Integrated in vivo multiomics analysis identifies p21-activated kinase signaling as a driver of colitis

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="https://doi.org/10.1126/scisignal.aan3580">https://doi.org/10.1126/scisignal.aan3580</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Association for the Advancement of Science (AAAS)</td>
</tr>
<tr>
<td>Version</td>
<td>Author’s final manuscript</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/118787">http://hdl.handle.net/1721.1/118787</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
Title: Integrated in vivo multi-omics identifies p21-activated kinase signaling as a driver of colitis

Authors: Jesse Lyons\textsuperscript{a,h,†}, Douglas K. Brubaker\textsuperscript{a,h,†}, Phaedra C. Ghazi\textsuperscript{a}, Katherine R. Baldwin\textsuperscript{a,c}, Amanda Edwards\textsuperscript{d,e}, Myriam Boukhali\textsuperscript{a,e}, Samantha Dale Strasser\textsuperscript{a,b,f}, Lucia Suarez-Lopez\textsuperscript{g}, Yi-Jang Lin\textsuperscript{a}, Vijay Yajnik\textsuperscript{h}, Joseph Kissil\textsuperscript{i}, Wilhelm Haas\textsuperscript{d,e}, Douglas A. Lauffenburger\textsuperscript{b,g}, and Kevin M. Haigis\textsuperscript{a,e,j,*}

One Sentence Summary: Integrated transcriptomics, proteomics, and phospho-proteomics reveals mechanisms of signal transduction and therapeutic targets in physiologic colitis.

Affiliations:

\textsuperscript{a} Cancer Research Institute and Department of Medicine, Beth-Israel Deaconess Medical Center, Boston, Massachusetts, USA
\textsuperscript{b} Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
\textsuperscript{c} Department of Pediatric Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts, USA
\textsuperscript{d} Center for Cancer Research, Massachusetts General Hospital, Boston, Massachusetts, USA
\textsuperscript{e} Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA
\textsuperscript{f} Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
\textsuperscript{g} David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
\textsuperscript{h} Department of Medicine, Division of Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts, USA
\textsuperscript{i} Department of Cancer Biology, The Scripps Institute, Jupiter, Florida, USA
\textsuperscript{j} Harvard Digestive Disease Center, Harvard Medical School, Boston, Massachusetts, USA

* To whom correspondence should be addressed: khaigis@bidmc.harvard.edu
† Equally contributing authors.
Abstract: Inflammatory bowel disease (IBD) is a chronic disorder of the gastrointestinal tract. The molecular mechanisms underlying IBD are poorly characterized and treatment options are limited. To gain insight into the pathogenesis of chronic colonic inflammation (colitis), we performed a multi-omic analysis that integrates RNA microarray, total protein mass spectrometry (MS), and phospho-protein MS (pMS) measurements from a mouse model of the disease. Because we collected all three types of data from individual samples, we were able to track information flow from RNA to protein to phospho-protein to identify species that were coordinately or discordantly regulated. We used this information in two ways. First, we identified pathways that demonstrate complex in vivo regulation. For example, the genes encoding acute phase proteins were expressed in the liver, but the proteins were detected via MS in colons during inflammation. Second, we used the multi-dimensional dataset to ascertain which facets of chronic inflammation are described by each type of data. Using gene set enrichment analysis and trans-omic co-expression network analysis, we found that each data set provides a unique viewpoint on the molecular pathogenesis of colitis. Nevertheless, all of the mouse data sets, as well as human transcriptomic data, implicated increased p21-activated kinase (Pak) signaling as a driver of colitis and chemical inhibition of Pak1/2 with FRAX597 suppressed active colitis in mice. These studies provide a comprehensive view of the state of signaling in the context of colitis and identify Pak as a therapeutic target in IBD.
Introduction

Inflammatory bowel disease (IBD), composed of Crohn’s disease (CD) and ulcerative colitis (UC), affects more than 5 million people worldwide. Sufferers experience a variety of debilitating gastrointestinal symptoms that require medical and, eventually, surgical intervention. The ultimate target of medical treatment is mucosal healing, which has been demonstrated to lead to improved outcomes. However, this goal remains elusive in many patients. While there have been recent advances in therapeutics, the current treatment options for IBD are limited and include general immunomodulators and targeted antibody-based biologics such as anti-tumor necrosis factor alpha (TNF-α), anti-α4β7 integrin, and anti-IL-12/23 (1). All of these therapies suffer from variable efficacy and non-durable response, as well as significant side effect profiles. A better understanding of the molecular pathogenesis of IBD would lead to new therapeutic strategies that could lead to more effective treatments.

Genetic and epidemiological studies in human patients (2, 3), as well as experimental studies in animal models (4), have identified numerous genetic and environmental risk factors for CD and UC, but have not identified clear driver mutations that lead to therapeutic opportunities. In order to account for the complexity in IBD etiology, researchers have taken transcriptomic (5, 6), proteomic (7, 8), metabolomic (9, 10) and metagenomic (11, 12) approaches in an attempt to understand global disease networks and to identify novel differentially expressed genes, proteins, and metabolites that may be involved in disease pathogenesis. While these approaches have provided valuable insight, they have fallen short of identifying potential high value therapeutic targets in IBD.

In this study, we have generated a unique multi-omic dataset in which transcriptomic, proteomic, and phospho-proteomic measurements were made from individual colons of mice with and without colitis. We have used this dataset to understand the relationships between RNA expression, protein expression, and protein phosphorylation and to determine what each measurement reveals about gut inflammation. Since all three types of data were collected from each individual sample, we were able to identify discrepancies between transcriptomic and proteomic measurements, allowing us to predict post-translational protein regulation and to identify expression changes originating at distant organ sites. Finally, we performed co-expression network analysis to identify signaling pathways that were coordinately or uniquely dysregulated in the different data sets and we computationally inferred kinase activation from global pMS data through collation of kinase substrate lists inputted in the gene set enrichment analysis (GSEA) algorithm. These complementary computational approaches implicated Pak signaling as a potential driver of colitis, which we validating by performing a preclinical therapeutic study in mice. Altogether, these studies provide an unprecedented view of dysregulated signaling in colitis and identify a previously unrecognized pathogenic signaling pathway that represents a viable therapeutic opportunity.

Results

Collection of multi-omic data from mouse colon

The initial goal of this study was to quantify global transcriptomic and proteomic changes that occur during chronic colitis. Because this requires a large amount of starting material, we chose
to use a mouse model of IBD, namely the adoptive transfer mouse model of CD, known for high penetrance and relatively short latency (13). Rag1 null animals on a C57BL/6J genetic background were injected with 400,000 CD45RBhi naïve T cells or, as a negative control, 200,000 regulatory T cells (Tregs) from isogenic wild-type (WT) animals and then weighed bi-weekly and assessed for symptoms related to the onset of colitis, such as diarrhea and rectal prolapse. Animals were sacrificed following sustained weight loss of greater than 1.5 grams for one week, which we found to be indicative of severe colitis. Control animals were sacrificed concomitantly. Upon sacrifice, 3mm of tissue from the medial colon was removed and fixed for histological assessment (fig. S1). The remaining colon was opened longitudinally and approximately 1/8th was snap frozen for microarray analysis, while the remaining matched tissue was snap frozen for mass spectrometry (fig. S2). This tissue processing strategy allowed us to collect RNA, total protein, and phospho-protein from an individual colon, and then to perform relative quantification among control and experimental animals.

Following sample processing and data collection, we quantified 39,325 named (38,666 unique) RNA transcripts, 8,131 proteins (7,951 unique proteins), and 3,159 phospho-peptides representing 3,325 unique phosphorylation sites on 1,711 proteins (tables S1-S3). Our initial analysis indicated that each of the three measurements segregated the inflamed mice from the non-inflamed via unsupervised hierarchical clustering (Fig. 1A). Of the 7,951 proteins measured by total MS, 7,611 (96%) were represented in the microarray data set. RNA transcripts were measured for 1,634 (95%) of the 1,711 proteins in the pMS data set. Total protein MS data was obtained for 1,474 (86%) of the 1,711 proteins measured by pMS and 1,415 species were measured in all three data sets.

