Gastro Hep Advances 2022;1:689–702

ORIGINAL RESEARCH—BASIC

Proteome-Wide Analysis Using SOMAscan Identifies and Validates Epidermal Growth Factor as a Disease Marker of Collagenous Gastritis

Debora Curci,[1] Simon T. Dillon,[2,3,4] Xuesong Gu,[2,3,4] Harland Winter,[4,5,*] and Towia A. Libermann[2,3,4,*]

[1]Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Advanced Diagnostic and Translational Medicine Laboratory, Trieste, Italy; [2]Division of Interdisciplinary Medicine and Biotechnology, Beth Israel Deaconess Medical Center, Boston, Massachusetts; [3]Beth Israel Deaconess Medical Center Genomics, Proteomics, Bioinformatics and Systems Biology Center, Boston, Massachusetts; [4]Department of Medicine, Harvard Medical School, Boston, Massachusetts; and [5]Center for Pediatric Inflammatory Bowel Disease, Massachusetts General Hospital, Boston, Massachusetts

BACKGROUND AND AIMS: Collagenous gastritis (CG) is a rare disorder characterized by increased subepithelial collagen deposition and inflammatory infiltrates. The mechanisms involved in CG pathogenesis are poorly understood, and no CG-associated biomarkers have been identified. This proteomics study identified serum biomarkers and pathogenic pathways to provide new knowledge about the pathobiology of CG, a disease reported in less than 100 patients. **METHODS:** Nine serum samples from pediatric patients diagnosed with CG were evaluated using novel aptamer-based proteomic technology and systems biology to generate new knowledge about the complex interactions between the differentially expressed proteins and candidate upstream regulators, using the Ingenuity Pathway Analysis in patients with non-CG and patients with normal gastric biopsies or nongastritis (NG). **RESULTS:** SOMAscan analysis identified 63 proteins significantly dysregulated in CG as compared to non-CG or NG patients that converged around enhanced inflammatory response and immune cell migration but reduced vascular functions. Principal component analysis using 15 of those proteins accurately separated the CG cases from the 2 comparator control groups. Using immunoassays, serum epidermal growth factor concentrations in CG patients, a protein involved in collagen production, were confirmed to be significantly lower than those in gastritis/NG patients. **CONCLUSION:** This is the first comprehensive analysis of the proteome in CG patients that reveals metabolic pathways relating inflammation and fibrosis as well as a new potential role of epidermal growth factor as a disease biomarker.

Keywords: Collagenous Gastritis; Proteomics; Biomarker; Epidermal Growth Factor

Introduction

Collagenous gastritis (CG) is a rare disorder characterized by subepithelial collagen deposition and infiltration of the lamina propria with a mixture of inflammatory cells.[1] Approximately 100 cases are reported[2,3] with a childhood-onset CG incidence of 1/400,000 person-years.[4] Two phenotypes of the disease, pediatric and adult, are identified. Both types present with anemia and abdominal pain, but nausea/vomiting are more common in children, whereas diarrhea is more common in adults. In contrast to pediatric-onset disease, adult-onset CG is more heterogenous and may involve other areas of the gastrointestinal (GI) tract.[4] The etiopathogenesis of CG remains unknown, and no infectious agents or diagnostic biomarkers are known. Immune deficiency as well as intestinal and autoimmune disorders, including celiac disease, Sjogren syndrome, systemic lupus erythematosus, juvenile and rheumatoid arthritis, Graves' disease, and diabetes mellitus type 1, have been associated with CG in both pediatric and adult patients.[2] No standard therapy is recommended for CG except to correct anemia with oral or intravenous iron replacement.[3]

Specific metabolic pathways associated with CG remain unknown. The objectives of this study were to assess serum biomarkers and pathogenic pathways in individuals with pediatric-onset CG to identify a serum protein signature that distinguishes CG from patients with noncollagenous gastritis (NCG) and nongastritis (NG). The multiplex aptamer-based SOMAscan proteomics platform that measures 1305 biologically relevant proteins across the whole dynamic range was used to identify serum proteins that are associated with CG. Proteins selected from the SOMAscan-derived results were validated using an orthogonal immunoassay method (Ella). Pathway analysis of the differentially

*Co-senior authors.

Abbreviations used in this paper: **BA, bile acid; CG, collagenous gastritis; EGF, epidermal growth factor; EMT, epithelial-to-mesenchymal transition; FXR, farnesoid X receptor; GI, gastrointestinal; IGF, insulin-like growth factor; IPA, Ingenuity Pathway Analysis; NCG, noncollagenous gastritis; NG, nongastritis; PCA, principal component analysis.**

Most current article

2772-5723
https://doi.org/10.1016/j.gastha.2022.04.016

expressed CG-associated proteins generated new knowledge into potential pathophysiological mechanisms linked to CG that could lead to the discovery of new treatments.

Material and Methods

Population Cohort

All patients were enrolled in the Massachusetts General Hospital, and the institutional review board approved the pediatric GI biorepository and signed consent/assent (Protocol Title: Pediatric Gastrointestinal Tissue, Stool, Saliva and Blood Registry; IRB Protocol No: 2009P001287). Clinical characteristics are summarized in Table 1. Serum samples were aliquoted and stored at −80 °C until assayed.

SOMAscan Assay

SOMAscan analysis (SomaLogic, Inc, Boulder, CO) using serum samples was performed on 9 patients with CG, 12 patients with NCG, and 8 patients with normal gastric biopsies (NG) at the BIDMC Genomics, Proteomics, Bioinformatics, and Systems Biology Center, an experienced SomaLogic-certified service provider, according to standard protocols for serum from SomaLogic that have been described elsewhere.[5] Using the recommended protocol from the manufacturer, 50-μL serum was run on the SOMAscan Assay Kit for human serum 1.3k (cat. #900-00012), which measures the expression of 1305 human proteins using highly selective single-stranded modified slow off-rate modified DNA aptamers (SOMAmer). Five pooled human serum controls and 1 no-protein buffer control were run in parallel with the serum test samples. Sample-to-sample variability was further controlled by several hybridization spike-in controls. Data quality control, calibration, and normalization were done by SomaLogic according to the manufacturer's protocol, as previously described.[6] All samples passed the SomaLogic standard quality control and normalization criteria for the manual 1.3k assay.

Statistical Analysis

Mean and median fold changes of protein expression were calculated for proteins with statistically significantly different expression between CG cases and NCG and NG cases. Statistical significance was determined by using a t-test to compare SOMAscan relative fluorescence units. A protein was considered to be significantly dysregulated if the P value for expression between cases and controls was $<.05$, as described elsewhere.[7] Hierarchical clustering heat maps of the most significantly dysregulated proteins were generated with Morpheus (Broad Institute, Cambridge, MA). A principal component analysis (PCA) was performed for the top dysregulated proteins to evaluate their ability to discriminate cases from controls using XLSTAT (Addinsoft, Long Island City, NY). Box-whisker plots of median protein expression values and receiver operating characteristic curves were created using XLSTAT.

Systems Biology Analysis

To acquire new insights into potential pathophysiological pathways and biological functions underlying the CG-related serum protein signature and to more precisely understand the complex interactions between the differentially expressed proteins and candidate upstream regulators, we performed functional category, canonical pathway, interactive network, upstream regulator, and regulator effect analyses of all dysregulated proteins with a P value $<.05$ using the Ingenuity Pathway Analysis (IPA) software tool (QIAGEN, Redwood City, CA),[8] a repository of biological interactions and functions created from millions of individually modeled relationships ranging from the molecular (proteins, genes) to organism (diseases) level.

The 63 dysregulated CG-associated proteins with a $P < .05$ were included in further network analysis using the STRING database version 11.0 for protein-protein functional and physical interactions, the results of which were displayed as a functional network.[9] Interactions were considered with a high STRING confidence score of 0.7 or higher garnered from the "experimental" and "databases" categories. Proteins without associations to other proteins in the network were removed. A k-means clustering algorithm was performed to select connected proteins (k-means = 6). Functional description of clusters was assigned based on a manually curated evaluation of enriched Kyoto Encyclopedia of Genes and Genomes pathway, Gene Ontology, Reactome, STRING local network clusters terms, and PubMed literature search.

Table 1. Demographic and Clinical Characteristics of the Population Enrolled

Characteristics	Collagenous gastritis (n = 14)	Gastritis (n = 36)	Nongastritis (n = 25)
Age, y (Std)	13.9 (3.1)	16.0 (4.50)	16.9 (3.53)
Gender, n (%)			
Female	5 (36)	9 (25)	12 (48)
Male	9 (64)	27 (75)	13 (52)
Symptoms			
Anemia	10	1	0
Abdominal pain	10	12	0
GI bleeding	4	1	0
Abdominal perforation	2	0	0
Endoscopy			
Corpus nodular	14	0	1
Antrum nodular	3	0	1
Corpus erythema	0	8[a]	2
Antral erythema	0	10	2
Corpus biopsy			
Collagen deposition	14	0	0
Gastritis	11	7	0
Eosinophils	7	0	0
Antral biopsy			
Collagen deposition	2	0	0
Gastritis	3	35	0
Eosinophils	0	0	0

[a]One gastric ulcer.
GI, gastrointestinal; Std, standard deviation.

