Title Page

Integrating Lung Tissue and Lavage Proteomes Reveals Unique Pathways in Allergen-Challenged Mice

Thomas H Mahood[1,2,3,6], Christopher D Pascoe[1,2,3,6], Aruni Jha[1,2,3,6], Sujata Basu[1,2,3,6], Peyman Ezzati[4], Victor Spicer[4], Neeloffer Mookherjee[2,3,4,5,6], Andrew J Halayko[1,2,3,6]

[1]*Department of Physiology & Pathophysiology, University of Manitoba, Winnipeg, Manitoba, R3E 0J9, Canada*

[2]*DEVOTION Network, Winnipeg, Manitoba, Canada*

[3]*Biology of Breathing Group, Children's Hospital Research Institute of Manitoba, Winnipeg, Manitoba, R3E 3P4, Canada*

[4]*Manitoba Centre for Proteomics and Systems Biology, Department of Internal Medicine, University of Manitoba, University of Manitoba, Winnipeg, Manitoba, R3E 3P4, Canada*

[5]*Department of Immunology, University of Manitoba, Winnipeg, Manitoba, R3E 0T5, Canada*

[6]*on behalf of the Canadian Respiratory Research Network, http://respiratoryresearchnetwork.ca/, Ottawa, Ontario, Canada*

Running Title: Lung Tissue and Lavage Proteome

Abstract

Independent proteomic analysis do not capture the biological interactions between the tissue and extracellular biological compartments when examined in isolation. To address this, we analyzed and compared the proteome from lung tissue and from bronchoalveolar lavage fluid (BALF) of individual allergen-naïve and allergen-challenged BALB/c mice, a common pre-clinical model of allergic asthma. Collectively, we quantified 2,695 proteins from both tissue and BALF of allergen-naïve and -exposed mice. We created an integrated dataset to examine tissue-BALF proteome interactions. Multivariate analysis identified this Integrated-Tissue-BALF (ITB) dataset as being distinct from either the lung tissue or the BALF dataset. Pathway and network analysis of the ITB dataset uncovered protein hubs that span both the tissue and BALF, but are not significantly enriched in either sample dataset alone. This work reveals that by combining individual datasets provides insight about protein networks that integrate biology in lung tissue and the airway space in response to allergen challenge.

Keywords: bronchoalveolar lavage / asthma / bioinformatics / allergen / proteomics

Introduction

At the organ level, the biological underpinnings of chronic inflammatory disease involves complex molecular interplay between local structural cells, recruited cells and the extracellular mediators that they each release. Such is the case for asthma, a chronic disorder of the airways that requires new therapies to fully control steroid-resistant inflammation. Mice challenged with allergen are a mainstay for pre-clinical asthma research. A common approach is to use house dust mite (HDM) as an aeroallergen as it is clinically relevant to human disease, and represents a complex stimulus that includes multiple immunogens and stressors, including fungal spores, bacterial endotoxins, lipid-binding proteins and proteases (Calderón *et al.* 2015; Choopong *et al.* 2016). Repeated HDM challenge induces pathophysiologic symptoms that are hallmarks of human asthma (Jha *et al.* 2018), however the scope and complexity of the interplay between lung cells, recruited inflammatory cells, extracellular mediators of autocrine and paracrine significance, and the integrated signaling pathways that lead to pathobiology and can determine the efficacy of pre-clinical asthma therapeutics, has not been refined as a systems biology model.

To understand complex and integrated disease mechanisms, a number of omics technologies have been employed, including proteomics. Some studies examine the proteome of individual biological compartments, including airway spaces - collected as bronchoalveolar lavage fluid (BALF) or sputum - and lung tissue from asthmatic patients and murine models of the disease (Wu *et al.* 2005; O'Neil *et al.* 2011; Burg *et al.* 2018). These studies have been important for endotyping patients and animal models, identifying biomarkers of disease and providing direction for new therapeutic strategies. However, analysis of these biological compartments in isolation does not enable identification of integrated molecular networks that that are critical for disease expression. Thus, despite some advances in identifying extracellular biomarkers or tissue response networks, understanding the molecular systems that are affected by the interaction between secreted proteins in the lung and the pathways that are regulated in the

resident cells of the lung tissue has not been fully established.

To address this, we used unbiased proteomic analysis to establish a molecular signature in the lung through integrated analysis of matched lung tissue and BALF, and comparing allergen-naïve and HDM-challenged mice. We found that inhaled HDM challenge induces relatively distinct proteome signatures in lung tissue and BALF. Using a bioinformatic approach, we developed an Integrated-Tissue-BALF (ITB) proteome that uncovers signaling networks that are not evident from proteomic datasets of lung tissue and BALF individually.

Results

Lung function & differential cell count analysis

For all mice, we characterized hallmark pathophysiological features that developed as a result of repeated HDM challenge for two weeks in 6-8-week-old female BALB/c mice. We performed lung function and differential immune cell count analyses to assess the phenotype of individual animals (Supplemental Figure S1). HDM challenge resulted in features that are consistent with human asthma, including increased methacholine-induced airway resistance (adjusted p-value, p.adj ≤ 0.001), tissue elastance (p.adj ≤ 0.001) and tissue resistance (p.adj ≤ 0.001) at 50 mg/mL methacholine (two-way nested ANOVA with Tukey's multiple comparison and FDR correction, $n = 6$) (Supplemental Figure S2A). Differential counting of BALF immune cells revealed that HDM challenge triggered their accumulation, including eosinophils and neutrophils (Supplemental Figure S2B).

Lung tissue, BALF and Integrated-Tissue-BALF proteomes have distinct profiles

In total, our proteomic analysis of lung tissue and BALF samples across all mice, yielded 2695 protein identifications (IDs). We obtained 1595 ± 155 protein IDs from lung tissue and 581 ± 201 (mean ± SD) protein ID's from BALF (Supplemental Figure S3A,B).

HDM exposure increased the absolute number of proteins in the lung tissue by 16 % (357), and by 53 % (621) in the BALF. The relative proportion of protein ID's that are unique to tissue decreased from 78.4 % to 61.4 % after HDM exposure. In contrast, the proportion of unique BALF proteins was relatively unchanged by HDM challenge (29.0 % and 28.2 %, respectively). The relative proportion of protein ID's that are shared between the BALF and lung tissue increased from 19.7 % to 33.47 % after HDM exposure (Figure 1A).

To confirm that secreted proteins are enriched in BALF, we characterized the protein ID's using Uniprot annotation keywords. Of the proteins that could be annotated as "secreted" we found that 29 % and 81 % of the protein ID's were classified as tissue and BALF respectively. Therefore, the BALF is enriched in secreted proteins by 279 %, compared to lung tissue. Of the proteins that could be annotated as "transmembrane" we found that 71 % of proteins were associated with the lung tissue proteome, compared to only 19 % of BALF proteins. Therefore, the tissue is enriched in cell-associated proteins by approximately 373 %, compared to BALF.

Using the normalized log2 protein expression values, we calculated z-scores (on a sample-by-sample basis) to assess the relative contribution of protein signatures from the tissue or BALF on a common scale. Once calculated for each sample, z-scores for HDM exposure were corrected by subtracting each sample by the mean z-score of the naïve samples. To assess the integrated biological contribution of BALF and tissue proteomes from HDM challenged mice, z-scores from individual datasets were summed. This Integrated-Tissue-BALF (ITB) dataset includes 2,695 protein IDs.

To investigate if the proteome from lung tissue, BALF and ITB datasets are distinct we performed Partial Least Squares Discriminant Analysis (PLSDA). Figure 1B shows that the individual 95 % confidence interval ellipse for lung tissue, BALF and ITB proteome are segregate. Variance Importance in Projection (VIP) scoring identified the proteins that are the

most significant discriminators of the three datasets. We identified 30 targets that had a VIP score ≥ 2.37. Of these, eight proteins (Rab1a, Fn1, Sftpa1, Clic4, Npepps, Iqgap1, Flna, Actr3) comprised a group with the highest VIP scores, positioned above a natural inflection point set by Actr3 (VIP = 2.64) (Figure 1C). We used matched lung tissue and BALF from individual mice, thus we were not only able to discriminate proteins that were changed after HDM challenge, but also detected those that were enriched or diminished specifically in BALF or lung tissue. For example, the VIP Score projection plot (Figure 1C) reveals that fibronectin-1 (Fn1) was diminished in tissue (z-score -0.816), enriched in BALF (z-score 0.91), and appeared to be unchanged if considered in the ITB proteome (z-score 0.093). This suggests that the change in protein abundance in one biological compartment, may be a reflection of change in another compartment.

Unique and significantly enriched proteins from tissue, BALF and ITB proteomes

We identified 1,818 proteins unique to lung tissue and BALF datasets across both the HDM challenged and allergen-naïve mice. These are revealed in Figure 2 by plotting individual protein z-scores (assigning a z-score of -4 for any absent value) from the naïve and HDM treatments across the lung tissue, BALF, and ITB datasets. This approach classified proteins in lung tissue, BALF, and ITB datasets as being: 1) unique to HDM-challenge (located along y-axis of figure inserts in panels 2A-C); 2) unique to allergen-naïve (located along the x-axis of figure inserts in panels 2A-C); or, 3) common to HDM-challenged and allergen-naïve mice (located within the ellipse regions of figure inserts in panels 2A-C). Of the proteins that are unique to HDM exposed mice, we detected 529 in lung tissue, 654 in BALF, and 699 in the ITB dataset. For the proteins identified only in allergen-naïve mice, 174 existed in lung tissue, 34 in BALF, and 171 in the ITB dataset. Of the proteins that were common in samples from HDM challenged and allergen-naïve mice, we detected 1671 in tissue, 534 in BALF, and 1825 in the ITB dataset.

Using a statistical analysis of microarrays (SAM) workflow on the proteins that were common to samples from allergen-naïve and HDM-challenged mice, we identified those proteins that were significantly enriched or diminished in tissue (124 enriched; 134 diminished), BALF (92 enriched; 56 diminished), and the ITB dataset (51 enriched; 35 diminished). These proteins are highlighted in the volcano plots depicted in Figure 2A-2C. Of these differentially regulated proteins, a number were uniquely affected in lung tissue or BALF, while only a small number emerged as being differentially regulated exclusively in the ITB dataset.

To identify proteins that are common in the lung tissue and BALF proteomes, but differentially regulated in each dataset by HDM exposure, we plotted the average BALF z-score *versus* the average lung tissue z-score for each protein (Figure 2D). Of 27 proteins *enriched in tissue, but diminished in the BALF proteome,* 10 were confirmed to be significantly enriched (Benjamini-Hochberg multiple comparison correction; Limma R package, p.adj ≤ 0.05) (Table 1). Reactome pathway analysis using the 10 most significantly enriched proteins, identified glutathione synthase deficiency as the most significantly enriched pathway (p.adj = 6.43×10^{-4}), (Figure 2E). Of the 200 proteins *enriched in BALF but diminished in the lung tissue proteome,* 141 were confirmed to be significantly enriched (Benjamini-Hochberg multiple comparison correction; Limma R package, p.adj ≤ 0.05) (Table 2). InnateDB pathway analysis from these 141 proteins identified platelet degranulation as the most significantly enriched pathway (p.adj = 8.17×10^{-11}) (Figure 2F).

Unique biological processes induced by allergen challenge in tissue, BALF and ITB

Using InnateDB we identified biological responses that are significantly altered by HDM challenge as represented in the lung tissue, BALF, and ITB datasets (Figure 3). In the lung tissue proteome a number of dataset-unique processes were induced. The top five statistically significant processes were: small molecule metabolism (p.adj = 1.04×10^{-10}); antigen processing

and presentation (p.adj = 2.19 $x10^{-9}$); activation of PKC (p.adj = 5.82 $x10^{-7}$); activation of the adaptive immune system (p.adj = 6.57 $x10^{-7}$); and, platelet degranulation (p.adj = 9.50 $x10^{-7}$) (Figure 3A).

In BALF several dataset-unique biological processes were also evident, with the top five responses being: cell cycle arrestment through degradation of Cyclin D (p.adj = 5.39 $x10^{-21}$); dysfunctional binding of β-catenin through altered AMER1 and APC (p.adj = 6.07 $x10^{-21}$); altered WNT signaling through TCF7L2 frameshift mutations (p.adj = 6.07 $x10^{-21}$); and, GSK3 stabilization and nuclear localization through β-catenin mutations (p.adj = 6.07 $x10^{-21}$) (Figure 3B).

In the ITB dataset we identified 19 pathways for which tissue was the major contributing factor (Figure 3C). The top 3 most significant pathways included, trafficking and processing of endosomal TLR (p.adj = 1.43 $x10^{-4}$), antigen processing and presentation (p.adj = 4.85 $x10^{-4}$), and cross-presentation of particulate exogenous antigens (phagosomes) (p.adj = 4.97 $x10^{-4}$). We also identified 6 pathways for which BALF was the major contributing factor in the ITB proteome (p.adj ≤ 0.05) (Figure 3D), with the top three being antigen processing (ubiquitination & proteasome degradation, p.adj = 2.22 $x10^{-13}$), recycling pathway of L1 (p.adj = 3.14 $x10^{-7}$), and post-chaperonin tubulin folding pathway (p.adj = 2.35 $x10^{-5}$).

We identified 11 pathways that were only significantly enriched after the proteome datasets from the lung tissue and BALF were integrated, thus were unique to the ITB dataset (p.adj ≤ 0.05) (Figure 3E). The top processes included: endosomal transport (ESCRT) (p.adj = 0.0217); phospholipid metabolism (p.adj = 0.0332); glycerophospholipid metabolism (p.adj = 0.0334); synthesis of IP3 and IP4 in the cytosol (p.adj = 0.0352); and, degradation of the ECM (p.adj = 0.0353).

Disparate protein-protein interactions in BALF and tissue datasets

The underpinnings of the biological processes affected by allergen challenge in lung tissue and BALF proteome datasets lies in the protein interaction hubs and networks that are engaged. To decipher these pathways we used NetworkAnalyst and identified the most significantly induced networks, based on the number of first order protein-protein interactions, but independent of the fold-change in abundance of individual proteins. We extracted the top 3 protein-protein interaction hubs in the individual datasets for the lung tissue and BALF datasets (Figure 4). In lung tissue, the top three protein interaction hubs were for Hdac1, Ctnnb1 and Smarca4 (Figure 4A). Collectively, these lung tissue protein interaction hubs create a network associated with adherens junctions interactions (p.adj = 7.44 $x10^{-5}$). In the BALF dataset, the top three first order interaction hubs were Akt1, Dnm1 and Csf1r (Figure 4B). The biological pathways associated with these interaction hubs were distinct from those of lung tissue proteome. The resulting network formed from these three BALF hubs was associated with immune system pathways (p.adj = 2.00 $x10^{-14}$).

