-RNA-seq gene expression studies because these analyses require the largest sample size to achieve a genome-wide false discovery rate probability <0.05 (Table 1). Assuming that genes less than 2-fold change between two groups are regarded as “no change,” Sizepower (R package) was used to calculate the power with various values of the parameters – false discovery rate (FDR) and the anticipated standard deviation (SD) of the difference in log-expression between groups. The SD of sputum gene expression ranges from 0.4-0.6. FDR is controlled by the mean number of false positives, fixing the anticipated number of genes in an experiment that are not differently expressed at 10,000. Table 1 demonstrates that we have adequate power for an FDR of 0.05 if we generate bulk RNAseq from 150 sputum samples in the YCAAD biorepository to analyze. Power for single cell RNA-seq studies will be significantly higher compared to bulk RNA-seq as there will be many cells captured from each subject.

***Deliverables/expected results.*** The translational infrastructure in place and the preliminary data we have generated demonstrate the feasibility of CyTOF and Single Cell RNA-seq on the sputum. This highlights the strength of YCAAD’s multidisciplinary team and shows that we will achieve our aims. Specifically, we will define the cell populations and signatures associated with YKL-40 endotypes. This will significantly enhance our understanding of the role that YKL-40 plays in asthma pathogenesis. We expect to find that YKL-40 drives the activation of MAPK, AKT, and Wntcatenin and increase IL-8 and TGF-expression, primarily in macrophages of M1 lineage.46 We may also detect novel cell types associated with YKL-40 expression such as cell of mast cell lineage or dendritic cells. While we know that YKL-40 is expressed in many cell types, CyTOF and single cell transcriptome studies will likely show that cells shifting towards the development of remodeling responses will have the highest level of YKL-40 expression. YKL-40 will be modulated by both Th1 and Th2 cytokines, and activation of STAT 6 by IL-13 will drive YKL-40 expression.47 Single cell analysis will reveal cellular lineages and might show YKL-40 is associated with a specific subset of cells or myofibroblast cell lineage consistent with its pro-fibrotic, tissue remodeling properties.38,48,49 We suspect the YKL-40 endotype classifier will show that YKL-40 E3, the remodeled endotype, is associated with remodeling responses while YKL-40 E4 will be associated with sirtuin 1 expression and more Th2 inflammation in the airway.27

***Points of Discussion and Alternative approaches****:*

***Adequate sputum samples.*** The protocols we have developed with CyTOF and single cell RNA-seq are extremely well suited for studies of sputum. CyTOF and single cell capture can be done on as little as 10,000 cells and with single cell capture, we can exclude buccal squamous cell collection base on their very large size compared to inflammatory and bronchial epithelial cells. This is a major advantage and step forward for our program and will increase our yield of high quality sputum RNA samples at least by a factor of 20%. This suggests there will be at least 150 baseline samples available for RNA-seq analysis. We have also demonstrated the feasibility of low input for RNA-seq and single cell capture from the sputum (detailed in Precision Profiling Core C), so there will be adequate cells for all these studies. In addition, we have patients returning for follow-up visits 2 and 3 as part of the NextGen study protocol, so additional opportunities will present themselves for the acquisition of samples if cell isolation or sample quality is an issue.

***Batch effect & statistical issues.*** Batch effect can be a significant issue, so we try to minimize any alterations in protocols and track if any change in processing could affect results. In addition, when detected, we adjust using several statistical approaches as we’ve previously reported.8

***Aim 2: Determine the single cell signatures and functional responses associated with TEA clusters of asthma from baseline and follow-up samples from the NextGen Study (Core B).*** *Studies will be conducted on NextGen study samples to determine the cell populations associated with TEA clusters of asthma determined by single cell CyTOF analysis of sputum cell populations at baseline, b) the single cell transcriptomic signatures and networks associated with TEA clusters and c) the functional responses of the cell populations defined in Aim 1a and 1b to cytokines including TNF, IL-13 and YKL-40.*

**Rationale.** We have generated a novel approach to defining endotypes of asthma by conducting unsupervised clustering analysis of microarray measured genome-wide gene expression in bulk RNA isolated from the sputum. These samples were collected from a heterogeneous population of individuals with asthma.8,50 We have focused on this tissue compartment because it examines gene expression at the site of disease and can be obtained non-invasively. This has allowed for the enrollment of a large number of subjects and captures the “real world” spectrum asthma heterogeneity. This analysis identified 3 transcriptomic endotypes of asthma (TEA), each associated with distinct, clinically meaningful, features of disease and pathways that contribute to the pathophysiology of asthma. In this aim, we will characterize the cell populations and gene expression at the single cell level by CyTOF and single cell RNAseq that are associated with the TEA clusters. Ultimately, these studies will advance our understanding of the TEA clusters that we have identified and the pathways and genes that are driving disease pathogenesis in these subgroups of disease and, potentially, asthma in general. Adding single cell analysis to TEA cluster modeling provides the next level of resolution needed to identify specific cell populations and has the capacity to identify novel mechanistic drivers of asthma heterogeneity.

Figure 9: Pathway based unsupervised clustering of sputum gene expression reveals 3 TEA clusters (columns are samples and rows are pathways).

