Original research

Lung function, airway and peripheral basophils and eosinophils are associated with molecular pharmacogenomic endotypes of steroid response in severe asthma

Alvin T Kho1,2, Michael J McGeachie2, Jiang Li1,3, Robert P Chase2, Sami S Amr2,4, Annette T Hastie5, Gregory A Hawkins5, Xingnan Li6, Geoffrey L Chupp7, Deborah A Meyers6, Eugene R Bleecker6, Scott T Weiss2,4, Kelan G Tantisira8

ABSTRACT

Introduction Asthma is a complex disease with heterogeneous expression/severity. There is growing interest in defining asthma endotypes consistently associated with different responses to therapy, focusing on type 2 inflammation (Th2) as a key pathological mechanism. Current asthma endotypes are defined primarily by clinical/laboratory criteria. Each endotype is likely characterised by distinct molecular mechanisms that identify optimal therapies.

Methods We applied unsupervised (without a priori clinical criteria) principal component analysis on sputum airway cells RNA-sequencing transcriptomic data from 19 asthmatics from the Severe Asthma Research Program at baseline and 6–8 weeks follow-up after a 40 mg dose of intramuscular corticosteroids. We investigated principal components PC1, PC3 for association with 55 clinical variables.

Results PC3 was associated with baseline Th2 clinical features including blood (rank-sum p=0.0082) and airway (rank-sum p=0.0024) eosinophilia, FEV1 change (Kendall tau-b R=−0.333 (−0.592 to −0.012)) and follow-up FEV1, albuterol response (Kendall tau-b R=−0.392 (0.079 to 0.634)). PC1 with blood basophilia (rank-sum p=0.0191). The top 5% genes contributing to PC1, PC3 were enriched for distinct immune system/inflammation ontologies suggesting distinct subject-specific clusters of transcriptomic response to corticosteroids. PC3 association with FEV1 change was reproduced in silico in a comparable independent 14-subject (baseline, 8 weeks after daily inhaled corticosteroids (ICS)) airway epithelial cells microRNAome dataset.

Conclusions Transcriptomic PCs from this unsupervised methodology define molecular pharmacogenomic endotypes that may yield novel biology underlying different subject-specific responses to corticosteroid therapy in asthma, and optimal personalised asthma care. Top contributing genes to these PCs may suggest new therapeutic targets.

INTRODUCTION

Asthma is a common complex disease with heterogeneity in both expression and severity. Multiple studies have addressed clinically based endotypes of asthma.1–8 Despite heterogeneity in asthma study populations and their defining clinical features/criteria, some universally recognised common clinical subgroups have emerged. These include mild atopic asthma originating in childhood, atopic asthma with lower lung function, and medication-resistant exacerbation-prone asthma. Clinical endotypes have also been consistently associated with different responses to therapy.9 With the advent of next generation biologics for the treatment of severe asthma, endotyping efforts have focused on identifying type-2 airway inflammation (Th2) asthma.10–14 Currently, asthma endotypes have been defined primarily by clinical laboratory or biomarkers criteria.9 For example, the Th2 clinical...
features include fractional exhaled nitric oxide (FeNO) >25 ppb, peripheral blood eosinophil count >300 (or >150, 2 cut-offs considered separately) cells/µL, and sputum eosinophils >2%. There has been no reported molecular clustering in asthma without these a priori clinical criteria. Indeed, several studies have reported distinct transcriptomic differences that appear to correlate with clinical subgroups. The aim of this study is to extend the concept of molecular endotype identification in asthma in novel directions through RNA-sequencing (RNA-seq) of sputum airway cells derived from a subset of 19 asthmatic subjects in the Severe Asthma Research Program (SARP, Wake Forest University centre) at baseline before (baseline), and 6–8 weeks after (follow-up) administration of a 40 mg dose of intramuscular corticosteroid (CS). We used the unsupervised principal components (PCs) of transcriptomic changes following systemic CS treatment to define distinct molecular pharmacogenomic endotypes, derived independently of any a priori clinical criteria and evaluated these endotypes for their association with clinical features (focusing on Th2 ones) and clinical response outcomes.