To assess the impact of integrating the lung tissue and BALF datasets on the top three pathways predicted from each dataset individually, we reassessed the degree of networking for each of these six hubs in the ITB dataset (Figure 4C). Dataset integration increased the average number of first order protein-protein interactions by 2.83 per hub (a 7.7 % increase). For the top interaction nodes first identified in the lung tissue dataset alone (Figure 4A), the biological pathways associated to each protein hub were unchanged in the ITB dataset (Figure 4C). However, for the most significant interaction hubs first identified in the BALF dataset alone (Akt1, Dnm1 and Csf1) (Figure 4B), the increased number and diversity of protein-protein interactions created in the ITB dataset changed the biological process predicted from “immune response” to “adaptive immune system pathways” (p.adj = 9.70 $x10^{-13}$). Finally, for the ITB dataset, “developmental biology” (p.adj = 1.49 $x10^{-13}$) emerged as the most significantly associated biological pathway when we assessed the integration of the six interaction hubs

identified in lung tissue and BALF proteomes individually. These data demonstrate the diversity of insight that can be obtained from individual biological compartments of the lung, and from an ITB dataset that includes both components.

Protein-protein interaction enrichment in the Integrated-Tissue-BALF dataset

We catalogued the new interactions that emerge in the ITB dataset to explore how unifying lung tissue and BALF datasets affects the scope of the predicted protein-protein interactions (Table 3A and 3B). For this purpose, we developed a combined dataset that included 1,346 proteins unique to the lung tissue, and 333 proteins unique to the BALF proteome after HDM challenge. In total this comprised 1,679 proteins and is hereafter called the "Unique-to-Combined" (UtC) dataset. The UtC dataset is distinct from the ITB dataset as it *excludes* the 845 proteins that we identified in both lung tissue and BALF. This enables more direct critical assessment of the potential integration of networks between compartments. We limited our analysis of protein hubs that had at least five first order interactions in the UtC dataset, and assessed the number of new first order interactions that emerged for lung tissue- or BALF-specific proteins.

First order interactions specifically identified for BALF-specific proteins were enriched 42.1 ± 21.8% (range of 12-85%) in the UtC dataset (Table 3A), whereas protein hubs identified from the lung tissue-specific proteome were only enriched 20.4 ± 5.8% (range of 14-33%) (Table 3B). To more clearly decipher the impact of combining unique to-lung tissue or –BALF proteins we determined first order protein interaction networks in the UtC dataset using the Reactome module within NetworkAnalyst. To streamline our analysis, we specifically examined the effects on networks developed from the top three protein hubs that emerged from unique to-lung tissue data, and top four unique-to-BALF dataset (the 3rd and 4th ranked hubs had identical enrichment scores).

For hubs involving lung tissue-unique proteins, Hgs, Arhgef7 and Akap8 were most enriched by integrating BALF proteins (Figure 5A), however there was limited interaction evident between individual hubs. The network of these three interaction hubs identified "signaling by interleukins" (p.adj = 1.49 $x10^{-13}$) as the most significantly associated biological process. For hubs formed by unique-to-BALF proteins, all hubs were interconnected, for example, two independent connections were evident between Casp7 and Spp1 (Figure 5B). The network formed by the four BALF-unique proteins (Dnm1, Tceb1, Spp1 and Casp7) identified "axon guidance" (p.adj = 0.0224) as the most significantly associated biological process.

To further assess the impact of integrating unique-to-lung tissue and -BALF proteins in the UtC, we reassessed the degree of networking for each hubs formed by lung tissue- and BALF-unique proteins (Figure 5C). In the UtC dataset we identified a large number of new connections between individual proteins and between protein hubs, highlighting the strong potential for protein-protein interactions between the lung tissue and BALF compartments. The resulting increased number of protein-protein interactions at these hubs in the UtC dataset predicted new biological functions. For example, new interactions in the UtC dataset for the lung protein hubs (Hgs, Arhgef7 and Akap8) changed pathway association from signalling by interleukins (Figure 5A) to being "SMC binds to IAPs" (p.adj = $7.24x10^{-5}$). The effect of data integration in the UtC dataset also changed the biological function predicted for unique-to-BALF protein hubs (Dnm1, Tceb1, Spp1 and Casp7), resulting in a prediction for EGFR downregulation ($4.31x10^{-7}$). Together these results suggest that the UtC dataset enriches the number of protein-protein interactions within individual networks. This enrichment enhances interconnectivity of proteins between biological compartments of the lung and refines that predicted biological significance of these networks.

Discussion

Murine models of allergic airways inflammation employ inhaled aeroallergen such as HDM to support preclinical and discovery research. Asthma pathobiology involves recruited inflammatory cells and lung structural cells that interact to define pathobiological processes. In this study, we used label-free proteomics and multivariate bioinformatics to describe and compare the molecular interactome in BALF and lung tissue specimens from HDM challenged mice. We used matched samples from a cohort of individual mice for this process, and in so doing have been able to discriminate responses in tissue and the airspaces, and through *in silico* re-integration, predict the interactions between the lung tissue and the BALF. We demonstrate that the proteome between the lung tissue and BALF is significantly different in mice, with each sample exhibiting unique proteins that are differentially changed by HDM challenge, revealing unique biological responses in each. We generated an Integrated-Tissue-BALF dataset to enable network analysis that reveals points of interaction between lung tissue and airspace proteins and pathways. Our study provides a platform that reveals the scope and limits of biological insights that can be obtained from lung tissue or BALF sample proteins alone and offers the potential to interrogate network interactions between the lung tissue and extracellular airway space during allergen challenge.

Most ‘omics’ bases studies, including proteome profiling, usually focuses on what is enriched/depleted or up-/down-regulated in a treatment in a tissue or individual bio-sample, compared to a control. Though this is insightful, it is often not possible to distinguish whether proteins are from cells, released by cells, or both in a tissue. The source of the proteins can be predicted using most available informatics tools, though their annotation is completely based on if that protein resides within the tool being used. In this study we demonstrate that investigating individual compartments of one organ, the lung, then integrating them post hoc can provide clarity about the source of the biological processes in the whole system. In this study we

developed a methodology to integrate proteomes from two separate, but biologically linked compartments. To enable this, we sampled lung tissue and airway lavage from each animal to separate sample-unique proteomes. We also used untargeted proteomics to provide a broad, unbiased survey of changes in BALF and lung tissue pre- and post- allergen challenge. Once proteome data was acquired we employed non-biased statistics to integrate the two datasets, specifically using z-score analysis that normalized for differences in total protein abundance in different samples. This approach allowed us to consider the biological interaction between the tissue and airspace compartments of the murine lung.

We used an untargeted approach to compare the lung tissue and BALF proteomic changes which occur during repeated aeroallergen challenge. To do this, we collected animal-matched samples, meaning that the lung tissue proteome was determined in lungs after BALF was collected; as such, the protein fingerprint linked to secretory function of egressed inflammatory cells and structural cells is enriched in BALF, and diminished in the lung tissue proteome. To create a ‘pseudo whole lung’ proteome, we integrated individual datasets post hoc to create the ITB dataset, rather than collecting a proteome dataset from lungs that had not undergone BALF collection. Our design was necessary to enable discrimination of unique effects in the lung tissue and BALF compartments. Importantly, prior to proteomic analysis, BALF samples were centrifuged to remove immune cells for counting and differential analysis. Thus, our integrated dataset does not capture intracellular proteins from immune cells that had migrated into the airway space.

Biological pathways in the lung tissue proteome

The lung tissue dataset had a 447 % more unique protein ID’s compared to BALF. When examining the top 10 unique tissue pathways, the immune system was found to be a significant biological signature. These pathways were associated with an acute immune response signature,

specifically allergen detection and antigen processing/presentation. This likely reflects the significant immune cell population that resides in the airway wall and the lung interstitium, but may also include the small number of cells remaining in the pulmonary circulation. Upon allergen challenge, epithelial, mesenchymal and resident immune cells are stimulated and secrete cytokines, chemokines and other pro-inflammatory mediators that have auto- and paracrine effects on structural cells such as fibroblasts and airway smooth muscle cells (Halayko and Amrani, 2003). These cells have a significant immunomodulatory capacity, and also secretion mediators to orchestrate local inflammation and epithelial cell biology. This promotes inflammatory and fibro-proliferative processes in lung tissue that are hallmarks of the response to allergen exposure (Halayko and Amrani, 2003). An immune signature in lung tissue and BALF has been described to represent a Th2-polarizied response after challenge with inhaled HDM (Piyadasa *et al.* 2016), and as we confirmed in the current study, includes significant tissue infiltration by eosinophils and neutrophils. Together this data suggests that immune responses in lung tissue likely reflect the coordinated activity of resident structural cells and immune cells that has infiltrated lung tissue.

Biological pathways in the BALF proteome

In our BALF proteomic dataset, we detected a strong biosignature for pathways involving β-catenin, which has recognized roles in regulation of cell-cell adhesion and gene transcription (Baarsma *et al.* 2013; Koopmans *et al.* 2017). Within the airways, epithelial barrier function is maintained through adherens junctions in which β-catenin interacts with its neighbouring partners α-catenin, p120 and e-cadherin. Through repeated allergen exposure, e-cadherin is down-regulated to disrupt β-catenin function and barrier repair signaling through epithelial-growth-factor-receptor mediated signaling (Heijink *et al.* 2007). This is also associated with the secretion of Th2 immune mediators (Heijink *et al.* 2007), and local cell damage and necrosis in

response to HDM challenge (Chan *et al.* 2016; O'Neil *et al.* 2011; Petta *et al.* 2017). Our proteome analysis indicates that epithelial cell denudation and barrier disruption during allergen challenge contributes significantly to molecular mechanisms that can be detected in BALF.

Biological pathways in the Integrated-Tissue-BALF proteome

We identified several biological mechanisms that emerge as significant processes only after integrating the proteome of the lung tissue and BALF. This appears to arise from an enrichment of the number of pathway-specific proteins, and the additive effect on the abundance of proteins that are common to lung tissue and BALF. As an example, the ITB dataset uniquely reveals processes for generating multi-vesicular bodies that both target protein for ubiquitination leading to turnover in lysosomes (Karim *et al.* 2018), and for the biogenesis of extracellular vesicles (Colombo *et al.* 2013). This specifically relates to proteins associated with multiunit Endosomal Sorting Complexes Required for Transport (ESCRT), including endosomal sorting protein Chmp6, and vesicle sorting (Vps4b) and trafficking (Vta1) proteins from lung tissue, and ubiquination proteins (Tonsoku Like, DNA Repair Protein (Tonsl) and ubiquitin B) and vacuole sorting proteins (Vps25 and Vps4a) from BALF. Interestingly, ESCRT is associated with secretion of inflammatory mediators and protein turnover, key processes in inflammation and tissue remodeling processes (Kulshreshtha *et al.* 2013; Karim *et al.* 2018).

Another example of a biological pathway that is significantly enriched only in the ITB dataset relates to the degradation of the extracellular matrix (ECM). BALF included a number of ECM proteins, including collagen (Col17a1) and fibronectin (Fn1). Lung tissue included multiple enzymes that modulate ECM homeostasis, such as nicastrin (Ncstn, a member of the gamma secretase complex), cathepsins G, L1, and S, neutrophil elastase (Elane), and matrix metalloproteinase-9 (MMP9). The emergence of both ESCRT signaling and ECM turnover only after combining the lung tissue and BALF proteomes suggests that mechanisms that define

interactions between lung tissue cells and the lung airspace are underrepresented in the individual proteomes for lung tissue or BALF.

To better understand how the proteins in BALF and lung tissue may influence the biological activity in each compartment we created a subset of ITB data that merged only those proteins that were unique to either BALF of lung tissue. The so-called, UtC enabled identification of protein hubs in BALF or lung tissue that could be most significantly influenced by signals in the other compartment. Through this approach predictive assessment of pathways that link biological processes in lung tissue and the airspace compartment is possible. In our analysis we showed that a unique lung tissue protein network predicted to support signaling by interleukins, was refined to one associated with SMAC binding to IAPs, which releases caspases that mediate apoptotic cell death (Du *et al.* 2000). In parallel, a unique BALF protein network predicted to support axonal guidance, was refined to one associated with EGFR down regulation on the basis of the influence of lung tissue proteins. Furthermore, in the UtC, developmental biology emerged as the primary response pattern, a result that was not predicted from lung tissue or BALF proteome datasets individually. Overall integrating lung tissue and BALF proteomes strengthens independent connectedness and yields new biological insight for the whole lung in response to allergen exposure.

The interpretation of our study is limited by a number of factors. Though we carefully controlled our collection methods to reduce variability and decrease damaging effects of BALF collection on tissue cells, but we cannot discount that BALF samples included a small fraction of proteins from damaged cells. Our samples were only collected at a single time point (48 hours after final allergen challenge), a strategic choice, as this represents the time when lung dysfunction is greatest, including airway hyperresponsiveness. Thus, our work provides only a snapshot of the dynamics of proteome response to a specific allergen challenge, and future studies looking at temporal patterns in response to HDM and other allergens are needed for a

more robust resource to delineate the molecular responses that contribute to lung inflammation and dysfunction in allergen-challenge mouse models.

Conclusion

We characterized the proteome of lung tissue and BALF from HDM-challenged mice that mimic allergic asthma pathophysiology. Using matched samples from individual animals, our work reveals that lung tissue and BALF proteomes are diverse, and that integrating both datasets reveals additional novel biological processes and protein interaction hubs. This work provides a resource and approach for identifying new proteins and pathways, and a basis to interrogate interactions between sample compartments to identify mechanisms for airways pathophysiology and, perhaps, new targets for developing therapeutic approaches.

Methods

Animal experiments

(a) Murine HDM allergen challenge

All animal experiments were planned and performed following the approved protocols and guidelines of the animal ethics board at the University of Manitoba. Female, BALB/c mice (6-8 weeks, $n = 3$) were intranasally challenged with HDM (25 µg per mouse, in a total volume of 35 µL saline) five times a week for two weeks (Supplemental 1). Our HDM formulation consisted of HDM extract (Greer Labs, Lenoir, NC) prepared in sterile phosphate buffered saline (PBS, pH 7.4; Life Technologies, Waltham, MA). The HDM extract we used contained 36,000 endotoxin units (EU) per vial (7877 EU/mg of protein or 196.9 EU/dose) containing 4.9 % Der p 1 protein.

(b) Lung function, inflammatory differential cell counts and sample collection

Lung function was performed 48 h after the last HDM challenge. Mice were anesthetized with sodium pentobarbital (90 mg/kg), given intraperitoneally and tracheotomized with a 20-gauge polyethylene catheter. The polyethylene catheter was connected to a flexiVent small animal ventilator (Scireq, Montréal, Canada) and mice were mechanically ventilated with a tidal volume of 10 mL/kg body weight, 150 times/min. Forced oscillation technique and positive end expiratory pressure of 3 cm·H_2O was used for the entire study. Mice were subjected to an nebulized methacholine (MCh) challenge (0 to 50 mg/mL) to assess concentration dependent response of the respiratory mechanics. Measures of newtonian resistance (Rn), peripheral tissue damping (G) and tissue elastance (H), and total resistance (R) were collected. Values for each parameter were calculated as the peak of all 12 perturbation cycles performed after each MCh challenge.