***Preliminary Studies.*** Clustering endotypes using gene expression profiles has been used effectively to sub-classify a number of diseases, in particular malignant diseases such as breast cancer and lymphoma, but to date only a few studies have attempted this in asthma.51,52 Prior to our description of the TEA clusters, two studies analyzed the sputum transcriptome by microarray analysis: a study of childhood asthma exacerbations and a small adult asthma study that included smokers and did not adjust for batch effect or RNA quality, limiting the conclusions of the study.51,52 In addition, both studies focused on large differences in gene expression between asthmatics and normals to select genes to cluster, a filtering approach that is commonly used but has the effect of clustering only high abundance genes that may be downstream “danger” signals and not constitutively expressed genes that drive endotypes of disease. In contrast to this computational approach, the novel method that Dr. Yan has developed selects genes for cluster analysis that relies on knowledge-based significance and clusters pathways as opposed to individual genes. This method assumes that genes expressed in a given asthma endotype belong to the same or related biological pathways and has proven to be more robust and less biased than those used previously (Figure 9).8

We applied pathway based cluster analysis to microarray measured gene expression from mixed cell sputum RNA sputum collected from subjects enrolled in the YCAAD phenotyping protocol (PI Chupp, R01 HL118346-01, HL095390-01) that included samples that pass quality control criteria for gene expression studies: < 20% squamous cell in the sample (indicating oral contamination) and an RNA integrity number (RIN) > 4.0.8 112 sputum RNA samples fulfilling the inclusion criteria for asthma had global gene expression measured by microarray analysis using Affymetrix ST 1.0 Gene arrays. For this analysis, the distance between individual samples was calculated based on the expression levels of gene sets defined by publically accessible, annotated pathways and uses a pathway expression metric to cluster samples rather than gene expression. We applied this approach using the KEGG pathways in MsigDB. This reduced the gene number to approximately 5,500 genes (compared to ~8,000) that were used to redefine the heatmap and distance matrix of sputum gene expression. As can be seen in the figure, this method is more effective than the traditional approach, showing 3 distinct sputum TEA clusters (Figure 9).

***Correlations between sputum TEA clusters and the clinical, physiologic and biologic phenotypes of asthma.*** To understand the pathobiologic relevance of the TEA clusters, we compared the TEA cluster’s clinical, physiologic, and biologic characteristics (Table 2, Figure 10, where P values are compared among the TEA clusters by Kruskal-Wallace test). While there were no differences in demographic characteristics of disease, there were significant differences between the clusters in post-bronchodilator (FEV1 P=0.006), bronchodilator response (BDR, P=0.03), FeNO (P = 0.03), IL-13 (P=0.04), and YKL-40 (P=0.04)(data not shown).8 Importantly, TEA clusters 1 and 2 were associated with an increased likelihood of intubation or hospitalization for asthma, respectively. Therefore, sputum TEA clusters correlate with severe disease characteristics, airway remodeling, FeNO, Th2 cytokines and a biomarker of remodeling independent of a specific cell population in the sputum (figure 10).



Figure 10: Mean Th2 gene signature (IL-4, 5, 13) among the TEA clusters (P<0.05)

***Validation of TEA clusters using a blood derived gene signature.*** To determine if there is a gene expression signature in the blood that can accurately predict the TEA cluster an individual belongs to, we built a multinomial logistic regression prediction model using gene expression data from matching blood arrays. Genes in the blood that were significantly different between the clusters were selected for this analysis (330 genes). The LASSO algorithm (least squares) was employed to estimate the parameters in the model, which adds a penalty function to the likelihood function to shrink the estimated parameters. The tuning parameter was chosen based on the leave-one-out cross-validation process that provides an estimation for the prediction error. Using this method and gene set, sputum TEA clusters can be predicted with an internal accuracy of 89%. The prediction model was applied to a second set of blood gene expression samples selected from the Asthma BioRepository for Integrative Genomic Exploration (Asthma BRIDGE, N=870) in collaboration with Benj Raby, at the Channing Laboratory at Harvard University School of Medicine. This childhood asthma cohort is 3-4 decades younger than the YCAAD cohort, has a higher percentage of male subjects, and a lower percentage of atopic patients compared to the YCAAD cohort. Gene expression profiling was conducted using HumanHT12 BeadChips (Illumina, San Diego CA), employing a randomized-array allocation strategy to minimize potential confounding by technical batch effects. Quantlie normalization and log2 transformation were conducted with the *lumi* package https://biolincc.nhlbi.nih.gov/home/. Cross-platform selection of genes identified 50 out of 53 genes between the Illumina and Affymetrix platforms for the prediction model. Comparison of the TEA clusters in the BRIDGE cohort validated several observations made in the YCAAD cohort (Table 3). First, the prevalence of each of the TEA clusters was similar to the YCAAD cohort, with TEA cluster 2 and 3 being the least and the most prevalent, respectively (P=2.2x10-16). Second, also similar to the YCAAD cohort is that TEA cluster 1 was significantly more likely to have a history of intubations (P=6.0 x 10-6), and TEA cluster 2 was more likely to have required hospitalization (P=0.01) compared to the other TEA clusters. TEA cluster 3 was less likely to have been intubated or hospitalized in both cohorts. This data suggests that there are stable transcriptomic profiles in the airway associated with transcriptomic profiles in the blood in adults and children with asthma that are linked to gene pathway expression in the airway.

