Resource

Cell Chemical Biology

Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis

Highlights

- Cell Painting can aid phenotypic drug discovery campaigns for challenging diseases
- Our roadmap helps scientists analyze complex Cell Painting data for phenotypic screening
- Potential drugs and targets for intestinal fibrosis treatment were identified

Authors

Shan Yu, Alexandr A. Kalinin, Maria D. Paraskevopoulou, ..., Deborah Nguyen, Anne E. Carpenter, Shane R. Horman

Correspondence

shan.yu@takeda.com (S.Y.), shane.horman@takeda.com (S.R.H.)

In brief

Yu et al. established a multiparametric high-throughput screening approach incorporating inflammatory biomarker and Cell Painting methods to identify therapeutic targets and small molecules for treating intestinal fibrosis. This method can be broadly applied to drug discovery campaigns for other complex diseases.

Yu et al., 2023, Cell Chemical Biology *30*, 1–14
September 21, 2023
https://doi.org/10.1016/j.chembiol.2023.06.014

CellPress

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

Cell Chemical Biology

CellPress

Resource

Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis

Shan Yu,[1,*] Alexandr A. Kalinin,[3] Maria D. Paraskevopoulou,[2] Marco Maruggi,[1] Jie Cheng,[2] Jie Tang,[1] Ilknur Icke,[2] Yi Luo,[1] Qun Wei,[1] Dan Scheibe,[1] Joel Hunter,[1] Shantanu Singh,[3] Deborah Nguyen,[1] Anne E. Carpenter,[3] and Shane R. Horman[1,4,*]

[1]Takeda Development Center Americas, Inc., San Diego, CA 92121, USA
[2]Takeda Development Center Americas, Inc., Cambridge, MA 02142, USA
[3]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
[4]Lead contact
*Correspondence: shan.yu@takeda.com (S.Y.), shane.horman@takeda.com (S.R.H.)
https://doi.org/10.1016/j.chembiol.2023.06.014

SUMMARY

Intestinal fibrosis, often caused by inflammatory bowel disease, can lead to intestinal stenosis and obstruction, but there are no approved treatments. Drug discovery has been hindered by the lack of screenable cellular phenotypes. To address this, we used a scalable image-based morphology assay called Cell Painting, augmented with machine learning algorithms, to identify small molecules that could reverse the activated fibrotic phenotype of intestinal myofibroblasts. We then conducted a high-throughput small molecule chemogenomics screen of approximately 5,000 compounds with known targets or mechanisms, which have achieved clinical stage or approval by the FDA. By integrating morphological analyses and AI using pathologically relevant cells and disease-relevant stimuli, we identified several compounds and target classes that are potentially able to treat intestinal fibrosis. This phenotypic screening platform offers significant improvements over conventional methods for identifying a wide range of drug targets.

INTRODUCTION

Intestinal fibrosis is a pathophysiological mechanism of intestinal tissue repair that leads to the deposition of desmoplastic connective tissue after injury. This process can be triggered by noxious agents, including infections, autoimmune reactions, and physical, chemical, and mechanical injuries. Under normal physiological conditions, intestinal immune components can help to clear foreign pathogens and facilitate tissue repair through canonical wound healing processes. However, fibrogenesis may occur when the immune response is uncontrolled and persistent, or when injuries repeat, resulting in chronic damage.[1,2] Intestinal fibrosis is one of the most common complications of patients who suffer from inflammatory bowel disease (IBD), occurring in approximately 5% of ulcerative colitis (UC) patients and more than 30% of Crohn's disease patients. The prevalence of IBD increased from 0.5% in 2010 to 0.75% in 2022 in Western countries and is projected to reach 1% in 2030.[3,4] Fibrostenotic complications, including stricture formation and subsequent intestinal obstruction, significantly increase morbidity and hospitalization, surgical intervention, and health care costs.[1] Despite advances in the development of therapeutics for treating IBD, including small molecular weight immunomodulators (prednisone, 5-aminosalicylic acid, tofacitinib, and ozanimod), DNA/RNA replication inhibitors (azathioprine, methotrexate, and 6-mercaptopurine), and large molecular weight anti-inflammatory biologics (anti-TNFα, anti-integrins, and anti-IL-12/IL-23), the high incidence of intestinal strictures and requirement for surgical interventions remain.[5] The lack of effective drug therapies for fibrostenotic IBD represents an increasing and significant unmet medical need.

At a molecular basis, intestinal fibrosis in IBD is a dynamic and multifactorial process. It is a consequence of local chronic inflammation and subsequent activation of fibroblasts. Mucosal inflammation occurs when the mucosal integrity is compromised resulting in the influx of micro-organisms from the gut lumen. Myeloid cells, such as macrophages and dendritic cells, recognize these pathogen-associated molecular patterns via Toll-like and NOD-like pattern recognition receptors and propagate the immune signaling by recruiting other immune cells to clear the offending pathogens by releasing cytokines and chemokines, such as TNFα, IL-1β, IL-36, and Oncostatin-M (OSM).[6] Tissue repair and wound healing occurs in the resolution of the inflammation process after initial inflammatory responses. However, in the context of chronic inflammation, cytokines and chemokines drive the differentiation and activation of fibroblasts and their subsequent production of extracellular matrix (ECM) proteins. When the balance between production and enzymatic degradation of ECM proteins is lost, intestinal fibrosis occurs.[5] TGFβ is a key cytokine that is produced in response to inflammation, and is a well-known driver of fibrogenesis.[5,7] Numerous studies have been carried out to address TGFβ-induced

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

fibrosis.[7–10] However, due to the broad physiological functions, TGFβ inhibition induces undesirable toxicities, which override its therapeutic benefits.[11] In contrast, inflammation-associated fibroblasts (IAFs), enriched for expression of many genes associated with colitis and fibrosis, represent another paradigm in addressing IBD-related fibrosis.[12,13]

Due to the failure rate of translational efficacy for many clinical candidates for IBD,[14] there is an increased interest in the exploratory phase of drug discovery, to utilize disease-relevant phenotypic screening to provide more confidence to identify drug targets or small molecules.[15–17] However, lead molecules derived from phenotypic screening campaigns may be difficult to follow up due to intrinsic complexities of generating useful structure-activity relationships, and lack of structure-based drug design input, coupled to the difficulties in predicting and successfully navigating mechanism-associated toxicities. Chemogenomic screening utilizes a library of selective small molecules with annotated targets. The benefit of phenotypically profiling compounds with known targets and mechanisms is to assist generation of mechanistic hypotheses that can initiate ensuing target validation studies. Although focused chemogenomics libraries restrict the surveyable mechanistic space, hit molecules identified from such screens can suggest that their targets are amenable to functional pharmacological modulation, thus providing evidence of the druggability of the targets.[17]

Due to practicality and affordability, drug discovery campaigns typically employ one or a few readily interpretable biomarkers, such as secretory or intracellular markers or gene-of-interest-driven reporters that reflect known biology. Recently, significant interest has arisen in the drug discovery industry to capture high-dimensional cellular morphological changes to stimuli and drug treatments by using an image-based profiling with automated microscopy.[18] This unbiased, inexpensive, and scalable image-based method, most often using the Cell Painting assay, combines multiple organelle stains in a robust assay yielding single-cell profiles composed of thousands of features.[18] Integrated into machine learning and data mining, Cell Painting offers the potential to accelerate therapeutic discovery by identifying drug-induced cellular phenotypes, elucidating modes of action, and characterizing drug toxicities.[18]

In this study, we describe a chemogenomic library screen in human intestinal fibroblasts using both disease-relevant biomarkers and Cell Painting readouts to interrogate targeted small molecules that can alleviate the fibrotic phenotype. We identified clinically relevant hits from both assay readouts, though the mechanisms-of-action of hits from each assay represent distinct fibrotic biology. We identified inflammatory response regulators with the biomarker assay, and tissue plasticity, remodeling, fibrosis, and angiogenesis signaling modulators with the Cell Painting assay. The hits were further confirmed and validated in colonic fibroblasts treated with other pro-fibrotic stimuli. With this integrated approach using both high throughput biomarker analysis and artificial intelligence-enabled morphological profiling, we were able to discover a wide spectrum of physiologically and clinically relevant small molecules and targets for intestinal fibrosis. Typically, such high-dimensional datasets require extensive data mining and analysis with trained informatics experts to dissect the information. Here, this study serves as a general roadmap to bench scientists without machine learning skills to identify targets and hits for other complex and challenging phenotypes and polyetiological disease areas.

RESULTS

Development of an *in vitro* cellular disease model that mimics human intestinal fibrosis pathogenic cell population

The CCD-18co human colon fibroblast cell line was identified as a physiologically relevant model for human intestinal fibroblasts.[19] In order to identify culture conditions that yielded the most clinically relevant response to disease-associated stimuli, we performed single-cell transcriptomic analysis of CCD-18co cells that were treated with various pro-fibrotic stimuli, including TNFα, IL-1β, TGFβ, TL1a, OSM, and IL-36, for 16 h. We combined data from each treatment in an integrated UMAP (Figure 1A) and compared their single-cell RNA sequencing profiles side-by-side (Figure 1B). We identified seven distinct clusters of cells in total, of which several common clusters were shared among all treatments, as well as unique clusters corresponding to particular treatment groups (Figure 1B).

Within these clusters, we performed functional characterization by mapping the enriched canonical pathways and upstream regulators. Clusters 2 and 6 were predominant in TNFα and IL-1β treatment groups (Figure 1C). Genes upregulated in these clusters represented IL-17 signaling, wound healing, TREM1 signaling, cytokine-mediated fibroblast crosstalk, leukocyte migration, and tumor microenvironment pathways; as well as genes involved in mediating inflammatory pathways associated with cancer (Figure 1D). Cluster 3 and cluster 5 were mainly found in TGFβ and OSM treatments, respectively (Figure 1C). Genes upregulated in cluster 5 represented IL-6 signaling and acute phase response signaling, while genes upregulated in cluster 3 represented tissue fibrosis activities (Figure 1D). IL-36 and TL1A treatment profiles were similar to the control, suggesting neither stimulus exerted a significant effect on the cells (Figure 1B). Upstream regulator detection analysis corroborated that the clusters 2 and 6 are modulated by TNFα and IL-1β, while cluster 3 by TGFβ and cluster 5 by OSM.

To identify which CCD-18co population exhibited the most disease-mimetic gene expression profile, we mapped activated CCD-18co clusters (clusters 2, 6, 3, and 5) to cell populations from primary human colon stromal biopsies from healthy and UC patients[12] (Figure 1E). We found that clusters 2 and 6, most prevalent in TNFα and IL-1β treatments, and cluster 5, unique to OSM treatment, had signatures that closely overlapped with those of IAFs in diseased human colon biopsies. Cluster 3, specific to TGFβ treatment, corresponded to both IAFs and myofibroblasts in human colon biopsies. Because IAFs are the immunological hub of multiple signaling pathways that play important roles during the onset of intestinal inflammation and fibrosis,[7] and IAFs are associated with anti-TNFα drug resistance in IBD patients,[12] we sought to address this key unmet medical need for intestinal fibrosis and perform the primary screen with TNFα as stimulus, as it was found to induce an IAF phenotype.

To quantify the effects of TNFα signaling on morphological fibrosis in CCD-18co cells, we knocked out the *TNFRSF1A* and *TNFRSF1B* genes, which encode TNFR1 and TNFR2 (TNFα cell surface receptors), respectively, individually or together using

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

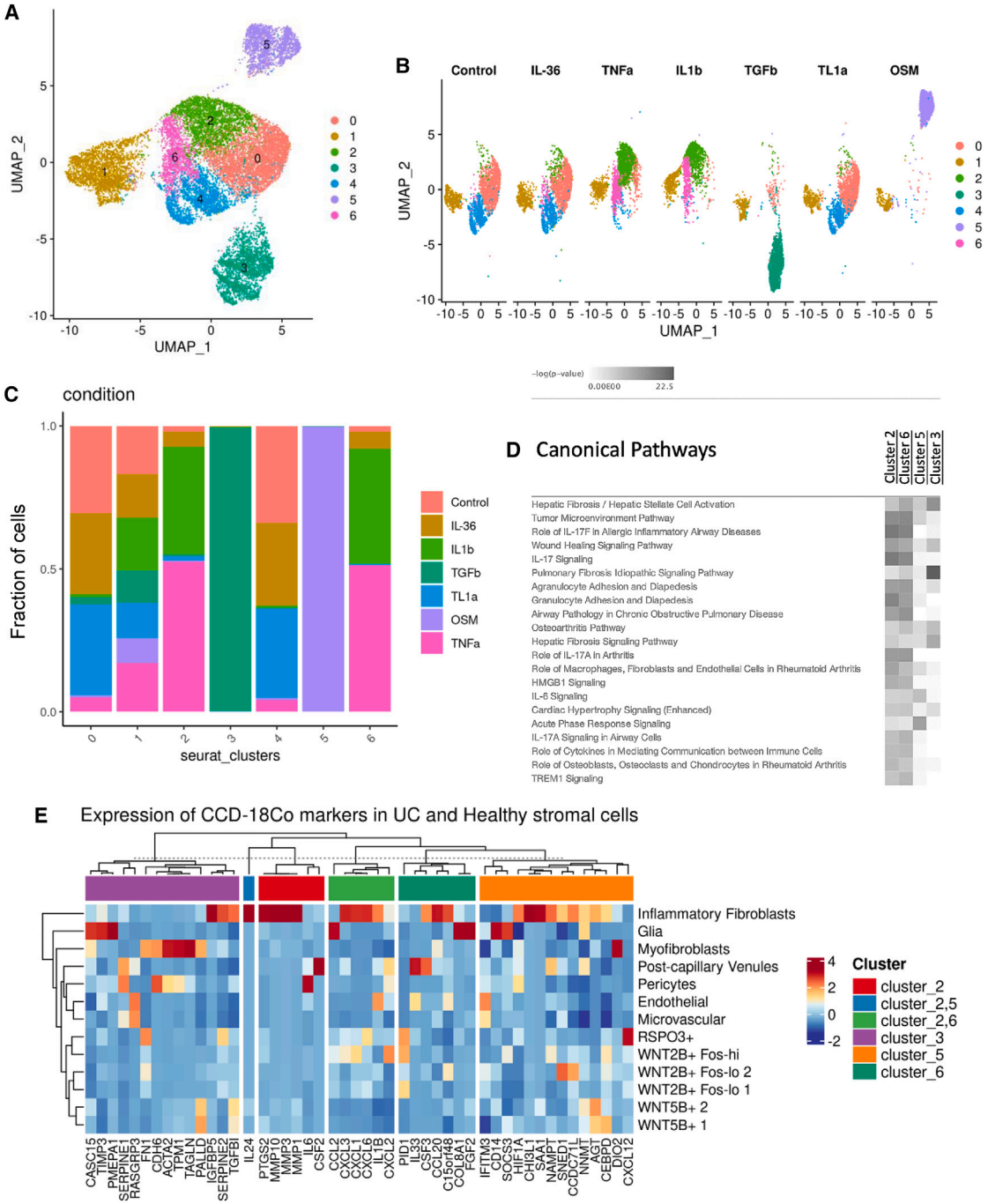


Figure 1. Bioinformatic analysis of transcriptome profile of CCD-18co cells and comparison with human colon biopsies

(A and B) UMAP embedding of 16,750 single-cell RNA sequencing (scRNA-seq) profiles from CCD-18co fibroblast cell cultures with different stimuli, including TNFα, IL-1β, TGFβ, TL1a, OSM, and IL-36, for 16 h. Seven identified single-cell clusters are indicated by colors.