Because we measured RNA, protein, and phospho-protein from individual samples, we were able to perform one-to-one matched correlation of individual genes across measurements (Fig. 1B). The smoothed probability density functions for Spearman correlations showed a correlation landscape for RNA/MS comparison that was distinct from RNA/pMS and MS/pMS. Most RNA/MS gene pairs were positively correlated with each other with only a few species showing inverse correlation (Fig. 1B). The RNA/pMS probability density function showed a relatively flat, slightly bimodal, distribution indicating that there were similarly sized sub-groups of RNA molecules for which there was positive correlation, no correlation, or negative correlation between RNA abundance and phospho-peptide abundance. There were two peaks of weakly positive and weakly negative correlation between matched RNA/pMS species. Overall, the shape of the RNA/pMS distribution indicated that there were more species for which there was no correlation between RNA and pMS than positive correlation and that the group of inverse correlation between RNA and pMS was the smallest.

The MS/pMS probability distribution was similar to the RNA/pMS distribution, but there were more positively correlated MS to phospho-peptide species than RNA/pMS species. There were also fewer MS to pMS species with no correlation between total protein and phospho-peptide abundance (Fig. 1B). We reasoned that the inverse correlation between some MS/pMS species might suggest proteins whose stability is post-translationally regulated, with phosphorylation marking the protein for degradation. Eplin (encoded by the Lima1 gene), which has a known ubiquitin-priming phosphorylation site, provides an example of this type of regulation. Lima1 transcript levels were essentially unchanged (1.1-fold down-regulated) in inflamed versus non-inflamed tissue, yet the protein level was nearly two-fold decreased (Fig. 1C). At the same time there was a 3.3-fold induction in phosphorylation at Ser360 (Fig. 1C and D), which targets the
protein for ubiquitination and degradation (14).

Differential expression analysis

To identify the RNAs, proteins, and phospho-peptides that were differentially expressed in inflamed colons, we used the Wilcoxon-Mann Whitney (WMW) test with a Benjamini-Hochberg false discovery rate (BH FDR) correction (WMW p<0.05, BH FDR q<0.25). Overall, 7,752 of 38,666 RNA transcripts, 4,443 of 7,951 proteins, and 2,346 of 3,325 phospho-peptides were differentially expressed (table S4). Of the 7,611 overlapping RNAs and proteins, 1,858 were similarly differentially expressed, while 1,064 RNAs and 2,401 proteins were differentially expressed in only one data set (Fig. 1E). All of the 1,474 proteins measured by both pMS and MS were differentially expressed at either the total protein or phospho-peptide level or both (Fig. 1E). Finally, none of the 1,634 RNAs with co-measured phospho-peptides were differentially expressed, resulting in limited differential expression similarity between all three data sets (Fig. 1E). This comparison provides the first demonstration that the pMS data set tells a truly unique story about the molecular pathogenesis of colitis.

Because our goal was to characterize changes to the tissue-level signaling network during colitis, we isolated RNA and protein from whole colon, which includes the epithelium, lamina propria (including immune, stromal, and vascular cells), and muscularis as the starting material for our RNA and protein datasets. The cellular representation of an inflamed colon is different from that of a normal colon, most significantly with respect to the influx of inflammatory cells (fig. S1). To explore whether the changes in signaling we detected via transcriptomics and proteomics simply reflected the influx of inflammatory cells, we used immunohistochemistry to analyze the cellular localization of signals that were up-regulated in animals with colitis. For example, the phosphorylation of Trim28 (Ser473) and Map3k3 (Ser337) was significantly greater during colitis (Fig. 1F). We found that Trim28 phosphorylation was increased in the colonic epithelium, while Map3k3 phosphorylation was increased in all components of the colonic environment (Fig. 1F). These observations indicate that the changes to RNAs and proteins that occur during colitis are not just a reflection of a major change in the cellular representation of the tissue, but rather reflect broad, and potentially pathologic, changes to the tissue-level signaling network.

Pathway enrichment analysis

While these analyses described how the data sets related to one another on a species-by-species level, we wanted to determine whether the differences between inflamed and non-inflamed samples in each data set represented similar pathways and high level functional categories. We ran GSEA (15) on the overlapping 7,611 genes from the RNA and total protein MS expression data sets to identify differentially regulated pathways between inflamed and non-inflamed mice. At the RNA level, 19 pathways were positively enriched and 8 pathways were negatively enriched in inflamed mice (p<0.05, FDR q<0.25) (Fig. 2A and table S5). These largely overlapped with 15 positively enriched and 10 negatively enriched pathways in inflamed mice at the total protein level, suggesting strong functional concordance between the transcriptomics and proteomics data (Fig. 2B and table S5). The finding that 14/20 positively enriched and 8/10 negatively enriched pathways were similarly regulated in inflamed mice was consistent with the Spearman correlation histogram of the RNA/MS data (Fig. 1B). Coordinately positively enriched
pathways included those involved in inflammatory response and tumor necrosis factor alpha signaling, and also pathways controlling epithelial proliferation, such as E2F targets, KRAS signaling, and MYC targets (Fig. 2C). Interestingly, the “epithelial to mesenchymal transition” gene set was positively enriched at the RNA level and negatively enriched at the protein level, suggesting complex regulation of this particular pathway (Fig. 2C). This observation highlights the potential for RNA analysis to provide misleading information about the role of a particular pathway in driving a disease.

Analysis of non-correlated RNA and protein signals

Comparative analysis of RNA and protein levels allowed us to identify pathways that were coordinately regulated during colitis, but also provided an opportunity to explore if and how each data set provided unique information on genes, proteins, and pathways. For example, in our GSEA analysis, “interferon alpha response” was positively enriched in total protein dataset, but not in the RNA dataset (table S5). To expand upon this observation, we examined the correlation between RNA and protein expression for the 7,611 species for which we collected both types of data. Although we observed general concordance in fold-change differences (control vs. colitis) between RNA and protein measurements, a small number of species were induced at the protein level, but exhibited no change at the RNA level (Fig. 3A, colored dots). To determine if these species fell into particular functional categories, we performed Gene Ontology (GO) enrichment analysis and found enrichment for genes involved in defense response (positively enriched in the MS dataset) and extra-cellular matrix (negatively enriched in the MS dataset). The extra-cellular matrix category of proteins was composed primarily of collagens (Fig. 3A, purple dots) and when we surveyed all of the collagens shared between the microarray and total MS data sets, we found that 12/17 collagens were reduced by at least 2-fold in protein expression, but unchanged or weakly increased in abundance at the RNA level (Fig. 3B). Interestingly, we observed strong induction of several extracellular matrix metalloproteinases (MMPs), including MMPs 3,7,9 and 10, in the MS dataset, many of which play known roles in degrading collagen extracellular matrix. This provides a powerful example of how comparative RNA/protein expression analysis can infer genes for which the corresponding protein is regulated via degradation.

Within the defense response GO category, we found two interesting groups of proteins. The first group was made up of acute phase proteins (Fig. 3A, red dots). Given that the liver coordinates the acute phase response, we hypothesized that these genes may be transcribed and translated in the liver and then travel to the colon through the blood stream. Indeed, when tissue expression of these genes was assessed using the publicly available mouse gene atlas (J6), all were reported to be highly expressed in the liver, with minimal expression in other tissues (Fig. 3C). In order to confirm this expression pattern in our experimental model, two of the genes, Fga and Orm1, were assayed for expression in the colons and livers of animals with and without colitis. For both genes, there was at least 5,000-fold higher expression in the liver than in the colon (Fig. 3D). In inflamed animals, this ratio increased to ~12-50,000-fold (Fig. 3D), indicating that the transcriptional events that ultimately resulted in increased colonic protein expression of acute phase proteins most likely originated in the liver.

Another subset of defense response genes included five neutrophil-specific genes (Fig. 3A, orange dots). This group of genes was somewhat surprising given that the adoptive transfer model of IBD exhibits strong neutrophil recruitment to the colon, and there were many other
neutrophil-specific genes that were increased in abundance at the RNA and protein levels in animals with colitis. According to the mouse gene atlas (16), these five genes exhibit bone marrow-specific expression in normal mice (Fig. 3C), and a recent study demonstrated that there are subsets of genes, including those identified by our differential regulation analysis, that show decreasing RNA expression as neutrophils travel from the bone marrow to the circulating blood (17). To assess whether this expression pattern underlies the discordant colonic RNA and protein expression in our model, we isolated neutrophils by flow cytometry from the bone marrow and colons of animals with colitis. RT-PCR for Camp and Elane revealed an average of 2,257- and 6,142-fold higher expression in the bone marrow derived neutrophils compared to colonic neutrophils from inflamed animals (Fig. 3E). These data suggest that neutrophils transcribe and translate these proteins in the bone marrow, but that the RNA message is degraded by the time the neutrophils reach the site of inflammation, leaving only the protein. Together, these examples demonstrate how overlapping transcriptomic and proteomic data can be used to infer regulation of expression in distant organ sites. In both cases, the message and protein were initially expressed in a distant organ, but carried to the colon via the blood (acute phase proteins) or within cells (neutrophil proteins) (Fig. 3F).