Table 2: Clinical characteristics among the TEA clusters in YCAAD



Table 3: Clinical characteristics of TEA clusters in Asthma Bridge cohort



***Summary.*** *Our seasoned, multidisciplinary team of experts in asthma, genomics, and computational biology has developed a high throughput system to perform transcriptomic analysis of airway samples collected non-invasively. The data generated using this system has identified TEA clusters that are associated with pathobiologically meaningful features of disease. This demonstrates the utility of sputum transcriptomics to resolve the heterogeneity of asthma. However, to obtain a deeper understanding of the molecular mechanisms that underlie the TEA clusters we will define the cell populations and their functional characteristics using CyTOF and single cell RNA-seq. These studies will provide a critical level of detail that will reveal the pathogenetic underpinnings of the TEA clusters and the genes, pathways, networks, and cellular populations that are contributing to these endotypes.*

**Approach.** For these studies, total RNA samples from subjects enrolled in the NextGen Endotyping Study will be collected and processed as outlined in the Clinical Recruitment and Biostatistics Core B. The TEA cluster assignment of each subject will be determined by measuring the blood levels of 50 gene signature outlined above using Nanostring nCounter technology (Seattle WA), in the Center for Precision Pulmonary Medicine. The nCounter Analysis System is ideal for this application as it is high throughput, multiplex capable, and utilizes a digital color-coded barcode technology that offers high levels of precision and sensitivity (>1 copy per cell). The technology uses single molecule imaging to detect and count hundreds of unique transcripts in a single reaction. Each color-coded barcode is attached to a single target-specific probe corresponding to a gene of interest. Mixed together with controls, the probes form a multiplexed CodeSet. We will use this system to profile each individual’s TEA cluster blood signature. The cellular immunophenotype characterized by CyTOF will be determined as outlined above and in the Precision Profiling Core C. Baseline sputum CyTOF immunophenotypes acquired from the Precision Profiling Core C and pathophysiologic features of disease acquired from Core B will be compared among the TEA clusters as described in aim 1. As data becomes available from Project 2, we will also examine the relationship between Tfh cell populaitons, Dkk-1 and IgA-seq results with the TEA clusters. For functional studies of cellular responses, cells will be stimulated as described in aim 1 and customized CyTOF panels will be developed based on findings associated with the TEA clusters at baseline. This will be done in collaboration with Dr. Montgomery and will be used on sputum samples that will be available from NextGen study longitudinal visits (see Precision Profiling Core C for preliminary data on sputum cell stimulation experiments). TNF, Th1/Th2, YKL-40, will be used to determine functional response of cellular subpopulations. Transcriptomic analyses will be conducted on bulk and single cell RNA-seq data made available as outlined in Project 3 to identify the genes that are associated with cellular subpopulations associated with each of the TEA clusters. For these analyses, TEA cluster associated genes in the sputum will be evaluated in both an unsupervised manner and supervised. We will determine if there are cell populations, defined by transcriptomic clustering of single cells, CyTOF immunophenotyping, or traditional methods that are associated with the TEA clusters. This process will be iterative and in collaboration with the Project 3 team. We will identify the molecular phenotype of cells associated with the TEA clusters, the relevant TEA cluster genes and how they relate to phenotypic features of asthma.

***Statistical considerations.*** The statistical approached related to this aim are the same as outlined in detail in aim 1 and will be conducted by Dr. Yan, lead of the Data Management and Biosatatistics sub-Core and in Collaboration with Project 3 investigators Drs. Krishnaswamy and Gerstein. Gene expression difference among the TEA clusters will be compared on bulk RNA-seq and SC RNA-seq samples. Most of the measures made in human asthma demonstrate a non-parametric distribution, so we routinely use non-parametric tests and control for multiple comparisons to be statistically robust even when in combination with normalized gene expression data. This will be the case for endpoints that compare phenotypic differences among the TEA clusters such as cell concentration in the airway, cell viability, post-bronchodilator FEV1/FVC ratio, bronchodilator response, and sputum cytokine levels. Multiple comparisons will be controlled using the Hochberg method.53-55 For the multidimensional nature of CyTOF, we are fortunate to have the expertise of Dr. Krishnaswamy to develop novel tools for the CyTOF analysis. For sputum cell stimulation experiments fold change compared to control will be utilized so that data from different subjects can be combined.

***Power calculation.*** Power considerations are identical to aim 1 for gene expression studies. Since we know the distribution of the TEA clusters in 2 populations of asthmatics are similar, we can assume a similar distribution for these studies. Based on our preliminary data; the accuracy of the nanostring TEA cluster classification tool will be approximately 85%, the predicted prevalence of each TEA cluster, and a difference of 20% for any endpoint, we will have a power of 0.90 to show a significant (p=0.05) difference in the endpoints between the TEA clusters.