(C) Stacked bar graph showed cluster cell composition, with cluster 3 corresponding to cells activated by TGFβ, cluster 5 corresponding to cells stimulated by OSM, while clusters 2 and 6 captured cellular responses upon TNFα and IL-1β treatments. The remaining clusters were not overrepresented in any of the conditions and were considered baseline state.

(D) Ingenuity pathway analysis (IPA) canonical pathways associated with the upregulated genes in clusters 2 and 6 (TNFα and IL-1β stimuli), cluster 5 (OSM), and cluster 3 (TGFβ). Cluster 1, not shown in Figure 1D, exhibited high expression of cell cycle phase genes. Grayscale represents p-score = − log10 (p value).

(E) Top 15 markers from CCD-18co fibroblast cell clusters 2, 3, 5, and 6 were analyzed in human colon fibroblasts from UC and healthy patients, retrieved from the published stromal single cell atlas.[12] Highly expressed genes in CCD-18co clusters 2, 5, and 6 (TNFα, OSM, and IL-1β treatments) were enriched in inflammatory fibroblasts, and highly expressed genes in CCD-18co cluster 3 (TGFβ cellular treatment) were elevated in myofibroblasts from colonic biopsies.

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

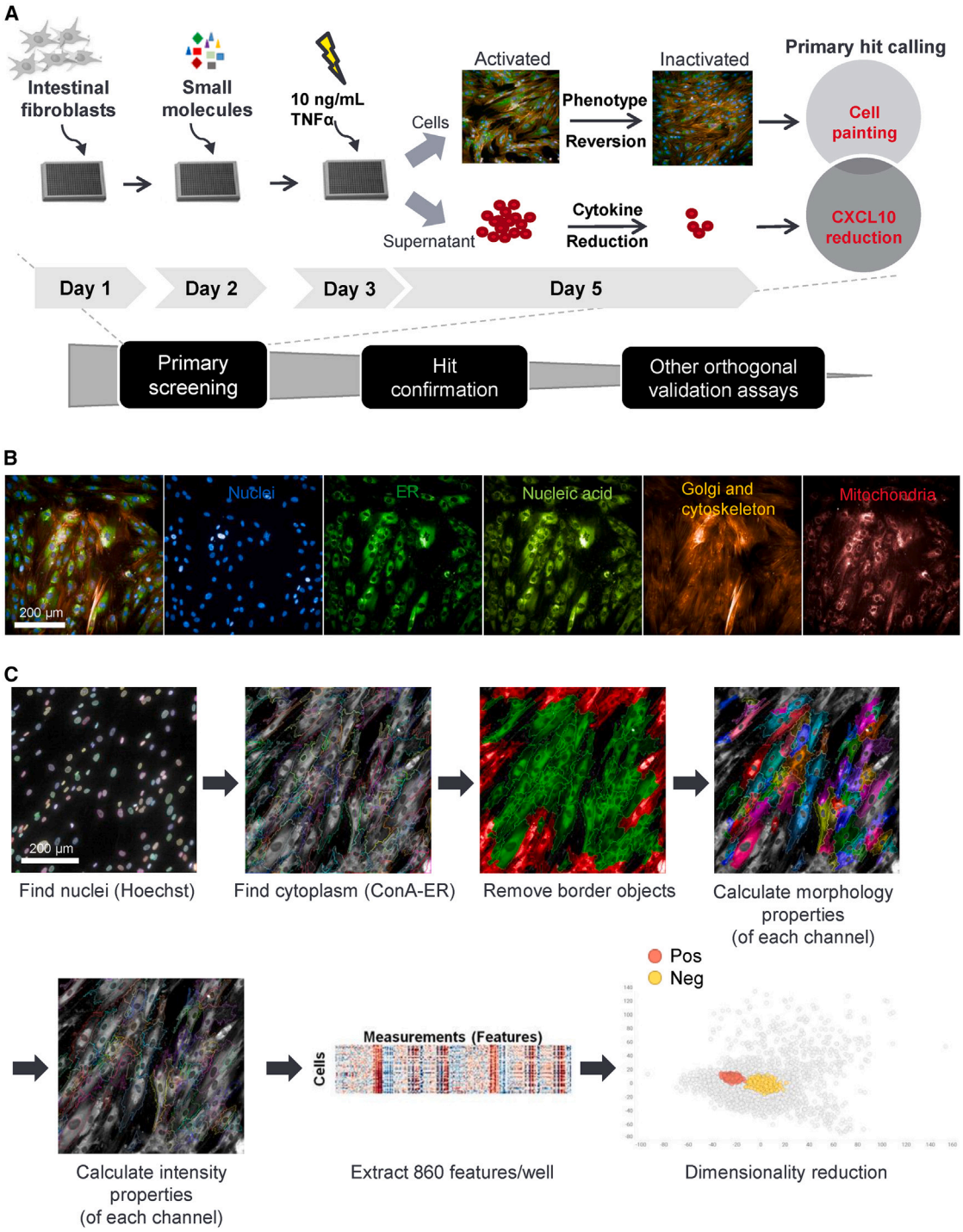


Figure 2. Chemogenomic library screen workflow

(A) The screen was conducted through a process including primary screening, hit confirmation, and orthogonal validation assays. For the primary screen, colonic fibroblasts CCD-18co cells were plated on day 1, followed by small molecule transfer on day 2, and 10 ng/mL TNFα stimulation on day 3. The supernatant samples were collected for the CXCL10 reduction assay and cells were stained with the Cell Painting dyes for the high content imaging assay. Hits from both assays were called and analyzed individually and collectively.

(B) CCD-18co cells that were stained with Cell Painting dyes including Hoechst 33342 (nuclei), Concanavalin A-Alexa 488 (ER), SYTO 14 (nucleic acid), WGA-Alexa 555 (Golgi), phalloidin-Alexa 568 (cytoskeleton) and MitoTracker Deep Red (mitochondria), and imaged with Operetta CLS. The image on the far left represents the merged image of all channels.

(legend continued on next page)

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

CRISPR/Cas9 gene editing (Figure S1A, related to Figures 1 and 2), then evaluated the response of the cells to TNFα. Upon activation of NF-kB by TNFα signaling, p65, a subunit of NF-kB also known as RELA, was observed to translocate from the cytoplasm to the nucleus (Figure S1B, related to Figures 1 and 2). However, cells transfected with individual or pooled TNFRSF1A guide RNAs (gRNAs) showed that p65 remained, at least partially, in the cytoplasm (Figure S1C, related to Figures 1 and 2), indicating reduced NF-kB signaling. Further, CCD-18co cells transfected with individual or pooled TNFRSF1A gRNAs showed a trend toward diminished CXCL10 secretion compared to control cells (Figure standard deviations (S.D.) related to Figures 1 and 2). The effect of dual TNFRSF1A and TNFRSF1B knockout was similar to TNFRSF1A knockout alone indicating TNFα signaling was mediated, at least partially, through TNFR1 instead of TNFR2 in CCD-18co cells.

In high-throughput screening, it is important to use clinically proximal readouts whenever possible to ensure the observed phenotype is a robust surrogate for disease pathology. To that end, we assessed protein and mRNA expression levels of a panel of inflammation-related biomarkers in CCD-18co cells that were treated with disease-relevant pro-fibrotic stimuli. We identified CXCL10 as a significantly upregulated biomarker at both protein and mRNA levels by multiple stimuli, including TNFα, IL-1β, and IL-36 (Figure S2, related to Figure 2). Because CXCL10 contributes to fibrosis by supporting monocyte/macrophage recruitment, angiogenesis, fibroblast collagen synthesis, myofibroblast activation and differentiation, and modulation of CXCL10 and its receptor CXCR3 has been reported to be associated with inflammatory signaling-driven fibrogenesis,[20–24] we chose it as a readout for efficacy in the ensuing screen. Though we did profile more conventional biomarkers of fibrosis, including ACTA2 and COL1A1, neither was induced by pro-fibrotic stimuli at either protein or mRNA level to yield an acceptable assay window for a high throughput screen (Figure S3, related to Figure 2). This is likely due to the fact that they are biomarkers of canonical TGFβ signaling instead of other pro-inflammatory stimuli (e.g., TNFα, IL-1β, and IL-36).

In addition to CXCL10 secretion as a readout for efficacy, we also used the Cell Painting assay to serve as a morphological readout of cellular fibrosis. Morphologies of CCD-18co cells treated with different pro-fibrotic stimuli were visually distinct (Figure S4A, related to Figure 2) and this translated to cellular features that yielded equally distinct principal component analysis (PCA) plots (Figure S4B, related to Figure 2). Interestingly, the Cell Painting PCA plot strongly resembled the transcriptomic PCA plot (Figure S4C, related to Figures 1 and 2), suggesting CCD-18co cellular morphology might be tightly correlated with gene expression and subsequent biological activities.

Automated high throughput chemogenomic library screen to identify targeted perturbagens of intestinal fibrosis

To comprehensively profile diverse biological and functional space (Figure 2A), we sourced two small molecule libraries totaling 4,871 compounds annotated with either their reported targets and/or mechanisms of action and have been either tested in clinical trials or approved by the FDA (Selleckchem; Figure S5A, related to Figure 2). The molecular weight and ALogP of these compounds were within the standard range for "drug-like" molecules (Figure S5B, related to Figure 2).[25]

For the primary high throughput screening assay, 1,200 CCD-18co cells/well were plated on the first day, followed by compounds and controls after 24 h (Figure S5C, related to Figure 2). Each compound was tested at 3 μM in biological triplicate. 1 ng/mL anti-TNFα antibody adalimumab was used as the positive control, because adalimumab was able to effectively suppress TNFα signaling in the CXCL10 assay (as well as in the Cell Painting assay, as discussed later, Figure S.D. related to Figures 2 and 3). Cells were then treated with 10 ng/mL TNFα on the third day for 48 h, after which time the cell culture supernatants were collected for CXCL10 protein quantitation using a homogeneous time-resolved fluorescence (HTRF) assay.

For the Cell Painting assay, cells from the exact same samples were stained with Cell Painting dyes followed by high-content image acquisition and analysis. The assay includes six fluorescent dyes to highlight different organelles of CCD-18co cells, including MitoTracker Deep Red FM for mitochondria, Concanavalin A-Alexa 488 for endoplasmic reticulum, SYTO 14 for nucleoli and cytoplasmic RNA, WGA-Alexa 555 and phalloidin-Alexa 568 for F-actin cytoskeleton, Golgi, and plasma membrane, Hoechst 33342 for nucleus[26] (Figure 2B). High-content images were captured and cellular morphological features were extracted and then analyzed using a dimensionality reduction method. Compounds that clustered around the positive controls were categorized as Cell Painting hits (Figure 2C). For dimensionality reduction, we used either a supervised PCA or a linear predictive model. For both methods, the medians of positive controls and negative controls were normalized to 0 and 1, respectively. Compounds were then binned into positive or negative bins depending on the projection scores (STAR Methods, Figure 3A, left). Compounds positive for >2 out of 3 replicates in the positive bins and projection scores within the range of Average(pos ctrl) ± 3 x S.D. were picked as preliminary hits. Compounds exhibiting cytotoxic profiles were then further filtered based on cell count.

In total, 160 and 152 compounds were picked as hits from supervised PCA and linear predictive models of the Cell Painting data, respectively (Figure 3A, right). There were 100 hits that overlapped between both models for Cell Painting analysis (Figure 3A, right), suggesting the two analytical methods yielded mainly convergent results. In addition, we assessed three other metrics for picking Cell Painting hits; namely using the top 50 features, top 5 features, or top 3 features per channel that separate positive and negative controls, though the hits and targets that were identified were mostly similar (Figure S6, related to Figure 4). To determine whether these cellular features correlate with their biological functions, we projected the cellular features of the targets that were most distinct from the negative controls onto a two-dimensional t-SNE map (Figure 3B). This

(C) Workflow of cellular compartment segmentation of high content images using PerkinElmer Harmony software. Nuclei were identified by Hoechst 33342 stain. Cytoplasm was then identified by Concanavalin A-Alexa 488 stain. The border objects were excluded from analysis. Different morphology and intensity properties of each channel were calculated and 860 features were extracted at the well-level. The profiling dataset was then analyzed with a dimensionality reduction method, such as PCA.