Trans-omic co-expression network analysis

To further address the relative information content of the three data sets, we assessed global co-expression behavior of the 1,415 species (1,429 RNA transcripts, 1,452 proteins, 3,080 phosphopeptides) that were present in all data sets. Spearman correlation coefficients were calculated for all pairs of variables within each data set and dichotomized with correlations greater than 0.9 or less than -0.9 set to 1 and all others set to 0. We visualized both the correlation networks and the two-step generalized topological overlap matrix (GTOM2) to explore the correlation landscape, modularity, and existence of highly correlated subsets of variables of each data set (Fig. 4A and B). GTOM2 relates the interconnectedness of two genes by computing the number of shared-two step network neighbors. Clustering the GTOM2 matrix facilitates identification of groups of highly correlated genes called modules (18, 19). At the RNA level we noted several smaller clusters of highly correlated transcripts, while the protein expression and phospho-peptide levels were much less modular (Fig. 4B). To estimate the true number of clusters in the RNA, MS, and pMS data sets, we computed the gap statistic for up to 15 clusters, via k-means clustering, of the GTOM2 matrix (20). The gap statistic measures the within-cluster variation of each cluster in dataset compared to a null distribution of expected within-cluster variation. When the gap curve begins to level off or decrease as a function of the number of clusters, then we can approximately determine the number of clusters in each data set. Based upon the gap curves and clustered GTOM2 matrices, we estimated that there were 8 RNA clusters, 5 MS clusters, and 5 pMS clusters in our data (Fig. 4C).

We extracted the genes in each of these clusters (table S6) and compared the gene lists between clusters using Fisher’s exact test with a BH FDR correction (p<0.05, q<0.25) (table S7). We visualized the significant overlap between modules in a network diagram and found that most RNA modules shared a significant overlap with multiple MS and pMS modules (Fig. 4D). Though some MS and pMS modules overlapped, the connections were sparser and the partially isolated modules (RNA_3, MS_5, and pMS_4) might present interesting cases of specific trans-omic regulation in colitis. To functionally interpret the individual modules in each data set and to examine function across data sets, we analyzed them using the network analysis tool.
YourCrosstalker (21, 22). This allowed us to identify the enriched pathways, filter out network edges not associated with protein-protein interactions, and identify topologically relevant genes not in our original cluster gene lists. The YourCrosstalker random walk was performed on the STRING protein-protein interaction network and a combination of Reactome and the National Cancer Institute Pathway Interaction Database (NCI-PID) were used for pathway enrichment (23-26).

The correlation structure of module RNA_3 was selectively associated with the phosphorylation modules pMS_3 and pMS_4. The 54 genes in RNA_3 showed enrichment only for nerve growth factor (Ngf) signaling, while the 121 genes in pMS_4 showed enrichment only for processing of capped intron-containing pre-mRNAs. In contrast, we found a greater diversity of pathway enrichments in the YourCrosstalker modules of the 397 genes in pMS_3 (Fig. 4E). Here we found a central network that connected multiple Vegf signaling, p38 signaling, cell cycle, and pre-mRNA processing pathways (Fig. 4E). We next analyzed the 221 proteins in the module MS_5, a protein module that did not significantly associate with any pMS modules, but significantly overlapped with the RNA modules 2, 4, 5, and 6 (Fig. 4F). Interestingly, we found that similar pathways were implicated by the genes in MS_5 as in pMS_3. In particular, we noted the shared MS_5-pMS_3 pathways of Vegf, pre-RNA processing, cell cycle, and p38 signaling, as well as the presence of Ngf signaling shared with RNA_3 (Fig. 4F). This observation suggests that the signaling network at the intersection of these pathways plays a role in colitis and that the components of this network are dysregulated in distinct ways in the different -omic data sets. Importantly, this discovery was enabled by the integrated picture provided by our multi-omic data set and trans-omic analysis approach.

Inferring kinase activity from phospho-proteomic data

Our overarching goal is to understand how dysregulated signaling contributes to the pathogenesis of colitis and to determine whether there are signaling pathways that could represent novel therapeutic targets for IBD. To this end, we reasoned that our pMS data set would most accurately reflect therapeutically tractable changes in signal transduction that occur during colitis. Our pMS analysis identified 2,346 differentially phosphorylated peptides in animals with colitis and, although 80% of the phospho-peptides had been previously identified, only 1.68% were functionally annotated (fig. S3). In essence, the biological significance of the vast majority of measured phosphorylation events is presently unknown. We sought to overcome this deficiency in prior knowledge by utilizing known kinase-substrate relationships to computationally infer kinase activity from pMS data. Lists of kinase-substrate relationships for 348 kinases across a range of kinase families were curated from PhosphoSitePlus (table S8) (27). These were inputted as ‘gene’ lists into the GSEA algorithm, allowing us to perform a “kinase activation” analysis from pMS data. Although many kinases showed substrate enrichment in inflamed colons (table S9 and Fig. 5A), only Pak1, which had an uncertain role in IBD pathogenesis, reached significance as defined by an FDR less than 0.25 (Fig. 5A,B). Interestingly, Pak1 and Pak2 were part of several of the YourCrosstalker modules that we identified in our co-expression network analysis (Fig 4E and F and table S6), providing a cross validation of the different computational analyses.

In addition to predicted activation of Pak1 during colitis, GSEA predicted that six kinases – Casein Kinase 2A1 (Csnk2a1), Gsk3α/β, p38α (Mapk14), Casein Kinase 1D (Csnk1d), and
Dyrk1a — were activated in the control animals relative to animals with colitis (Fig. 5A and B). Of these kinases, only p38, one of the stress-induced mitogen-activated protein kinases (MAPKs), has been linked to inflammation (28). To confirm the predicted kinase activation, we measured the phosphorylation of sites on Gsk3α/β and p38 that are known to regulate their activation states. Here we used a Luminex-based phosphorylation assays on samples from an independent cohort of animals with adoptive transfer-induced colitis. This analysis revealed that animals with colitis had decreased phosphorylation of p38 (Thr180/Tyr182) and increased phosphorylation of Gsk3α/β (Ser21/Ser9) (Fig. 5C). Phosphorylation of p38 is indicative of kinase activation, while phosphorylation of Gsk3α/β is a negative regulatory modification. As such, this experiment validated the down-regulation of p38 and Gsk3α/β activity as predicted by GSEA on the pMS dataset.

Evidence of Pak signaling in human IBD patients

An important consideration with any experimental model system is the generalizability of the findings of that system to the human in vivo context. To compare our mouse data to human, we obtained a publicly available gene expression dataset of inflamed (n=12) and uninflamed (n=16) IBD patient colonic biopsies (6). Our aim was to assess global concordance between human IBD differential gene expression and, in particular, whether the identification of signaling events in Pak1 and Pak2 were conserved in the human in vivo context. We performed differential expression analysis on the entire human dataset (WMW p<0.05, FDR q<0.25) and compared differential expression of human RNA to mouse RNA, MS, and pMS. There were 1,708 genes differentially expressed between inflamed and uninflamed human samples. All 1,708 human RNAs were represented in the mouse RNA data, 1,040 genes were represented in the mouse MS data, and 529 genes were found in the mouse pMS data. Of the 2,710 homologous mouse RNA transcripts that were differentially expressed, 751 were also differentially expressed in humans (Fig. 6A). Of the 2,608 homologous mouse differentially expressed total protein measurements, 613 were also differentially expressed in the human RNA data (Fig. 6A). Finally, of the 867 homologous differentially phosphorylated proteins measured in the mouse pMS data, 197 were differentially expressed in the human RNA data (Fig. 6A).

In general, more species (RNA, MS, and pMS) tended to be differentially expressed in the mouse relative to the human RNA data. The larger set of differentially expressed mouse RNA, MS, and pMS species was able to capture 43.9%, 58.9%, and 37.2% of the possible homologous differential expression events. This suggests that the mouse MS dataset offers the best experimental representation of the human in vivo disease context. Further, it shows that genes differentially expressed in the human tended to not be differentially phosphorylated in the mouse, a reflection of the generally weak correspondence between RNA and phospho-peptide differential activity in the mouse.