***Deliverables/expected results.*** We have demonstrated the feasibility of single cell RNA and CyTOF on sputum cell populations and have shown the ability of a blood signature to determine the TEA cluster of enrolled subjects, so these studies will provide important details on the cell populations that are associated with the TEA clusters, CyTOF studies will likely demonstrate specific subpopulations of cells that are associated with each of the TEA clusters. Since TEA cluster 1 is associated with differential expression of L-histidine decarboxylase (HDC), an enzyme in the histamine metabolism pathway that converts histidine to histamine, we may detect subpopulations of cells with mast cell lineage markers and transcriptional signatures that include C-kit. We anticipate increased responses to Th2 cytokines in these individuals, linking this endotype to higher levels inflammation and severe disease. Since Dr. Bothwell’s laboratory has detected higher levels of Dkk-1 in TEA 1, we expect that protein levels be higher in this cluster (see Project 2, aim 2 for preliminary data). TEA cluster 2 is the most heterogeneous cluster with elevated levels of YKL-40 in the sputum, so we anticipate some overlap between this cluster and YKL-40 endotypes. In TEA cluster 3, the largest, least heterogeneous cluster, with the mildest disease, had the most significant differentially expressed genes compared to control subjects including Defensin DEFB1), an antimicrobial peptide.

***Points of Discussion and Alternative approaches.*** In addition to the discussion points outlined in aim 1 that apply to this aim additional points include:

***Generalizability of the TEA clusters to airway biology.*** We have identified and validated the TEA clusters in adults and children with asthma. One interpretation of these results is that there are common phenomena in the lung amongst individuals that have broader implications to endotypes of other lung diseases.To examine this, we will examine the TEA clusters in publicly accessible datasets from GEO and the Lung Genomics Research Consortium (LGRC). The latter consortium provides access to data from a comprehensive genomic analysis of lung tissue samples from 400 patients with chronic obstructive pulmonary disease (COPD) or interstitial lung disease (ILD), as well as detailed clinical information about those patients.In addition, Next Generation endotypes identified in Project 3 will be compared to the TEA cluster assignments and YKL-40 endotypes, as all of these will be defined in the recruited cohort.

***Aim 3: Define the functional responses associated with endotypes of asthma identified in Projects 2 and 3 compared to YKL-40 endotypes and TEA clusters.*** *Studies will be conducted on samples collected from subjects enrolled in the NextGen study to determine the functional responses of cell populations associated with novel asthma clusters of disease identified in Project 3 by integrated modeling, b) how the immunophenotype and cellular responses differ among different endotypes compared to the integrated clusters, and c) define the relationship between YKL-40 endotypes and TEA clusters with Tfh cells, Dkk-1, and immunogenic microbiota endotypes defined in Project 2.*

**Rationale.** The clustering efforts to date, of clinical variables and the sputum transcriptome, have generated clear evidence that machine learning improves our resolution of asthma heterogeneity and indicates that more asthma specific computational tools are needed to define the full breadth of endotypes that are related to this complex disease. This approach provides us an opportunity to redefine how do sub-classify individuals with asthma. Only then will we be able to truly achieve precision medicine for this disease and achieve curative treatments. The evolution of high throughput sequencing, mass spectrometry analytical devices, and single capture technologies provide an opportunity to define individual responses as opposed to averaging across cell populations and individuals. By examining cellular expression and responses to stimuli at the single cell level, we have the capacity to clearly define the cell populations driving the clinical phenotypes that we currently consider as severity. We are clearly on the precipice of a paradigm shift in how we will look at and study asthma, from a single disease defined by variable airflow obstruction and chronic symptoms to a complex physio-immunologic inflammatory disease juxtaposed between the environment and the genome consisting of numerous endotypes. In this aim, we will leverage to combine expertise of our multidisciplinary team of scientists to characterize the pathophysiologic features and cellular functional responses of the asthma subgroups determined by the integrative clustering outlined in Project 3 as well as endpoints measured in project 2. In contrast to the clustering approaches we used for identification of the TEA clusters, these modeling approaches will be developed using single cell transcriptomic data (SC clusters, see project 3, Table 1) and models that include both clinical features (such as those used to identify the YKL-40 endotypes), cell based signals (protein or transcription), and by gene (logic modeling of disease mechanism). This will entail the integration data described in Aims 1 and 2 of Project 3 to generate integrative clusters of asthma based using clinical data, single cell data, CyTOF and endotype data from Projects 1 and 2 including biomarker and immunogenic microbiome data. This will include integration with clinical and external datasets and use logical modeling to define the regulatory differences between asthma endotypes. Cell-type signatures defined in Aim 2 will be used to deconvolve the bulk-cell RNA-seq data to its component cell transcripts, increasing the effective dynamic range of the cell-specific transcriptional data and facilitating integration with the abundance of bulk-RNA-seq datasets as described in Project 3. These external datasets and the SCyG clusters from Aim 2 will be used to build regulatory network logical models. Each of these clusters, networks and models will be evaluated in the context of established clinical measurements (e.g. FEV1 and FeNO) to identify effective measures to stratify patients and how they might give insight to the mechanisms of asthma disease and heterogeneity. In this aim, we will evaluate the resulting clusters and compare these clusters to YKL-40 endotypes and TEA clusters of asthma. Furthermore, using the iterative approach, we will develop customized experiments and, using samples collected from follow-up visits from the NextGen study, functional studies will be conducted to define the inflammatory response of the integrated clusters. All of these clusters and analyses will be provided to the community on a publicly accessible, searchable, integrated asthma MAP website that will be dedicated to this effort.