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

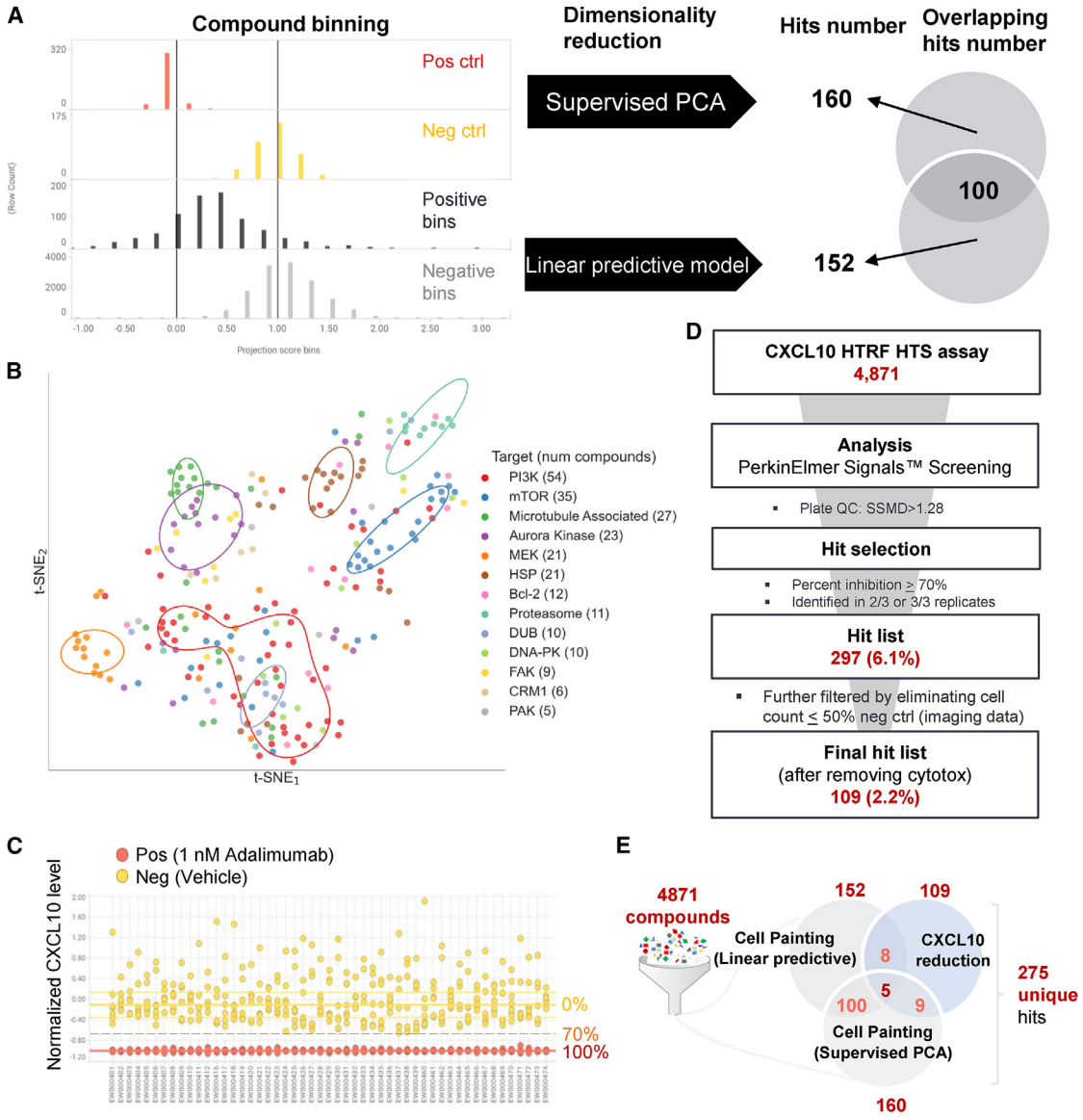


Figure 3. Primary screen hit picking strategies for the CXCL10 reduction assay and Cell Painting assay

(A) The Cell Painting dataset was analyzed with both supervised PCA and linear predictive model methods. Projection scores of Cell Painting controls and samples help to determine the similarities between compounds and controls. Compounds in positive bins in the range between Average(projection score) ± 3 x S.D. were picked as hits.

(B) t-SNE plot shows the phenotypic space of top compound target categories that are farthest from the negative controls.

(C) Pos and neg ctrl data points of CXCL10 HTRF assay. X axis shows the plate barcode, y axis shows the normalized CXCL10 level. Solid yellow line shows 0% inhibition representing the median of the neg ctrl (vehicle), and solid red line shows 100% inhibition representing the median of pos ctrl (1 nM adalimumab). Dotted orange line shows 70% cutoff for hit picking.

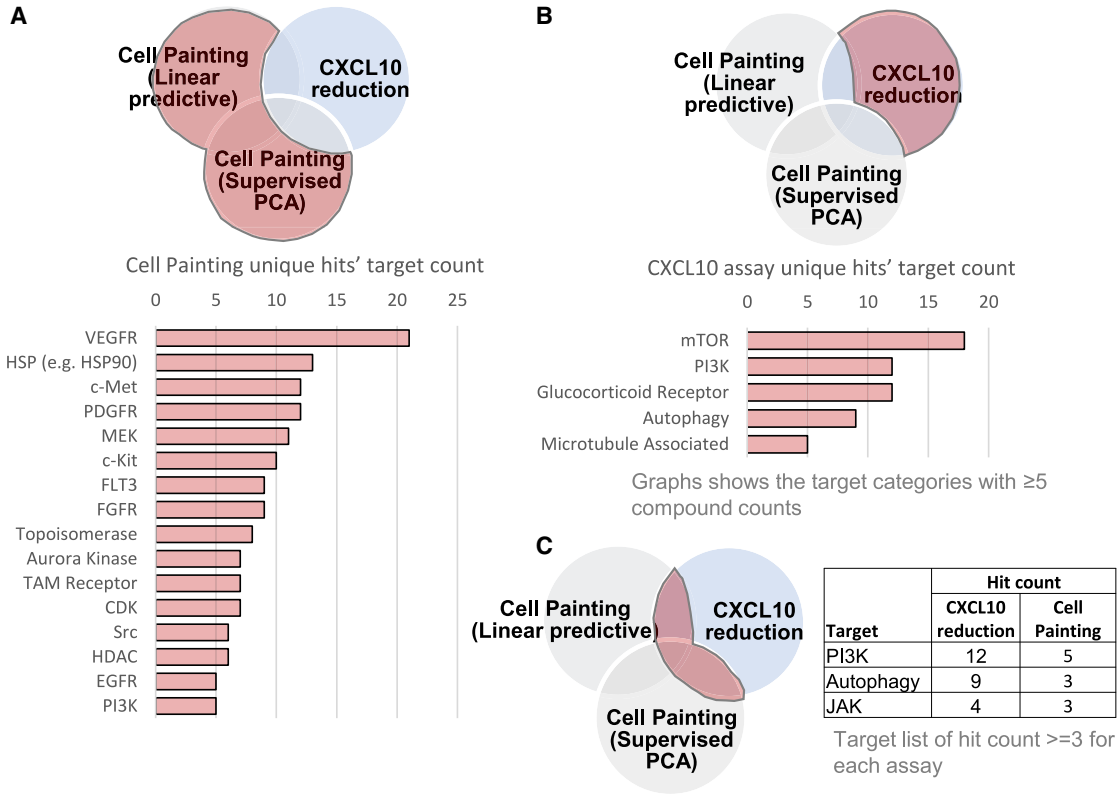
(D) The CXCL10 HTRF assay screening funnel.

(E) Overview of small molecule hit numbers from each assay/analysis.

map showed that some co-annotated compounds form coherent clusters (e.g., MEK and HSP) in phenotypic space whereas others do not (e.g., Bcl-2, FAK, CRM1, and DNA-PK).

For the CXCL10 assay, luminescence intensities of positive and negative controls of each plate were fit on a 0 to 1 scale and were then normalized for their percent inhibition, with the mean of positive control being 100% and the mean of negative control being 0% (Figure 3C). The strictly standardized mean difference (SSMD) was used to measure the effect size and gauge the assay quality.[27] Plates with SSMD>1.28 (the SSMD quality cutoff) then proceeded to hit selection. Compounds positive for >2 out of 3 replicates with CXCL10 inhibition >70% were identified as preliminary hits, and then filtered by eliminating cytotoxic compounds (dependent on cell count). After applying this gating strategy, 109 compounds were identified, resulting in a 2.2% hit rate for the CXCL10 screen (Figure 3D).

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014



Target	Hit count	
	CXCL10 reduction	Cell Painting
PI3K	12	5
Autophagy	9	3
JAK	4	3

Target list of hit count >=3 for each assay

Figure 4. Hit category analysis of Cell Painting and CXCL10 reduction assays

(A) Hit number of target categories for linear predictive model and supervised PCA analysis of Cell Painting. Bar chart shows the target categories with ≥5 compounds in each.

(B) Hit number of target categories for CXCL10 reduction assay. Bar chart shows the target categories with ≥5 compounds in each.

(C) Hit number of target categories for the overlapping hits between CXCL10 reduction assay and Cell Painting assays. Table shows target categories with ≥3 compounds in each assay.

Surprisingly, there were only 8 hits that overlapped between the Cell Painting linear predictive model and the CXCL10 assay, and only 9 hits that overlapped between the Cell Painting supervised PCA model and the CXCL10 assay. Only 5 hits overlapped among all three methods. In the end, after removing duplicate compounds, 275 unique hits from either Cell Painting or CXCL10 assay were advanced for further confirmation and validation (Figure 3E).

Target discovery through the integration of cytokine biomarker and morphological profiles

It was intriguing that the CXCL10 assay and the Cell Painting assay identified vastly different pools of hit compounds. For hit compounds that were unique to Cell Painting, the top targets included VEGFR, HSP, c-Met and PDGFR, MEK, c-Kit, FLT3, and FGFR (Figure 4A); while hits that were unique to the CXCL10 assay included the targets mTOR, PI3K, glucocorticoid receptor, and several components of the autophagy and microtubule pathways (Figure 4B). For hits that were shared between the two assays, the top targets included PI3K, autophagy, and Janus kinase (JAK) (Figure 4C).

In several contexts, image-based profiles have proven to show predictive abilities for other assays.[28] We wondered whether any particular cellular morphology features from the Cell Painting assay could be used to predict CCD-18co cells' response to TNFα, in terms of secreting CXCL10. We studied the statistical dependence between CXCL10 levels and each of the 860 individual cellular features. Overall, 752 out of 860 features had some linear relationship with the CXCL10 level (F-test, $p < 0.01$, Bonferroni-corrected with $\alpha = 0.01$). In particular, we found that a few categories of cellular features including axial small length (the length of the cell's shorter axis in pixel units) and Radial Mean (the mean object radius based on the intensity values weighted by the distance from the mass center) from the ER, mitochondria and F-actin, Golgi and PM channels (n = 54 features) had strong relationships with CXCL10, as indicated by higher average F statistic values (97th percentile of distribution of F statistic, all adjusted p values = 0.0), which capture the linear dependency between features and the CXCL10 (Figure 5A), We further confirmed this finding by also calculating averaged mutual information (MI), which is a nonparametric measure that can capture any kind of statistical dependency, and demonstrated that these feature categories have strongest relationships with the CXCL10 level (98th percentile of distribution of MI values) (Figure 5B). To focus on the subcategories and examine which particular features had the strongest statistical dependency with CXCL10 level, we found that several Radial Mean features including Edge, Ridge, and Spot of the

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

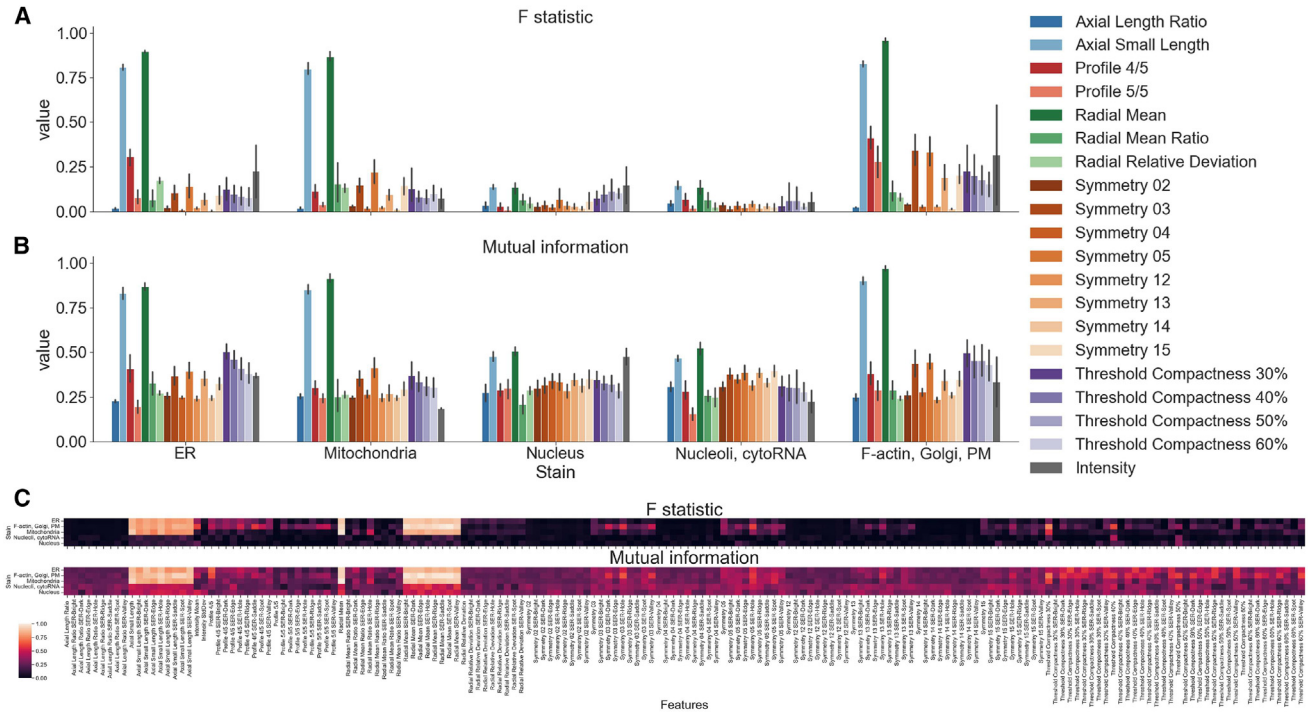


Figure 5. Correlation analysis of morphological features with CXCL10 level

(A and B) F statistic (which shows the linear dependency) and mutual information (which shows any type of dependency, including linear dependency) between cellular feature subcategories and the CXCL10 level. Error bars show a bootstrap-estimated 95% confidence interval.