Enzyme-Linked Immunosorbent Assay Validation of Proteomic Data

Serum levels of 2 proteins (epidermal growth factor [EGF], FGF19), identified by SOMAscan to be associated with CG, were

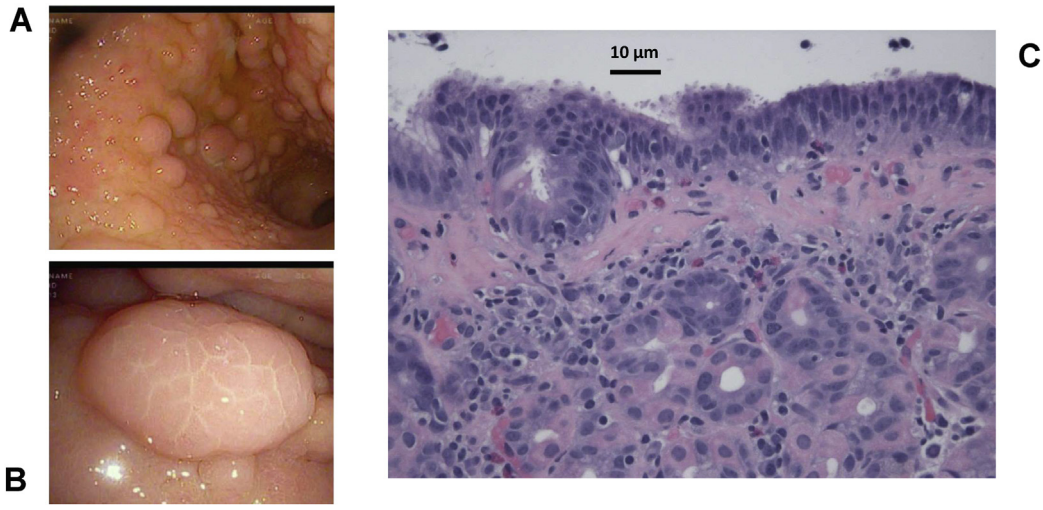

Figure 1. (A) The endoscopic view of the gastric corpus with diffuse nodularity. (B) A large sessile polyp in the gastric corpus characteristic of CG. (C) Hematoxylin and eosin stain of gastric corpus biopsy demonstrating a thickened band of subepithelial collagen (40×).

measured using the fully automated immunoassay platform Ella (ProteinSimple, San Jose, CA) and commercial immunoassays optimized for Ella, following the manufacturer's protocols. Interassay coefficients of variation of triplicate measures were generally <5%. If a coefficient of variation was >10%, the assay was repeated. Assay personnel were blinded to case and control status.

Immunoassay concentrations of the biomarkers were reported as median with interquartile range, and t-test was used to determine the statistical significance of the difference among cases and controls. *P* values < 0.05 were considered significant. For this initial exploratory study, we selected a sample of 75 subjects, 14 with CG, 36 with NCG, and 25 with NG. Given this was an exploratory study, we did not perform a sample size calculation.

Results

Patient Cohort Characteristics

CG is a rare form of gastritis that is usually characterized by signs and symptoms of anemia with or without abdominal pain. The diagnosis is suspected by characteristic nodularity in the gastric corpus (Figure 1A) often with large sessile polyps (Figure 1B) and confirmed by standard criteria with at least 1 area of subepithelial collagen that is ≥10 microns in thickness (Figure 1C).

Demographic information from the 75 subjects whose serum was included in the final analysis is summarized in Table 1.

Subjects with CG had a mean age of 13.9 years, which was slightly less than that of the group with NCG (16.0 years) and the NG group (16.9 years). There was a male predominance in each group. Gastric biopsies from all 14 patients with CG were reviewed by a single pathologist who confirmed the diagnosis. In addition to the collagen deposition in the gastric corpus, 2 patients also had collagen deposition in the antrum. Half of the patients with CG had increased eosinophils in the gastric corpus, but none had increased eosinophils in the antrum. All biopsies were negative for *Helicobacter pylori*. All 14 subjects with CG had nodularity in the gastric corpus, and 3 had nodularity in the antrum. The CG patients did not have other autoimmune disorders, and only 1 patient had an Immunoglobulin A deficiency.

There were 36 subjects who had NCG, 33 had chronic inactive gastritis, 1 had focal active gastritis, 1 had a gastric ulcer, and 1 had *H. pylori*. None of the patients had nodularity in the gastric corpus/antrum or subepithelial collagen. All the patients had antral gastritis. Seven patients had gastritis in the corpus, 11 had normal corpus biopsies, and 17 did not have a biopsy of the corpus. Increased eosinophils were not noted in any of the biopsies from subjects with NCG. The NG group (n = 25) had no gastritis, but they had a variety of associated comorbidities including sinusitis, rectal bleeding, weight loss, gastroesophageal reflux, eosinophilic esophagitis, Pitt-Hopkins syndrome, hereditary alpha tryptasemia, irritable bowel syndrome, mast cell activation syndrome, solitary rectal ulcer, *Clostridium difficile*, juvenile polyposis, tubular adenoma, celiac disease, constipation, functional abdominal pain, or juvenile idiopathic arthritis. One patient had a nodular appearance in the stomach, and 2 patients had gastric erythema, but all patients had normal antral biopsies, and 15 patients had normal biopsies of the corpus.

A correlation analysis between time of diagnosis and blood draw which ranged from 1 month to 162 months was performed, but no significant correlation with EGF levels was found in our cohort of analysis.

SOMAscan Proteomics for Collagenous Gastritis-Related Protein Biomarker Discovery

SOMAscan analysis of 9 pediatric patients with CG, 12 with NCG, and 8 with NG identified 63 out of 1305 plasma

proteins with the greatest degree of differential expression ($P < .05$) in CG cases as compared to patients with NCG or NG; 39 proteins were increased and 24 decreased in CG cases compared to those in patients with NCG or NG. The 63 upregulated and downregulated proteins that were statistically significant ($P < .05$) between CG subjects and NCG/NG patients are listed in Table 2, with the *P* values and mean and median fold changes. Hierarchical clustering of the 17 most significantly differentially expressed proteins ($P < .01$) discriminated well between CG and NCG/NG samples as shown in Figure 2, with 2 CG samples clustering with NCG/NG and 1 NCG sample clustering with CG.

Evaluating in more detail which of these proteins retain statistical significance when comparing CG to NCG alone or CG to NG alone resulted in 15 of the 17, $P < .01$, proteins remaining significant with $P < .05$ (Table A1). Table A2 lists raw data for the entire data set among all participants. Applying unsupervised PCA to all samples using the 17 most dysregulated proteins ($P < .01$) as in Figure 3 resulted in excellent separation of CG cases from NCG cases and NG, with only 2 NCG samples clustering together with the CG phenotype and 1 CG case clustering with NCG/NG.

The first principal component accounts for 41% of the variance, and the second principal component for 12% of the variance. This analysis demonstrates that the SOMAscan-derived proteomics data contain a significant component that differentiates CG cases from NCG/NG cases using the first 2 principal components. A similar PCA using the 15 proteins discriminating between CG and NCG or CG and NCG separated all CG samples from all NCG and NG cases (Figure 4).

Box-whisker plots of the CG, NCG, and NG cases illustrate the degree of difference in SOMAscan relative fluorescence unit expression levels for 13 representative targets: EGF, LTA4H, TIMP2, FGF19,HAMP, C3, SCARF1, CCL25, EPHA2, CTSF, GCG, HIST1H3A, and CNDP1 (Figure 5 and Figure A1). Among these proteins, 9 are increased, and 4 decreased in CG. Several of them such as EGF, GCG, CNDP1, and CCL25 show a gradual expression change from CG to NCG to NG.

Systems Biology Analysis of Pathways Dysregulated in CG

To identify pathways that are dysregulated in CG and to gain new insights into the biological mechanisms associated with CG, we performed an IPA using the 63 CG-associated proteins as input ($P < .05$). Interactive network analysis indicated a link between collagen and multiple of the proteins discriminating CG from NCG and NG such as one of the focus hubs, EGF (Figure A2). Especially informative was modeling the links between CG-associated proteins based on their established connections with predicted upstream regulatory proteins. The predicted upstream regulators with the highest statistical significance converged on activation of proinflammatory and immunoregulatory cytokines (TNF, IL6, IL13, IFNG, IL1B, IL4, IL2, CSF2, CHUK) and extracellular matrix and fibrosis regulators (TGFB1) (Figure 6A), indicating that these signaling nodes may be involved in the dysregulation of a sizeable portion of the top 63 proteins in the serum signature for CG. Key positive regulators were the proinflammatory cytokines TNF and IFNG that are predicted to regulate 22 and 15 of the 63 CG-associated proteins, respectively (Figure 6B). Eighteen of the 63 CG-associated proteins were predicted to be regulated by TGFB1, a key profibrogenic growth factor and potent stimulator of tissue collagen production and deposition (Figure 6B).