**Approach.**Experimental approaches will be similar to that described in aims 1 and 2. Cluster determination of each individual enrolled in the NextGen study will be determined as described in Project 3. There will be several different types of clusters identified depending on the type of data clustered (clinical, CyTOF, or RNAseq, or integrated cluster). Customized CyTOF panels will be developed in conjunction with the Precision Profiling Core C and stimulation experiments will be designed accordingly. Analyses of cellular sub-populations and transcriptomic signatures will be characterized by CyTOF analyses as described in the Precision Profiling Core C proposal and Project 3, aim 3. Cellular stimulation studies will be conducted on visit 2 and 3 samples and customized CyTOF panels will be developed based on the results of the baseline analyses. Stimulation experiments will include Th1/Th2 cytokines, YKL-40, and microbiota antigens identified in project 2. Clinical phenotypes associated with the ISC Clusters will be compared to those associated with YKL-40 endotypes and TEA clusters. Tfh cell, Dkk-1, and IgA-seq measurements will also be integrated into the modeling process as data is generated.

***Statistical considerations.*** Statistical approaches will be similar to those outlined above in aim 1 and 2. We will rely on the expertise of the investigators of all the Projects to enhance these analyses and modeling. The conversations will occur at weekly and ad hoc lab meetings as described in Core A.

***Power calculation.*** Power for these studies is outlined in Aim 1 above.

***Deliverables/expected results.*** These studies will define the clinical phenotypes of the next generation clusters and in comparison to the YKL-40 endotypes and TEA clusters as well as their relationship to Dkk-1, Tfh cells, and immunogenic microbiota. The functional response of airway inflammatory cells among these cells will be defined and used to enhance the specificity of the clustering algorithm and definition of the integrated clusters. We anticipate that there will be some overlap among the various clusters, but anticipate that the integrated clusters will show specific associations to clinical phenotypes such as remodeling and history of flares and possible near fatal asthma. We anticipate that integrated clusters may be associated with specific blood biomarkers and can be tracked from the blood. Importantly, the next generation clusters will add but not replace other clusters of asthma such as the SARP clusters by providing subgroups based on cellular mechanics rather than clinical variables.

***Points of Discussion and Alternative approaches.*** The points of discussion that are associated with this aim are the same as outlined above in aims 1 and 2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table 4: Timeline for Project 1 | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
| Aim 1: Mechanistic Studies of YKL-40 Endotypes |  |  |  |  |  |
| a. Baseline CyTOF studies, YKL-40 assays | XXXXX | XXXXX | XXXXX |  |  |
| b. Bulk and single cell (SC) RNAseq and CyTOF | XXXXX | XXXXX | XXXX | XXXX | XXXXX |
| c. CyTOF and SC RNAseq Stimulation studies |  | XXXXX | XXXX | XXXX |  |
| Aim 2: Mechanistic Studies of TEA clusters |  |  |  |  |  |
| a. Baseline CyTOF studies, Nanostring TEA cluster assay | XXXXX | XXXXX | XXXXX |  |  |
| b. Bulk and single cell (SC) RNAseq studiesCyTOF | XXXXX | XXXXX | XXXXX | XXXXX | XXXXX |
| c. CyTOF and SC RNAseq Stimulation studies | XXXX | XXXX | XXXXX | XXXXX |  |
| Aim 3: Mechanistic Studies of Integrated SC clusters |  |  |  |  |  |
| a. Baseline CyTOF studies, identificatin of ISC clusters | XXXX | XXXX | XXXXX | XXXXX | XXXXX |
| b. Bulk and single cell (SC) RNAseq studiesCyTOF | XXXX | XXXX | XXXXX | XXXXX | XXXXX |
| c. Customized CyTOF and SC RNAseq Stimulation studies | XXXX | XXXX | XXXXX | XXXXX | XXXXX |
| DISSEMINATION Documentation preparation, presentation at national meetings, manuscript submission. Development of Asthma MAP | XXXXX | XXXXX | XXXXX | XXXXX | XXXXX |

**Interactions with Cores and other driving projects.** Project 1 will have extensive interaction with all Next Generation Endotyping of Asthma Heterogeneity Cores and Projects. This speaks to the synergy of the multidisciplinary team and transformative technological and computational approaches outlined in the Yale Asthma U19 proposal.

**Timeline**. The timeline for Project 1 is outlined in Table 4

**References**

1. Centers for Disease C, Prevention. Vital signs: asthma prevalence, disease characteristics, and self-management education: United States, 2001--2009. MMWR Morbidity and mortality weekly report 2011;60:547-52.

2. Fahy JV. Type 2 inflammation in asthma - present in most, absent in many (vol 15, pg 57, 2015). Nat Rev Immunol 2015;15:130-.