(C) Heatmap of top highly correlated features of each subcategory with CXCL10.

Spots, Edges and Ridges (SER) texture analysis in the F-actin, Golgi, and plasma membrane channel have nearly perfect statistical dependency with the CXCL10 level (e.g., Radial Mean SER-Spot has F-statistic of 1.0 and Radial Mean SER-Edge has MI of 1.0) (Figure 5C); indicating these features have strong dependency with CXCL10 and can be considered as potential predictors of CXCL10 level.

Target validation using pro-fibrotic stimuli-treated cell models

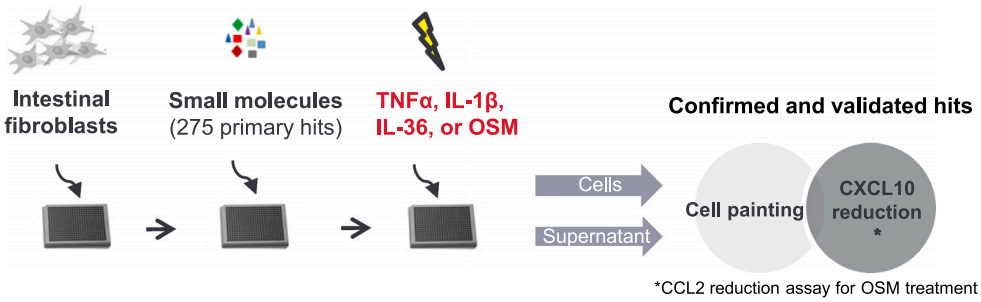
To further characterize hit compounds according to their ability to ameliorate fibrosis from pro-fibrotic stimuli other than TNFα, we profiled the 275 unique hit compounds at three doses (3 μM, 0.6 μM, and 0.125 μM) in assays with different stimuli (IL-1β, IL-36, or OSM) in addition to TNFα (Figure 6A). The CXCL10 assay was used for TNFα-, IL-1β-, and IL-36-treated cells, while a CCL2 assay conducted 2 h post-OSM treatment was used for OSM-treated cells, because CCL2 (Figure S.D. related to Figure 6) but not CXCL10 (Figure S.D. related to Figure 6) is a functional biomarker for OSM stimulation. Similar to CXCL10, CCL2 contributes to fibrosis by recruiting monocyte/macrophage and myofibroblast activation and differentiation.[23] The Cell Painting assay was only used for TNFα and IL-1β stimulation, as there were no viable assay windows for cells treated with either IL-36 or OSM (Figure S4B, related to Figure 6), leaving four cytokine assay and two Cell Painting assay results available for analysis.

The TNFα-stimulated reconfirmation screen of 275 unique hit compounds yielded a 51% reconfirmation rate for reducing CXCL10 expression/secretion and a 47% reconfirmation rate for Cell Painting, suggesting the robustness of the primary screening assays (confirmed and validated hit results are shown in Table S1, and details of example hits are shown in Figure S7, related to Figure 6). Using a combinatory approach to examine the target categories, we pooled the four cytokine stimulation results and identified glucocorticoid receptor as the top target with 16 hits. This was followed by autophagy, inflammatory-related mechanisms, JAK, PDGFR, and SYK (Figure 6B). The two Cell Painting reconfirmation assays (TNFα and IL-1β) similarly showed glucocorticoid receptor to be the top target, followed by MEK, PDGFR, VEGFR, and inflammatory-related mechanisms (Figure 6C).

When considering all six compound lists, the hits were binned into three buckets depending on the number of assays in which they were identified as hits. Bucket one included compounds that were picked as hits in six out of six assays. All hits in this bucket were glucocorticoid receptor modulators (steroids). Bucket two included compounds that were picked as hits in five out of six assays and similarly, all hits in bucket two were mainly glucocorticoid receptor modulators. Bucket three included compounds that were identified as hits in three or four out of six assays and this bucket represented the largest variety of biological functions with different mechanisms of action (Figure 6D).

To understand these targets in the context of signaling pathways, we mined the literature and identified any associations between targets in bucket three and intestinal fibrosis. Overall, three main pathways were identified: ER stress response,

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014



B

4 cytokine assays	
Target	Hit count
Glucocorticoid Receptor	16
Autophagy	4
Inflammatory related	4
JAK	3
PDGFR	3
Syk	3

C

2 Cell Painting assays	
Target	Hit count
Glucocorticoid Receptor	16
MEK	6
PDGFR	5
VEGFR	4
Inflammatory related	3

D

4 cytokine reduction assays + 2 Cell Painting assays			
	Target	Hit count	Ref.
Bucket 1 (6/6)	Glucocorticoid Receptor	15	
Bucket 2 (5/6)	Glucocorticoid Receptor	5	33-35
Bucket 3 (4/6 or 3/6)	c-Met	9	36-38
	VEGFR	9	29, 32
	PI3K	8	39-43
	TAM Receptor	5	44, 45
	mTOR	5	39, 40, 42
	Src	4	46-48
	EGFR	4	31, 38, 49, 50
	CDK	4	51, 52
	MEK	3	41, 53
	Others	3	
	ALK	3	54
	c-Kit	3	55, 56
	Akt	3	39, 42, 43
	LPA Receptor	3	57, 58
	Dehydrogenase	3	59, 60
	PDGFR	3	32, 61
	FGFR	3	30, 62

Figure 6. Hit confirmation and validation assay workflow and hit categories

(A) Hit confirmation and validation experimental workflow.

(B) Top target categories across the four cytokine reduction assays. Table shows target categories with ≥3 compounds for each assay.

(C) Top target categories for the Cell Painting results of TNFα and IL-1β stimulation.

(D) Top target categories for all six assay results. The results were further bucketed into three categories. Bucket 1 includes compounds that showed effects in all 6 assays. Bucket 2 includes compounds that showed effects in 5 out of 6 assays. Bucket 3 includes compounds that showed effects in 3 or 4 out of 6 assays. Table only shows target categories with ≥3 compounds in each.[29–62]

fibrosis/angiogenesis, and inflammation (Figure 7). All three pathways were shown to play a role in tissue fibrosis.[7,9,29–31,63–66] Interestingly, we identified and confirmed both nintedanib (targets PDGFR, VEGFR, and FGFR) and pirfenidone (targets NF-kB), approved drugs for treating idiopathic pulmonary fibrosis (IPF),[32,66] as potent antagonists of myofibroblast activation[67] (Figure 7). These data suggest that the small molecules, targets, and signaling pathways identified through our multi-parametric biomarker and cellular feature profiling approach were physiologically and clinically relevant. Further, this screening platform was able to identify molecules from a wide spectrum of mechanisms of action.

DISCUSSION

IBD-associated intestinal fibrosis represents a highly invasive and deleterious disease that currently has no approved pharmacological intervention. In order to address this, we developed a clinically relevant humanized intestinal fibrosis model composed of TNFα-activated colon fibroblasts. In order to leverage large collections of small molecules for therapeutic profiling efforts, we miniaturized the human IBD fibrosis model to accommodate a scalable phenotypic screening platform for fully automated drug discovery. Employing transcriptomics as a surrogate characteristic for comparing our CCD-18co *in vitro* model to IBD patient biopsies, we identified several distinct transcriptional clusters corresponding to different pro-inflammatory cytokine stimuli. Although TGFβ treatment of CCD-18co cells produced a canonical gene expression profile that overlapped with myofibroblast components of patient biopsies, discovery of therapeutics targeting the TGFβ pathway has not yielded any clinical treatment due to undesirable toxicities. In recent years, IAFs have been shown to be critical to fibrogenesis associated with chronic inflammatory diseases.[12,13,22] Here, we intended to identify potential therapeutics by targeting IAFs.

As intestinal fibrosis is a result of a complex interplay of immune-mediated inflammatory processes as well as modulation of pro-inflammatory cytokine-mediated signaling pathways,

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

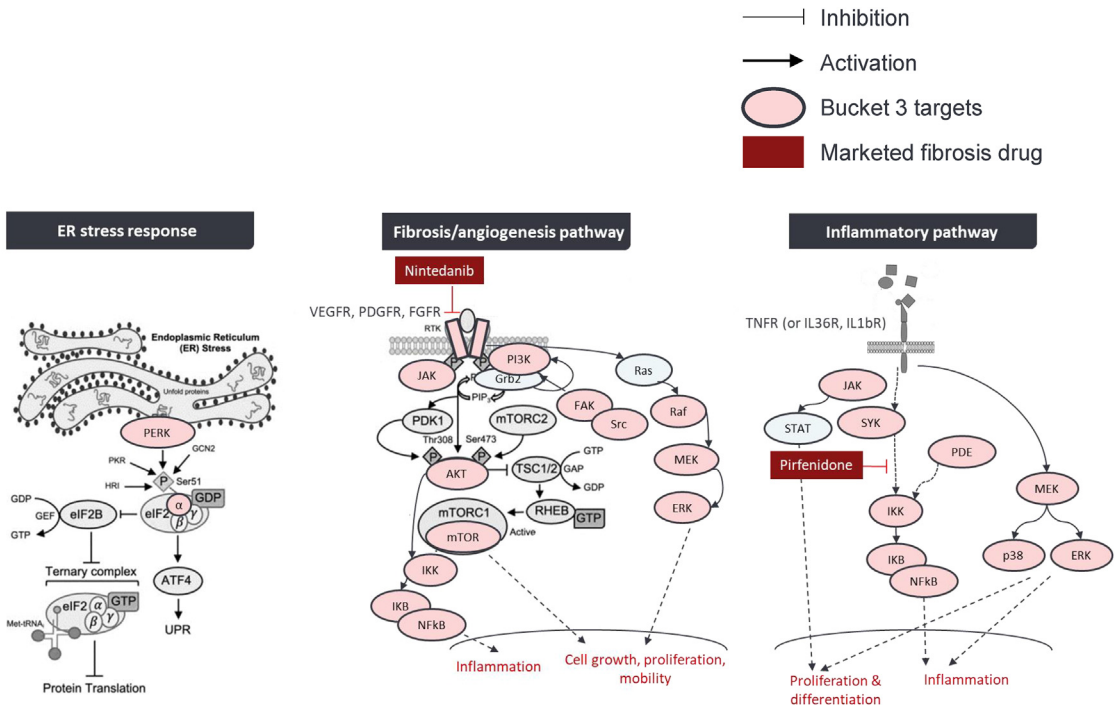


Figure 7. Major pathways of the bucket 3 compound targets

Three major pathways, including ER stress response, fibrosis/angiogenesis pathway, and inflammatory pathway were identified by analyzing the targets of bucket 3 compounds. Pink bubbles show the targets that were identified in the bucket 3 compounds. Gray bubbles show other intermediate targets in the pathway. Nintedanib, a marketed drug for idiopathic pulmonary fibrosis, was identified as a hit in the screen. The screen also identified inflammatory pathway targets through which pirfenidone, another marketed drug for idiopathic pulmonary fibrosis, exerts its effect.

our screening platform required a sophisticated series of assay readouts to account for these polyetiological causes. We first chose to use CXCL10 (IP10) as the primary screen readout due to its well-characterized association with intestinal fibrotic pathology and because compared to other biomarkers, both its mRNA and protein levels were significantly increased by multiple pro-fibrotic stimuli (Figure S2, related to Figure 2). However, to fully assess changes in the fibrotic morphological phenotype, we applied an unbiased image-based profiling technique called Cell Painting. Although Cell Painting has not been widely adapted in the drug discovery industry as a phenotypic readout for efficacy, its scalable ease of use as well as its ability to quantitate changes in thousands of cellular features makes it an ideal method for studying complex biology such as intestinal fibrosis. Cell Painting produces vast morphological information as a collection of extracted cellular features, but by integrating artificial intelligence analytical methods, such as machine learning, we can mine these data to reveal important biological activities of potentially therapeutic small molecules.[18] For example, we found that the relative positions of pro-fibrotic stimuli-treated clusters to vehicle controls in Cell Painting PCA plots were similar to those from RNA-seq PCA plots, suggesting transcriptome profiles and related biological activities strongly correlate with cellular morphological profiles. We also examined whether any specific cellular features were highly correlated with CXCL10 level, because these features may potentially be used as sentinel readouts for CXCL10 in future studies. We identified several subcategories of features, such as Axial Small Length and Radial Mean in ER, mitochondria and F-actin, Golgi and plasma membrane channels that had high correlations with CXCL10 level (Figure 5).

Surprisingly, we observed divergent hit distribution profiles between the CXCL10 and Cell Painting assay readouts. The reason might be attributed to the fact that only a few cellular features from the Cell Painting assay had a strong statistical correlation to the CXCL10 level (Figure 5). Different cellular features were chosen that better represented the TNFα-stimulated phenotype though they had a lower correlative relationship with CXCL10. These features were chosen for Cell Painting hit selection because they were more prominent in differentiating TNFα-treated and non-treated cells. While the CXCL10 readout identified well-characterized regulators of fibrosis such as mTOR and glucocorticoid receptor, the targets identified through the Cell Painting readout were mechanistically more diverse (e.g., VEGFR, PDGFR, FGFR, c-Met, c-Kit, and MEK) and included such cellular processes as fibrosis, tissue plasticity and remodeling, and angiogenesis. In short, the CXCL10 assay conferred a confidence metric to the biological relevance of our assay platform by identifying several steroid molecules as alleviators of the fibrotic phenotype. However, the Cell Painting assay was able to reveal a diverse array of potential mediators implicated in intestinal fibrosis pathology, expanding the scope of actionable targets. Overall, this high-throughput screening platform combining CXCL10 and Cell Painting readouts was able to identify small molecule hits with proven clinical relevance. For example, our screen identified and confirmed nintedanib, a

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

drug for treating IPF, may be repurposed to treat intestinal fibrosis. We also identified small molecules that modulate other known fibrosis targets (Figure 6). This suggests the screening platform may be used for repurposing approved or clinical-stage drugs or discovering novel small molecules for intestinal fibrosis.