METHODS
Study population and biological samples
The study population consisted of 19 subjects from the National Heart, Lung and Blood Institute Severe Asthma Research Program 3 (SARP, 3, Wake Forest University centre Institutional Review Board 1507) ClinicalTrials.gov identifier NCT01750411. Twelve subjects in this longitudinal cohort have severe asthma and seven have non-severe (mild to moderate) asthma. All subjects underwent extensive phenotyping including an induced phenotype where each subject had systemic CS responsiveness characterised before (‘baseline’), and 6–8 weeks after (‘follow-up’) the administration of a 40 mg dose of intramuscular CS triamcinolone acetonide. Sputum was obtained at both baseline and follow-up using standardised SARP protocols.

RNA-seq and data analysis
Expanded in online supplemental methods M1. Briefly, unsupervised principal component analysis (PCA) was applied on the subject-wise sample difference (follow-up minus baseline) column vectors of expression of 16 944 genes to identify the dominant directions of sample variation in transcriptome space called PCs focusing on PC1–3. Sample PC coordinates were tested for association with the 55 core clinical variables, online supplemental table S1. Gene ontology (GO) enrichment analysis was performed on the top 5% contributing genes of each PC. All data used in this study will be publicly available on NCBI GEO GSE184433.

RESULTS
The 19 SARP subjects in this study are listed on online supplemental table S2 with select clinical variables at baseline, and their change (follow-up minus baseline) 6–8 weeks after a 40 mg dose of intramuscular CS. Sputum cell samples were obtained at baseline and follow-up, and the 19×2 sputum transcriptome profile were obtained using RNA-seq and analysed. Figure 1 outlines the analysis workflow and key results on these transcriptomic data with select clinical variables, and on a comparable independent 14-subject airway epithelial cells (HBE) microRNAome dataset GSE34466 on NCBI GEO for assessing the generalisability of our strategy and results in silico.

Sputum cell transcriptomic PC1 and PC3 sample coordinate dichotomies are associated with Th2 clinical features
We applied PCA on the CS treatment sample difference/ change (follow-up minus baseline, paired) column vectors of 16 944 genes for all 19 subjects. Without using any clinical variables, PCA identifies the directions of dominant global variance among the 19 subjects in this 16 944-gene dimensional sputum cell transcriptome space. PC1–3 accounted for 28.93%, 19.95% and 11.93% of the global transcriptomic variance, respectively. We plotted the 19 subjects in PC1–3 space, and visual inspection suggests transcriptomic clusters within quadrants of PC1/PC3 space, figure 2A–B.

We dichotomised PC1 and PC3 coordinates at 0, respectively, to define transcriptomic endotypes, and we checked whether these dichotomies/endotypes were associated with CS treatment response in particular Th2 clinical features (eg, FeNO, sputum eosinophil %) and lung function change, table 1.

PC1 dichotomy was significantly associated with baseline blood basophil count (p=0.0191, Wilcoxon rank-sum). PC3 dichotomy was significantly associated with baseline blood eosinophil count (p=0.0082), baseline (p=0.0024) and change (p=0.0181) in sputum eosinophil %. Thus, the molecular endotypes of CS treatment response defined by PC1 and PC3 dichotomies are directly associated with baseline blood basophilia (PC1), baseline blood eosinophilia (PC3), and baseline change in sputum/airway eosinophilia (PC3). We also noted these were no significant associations with FeNO or pre-bronchodilator (BD) FEV1 percent predicted (%p).
We regard the top 5% PC loading coefficient magnitude (~5%) such genes ranked by loading coefficient magnitude as the top contributing genes driving variation along each respective PC coordinate axis. There are ~847/16 944 (~5%) such genes as the top contributing genes driving variation along each respective PC coordinate axis. There are ~847/16 944 (~5%) such genes ranked by loading coefficient magnitude per PC after adjusting for ties and repeat genes. We investigated their GO enrichments separately. In contrast to a supervised analysis, no baseline/follow-up change in sputum macrophage % (R=−0.485 (−0.699 to −0.193)), table 2, figure 4C–D. Change in sputum neutrophil and macrophage % associations remain significant within the 12/19 severe asthma subpopulation. Blood neutrophil count is not associated with PC1–3.

Sputum cell transcriptomic PC3 sample coordinates associated with baseline blood eosinophil count

We consider a subject to have a high blood eosinophil count if their change in pre-BD FEV1 is >10% increase in pre-BD FEV1%p (Kendall tau-b R=−0.333 (−0.592 to −0.012)), follow-up BD response (R=0.392 (0.079 to 0.634)). Within the 12/19 severe asthma subpopulation, both associations became greater, and interestingly PC3 is now significantly associated with baseline (R=0.485 (0.071 to 0.756)) and follow-up (R=0.576 (0.195 to 0.806)) BD response.