3. Huang YJ, Boushey HA. The microbiome in asthma. J Allergy Clin Immun 2015;135:25-30.

4. Thorburn AN, McKenzie CI, Shen S, Stanley D, Macia L, Mason LJ, Roberts LK, Wong CHY, Shim R, Robert R, Chevalier N, Tan JK, Marino E, Moore RJ, Wong L, McConville MJ, Tull DL, Wood LG, Murphy VE, Mattes J, Gibson PG, Mackay CR. Evidence that asthma is a developmental origin disease influenced by maternal diet and bacterial metabolites. Nat Commun 2015;6.

5. Levy BD, Noel PJ, Freemer MM, Cloutier MM, Georas SN, Jarjour NN, Ober C, Woodruff PG, Barnes KC, Bender BG, Camargo CA, Jr., Chupp GL, Denlinger LC, Fahy JV, Fitzpatrick AM, Fuhlbrigge A, Gaston BM, Hartert TV, Kolls JK, Lynch SV, Moore WC, Morgan WJ, Nadeau KC, Ownby DR, Solway J, Szefler SJ, Wenzel SE, Wright RJ, Smith RA, Erzurum SC. Future Research Directions in Asthma: An NHLBI Working Group Report. Am J Respir Crit Care Med 2015.

6. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. Nat Med 2012;18:716-25.

7. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, D'Agostino R, Jr., Castro M, Curran-Everett D, Fitzpatrick AM, Gaston B, Jarjour NN, Sorkness R, Calhoun WJ, Chung KF, Comhair SA, Dweik RA, Israel E, Peters SP, Busse WW, Erzurum SC, Bleecker ER, National Heart L, Blood Institute's Severe Asthma Research P. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. Am J Respir Crit Care Med 2010;181:315-23.

8. Yan X, Chu JH, Gomez J, Koenigs M, Holm C, He X, Perez MF, Zhao H, Mane S, Martinez FD, Ober C, Nicolae DL, Barnes KC, London SJ, Gilliland F, Weiss ST, Raby BA, Cohn L, Chupp GL. Noninvasive analysis of the sputum transcriptome discriminates clinical phenotypes of asthma. Am J Respir Crit Care Med 2015;191:1116-25.

9. Gomez JL, Crisafi GM, Holm CT, Meyers DA, Hawkins GA, Bleecker ER, Jarjour N, Severe Asthma Research Program I, Cohn L, Chupp GL. Genetic variation in chitinase 3-like 1 (CHI3L1) contributes to asthma severity and airway expression of YKL-40. The Journal of allergy and clinical immunology 2015;136:51-8 e10.

10. Chupp GL, Lee CG, Jarjour N, Shim YM, Holm CT, He S, Dziura JD, Reed J, Coyle AJ, Kiener P, Cullen M, Grandsaigne M, Dombret MC, Aubier M, Pretolani M, Elias JA. A chitinase-like protein in the lung and circulation of patients with severe asthma. The New England journal of medicine 2007;357:2016-27.

11. Proceedings of the ATS workshop on refractory asthma: current understanding, recommendations, and unanswered questions. American Thoracic Society. Am J Respir Crit Care Med 2000;162:2341-51.

12. Woodruff PG, Modrek B, Choy DF, Jia GQ, Abbas AR, Ellwanger A, Koth LL, Arron JR, Fahy JV. T-helper Type 2-driven Inflammation Defines Major Subphenotypes of Asthma (vol 180, pg 388, 2009). Am J Resp Crit Care 2009;180:796-.

13. Buss G. [Biologics in asthma: what's new?]. Revue medicale suisse 2015;11:20, 2-4.

14. Darveaux J, Busse WW. Biologics in asthma--the next step toward personalized treatment. The journal of allergy and clinical immunology In practice 2015;3:152-60; quiz 61.

15. Hartl D, Lee CG, Da Silva CA, Chupp GL, Elias JA. Novel biomarkers in asthma: chemokines and chitinase-like proteins. Curr Opin Allergy Cl 2009;9:60-6.

16. Zhu Z, Lee CG, Zheng T, Chupp G, Wang JM, Homer RJ, Noble PW, Hamid Q, Elias JA. Airway inflammation and remodeling in asthma - Lessons from interleukin 11 and interleukin 13 transgenic mice. Am J Resp Crit Care 2001;164:S67-S70.

17. Moore WC. The natural history of asthma phenotypes identified by cluster analysis. Looking for chutes and ladders. Am J Respir Crit Care Med 2013;188:521-2.

18. Hastie AT, Moore WC, Meyers DA, Vestal PL, Li H, Peters SP, Bleecker ER, National Heart L, Blood Institute Severe Asthma Research P. Analyses of asthma severity phenotypes and inflammatory proteins in subjects stratified by sputum granulocytes. The Journal of allergy and clinical immunology 2010;125:1028-36 e13.

19. Ober C, Chupp GL. The chitinase and chitinase-like proteins: a review of genetic and functional studies in asthma and immune-mediated diseases. Curr Opin Allergy Clin Immunol 2009;9:401-8.