As our collective understanding of the causes and mediators of disease biology increase, so must our ability to interrogate those causes to discover the next generation of small molecule therapeutics. A complex image-based profiling technique like Cell Painting integrated with state-of-the-art machine learning algorithms to translate thousands of cellular features into disease-relevant targets and pathways may represent a giant leap forward in industrialized drug discovery. Although it may be unlikely that image-based profiling will completely replace conventional biochemical, transcriptional, or proteomic profiling methods, when incorporated into exploratory phases of the drug discovery pipeline, Cell Painting may accelerate the identification of novel therapeutics and expand the targeting space of polyetiological and poorly understood diseases like intestinal fibrosis.

Limitations of the study

In this study, we utilized CXCL10 and CCL2 as functional readouts for CCD-18co cells due to their robust response to pro-fibrotic stimuli, resulting in an up-regulation of mRNA and protein expression levels (Figure S2B, related to Figure 2). The evidence suggests that CXCL10 and CCL2 play a role in fibrosis by supporting monocyte/macrophage inflammatory response, angiogenesis, fibroblast collagen synthesis, myofibroblast differentiation, and fibroblast recruitment and survival.[23,24] However, it should be noted that the role of CCL2 in fibrosis is somewhat controversial, as there have been reports of CCL2 mediating anti-fibrotic effects in human fibroblasts independently of CCR2.[68] Because of the complexity of intestinal fibrosis and translatability and feasibility of using other validated biomarkers in the cellular screening system, we selected CXCL10 and CCL2 as functional readouts in our screen.

SIGNIFICANCE

Our study showed that the integration of Cell Painting morphological profiling with biomarker analysis can be used to identify potential targets and small molecule drugs for a broad spectrum of polyetiological and poorly understood diseases, such as intestinal fibrosis. Here, we provide a roadmap for bench scientists without sophisticated informatics tools or machine learning skills to analyze high dimensional Cell Painting datasets and incorporate image-based profiling into an industrial phenotypic high throughput screening campaign.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - CCD-18co cell line
  - CCD-18co cell line maintenance and stimulation
- METHOD DETAILS
  - Single Cell RNA-seq
  - Analysis of CCD-18Co scRNA-Seq data
  - High throughput screening assay
  - Cell Painting high content imaging, feature extraction and high dimensionality data analysis
  - Cell Painting data projection score calculation
  - CCD-18co cell hit confirmation and validation hit picking strategy
  - Visualization of primary screen Cell Painting dataset with t-SNE
  - CXCL10 and cellular feature statistical dependence analysis
  - CRISPR/Cas9, mismatch detection and T7E1 assays
  - Olink Target 96 inflammation assay
  - RNA-seq
  - Pro-fibrotic biomarker detection with Luminex®
  - Immunofluorescence detection for ACTA2 and COL1A1
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chembiol.2023.06.014.

ACKNOWLEDGMENTS

We would like to thank our Takeda Development Center of Americas' colleagues Darren Ruane and Christopher Haines for suggesting the pro-fibrotic stimuli to be used for this study, Tony Li for CCD-18co cell culture for RNA-seq experiment, and Narender Gavva for the general supervision of the Early Target Discovery/Core Biology group and administrative support. We would also like to thank Joseph Z. Chen from California Institute for Biomedical Research, a division of Scripps Research for designing the graphic abstract.

Authors from the Broad Institute of Harvard and MIT were funded by a grant from Takeda Development Center Americas for part of this work, as well as by the National Institutes of Health (R35 GM122547 to A.E.C.).

AUTHOR CONTRIBUTIONS

S.Y., S.R.H., and D.N. concepted and designed the project; S.Y., A.K., M.P., M.M., and S.R.H. contributed to the result interpretation and drafted the manuscript; S.Y. performed assay development and high throughput screen experiments; Y.L. participated in high throughput screen experiments; S.Y., Y.L., and Q.W. performed CRISPR/Cas9 experiment; M.M. designed, performed, and analyzed scRNA-seq experiment; M.P. performed transcriptomic analysis of CCD-18co datasets and public human colon biopsies datasets; I.I. performed PCA on pro-fibrotic stimuli treated CCD-18co Cell Painting data; J.T. performed PCA on CCD-18co RNA-seq data; J.C. performed Cell Painting primary hit picking with linear predictive model; D.S. and J.H. assisted with laboratory automation and compound management; S.S. and A.E.C. provided input on morphological profiling analysis and its result interpretation, and edited the manuscript.

DECLARATION OF INTERESTS

We declare competing interests. The authors who were affiliated with Takeda Development Center Americas were employees of Takeda Pharmaceuticals during the course of this work, and have real or potential ownership interest in Takeda. AEC serves on the Scientific Advisory Board of, and has ownership interest in, Recursion, a pharmaceutical company using image-based profiling

for drug discovery. Authors from the Broad Institute of Harvard and MIT were funded by a grant from Takeda for part of this work.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: July 24, 2022
Revised: March 11, 2023
Accepted: June 13, 2023
Published: July 11, 2023

REFERENCES

1. Lenti, M.V., and Di Sabatino, A. (2019). Intestinal fibrosis. Mol. Aspects Med. *65*, 100–109.
2. Pariente, B., Cosnes, J., Danese, S., Sandborn, W.J., Lewin, M., Fletcher, J.G., Chowers, Y., D'Haens, G., Feagan, B.G., Hibi, T., et al. (2011). Development of the Crohn's disease digestive damage score, the Lemann score. Inflamm. Bowel Dis. *17*, 1415–1422.
3. Rieder, F., and Fiocchi, C. (2009). Intestinal fibrosis in IBD–a dynamic, multifactorial process. Nat. Rev. Gastroenterol. Hepatol. *6*, 228–235.
4. Kaplan, G.G., and Windsor, J.W. (2021). The four epidemiological stages in the global evolution of inflammatory bowel disease. Nat. Rev. Gastroenterol. Hepatol. *18*, 56–66.
5. Abdulla, M., and Chew, T.S. (2021). Molecular targets and the use of biologics in the management of small bowel fibrosis in inflammatory bowel disease. Curr. Opin. Gastroenterol. *37*, 275–283.
6. Friedrich, M., Pohin, M., and Powrie, F. (2019). Cytokine Networks in the Pathophysiology of Inflammatory Bowel Disease. Immunity *50*, 992–1006.
7. Hayashi, Y., and Nakase, H. (2022). The Molecular Mechanisms of Intestinal Inflammation and Fibrosis in Crohn's Disease. Front. Physiol. *13*, 845078.
8. Walton, K.L., Johnson, K.E., and Harrison, C.A. (2017). Targeting TGF-beta Mediated SMAD Signaling for the Prevention of Fibrosis. Front. Pharmacol. *8*, 461.
9. Kim, M.H., Jung, S.Y., Song, K.H., Park, J.I., Ahn, J., Kim, E.H., Park, J.K., Hwang, S.G., Woo, H.J., and Song, J.Y. (2020). A new FGFR inhibitor disrupts the TGF-beta1-induced fibrotic process. J. Cell Mol. Med. *24*, 830–840.
10. Yu, S., Ericson, M., Fanjul, A., Erion, D.M., Paraskevopoulou, M., Smith, E.N., Cole, B., Feaver, R., Holub, C., Gavva, N., et al. (2022). Genome-wide CRISPR Screening to Identify Drivers of TGF-beta-Induced Liver Fibrosis in Human Hepatic Stellate Cells. ACS Chem. Biol. *17*, 918–929.
11. Dewidar, B., Meyer, C., Dooley, S., and Meindl-Beinker, A.N. (2019). TGF-beta in Hepatic Stellate Cell Activation and Liver Fibrogenesis-Updated 2019. Cells *8*, 1419.
12. Smillie, C.S., Biton, M., Ordovas-Montanes, J., Sullivan, K.M., Burgin, G., Graham, D.B., Herbst, R.H., Rogel, N., Slyper, M., Waldman, J., et al. (2019). Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell *178*, 714–730.e22.
13. Wei, K., Nguyen, H.N., and Brenner, M.B. (2021). Fibroblast pathology in inflammatory diseases. J. Clin. Invest. *131*, e149538.
14. Yamaguchi, S., Kaneko, M., and Narukawa, M. (2021). Approval success rates of drug candidates based on target, action, modality, application, and their combinations. Clin. Transl. Sci. *14*, 1113–1122.
15. Bunnage, M.E. (2011). Getting pharmaceutical R&D back on target. Nat. Chem. Biol. *7*, 335–339.
16. Cong, F., Cheung, A.K., and Huang, S.M.A. (2012). Chemical genetics-based target identification in drug discovery. Annu. Rev. Pharmacol. Toxicol. *52*, 57–78.
17. Jones, L.H., and Bunnage, M.E. (2017). Applications of chemogenomic library screening in drug discovery. Nat. Rev. Drug Discov. *16*, 285–296.
18. Chandrasekaran, S.N., Ceulemans, H., Boyd, J.D., and Carpenter, A.E. (2021). Image-based profiling for drug discovery: due for a machine-learning upgrade? Nat. Rev. Drug Discov. *20*, 145–159.
19. Hinterleitner, T.A., Saada, J.I., Berschneider, H.M., Powell, D.W., and Valentich, J.D. (1996). IL-1 stimulates intestinal myofibroblast COX gene expression and augments activation of Cl- secretion in T84 cells. Am. J. Physiol. *271*, C1262–C1268.
20. Groover, M.K., and Richmond, J.M. (2020). Potential therapeutic manipulations of the CXCR3 chemokine axis for the treatment of inflammatory fibrosing diseases. F1000Res. *9*, 1197.
21. Julian, D.R., Kazakoff, M.A., Patel, A., Jaynes, J., Willis, M.S., and Yates, C.C. (2021). Chemokine-Based Therapeutics for the Treatment of Inflammatory and Fibrotic Convergent Pathways in COVID-19. Curr. Pathobiol. Rep. *9*, 93–105.
22. Korsunsky, I., Wei, K., Pohin, M., Kim, E.Y., Barone, F., Major, T., Taylor, E., Ravindran, R., Kemble, S., Watts, G.F.M., et al. (2022). Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast phenotypes in four chronic inflammatory diseases. Med (N Y) *3*, 481–518.e14.
23. Raghu, G., Martinez, F.J., Brown, K.K., Costabel, U., Cottin, V., Wells, A.U., Lancaster, L., Gibson, K.F., Haddad, T., Agarwal, P., et al. (2015). CC-chemokine ligand 2 inhibition in idiopathic pulmonary fibrosis: a phase 2 trial of carlumab. Eur. Respir. J. *46*, 1740–1750.
24. Berres, M.L., Trautwein, C., Schmeding, M., Eurich, D., Tacke, F., Bahra, M., Neuhaus, P., Neumann, U.P., and Wasmuth, H.E. (2011). Serum chemokine CXC ligand 10 (CXCL10) predicts fibrosis progression after liver transplantation for hepatitis C infection. Hepatology *53*, 596–603.
25. Bickerton, G.R., Paolini, G.V., Besnard, J., Muresan, S., and Hopkins, A.L. (2012). Quantifying the chemical beauty of drugs. Nat. Chem. *4*, 90–98.
26. Bray, M.A., Singh, S., Han, H., Davis, C.T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir, S.M., Gibson, C.C., and Carpenter, A.E. (2016). Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat. Protoc. *11*, 1757–1774.
27. Zhang, X.D. (2011). Illustration of SSMD, z score, SSMD*, z* score, and t statistic for hit selection in RNAi high-throughput screens. J. Biomol. Screen *16*, 775–785.
28. Simm, J., Klambauer, G., Arany, A., Steijaert, M., Wegner, J.K., Gustin, E., Chupakhin, V., Chong, Y.T., Vialard, J., Buijnsters, P., et al. (2018). Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery. Cell Chem. Biol. *25*, 611–618.e3.
29. Amano, H., Matsui, Y., Hatanaka, K., Hosono, K., and Ito, Y. (2021). VEGFR1-tyrosine kinase signaling in pulmonary fibrosis. Inflamm. Regen. *41*, 16.
30. Yang, L., Zhou, F., Zheng, D., Wang, D., Li, X., Zhao, C., and Huang, X. (2021). FGF/FGFR signaling: From lung development to respiratory diseases. Cytokine Growth Factor Rev. *62*, 94–104.
31. Liang, D., Chen, H., Zhao, L., Zhang, W., Hu, J., Liu, Z., Zhong, P., Wang, W., Wang, J., and Liang, G. (2018). Inhibition of EGFR attenuates fibrosis and stellate cell activation in diet-induced model of nonalcoholic fatty liver disease. Biochim. Biophys. Acta, Mol. Basis Dis. *1864*, 133–142.
32. Hostettler, K.E., Zhong, J., Papakonstantinou, E., Karakiulakis, G., Tamm, M., Seidel, P., Sun, Q., Mandal, J., Lardinois, D., Lambers, C., and Roth, M. (2014). Anti-fibrotic effects of nintedanib in lung fibroblasts derived from patients with idiopathic pulmonary fibrosis. Respir. Res. *15*, 157.
33. Rebeyrol, C., Saint-Criq, V., Guillot, L., Riffault, L., Corvol, H., Chadelat, K., Ray, D.W., Clement, A., Tabary, O., and Le Rouzic, P. (2012). Glucocorticoids reduce inflammation in cystic fibrosis bronchial epithelial cells. Cell. Signal. *24*, 1093–1099.
34. Vaglio, A., Palmisano, A., Alberici, F., Maggiore, U., Ferretti, S., Cobelli, R., Ferrozzi, F., Corradi, D., Salvarani, C., and Buzio, C. (2011). Prednisone versus tamoxifen in patients with idiopathic retroperitoneal fibrosis: an open-label randomised controlled trial. Lancet *378*, 338–346.
35. van Bommel, E.F.H., Siemes, C., Hak, L.E., van der Veer, S.J., and Hendriksz, T.R. (2007). Long-term renal and patient outcome in idiopathic

retroperitoneal fibrosis treated with prednisone. Am. J. Kidney Dis. *49*, 615–625.