IPA for enriched bio functions using the 63-protein CG signature converged on enrichment of proteins associated with enhanced migration of immune cells (leukocyte migration, cell movement of leukocytes, cell movement of phagocytes, migration of phagocytes, cell movement of myeloid cells), decreased vascular functions (vasculogenesis, development of vasculature, development of endothelial tissue, angiogenesis, endothelial cell development, proliferation of endothelial cells, vascularization), enhanced cell death (necrosis, necrosis of epithelial tissue, apoptosis, cell death of epithelial cells), enhanced inflammatory processes (inflammatory response), epithelium (growth of epithelial tissue, development of epithelial tissue, proliferation of epithelial cells), coagulation (thrombus, coagulation of blood, aggregation of blood platelets, binding of blood platelets, activation of blood platelets), lymphatic system (stimulation of lymphatic system cells, quantity of lymphatic system cells, expansion of lymphatic system cells), and enhanced fibrosis (interstitial fibrosis, fibrosis) (Figure 7A). Figure 7B highlights in detail the 22 proteins linked to a predicted enhanced inflammatory response among the 63 CG proteins and the 27 proteins linked to reduced vascularization.

Seventeen proteins were predicted to be linked to fibrosis (Figure A3). Additional relevant pathways associated with immune cell movement, cell death, aggregation of platelets, and thrombus are highlighted in Figure A3. Next, we performed network and cluster analysis using the STRING database of functional and physical protein associations curated across major data repositories. Thirty-six of the 63 proteins formed distinct interacting protein clusters that were enriched in pathways associated with enteroendocrine cell and pancreatic hormones, metabolism, activin signaling, complement and coagulation, insulin-like growth factor (IGF) transport and uptake, iron metabolism, platelet activation, B-cell activation, epithelial-to-mesenchymal transition (EMT), vascularization, bile acid (BA) and glycogen synthesis, extracellular matrix, and collagen turnover and fibrillogenesis.

EGF occupied a key focus hub in this analysis linking complement and coagulation, IGF transport and uptake, B-cell activation, EMT, vascularization, BA and glycogen synthesis, extracellular matrix, and collagen turnover and fibrillogenesis and was itself reduced in serum of pediatric CG patients compared with NCG and NG patients (Table 2 and Table A1).

Table 2. Upregulated and Downregulated Proteins Between Collagenous Gastritis Subjects and Gastritis/Nongastritis Subjects

Increased in collagenous gastritis

SomaId	Target Fullname	Target	Entrez Gene Sym	*P* value	Mean FC	Median FC
SL007100	Leukotriene A-4 hydrolase	LKHA4	LTA4H	.004404	2.68	8.20
SL000313	C3a anaphylatoxin	C3a	C3	.002201	2.01	2.77
SL008945	Protein-glutamine gamma-glutamyltransferase E	TGM3	TGM3	.022971	2.27	1.86
SL008382	Cystatin-D	CYTD	CST5	.002297	1.96	1.85
SL000428	Follicle stimulating hormone	FSH	CGA FSHB	.040340	1.36	1.82
SL000456	Complement C3b, inactivated	iC3b	C3	.017763	1.35	1.65
SL000433	Glucagon	Glucagon	GCG	.000545	1.57	1.64
SL008158	Histone H3.1	H31	HIST1H3A	.000846	1.92	1.58
SL003685	Nicotinamide phosphoribosyltransferase	PBEF	NAMPT	.003429	1.72	1.58
SL003993	Bone morphogenetic protein 6	BMP-6	BMP6	.005889	1.30	1.55
SL003197	C-C motif chemokine 25	TECK	CCL25	.001355	1.66	1.49
SL006694	Beta-Ala-His dipeptidase	CNDP1	CNDP1	.006625	1.48	1.48
SL004337	**Fibroblast growth factor 19**	**FGF-19**	**FGF19**	**.017681**	**1.35**	**1.43**
SL002762	Chromogranin-A	CgA	CHGA	.006364	1.51	1.40
SL001800	Tumor necrosis factor receptor superfamily member 1B	TNF sR-II	TNFRSF1B	.018875	1.21	1.35
SL002654	Ephrin type-A receptor 2	Epithelial cell kinase	EPHA2	.007750	1.31	1.35
SL006523	Lactadherin	MFGM	MFGE8	.019759	1.39	1.34
SL004492	Toll-like receptor 2	TLR2	TLR2	.003427	1.26	1.34
SL005157	CD209 antigen	DC-SIGN	CD209	.006225	1.30	1.33
SL001938	Inhibin beta A chain	Activin A	INHBA	.031879	1.29	1.32
SL006230	Lumican	Lumican	LUM	.034743	1.22	1.30
SL014269	Killer cell immunoglobulin-like receptor 3DL2	KI3L2	KIR3DL2	.025296	1.72	1.30
SL000254	Serum albumin	Albumin	ALB	.015114	1.20	1.28
SL000124	72-kDa Type IV collagenase	MMP-2	MMP2	.023929	1.28	1.26
SL000592	Metalloproteinase inhibitor 2	TIMP-2	TIMP2	.000341	1.23	1.24
SL011499	Pancreatic hormone	PH	PPY	.033683	1.26	1.23
SL000022	D-dimer	D-dimer	FGA FGB FGG	.033727	1.12	1.18
SL007756	Growth/differentiation factor 2	GDF2	GDF2	.037642	1.15	1.17
SL003066	Pigment epithelium-derived factor	PEDF	SERPINF1	.042865	1.14	1.16
SL009324	Follistatin-related protein 3	FSTL3	FSTL3	.035470	1.17	1.15
SL007306	Protein FAM3B	FAM3B	FAM3B	.044092	1.29	1.14
SL005235	Homeobox protein NANOG	NANOG	NANOG	.005030	1.27	1.14
SL014188	Natural killer cell receptor 2B4	CD244	CD244	.035937	1.91	1.14
SL013988	Carbohydrate sulfotransferase 2	CHST2	CHST2	.033292	1.11	1.14
SL003060	Fibroblast growth factor receptor 1	bFGF-R	FGFR1	.023872	1.22	1.14
SL010328	Mediator of RNA polymerase II transcription subunit 1	MED-1	MED1	.035013	1.13	1.13
SL016555	Dual 3′,5′-cyclic-AMP and -GMP phosphodiesterase 11A	PDE11	PDE11A	.037909	1.12	1.13
SL003735	Tumor necrosis factor ligand superfamily member 9	4-1BB ligand	TNFSF9	.013437	1.36	1.10
SL003341	Fibrinogen gamma chain	Fibrinogen g-chain dimer	FGG	.046962	1.22	1.10

Decreased in collagenous gastritis		Target	Entrez Gene Sym	*P* value	Average FC	Median FC
SomaId	Target Fullname					
SL003785	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH, liver	GAPDH	.020790	−5.13	−5.55
SL004536	Hepcidin	LEAP-1	HAMP	.017358	−3.77	−4.17
SL010521	Tyrosine-protein kinase BTK	BTK	BTK	.041451	−2.06	−2.25
SL018938	Small ubiquitin-related modifier 3	SUMO3	SUMO3	.025776	−1.74	−2.24
SL005679	Translationally-controlled tumor protein	TCTP	TPT1	.045976	−1.84	−2.08
SL004869	Carbonic anhydrase 13	Carbonic anhydrase XIII	CA13	.033800	−2.39	−1.90
SL004823	Peptidyl-prolyl cis-trans isomerase A	Cyclophilin A	PPIA	.027756	−1.63	−1.88
SL010516	Proto-oncogene tyrosine-protein kinase Src	SRCN1	SRC	.019832	−1.85	−1.74
SL008759	Platelet glycoprotein VI	GPVI	GP6	.038284	−1.43	−1.71
SL011529	Ribosome maturation protein SBDS	SBDS	SBDS	.047291	−1.84	−1.70
SL006917	Tyrosine-protein kinase Lyn	LYN	LYN	.042330	−1.66	−1.64
SL005588	Tyrosine-protein kinase Fer	FER	FER	.047355	−1.96	−1.64
SL010500	Tyrosine-protein kinase Lyn, isoform B	LYNB	LYN	.031755	−1.65	−1.59
SL010529	Ubiquitin-fold modifier-conjugating enzyme 1	UFC1	UFC1	.033199	−1.38	−1.40
SL008381	Cathepsin F	CATF	CTSF	.005022	−1.31	−1.32
SL003862	CD40 ligand	CD40 ligand, soluble	CD40LG	.048036	−1.38	−1.29
SL005221	Scavenger receptor class F member 1	SREC-I	SCARF1	.015007	−1.29	−1.23
SL005629	N-acetyl-D-glucosamine kinase	NAGK	NAGK	.037621	−1.29	−1.22
SL013488	C-type lectin domain family 1 member B	CLC1B	CLEC1B	.046526	−1.31	−1.22
SL007310	Ras-related C3 botulinum toxin substrate 3	RAC3	RAC3	.046616	−1.30	−1.19
SL010512	PIK3CA/PIK3R1	PIK3CA/PIK3R1	PIK3CA PIK3R1	.031175	−1.25	−1.19
SL000049	Vitamin K-dependent protein S	Protein S	PROS1	.047175	−1.15	−1.17
SL012561	Regenerating islet-derived protein 4	REG4	REG4	.038200	−1.18	−1.14
SL000084	**Epidermal growth factor**	**EGF**	**EGF**	**.001553**	**−1.31**	**−1.13**

Proteins measured by ELISA are highlighted in bold.
FC, fold change.