20. Dickey BF. Exoskeletons and exhalation. The New England journal of medicine 2007;357:2082-4.

21. He CH, Lee CG, Dela Cruz CS, Lee CM, Zhou Y, Ahangari F, Ma B, Herzog EL, Rosenberg SA, Li Y, Nour AM, Parikh CR, Schmidt I, Modis Y, Cantley L, Elias JA. Chitinase 3-like 1 regulates cellular and tissue responses via IL-13 receptor alpha2. Cell reports 2013;4:830-41.

22. Cho WK, Lee CM, Kang MJ, Huang Y, Giordano FJ, Lee PJ, Trow TK, Homer RJ, Sessa WC, Elias JA, Lee CG. IL-13 receptor alpha2-arginase 2 pathway mediates IL-13-induced pulmonary hypertension. American journal of physiology Lung cellular and molecular physiology 2013;304:L112-24.

23. Zhu Z, Zheng T, Homer RJ, Kim YK, Chen NY, Cohn L, Hamid Q, Elias JA. Acidic mammalian chitinase in asthmatic Th2 inflammation and IL-13 pathway activation. Science 2004;304:1678-82.

24. Zhu Z, Homer RJ, Wang Z, Chen Q, Geba GP, Wang J, Zhang Y, Elias JA. Pulmonary expression of interleukin-13 causes inflammation, mucus hypersecretion, subepithelial fibrosis, physiologic abnormalities, and eotaxin production. J Clin Invest 1999;103:779-88.

25. Lee CG, Hartl D, Lee GR, Koller B, Matsuura H, Da Silva CA, Sohn MH, Cohn L, Homer RJ, Kozhich AA, Humbles A, Kearley J, Coyle A, Chupp G, Reed J, Flavell RA, Elias JA. Role of breast regression protein 39 (BRP-39)/chitinase 3-like-1 in Th2 and IL-13-induced tissue responses and apoptosis. The Journal of experimental medicine 2009;206:1149-66.

26. Elias JA, Zheng T, Lee CG, Homer RJ, Chen Q, Ma B, Blackburn M, Zhu Z. Transgenic modeling of interleukin-13 in the lung. Chest 2003;123:339S-45S.

27. Ahangari F, Sood A, Ma B, Takyar S, Schuyler M, Qualls C, Dela Cruz CS, Chupp GL, Lee CG, Elias JA. Chitinase 3-like-1 Regulates Both Visceral Fat Accumulation and Asthma-like Th2 Inflammation. Am J Resp Crit Care 2015;191:746-57.

28. Gavala ML, Kelly EA, Esnault S, Kukreja S, Evans MD, Bertics PJ, Chupp GL, Jarjour NN. Segmental allergen challenge enhances chitinase activity and levels of CCL18 in mild atopic asthma. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology 2013;43:187-97.

29. Ober C, Tan Z, Sun Y, Possick JD, Pan L, Nicolae R, Radford S, Parry RR, Heinzmann A, Deichmann KA, Lester LA, Gern JE, Lemanske RF, Nicolae DL, Elias JA, Chupp GL. Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function. New Engl J Med 2008;358:1682-91.

30. Rathcke CN, Holmkvist J, Husmoen LL, Hansen T, Pedersen O, Vestergaard H, Linneberg A. Association of polymorphisms of the CHI3L1 gene with asthma and atopy: a populations-based study of 6514 Danish adults. PloS one 2009;4:e6106.

31. Ortega H, Prazma C, Suruki RY, Li H, Anderson WH. Association of CHI3L1 in African-Americans with prior history of asthma exacerbations and stress. The Journal of asthma : official journal of the Association for the Care of Asthma 2013;50:7-13.

32. Cunningham J, Basu K, Tavendale R, Palmer CN, Smith H, Mukhopadhyay S. The CHI3L1 rs4950928 polymorphism is associated with asthma-related hospital admissions in children and young adults. Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology 2011;106:381-6.

33. Yao Y, Liu R, Shin MS, Trentalange M, Allore H, Nassar A, Kang I, Pober JS, Montgomery RR. CyTOF supports efficient detection of immune cell subsets from small samples. Journal of immunological methods 2014;415:1-5.

34. Chupp G, Elias JA. Chitinase-like protein and asthma - Reply. New Engl J Med 2008;358:1075-.

35. Levin JC, Gagnon L, He X, Baum ED, Karas DE, Chupp GL. Improvement in asthma control and inflammation in children undergoing adenotonsillectomy. Pediatric research 2014;75:403-8.

36. Krishnaswamy S, Spitzer MH, Mingueneau M, Bendall SC, Litvin O, Stone E, Pe'er D, Nolan GP. Systems biology. Conditional density-based analysis of T cell signaling in single-cell data. Science 2014;346:1250689.

37. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, Desai TJ, Krasnow MA, Quake SR. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature 2014;509:371-5.

38. Tang H, Sun Y, Shi Z, Huang H, Fang Z, Chen J, Xiu Q, Li B. YKL-40 induces IL-8 expression from bronchial epithelium via MAPK (JNK and ERK) and NF-kappaB pathways, causing bronchial smooth muscle proliferation and migration. Journal of immunology 2013;190:438-46.