Focusing on the change in pre-BD FEV1%p with PC3, we used a ROC curve analysis to assess the diagnostic performance of PC3 for discriminating lung function improvement. We consider a subject to have improved lung function if their change in pre-BD FEV1%p following CS is >0. In this mainly severe asthmatic population (12/19 severe), just 2/19 had >10% increase in pre-BD FEV1%p and 8/19 had >0. The AUCCH value of 0.864 indicate PC3 to be a robust predictor of lung function improvement following CS treatment, figure 3D. That is, PC-defined molecular endotypes resulting from CS treatment response can be used to as a pharmacogenomics predictor of clinical response.

In contrast to a supervised analysis, we similarly investigated top 5% genes ranked correlated (magnitude) with Th2 clinical features resulting from CS treatment response can be used to as a pharmacogenomics predictor of clinical response.

Top contributing genes to sputum cell transcriptomic PC1 and PC3 enriched for immune system and inflammation GOs

We regard the top 5% PC loading coefficient magnitude genes as the top contributing genes driving variation along each respective PC coordinate axis. There are ~847/16 944 (~5%) such genes ranked by loading coefficient magnitude per PC after adjusting for ties and repeat genes. We investigated their GO enrichments separately. In contrast to a supervised analysis, we similarly investigated the top 5% genes rank correlated (magnitude) with Th2 clinical features resulting from CS treatment response can be used to as a pharmacogenomics predictor of clinical response.

Sputum cell transcriptomic PC3 sample coordinates associated with change in pre-BD FEV1 following CS treatment

A notable result when the subject PC3 coordinates is treated as a continuous variable (not dichotomised at 0) is that PC3 is significantly associated with two lung function measurements following CS, table 2, figure 4A–B, namely change in pre-BD FEV1%p (Kendall tau-b R=−0.333 (−0.592 to −0.012)), follow-up BD response (R=0.392 (0.079 to 0.634)). Within the 12/19 severe asthma subpopulation, both associations became greater, and interestingly PC3 is now significantly associated with baseline (R=0.485 (0.071 to 0.756)) and follow-up (R=0.576 (0.195 to 0.806)) BD response.

Focusing on the change in pre-BD FEV1%p with PC3, we used a ROC curve analysis to assess the diagnostic performance of PC3 for discriminating lung function improvement. We consider a subject to have improved lung function if their change in pre-BD FEV1%p following CS is >0. In this mainly severe asthmatic population (12/19 severe), just 2/19 had >10% increase in pre-BD FEV1%p and 8/19 had >0. The AUCCH value of 0.864 indicate PC3 to be a robust predictor of lung function improvement following CS treatment, figure 3D. That is, PC-defined molecular endotypes resulting from CS treatment response can be used to as a pharmacogenomics predictor of clinical response.

Top contributing genes to sputum cell transcriptomic PC1 and PC3 enriched for immune system and inflammation GOs

We regard the top 5% PC loading coefficient magnitude genes as the top contributing genes driving variation along each respective PC coordinate axis. There are ~847/16 944 (~5%) such genes ranked by loading coefficient magnitude per PC after adjusting for ties and repeat genes. We investigated their GO enrichments separately. In contrast to a supervised analysis, we similarly investigated the top 5% genes rank correlated (magnitude) with Th2 clinical features resulting from CS treatment response can be used to as a pharmacogenomics predictor of clinical response.

Top contributing genes to sputum cell transcriptomic PC1 and PC3 enriched for immune system and inflammation GOs

We regard the top 5% PC loading coefficient magnitude genes as the top contributing genes driving variation along each respective PC coordinate axis. There are ~847/16 944 (~5%) such genes ranked by loading coefficient magnitude per PC after adjusting for ties and repeat genes. We investigated their GO enrichments separately. In contrast to a supervised analysis, we similarly investigated the top 5% genes rank correlated (magnitude) with Th2 clinical features resulting from CS treatment response can be used to as a pharmacogenomics predictor of clinical response.