We next sought to determine the extent to which the Pak1 and Pak2 signaling network neighborhood was differentially active in human IBD. We assembled a human protein-protein interaction (PPI) network by querying PAK1 and PAK2 in the Pathway Commons database (25). The Pak network neighborhood contained 529 unique genes and 3,666 interactions. We then filtered the human expression data and PAK network for overlapping genes. After filtering for expression array coverage, the final PAK neighborhood contained 431 genes and 2,534 interactions (Fig. S4). When we overlaid the differentially expressed genes from human IBD
patients onto the Pak network neighborhood, we found that 95 genes were differentially expressed (Fig. 6B) and the largest connected network of these genes implicated PAK2, STAT1, and STAT3 as key hub nodes. This analysis suggests that the inferred PAK signaling mechanism from the mouse model does translate to the human in vivo disease and may be a viable therapeutic target in IBD.

In order to assess the relative importance of PAK signaling compared to other kinases implicated by phospho-peptide based GSEA, we repeated the analysis of constructing a network neighborhood of the kinase and searching for differentially expressed human genes. A hypergeometric test was applied to assess the significance of the overlap between differentially expressed human genes and the network neighborhood of the kinases MAPK14, CSNK2A1, GSK3A, and GSK3B (Figure 5). The Pak neighborhood contained 95 differentially expressed genes and was the most human significant kinase ($p < 10^{-15}$). The MAPK14 neighborhood contained 48 differentially expressed genes ($p < 10^{-6}$), CSNK2A1 contained no differentially expressed genes ($p = 0.097$), GSK3a contained 38 differentially expressed genes ($p < 10^{-7}$), and GSK3b contained 43 differentially expressed genes ($p < 10^{-7}$). While MAPK14, GSK3A, and GSK3B also have significantly active network neighborhoods, the PAK neighborhood is statistically the most significant of the kinases.

Validation of Pak as a therapeutic target in colitis

With multiple lines of evidence pointing toward dysregulation of Pak signaling in colitis, we investigated whether pathway activation was a cause or consequence of the disease. First, we sought to validate, in an independent cohort of animals, that Pak was activated during colitis. Phosphorylation of Pak1/2 on Ser144/Ser141 is associated with kinase activation [29] and western blotting of colonic protein lysates from the new cohort confirmed increased phosphorylation in inflamed colons (Fig. 7A). This observation validated our prior MS analysis predicting Pak1 activation based on increased phosphorylation of its substrates in inflamed colons (Fig. 5A and B). Note that Pak2 was not predicted by GSEA to be activated because not enough of its known substrates (table S8) were represented in the pMS dataset. We also found that Pak1 auto-phosphorylation was increased in animals with acute colitis due to treatment with dextran sodium sulfate, indicating that Pak activation is not specific to the adoptive transfer mouse model of colitis (fig. S5).

Next, we investigated whether inhibition of Pak signaling could suppress colitis. We chose to focus on animals that already had severe inflammation because our goal was determine whether inhibition of the pathway could be effective in patients with active disease. We found that colons from animals with colitis that were treated acutely with FRAX597, a Pak1/2 inhibitor [30], exhibited reduced phosphorylation of Merlin on Ser518, a Pak1/2 substrate (Fig. 7B). Next, we treated sick animals with FRAX597 for 7 days (100 mg/kg/day) and assessed their phenotype by endoscopic monitoring. Following Pak inhibition with FRAX597, animals exhibited decreased mucosal thickness, return of both small and large vessel visible vascular markings, and resolution of contact friability and bleeding (Fig. 7C). These are all signs that the active colitis had been diminished follow Pak inhibition.

At the histologic level, animals treated with FRAX597 exhibited reduced immune cell infiltrate in their colons, and a return to more normal epithelial crypt morphology (Fig. 7D). The colitis that arises in the T cell transfer model is characterized by increased numbers of colonic
macrophages and neutrophils in the lamina propria (31). Interestingly, Pak is known to regulate the chemotaxis of neutrophils (32). To determine whether Pak inhibition alters the immune milieu in the colon, we used fluorescence-activated cell sorting (FACS) to quantify immune cells. Consistent with the reduction in the clinical and histologic presentation of colitis, we found that animals treated with FRAX597 had reduced numbers of colonic macrophages and neutrophils relative to vehicle-treated controls (Fig. 7E). Altogether, our computational and experimental studies implicate Pak signaling as a driver of chronic inflammation in the colon.

Discussion

Here we present the first instance in which RNA microarray, total proteomic, and phosphoproteomic measurements have been analyzed and integrated in matched tissue from single animals in a mouse model of colitis. This “all measurements from a single animal” approach enabled us to determine how changes in gene expression are carried through to protein expression and modification. We found that genes that were differentially expressed at the RNA level showed similar patterns of differential regulation in the MS data set, but that differential expression at the RNA level did not predict pMS phosphorylation status. Furthermore, though most genes differentially expressed at the total protein level were also differentially phosphorylated at the phospho-protein level, the pMS data contained many differentially expressed phospho-peptides that were unchanged at the RNA and total protein level (Fig. 1E). Although there are typically a greater number of molecular species measured via transcriptomics, our finding demonstrates that there are additional layers of molecular regulation that are not represented within transcriptomic datasets, underscoring the importance of proteomic measurement for understanding disease pathogenesis at the molecular level.

We observed stronger concordance between the RNA and MS data at the pathway level than the single gene level (Fig. 2B). The enriched pathways pointed to several dysregulated signaling pathways, such as oxidative phosphorylation and signaling through E2F, KRAS, and MYC, that might provide some therapeutic targets in IBD. Work by Bar et al. demonstrated that mice with DSS and TNBS induced colitis were protected against more severe symptoms when oxidative phosphorylation was more active (33). Their work suggested that increasing the activity of this pathway could reduce inflammation and crypt formation in the intestinal epithelium through up-regulated nuclear factor kappa-light-chain-enhancer of activated B cells (NFkB) signaling (33). Since our analysis identified a link to oxidative phosphorylation in the T cell transfer model of colitis, it appears that the anti-inflammatory activity of this pathway is conserved among the DSS, TNBS, and adoptive transfer models.

Interestingly, while the RNA and MS measurements exhibited conserved patterns of expression and higher-order process enrichment, there were also many proteins that were differentially regulated in inflamed versus non-inflamed tissue, but had unchanged RNA expression. Using GO enrichment analysis, we identified differential regulation in defense response and extracellular matrix. This led us to infer post-translational regulation of collagens via matrix metalloproteinases (Fig. 3B). Additionally, we found a subset of neutrophil and acute-phase proteins that were transcribed and translated in other tissues (bone marrow and liver) and transported to the colon in the blood stream or via infiltrating immune cells (Fig. 3F). These analyses leveraged the transcriptomic and proteomic data to produce hypotheses regarding organismal scale gene and protein regulation that could not have been made with either data set.
alone, underscoring the value of multi-omic measurement. Only multi-omic analysis of an intact organism could identify these instances of physiologic regulation of protein expression.

The collection of multi-omics data enabled us to perform trans-omic co-expression network analysis to examine the correlation structure between all pairs of RNA transcripts, proteins, and phospho-peptides measured in all three data sets (Fig. 4). By clustering the data sets into modules and performing targeted network analysis on the modules with YourCrosstalker, we were able to identify important functional commonalities between the phospho-peptide module pMS_3 and the total protein module MS_5. pMS_3 associated with all 8 RNA modules and was characterized by an intriguing core network of hyper-phosphorylated proteins in the RNA metabolism and pre-RNA processing pathways coupled to a de-phosphorylated network of Vegf signaling. MS_5 was a module that did not associate with any pMS modules, and yet it was characterized by similar coupling of pre-RNA processing pathway proteins to Vegf signaling proteins via the cell cycle signaling pathway (Fig. 4E and F). Though the characteristic proteins of MS_5 and pMS_3 did not overlap, the same signaling pathways and core network architecture were identified by YourCrosstalker, suggesting an important role for these pathways in colitis. RNA processing and cell cycle are general terms and difficult to ascertain specific mechanisms potentially involved in colitis, but Vegf is known to play a role in angiogenesis and lymphangiogenesis during colitis (34). Indeed, Vegf therapy has been shown to be beneficial in mouse models of the disease (35, 36). By extension, Vegf signaling in experimental colitis is likely to be a result of the inflammation as the tissue attempts to repair itself.