Following the lung function measurement, lungs were lavaged with 1.0 mL of saline two times, for a total of 2 mL containing 0.1 % ethylenediaminetetraacetic acid (EDTA; Sigma-Aldrich, St. Louis, MO). BALF was centrifuged to collect the immune cell pellet (1,000 xg, 10 min, 4 ºC) and the supernatant was collected and aliquoted prior to flash freezing in liquid nitrogen and storage at -80 ºC. Immune cell pellet was resuspended in saline and the total immune cell count was estimated using a hemocytometer. For differential counts, cells were stained with a modified Wright-Giemsa stain (HEMA 3 STAT PACK, Fisher Scientific, Waltham, MA). Cell distribution was analyzed by manually identifying and counting eosinophils, neutrophils, macrophages and lymphocytes in six randomly chosen fields of view examined under a light microscope at 200 x magnification. Post-BAL lung tissue from the left lung and half the right lung were excised, portioned (~35 mg/each), wrapped in aluminum foil, placed in a 2.0 mL centrifuge tube prior to flash freezing in liquid nitrogen and storage at -80 ºC until processed.

Assessment of protein extraction proficiency

From each randomly chosen portion of frozen lung tissue (35 mg, wt/wt) our extraction process yielded an average of 1.146 mg of total protein. BALF yielding an average of 600 µg of total protein (100 µg per 250 µL aliquot) per mouse. Qualitative assessment of total protein molecular weight diversity was performed by gradient SDS-PAGE of both BALF and tissue protein lysates followed by coomassie blue staining. Using our extraction protocol, we obtained protein homogenates from both tissue (Supplemental 4A) and BALF (Supplemental 4B) rich in diverse molecular weight proteins. The dark band in BALF shown at ~66.5 kDa approximates the molecular weight of albumin as no high abundance protein depletion methods were employed for either BALF or tissue samples to reduce potential elimination bias.

To determine the quality of both technical and biological replicates we used mouse

BALF samples and performed two in parallel protein Filter Assisted Sample Preparation (FASP) procedures. We first examined our technical variation by assessing as early as possible the variation that might accumulate as we move identical samples through the FASP process. We divided the same sample into equal parts (100 µg total protein each) and processed the samples individually but in parallel through the entire FASP protocol and informatics pipeline. Our second experiment mirrored the first with exception to examining the effect that biological variation has on our FASP workflow. Our results show that technical variation (Supplemental 4C,D) is lower than our biological variation (Supplemental 4E,F). Therefore, we negated the use of technical replicates for our proteomic analysis.

Preparation of lung tissue

A randomly selected portioned lung tissue was thawed and weighed. The tissue surface was marred to increase surface area using scissors. To wash the lung tissue of residual blood contamination, each tissue sample was placed in a 15 mL centrifuge tube containing inhibitors dissolved in 15 mL PBS (-$CaCl_2$, -$MgCl_2$, pH 7.4; Invitrogen) and placed on an end-over-end mixer at 4 ºC for 30 min. Inhibitors including Phenylmethylsulfonyl Fluoride (PMSF, 100 mM stock), Phosphatase Inhibitor Cocktail 2 (Sigma-Aldrich, St. Louis, MO), and Protease Inhibitor (Sigma-Aldrich) each at 1:100 dilution. Tissues were then removed and placed into siliconized 2.0 mL centrifuge tubes (Thomas Scientific, Swedesboro, NJ) along with 100 µL of lysis buffer. Lysis buffer composition: 150 mM NaCl (Fisher Scientific), 50 mM Tris-HCL (pH 7.5; Fisher Scientific, Waltham, MA), 5 % glycerol (Sigma-Aldrich), 1 % sodium deoxycholate (Sigma-Aldrich), 1 % benzonase (25 U/µL; Merck, Kenilworth, NJ), 1 % sodium dodecyl sulfate (SDS; Fisher Scientific), Protease Inhibitor Cocktail 2 (1:100 dilution, Sigma-Aldrich), 1 mM PMSF (Sigma-Aldrich), Phosphatase Inhibitor Cocktail 2 (1:100 dilution, Sigma-Aldrich), 2 mM $MgCl_2$ (Fisher Scientific) built up with molecular grade water (Invitrogen). Tissue samples were

kept on ice during homogenization. Additional 100 µL of homogenization buffer was added sequentially until an optimal ratio was obtained (400 µL lysis buffer / 0.0335g lung tissue). Foaming was kept to a minimum by centrifugation throughout the homogenization process (10,000 xg, 30 s). Tissue homogenate was centrifuged (21,000 xg, 10 min, 4 ºC, no break) and the supernatant was transferred to a separated siliconized centrifuge tube before incubating at room temperature for 30 min to permit benzonase activity before storing on ice. All chemicals used for tissue preparation were of molecular/electrophoresis grade. Tissues samples and homogenates were transferred and manipulated on wet ice (~4 ºC) whenever possible.

Protein sample quality control

BALF, samples were thawed on ice and spun (21,000 xg, 5 min, 4 ºC, no break) and the pellets stored at -80 ºC. The resulting supernatant was used for protein quantification using a micro bicinchoninic acid (µBCA) protein assay (Pierce, Waltham, MA) as per the manufacturer's instructions. Lung tissue homogenate was quantified using a DC-Lowery Kit (Bio-Rad DC Assay; Hercules, CA) as per the manufacturer's instructions. Both colourmetric kits were read using a UV-Vis spectrophotometer (BioTek; Winooski, VT) using quadrupole technical replicates. Homogenates were subsequently stored at -80 ºC.

Identification of protein complexity by molecular weight for both BALF and tissue was discerned using SDS-PAGE and coomassie total protein stain. Briefly, (20 µg total protein) was prepared in LDS-Loading buffer (Invitrogen) combined with 100 mM DTT (final conc.) and boiled for 5 min. After sample was cooled, proteins were loaded into a pre-cast Bis-Tris gradient gel (4-12 %, 1.0 mm, 10 well; NuPAGE; Invitrogen). With 20 µg per tissue sample (7 µg for BALF) and 5 µL of protein standard (Page Ruler Plus, ThermoFisher Scientific, Cat no: 26619), samples were run at 75 V for 15 min before final separation at 150V until dye was run off the gel. Gels were washed to remove salt (MiliQ H_2O, 50 rpm, 3 x 20 min wash) prior to coomassie

stain overnight (GelCode, Invitrogen) and destained for 1h (MiliQ H_2O, 50 rpm, 3x20 min wash) to remove background before white field imaging (ChemiDock, Bio-Rad), (Supplemental Figure S4A,B).

Filter assisted sample preparation (FASP) of lung tissue

We modified a previously used FASP protocol for use in lung tissue (Wisniewski *et al.* 2009). Tissue homogenate (300 µg total protein) was supplemented with DTT (final concentration 100 mM) before boiling for 5 min, cooled and centrifuged (21,000 xg, 10 min, 4 ºC, no break) before transferring the supernatant to a siliconized centrifuge tube. Molecular weight cut off filters at 30 kDa (Amicron Ultra 0.5; Milipore; Burlington, MA) were tested for efficiency by adding 450 µL of 8 M urea (in 100 mM Tris in LCMS grade water) to each column and centrifuging them through (10,000 xg, 10 min, room temperature). Tissue homogenates were built with urea buffer (8 M urea, in LCMS grade 100 mM Tris) to 800 µL (to dilute SDS and deoxycholate in the sample) before loading onto tested 30 kDa molecular weight cut off columns (Milipore). Samples were repeatedly spun through the column (10,000 xg, 10 min, room temperature) until samples were fully loaded onto the column. Once samples were bound, columns were washed twice with urea buffer (450 µL) to wash out excess DTT (10,000 xg, 10 min, room temperature). Samples were alkylated by adding 400 µL iodoacidimide (IAA; 50 mM in urea buffer; Sigma-Aldrich) to the columns (45 mins, protected from light, room temperature). To halt the cysteine residue modifications by the IAA reaction, 20 mM DTT was added before centrifugation (13,000 xg, 10 min, room temperature). Columns were then washed twice with 450 µL of urea buffer (13,000 xg, 10 min, room temperature) before one last hard spin to reduce the volume to a minimum within the column (14,000 xg, 15 min, room temperature). A fresh collection tube filled with 150 µL (98 % ACN, 2 % TFA) was then prepared used prior to trypsinization. Using a protein:trypsin ratio of 50:1 (Trypsin Gold; Promega; Madison, WI), 6 µg

of fresh trypsin (trypsin was dissolved in 1x digestion buffer; 50 mM Tris, 2 mM $CaCl_2$, LCMS water) and filled to the top of the filter (~350 µL). Samples were sealed with parafilm and placed under shaking conditions at 37 ºC for 16 h before the reaction was halted by the addition of TFA (1 % final concentration) and placing the samples at 4 ºC. A saline solution (500 mM NaCl, final concentration) was added to each column prior to a test centrifugation (5,000 xg, 10 min, room temperature) to ensure filter integrity. Columns were then washed by adding 400 µL 50 % methanol (LCMS grade water), incubating for 5 min at room temperature prior to centrifugation (10,000 xg, 10 min, room temperature). A final wash was completed by adding 300 µL 15 % acetonitrile (LCMS grade water), incubating for 5 min at room temperature prior to centrifugation (10,000 xg, 10 min, room temperature). Flow through was transferred to a clean siliconized tube and a hard spin (13,000 xg, 10 min, room temperature) was then completed and merged with the remainder of the centrifuge tube. Samples were then frozen with the lids open at -80 ºC until perfectly frozen to assist in sample drying by speed vac until dry (~4 h). Dried samples were sealed and frozen at -80 ºC until ready for reconstitution.

Filter assisted sample preparation of BALF

Mouse BALF samples were processed in a similar manner to that of mouse lung tissue with some minor changes to the protocol. Thawed BALF supernatant samples were centrifuged (10,000 xg, 5 min, 4 ºC), 100 µg of total protein was denatured in 8M urea buffer with DTT (100 mM final concentration) to a volume of 450 µL under mixing conditions for 1 h (room temperature) before loading onto 30 kDa molecular weight columns (Milipore) and proceeded to protein digestion steps mentioned previously.

Peptide desalting by reverse phase 1D-HPLC

BALF or tissue lysate samples were warmed, inspected for condensation (if present put

on speed vac) and re-suspended in 800 µL TFA (0.5 %, LCMS grade water), vortexed for 15 min until peptides dissolved, and centrifuged (21,000 xg, 10 min, 4 ºC) to check for undissolved peptides before the supernatant was injected. BALF samples were loaded onto a C18 column (Luna 10 µM C18(2), 100 Å, 50 x 4.6 mm; Phenomenex, Torrance, CA) while tissue lysate samples were loaded onto a separate C18 column (Phenomenex). Column efficiency and elution conditions were tested using a specialized 6 peptide solution prior to starting the samples (Krokhin and Spicer, 2009). Samples were collected at a flow rate of 500 µL/min with an additional 30 s before and after the eluted peptide spectra was detected. Manual loading was used with no gradient. Agilent 1110 HPLC System using ChemStation for Control and Data Analysis (Santa Clara, CA) was used to analyze the chromatograms from Reverse Phase – HPLC (high pressure liquid chromatography). Samples were frozen at -80 ºC with the lids open. All chemicals used were mass spec grade.

Reconstitution of sample & LC-MS/MS run

Thawed samples were desiccated by speed vac (as previously mentioned) and reconstituted in 50 µL formic acid (0.1 %, LCMS grade water). Samples were vortexed for 15 min to encourage dissolving of the peptides. Peptide concentration was determined by UV spectrophotometry (Nanodrop Spectrophotometer 2000, Thermofisher) at 280 nm. Spectrophotometer was checked for contamination between samples to ensure accurate measurements ± 10 nm. Peptides (2 µg) were diluted in formic acid (0.1 %, LCMS grade water) and injected into the LC-MS/MS analysis at a flow rate of (500 nL/min). Samples were injected into an online LC-MS/MS workflow using a 3 h gradient run resulting in a 180 min run on a Sciex TripleTOF 5600 instrument (Sciex; Framingham, MA). Raw spectra files were converted into Mascot Generic File format (MGF) for protein identification using the tools bundled by the manufacturer. All chemicals used were mass spec grade.

Bioinformatic & statistical analysis

The MGF files were processed by X!Tandem (Craig and Beavis, 2004) against single-missed-cleavage tryptic peptides from the *Mus musclus* Uniprot database (16704 proteins). The following X!Tandem search parameters were used: 20 ppm and 50 ppm mass tolerance for parent and fragment ions, respectively; constant modification of Cys with iodoacetamide; default set post-translational modifications: oxidation of Met, Trp; N-terminal cyclization at Qln, Cys; N-terminal acetylation, phosphorylation (Ser, Thr, Tyr), deamidation (Asn and Gln); an expectation value cut-off of $\log_e < -1$ for both proteins and peptides.

Each MS run in yielded a list of protein expression values in a $\log_2$ scale, quantified based on their member peptide MS2 fragment intensity sums. The simple rule of a quantified protein needing at least two non-redundant (unique) peptides with identification scores $\log_e < -1.5$ each followed from our prior approaches. To correct for total protein loading differences when comparing between BALF and tissue, z-scores were calculated across the datasets. The Integrated-Tissue-BALF dataset is described as the combined proteomic profile of both the lung tissue and BALF datasets.

Partial least squares discriminant analysis (PLSDA) and its accompanied variable importance in projection (VIP) scores were calculated using the mixOmics package (v.6.3) in R using both z-scores and delineating 5 components for the discriminant analysis (Rohart *et al.* 2017).

To separate proteins that are found only in either HDM or Naïve mice from those common to both, we plotted the z-scores from each dataset and assigned a z-score of -4 for any absent value. Statistical Analysis of Microarrays (SAM) was conducted using the SAMR (v2.0) package in R using only proteins common to both HDM & Naive groups across tissue and BALF and Integrated-Tissue-BALF datasets (Tusher *et al.* 2001). Delta values were selected based

upon an FDR ≤ 10 % after 1,000 permutations. No $\log_2$ correction or median centering was performed.

Univariate analysis was conducted using the LIMMA package (v3.3) in R (Ritchie *et al.* 2015). Pathway Analysis was conducted using either InnateDB or Reactome with Benjamini-Hochberg multiple comparison adjustment (Breuer *et al.* 2013). For this analysis, Uniprot ID's alongside z-score fold changes (FC) were used from both the unique to treatment datasets and significantly identified proteins from SAM analysis. Missing values were assigned a value of -4 prior to fold change calculation. Fold changes were calculated as the delta z-score (HDM-average Naïve).