Top contributing genes to sputum cell transcriptomic PC1 and PC3 enriched for immune system and inflammation GOs

We regard the top 5% PC loading coefficient magnitude genes as the top contributing genes driving variation along each respective PC coordinate axis. There are ~847/16 944 (~5%) such genes ranked by loading coefficient magnitude per PC after adjusting for ties and repeat genes. We investigated their GO enrichments separately. In contrast to a supervised analysis, we similarly investigated the top 5% genes rank correlated (magnitude) with Th2 clinical features resulting from CS treatment response can be used to as a pharmacogenomics predictor of clinical response.
features: baseline blood basophil eosinophil counts and FeNO, and change in FeNO.

First, we found strong overlap between top 5% PC1 genes (OR 4.07 (3.22 to 5.13) Fisher exact 95% CI) and top 5% PC3 genes (OR 3.73 (2.94 to 4.73)) with our previously reported glucocorticoid gene set (GCGS), table 3. In contrast, the supervised top 5% Th2 clinical features correlated gene overlaps with GCGS were not significant or had smaller ORs. Also notable is that the top PC1 and PC3 correlated gene overlaps with GCGS were not significant or had lesser enrichments for immune system and inflammation related GO clusters, table S4.

For continuous-valued clinical variables, the median (IQR) and 2-sided Wilcoxon rank-sum p value are shown. For binary-valued clinical variables (eg, asthma severity, change in pre-BD FEV1%) predicted dichotomised at 0), the ratios, ORs and Fisher exact test (95% CI) are shown.

*indicates significance at p<0.05.
First, we applied PCA on the 14-subject 8 weeks minus baseline sample difference column vectors of 862 microRNAs with PC1–3 accounting for 38.64%, 14.71% and 12.13% of the global microRNAome variance respectively, as we had done with the SARP transcriptome above. Second, we found PC3 to be associated with the Severe Asthma Research Program study using Kendall tau-b correlation where * indicates significance where 0 lies outside the respective 95% CI. There associations were also studied within the 12/19 severe asthma subpopulation. *Indicates statistical significance.

![Table 2: Select clinical variable associations of sputum transcriptomic change PC1 and PC3 sample coordinates and severity (binary) variable in the Severe Asthma Research Program study using Kendall tau-b correlation where * indicates significance where 0 lies outside the respective 95% CI.](http://thorax.bmj.com/)

**DISCUSSION**

We identified molecular pharmacogenomic endotypes to an induced phenotype of response to systemic Cs in mainly severe asthmatic subjects using unsupervised PCA of a subject’s sputum airway cell transcriptomic changes.\(^{15}^{16}\) We used resulting the PC1 and PC3 coordinates (figure 2) to define endotypes in two ways: dichotomised at the 0 coordinate point (binary variable) and as is (continuous variable). Respectively, we investigated PC1 and PC3 separately for associations with a set of 55 clinical variables, focusing on Th2 clinical features. We used the top 5% genes contributing to these PCs to uncover biologically and clinically meaningful treatment response genes and ontologies. Crucially, the molecular endotypes defined in this fashion were associated with clinical response to Cs. Namely, PC3 is associated with treatment-related change in pre-BD FEV1 and follow-up BD response (table 2). While the molecular endotypes were also related to baseline Th2 status, with PC1 associated with baseline blood basophils, and PC3 with baseline blood...
and sputum eosinophils (tables 1–2, figures 2–3), none of these baseline Th2 clinical features were themselves associated clinical response to CS (online supplemental table S3). Within the 12/19 severe asthma subpopulation, the baseline blood basophil (PC1) and eosinophil (PC3) associations were diminished, PC3 associations with sputum eosinophil % and change in pre-BD FEV1 and follow-up BD response became greater. The differences between the molecular treatment response endotypes and Th2 status is further emphasised through different biologies represented by ontology enrichment analysis of top loading (contributing)

Figure 3  Areas under receiver operator characteristic curve (AUCCH) analyses for: (A) PC1 sample coordinate and baseline blood basophil count >10 cells/μL. (B) PC3 sample coordinate and baseline blood eosinophil (EOS) count >300 cells/μL. (C) PC3 sample coordinate and baseline blood EOS count >150 cells/μL. (D) PC3 sample coordinate and change in pre-BD FEV1 % predicted >0.