36. Giebeler, A., Boekschoten, M.V., Klein, C., Borowiak, M., Birchmeier, C., Gassler, N., Wasmuth, H.E., Müller, M., Trautwein, C., and Streetz, K.L. (2009). c-Met confers protection against chronic liver tissue damage and fibrosis progression after bile duct ligation in mice. Gastroenterology *137*, 297-308–308.e1-4.
37. Kawaguchi, Y., Harigai, M., Hara, M., Fukasawa, C., Takagi, K., Tanaka, M., Tanaka, E., Nishimagi, E., and Kamatani, N. (2002). Expression of hepatocyte growth factor and its receptor (c-met) in skin fibroblasts from patients with systemic sclerosis. J. Rheumatol. *29*, 1877–1883.
38. Zbodakova, O., Chalupsky, K., Sarnova, L., Kasparek, P., Jirouskova, M., Gregor, M., and Sedlacek, R. (2021). ADAM10 and ADAM17 regulate EGFR, c-Met and TNF RI signalling in liver regeneration and fibrosis. Sci. Rep. *11*, 11414.
39. Ji, D., Zhao, Q., Qin, Y., Tong, H., Wang, Q., Yu, M., Mao, C., Lu, T., Qiu, J., and Jiang, C. (2021). Germacrone improves liver fibrosis by regulating the PI3K/AKT/mTOR signalling pathway. Cell Biol. Int. *45*, 1866–1875.
40. Lukey, P.T., Harrison, S.A., Yang, S., Man, Y., Holman, B.F., Rashidnasab, A., Azzopardi, G., Grayer, M., Simpson, J.K., Bareille, P., et al. (2019). A randomised, placebo-controlled study of omipalisib (PI3K/mTOR) in idiopathic pulmonary fibrosis. Eur. Respir. J. *53*, 1801992.
41. Madala, S.K., Edukulla, R., Phatak, M., Schmidt, S., Davidson, C., Acciani, T.H., Korfhagen, T.R., Medvedovic, M., Lecras, T.D., Wagner, K., and Hardie, W.D. (2014). Dual targeting of MEK and PI3K pathways attenuates established and progressive pulmonary fibrosis. PLoS One *9*, e86536.
42. Qin, W., Cao, L., and Massey, I.Y. (2021). Role of PI3K/Akt signaling pathway in cardiac fibrosis. Mol. Cell. Biochem. *476*, 4045–4059.
43. Wang, J., Hu, K., Cai, X., Yang, B., He, Q., Wang, J., and Weng, Q. (2022). Targeting PI3K/AKT signaling for treatment of idiopathic pulmonary fibrosis. Acta Pharm. Sin. B *12*, 18–32.
44. Espindola, M.S., Habiel, D.M., Narayanan, R., Jones, I., Coelho, A.L., Murray, L.A., Jiang, D., Noble, P.W., and Hogaboam, C.M. (2018). Targeting of TAM Receptors Ameliorates Fibrotic Mechanisms in Idiopathic Pulmonary Fibrosis. Am. J. Respir. Crit. Care Med. *197*, 1443–1456.
45. Mukherjee, S.K., Wilhelm, A., and Antoniades, C.G. (2016). TAM receptor tyrosine kinase function and the immunopathology of liver disease. Am. J. Physiol. Gastrointest. Liver Physiol. *310*, G899–G905.
46. Hu, M., Che, P., Han, X., Cai, G.Q., Liu, G., Antony, V., Luckhardt, T., Siegal, G.P., Zhou, Y., Liu, R.M., et al. (2014). Therapeutic targeting of SRC kinase in myofibroblast differentiation and pulmonary fibrosis. J. Pharmacol. Exp. Ther. *351*, 87–95.
47. Li, H., Zhao, C., Tian, Y., Lu, J., Zhang, G., Liang, S., Chen, D., Liu, X., Kuang, W., and Zhu, M. (2020). Src family kinases and pulmonary fibrosis: A review. Biomed. Pharmacother. *127*, 110183.
48. Lu, Y.Y., Zhao, X.K., Yu, L., Qi, F., Zhai, B., Gao, C.Q., and Ding, Q. (2017). Interaction of Src and Alpha-V Integrin Regulates Fibroblast Migration and Modulates Lung Fibrosis in A Preclinical Model of Lung Fibrosis. Sci. Rep. *7*, 46357.
49. Chen, J., Chen, J.K., Nagai, K., Plieth, D., Tan, M., Lee, T.C., Threadgill, D.W., Neilson, E.G., and Harris, R.C. (2012). EGFR signaling promotes TGFbeta-dependent renal fibrosis. J. Am. Soc. Nephrol. *23*, 215–224.
50. Vallath, S., Hynds, R.E., Succony, L., Janes, S.M., and Giangreco, A. (2014). Targeting EGFR signalling in chronic lung disease: therapeutic challenges and opportunities. Eur. Respir. J. *44*, 513–522.
51. Kim, S.K., Jung, S.M., Park, K.S., and Kim, K.J. (2021). Integrative analysis of lung molecular signatures reveals key drivers of idiopathic pulmonary fibrosis. BMC Pulm. Med. *21*, 404.
52. Liu, Y., Li, J., Liao, L., Huang, H., Fan, S., Fu, R., Huang, J., Shi, C., Yu, L., Chen, K.X., et al. (2021). Cyclin-dependent kinase inhibitor roscovitine attenuates liver inflammation and fibrosis by influencing initiating steps of liver injury. Clin. Sci. *135*, 925–941.
53. Madala, S.K., Schmidt, S., Davidson, C., Ikegami, M., Wert, S., and Hardie, W.D. (2012). MEK-ERK pathway modulation ameliorates pulmonary fibrosis associated with epidermal growth factor receptor activation. Am. J. Respir. Cell Mol. Biol. *46*, 380–388.
54. Terashima, H., Aonuma, M., Tsuchida, H., Sugimoto, K., Yokoyama, M., and Kato, M. (2019). Attenuation of pulmonary fibrosis in type I collagen-targeted reporter mice with ALK-5 inhibitors. Pulm. Pharmacol. Ther. *54*, 31–38.
55. Beyer, C., and Distler, J.H.W. (2013). Tyrosine kinase signaling in fibrotic disorders: Translation of basic research to human disease. Biochim. Biophys. Acta *1832*, 897–904.
56. Mansuroglu, T., Ramadori, P., Dudás, J., Malik, I., Hammerich, K., Füzesi, L., and Ramadori, G. (2009). Expression of stem cell factor and its receptor c-Kit during the development of intrahepatic cholangiocarcinoma. Lab. Invest. *89*, 562–574.
57. Huang, L.S., Fu, P., Patel, P., Harijith, A., Sun, T., Zhao, Y., Garcia, J.G.N., Chun, J., and Natarajan, V. (2013). Lysophosphatidic acid receptor-2 deficiency confers protection against bleomycin-induced lung injury and fibrosis in mice. Am. J. Respir. Cell Mol. Biol. *49*, 912–922.
58. Sakai, N., Chun, J., Duffield, J.S., Lagares, D., Wada, T., Luster, A.D., and Tager, A.M. (2017). Lysophosphatidic acid signaling through its receptor initiates profibrotic epithelial cell fibroblast communication mediated by epithelial cell derived connective tissue growth factor. Kidney Int. *91*, 628–641.
59. Goodwin, J., Choi, H., Hsieh, M.H., Neugent, M.L., Ahn, J.M., Hayenga, H.N., Singh, P.K., Shackelford, D.B., Lee, I.K., Shulaev, V., et al. (2018). Targeting Hypoxia-Inducible Factor-1alpha/Pyruvate Dehydrogenase Kinase 1 Axis by Dichloroacetate Suppresses Bleomycin-induced Pulmonary Fibrosis. Am. J. Respir. Cell Mol. Biol. *58*, 216–231.
60. Hamanaka, R.B., Nigdelioglu, R., Meliton, A.Y., Tian, Y., Witt, L.J., O'Leary, E., Sun, K.A., Woods, P.S., Wu, D., Ansbro, B., et al. (2018). Inhibition of Phosphoglycerate Dehydrogenase Attenuates Bleomycin-induced Pulmonary Fibrosis. Am. J. Respir. Cell Mol. Biol. *58*, 585–593.
61. Buhl, E.M., Djudjaj, S., Klinkhammer, B.M., Ermert, K., Puelles, V.G., Lindenmeyer, M.T., Cohen, C.D., He, C., Borkham-Kamphorst, E., Weiskirchen, R., et al. (2020). Dysregulated mesenchymal PDGFR-beta drives kidney fibrosis. EMBO Mol. Med. *12*, e11021.
62. Xie, Y., Su, N., Yang, J., Tan, Q., Huang, S., Jin, M., Ni, Z., Zhang, B., Zhang, D., Luo, F., et al. (2020). FGF/FGFR signaling in health and disease. Signal Transduct. Target. Ther. *5*, 181.
63. Tanjore, H., Blackwell, T.S., and Lawson, W.E. (2012). Emerging evidence for endoplasmic reticulum stress in the pathogenesis of idiopathic pulmonary fibrosis. Am. J. Physiol. Lung Cell Mol. Physiol. *302*, L721–L729.
64. Kropski, J.A., and Blackwell, T.S. (2018). Endoplasmic reticulum stress in the pathogenesis of fibrotic disease. J. Clin. Invest. *128*, 64–73.
65. Santacroce, G., Lenti, M.V., and Di Sabatino, A. (2022). Therapeutic Targeting of Intestinal Fibrosis in Crohn's Disease. Cells *11*, 429.
66. Flaherty, K.R., Wells, A.U., Cottin, V., Devaraj, A., Walsh, S.L.F., Inoue, Y., Richeldi, L., Kolb, M., Tetzlaff, K., Stowasser, S., et al. (2019). Nintedanib in Progressive Fibrosing Interstitial Lung Diseases. N. Engl. J. Med. *381*, 1718–1727.
67. Kurita, Y., Araya, J., Minagawa, S., Hara, H., Ichikawa, A., Saito, N., Kadota, T., Tsubouchi, K., Sato, N., Yoshida, M., et al. (2017). Pirfenidone inhibits myofibroblast differentiation and lung fibrosis development during insufficient mitophagy. Respir. Res. *18*, 114.
68. Kalderén, C., Stadler, C., Forsgren, M., Kvastad, L., Johansson, E., Sydow-Bäckman, M., and Svensson Gelius, S. (2014). CCL2 mediates anti-fibrotic effects in human fibroblasts independently of CCR2. Int. Immunopharmacol. *20*, 66–73.
69. Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. *9*, 90–95.
70. Waskom, M. (2021). seaborn: statistical data visualization. J. Open Source Softw. *6*, 3021.

71. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., and Thirion, B. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

72. McKinney, W., et al. (2010). Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference, *445*, pp. 51–56.

73. Wali, G., Berkovsky, S., Whiten, D.R., Mackay-Sim, A., and Sue, C.M. (2021). Single cell morphology distinguishes genotype and drug effect in Hereditary Spastic Paraplegia. Sci. Rep. *11*, 16635.

74. Wattenberg, M., Viégas, F., and Johnson, I. (2016). How to Use t-SNE Effectively. Distill *1*.

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-ACTA2 antibody	Abcam	Cat#:ab7817; RRID:AB_262054
anti-COL1A1 antibody	Abcam	Cat#:Ab34710; RRID:AB_731684
goat anti-mouse Alexa 594	ThermoFisher	Cat#:A-11032; RRID:AB_2534091
goat anti-rabbit Alexa 488	ThermoFisher	Cat#:A-11008; RRID:AB_143165
Chemicals, peptides, and recombinant proteins		
MitoTracker™ Deep Red FM	ThermoFisher	Cat#:M22426
Hoechst 33342	Invitrogen	Cat#:H3570
Concanavalin A-Alexa 488	Invitrogen	Cat#:C11252
SYTO™ 14	Invitrogen	Cat#:S7576
WGA-Alexa 555	Invitrogen	Cat#:W32464
phalloidin-Alexa 568	Invitrogen	Cat#:A12380
Sodium bicarbonate	Sigma Millipore	Cat#:S6014
Recombinant human OSM	Bio-techne	Cat#:295-OM-050
Recombinant human TNFα	Bio-techne	Cat#:10291-TA
Recombinant human IL-1β	Bio-techne	Cat#:201-LB
Recombinant human TL1A	Bio-techne	Cat#:1319-TL
Recombinant human IL-36	Bio-techne	Cat#:6835-IL
Recombinant human TGFβ	Bio-techne	Cat#:240-B-002
Trypsin EDTA	ThermoFisher	Cat#:25200056
Fetal bovine serum	Gibco	Cat#:10082147
Chemogenomic screen library	Selleckchem	Cat#:L1100 and L3800
32% Paraformaldehyde	Electron Microscopy Sciences	Cat#:15714S
PBS	ThermoFisher	Cat#:10010023
Lentiviral particles of inducible Cas9 nuclease with hEF1a promoter	Horizon discovery	Cat#:VCAS11227
blastcidin	ThermoFisher	Cat#:A1113903
doxycycline	Sigma Millipore	Cat#:D9891
Gene Knockout Kit v2 - human - *TNFRSF1A*	Synthego	https://www.synthego.com/order/crispr-kits/gene-knockout-kit
Gene Knockout Kit v2 - human - *TNFRSF1B*	Synthego	https://www.synthego.com/order/crispr-kits/gene-knockout-kit
Lipofectamine™ RNAiMAX	ThermoFisher	Cat#:13778075
Opti-MEM	ThermoFisher	Cat#:31985062
Phusion high-fidelity buffer	ThermoFisher	Cat#:F-518L
RNase A	ThermoFisher	Cat#:EN0531
dNTPs	ThermoFisher	Cat#:R0181
Phusion hot start II high-fidelity DNA polymerase	ThermoFisher	Cat#:F-549S
NEBuffer 2	NEB	Cat#:B7002S
T7 endonuclease I	NEB	Cat#:M0302S
Triton X-100	Sigma Millipore	Cat#:X100
Bovine serum albumin (BSA)	VWR	Cat#:97061-420
Eagle's Minimum Essential Medium	ATCC	Cat#:30-2003
Other		
OptiPlate-384 microplates	PerkinElmer	Cat#:6007290

(*Continued on next page*)

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CellCarrier-384 Ultra microplates	PerkinElmer	Cat#:6057302
Critical commercial assays		
10X Genomics Next-GEM chip	10X Genomics	Cat#:1000127
10X Genomics Next-GEM 3' GEM kit	10X Genomics	Cat#:1000121
Agilent Tapestation D1000 and D5000 tapes	Agilent	Cat#:5592 and 5584
Human CXCL10 HTRF kit	Cisbio	Cat#:62HCX10PEH
Human CCL2 HTRF kit	Cisbio	Cat#:62HCCL2PEG
Luminex xMAP kit	Luminex	Customized
Experimental models: Cell lines		
CCD-18co	ATCC	Cat#:CRL-1459; RRID:CVCL_2379
Software and algorithms		
Cell ranger mkfastq	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/tutorial_fq
Seurat v3	Butler et al., 2018	https://satijalab.org/seurat/
Ingenuity Pathway Analysis (IPA)	Qiagen Inc	https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/
Green Button Go scheduler	Biosero	https://biosero.com/software/green-button-go-scheduler/
Harmony® PhenoLOGIC™ software	PerkinElmer	https://www.perkinelmer.com/product/harmony-4-9-office-license-hh17000010
TIBCO® SpotFire® Signals VitroVivo	TIBCO / PerkinElmer	https://perkinelmerinformatics.com/products/research/signals-vitrovivo
Matplotlib and seaborn Python libraries	Hunter et al.[69] Waskom et al.[70]	N/A
scikit-learn and pandas Python libraries	Pedregosa et al.[71] Mckinney et al.[72]	N/A
Graphpad Prism software	GraphPad software Inc	https://www.graphpad.com/scientific-software/prism/
Deposited data		
Single cell analysis of the colon mucosa of UC patients and healthy individuals	Smillie et al.[12]	Single Cell Portal: SCP259 (https://singlecell.broadinstitute.org/single_cell/study/SCP259)
CCD18co colonic fibroblast cell line scRNAseq, with treatment of pro-inflammatory stimuli	This paper	GEO accession: GSE233063 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE233063)

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Shane R. Horman (shane.horman@takeda.com).