39. Gerstein MB, Rozowsky J, Yan KK, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Li JJ, Pei B, Harmanci AO, Duff MO, Djebali S, Alexander RP, Alver BH, Auerbach R, Bell K, Bickel PJ, Boeck ME, Boley NP, Booth BW, Cherbas L, Cherbas P, Di C, Dobin A, Drenkow J, Ewing B, Fang G, Fastuca M, Feingold EA, Frankish A, Gao G, Good PJ, Guigo R, Hammonds A, Harrow J, Hoskins RA, Howald C, Hu L, Huang H, Hubbard TJ, Huynh C, Jha S, Kasper D, Kato M, Kaufman TC, Kitchen RR, Ladewig E, Lagarde J, Lai E, Leng J, Lu Z, MacCoss M, May G, McWhirter R, Merrihew G, Miller DM, Mortazavi A, Murad R, Oliver B, Olson S, Park PJ, Pazin MJ, Perrimon N, Pervouchine D, Reinke V, Reymond A, Robinson G, Samsonova A, Saunders GI, Schlesinger F, Sethi A, Slack FJ, Spencer WC, Stoiber MH, Strasbourger P, Tanzer A, Thompson OA, Wan KH, Wang G, Wang H, Watkins KL, Wen J, Wen K, Xue C, Yang L, Yip K, Zaleski C, Zhang Y, Zheng H, Brenner SE, Graveley BR, Celniker SE, Gingeras TR, Waterston R. Comparative analysis of the transcriptome across distant species. Nature 2014;512:445-8.

40. Krishnaswamy S, Kanteti R, Duke-Cohan JS, Loganathan S, Liu W, Ma PC, Sattler M, Singleton PA, Ramnath N, Innocenti F, Nicolae DL, Ouyang Z, Liang J, Minna J, Kozloff MF, Ferguson MK, Natarajan V, Wang YC, Garcia JG, Vokes EE, Salgia R. Ethnic differences and functional analysis of MET mutations in lung cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 2009;15:5714-23.

41. Chupp GL, Lee CG, Jarjour N, Shim YM, Holm CT, He S, Dziura JD, Reed J, Coyle AJ, Kiener P, Cullen M, Grandsaigne M, Dombret M, Aubier M, Pretolani M, Elias JA. A chitinase-like protein in the lung and circulation of patients with severe asthma. New Engl J Med 2007;357:2016-27.

42. Agarwal A, Koppstein D, Rozowsky J, Sboner A, Habegger L, Hillier LW, Sasidharan R, Reinke V, Waterston RH, Gerstein M. Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays. BMC genomics 2010;11:383.

43. Habegger L, Sboner A, Gianoulis TA, Rozowsky J, Agarwal A, Snyder M, Gerstein M. RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. Bioinformatics 2011;27:281-3.

44. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics 2009;10:57-63.

45. Qian F, Chung L, Zheng W, Bruno V, Alexander RP, Wang Z, Wang X, Kurscheid S, Zhao H, Fikrig E, Gerstein M, Snyder M, Montgomery RR. Identification of genes critical for resistance to infection by West Nile virus using RNA-Seq analysis. Viruses 2013;5:1664-81.

46. Bonneh-Barkay D, Bissel SJ, Kofler J, Starkey A, Wang G, Wiley CA. Astrocyte and macrophage regulation of YKL-40 expression and cellular response in neuroinflammation. Brain pathology 2012;22:530-46.

47. Singh SK, Bhardwaj R, Wilczynska KM, Dumur CI, Kordula T. A complex of nuclear factor I-X3 and STAT3 regulates astrocyte and glioma migration through the secreted glycoprotein YKL-40. The Journal of biological chemistry 2011;286:39893-903.

48. Konradsen JR, James A, Nordlund B, Reinius LE, Soderhall C, Melen E, Wheelock AM, Lodrup Carlsen KC, Lidegran M, Verhoek M, Boot RG, Dahlen B, Dahlen SE, Hedlin G. The chitinase-like protein YKL-40: a possible biomarker of inflammation and airway remodeling in severe pediatric asthma. The Journal of allergy and clinical immunology 2013;132:328-35 e5.

49. Lee CG, Dela Cruz CS, Herzog E, Rosenberg SM, Ahangari F, Elias JA. YKL-40, a chitinase-like protein at the intersection of inflammation and remodeling. Am J Respir Crit Care Med 2012;185:692-4.

50. Steiling K, Christenson SA. Targeting 'types: Precision Medicine in Pulmonary Disease. Am J Respir Crit Care Med 2015;191:1093-4.

51. Baines KJ, Simpson JL, Wood LG, Scott RJ, Gibson PG. Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples. The Journal of allergy and clinical immunology 2011;127:153-60, 60 e1-9.

52. Bosco A, Ehteshami S, Stern DA, Martinez FD. Decreased activation of inflammatory networks during acute asthma exacerbations is associated with chronic airflow obstruction. Mucosal immunology 2010;3:399-409.

53. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behavioural brain research 2001;125:279-84.

54. Benjamini Y, Hechtlinger Y. Discussion: An estimate of the science-wise false discovery rate and applications to top medical journals by Jager and Leek. Biostatistics 2014;15:13-6; discussion 39-45.

55. Benjamini Y, Yekutieli D. Quantitative trait Loci analysis using the false discovery rate. Genetics 2005;171:783-90.