The primary goal of this study was to use the multi-dimensional dataset to identify new drivers of colitis and we noted that the pMS provided additional mechanistic insight that was not revealed by RNA or protein expression studies. Because the majority of individual sites identified by phospho-MS were not functionally annotated, we reasoned that phospho-proteomic data coupled with prior knowledge of kinase substrates would allow for computational inference of kinase activity. Public databases and software packages provide information on kinase-substrate interactions, kinase recognition motifs, and kinase substrate predictions and a variety of algorithms have been used to determine a kinase activity metric based on these relationships. For example, similar to our analysis, Drake and colleagues compiled substrate sets and utilized an algorithm analogous to GSEA to quantify enrichment (37). In our study, GSEA predicted one significant positively enriched and six significant negatively enriched kinases from our pMS dataset (Fig. 5A), most of which had unknown or poorly characterized roles in IBD. The activation or inhibition of several of these predicted kinases was confirmed via measurement of regulatory phosphorylation sites on those kinases (Figs. 5C and 7A). Pak1 was of particular interest because our enrichment analysis indicated that it was activated during colitis, suggesting that inhibition could be a therapeutic strategy. Although Pak2 was not implicated by GSEA, western blotting and YourCrosstalker network analysis revealed that it is also activated in animals with colonic inflammation and in human patients (Fig. 4E and F and 6B). Inhibition of Pak1/2 with FRAX597 suppressed inflammation in animals with active colitis, indicating that this pathway plays an active role in the pathogenesis of the disease (Fig. 7C to E). This was the first demonstration that chemical inhibition of Pak signaling can revert inflammatory disease in the colon.

Pak1 and Pak2 are members of the family of group I p21-activated kinases that regulate inflammatory responses, in part by regulating NADPH oxidase in neutrophils (38). Pak1 was previously reported to be over-expressed in epithelial cells during colitis and is thought to be
regulated by mTOR signaling (39). Our YourCrosstalker network analysis identified the mTOR pathway as being activated in the T cell transfer model (Fig. 4E), although our GSEA analysis failed to identify enrichment for mTOR substrates in pMS data from animals with colitis (table S9). As such, the mechanism of activation of Pak1 in the T cell transfer model is not clear. Nevertheless, the effect of Pak inhibition validates the prediction made by our computational modeling approaches: that Pak plays a critical role in the pathogenesis of colitis.

In sum, our analyses demonstrate the added value of multi-omic measurements by showing how different molecular species in each data set may be acting on similar pathways in distinct ways. We used multi-omics comparisons to obtain mechanistic insight into the pathogenesis of chronic inflammation in the colon in particular identifying Pak signaling as a bona fide therapeutic target. This work highlights the power of analyzing the global proteome and phospho-proteome for uncovering dysregulated signaling pathways that are not revealed by transcriptomic studies alone.

**Materials and Methods**

**T cell transfer model of colitis**

T cell transfer (TCT) was performed according to established methods (40). Briefly, splenocytes were isolated from wild-type C57BL/6J animals (Jackson Laboratory) and depleted for red blood cells via treatment with ACK lysis buffer. CD4+ T cells were enriched using Dynal CD4 untouched kit (Thermo-Fisher). Naïve T cells (CD4+;CD45RBhi) and regulatory T cells (CD4+; CD25+) were isolated by FACS. 400,000 Naïve T cells or 200,000 Tregs were injected IP in PBS vehicle into C57BL/6J Rag1 null mice (Jackson Laboratory). Recipients were weighed bi-weekly. Animals treated with naïve T were sacrificed following sustained weight loss of 1.5g for one week. Treg-injected animals were sacrificed at these time points. Upon sacrifice, medial colon was formalin fixed for histology and then matched tissue was snap frozen for microarray analysis and mass spectrometry (fig. S2). All animal work was approved by the Institutional Care and Use Committees of Massachusetts General Hospital and Beth Israel Deaconess Medical Center.

**Microarray analysis**

RNA was isolated from snap frozen tissue with Qiagen RNeasy microarray tissue mini kit (Qiagen: 73304). RNA expression was quantified on Affymetrix Mouse Transcriptome 1.0 Arrays, and data were processed using the Affymetrix Expression Console software. Subsequent analysis was performed on named transcripts. Microarray data were submitted to Gene Expression Omnibus (GEO accession GSE95705).

**Protein digestion and tandem mass tag (TMT) labeling for mass spectrometry**

Excised colon tissue was re-suspended in mammalian cell lysis buffer (75mM NaCl, 50mM HEPES [pH 8.5], 10mM sodium pyrophosphate, 10mM NaF, 10mM β-glycerophosphate, 10mM sodium orthovanadate, 1mM PMSF, 3% SDS, and complete mammalian protease inhibitor tablet [Roche]). Suspensions were mixed with zirconium oxide beads (1mM diameter) and lysed on a mini bead beater (Biospec) four times for 45 seconds, cooling the sample in between. Beads
were removed, the lysate was centrifuged at 15,000xg for 5 minutes at 4°C, and insoluble debris was discarded. Dithiothreitol (DTT) was used to reduce disulfide bonds and free thiols were alkylated with iodoacetamide (IAA) as described previously (41). Reduced and alkylated proteins were then precipitated following the methanol/chloroform method as described previously (42). Precipitated proteins were reconstituted in 300µL of 1M urea in 50mM HEPES, pH 8.5. Vortexing and sonication were used to aid solubility. Proteins were then digested in a two-step process, first with 3µg endoproteinase Lys-C (Wako) for 17 hours at room temperature (RT) and then with 3µg sequencing-grade trypsin (Promega) for 6 hours at 37°C. The digest was acidified with trifluoroacetic acid (TFA). Peptides were desalted over Sep-Pak C18 solid-phase extraction (SPE) cartridges (Waters). The peptide concentration was determined using a BCA assay (Thermo Scientific) and a maximum of 50µg of peptides were aliquoted, then dried under vacuum and stored at -80°C prior for labeling with TMT reagents. Peptides were labeled with 10-plex TMT reagents (Thermo Scientific). TMT reagents were suspended in dry acetonitrile (43) at a concentration of 20µg/µL. Dried peptides were re-suspended in 30% dry ACN in 200mM HEPES, pH 8.5, and 5µL of the appropriate TMT reagent was added to the sample, which was incubated at RT for one hour. The reaction was then quenched by adding 6µl of 5% (w/v) hydroxylamine in 200mM HEPES (pH 8.5) and incubated for 15 min at RT. The solutions were acidified by adding 50µl of 1% TFA, combined into one sample, and desalted.

Basic pH reversed-phase liquid chromatography (bRPLC) sample fractionation

bRPLC was used to perform sample fractionation with concatenated fraction combining. Briefly, samples were re-suspended in 5% formic acid (FA)/5% ACN and separated over a 4.6 mm x 250 mm ZORBAX Extend C18 column (5µm, 80 Å, Agilent Technologies) on an Agilent 1260 HPLC system outfitted with a fraction collector, degasser, and variable wavelength detector. A two buffer system (Buffer A: 5% ACN, 10mM ammonium bicarbonate; Buffer B: 90% ACN, 10mM ammonium bicarbonate) was used for separation, with a 20-35% gradient of Buffer B over 60 minutes at a flow rate of 0.5 mL/minute. A total of 96 fractions were collected, which were combined in a total of 24 fractions. The combined fractions were dried under vacuum, reconstituted with 8µL of 5 % FA/5 % ACN, 3µL of which were analyzed by LC-MS2/MS3.

Phospho-peptide enrichment

For each sample, 450µg of total peptides were subjected to phospho-peptide enrichment using a 4 : 1 ratio of titanium dioxide beads : peptide (w/w). Peptides were resuspended in 2M lactic acid in 50% ACN and added to 1.8mg of titanium dioxide beads. The mixture was shaken gently for 1 hour. Beads were collected by centrifugation and washed 3 times with 2M lactic acid in 50% ACN and 3 times with 50% ACN/ 0.1% TFA. Phospho-peptides were eluted with 2 x 200µL of 50mM KH2PO4, pH 10, and acidified with 1% TFA. Eluted phospho-peptides were desalted, lyophilized, and labeled with 2µL of 10-plex TMT reagents 127n-130c as described above. The combined sample was enriched for phospho-tyrosine-containing peptides using phospho-tyrosine antibody-conjugated beads (Cell Signaling Technology) following the protocol provided by the manufacturer. Unbound peptides (phospho-serine and phospho-threonine peptides) were desalted, lyophilized, and fractionated by bRPLC using a gradient of 5-28% Buffer B. A total of 96 fractions were collected, and fractions were combined into 12 fractions. Bound peptides (phospho-tyrosine peptides) were eluted and desalted. All 13 fractions were re-suspended in 5%
ACN / 5% formic acid and analyzed on an Orbitrap Fusion mass spectrometer using LC-MS2/MS3 for identification and quantification of the phospho-peptides.