Immunoassay Validates Reduced Expression of EGF in Patients With Collagenous Gastritis

To confirm the results from the SOMAscan analysis with an orthogonal, gold standard immunoassay technology, we performed immunoassays for absolute quantification on the fully automated Ella platform using an expanded cohort of patients. Selection criteria for the proteins tested by enzyme-linked immunosorbent assay were based on statistical significance in the SOMAscan data and, most importantly, availability of enzyme-linked immunosorbent assay from a reputable commercial provider. Secondary criteria were biological plausibility which finally resulted in selection of EGF and FGF19 for further evaluation.

We selected EGF, which SOMAscan showed to be decreased with statistical significance ($P = .0016$) in CG as compared to that in NCG and NG for further validation based on the STRING analysis placing EGF as a focus hub linking various enriched pathways impacted by CG as well as its known role in regulation of collagenases and collagen expression levels (Figure 8).[10–12]

As shown in Table A3 and Figure 9, the immunoassay data for EGF demonstrated a significant, median ∼2-fold decrease of EGF expression in CG cases (N = 14) compared to both patients with NCG (N = 36, $P = .016$) or NG (N = 25, $P = .047$), demonstrating that the SOMAscan data accurately reflect the relative protein expression levels.

FGF19, shown by SOMAscan to be significantly increased in CG as compared to NCG and NG, was picked as a second protein for immunoassay analysis. While FGF19 concentrations showed a trend to increase in CG compared to NCG (Table A3 and Figure 9), neither comparison resulted in

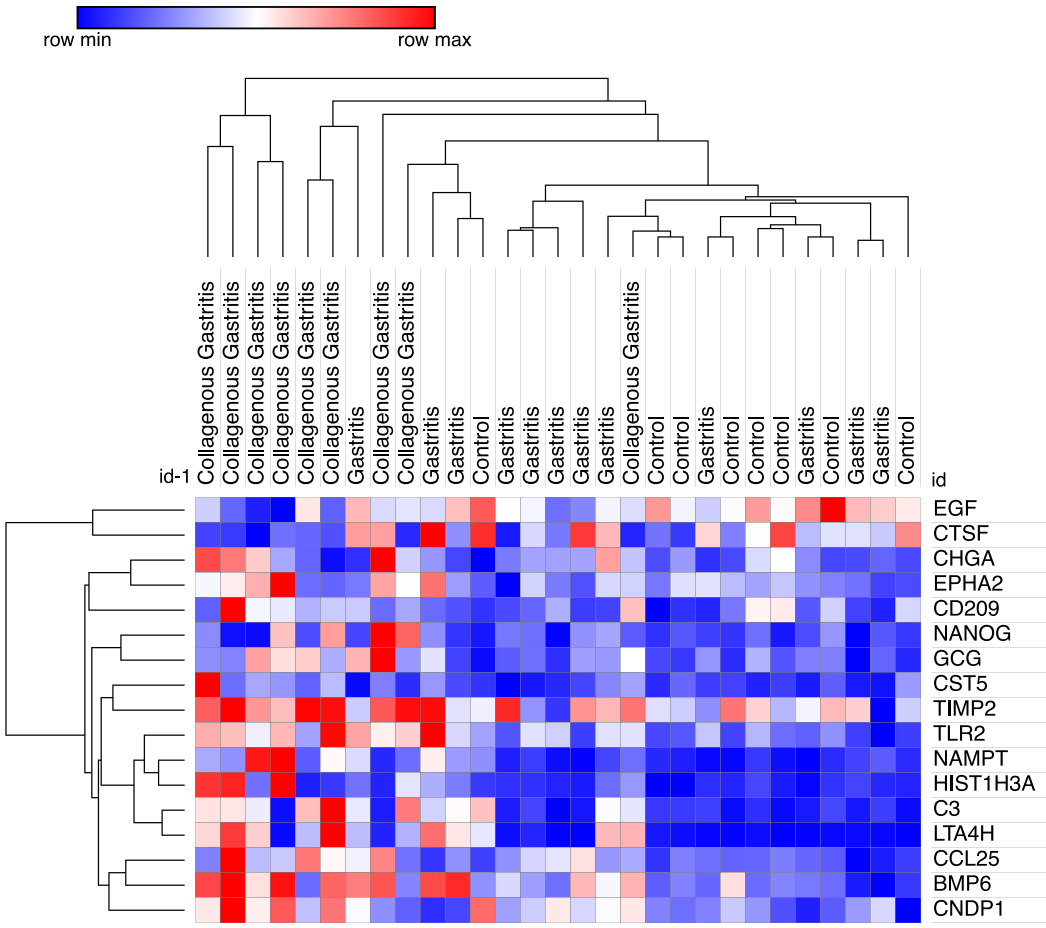

Figure 2. Hierarchical clustering (HC) of 17 proteins ($P < .01$) for CG vs NCG vs NG. In the HC colormap, red denotes upregulation, and blue denotes downregulation. The relative color scheme uses the minimum and maximum values in each row to convert values to colors.

statistical significance ($P > .05$) within this relatively small patient cohort.

Discussion

CG is an extremely rare, chronic GI disease with both adult- and childhood-onset. Little is known about the etiopathology and pathophysiological mechanisms underlying CG development. We believe that this study is the first to identify circulating serum proteins and pathways associated with pediatric CG, providing new insights into biological pathways involved in CG pathogenesis. Using SOMAscan proteomics to survey 1305 proteins, we provide the first evidence that pediatric-onset CG has a distinct serum protein profile. We identified 63 proteins (P value $< .05$) by unbiased proteomics discovery whose expression levels were consistently altered between CG cases and NCG and NG cases. Among these 63 proteins, we identified EGF as a novel, potentially clinically relevant circulating biomarker for CG. EGF is a ubiquitous protein that is associated with other inflammatory conditions. Recent works showed a significant reduction of urinary EGF levels at the very early stage of lupus nephritis and in patients with antineutrophil cytoplasmic antibody-associated vasculitis suggesting its possible role in the early management and monitoring of disease progression.[13,14] In the context of CG, the following aspects support the hypothesis that EGF may play a pathophysiological role: (1) EGF, one of the statistically most significant regulators emerging from our analysis, was found by SOMAscan to be decreased in CG compared to NCG and NG controls; (2) reduced EGF expression in CG was validated by an orthogonal, gold standard immunoassay method on an expanded study cohort; (3) STRING analysis places EGF as a central focus hub that interacts with various biological pathways (extracellular matrix collagen turnover and fibrillogenesis, vascularization, BA and glycogen synthesis, EMT, IGF transport/uptake coagulation; complement); (4) under several pathophysiological conditions, EGF has been reported to induce collagen proteolysis due to its enhancement of collagenase expression, while at the same time reducing collagen gene transcription.[10,15] Taken together, these results support the concept that suppression of EGF is a pivotal contributor to the pathophysiology and collagen deposition in CG.