Materials availability

All materials used in this paper are commercially available. If any additional request, please direct to the lead contact.

Data and code availability

- This paper does not report original code.
- Cell Ranger matrix loupe tar files of the single cell RNA-seq data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. This paper analyzes existing, publicly available

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

data. These accession numbers for the datasets are also listed in the key resources table.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

CCD-18co cell line

CCD-18co cell line (Cat #: CRL-1459) was obtained from American Type Culture Collection (ATCC). CCD-18Co is a cell line exhibiting fibroblast morphology that was isolated from the normal colon tissue of a 2.5-month-old, Black, female.

CCD-18co cell line maintenance and stimulation

A large batch of CCD-18co cells were purchased from ATCC (Cat #: CRL-1459), and were cultured with ATCC-formulated Eagle's Minimum Essential Medium (Cat #: 30-2003) with 10% fetal bovine serum (FBS, Gibco, Cat#:10082147) as recommended by ATCC. The cells earlier than passage-6 were used in experiments. For pro-fibrotic stimuli treatments, if not specified otherwise, the concentrations of stimuli used to treat CCD-18co cells were as follow: 10 ng/mL OSM (Bio-techne, Cat#:295-OM-050), 10 ng/mL TNFα (Bio-techne, Cat#:10291-TA), 10 ng/mL IL-1β (Bio-techne, Cat#:201-LB), 100 ng/mL TL1A (Bio-techne, Cat#:1319-TL), 200 ng/mL IL-36 (Bio-techne, Cat#:6835-IL) and 10 ng/mL TGFβ (Bio-techne, Cat#:240-B-002).

METHOD DETAILS

Single Cell RNA-seq

CCD-18co cells were cultured as previously described, and were treated with vehicle, 10 ng/mL OSM, 10 ng/mL TNFα, 10 ng/mL IL-1β, 100 ng/mL TL1A, 200 ng/mL IL-36 or 10 ng/mL TGFβ for 16 hours. Cells were detached from plastic culture media using Trypsin EDTA (ThermoFisher, Cat#:25200056) and manually counted. 10,000 cells of each sample were loaded in individual lanes of a 10X Genomics Next-GEM chip (10X Genomics, cat #: 1000127). Libraries were prepared using the 10X Genomics Next-GEM 3' GEM kit (10X Genomics, cat #: 1000121). Library quality control was performed using the Agilent Tapestation D1000 and D5000 tapes (Agilent, part #: 5592 and 5584). Sequencing was performed externally at a vendor (Novogene Inc.) at a depth of 50,000 reads per cell. All sequencing raw files are available (GEO accession code GSE233063).

Analysis of CCD-18Co scRNA-Seq data

Cell Ranger mkfastq was used to demultiplex raw sequencing reads, and Cell Ranger count was used to align reads to Human GRCh38 transcriptome, and generate gene-cell expression count matrices. Expression matrices were filtered to remove low quality cells with less than 200 genes detected or more than 0.25% mitochondrial mRNAs. Cell filtering resulted on an average of ~2,400 cells per condition and a total number of 16,750 cells in the integrated dataset. Seurat v3 workflow was employed to perform normalization, detection of variable genes, dimensionality reduction, and graph-based clustering with Louvain algorithm. Upon log-normalization and scaling of gene expression, variable genes were identified using the vst method and then subjected to principal component analysis (PCA). The number of principal components (PCs) used for nonlinear dimensional reduction analysis (UMAP) was chosen according to the PCElbowPlot function and JackStrawPlot function. For cell clustering, FindClusters method was parameterized with different resolutions to optimize cluster granularity. Sample integration was performed with the IntegrateData function using anchors set by FindIntegrationAnchors. Cells with high mitochondrial RNA expression (greater than 5% of total cell reads) were excluded from downstream analysis. FindAllMarkers function in Seurat was utilized to detect top gene markers per cluster with the default Wilcoxon rank-sum test setting. The results of FindAllMarkers were subjected to functional pathway analysis. Up-regulated gene markers per cluster with logFC>0.4 and adjusted p-value <0.05 were subjected to functional enrichment with Ingenuity Pathway Analysis (IPA) (Qiagen Inc.). IPA canonical pathways and upstream transcriptional regulators were predicted per cluster and then integrated via the Comparison Analysis function. All analyzed files are available (GEO accession code GSE233063).

High throughput screening assay

The chemogenomic screen library was assembled with two screening libraries from Selleckchem (cat #: L1100 and L3800, the complete list of libraries and batch specific certificates of analysis, MSDS, HNMR data and HPLC data can be found on www.selleckchem.com). The high throughput screen was conducted with an automation platform consisting of multiple instruments as shown below , which were operated by Green Button Go software (Biosero).

Instrument name	Purpose	Brand	Count	Cat #
Precise PF400 Arm	Robot arm	Precise (ordered through Biosero)	1	INT3080006
Precise rail 2M linear	Axis rail for PF400	Precise (ordered through Biosero)	1	INT3080001
Cytomat™ 10 C425 automated incubator	4°C incubator	ThermoFisher	1	51031161
Cytomat™ 10 C450 automated incubator	37°C incubator	ThermoFisher	1	51031166

(*Continued on next page*)

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

Continued

Instrument name	Purpose	Brand	Count	Cat #
Biosolutions AmbiStore™ D	Room temperature plate storage	HighRes	1	N/A
BioShake 5000 elm shaker	Plate shaker	Qinstrument	3	2016-0022
Microplate centrifuge	Microplate centrifuge	Agilent	1	G5582AA
PlateLoc thermal microplate sealer	Microplate sealer	Agilent	1	G5585BA
XPeel	Microplate peeler	Brooks	1	XP-A
EL406 combination microplate washer dispenser	Washer and dispenser	BioTek	1	406PSUB1
Bravo liquid handler bundle	Dispenser	Agilent	1	G5563AA
Blue®Washer	Washer and dispenser	Blue Cat Bio	1	N/A
Operetta CLS high content analyis system	High content imager	PerkinElmer	1	N/A
EnVision® HTS plate reader	Multi-mode plate reader	PerkinElmer	1	2105-0010
Regrip nest	Plate staging and regripping	Biosero	1	INT3020012

For the primary screen, 1,200 cells/16 μL/well of CCD-18co cells were seeded in CellCarrier-384 Ultra microplates (Perkin Elmer, Cat #: 6057302) with an EL406 combination microplate washer dispenser (EL406, Biotek). After overnight incubation at 37°C, 37.5 nL/well (for 2 mM stock) or 7.5 nL/well (for 10 mM stock) small molecules were dispensed by ATS GEN5 (EDC Biosystems). The wells were then backfilled with 10 μL/well cell culture medium with an EL406 (except for column 12). Column 12 was filled with 10 μL/well adalimumab (final concentration 1 nM) for positive control wells or 10 μL/well medium for negative controls. After overnight incubation, each well was then filled with 25 μL/well recombinant human TNFα (R&D systems, cat #: 210-TA-100) diluted in cell culture medium (final concentration 10 ng/mL) with an EL406. After 48-hour stimulation, 4 μL/well of cell culture supernatant was taken out and dispensed into OptiPlate-384 microplates (PerkinElmer, cat #: 6007290) that was pre-filled by 12 μL/well PBS by an EL406. CXCL10 HTRF assay antibodies were prepared according to the manufacturer's recommendation (Cisbio, cat #: 62HCX10PEH). 4 μL of the mixture of equal part of diluted CXCL10 Eu Cryptate antibody and CXCL10 d2 antibody was added to PBS diluted supernatant in OptiPlate-384 microplates by a Bravo liquid handler. The OptiPlate-384 microplate was then sealed by a PlateLoc thermal microplate sealer (Agilent) and shaken at 1000 rpm for 30 seconds on a BioShake 5000 elm shaker (QInstrument) followed by centrifugation at 500xg for 30 seconds on a Microplate centrifuge (Agilent). The OptiPlate-384 plate was incubated at room temperature in a Biosolutions AmbiStore™ D (HighRes) for 2 hours before the HTRF signal was measured on an EnVision® HTS plate reader (PerkinElmer). For OSM-induced CCL2 assay in hit validation step, the cell supernatant was collected after 2-hour treatment with OSM. The experimental procedure was similar to CXCL10 assay, except the signal was detected using a CCL2 kit (Cisbio, Cat #: 62HCCL2PEG). The OSM treated cells were then returned back to the incubator, and the Cell Painting assay was carried out after 48 hours post OSM treatment. After supernatant being transferred to the OptiPlate-384 microplates, the cells in CellCarrier-384 Ultra microplates were treated with 25 μ L of 0.71 μM MitoTracker™ Deep Red FM (ThermoFisher, cat #: M22426) that was prepared in cell culture medium from 1 mM DMSO stock (final concentration of MitoTracker™ Deep Red FM is 0.25 μM). MitoTracker™ Deep Red FM was dispensed by an EL406. After incubation with MitoTracker™ Deep Red FM for 30 minutes at 37°C, the cells were then fixed with 4% paraformaldehyde (made from a 32% stock, Electron Microscopy Sciences, Cat#:15714S) for 10 minutes. Paraformaldehyde was dispensed by an EL406. The CellCarrier-384 Ultra microplates were then washed with PBS (ThermoFisher, Cat#:10010023) twice on a Blue®Washer, and further stained with a cocktail of 15 μL/well including 2 μg/mL Hoechst 33342 (Invitrogen, cat #: H3570), 100 μg/mL Concanavalin A-Alexa 488 (Invitrogen, cat #: C11252), 3 μM SYTO™ 14 (Invitrogen, cat #: S7576), 5 μg/mL WGA-Alexa 555 (Invitrogen, cat #: W32464), 2.5 u/mL phalloidin-Alexa 568 (Invitrogen, cat #: A12380) for 30 minutes at room temperature. The dye cocktail was dispensed by a Blue®Washer. After incubation, the plates were washed with 1X PBS three times. The cells were then imaged with an Operetta CLS High Content Analysis System (PerkinElmer).

MitoTracker™ Deep Red FM was reconstituted with DMSO and was always prepared fresh from lyophilized stock. Concanavalin A-Alexa 488 was diluted with 0.1 M sodium bicarbonate and WGA-Alexa 555 was diluted with ddH2O, and unused stock of both could be frozen in -20°C up to 1 month. Phalloidin-Alexa 568 was reconstituted with DMSO. Both phalloidin-Alexa 568 and SYTO™ 14 could be frozen in -20°C up to 1 year. Hoechst 33342 was stored at 4°C.

Cell Painting high content imaging, feature extraction and high dimensionality data analysis

Cells stained with Cell Painting dyes on CellCarrier-384 Ultra microplates were imaged with an Operetta CLS High Content Analysis System. The details of the Operetta CLS channels and stains imaged in the Cell Painting assay are shown below.

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

Dye	Filter (Excitation)	Filter (Emission)	Organelle or cellular component
Hoechst 33342	355-385 nm	430-500 nm	Nucleus
Concanavalin A-Alexa 488	460-490 nm	500-550 nm	Endoplasmic reticulum
SYTO™ 14	490-515 nm	525-580 nm	Nucleoli, cytoplasmic RNA
WGA-Alexa 555 and phalloidin-Alexa 568	530-560 nm	600-640 nm	F-actin cytoskeleton, Golgi, plasma membrane
MitoTracker™ Deep Red FM	615-645 nm	685-760 nm	Mitochondria

20x water objective with non-confocal mode was used. Binning value of 2 was chosen to boost camera frame rate and dynamic range. The middle 9 fields of view were imaged for each well. The exposure time and power were adjusted to ensure the max of signal level under auto contrast for each channel was between the range of 8,000 – 15,000.

For Cell Painting feature extraction, Harmony® PhenoLOGIC™ software was used. The basic flatfield correction and bright-field correction were applied to the images. Hoechst 33342 channel was used to identify nuclei, followed by using concanavalin A-Alexa 488 channel to find cytoplasm. Border objects of each image were removed to ensure the cells with complete image were used for feature extraction. Properties of ell features were then quantified, including fluorescence intensities within different cell regions, basic morphological features (cell area, roundness, length, width and width/length ratio), advanced morphological features (Symmetry, Threshold compactness, Axial, Radial and profile, or STAR properties), cell texture features (Spot, Hole, Edge, Ridge, Valley, Saddle, Bright and Dark, or SER properties).[73] In total, there were 860 cellular features being extracted.