Mass spectrometry acquisition and analysis

TMT-labeled peptides were subjected to multiplexed quantitative proteomics analysis on an Orbitrap Fusion mass spectrometer (Thermo Scientific) coupled to an EASY-nLC 1000 integrated autosampler and HPLC pump system. Peptides were separated over a 100µm inner diameter microcapillary column, packed in-house with 0.5cm of Magic C4 resin (5µm, 100Å, Michrom Bioresources), followed by 0.5cm of Maccel C18 resin (3µm, 200Å, Nest Group), followed by 29cm of GP-C18 resin (1.8µm, 120Å, Sepax Technologies). Samples were eluted over 165 minutes at a flow rate of 300nL/minute over a gradient of 6-25% ACN/0.125% formic acid.

TMT-labeled peptides were identified using MS2 spectra and quantified using a MultiNotch (Simultaneous Precursor Selection, SPS) MS3 method (41, 44) in a data dependent mode. Each scan sequence began with acquisition of a full MS spectrum (MS1) acquired in the Orbitrap (m/z range: 500-1,200; resolution: 60,000; AGC target: 5x105; maximum injection time: 100ms). From this spectrum, the ten highest intensity peptide ions were subjected to MS2 analysis, where acquisition time was optimized in an automated fashion (top speed: 5 seconds). Peptides were fragmented by CID (normalized collision energy: 30%), and low resolution MS2 scans were performed in the linear ion trap (quadrupole isolation width: 0.5 Th; AGC target: 1x104; maximum injection time: 35ms). From each MS2 spectrum, the ten highest intensity fragment ions were selected for SPS MS3 analysis. Fragment ions were restricted to an m/z range of 400-2000, an m/z range of -40 to +15 around the precursor peptide ion m/z was excluded from selecting fragment ions, and “TMT” was selected for Isobaric Tag Loss exclusion settings. This group of MS2 fragment ions was further fragmented by HCD (normalized collision energy: 50%), and high resolution MS3 scans were performed in the Orbitrap (resolution: 60,000; AGC target: 5x104; maximum injection time: 250ms). When analyzing phospho-peptide samples, two MS2 spectra were acquired per peptide, a 15,000 resolution spectrum in the Orbitrap upon HCD fragmentation (normalized collision energy = 40%) and a low resolution CID-MS2 spectrum as described above. Precursor ion selection for MS3 spectra was done based on the low resolution MS2 spectral data using the top ten intensity fragment ions.

Data analysis was performed on an in-house, SEQUEST-based software platform (45, 46). RAW files were converted into the mzXML format using a modified version of ReAdW.exe. Peptide identification was performed as reported previously (47), searching against a protein sequence database containing all protein sequences in the mouse ORF database (downloaded 01/14/2014), as well as that of known contaminants. For phospho-peptide data, high and low resolution spectra were annotated in two separate searches and subsequently combined. Phosphorylation of serine, threonine, and tyrosine residues (79.966331 Da) was set as a variable modification, and up to 3 modifications were allowed.

Peptides were quantified based on the TMT reported ion intensities in the collected MS3 spectra, as reported previously (47, 48). Quantified peptides were required to have a summed signal-to-noise value greater than 386 and an isolation specificity greater than 0.75 (41). TMT intensities for all peptides assigned to a protein were summed for protein quantification. Both protein and phospho-peptide quantitative data were normalized in a two-step procedure. First, the average
intensity of each species (protein or phospho-peptide) was calculated and normalized to the median of all of these average intensities. Second, to account for any mixing errors, the intensity of each species was normalized to the ratio of the median intensity for a given TMT channel to the median of all species intensities. Total protein and phospho-protein MS datasets were deposited into the Mass Spectrometry Interactive Virtual Environment (MassIVE, accession number MSV000081198).

RT-PCR confirmation of acute phase and neutrophil genes

To confirm distant tissue expression, new mice were generated using the protocol described above. Infamed and negative control animals were sacrificed and whole liver and colon tissue were snap frozen in liquid nitrogen. In the same animals, bone marrow was collected from one femur from each animal. In brief, femurs were flushed with HBSS and passed through a 45 mm filter. Red blood cells were lysed in ACK buffer for three minutes and remaining cells were stained with APC-Cy7 CD45 (BioLegend) and Alexa-700 LY6G 1A8 (BioLegend) for ten minutes at room temperature. In addition, colon was dissociated in a collagenase solution for 1hr at 37 degrees C with agitation. Tissue was then passed through a 45 mm filter and stained with APC-Cy7 CD45 (BioLegend) and Alexa-700 LY6G 1A8 (BioLegend) as described above. Cells were gated on CD45+ and LY6G\textsuperscript{hi} population (neutrophils) and sorted directly into trizol on the BD Biosciences ARIA flow cytometer and stored at -80C until processing.

RNA was isolated from whole tissue and sorted neutrophils via trizol extraction. For whole tissue extraction, liver and colon segments were homogenized in trizol via chopping with a razor blade. Following RNA extraction, cDNA was produced using the Taqman High Capacity cDNA Reverse Transcription Kit (Applied Biosystems), and pre-amplification was performed according to manufacturers’ protocols (Applied Biosystems). For liver genes, we assayed Fga (TaqMan Mm00802584_m1) and Orm1 (TaqMan Mm00435456_g1). For neutrophils, we assayed Camp (Mm00438285_m1) and Elane (Mm01168928_g1). 18s rRNA (Mm04277571_s1) was used as a standard in each case and expression levels were determined by dCT/dCT method. For liver genes, graphs represent the ratio of liver:colon transcripts in inflamed and non-inflamed tissue. For neutrophil genes, all measurements were normalized with the lowest value set to one within each gene. Graph represents Log2 normalized expression.

Western blotting for Pak1 and Merlin

Lysates from additional inflamed and non-inflamed animals (generated using the protocol described above) were run on 4-12% Novex tris-gly gels (Invitrogen). Following transfer, membranes were probed with primary antibodies as follows: phospho-Merlin (Ser518) from Cell Signaling (13281) used at 1:1,000; phospho-Pak1/2 Ser144/Ser141 from Cell Signaling (2606) used at 1:1,000; Actin from Santa Cruz Biotechnology (sc-1616) used at 1:10,000; Gapdh used at 1:10,000 from Cell Signaling (D16H11). Secondary antibodies were anti-mouse and anti-rabbit IgG HRP-linked as appropriate (Cell Signaling 7076S, 7074S; used at 1:5,000).

Luminex analysis for p38 and Gsk3\textsuperscript{\textasciitilde}/\textasciitilde phosphorylation

BioplexPro (Bio-Rad) phospho-protein measurements were carried out according to
manufacturer’s instructions with the following kits: Gsk3\textsuperscript{\textcircled{a}/\textcircled{b}}, Ser21/Ser9 (171-V50007M), p38 MAPK Thr180/Tyr182 (171-V50014M).

\textit{Pak inhibition in vivo}

Animals that had received adoptive transfer of naïve T cells were monitored for the development of colitis with regular weight measurements. Once mice developed signs of colitis, they underwent rigid endoscopy to confirm evidence of inflammation, using a validated endoscopic scoring system for colitis (49). Once inflammation was demonstrated, mice were treated with FRAX597 (100 mg/kg QD) or vehicle by oral gavage for 7 days (30). A subset of mice underwent post-treatment endoscopy and then all mice were sacrificed at 7 days. The colons were resected and opened longitudinally; two side portions (1/5th) from the entire length were reserved, one for flow cytometry and one placed in Bioplex lysis buffer. The rest of the tissue was utilized for flow cytometry.

\textit{Flow cytometry}

Tissue was homogenized in serum-free DMEM with 2mg/ml collagenase type I C (VWR 234153-100MG) and incubated for 1hr at 37C. Following incubation, sample was strained through a 45\textmu m filter and spun down for 5 minutes at 700g. Cells were then stained with the following antibodies from BioLegend (1:300 in FACS buffer) for 10 minutes: FITC CD4 (116004) (BV-421 F4/80 (123131), BV-605 CD4 (100547), BV-510 cd11b (101245), Alexa-700 Ly6G (127622), PE/Cy7 cd11c (117317), APC/Cy7 CD45 (103116), and PE CD45RB (103308). Cells were analyzed on a 5-laser LSR II (Becton Dickinson SORP).