Figure 4  (A) PC3 sample coordinate versus follow-up bronchodilator response. (B) PC3 sample coordinate versus change in pre-bronchodilator FEV1 % predicted. (C) PC3 sample coordinate versus change in sputum macrophage %. (D) PC3 sample coordinate versus change in sputum neutrophil %. Each subject is indicated by their unique 2-digit identifier. 7/19 non severe asthmatics are in cyan. The linear regression line is drawn in each and the Kendall tau-b correlation and 95% CI are indicated.

genes from the PC-defined endotypes (table 5) versus the top correlated genes with any Th2 clinical feature (online supplemental table S4). Finally, we demonstrated that the molecular endotypes are generalisable at the level of the airway epithelial cell, in an independent inhaled CS treatment cohort analysis of microRNA data. Overall, molecular treatment endotypes appear to be more representative of CS induced changes in biology and more specific predictors of clinical CS treatment response.

Underscoring the Th2 biologic significance of our findings, the endotype defined by positive PC1 coordinates mainly has subjects with baseline peripheral blood basophilia, figure 3A. Basophils are underappreciated mediators of Th2 immune response that are increasingly recognised as mediators of thymic stromal lymphopoietin induced Th2 inflammation.30 31 In contrast, the endotype defined by negative PC3 coordinates has subjects with the classic Th2 phenotype of blood and sputum eosinophilia,32 33 figure 3B–C. This emphasises the role that Th2 status may play in response to CS and supports Th2 status as a key underlying feature of our molecular endotypes. From a clinical response perspective, as the PC3 coordinates increased, there was a significant decrease in lung function change following CS administration, figure 4B. Given that blood/sputum eosinophils were not found to be associated with lung function change, the PC3 molecular endotype contains features that represent a more comprehensive biological picture of treatment response than blood/sputum eosinophils alone. This is consistent with previous work suggesting that molecular features might be superior to clinical characteristics in predicting response to CS in severe asthma from children in SARP.35 The top 5% PC1 and PC3 contributing genes were enriched for CS signalling indicated by their significant enrichment for our previously reported GCGS28 table 3. Thus, these mainly severe asthmatic subjects appear to cluster into at least two distinct types of clinically and biologically relevant molecular responses to CS treatment.

We previously applied unsupervised PCA to identify top 5% PC contributing genes to be associated with human fetal lung development in the context of embryonic gestational age, vitamin D signalling and CS response.20 21 36 While those analyses were aimed at expanding basic understanding of lung development biology, our current study suggests that molecular subgroups resulting from such unsupervised analysis may also reflect distinct biologic mechanisms in response to CS treatment. The method may be generally applied to classifying and predicting CS treatment response. In contrast to supervised analyses using clinical variables, unsupervised PCA (and more generally surrogate variable analysis SVA22) has two distinct advantages. First, a gene’s contribution is not limited to a single PC as shown by the strong overlap between top PC1 and PC3 contributing genes, table 4. Indeed this might reflect the interactome being the basis of a systems biology model, or pleiotropy in general. Second, we can systematically identify and prioritise the top N contributing genes per PC to clarify both individual genes and their ontologies that underlie the endotypic subgroups. We highlight the effectiveness of our method by contrasting it with supervised analysis using the top 5% Th2 clinical features (ie, blood and sputum eosinophilia, high FeNO) correlated genes which ultimately had fewer overlapping genes with each other (table 4) and lesser enrichment for GCGS (table 3), immune system and inflammation related ontologies (online supplemental table S4) suggesting less biologic certainty for prominent involvement of these processes.

Our work complements current clinical cluster analyses. A recent SARP analysis focused on CS response in severe asthmatics18 suggested four potential CS response clusters, with the
cluster associated with the greatest CS response demonstrating the highest levels of blood eosinophils. Another current SARP-3 analysis identified differences in pre-BD and post-BD lung function longitudinal trajectories between subjects stratified into four groups by baseline sputum eosinophil % high/low+neutrophil% high/low. While our subset analysis does not have sufficient sample size to assess the presence of such clinical clusters, the association of blood eosinophils and sputum neutrophils with PC3 supports a possible overlap of a portion of the clinical endotype with molecular features. Nonetheless, given the deficiencies of supervised analysis to fully capture meaningful ontological associations, it is also likely that molecular endotyping can independently yield additional novel biology when used as an independent metric, as supported by our ontology enrichment analysis. Others and we previously cited the advantage of molecular endotyping for predicting therapeutic response in asthma. However, these have generally focused on circulating blood cells rather than airway cells, and on baseline differential gene expression differences between predefined responding and non-responding subject populations. Moreover, their prior expression analyses focused on panels of preselected candidate genes. In contrast, our molecular endotyping of each subject is based on whole transcriptomic change in sputum cells following CS treatment. Our model has the unique advantage of capturing dynamic changes inherent in transcriptional response to CS treatment that is agnostic to clinical outcomes/features a priori and gene/molecular preselection biases, and yet our resulting molecular endotypes were found to be associated with clinical variables at baseline and follow-up. Interestingly, 3 genes* from a 6-gene baseline sputum gene expression signature (CLC, CPA3, DNASE1L3, IL1B*, ALPL*, CXCR2*) reported to discriminate inflammatory phenotypes of asthma and predict ICS treatment response are among the top 5% genes contributing to PC1 and PC3 with IL1B ranked 2, online supplemental table S5, Berthon et al noted the 3 genes to be neutrophil markers and