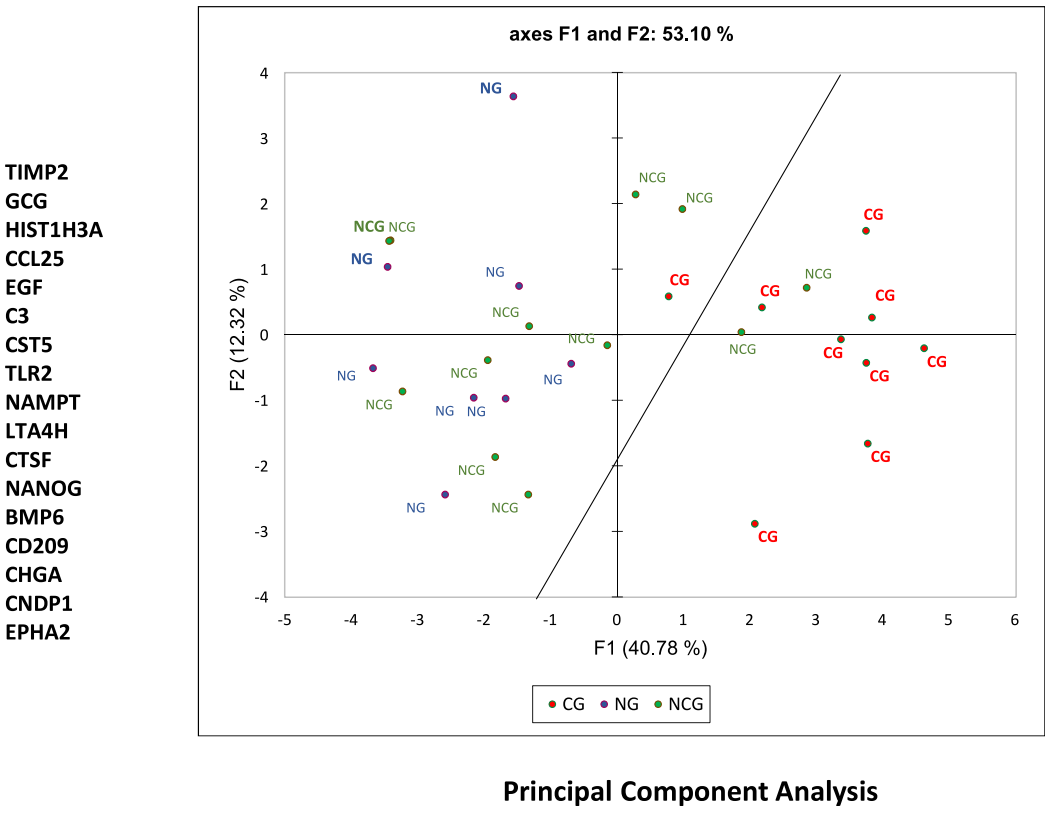

Figure 3. Discrimination between CG, NCG, and NG cases. PCA plot, using the top 17 proteins from SOMAscan analysis, demonstrates excellent separation into 2 clusters. Red circles, CG patients; blue circles, NG; and green circles, NCG.

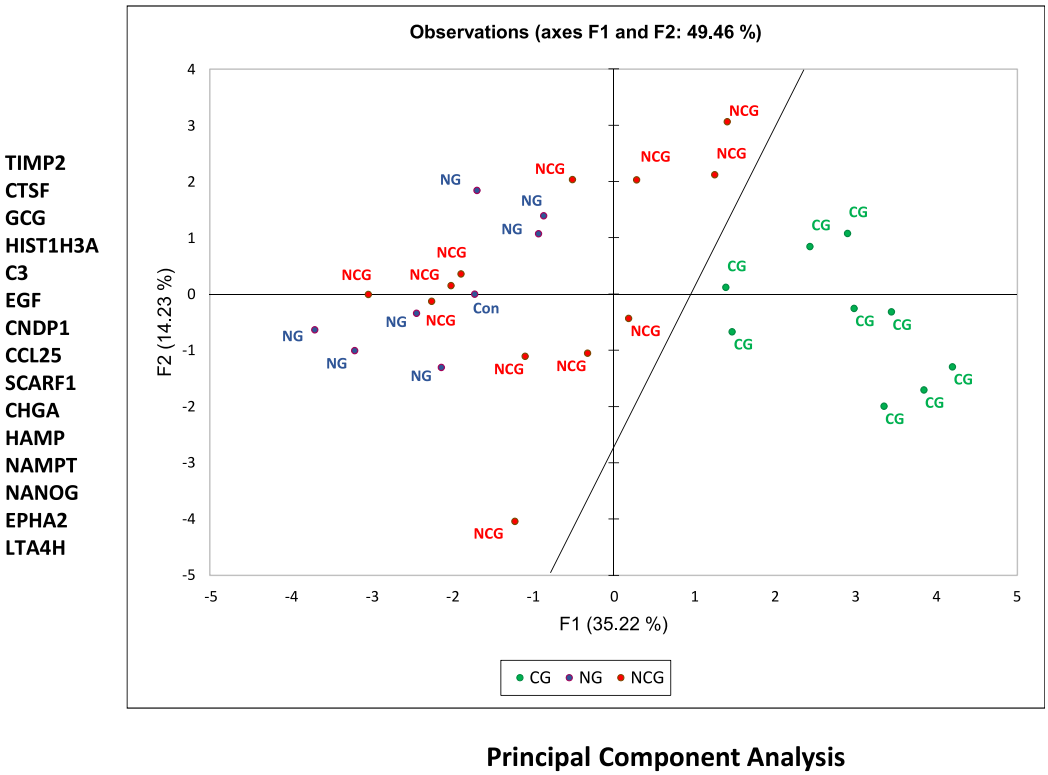

Figure 4. PCA plot, using the 15 proteins that are statistically significant when comparing CG to NCG or CG to NG, demonstrates excellent separation into 2 clusters. Red circles, CG patients; blue circles, NG; and green circles, NCG patients.

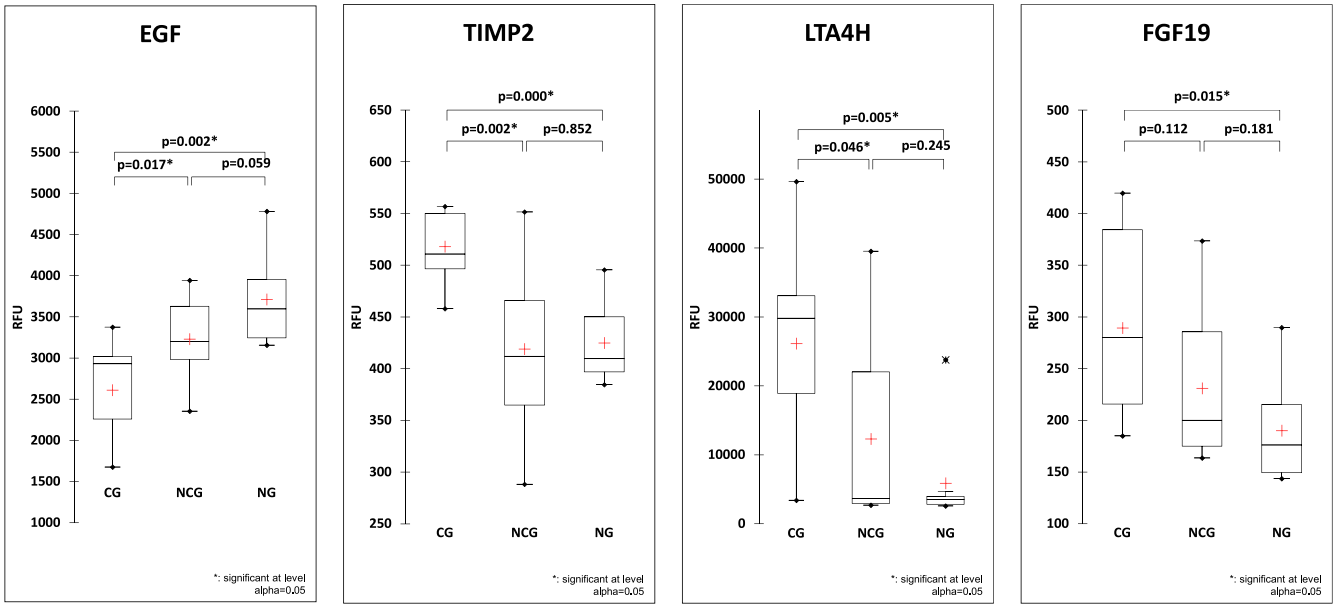

Figure 5. Box-whisker plots for differentially expressed proteins in CG, NCG, and NG cases. Expression pattern of 4 representative proteins in CG, NCG, and NG cases among the top dysregulated proteins ($P < .05$), using relative fluorescence units derived from SOMAscan. Data are shown as a box-whisker plot with mean expression depicted as + and median expression indicated by a horizontal line.

This proteomic study contributes important new insights into dysregulated pathways and expression of proteins associated specifically with CG with a specific focus on the apparent involvement of EGF. EGF is a multifunctional growth factor which, in addition to its effect on collagen, contributes among many other functions to epithelial development, inhibition of gastric acid secretion, acceleration of wound healing, and promotion of angiogenesis.[16] Decreased serum EGF levels are observed in patients with inflammatory bowel disease. This downregulation may be correlated with different patterns of bowel inflammation, epithelial development, and wound healing in inflammatory bowel disease.[16] Once the surface epithelium is injured, epithelial restitution and epithelial proliferation occur. EGF is known to enhance epithelial cell restitution, through a transforming growth factor beta-dependent pathway, and to stimulate epithelial cell proliferation.[17] However, deeper lesions require additional repair mechanisms, including inflammatory processes, angiogenesis, and the deposition of extracellular matrix components.[17] Our results have shown a significantly lower serum EGF levels in CG than in controls with NCG/NG. No difference in EGF was observed between NCG patients and NG patients which may be explained by different pathways of inflammation and restoration between CG and NCG patients and the collagen production and deposition unique to CG.