\textit{Unsupervised clustering}

Unsupervised hierarchical clustering was performed using the clustergram function in Matlab, using default parameters. Input data was log2 transformed RNA, MS and pMS datasets. For the purposes of visualization, RNA data was masked using the genevarfilter, genelowvalfilter, and geneentropyfilter as described:


\textit{Differential gene expression and pathway analysis}

To compare univariate differential expression in each data set, RNA, MS, and pMS data were analyzed using the Wilcoxon-Mann Whitney test with Benjamini-Hochberg False Discovery Rate Correction, \(p<0.05, q<0.25\). Pathway analysis was performed on the RNA and MS data sets using Gene Set Enrichment Analysis (GSEA). GSEA was performed with weighted log2ratio ranking and 1000 gene set permutations using the “Hallmarks Gene Sets” from MSigDb.

\textit{Co-expression network analysis}

Spearman correlations were calculated for all pairs of species represented in all three data sets. Correlation coefficients less than 0.9 or greater than -0.9 were set to 0 and all others were set to 1.
to define the undirected network adjacency matrix. The topological similarity of nodes in each data set was calculated using the two-step generalized topological overlap matrix (GTOM2) (18, 19). We then clustered the GTOM2 using k-means clustering and determined the number of GTOM2 clusters in each data set by calculating the gap statistic for 1 to 15 possible clusters. The genes were then extracted and module overlap was computed using Fisher’s exact test with Benjamini-Hochberg False Discovery Rate correction (p<0.05, q<0.25). All analysis was performed in MATLAB R2016a and networks were generated using Cytoscape 3.4.0 (50). GTOM2 visualizations and calculations were performed using the implementation described in: https://www.mathworks.com/matlabcentral/fileexchange/17668-gtom-generalized-topological-overlapping-measure

YourCrosstalker network analysis

Co-expression network modules were then analyzed using the YourCrosstalker network analysis tool (www.youromics.com) (21, 22). Given an input set of gene seeds and a protein-protein interaction network database, Crosstalker identifies subnetworks of highly connected genes as scored by a random walk with restarts at the seed genes. As the random walk reaches a steady state, seed genes are removed if they are not topologically related in the PPI network and additional genes are recruited if significantly traversed (“Crosstalkers”) indicating possible association with the seed genes. Pathway enrichment is then tested using Fisher’s exact test against a database of curated pathways and network edges are colored by pathway. The Reactome and National Cancer Institute Pathway Interaction Database pathway databases and STRING protein-protein interaction network (high confidence interactions, edge weights >0.7) were used in the analysis (23-26).

Differential regulation analysis

122 genes/proteins were selected as having greater than 2-fold up or down-regulation at the protein level and less than 2.75-fold up or down-regulation at the RNA level. Gene Ontology (GO) enrichment was performed using Gorilla (http://cbl-gorilla.cs.technion.ac.il/) with the background list as all of the shared RNA and protein species. GO terms that were enriched with a p value < 10^{-3} were deemed significant. Tissue expression of acute phase and neutrophil genes was assessed using the mouse tissue expression atlas. Each gene was normalized to 1 by the highest expression for that gene. Following normalization, the average value for all of the genes of a given category was calculated and plotted for each of the indicated tissues.

Gene set enrichment analysis of kinase substrates

Functional annotation was assessed based on prior knowledge pulled from downloads available on PhosphoSitePlus. IDed proteins were pulled from Phosphorylation_site_dataset which contains all of the published mouse phosphorylation sites found in their database; functional annotation was pulled from Regulatory_sites and we included all of the phospho-sites that had known effects on function; upset kinase information was pulled from kinase lists assembled as described below. Kinase substrate lists were pulled from PhosphoSitePlus (http://phosphosite.org). Lists were composed of known mouse, rat and human kinase substrate sites. Human and rat sites
were converted to mouse numbering. This produced a total of 348 kinases substrate sets with a maximum of 558 unique phospho-sites and a median of 6. Enrichment was performed with a rank-based test using GSEA software developed at the Broad Institute (http://www.broadinstitute.org/gsea/index.jsp). Parameters can be found along with the output of the analysis in supplemental table 4. Briefly, out of 348 kinases sets, 29 met the minimum size threshold of overlap with our dataset (4 phospho-sites). Input data was the ratio of phosphopeptides to total peptides from match samples. The ratio measurement was used because this will be more reflective of specific activation of phosphorylation. Phospho-sites were ranked by ‘Log2 ratio of classes’ and permuted 35,000 times to the gene set (because of the limited number of samples, phenotype permutation is not recommended). The kinase substrate lists have been added to the Msigdb so that they can be easily applied to the GSEA software.

**Supplementary Materials**

Fig. S1. Colonic histology from animals analyzed in the study.

Fig. S2. Schematic of experimental design.

Fig. S3. Functional annotation of phosphorylation sites for mammalian proteins.

Fig. S4. Human Pak1 and Pak2 signaling network neighborhood.

Fig. S5. Pak activation during acute colitis.

Table S1. Affymetrix microarray quantification of gene expression in individual samples.

Table S2. MS-based quantification of proteins in individual samples.

Table S3. MS-based quantification of phospho-peptides in individual samples.

Table S4. Differential expression analysis for each data set.

Table S5. Pathway analysis for each data set.

Table S6. Modules from trans-omic co-expression network analysis.

Table S7. Statistics for trans-omic co-expression network analysis.

Table S8. Phospho site lists for each kinase used in GSEA.

Table S9. Statistics for GSEA.

**References**


**Acknowledgments:** The authors would like to acknowledge Towia Libermann at the BIDMC Genomics Core and Stuart Levine at the MIT BioMicroCenter for microarray processing and instructive conversation. We would also like to thanks Marcia Haigis for helpful comments on the manuscript. **Funding:** J.L. was supported by a Research Fellowship Award (332968) from the Crohn’s & Colitis Foundation of America. S.D.S. was supported by a National Science Foundation Graduate Research Fellowship under Grant No. 1122374. K.M.H. was funded by grants from the NIH (R01-GM088827) and the DOD (W81XWH-16-1-0042). J.K. was funded by a grant from the NIH (R01-CA124495). D.A.L. and D.K.B. were funded by grants from Boehringer-Ingelheim as part of the SHINE program and from the Army Research Office Institute for Collaborative Biotechnologies (W911NF-09-0001). **Author contributions:** J.L., P.C.G., K.R.B., and L.S.L. performed mouse experiments; J.L. performed GSEA, and non-correlated RNA/MS analyses; D.K.B. performed statistical analyses and YourCrosstalker network analysis; A.E. and M.B. performed phospho-mass spectrometry; S.D.S. developed Pubmed searching algorithm and contributed to development of kinase substrate lists; Y.L. performed bone marrow isolation; J.K. provided PAK inhibitor for mouse studies; J.L., D.A.L., V.Y., J.K., W.H., and K.M.H. contributed to the conception, planning, and execution of experiments and analyses. J.L., D.K.B., and K.M.H. wrote the manuscript. **Competing interests:** The authors declare no competing financial interests.
Figure Legends

Fig. 1. Multi-omics of murine colitis. (A) Unsupervised clustering of microarray, mass spectrometry (MS), and phospho mass spectrometry (pMS) data. Non-inflamed samples are marked in blue in the dendrograms above the heat maps. Inflamed samples are marked in red. (B) Probability distributions of Spearman correlations for each pairwise comparison between –omic data sets. (C) *Lima1* (Eplin) average fold changes between inflamed and non-inflamed colons. (D) Scatter plot of phospho-Ser360 and total Eplin protein counts in individual samples. (E) Venn diagrams summarizing the unique and overlapping differential expression events between the RNA, MS, and pMS data sets. (F) Cellular localization of up-regulated phospho-signals in colitis. Left panels show expression levels in individual samples, as detected by MS, for Trim28 Ser473 (top) and Map3k3 Ser337 (bottom). ** denotes p < 0.01 and *** denotes p < 0.001 in an unpaired T test. Other panels show immunohistochemistry for phospho-Trim28 and phospho-Map3k3 in non-inflamed and inflamed colons. In all panels, N = 5 for control (non-inflamed) samples and N = 3 for inflamed samples.