### Table 3
Overlaps between previously reported glucocorticoid gene set (GCGS) with the top 5% PC1 and PC3 contributing genes (unsupervised), the top 5% Th2 clinical feature rank correlated genes (supervised) from Severe Asthma Research Program, and target genes of the top 5% PC3 contributing microRNAs (unsupervised) from GSE34466

| GCGS overlap with each test variable | PC1 28.93% | PC3 11.93% | Baseline blood baso | Baseline blood eos | Baseline FeNO | Change FeNO | GSE34466 PC3 | Test variable | GCGS |
|-------------------------------------|------------|------------|---------------------|-------------------|---------------|------------|-------------|--------------|-------|
| # genes in overlap                  | 7598       | 15592      | 15531               | 15542             | 15544         | 15556      | 16040       | No           | No    |
|                                    | 484        | 490        | 553                 | 545               | 545           | 526        | 522         | No           | Yes   |
|                                    | 753        | 759        | 820                 | 809               | 807           | 795        | 311         | Yes          | No    |
|                                    | 95         | 89         | 26                  | 34                | 34            | 53         | 57          | Yes          | Yes   |

OR (95% CI) 4.07 (3.22 to 5.13) * 3.73 (2.94 to 4.73) * 0.89 (0.60 to 1.33) 1.20 (0.84 to 1.71) 1.97 (1.47 to 2.64) * 5.63 (4.19 to 7.57) *

*Indicates statistical significance.

### Table 4
Overlaps between gene sets from the 19-subject Severe Asthma Research Program study: the top 5% PC1 and PC3 contributing genes (unsupervised), and the top 5% Th2 phenotypes rank correlated genes (supervised)

| Top 5% gene overlaps | PC1 28.93% | PC3 11.93% | Baseline blood baso | Baseline blood eos | Baseline FeNO | Change FeNO | # gene overlap |
|----------------------|------------|------------|---------------------|-------------------|---------------|------------|---------------|
|                       | 510        | 62         | 36                  | 23                | 109           | 0          | OR            |
|                       | 70.28*     | 1.54*      | 0.84                | 0.52*             | 3.06*         | (2.47 to 3.79)* | (95% CI) |
|                       | (59.02 to 83.70)* | (1.18 to 2.01)* | (0.60 to 1.18) | (0.34 to 0.79) |               |             |
|                       | 47         | 1.2        | 1.90*               | 0.84              | 2.14*         | 1.69 to 2.72)* |             |
|                       | (0.83 to 1.52) |             | (1.48 to 2.44)     | (0.60 to 1.18)   |               |             |
| PC3 11.93%            | 73         | 1.88*      | 1.30                | 0.20*             |               |             |
|                       | (1.46 to 2.41)* |             | (0.97 to 1.73)    | (0.10 to 0.38)   |               |             |
| Baseline blood baso   | 84         | 10         | 2.24*               | 0.22*             |               |             |
|                       |            |             | (1.77 to 2.84)*    | (0.12 to 0.41)   |               |             |
| Baseline blood eos    |            |            |                     |                   |               |             |
| Change FeNO           |            |            |                     |                   |               | 180         |
|                       |            |            |                     |                   |               | 6.29*       |
|                       |            |            |                     |                   |               | (5.24 to 7.54)* |

*Indicates statistical significance.

The number of overlapping genes, ORs and Fisher exact test 95% CI are indicated.
PC3 was significantly associated with baseline/follow-upchange in sputum neutrophils, table 2, figure 4D.