In addition to EGF, these SOMAscan results provided important new knowledge about potential mechanisms involved in CG pathogenesis. Using systems biology approaches, significant enrichment of CG-linked proteins in several biological functions was observed. Specifically, an enhanced inflammatory response, immune cell movement, cell death, and fibrosis as well as reduced vascular functions and platelet aggregation were predicted to involve a major portion of the 63 differentially expressed CG proteins. This was also reflected in the predicted upstream regulators with statistically significant enrichment of CG-associated proteins that included various proinflammatory cytokines such as TNF, IL6, and IFNG.

The pathogenic mechanisms responsible for subepithelial collagen deposition in enteric mucosa remain unclear. Several authors hypothesized that the changes of collagenous colitis were linked to decreased turnover of the fibroblasts of the colonic pericryptal fibroblast sheath[18] or that the surface epithelium of genetically or metabolically susceptible patients may be vulnerable to intraluminal toxic substances, resulting in reparative fibrosis limited to the superficial zone.[19] In contrast, in CG, a prominent eosinophilic infiltrate (noted in half of the CG subjects in this study) is noted with the implication that degranulation could cause tissue damage and the proliferation of fibroblasts. The presence of a repair type of collagen might be the result of activated immune cells stimulating the production of extracellular matrix.[20] The reason for this heterogeneity remains unclear, but it might be related to the etiology of gastritis.[21] TGFB1 was predicted by IPA analysis as a statistically significant upstream regulator of multiple proteins from the CG-linked proteins with enhanced activity. TGFB1 is a growth factor that plays a major role in collagen expression and fibrosis in various pathological conditions.[22,23]

Another protein with increased expression in CG based on SOMAscan analysis, FGF19, shows a similar trend of increased expression by immunoassay analysis of an

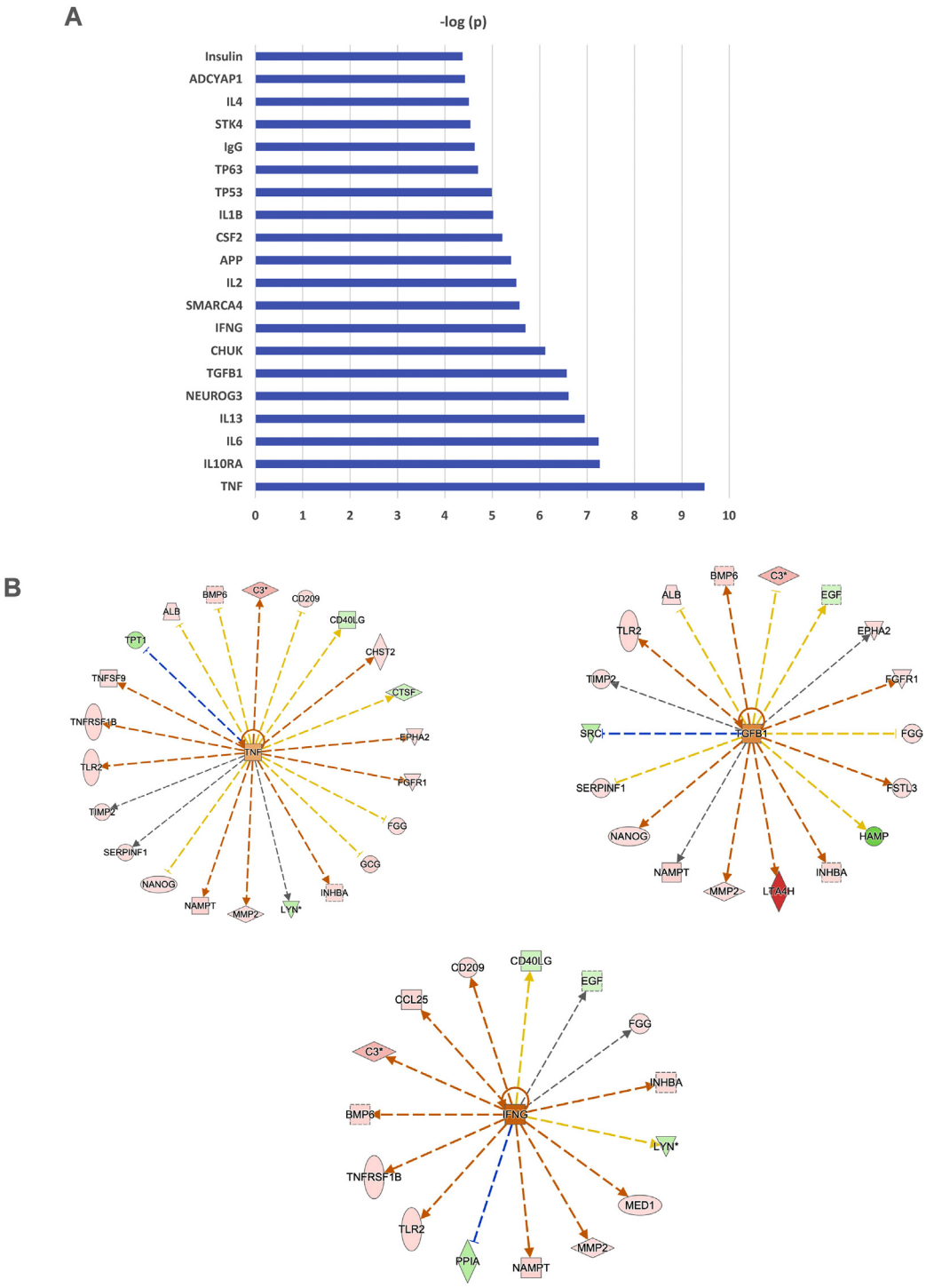

Figure 6. Systems biology analysis of the significant differentially expressed proteins was conducted using Ingenuity Pathway Analysis. (A) Upstream transcriptional regulators that best explain the observed expression changes in CG. (B) Upstream Regulator Analysis of TNF, TGFB, and IFNG interactions: upregulation (red) and downregulation (green) of the targets. Protein identifiers marked with an asterisk indicate that duplicate identifiers in the input protein list map to a single protein in the Global Molecular Network.

expanded set of samples but did not reach statistical significance. Recent studies in humans have shown a relationship among FGF19 levels, BA synthetic rates, and intestinal inflammation.[24,25] In particular, BAs have potent secretory effects on the colonic mucosa. When an excess of BA reaches the small bowel, the ileal capacity for BA absorption may be overwhelmed resulting in BAs spilling into the colonic lumen and accelerating colonic transit stimulating the colonic motility, secretion, and diarrhea. The farnesoid X receptor (FXR) is a nuclear receptor highly