Fig. 2. Differential expression and pathway analysis. (A) Heatmaps of RNA and total protein measurements most strongly contributing to pathway enrichment scores of GSEA. (B) Venn diagrams summarizing the unique and overlapping up and down-regulated pathway enrichment in the RNA and MS data sets. (C) GSEA plots from RNA and total protein MS data set. Gene lists corresponding to E2F targets and KRAS signaling are coordinately regulated in the two data sets. The gene list corresponding to EMT are oppositely regulated in the two data sets.

Fig. 3. Differential regulation of RNA and protein. (A) Scatter plot of fold change (inflamed vs. non-inflamed) RNA expression plotted against fold change in protein expression for species that were present in both data sets. Colored dots represent ECM proteins (purple), acute-phase proteins (red), and neutrophil proteins (orange), with the arrowheads indicating genes that were further investigated in panels D and E. (B) Collagen expression in RNA and protein datasets. Differential expression of MMPs and TIMPs are indicated in the heatmap inset. (C) Tissue expression patterns of acute phase and neutrophil transcripts. Each gene was normalized to a maximum of one, and all of the genes from each category averaged to generate bars. Acute phase proteins are depicted by red bars. Neutrophil proteins are depicted with orange bars. (D) Induction of acute phase protein RNA in the liver during inflammation. Bars represent the ratio of liver expression to colon expression for *Orm1* and *Fga* from N = 2 each of inflamed and non-inflamed animals. Assays were performed in duplicate. * denotes p < 0.05 in an unpaired T test. p = 0.085 for *Orm1*. (E) Loss of neutrophil gene expression by colonic neutrophils. Neutrophils isolated from the bone marrow express high levels of *Camp* and *Elane* relative to neutrophils isolated from colon. The plot represents log-transformed data from N = 2 inflamed animals, normalized to the smallest expression value for each gene. For both genes, p < 0.05 in a paired T test. (F) Model depicting RNA and protein (R and P) expression for acute phase proteins and neutrophils proteins in the colon and distant organ sites.

Fig. 4. Co-expression network landscape of RNA, MS, and pMS measurements. (A) Correlation networks for RNA, MS, and pMS data sets. Nodes indicate genes and edges connect two genes if the Spearman correlation between the expression of two genes is greater than 0.9
or less than -0.9. (B) Two-step generalized topological overlap matrices (GTOM2) of the RNA, MS, and pMS data sets clustered by unsupervised hierarchical clustering. Yellow indicates a high GTOM2 value while blue indicates a low value between 0 and 1. Square regions indicate highly connected clusters of genes. (C) Plot of the gap statistic versus the number of clusters in each data set. Clustering cutoff points are marked with a star for each data set based upon the gap statistic and GTOM2 topology. (D) Network visualization of module overlap. Nodes indicate particular modules while edges are present if significant there was overlap in genes between the two modules (Fisher Exact p<0.05, FDR q<0.25). (E) YourCrosstalker network module for pMS cluster 3. Nodes are colored by differential phosphorylation status in inflamed relative to uninflamed colons (red = hyper-phosphorylated, blue = de-phosphorylated, WMW p<0.05, FDR q<0.25) and edges are colored by pathway membership of the interaction. Pathways with higher statistical significance determine the interaction pathway association for interactions in multiple pathways. Dashed nodes were recruited by the algorithm during the random walk procedure as significantly traversed “crosstalker” nodes. (F) YourCrosstalker network for MS cluster 5. Nodes are colored by total protein differential expression status in inflamed relative to uninflamed mice (red = hyper-phosphorylated, blue = de-phosphorylated, WMW p<0.05, FDR q<0.25).

Fig. 5. Inferring Kinase Activity from pMS Measurements. (A) Volcano plot of Normalized Enrichment Score (NES) versus False Discovery Rate (FDR). Kinases with positive or negative enrichment and an FDR < 0.25 are specified. (B) Heatmaps of phospho-peptides corresponding to known kinase substrates. For each kinase, the Log2 differential expression (inflamed vs. non-inflamed) is highlighted for RNA (left box), protein (right box), and phosphorylation (circles). Normalized Enrichment Score (NES) is specified for each kinase. All of the kinases shown had FDR < 0.25 and are predicted to be either up-regulated (positive NES) or down-regulated (negative NES) in colitis. (C) Validation of p38 and Gsk3α/β phosphorylation in colitis. An activating phosphorylation event on p38 (Thr180/Tyr182) is reduced in colitis, while an inhibitory phosphorylation event on Gsk3α/β (Ser21/Ser9) in increased. Fluorescence intensity (FI) was measured using Luminex assays specific to each phosphorylation site. N = 12 for non-inflamed controls and N = 25 for inflamed samples. * denotes p < 0.05 and ** denotes p < 0.01 in an unpaired T test.

Fig. 6. Mouse model expression overlap with human IBD biopsies. (A) Venn diagrams representing the differential expression analysis of human IBD colonic biopsies in inflamed and uninflamed phenotypes (Wilcoxon Mann Whitney p<0.05, FDR q<0.25) compared to differentially expressed RNA, protein, and phospho-peptides between inflamed and uninflamed mouse colons. (B) Human IBD differentially expressed genes in the PAK signaling network neighborhood. Genes are colored by differential expression direction (red = up-regulated, blue = down-regulated) in inflamed relative to un-inflamed human colonic biopsies.

Fig. 7. Validation of Pak as a therapeutic target in colitis. (A) Validation of Pak activation in the colons of animals with inflammation. Western blotting revealed increased phosphorylation of both Pak1 and Pak2 at sites known to regulate kinase activity. (B)
Inhibition of Pak activity by FRAX597. Merlin phosphorylation on Ser518, a Pak substrate, is decreased in animals treated for 24hrs with FRAX597 (100 mg/kg single dose). (C) Colonoscopic monitoring of colitis. Colonoscopy images of a representative mouse before and after 7d of FRAX597 treatment (100 mg/kg QD). (D) Histological effects of FRAX597 treatment. Vehicle-treated animals have significant immune infiltrate in the colon, indicative of active inflammation. The inflammatory infiltrate is reduced after Pak inhibition. (E) Immunological effects of FRAX597. The number of macrophages and neutrophils is decreased in animals after 7d of FRAX597 treatment. * denotes p < 0.05 in one-tailed Mann-Whitney test. N = 3 for vehicle-treated animals and N = 4 for FRAX597-treated animals.
A microarray (39,325 named genes) A phospho MS (3,159 p-peptides) A total MS (8,131 proteins)

-2 -1 0 1 2

\( pMS \) \( \text{Counts (x10^5)} \)
\( \text{MS Counts (x10^5)} \)

\[ \text{Fold change (Log2 Inf vs. NI)} \]

\[ \text{pMS Counts (x10^5)} \]
\[ \text{MS Counts (x10^5)} \]

\[ \text{Lima1 (Eplin)} \]
\[ \text{Inflamed} \]
\[ \text{Non-inflamed} \]

\[ R = -0.91 \]

\[ \text{Probability density} \]
\[ \text{Spearman correlation} \]

B RNA/MS RNA/pMS MS/pMS RNA/MS RNA/pMS MS/pMS

C RNA/MS RNA/pMS MS/pMS

D RNA/MS RNA/pMS MS/pMS

E \( 7,611 \) species measured 5,323 differential expression events

- 1,474 species measured 1,474 differential expression events

- 1,634 species measured 1,300 differential expression events

- 1,415 species measured 1,293 differential expression events

F Trim28 Ser473

- **

- Non-inflamed

- Inflamed

anti-Trim28 Ser473

anti-Trim28 Ser473

Map3k3 Ser337

- ***

- Non-inflamed

- Inflamed

anti-Map3k3 Ser337

anti-Map3k3 Ser337
Processing of capped intron-containing pre-mRNA
p38 signaling mediated by Mapkap kinases
Signaling events mediated by Vegfr1/2
Regulation of p38-alpha and p38-beta
Signaling by VEGF
Metabolism of RNA
Cell cycle, mitotic

E

RNA
MS
pMS

RNA
MS
pMS

0.5
1
1.5
2

Gap statistic

Number of K-means clusters

D

RNA
MS
pMS

RNA
MS
pMS

RNA
MS
pMS

RNA
MS
pMS

E

F

RNA
MS
pMS

RNA
MS
pMS

RNA
MS
pMS

Processing of capped intron-containing pre-mRNA
Signaling by NGF
Signaling events mediated by Vegfr1/2
Regulation of p38-alpha and p38-beta
Signaling by VEGF
Cell-cell communication
2,708 mouse-human RNA transcript pairs

2,608 mouse-human protein-RNA pairs

867 mouse-human phospho-peptide-RNA pairs