Protein-protein interaction networks (first order with minimum connections) were identified using the InnateDB informatics source within NetworkAnalyst and accessed in July 2019 (Breuer *et al.* 2013; Xia *et al.* 2014). The top 3 most interconnected protein hubs which are unique to either tissue or BALF were selected. Pathway analysis of these top hubs was determined through the open-source Reactome database using only Uniprot ID. Network enrichment analysis was limited to hubs in the Unique-to-Combined (UtC) dataset that have ≥ 5 connections (Vallabhajosyula *et al.* 2009). The protein UBC was removed from all protein-protein interaction analysis.

All other visual tools were constructed using DataGraph (DataGraph v4.5, Visual Data Tools, Inc., Chapel Hill, NC, USA, https://www.visualdatatools.com/).

Acknowledgements

Work for this study was supported in part by the CIHR-Canadian Respiratory Research Network (CRRN), Research Manitoba and the DEVOTION Network, and the Biology of Breathing Group, Children's Hospital Research Institute of Manitoba. THM was supported by studentships from Research Manitoba, CRRN, AllerGen NCE Inc. and Asthma Canada. CDP was supported by a CIHR Banting Fellowship and a Research Manitoba Fellowship. AJ was supported by studentships from Research Manitoba and CRRN. AJH is supported through the Canada Research Chairs Program.

Author Contributions

THM performed all proteomics experiments, primary and secondary data analysis and prepared the manuscript draft. CDP provide supervision, guidance and assisted in the completed biostatistics and bioinformatics, design of figures, and editing and writing of the manuscript. AJ and SB were involved with experimental design and completion of animal studies. PE guided sample preparation and performed mass spectrometry. VS contributed to experimental design, primary proteomic data analysis, proteomic data quality control and primary statistics. NM contributed to experimental design and direction, and edited the manuscript. AJH conceived and led design of the study, including scope of biostatistics and bioinformatics, and contributed to writing and editing of the final manuscript.

Conflict of Interest

The authors declare no competing or financial interests.

References

Baarsma HA, Königshoff M, Gosens R (2013) The WNT signaling pathway from ligand secretion to gene transcription: molecular mechanisms and pharmacological targets. *Pharmacol Ther* 138: 66–83

Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock RE, Brinkman FS, Lynn DJ (2013) InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. *Nucleic Acids Res* 41: D1228–33

Burg D, Schofield JPR, Brandsma J, Staykova D, Folisi C, Bansal A, Nicholas B, Xian Y, Rowe A, Corfield J, Wilson S, Ward J, Lutter R, Fleming L, Shaw DE, Bakke PS, Caruso M, Dahlen SE, Fowler SJ, Hashimoto S, Horváth I, Howarth P, Krug N, Montuschi P, Sanak M, Sandström T, Singer F, Sun K, Pandis I, Auffray C, Sousa AR, Adcock IM, Chung KF, Sterk PJ, Djukanović R, Skipp PJ, The U-BSG (2018) Large-Scale Label-Free Quantitative Mapping of the Sputum Proteome. *J Proteome Res* 17: 2072–2091

Calderón MA, Linneberg A, Kleine-Tebbe J, De Blay F, Hernandez Fernandez de Rojas D, Virchow JC, Demoly P (2015) Respiratory allergy caused by house dust mites: What do we really know. *J Allergy Clin Immunol* 136: 38–48

Chan TK, Loh XY, Peh HY, Tan WNF, Tan WSD, Li N, Tay IJJ, Wong WSF, Engelward BP (2016) House dust mite-induced asthma causes oxidative damage and DNA double-strand breaks in the lungs. *J Allergy Clin Immunol* 138: 84–96.e1

Choopong J, Reamtong O, Sookrung N, Seesuay W, Indrawattana N, Sakolvaree Y, Chaicumpa W, Tungtrongchitr A (2016) Proteome, Allergenome, and Novel Allergens of House Dust Mite, Dermatophagoides farinae. *J Proteome Res* 15: 422–430

Colombo M, Moita C, van Niel G, Kowal J, Vigneron J, Benaroch P, Manel N, Moita LF, Théry C, Raposo G (2013) Analysis of ESCRT functions in exosome biogenesis, composition and secretion highlights the heterogeneity of extracellular vesicles. *J Cell Sci* 126: 5553–5565

Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 20: 1466–1467

Du C, Fang M, Li Y, Li L, Wang X (2000) Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition. *Cell* 102: 33–42

Halayko AJ, Amrani Y (2003) Mechanisms of inflammation-mediated airway smooth muscle plasticity and airways remodeling in asthma. *Respir Physiol Neurobiol* 137: 209–222

Jha A, Ryu MH, Oo O, Bews HJ, Carlson JC, Schwartz J, Basu S, Wong CS, Halayko AJ (2018) Prophylactic benefits of systemically delivered simvastatin treatment in a house dust mite challenged murine model of allergic asthma. *Br J Pharmacol* 175: 1004–1016

Karim MA, Samyn DR, Mattie S, Brett CL (2018) Distinct features of multivesicular body-lysosome fusion revealed by a new cell-free content-mixing assay. *Traffic* 19: 138–149

Koopmans T, Eilers R, Menzen M, Halayko A, Gosens R (2017) β-Catenin Directs Nuclear Factor-κB p65 Output via CREB-Binding Protein/p300 in Human Airway Smooth Muscle. *Front Immunol* 8: 1086

Krokhin OV, Spicer V (2009) Peptide retention standards and hydrophobicity indexes in reversed-phase high-performance liquid chromatography of peptides. *Anal Chem* 81: 9522–9530

Kulshreshtha A, Ahmad T, Agrawal A, Ghosh B (2013) Proinflammatory role of epithelial cell-derived exosomes in allergic airway inflammation. *J Allergy Clin Immunol* 131: 1194–203, 1203.e1

O'Neil SE, Sitkauskiene B, Babusyte A, Krisiukeniene A, Stravinskaite-Bieksiene K, Sakalauskas R, Sihlbom C, Ekerljung L, Carlsohn E, Lötvall J (2011) Network analysis of quantitative proteomics on asthmatic bronchi: effects of inhaled glucocorticoid treatment. *Respir Res* 12: 124

Petta I, Bougarne N, Vandewalle J, Dejager L, Vandevyver S, Ballegeer M, Desmet S, Thommis J, De Cauwer L, Lievens S, Libert C, Tavernier J, De Bosscher K (2017) Glucocorticoid

Receptor-mediated transactivation is hampered by Striatin-3, a novel interaction partner of the receptor. *Sci Rep* 7: 8941

Piyadasa H, Altieri A, Basu S, Schwartz J, Halayko AJ, Mookherjee N (2016) Biosignature for airway inflammation in a house dust mite-challenged murine model of allergic asthma. *Biol Open* 5: 112–121

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47

Rohart F, Gautier B, Singh A, Lê Cao KA (2017) mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol* 13: e1005752

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98: 5116–5121

Vallabhajosyula RR, Chakravarti D, Lutfeali S, Ray A, Raval A (2009) Identifying hubs in protein interaction networks. *PLoS One* 4: e5344

Wisniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Methods* 6: 359–362

Wu J, Kobayashi M, Sousa EA, Liu W, Cai J, Goldman SJ, Dorner AJ, Projan SJ, Kavuru MS, Qiu Y, Thomassen MJ (2005) Differential proteomic analysis of bronchoalveolar lavage fluid in asthmatics following segmental antigen challenge. *Mol Cell Proteomics* 4: 1251–1264

Xia J, Benner MJ, Hancock RE (2014) NetworkAnalyst--integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Res* 42: W167–74

Tables

Table 1: Proteins enriched in tissue, but diminished in the BALF proteome. Significantly enriched proteins (10) are highlighted in bold. Significance determined using a linear fit model in the R package LIMMA with Benjamini-Hochberg correction. Abbreviations used: log2 fold change (log_2FC), adjusted p-value (p.adj), Uniprot identification number (Uniprot ID).

Uniprot	Gene	Description	log_2FC	*p-value*	*p.adj*
P20065	**Tmsb4x**	**Thymosin beta-4**	**4.536**	**9.23E-07**	**1.25E-05**
Q62426	**Cstb**	**Cystatin-B**	**4.031**	**9.25E-07**	**1.25E-05**
Q99J08	**Sec14l2**	**SEC14-like protein 2**	**5.029**	**8.29E-06**	**7.46E-05**
P51855	**Gss**	**Glutathione synthetase**	**3.784**	**1.47E-05**	**9.37E-05**
Q2VLH6	**Cd163**	**Scavenger receptor cysteine-rich type 1 protein M130**	**2.772**	**1.73E-05**	**9.37E-05**
P70695	**Fbp2**	**Fructose-1,6-bisphosphatase isozyme 2**	**3.593**	**1.03E-04**	**4.16E-04**
P63254	**Crip1**	**Cysteine-rich protein 1**	**1.419**	**1.08E-04**	**4.16E-04**
Q9JHW2	**Nit2**	**Omega-amidase NIT2**	**3.261**	**2.18E-04**	**7.34E-04**
P15532	**Nme1**	**Nucleoside diphosphate kinase A**	**1.029**	**5.68E-03**	**0.0171**
P26883	**Fkbp1a**	**Peptidyl-prolyl cis-trans isomerase FKBP1A**	**0.659**	**1.81E-02**	**0.0488**
Q61646	Hp	Haptoglobin	0.806	3.50E-02	0.0818
Q9DBG5	Plin3	Perilipin-3	0.695	3.63E-02	0.0818
Q8VBW6	Nae1	NEDD8-activating enzyme E1 regulatory subunit	0.948	4.41E-02	0.0863
O35381	Anp32a	Acidic leucine-rich nuclear phosphoprotein 32 family member A	0.689	4.48E-02	0.0863
P35505	Fah	Fumarylacetoacetase	0.570	1.32E-01	0.2344
P17751	Tpi1	Triosephosphate isomerase	0.351	1.39E-01	0.2344
P10518	Alad	Delta-aminolevulinic acid dehydratase	0.352	1.58E-01	0.2431
P11352	Gpx1	Glutathione peroxidase 1	0.292	1.62E-01	0.2431
P13597	Icam1	Intercellular adhesion molecule 1	0.372	1.72E-01	0.2440
Q8K183	Pdxk	Pyridoxal kinase	0.499	1.92E-01	0.2591
Q9QYR9	Acot2	Acyl-coenzyme A thioesterase 2, mitochondrial	0.459	2.11E-01	0.2591
Q9CZ42	Carkd	ATP-dependent (S)-NAD(P)H-hydrate dehydratase	0.382	2.09E-01	0.2591
P05213	Tuba1b	Tubulin alpha-1B chain	0.743	2.31E-01	0.2594
P15626	Gstm2	Glutathione S-transferase Mu 2	0.421	2.30E-01	0.2594
P51125	Cast	Calpastatin	0.357	3.94E-01	0.4258
Q8BFZ3	Actbl2	Beta-actin-like protein 2	0.161	6.26E-01	0.6504
Q99JI6	Rap1b	Ras-related protein Rap-1b	0.134	7.10E-01	0.7096

Table 2: Proteins enriched in BALF but diminished in the lung tissue proteome. Significantly enriched proteins (141) are highlighted in bold. Significance determined using a linear fit model in the R package LIMMA with Benjamini-Hochberg correction. Abbreviations used: log2 fold change (log_2FC), adjusted p-value (p.adj), Uniprot identification number (Uniprot ID).