For CXCL10 HTRF and Cell Painting assay development, CCD-18co cells were treated with a range of concentrations of adalimumab biosimilar or infliximab biosimilar for 1 hour followed by treated of various concentrations of TNFα. After treatment with TNFα for 48 hours, the supernatant samples were taken for CXCL10 measurement by a HTRF kit, and the cells were stained with Cell Painting dyes and imaged on a high content imager. The Pearson's correlation between each cellular feature and Y=log(TNFα concentration+1) was calculated. Multiple testing adjusted p-value was also calculated for each feature using a random permutation method. The top features were used to fit a linear regression model, which were then used to calculate scores for the experiment run. The Cell Painting score was then normalized to 0 and 10 ng/mL TNFα treatments in a 0-to-100 scale.

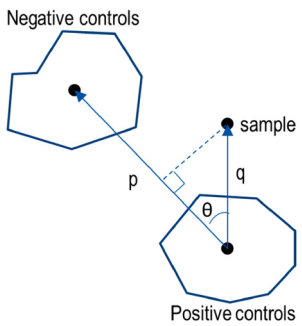
For the screening data analysis, the data including all extracted cellular features was imported to TIBCO® SpotFire® Signals VitroVivo (PerkinElmer) for analysis. Within the software, Editable Data Grid application was used to associate cellular feature data file with the compound transfer log file to annotate compounds. Grid Plate Editor application was used to designate the plate layout. High Content Profiler was then used to perform advanced feature normalization, selection, classification, profiling and hits selection. The compounds with score in the range of Average(neg ctrl) ± 3 S.D. were selected. If they appear 2 times among the tested triplicates, they were picked as hits.

In parallel, the Cell Painting data was analyzed with a linear predictive model. Briefly, the positive and negative controls were used as a training set to train a predictive model. Top features were ranked using area under the ROC curve. Multiple testing adjusted p-values were also calculated for each feature using a random permutation method. The top features that were highly predictive and could separate positive and negative controls were used to rank compounds. The compounds with scores in the range of Average(neg ctrl) ± 1 S.D. and appeared at least 2 times were picked as hits.

Cell Painting data projection score calculation

Projection score is a mathematical formalization of a heuristic, which predict sample's positivity depending on whether the feature values are closer to the positive controls and the negative controls. Projection score is calculated as

$$PS = \frac{q \times \cos(\theta)}{p}$$



Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

Where ***p*** is the distance between the center of positive and negative controls, ***q*** is the distance between sample and the center of negative controls, and θ is the angel between side ***q*** and side ***p***.

CCD-18co cell hit confirmation and validation hit picking strategy

For cytokine (CXCL10 and CCL2) assay hit picking, the assay plates first passed SSMD>1.28 QC criteria. Compounds' cytotoxicity effects were assessed by cell count from the imaging assay. If three or more cell count data points out of the six data points (three doses in replicates) of a compound show < 50% cell viability compared to the negative control, this compound was then considered toxic and filtered out. The area under the curve (AUC) of six normalized CXCL10 percent inhibition data points of each compound was then calculated for each stimulus. The compounds were ranked according to the values of AUC. For Cell Painting assay hit picking, after elimination of the cytotoxic compounds, the projection scores of Average(pos ctrl) ± 3 x S.D. were selected. If a compound had three or more data points out of the six data points that met this criterion, it was confirmed as a hit.

Visualization of primary screen Cell Painting dataset with t-SNE

To generate t-SNE projection of Cell Painting morphological feature profiles, compounds that have target annotations were selected and per-compound feature profiles were aggregated by calculating the mean of each feature across compound replicates. Per-target feature profiles were then aggregated by calculating the mean of each feature across all compounds annotated with this target. Per-target feature profiles were then rank-ordered by decreasing Euclidean distance from the mean feature profile of negative controls. Finally, the top-20 most distinct targets were selected and the targets that were associated with 3 or less compounds were filtered out, obtaining the final set of 13 targets.

The two-dimensional t-SNE embedding was generated using the scikit-learn Python library with random initialization,[71] perplexity of 20, early exaggeration of 30, automatic learning rate and squared cosine distance as a metric. Targets that formed coherent clusters were highlighted using kernel density estimation with empirically chosen thresholds. It should be noted that cluster sizes and inter-cluster distances should be interpreted with care when using t-SNE.[74] The t-SNE visualization was built using matplotlib[69] and seaborn[70] Python libraries.

CXCL10 and cellular feature statistical dependence analysis

To assess the correlation between CXCL10 level and cellular features extracted from Cell Painting imaging data, we first filtered out those compounds that demonstrated cytotoxic effects. We then selected those compounds that were present in both CXCL10 and imaging assays and aggregated per-compound imaging feature profiles by calculating the mean of each feature across compound replicates. Then, for each cellular morphological feature, we computed two metrics of relatedness to CXCL10 level: F-test statistic and Mutual information (MI). The method based on the F-test statistic estimates the degree of linear dependency between the feature and the outcome, while the mutual information method can capture any kind of statistical dependency. We plotted obtained results either as individual data points (Figure 5B) or as a bar plot that aggregates features by their type (Figure 5A), with error bars showing a bootstrap-estimated 95% confidence interval.

Computations of the F-test statistic and mutual information were performed using the scikit-learn[71] and pandas[72] Python libraries. Visualizations were built using matplotlib[69] and seaborn[70] Python libraries.

CRISPR/Cas9, mismatch detection and T7E1 assays

Lentiviral particles of inducible Cas9 nuclease with hEF1a promoter (Horizon discovery, cat #: VCAS11227) were transduced into CCD-18co cells with MOI=0.5 according to manufacturer's instruction. The Cas9 expressing cells were selected under 2 μg/mL blastcidin (ThermoFisher, Cat#:A1113903). The cells were added 10 ng/mL doxycycline (Sigma Millipore, Cat#:D9891) to induce Cas9 expression. The Cas9 expression could be readily detected by Western blot after 24 hours. Gene Knockout Kit v2 – Human or individual gRNA for targeting *TNFRSF1A* or *TNFRSF1B* were ordered from Synthego (https://www.synthego.com/order/crispr-kits/gene-knockout-kit). For gRNA transfection, 0.05 μL Lipofectamine™ RNAiMAX (ThermoFisher, cat#: 13778075) mixed with 5 μL opti-MEM was dispensed into each well of CellCarrier-384 Ultra microplates, followed by dispensing 2 pmol gRNA diluted in 5 μL opti-MEM. The RNAiMax and gRNA were then mixed by shaking the plate on a microplate shaker for 1 minute at 800 rpm. The plate was then centrifuged 1 minute at 500xg and incubated at room temperature for 20 minutes. After incubation, 1,200 cells/well with 10 ng/mL doxycycline were seeded in CellCarrier-384 Ultra microplates with RNAiMax and gRNA mixture. After 72 hours, the cells were collected for T7E1 assay.

For mismatch detection and T7E1 assay, the cells were lysed in 20 μL of 1x Phusion high-fidelity buffer (Thermofisher, cat #: F-518L) with 1 mg/mL proteinase K (Thermofisher, cat #: EO0492) and 0.5 mg/mL RNase A (cat #: EN0531). The plate was then incubated for 15 – 30 minutes at 56°C, followed by deactivation for 5 minutes at 96°C. Briefly centrifuge plate to collect liquid at bottom of wells. A 50 μL PCR reaction with the following condition was set up: 1X Phusion High-Fidelity buffer (cat #: F-549S), 500 nM forward primer, 500 nM reward primer, 200 μM each dNTPs (ThermoFisher, Cat#:R0181), 0.04 U/μL Phusion hot start II high-fidelity DNA polymerase (cat #: F-549S), and 5 μL cell lysate. The thermal cycling condition is as shown below:

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

Cycle step	Temperature (°C)	Time	Cycle(s)
Initial denaturation	98	3 min	1
Denature	98	10 s	10
Touchdown annealing	72 – 1/cycle	15 s	
Extension	72	30 s	
Denature	98	10 s	25
Annealing	62	15 s	
Extension	72	30 s	
Final extension	72	10 min	1

PCR samples were heated to 95°C for 10 minutes and then slowly (>15 minutes) cool to room temperature. 15 μL mismatch detection assay using T7E1 was then set up. The reaction includes PCR from gDNA (300 – 500 ng/reaction), 1X NEBuffer 2 (NEB, cat #: B7002S), and 0.33 U/μL T7 endonuclease I (NEB, Cat#: M0302S). The mixture was then incubated for 25 minutes at 37°C. The entire reaction was then loaded on 2% agarose gel to determine gRNA cleavage.

Olink Target 96 inflammation assay

CCD-18co cells were treated with vehicle, 10 ng/mL OSM, 10 ng/mL TNFα, 10 ng/mL IL-1β, 100 ng/mL TL1A, 200 ng/mL IL-36 or 10 ng/mL TGFβ for 16 or 48 hours. The cell culture supernatant samples (n=5 for each treatment) were collected and sent to Olink for the Olink Target 96 Inflammation assay. The samples were tested in neat or in 1:10 dilution. In total, 92 immune-related secretory biomarker protein levels were assessed. The results were presented in Normalized Protein Expression (NPX) value, which is an arbitrary unit on a Log2 scale. The linear portion of the NPX vs. concentration plot was used for calculation. The NPX values were further normalized to the vehicle treated control group. The heat map shows the average of the protein concentration of n=5. In parallel, the mRNA samples of the cells that were treated with various stimuli for 16 hours were sequenced. The matching gene expression levels of the 92 proteins that were assessed in the Olink assay were analyzed. The FPKM levels for each treatment was normalized to the vehicle control group. The heat map shows the average of the gene expression level of n=3.

RNA-seq

CCD-18co cells were treated with various stimuli for 16 hours in the same way as described above in Olink Target 96 Inflammation assay section. For mRNA-seq, the raw reads were aligned to the transcriptome using STAR (version 2.6.0)/RSEM (version 1.2.25) with default parameters, with a custom human GRCh38 transcriptome reference containing all protein coding and long noncoding RNA genes (based on human GENCODE version 32 annotation; downloaded from http://www.gencodegenes.org). The expression counts for each gene (i.e., transcripts per million) in all samples were normalized based on the sequencing depth. Differential expression genes (DEG) were identified using DESeq2 Bioconductor package at a minimum 2 fold change and false discovery rate <0.05 with Benjamini-Hochberg Procedure. The pathway enrichment analysis was performed using a hypergeometric test with Benjamini-Hochberg correction in Ingenuity pathway analysis (Qiagen). False discovery rate <0.05 was used as cutoff to identify significant signaling pathways after treatment.

Pro-fibrotic biomarker detection with Luminex®

6000 cells/100 μL/well of CCD-18co cells were seeded in 96 well plate. After overnight culture, they were treated with vehicle, 10 ng/mL OSM, 10 ng/mL TNFα, 10 ng/mL IL-1β, 100 ng/mL TL1A, 200 ng/mL IL-36 or 10 ng/mL TGFβ for 16 or 48 hours. After 16 hours, the cell culture supernatant was harvested. After 48 hours, upon collection of cell culture supernatant, the cells were washed once with PBS, and add 35 μL 0.1% tritonx100 (Sigma Millipore, Cat#:X100) in PBS to lyse cells. A customed kit for pro-fibrotic biomarker detection was ordered from R&D systems. The experiment was performed according to the manufacturer's instruction. Briefly, the supernatant and lysate samples were diluted 1:1 with the diluent provided in the Luminex kit. 50 μL of standard or sample was added to each well. Then 50 μL of diluted microparticle cocktail was added to each well, followed by incubation for 2 hours at room temperature on a shaker at 800 rpm. The plate was then washed by removing the liquid for each well when the plate was set on a magnetic plate separator to allow the magnetic beads to be sequestered, and filling with 100 μL wash buffer, and removing the liquid again. The wash was repeated 3 times. 50 μL of diluted biotin- antibody cocktail was added to each well, followed by covering and incubating for 1 hour at room temperature on a shaker at 800 rpm. Then the plate was washed 3 times. 50 μL of diluted streptavidin-PE was added to each well and incubated for 30 minutes at room temperature on a shaker at 800 rpm. Then the plate was washed 3 times. Then add 100 μL of wash buffer to each well and incubate for 2 minutes at room temperature on a shaker at 80 rpm. The fluorescence signal was read within 90 minutes using a Luminex 200 system.

Immunofluorescence detection for ACTA2 and COL1A1

1,500 cells/well CCD-18co cells were seeded in CellCarrier-384 Ultra microplates. After overnight culture, cells were treated with a range of concentrations of stimuli, including IL-11, OSM, TNFα, IL-1β, TL1A, IL-36 and TGFβ for 48 hours. Cells were then fixed with 2% paraformaldehyde for 10 minutes. After washing, the cells were permeabilized with 0.25% Triton X-100 in PBS for 10 minutes,

Please cite this article in press as: Yu et al., Integrating inflammatory biomarker analysis and artificial intelligence-enabled image-based profiling to identify drug targets for intestinal fibrosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.06.014

following by washing with PBS for 3 times. The cells were blocked with 1% BSA in PBST (0.1% Tween 20 diluted in PBS) for 30 minutes, and then incubated with anti-ACTA2 antibody (Abcam, cat #: ab7817) and anti-COL1A1 antibody (Abcam, cat #: ab34710). Both antibodies were used 1:1000 dilution in 1% BSA in PBST. After overnight incubation at 4°C, the cells were washed 3 times with PBS, followed by incubation with goat anti-mouse Alexa 594 (ThermoFisher, cat #: A-11032) and goat anti-rabbit Alexa 488 (ThermoFisher, cat #: A-11008) secondary antibodies for 30 minutes. The cells were then washed with PBS 3 times and stained with Hoechst 33342, before being imaged on an Operetta CLS High Content Analysis System.

QUANTIFICATION AND STATISTICAL ANALYSIS

Strictly standardized mean difference (SSMD) was used to assess plate quality. It is estimated as

$$\beta = \frac{\overline{X}_P - \overline{X}_N}{\sqrt{\frac{2}{K}\left((n_P - 1)S_P^2 + (n_N - 1)S_N^2\right)}}$$

Where $\overline{X}_P$ and $\overline{X}_N$ are means, S_P^2 and S_N^2 are variances, and n_P and n_N are sample sizes of positive and negative controls, respectively. $K \approx n_P + n_N - 3.48$.

For all experiments, statistical analysis was conducted using T-test or one-way ANOVA with Tukey's post hoc test on Graphpad Prism software with significance of * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$ if not specifically mentioned. Results are presented as means ± SD or means ± SEM, which is specified in the figure descriptions.