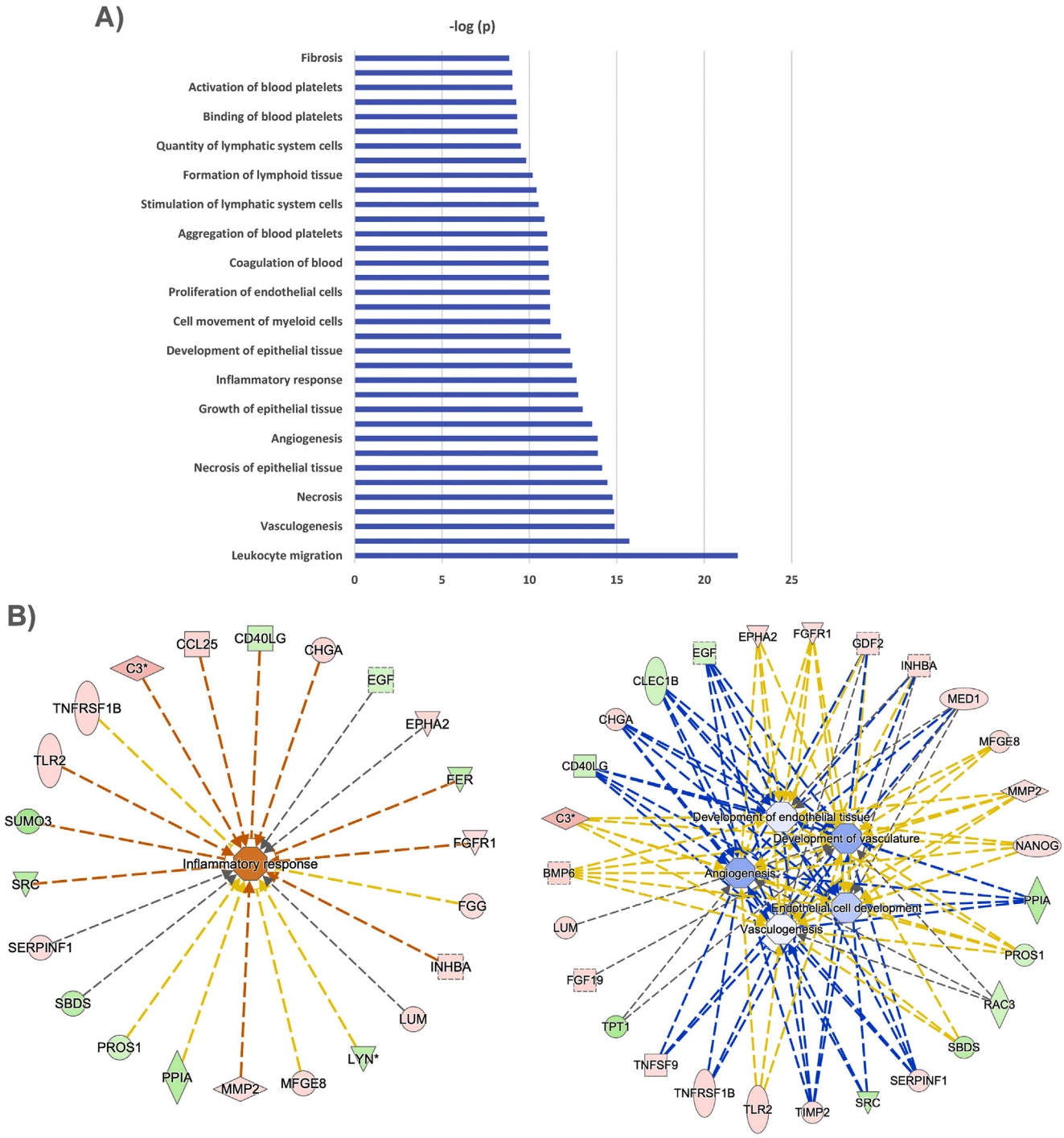

Figure 7. Systems biology analysis of the significant differentially expressed proteins was conducted using Ingenuity Pathway Analysis. (A) Biological functions that are significantly enriched by the input protein list. (B) Inflammatory response and vascular functions. Protein identifiers marked with an asterisk indicate that duplicate identifiers in the input protein list map to a single protein in the Global Molecular Network.

expressed in the gut-liver axis that acts as the master regulator of BA homeostasis, mainly via the gut hormone FGF19. FGF19 activates signaling pathways that inhibit BA synthesis, gluconeogenesis, and fatty acid oxidation, while promoting glycogen synthesis.[26] Strategies aimed at restoring BA homeostasis through activation of the intestinal FXR-FGF19 signaling system seems to be protective, but in our results, patients with CG have higher levels of FGF19 than NCG and NG patients. Most likely as reported by Lagorce-Pages et al,[27] clinical symptoms differ between pediatric and adult patients based on the severity of the disease and part of the GI tract involved. We hypothesize in CG patients, a disrupted BA metabolism, with an increased FXR nuclear receptor expression, leads to increased FGF19

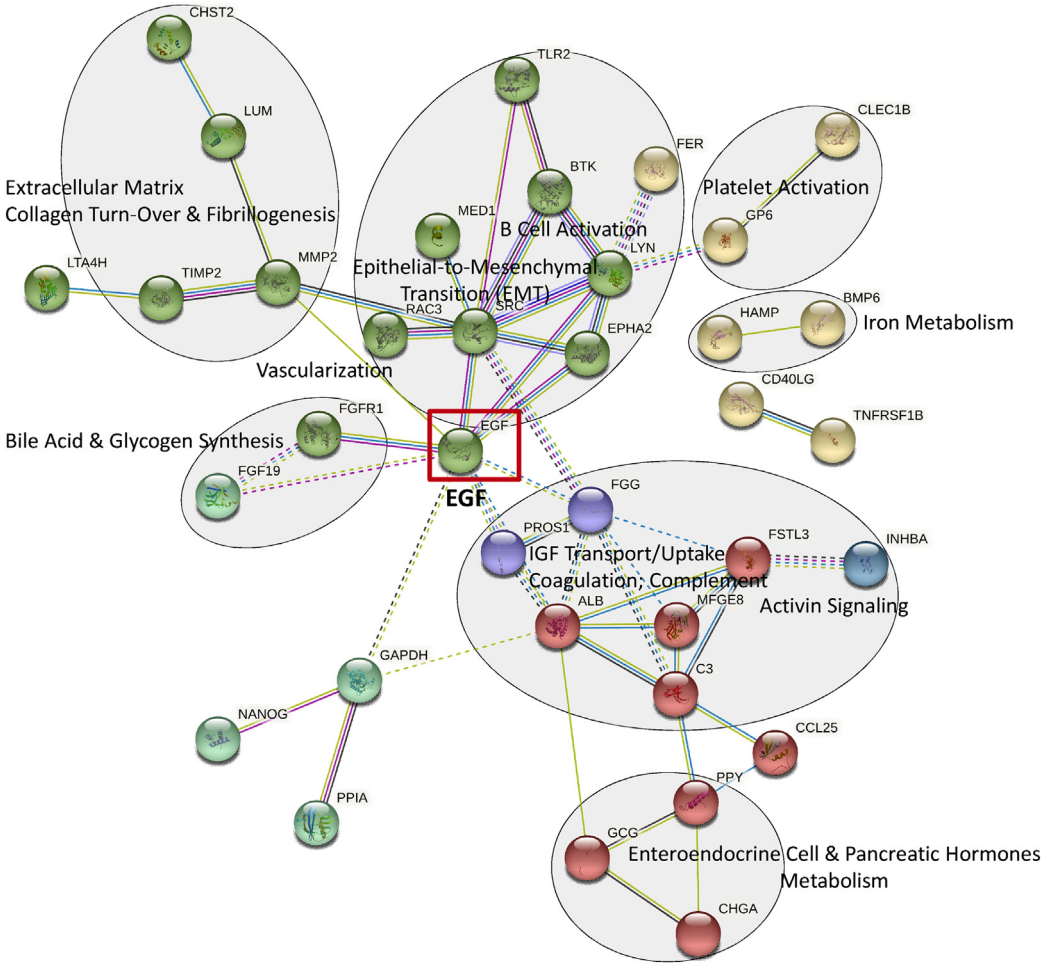

Figure 8. Network and cluster analysis using the STRING database of functional and physical protein associations.

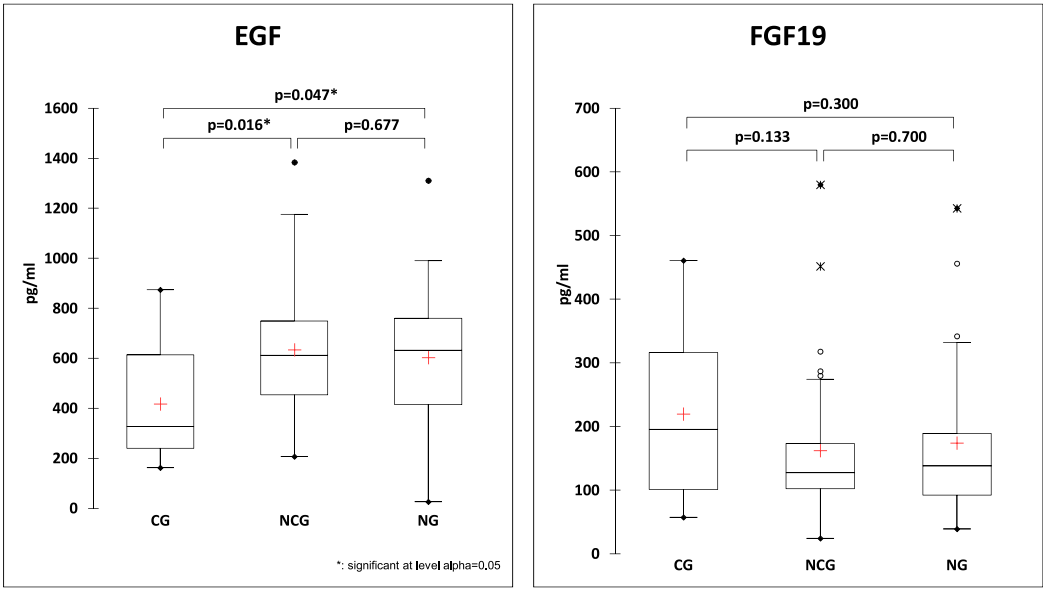

Figure 9. Box-whisker plots for EGF and FGF19 protein expression levels in CG, NCG, and NG cases. Data derived are shown as a box-whisker plot with mean expression depicted as + and median expression indicated by a horizontal line. The expression levels of EGF and FGF19 were determined by enzyme-linked immunosorbent assay from patient serum.

levels. Nevertheless, the potential role of FGF19 remains unclear.