Uniprot	Gene	Description	log_2FC	*p-value*	*p.adj*
Q64475	**Hist1h2bb**	**Histone H2B type 1-B**	**-7.500**	**1.00E-09**	**2.00E-07**
P48678	**Lmna**	**Prelamin-A/C**	**-4.004**	**5.72E-08**	**3.82E-06**
P62715	**Ppp2cb**	**Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform**	**-4.003**	**5.46E-08**	**3.82E-06**
P15508	**Sptb**	**Spectrin beta chain, erythrocytic**	**-4.851**	**8.00E-08**	**4.00E-06**
P31725	**S100a9**	**Protein S100-A9**	**-4.486**	**1.36E-07**	**4.25E-06**
P15864	**Hist1h1c**	**Histone H1.2**	**-4.333**	**1.53E-07**	**4.25E-06**
Q3THE2	**Myl12b**	**Myosin regulatory light chain 12B**	**-4.228**	**1.58E-07**	**4.25E-06**
Q8VCW8	**Acsf2**	**Acyl-CoA synthetase family member 2, mitochondrial**	**-4.052**	**1.86E-07**	**4.25E-06**
P46638	**Rab11b**	**Ras-related protein Rab-11B**	**-3.623**	**1.91E-07**	**4.25E-06**
Q8C1B7	**Sept11**	**Septin-11**	**-3.926**	**2.88E-07**	**5.75E-06**
Q8CIB5	**Fermt2**	**Fermitin family homolog 2**	**-3.941**	**3.54E-07**	**6.44E-06**
Q61879	**Myh10**	**Myosin-10**	**-4.212**	**5.80E-07**	**9.19E-06**
Q8BT60	**Cpne3**	**Copine-3**	**-4.166**	**7.19E-07**	**9.19E-06**
Q9WVK4	**Ehd1**	**EH domain-containing protein 1**	**-3.722**	**7.32E-07**	**9.19E-06**
P16546	**Sptan1**	**Spectrin alpha chain, non-erythrocytic 1**	**-3.700**	**6.37E-07**	**9.19E-06**
Q9EPC1	**Parva**	**Alpha-parvin**	**-3.486**	**7.35E-07**	**9.19E-06**
Q99NB1	**Acss1**	**Acetyl-coenzyme A synthetase 2-like, mitochondrial**	**-3.850**	**1.08E-06**	**1.14E-05**
P68368	**Tuba4a**	**Tubulin alpha-4A chain**	**-3.774**	**1.03E-06**	**1.14E-05**
Q8BHZ0	**Fam49a**	**Protein FAM49A**	**-2.793**	**1.02E-06**	**1.14E-05**
Q8K1B8	**Fermt3**	**Fermitin family homolog 3**	**-3.831**	**1.19E-06**	**1.19E-05**
O35215	**Ddt**	**D-dopachrome decarboxylase**	**-4.226**	**1.37E-06**	**1.24E-05**
P35282	**Rab21**	**Ras-related protein Rab-21**	**-3.751**	**1.43E-06**	**1.24E-05**
Q62261	**Sptbn1**	**Spectrin beta chain, non-erythrocytic 1**	**-3.636**	**1.40E-06**	**1.24E-05**
P68369	**Tuba1a**	**Tubulin alpha-1A chain**	**-5.702**	**1.62E-06**	**1.29E-05**
Q91ZX7	**Lrp1**	**Prolow-density lipoprotein receptor-related protein 1**	**-3.775**	**1.67E-06**	**1.29E-05**
Q91YR1	**Twf1**	**Twinfilin-1**	**-3.641**	**1.68E-06**	**1.29E-05**
P54071	**Idh2**	**Isocitrate dehydrogenase [NADP], mitochondrial**	**-3.655**	**1.84E-06**	**1.36E-05**
Q8CGB6	**Tenc1**	**Tensin-like C1 domain-containing phosphatase**	**-3.972**	**1.98E-06**	**1.36E-05**
P12382	**Pfkl**	**ATP-dependent 6-phosphofructokinase, liver type**	**-3.954**	**1.95E-06**	**1.36E-05**
Q61425	**Hadh**	**Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial**	**-3.781**	**2.09E-06**	**1.39E-05**
Q8BFW7	**Lpp**	**Lipoma-preferred partner homolog**	**-3.680**	**2.37E-06**	**1.48E-05**
Q9CZ44	**Nsfl1c**	**NSFL1 cofactor p47**	**-3.515**	**2.37E-06**	**1.48E-05**
P10630	**Eif4a2**	**Eukaryotic initiation factor 4A-II**	**-4.160**	**2.48E-06**	**1.50E-05**
Q8VI36	**Pxn**	**Paxillin**	**-3.701**	**3.11E-06**	**1.83E-05**
P08752	**Gnai2**	**Guanine nucleotide-binding protein G(i) subunit alpha-2**	**-3.762**	**3.23E-06**	**1.85E-05**
Q7TSV4	**Pgm2**	**Phosphoglucomutase-2**	**-3.709**	**3.56E-06**	**1.92E-05**
P48722	**Hspa4l**	**Heat shock 70 kDa protein 4L**	**-3.701**	**3.55E-06**	**1.92E-05**
Q62465	**Vat1**	**Synaptic vesicle membrane protein VAT-1 homolog**	**-3.820**	**4.02E-06**	**2.01E-05**
O35955	**Psmb10**	**Proteasome subunit beta type-10**	**-3.786**	**3.84E-06**	**2.01E-05**
Q9WUB3	**Pygm**	**Glycogen phosphorylase, muscle form**	**-3.655**	**3.99E-06**	**2.01E-05**
P70372	**Elavl1**	**ELAV-like protein 1**	**-3.904**	**4.33E-06**	**2.04E-05**
Q9CS42	**Prps2**	**Ribose-phosphate pyrophosphokinase 2**	**-3.355**	**4.31E-06**	**2.04E-05**
Q62348	**Tsn**	**Translin**	**-2.753**	**4.38E-06**	**2.04E-05**
Q9JJ28	**Flii**	**Protein flightless-1 homolog**	**-3.753**	**4.49E-06**	**2.04E-05**
Q8VE70	**Pdcd10**	**Programmed cell death protein 10**	**-3.831**	**4.84E-06**	**2.15E-05**
Q8QZT1	**Acat1**	**Acetyl-CoA acetyltransferase, mitochondrial**	**-3.760**	**5.84E-06**	**2.54E-05**
Q64331	**Myo6**	**Unconventional myosin-VI**	**-4.237**	**6.00E-06**	**2.54E-05**
Q8VIJ6	**Sfpq**	**Splicing factor, proline- and glutamine-rich**	**-3.654**	**6.10E-06**	**2.54E-05**
Q8VE97	**Srsf4**	**Serine/arginine-rich splicing factor 4**	**-3.882**	**6.25E-06**	**2.55E-05**
Q9WTI7	**Myo1c**	**Unconventional myosin-Ic**	**-3.517**	**6.72E-06**	**2.69E-05**
Q8C3J5	**Dock2**	**Dedicator of cytokinesis protein 2**	**-4.074**	**8.04E-06**	**3.03E-05**
Q3TCJ1	**Fam175b**	**BRISC complex subunit Abro1**	**-3.722**	**7.95E-06**	**3.03E-05**
Q99020	**Hnrnpab**	**Heterogeneous nuclear ribonucleoprotein A/B**	**-3.650**	**7.88E-06**	**3.03E-05**
P32067	**Ssb**	**Lupus La protein homolog**	**-3.664**	**8.79E-06**	**3.15E-05**
Q9QXZ0	**Macf1**	**Microtubule-actin cross-linking factor 1**	**-3.656**	**8.81E-06**	**3.15E-05**
O55234	**Psmb5**	**Proteasome subunit beta type-5**	**-2.439**	**8.52E-06**	**3.15E-05**
Q9Z0U1	**Tjp2**	**Tight junction protein ZO-2**	**-3.737**	**9.36E-06**	**3.23E-05**
Q9ET54	**Palld**	**Palladin**	**-3.689**	**9.44E-06**	**3.23E-05**
Q8BH95	**Echs1**	**Enoyl-CoA hydratase, mitochondrial**	**-3.642**	**9.69E-06**	**3.23E-05**
Q8BK64	**Ahsa1**	**Activator of 90 kDa heat shock protein ATPase homolog 1**	**-3.284**	**9.57E-06**	**3.23E-05**
Q8BFY9	**Tnpo1**	**Transportin-1**	**-3.695**	**1.02E-05**	**3.34E-05**
Q6P1F6	**Ppp2r2a**	**Serine/threonine-protein phosphatase 2A 55 kDa regulatory**	**-3.396**	**1.05E-05**	**3.40E-05**

		subunit B alpha isoform			
Q9JHU4	Dync1h1	Cytoplasmic dynein 1 heavy chain 1	-3.791	1.18E-05	3.74E-05
P84084	Arf5	ADP-ribosylation factor 5	-3.657	1.29E-05	4.03E-05
P26231	Ctnna1	Catenin alpha-1	-3.328	1.35E-05	4.15E-05
P33267	Cyp2f2	Cytochrome P450 2F2	-4.277	1.40E-05	4.23E-05
P35278	Rab5c	Ras-related protein Rab-5C	-3.161	1.87E-05	5.57E-05
Q01730	Rsu1	Ras suppressor protein 1	-3.466	2.06E-05	6.05E-05
Q9DBR7	Ppp1r12a	Protein phosphatase 1 regulatory subunit 12A	-3.784	2.32E-05	6.71E-05
Q9WTQ5	Akap12	A-kinase anchor protein 12	-3.878	2.51E-05	7.14E-05
P28660	Nckap1	Nck-associated protein 1	-3.545	2.53E-05	7.14E-05
O35737	Hnrnph1	Heterogeneous nuclear ribonucleoprotein H	-4.026	2.66E-05	7.38E-05
Q9D8T2	Gsdmdc1	Gasdermin-D	-2.587	3.34E-05	9.16E-05
Q62219	Tgfb1i1	Transforming growth factor beta-1-induced transcript 1 protein	-3.833	3.45E-05	9.29E-05
Q9Z0P5	Twf2	Twinfilin-2	-3.217	3.49E-05	9.29E-05
Q64514	Tpp2	Tripeptidyl-peptidase 2	-3.674	3.99E-05	1.05E-04
Q9D358	Acp1	Low molecular weight phosphotyrosine protein phosphatase	-3.384	4.13E-05	1.07E-04
Q99L45	Eif2s2	Eukaryotic translation initiation factor 2 subunit 2	-3.200	4.34E-05	1.11E-04
O88544	Cops4	COP9 signalosome complex subunit 4	-3.369	4.82E-05	1.22E-04
Q60994	Adipoq	Adiponectin	-2.786	5.36E-05	1.34E-04
Q6URW6	Myh14	Myosin-14	-3.386	5.60E-05	1.38E-04
Q9R1T2	Sae1	SUMO-activating enzyme subunit 1	-3.455	5.73E-05	1.40E-04
Q9QXS1	Plec	Plectin	-3.131	5.92E-05	1.43E-04
Q8BH61	F13a1	Coagulation factor XIII A chain	-2.444	6.01E-05	1.43E-04
Q8BYA0	Tbcd	Tubulin-specific chaperone D	-2.024	8.51E-05	2.00E-04
Q9JLV1	Bag3	BAG family molecular chaperone regulator 3	-3.280	8.97E-05	2.09E-04
Q61838	A2m	Alpha-2-macroglobulin	-1.743	1.08E-04	2.49E-04
P35123	Usp4	Ubiquitin carboxyl-terminal hydrolase 4	-2.167	1.10E-04	2.49E-04
P26043	Rdx	Radixin	-1.453	1.11E-04	2.49E-04
Q9CQD1	Rab5a	Ras-related protein Rab-5A	-3.723	1.13E-04	2.51E-04
Q3UDE2	Ttll12	Tubulin--tyrosine ligase-like protein 12	-2.927	1.19E-04	2.61E-04
Q9JM14	Nt5c	5'(3')-deoxyribonucleotidase, cytosolic type	-1.335	1.21E-04	2.62E-04
Q61151	Ppp2r5e	Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit epsilon isoform	-2.475	1.29E-04	2.78E-04
P28665	Mug1	Murinoglobulin-1	-1.768	1.74E-04	3.70E-04
P70335	Rock1	Rho-associated protein kinase 1	-3.247	1.82E-04	3.83E-04
Q99JI1	Mustn1	Musculoskeletal embryonic nuclear protein 1	-2.387	1.98E-04	4.14E-04
Q00623	Apoa1	Apolipoprotein A-I	-1.449	2.16E-04	4.45E-04
E9PV24	Fga	Fibrinogen alpha chain	-1.722	2.48E-04	5.07E-04
P07309	Ttr	Transthyretin	-1.336	3.49E-04	7.05E-04
Q8BGD9	Eif4b	Eukaryotic translation initiation factor 4B	-3.433	3.56E-04	7.11E-04
Q80TM9	Nisch	Nischarin	-2.752	4.07E-04	8.05E-04
Q8VCM7	Fgg	Fibrinogen gamma chain	-1.256	4.18E-04	8.19E-04
Q99JW4	Lims1	LIM and senescent cell antigen-like-containing domain protein 1	-1.242	6.39E-04	0.0012
O88456	Capns1	Calpain small subunit 1	-1.277	6.92E-04	0.0013
Q6P069	Sri	Sorcin	-1.454	7.38E-04	0.0014
P08249	Mdh2	Malate dehydrogenase, mitochondrial	-1.168	7.87E-04	0.0015
P00329	Adh1	Alcohol dehydrogenase 1	-1.144	1.43E-03	0.0027
P32261	Serpinc1	Antithrombin-III	-1.208	1.48E-03	0.0027
A2ASS6	Ttn	Titin	-1.049	1.52E-03	0.0028
Q07076	Anxa7	Annexin A7	-1.162	2.02E-03	0.0037
P26039	Tln1	Talin-1	-0.981	2.06E-03	0.0037
Q3UH68	Limch1	LIM and calponin homology domains-containing protein 1	-1.402	2.20E-03	0.0039
P63087	Ppp1cc	Serine/threonine-protein phosphatase PP1-gamma catalytic subunit	-0.984	2.43E-03	0.0043
Q8K0E8	Fgb	Fibrinogen beta chain	-1.311	2.48E-03	0.0043
Q9DC07	Nebl	LIM zinc-binding domain-containing Nebulette	-1.119	2.74E-03	0.0048
P52430	Pon1	Serum paraoxonase/arylesterase 1	-0.878	3.32E-03	0.0057
P97333	Nrp1	Neuropilin-1	-1.362	3.67E-03	0.0063
P48036	Anxa5	Annexin A5	-0.806	4.08E-03	0.0069
Q08879	Fbln1	Fibulin-1	-1.766	4.27E-03	0.0072
Q9WV60	Gsk3b	Glycogen synthase kinase-3 beta	-2.109	5.17E-03	0.0086
Q9EPK7	Xpo7	Exportin-7	-1.200	5.30E-03	0.0088
P58871	Tnks1bp1	182 kDa tankyrase-1-binding protein	-1.284	5.44E-03	0.0089
P47738	Aldh2	Aldehyde dehydrogenase, mitochondrial	-1.011	5.78E-03	0.0094
Q921I1	Tf	Serotransferrin	-0.806	6.48E-03	0.0105
Q91VC7	Ppp1r14a	Protein phosphatase 1 regulatory subunit 14A	-1.136	7.74E-03	0.0124
P24270	Cat	Catalase	-0.675	7.80E-03	0.0124
P97447	Fhl1	Four and a half LIM domains protein 1	-0.678	9.59E-03	0.0151
P40936	Inmt	Indolethylamine N-methyltransferase	-0.793	1.04E-02	0.0163
P68033	Actc1	Actin, alpha cardiac muscle 1	-2.423	1.09E-02	0.0168
P62204	Calm1	Calmodulin	-0.785	1.08E-02	0.0168
Q91X72	Hpx	Hemopexin	-0.684	1.22E-02	0.0187
P51174	Acadl	Long-chain specific acyl-CoA dehydrogenase, mitochondrial	-0.696	1.24E-02	0.0189
Q6IRU2	Tpm4	Tropomyosin alpha-4 chain	-0.737	1.30E-02	0.0195
P60766	Cdc42	Cell division control protein 42 homolog	-0.715	1.34E-02	0.0199
Q9CZY3	Ube2v1	Ubiquitin-conjugating enzyme E2 variant 1	-0.693	1.33E-02	0.0199