The small sample size of our study is an important multifaceted issue. Clinically, severe asthma is heterogeneous and sample size limits our ability to fully characterise this heterogeneity. With regards to generalising the pre-BD FEV₁ prediction to other asthma populations, we addressed this technically in silico in a comparable independent microRNA dataset GSE34466 [27] (cf. Result 6) where we found the resulting microRNAome change PC3 to be associated with change in pre-BD FEV₁ with an opposite signed correlation. This is intriguing considering that expression of microRNAs and their target genes are generally understood to be inversely correlated. Also, MALAT1 (ranked 6 in PC3, and a target of a top 5% GSE34466 microRNAome PC3 hsa-miR-429, online supplemental table S5) is one of five key long non-coding RNAs in an asthma-associated competing endogenous network targeted by eight potential new drugs. [28]

Despite sample size, we achieved potentially meaningful clinical endotypes. It is notable that neutrophils, eosinophils and basophils (peripheral blood/sputum inflammatory cells clinically known to be associated with asthma severity) are relevant to transcriptomic endotyping for steroid response thus suggesting that a simple Th2 versus non-Th2 dichotomy may be insufficient to describe response to inhaled steroids in severe asthma.

## CONCLUSIONS

Our analysis of sputum airway cell RNA-seq data in well-characterised mainly severe asthmatics identified distinct molecular pharmacogenomic endotypes in response to an induced phenotype with systemic CsSs. Since CsSs remain a mainstay of therapy for most asthmas with persistent symptoms, these endotypes may clarify biological differences underlying the heterogeneity in CS treatment response, and provide an objective useful model for predicting treatment response and developing novel therapies. Our analysis suggest that unsupervised PCA/SVA can be used to objectively define testable (across multiple datasets and populations) molecular endotypes that may yield novel biology underlying different responses to CS therapy in asthma, identify potential new therapeutic targets, and improve the identification of subjects most likely to respond to CsSs alone versus those who may benefit from a biologic added to their CsS regimen leading to optimal personalised asthma care. Potentially such molecular endotyping could be applied as a diagnostic test using sputum induction and rapid gene expression measurements designed to delineate therapy specific responses in asthma.

### Author affiliations
1. Computational Health Informatics Program, Boston Children’s Hospital, Boston, Massachusetts, USA
2. Channing Division of Network Medicine, Brigham and Women’s Hospital, Boston, Massachusetts, USA
3. Scientific Division of Network Medicine, Brigham and Women’s Hospital, Boston, Massachusetts, USA
4. Partners Personalized Medicine, Partners Healthcare, Boston, Massachusetts, USA
5. Center for Genomics and Personalized Medicine Research, Wake Forest Health Sciences, Winston Salem, North Carolina, USA
6. Division of Genetics, Genomics and Precision Medicine, University of Arizona Health Sciences, Tucson, Arizona, USA
7. Pulmonary & Critical Care Medicine, Yale School of Medicine, New Haven, Connecticut, USA
8. Department of Pediatrics, Division of Respiratory Medicine, University of California San Diego, La Jolla, California, USA

### Contributors

ATK and KGT have full access to all data in this study and take responsibility for the data integrity and accuracy of the data analyses. Acquisition, analysis or interpretation of the data: all authors. Drafting of the manuscript: ATK, MM, KGT. Study concept, design and supervision: KGT, DAM, ERG, STW.

### Funding

This study was supported by the National Heart, Lung and Blood Institute awards R01 HL127332 (KGT), R01 HL129935 (KGT), R01 HL092197 (KGT), U01 HL058899 (STW), R01 HL132825 (STW) R01 HL139634 (MM), U10 HL109164 (ERB), U1G HL139054 (ERB), R01 HL125734 (ATK), U19 AI118608 (ATK).

### Competing interests

None declared.

### Patient consent for publication

Not required.

### Ethics approval

The institutional review boards of all participating centers.
Supplemental material

This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access

This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, provided the original work is properly cited, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

ORCID IDs

Alvin T Kho http://orcid.org/0000-0002-9887-312X
Jiang Li http://orcid.org/0000-0002-9791-2553
Xingnan Li http://orcid.org/0000-0002-1208-8658