P97315	**Csrp1**	**Cysteine and glycine-rich protein 1**	**-0.731**	**1.67E-02**	**0.0245**
Q6A028	**Swap70**	**Switch-associated protein 70**	**-0.790**	**1.93E-02**	**0.0282**
P63260	**Actg1**	**Actin, cytoplasmic 2**	**-3.539**	**2.26E-02**	**0.0328**
P08226	**Apoe**	**Apolipoprotein E**	**-0.528**	**2.82E-02**	**0.0406**
Q08857	**Cd36**	**Platelet glycoprotein 4**	**-0.732**	**3.06E-02**	**0.0437**
Q01339	**Apoh**	**Beta-2-glycoprotein 1**	**-0.654**	**3.08E-02**	**0.0437**
Q9WVH9	Fbln5	Fibulin-5	-0.667	3.76E-02	0.0530
Q64471	Gstt1	Glutathione S-transferase theta-1	-0.683	3.85E-02	0.0538
O55222	Ilk	Integrin-linked protein kinase	-0.773	3.93E-02	0.0545
Q9JJU8	Sh3bgrl	SH3 domain-binding glutamic acid-rich-like protein	-0.610	4.39E-02	0.0606
Q61147	Cp	Ceruloplasmin	-0.496	4.79E-02	0.0657
Q9CPU0	Glo1	Lactoylglutathione lyase	-0.487	5.45E-02	0.0741
Q06890	Clu	Clusterin	-0.452	5.73E-02	0.0774
O55135	Eif6	Eukaryotic translation initiation factor 6	-0.614	6.25E-02	0.0838
P14602	Hspb1	Heat shock protein beta-1	-0.721	6.92E-02	0.0923
P51885	Lum	Lumican	-0.475	7.10E-02	0.0941
Q9JLJ2	Aldh9a1	4-trimethylaminobutyraldehyde dehydrogenase	-0.719	7.18E-02	0.0945
P35235	Ptpn11	Tyrosine-protein phosphatase non-receptor type 11	0.374	8.48E-02	0.1109
Q6PDN3	Mylk	Myosin light chain kinase, smooth muscle	-0.702	9.13E-02	0.1186
P34884	Mif	Macrophage migration inhibitory factor	-0.414	9.40E-02	0.1212
Q61171	Prdx2	Peroxiredoxin-2	-0.532	9.77E-02	0.1253
P00920	Ca2	Carbonic anhydrase 2	-0.467	9.90E-02	0.1261
O35639	Anxa3	Annexin A3	-0.382	9.99E-02	0.1264
P68510	Ywhah	14-3-3 protein eta	-0.351	1.03E-01	0.1292
P01942	Hba	Hemoglobin subunit alpha	-0.713	1.04E-01	0.1304
Q9D0J8	Ptms	Parathymosin	-0.406	1.06E-01	0.1317
P62806	Hist1h4a	Histone H4	-0.582	1.16E-01	0.1433
P70202	Lxn	Latexin	-0.458	1.17E-01	0.1433
Q8BPB5	Efemp1	EGF-containing fibulin-like extracellular matrix protein 1	-0.495	1.22E-01	0.1492
Q64727	Vcl	Vinculin	-0.376	1.24E-01	0.1509
Q9DC11	Plxdc2	Plexin domain-containing protein 2	0.394	1.54E-01	0.1858
Q62000	Ogn	Mimecan	-0.660	1.59E-01	0.1895
P24549	Aldh1a1	Retinal dehydrogenase 1	-0.431	1.59E-01	0.1895
P27546	Map4	Microtubule-associated protein 4	-0.355	1.67E-01	0.1976
P60710	Actb	Actin, cytoplasmic 1	-0.836	1.76E-01	0.2075
P23953	Ces1c	Carboxylesterase 1C	-0.405	1.83E-01	0.2136
P02089	Hbb-b2	Hemoglobin subunit beta-2	-0.580	1.99E-01	0.2300
Q8CG76	Akr7a2	Aflatoxin B1 aldehyde reductase member 2	-0.392	1.98E-01	0.2300
Q99KJ8	Dctn2	Dynactin subunit 2	-0.367	2.10E-01	0.2410
O09164	Sod3	Extracellular superoxide dismutase [Cu-Zn]	-0.275	2.12E-01	0.2419
P17563	Selenbp1	Selenium-binding protein 1	-0.260	2.38E-01	0.2708
P14152	Mdh1	Malate dehydrogenase, cytoplasmic	-0.270	2.41E-01	0.2716
P63330	Ppp2ca	Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	-0.266	2.42E-01	0.2716
O08677	Kng1	Kininogen-1	-0.432	2.46E-01	0.2735
P13020	Gsn	Gelsolin	-0.283	2.46E-01	0.2735
Q99K51	Pls3	Plastin-3	-0.228	2.66E-01	0.2941
P08228	Sod1	Superoxide dismutase [Cu-Zn]	-0.239	2.83E-01	0.3109
Q9WVA4	Tagln2	Transgelin-2	-0.226	2.95E-01	0.3221
P02088	Hbb-b1	Hemoglobin subunit beta-1	-0.477	3.16E-01	0.3434
Q60854	Serpinb6	Serpin B6	-0.214	3.24E-01	0.3502
P50396	Gdi1	Rab GDP dissociation inhibitor alpha	-0.202	3.26E-01	0.3506
P22599	Serpina1b	Alpha-1-antitrypsin 1-2	-0.266	3.93E-01	0.4205
Q99K30	Eps8l2	Epidermal growth factor receptor kinase substrate 8-like protein 2	-0.401	4.06E-01	0.4317
P30416	Fkbp4	Peptidyl-prolyl cis-trans isomerase FKBP4	-0.421	4.09E-01	0.4325
O08709	Prdx6	Peroxiredoxin-6	-0.163	4.32E-01	0.4551
Q06770	Serpina6	Corticosteroid-binding globulin	-0.377	4.41E-01	0.4619
O08553	Dpysl2	Dihydropyrimidinase-related protein 2	-0.144	5.03E-01	0.5240
Q9DBF1	Aldh7a1	Alpha-aminoadipic semialdehyde dehydrogenase	-0.117	5.63E-01	0.5834
P58389	Ppp2r4	Serine/threonine-protein phosphatase 2A activator	-0.138	5.67E-01	0.5849
P19221	F2	Prothrombin	-0.209	6.14E-01	0.6294
P49443	Ppm1a	Protein phosphatase 1A	-0.237	6.26E-01	0.6388
P63101	Ywhaz	14-3-3 protein zeta/delta	-0.099	6.68E-01	0.6785
Q8BH64	Ehd2	EH domain-containing protein 2	-0.144	6.72E-01	0.6785
P06728	Apoa4	Apolipoprotein A-IV	-0.085	7.44E-01	0.7477
P20918	Plg	Plasminogen	-0.053	8.79E-01	0.8790

723

Figure Legends

Figure 1: Characterization of the distinct proteomic signatures of the lung tissue and BALF. **A)** Venn Diagram showing distribution of unique protein ID's for the lung tissue and BALF across both naïve and HDM exposed mice. **B)** PLSDA of the proteomes from the tissue, BALF and Integrated-Tissue-BALF datasets as influenced by HDM treatment. **C)** Variance in Projection (VIP) score for the PLSDA shown in B identifies the top proteins that discriminate the proteomes of the tissue, BALF and Integrated-Tissue-BALF proteomes. Z-scores are shown on the right-hand side to show individual contributions from each proteomic dataset. Abbreviations used: Partial Least Squares Discriminant Analysis (PLSDA), House Dust Mite (HDM), BALF (B), tissue (T), Integrated-Tissue-BALF (I).

Figure 2: Differential expression analysis identifies both unique and significantly enriched proteins from tissue, BALF and Integrated-Tissue-BALF proteomes during HDM challenge. **A, B, C)** Insets: z-score plots from HDM treated and naïve mice. Proteins exclusive to either HDM or naïve samples are represented by assigning a z-score of -4 to the missing values. Proteins found in both HDM and naïve mice were used for SAM analysis and represented as volcano plots. Significance was taken at $p \leq 0.05$ after False Discovery Rate (FDR) correction. Proteins unique to the tissue or BALF proteomes are coloured red and blue respectively (A,B). Proteins unique to lung tissue or BALF proteomes when examined in isolation are colored in green (C). **D)** Proteins that are found in both lung tissue and BALF but are differentially enriched. **E-F)** Pathway analysis (top 10) using significantly enriched lung tissue proteins (diminished in BALF; E) and proteins that are enriched in BALF (diminished in tissue; F). Pathway analysis was conducted using InnateDB with an adjusted p-value by Benjamini-Hochberg multiple comparison. Vertical lines (E,F) indicate statistical significance (p.adj $p = 0.05$).

Figure 3: Individually, lung tissue or BALF proteomes do not recapitulate the biology of the Integrated-Tissue-BALF proteome. **A, B)** Significantly enriched pathways activated by HDM exposure for the lung tissue (A), and BALF proteomes (B). **C)** Pathways that are enriched in the tissue and the Integrated-Tissue-BALF proteome (p.adj ≤ 0.05) but not the BALF proteome (p.adj > 0.05). **D)** Pathways that are enriched in the BALF and the Integrated-Tissue-BALF proteome (p.adj ≤ 0.05) but not the lung tissue (p.adj > 0.05). **E)** Pathways that are enriched in the Integrated-Tissue-BALF proteome (p.adj ≤ 0.05) but are not enriched in either the lung tissue or BALF proteomes (p.adj > 0.05). Pathway analysis was conducted using InnateDB using Uniprot Protein IDs. Vertical black line indicates significance (p.adj ≤ 0.05) after Benjamini-Hochberg multiple comparison correction.

Figure 4: Protein-Protein interaction networks that drive the lung tissue, BALF and Integrated-Tissue-BALF proteome biology are distinct. **A)** Top 3 most connected protein hubs (protein-protein interactions) that are unique to the lung tissue dataset (Hdac1, Ctnnb1, Smarca4). **B)** Top 3 most connected protein hubs that are unique to the BALF dataset (Csf1r, Dnm1, Akt1). **C)** The Integrated-Tissue-BALF protein-protein interaction network using the top 3 hubs identified from either the lung tissue or BALF datasets (Hdac1, Ctnnb1, Smarca4, Csf1r, Dnm1, Akt1). The most significantly enriched biological associations (p.adj ≤ 0.05) determined using the Reactome module of NetworkAnalyst for each interactome is shown on the right-hand side of panel. Hubs are coloured to identify direct protein-protein interactions.

Figure 5: Integrating proteomes not only enrich protein-protein interactions but also connects otherwise separate proteomes together. (**A**) Interactome map based on proteins that are unique to tissue (**B**) Interactome map based on proteins that are unique to BALF. **(C)** Interactome map on unique to-lung tissue and -BALF protein hubs in the UtC dataset. The most significantly enriched biological associations (p.adj ≤ 0.05) determined using the Reactome module of NetworkAnalyst for each interactome is shown on the right hand side of panel. Hubs are coloured to identify direct protein-protein interactions. Black lines signify interactions between different protein hubs. Proteins which have split colouration are connected to two independent protein hubs.

Table 3A: Lung tissue protein hub enrichment in the Unique-to Combined dataset. We identified the protein interaction hubs that are unique to lung tissue and become enriched in the Unique-to-Combined dataset. For simplicity, the Unique-to-Combined dataset was filtered to have ≥ 5 connections and only the top 10 protein hubs (by % enrichment) are shown. Proteins in bold were selected for further analysis.

Unique to	Uniprot ID	Gene	# of Interactions		Protein Hub Enrichment in the Unique-to-Combined Dataset (%)
			Tissue	Unique-to-Combined	
Tissue	**Q99LI8**	**Hgs**	**6**	**8**	**33.33**
Tissue	**Q9ES28**	**Arhgef7**	**8**	**10**	**25.00**
Tissue	**Q9DBR0**	**Akap8**	**4**	**5**	**25.00**
Tissue	Q9CQU3	Rer1	5	6	20.00
Tissue	Q9Z2N8	Actl6a	5	6	20.00
Tissue	Q80TH2	Erbb2ip	6	7	16.67
Tissue	P09242	Alpl	6	7	16.67
Tissue	Q9Z277	Baz1b	6	7	16.67
Tissue	P70452	Stx4a	6	7	16.67
Tissue	Q68FF6	Git1	7	8	14.29

Table 3B: BALF Protein hub enrichment in the Unique-to-Combined dataset. We identified the protein interaction hubs that are unique to BALF and become enriched in the Unique-to-Combined dataset. For simplicity, the Unique-to-Combined dataset was filtered to have ≥ 5 connections and only the top 10 protein hubs (by % enrichment) are shown. Proteins in bold were selected for further analysis.

Unique to	Uniprot ID	Gene	# of Interactions		Protein Hub Enrichment in the Unique-to-Combined Dataset (%)
			BALF	Unique-to-Combined	
BALF	**P97864**	**Casp7**	**7**	**13**	**85.71**
BALF	**P83940**	**Tceb1**	**3**	**5**	**66.67**
BALF	**P10923**	**Spp1**	**4**	**6**	**50.00**
BALF	**P39053**	**Dnm1**	**14**	**21**	**50.00**
BALF	O70456	Sfn	5	7	40.00
BALF	P31750	Akt1	14	19	35.71
BALF	Q01279	Egfr	13	17	30.77
BALF	P33175	Kif5a	8	10	25.00
BALF	P97797	Sirpa	8	10	25.00
BALF	P68134	Acta1	8	9	12.50

Supplementary Information (Expanded View)

Supplemental 1: Experimental Workflow. Six mice were split into two groups and exposed to either HDM (House Dust Mite) or PBS (naïve) for two weeks. On day 14, mice were anesthetized and lung function data was collected using a small animal ventilator. Bronchial Alveolar Lavage Fluid (BALF) was collected, spun down to collect immune cells and flash frozen in liquid nitrogen. Post-BAL lung tissue was portioned into 5 equal segments and flash frozen in liquid nitrogen. Lung tissue was washed in PBS to remove excess blood and subsequently homogenized. Samples were processed using FASP (Filter Assisted Sample Preparation), trypsinized, desalted by 1D HPLC and quantified using an online LC-MS/MS proteomic system. Protein ID's were identified using our X!Tandem informatic pipeline and subsequent bioinformatic analysis was conducted on the data. On a per sample basis, log2 normalized protein abundance was converted to z-scores and then normalized to HDM exposure. Using these normalized datasets, the Tissue, BALF and the Integrated-Tissue-BALF and Unique-to-Combined datasets were used as inputs to assess the differences between the BALF and tissue datasets across multiple levels of analysis including proteins, pathways and networks.

Supplemental 2: HDM exposure induces altered lung function and increased inflammatory cell counts. **A)** Lung mechanics were measured 48 h after last HDM treatment using a flexivent small animal ventilator. Increasing doses of methacholine (3-50 mg/mL) intranasally administered to measure changes in airway resistance, tissue resistance and tissue elastance. **B)** Differential immune cell counts from Bronchial Alveolar Lavage Fluid (BALF) of naïve and HDM exposed mice. Each value is representative of the mean and SEM of three biological replicates. Statistical significance was determined by a two-way nested ANOVA with Tukey's post-hoc test for lung function and unpaired t-test with welch's correction for cell counts. FDR corrected p-values are reported. *(p.adj $\leq$ 0.05), **(p.adj $\leq$ 0.01), ****(p.adj $\leq$ 0.0001).

Supplemental 3: Summary of MS/MS analysis from X!Tandem informatic pipeline.

A) Distribution of spectra, non-redundant peptides and quantified protein IDs are significantly enriched in tissue compared to BALF samples. An EV cutoff of -1.5 and a minimum of two peptides were needed for each peptide and protein ID. **B)** Summary of MS/MS data. Unpaired one-way t-test with welch's correction and FDR adjustment was used for statistical analysis. Abbreviations used: Mouse (In House Mouse ID), number of spectra (SPEC), number of peptides (PEPS), number of non-pedundant peptides (NR-PEPS), number of proteins (PROTS), number of quantified proteins (QPROT), number of quantified peptides (QPEPS), Mean and Standard Deviation of Log_2 MS/MS Intensity (MEAN & SD), Bronchial Alveolar Lavage Fluid (BALF), House Dust Mite (HDM). *(p.adj $\leq$ 0.05), ****(p.adj $\leq$ 0.0001).

Supplemental 4: Proteomic quality control and reproducibility tests indicate proteomic variability resides in biological and not technical replicates. **A,B)** Representative SDS-PAGE gels stained for total protein loading with coomassie blue of tissue and BALF protein homogenates (20 µg loading for tissue, 7 µg for BALF). Each lane represents a different mouse. **C,D)** Technical replicate assessment of Log_2 MS/MS intensities from mouse BALF. **E,F)** Biological replicate assessment of Log_2 MS/MS intensities from mouse BALF. Deviations from the plot origin indicate variability (C,E). Red line indicates linear goodness of fit (D,F).

Figure 1

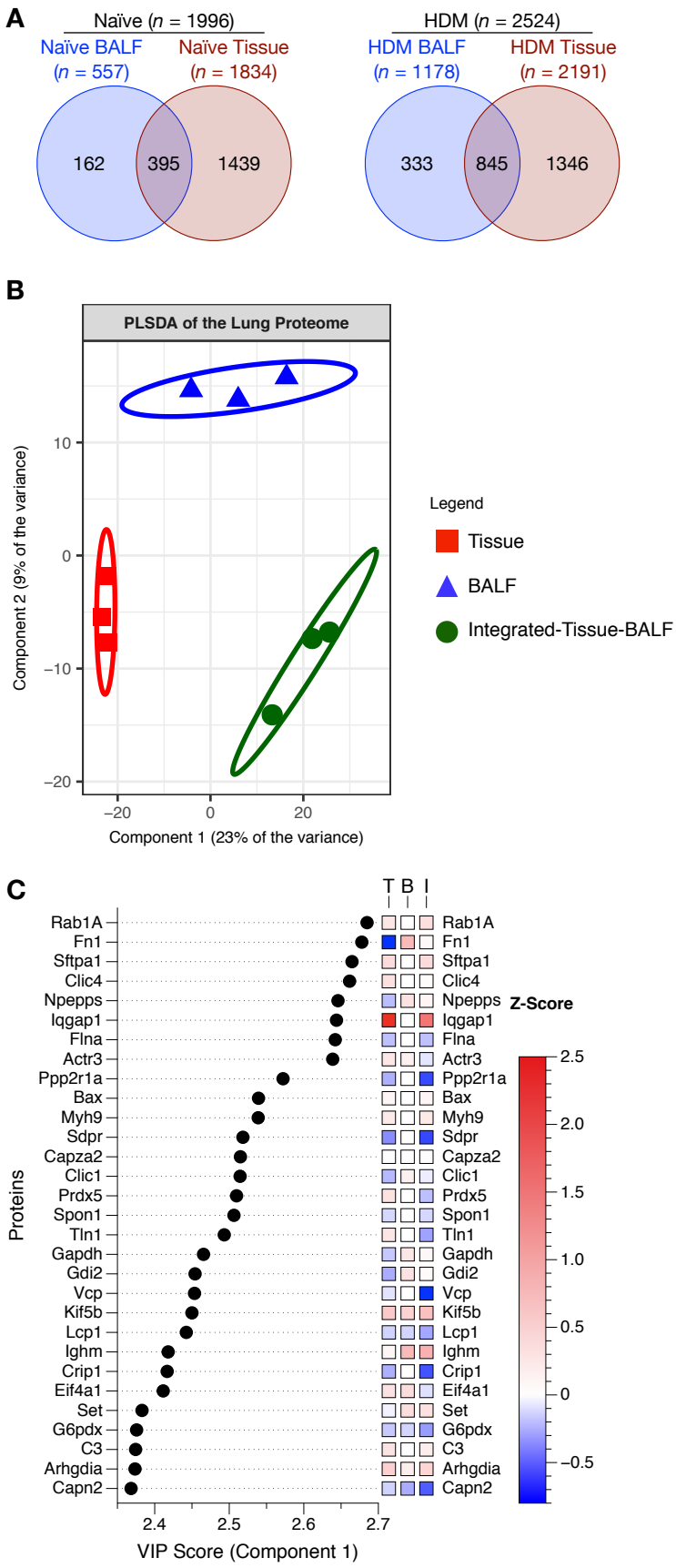

A
Naïve (n = 1996)
Naïve BALF (n = 557)
Naïve Tissue (n = 1834)
162
395
1439
HDM (n = 2524)
HDM BALF (n = 1178)
HDM Tissue (n = 2191)
333
845
1346
B
PLSDA of the Lung Proteome
Component 2 (9% of the variance)
Component 1 (23% of the variance)
Legend
Tissue
BALF
Integrated-Tissue-BALF
C
T B I
Proteins
Rab1A
Fn1
Sftpa1
Clic4
Npepps
Iqgap1
Flna
Actr3
Ppp2r1a
Bax
Myh9
Sdpr
Capza2
Clic1
Prdx5
Spon1
Tln1
Gapdh
Gdi2
Vcp
Kif5b
Lcp1
Ighm
Crip1
Eif4a1
Set
G6pdx
C3
Arhgdia
Capn2
VIP Score (Component 1)
Z-Score

Figure 2

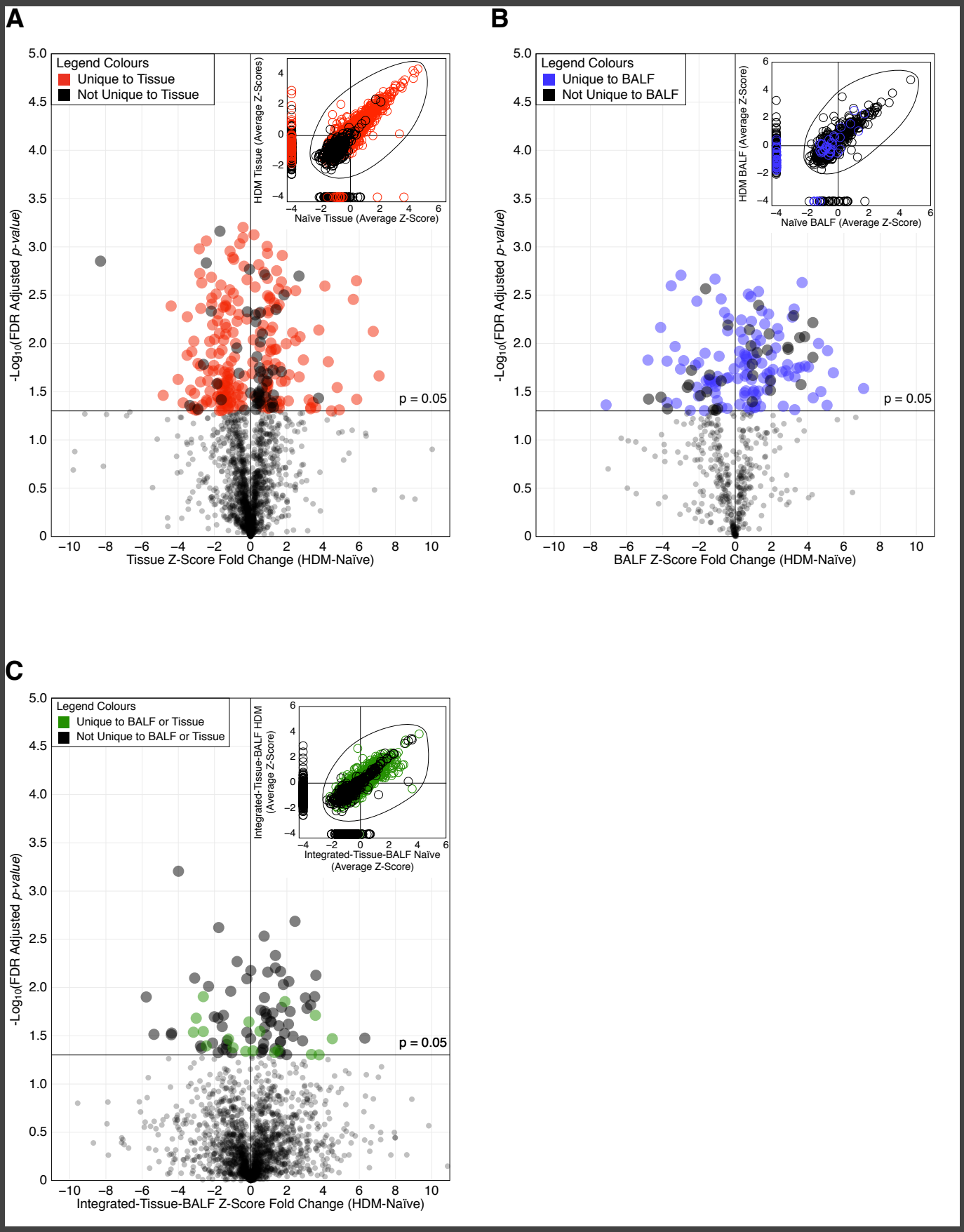

A
Legend Colours
Unique to Tissue
Not Unique to Tissue
HDM Tissue (Average Z-Scores)
Naïve Tissue (Average Z-Score)
-Log10(FDR Adjusted p-value)
p = 0.05
Tissue Z-Score Fold Change (HDM-Naïve)
B
Legend Colours
Unique to BALF
Not Unique to BALF
HDM BALF (Average Z-Score)
Naïve BALF (Average Z-Score)
-Log10(FDR Adjusted p-value)
p = 0.05
BALF Z-Score Fold Change (HDM-Naïve)
C
Legend Colours
Unique to BALF or Tissue
Not Unique to BALF or Tissue
Integrated-Tissue-BALF HDM (Average Z-Score)
Integrated-Tissue-BALF Naïve (Average Z-Score)
-Log10(FDR Adjusted p-value)
p = 0.05
Integrated-Tissue-BALF Z-Score Fold Change (HDM-Naïve)