REFERENCES

1. Agache I, Akdis C, Jutel M, et al. Untangling asthma phenotypes and endotypes. Allergy 2012;67:833–46.
2. Kiley J, Smith R, Noel P. Asthma phenotypes. Current Opinion in Internal Medicine 2007;6:196–200.
3. Lönhall J, Akdis CA, Bacharier LB, et al. Asthma endotypes: a new approach to classification of disease entities within the asthma syndrome. J Allergy Clin Immunol 2011;127:355–60.
4. Moore WC, Meyers DA, Wenzel SE, et al. Identification of asthma phenotypes using cluster analysis in the severe asthma research program. Am J Respir Crit Care Med 2010;181:315–23.
5. Wenzel S. Severe asthma: from characteristics to phenotypes to endotypes. Cln Exp Allergy 2012;42:650–8.
6. Woodruff PG. Subtypes of asthma defined by epithelial cell expression of messenger RNA and microRNA. Am J Respir Crit Care Med 2013;10 Suppl:S168–9.
7. Howrylak JA, Fuhligger AL, Strunk RC, et al. Classification of childhood asthma phenotypes and long-term clinical responses to inhaled anti-inflammatory medications. J Allergy Clin Immunol 2014;133:e1-12:1289–300.
8. Moore WC, Fitzpatrick AM, Li X, et al. Clinical heterogeneity in the severe asthma research program. Ann Am Thorac Soc 2013;10 Suppl:S118–24.
9. Bleeker E, Meyers DA, Wenzel S. SARRP: Dissecting subphenotypes and endotypes. In: Chung KF, Israel E, Gibson PG, eds. Asthma Research Program 3 (SARP 3) cohort. J Allergy Clin Immunol 2013;132:72–80.
10. Moore WC, Wenzel SE, Li X, et al. Sputum neutrophil counts are associated with more severe asthma phenotypes using cluster analysis. J Allergy Clin Immunol 2014;133:1557–63.
11. Haste AT, Mauger DT, Dentinger LC, et al. Baseline sputum eosinophil + neutrophil subgroups’ clinical characteristics and longitudinal trajectories for NHLBI Severe Asthma Research Program (SARP 3) cohort. J Allergy Clin Immunol 2020;146:222–6.
12. Solberg OD, Ostrin EJ, Love MJ, et al. Airway epithelial miRNA expression is altered in asthma. Am J Respir Crit Care Med 2012;186:965–74.
13. Sharma S, Kho AT, Chhabra D, et al. Glucocorticoid genes and the developmental origins of asthma susceptibility and treatment response. Am J Respir Cell Mol Biol 2013;52:S33–6.
14. Huang H-Y, Lin Y-C, Li J, et al. miRtarBase 2020: updates to the experimentally validated microRNA–target interaction database. Nucleic Acids Res 2019;116.
15. Salter BM, Oliveira JP, Nascu G, et al. Thymic stromal lymphopoietin activation of basophils in patients with allergic asthma is IL-3 dependent. J Allergy Clin Immunol 2015;136:1636–44.
16. Siracusa MC, Saerz SA, Hill DA, et al. TSLP promotes interleukin-3-independent basophil haematoopoiesis and type 2 inflammation. Nature 2014;147:229–33.
17. Castro M, Corren J, Favorov ID, et al. Dupilumab efficacy and safety in moderate-to-severe uncontrolled asthma. N Engl J Med 2018;378:2486–96.
18. Wenzel S, Castro M, Corren J, et al. Dupilumab efficacy and safety in adults with uncontrolled persistent asthma despite use of medium-to-high-dose inhaled corticosteroids plus a long-acting (I2 agonist: a randomized double-blind placebo-controlled pivotal phase 2B dose-ranging trial. Lancet 2016;388:31–44.
19. Dahlin AT, Tantisira KG. Personalized management of asthma exacerbations: lessons from genetic studies. Expert Rev Precis Med Drug Dev 2016;1:1487–95.
20. Persson H, Kwon AT, RamiloWski JA, et al. Transcriptome analysis of controlled and therapy-resistant childhood asthma reveals distinct gene expression profiles. J Allergy Clin Immunol 2015;136:638–48.
21. Baines KJ, Simpson JL, Wood LG, et al. Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. J Allergy Clin Immunol 2014;133:997–1007.
22. Berthon BS, Gibson PG, Wood LG, et al. A sputum gene expression signature predicts oral corticosteroid response in asthma. Eur Respir J 2017;49, doi:10.1183/13993003.00180-2017. (Epub ahead of print: 29 06 2017).
23. Liao Y, Li P, Wang Y, et al. Construction of asthma related competing endogenous RNA network revealed novel long non-coding RNAs and potential new drugs. Respir Res 2020;21:14.