Figure 2…continued

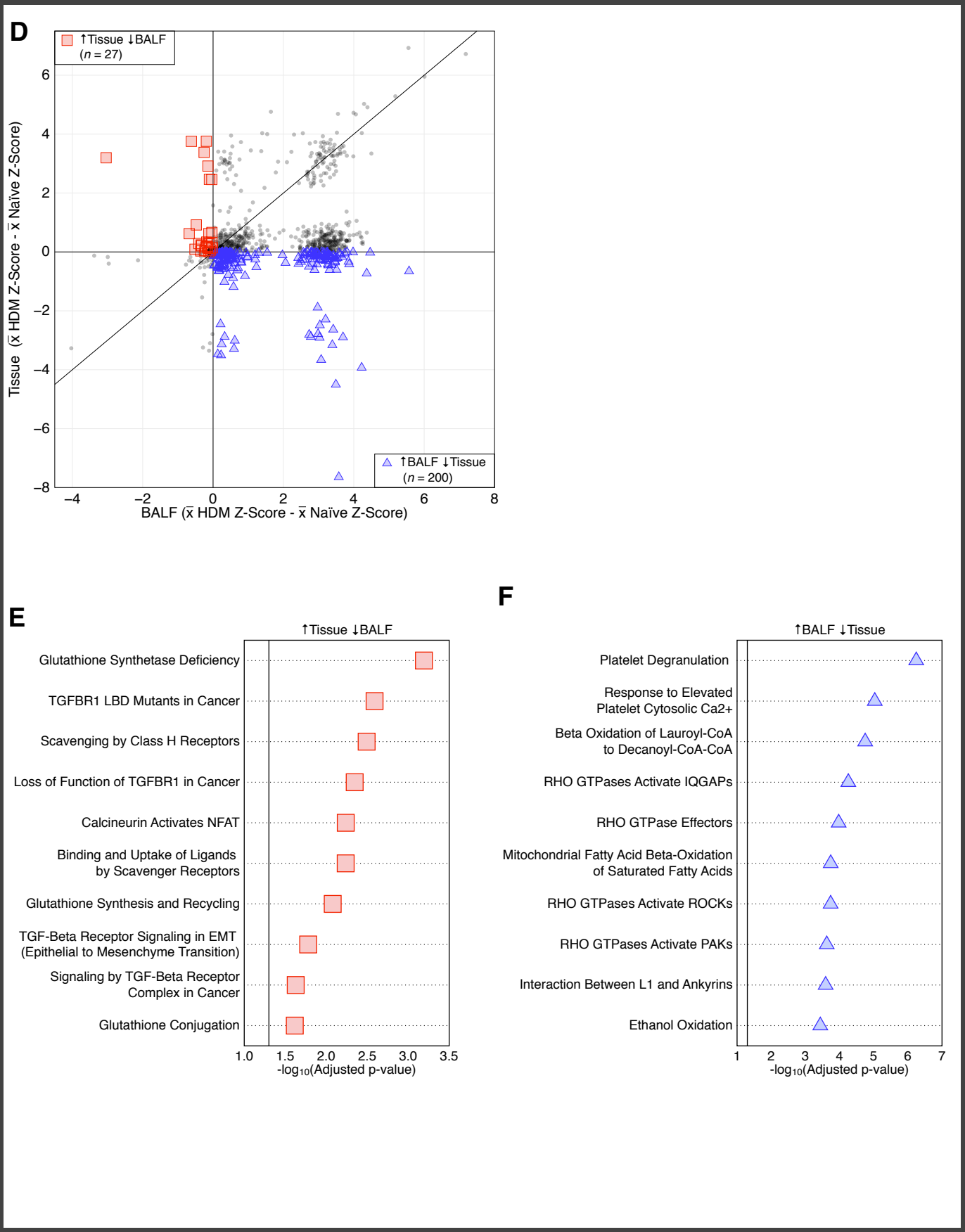


Figure 3

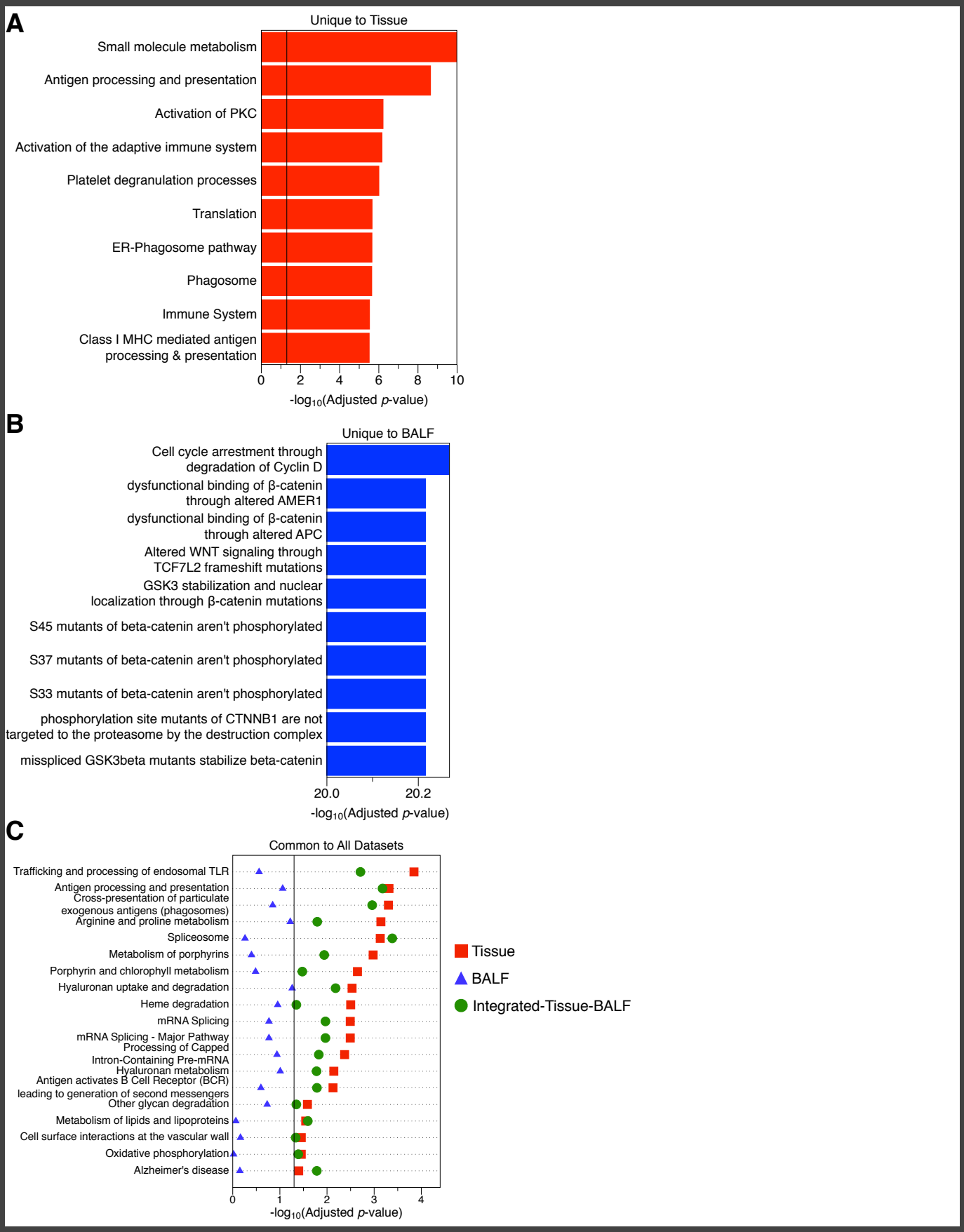
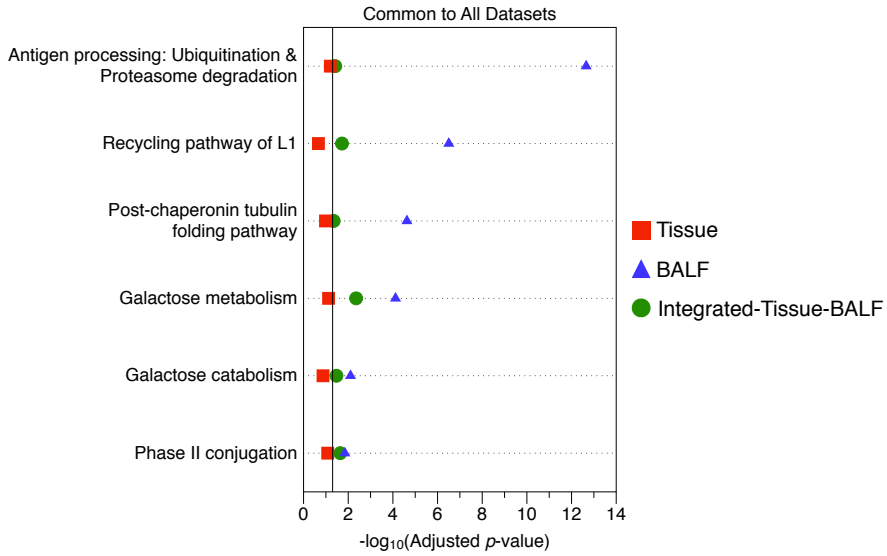

A
Unique to Tissue
Small molecule metabolism
Antigen processing and presentation
Activation of PKC
Activation of the adaptive immune system
Platelet degranulation processes
Translation
ER-Phagosome pathway
Phagosome
Immune System
Class I MHC mediated antigen processing & presentation
0
2
4
6
8
10
-log10(Adjusted p-value)
B
Unique to BALF
Cell cycle arrestment through degradation of Cyclin D
dysfunctional binding of β-catenin through altered AMER1
dysfunctional binding of β-catenin through altered APC
Altered WNT signaling through TCF7L2 frameshift mutations
GSK3 stabilization and nuclear localization through β-catenin mutations
S45 mutants of beta-catenin aren't phosphorylated
S37 mutants of beta-catenin aren't phosphorylated
S33 mutants of beta-catenin aren't phosphorylated
phosphorylation site mutants of CTNNB1 are not targeted to the proteasome by the destruction complex
misspliced GSK3beta mutants stabilize beta-catenin
20.0
20.2
-log10(Adjusted p-value)
C
Common to All Datasets
Trafficking and processing of endosomal TLR
Antigen processing and presentation
Cross-presentation of particulate exogenous antigens (phagosomes)
Arginine and proline metabolism
Spliceosome
Metabolism of porphyrins
Porphyrin and chlorophyll metabolism
Hyaluronan uptake and degradation
Heme degradation
mRNA Splicing
mRNA Splicing - Major Pathway
Processing of Capped Intron-Containing Pre-mRNA
Hyaluronan metabolism
Antigen activates B Cell Receptor (BCR) leading to generation of second messengers
Other glycan degradation
Metabolism of lipids and lipoproteins
Cell surface interactions at the vascular wall
Oxidative phosphorylation
Alzheimer's disease
0
1
2
3
4
-log10(Adjusted p-value)
Tissue
BALF
Integrated-Tissue-BALF

Figure 3…continued

D



E

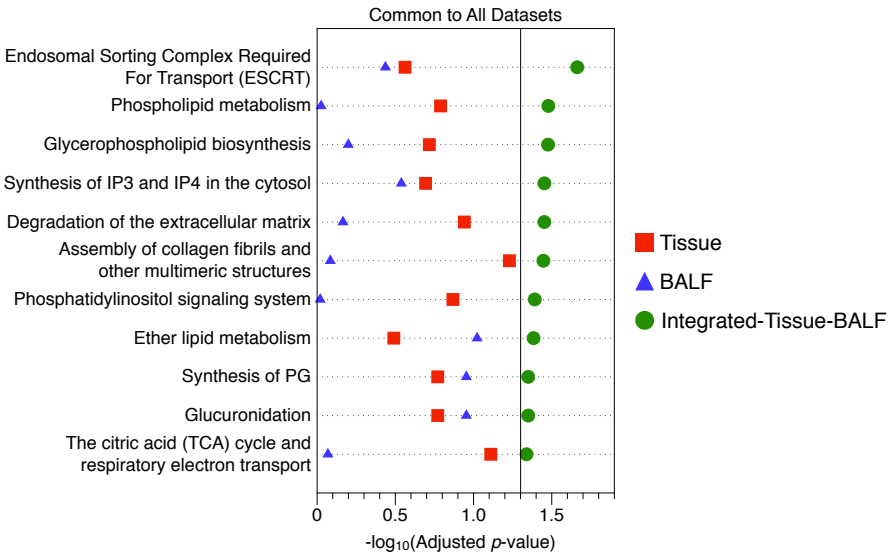


Figure 4

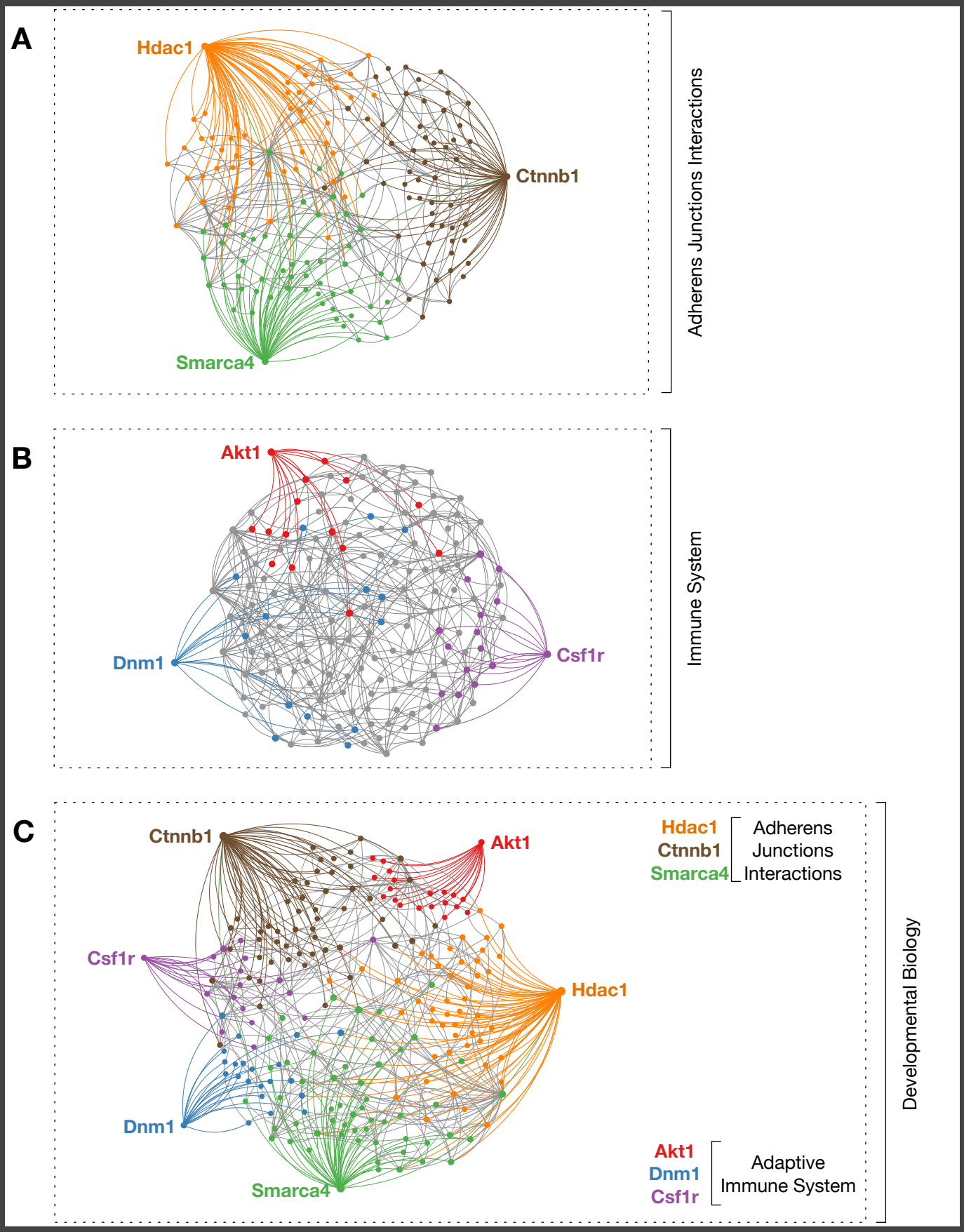
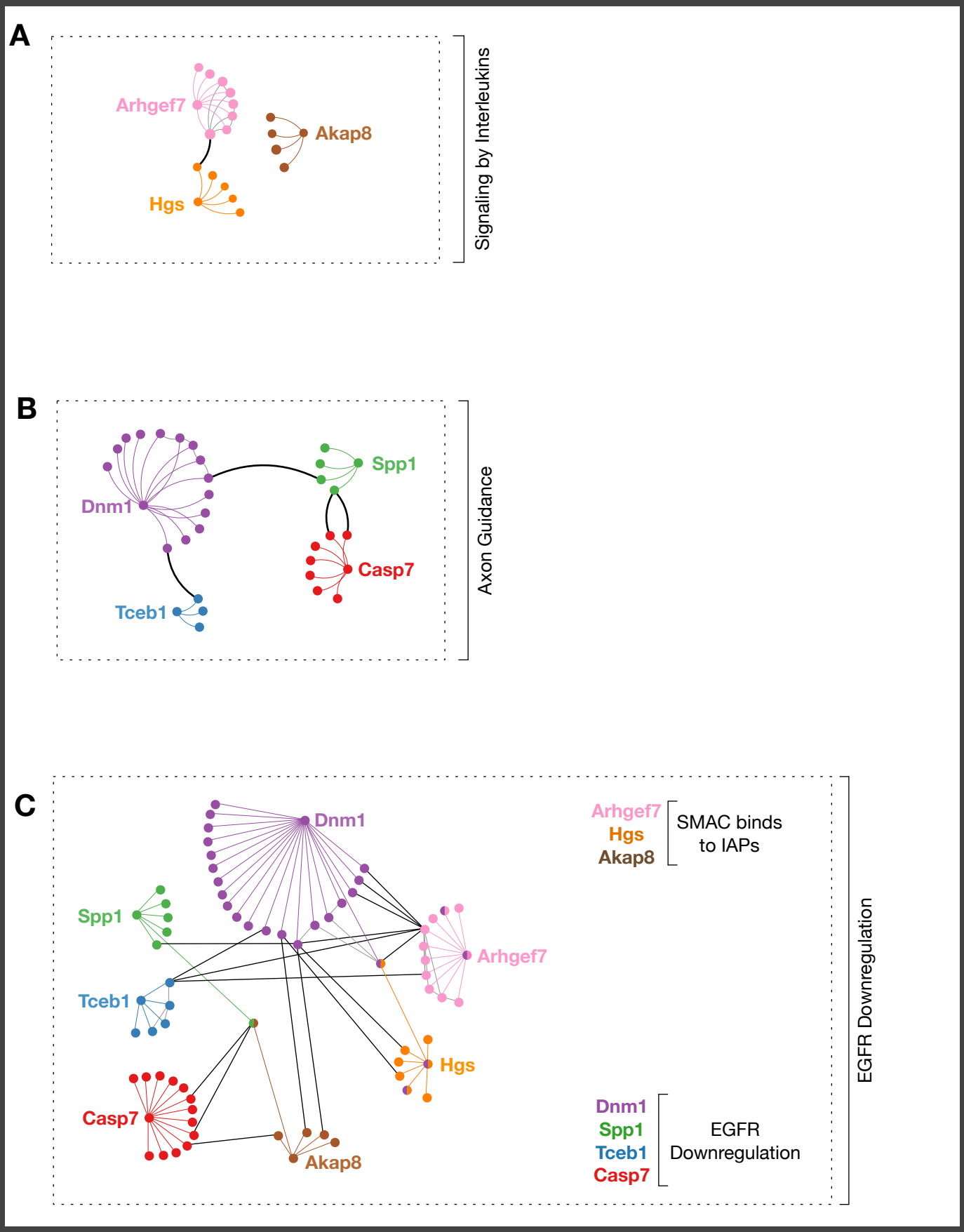

A
Hdac1
Ctnnb1
Smarca4
Adherens Junctions Interactions
B
Akt1
Dnm1
Csf1r
Immune System
C
Ctnnb1
Akt1
Csf1r
Hdac1
Dnm1
Smarca4
Hdac1
Ctnnb1
Smarca4
Adherens
Junctions
Interactions
Akt1
Dnm1
Csf1r
Adaptive
Immune System
Developmental Biology

Figure 5


A
Arhgef7
Akap8
Hgs
Signaling by Interleukins
B
Dnm1
Spp1
Casp7
Tceb1
Axon Guidance
C
Dnm1
Spp1
Tceb1
Arhgef7
Hgs
Casp7
Akap8
Arhgef7
Hgs
Akap8
SMAC binds to IAPs
Dnm1
Spp1
Tceb1
Casp7
EGFR Downregulation
EGFR Downregulation