Additional significantly dysregulated proteins in CG included CDNP1 and the transglutaminase TGM3. Histidine-containing dipeptides such as carnosine are key factors for protection against oxidate stress by reactive oxygen species and carbonyl scavenging.[28] Carnosine is predominantly degraded by CNDP1, a dipeptidase released by the liver and other organs that has been linked to diabetic nephropathy.[29,30] While the potential role of CNDP1 in CG is unknown, its overexpression in CG may result in reduced protection of the gastric mucosa against oxidative stress. Transglutaminases are known to help stabilizing collagen by crosslinking.[31] TGM3 is expressed in various epithelial tissues including the stomach, and its crosslinking activity may contribute to the CG phenotype.

Our study has several notable strengths. The first use of SOMAscan deep proteomics enabled measurement of over 1300 proteins for detecting and conducting high-dimensional functional characterization of pathophysiological correlates of CG—a platform that has been used to identify proteins associated with various diseases.[32–36] Functional enrichment analysis using IPA highlighted key upstream regulators and biological functions that appear to be associated with CG. This includes inflammation, other immune pathways, and vascular function.

We acknowledge some limitations of our study. First, the sample size is relatively small due to the rarity of CG. Nevertheless, this study involces the largest number of pediatric CG patients evaluated in any study of pathophysiology. Despite this, our EGF findings were validated by an orthogonal immunoassay method in a somewhat larger cohort. This finding supports the feasibility and efficiency of using a broad proteomics platform for the initial biomarker discovery phase. Second, the SOMAscan cohort is a subset of the Ella immunoassay validation cohort; thus, the validation set is not entirely independent. Third, our cohort included only pediatric-onset CG patients. Thus, our results may not be generalizable to patients with adult-onset CG.

In conclusion, the identification of EGF using SOMAscan combined with immunoassay validation places EGF as a promising biomarker of CG. The emergence of EGF from a comprehensive proteomic analysis points to a potential link to etiopathogenic mechanisms of CG. Although clinical relevance of reduction in EGF expression for CG requires further investigation and functional validation, our data include the first comprehensive analysis of the proteome in CG, reveal a new potential role of EGF in CG pathophysiology, and could lead to discovery of novel therapeutic targets.

Supplementary Materials

Material associated with this article can be found in the online version at https://doi.org/10.1016/j.gastha.2022.04.016.

References

1. Arnason T, Brown IS, Goldsmith JD, et al. Collagenous gastritis: a morphologic and immunohistochemical study of 40 patients. Mod Pathol 2015;28:533–544.
2. Ma C, Park JY, Montgomery EA, et al. A comparative clinicopathologic study of collagenous gastritis in children and adults: the same disorder with associated immune-mediated diseases. Am J Surg Pathol 2015;39:802–812.
3. Beinvogl BC, Goldsmith JD, Verhave M. Pediatric collagenous gastritis: clinical and histologic outcomes in a large pediatric cohort. J Pediatr Gastroenterol Nutr 2021;73:513–519.
4. Gopal P, McKenna BJ. The collagenous gastroenteritides: similarities and differences. Arch Pathol Lab Med 2010;134:1485–1489.
5. Rohloff JC, Gelinas AD, Jarvis TC, et al. Nucleic acid ligands with protein-like side chains: modified aptamers and their use as diagnostic and therapeutic agents. Mol Ther Nucleic Acids 2014;3:e201.
6. Kim CH, Tworoger SS, Stampfer MJ, et al. Stability and reproducibility of proteomic profiles measured with an aptamer-based platform. Sci Rep 2018;8:8382.
7. Vasunilashorn SM, Dillon ST, Chan NY, et al. Proteome-wide analysis using SOMAscan identifies and validates chitinase-3-like protein 1 as a risk and disease marker of delirium among older adults undergoing major elective surgery. J Gerontol A Biol Sci Med Sci 2022;77:484–493.
8. Krämer A, Green J, Pollard J Jr, et al. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics 2014;30:523–530.
9. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019;47:D607–D613.
10. Mimura Y, Ihn H, Jinnin M, et al. Epidermal growth factor affects the synthesis and degradation of type I collagen in cultured human dermal fibroblasts. Matrix Biol 2006;25:202–212.
11. Laato M, Kähäri VM, Niinikoski J, et al. Epidermal growth factor increases collagen production in granulation tissue by stimulation of fibroblast proliferation and not by activation of procollagen genes. Biochem J 1987;247:385–388.
12. Grande JP, Melder DC, Zinsmeister AR. Modulation of collagen gene expression by cytokines: stimulatory effect of transforming growth factor-beta1, with divergent effects of epidermal growth factor and tumor necrosis factor-alpha on collagen type I and collagen type IV. J Lab Clin Med 1997;130:476–486.
13. Mejia-Vilet JM, Shapiro JP, Zhang XL, et al. Association between urinary epidermal growth factor and renal prognosis in lupus nephritis. Arthritis Rheumatol 2021;73:244–254.
14. Wu L, Li XQ, Goyal T, et al. Urinary epidermal growth factor predicts renal prognosis in antineutrophil cytoplasmic antibody-associated vasculitis. Ann Rheum Dis 2018;77:1339–1344.

15. Rao VH, Kansal V, Stoupa S, et al. MMP-1 and MMP-9 regulate epidermal growth factor-dependent collagen loss in human carotid plaque smooth muscle cells. Physiol Rep 2014;2:e00224.
16. Oikonomou KA, Kapsoritakis AN, Kapsoritaki AI, et al. Downregulation of serum epidermal growth factor in patients with inflammatory bowel disease. Is there a link with mucosal damage? Growth Factors 2010; 28:461–466.
17. Sturm A, Dignass AU. Epithelial restitution and wound healing in inflammatory bowel disease. World J Gastroenterol 2008;14:348–353.
18. Groisman GM, Meyers S, Harpaz N. Collagenous gastritis associated with lymphocytic colitis. J Clin Gastroenterol 1996;22:134–137.
19. Eckstein RP, Dowsett JF, Riley JW. Collagenous enterocolitis: a case of collagenous colitis with involvement of the small intestine. Am J Gastroenterol 1988; 83:767–771.
20. Leung ST, Chandan VS, Murray JA, et al. Collagenous gastritis: histopathologic features and association with other gastrointestinal diseases. Am J Surg Pathol 2009;33:788–798.
21. Kamimura K, Kobayashi M, Sato Y, et al. Collagenous gastritis: review. World J Gastrointest Endosc 2015; 7:265–273.
22. Verrecchia F, Mauviel A. Transforming growth factor-beta and fibrosis. World J Gastroenterol 2007; 13:3056–3062.
23. Pan X, Chen Z, Huang R, et al. Transforming growth factor β1 induces the expression of collagen type I by DNA methylation in cardiac fibroblasts. PLoS One 2013; 8:e60335.
24. Nolan JD, Johnston IM, Walters JR. Altered enterohepatic circulation of bile acids in Crohn's disease and their clinical significance: a new perspective. Expert Rev Gastroenterol Hepatol 2013; 7:49–56.
25. Walters JR. Bile acid diarrhoea and FGF19: new views on diagnosis, pathogenesis and therapy. Nat Rev Gastroenterol Hepatol 2014;11:426–434.
26. Gadaleta RM, Moschetta A. Metabolic messengers: fibroblast growth factor 15/19. Nat Metab 2019; 1:588–594.
27. Lagorce-Pages C, Fabiani B, Bouvier R, et al. Collagenous gastritis: a report of six cases. Am J Surg Pathol 2001;25:1174–1179.
28. Prokopieva VD, Yarygina EG, Bokhan NA, et al. Use of carnosine for oxidative stress reduction in different pathologies. Oxid Med Cell Longev 2016;2016:2939087.
29. Ahluwalia TS, Lindholm E, Groop LC. Common variants in CNDP1 and CNDP2, and risk of nephropathy in type 2 diabetes. Diabetologia 2011;54:2295–2302.
30. Bellia F, Vecchio G, Rizzarelli E. Carnosinases, their substrates and diseases. Molecules 2014;19: 2299–2329.
31. Chau DY, Collighan RJ, Verderio EA, et al. The cellular response to transglutaminase-cross-linked collagen. Biomaterials 2005;26:6518–6529.
32. Shi L, Westwood S, Baird AL, et al. Discovery and validation of plasma proteomic biomarkers relating to brain amyloid burden by SOMAscan assay. Alzheimers Dement 2019;15:1478–1488.
33. Raffield LM, Dang H, Pratte KA, et al. Comparison of proteomic assessment methods in multiple cohort studies. Proteomics 2020;20:e1900278.
34. Shimada YJ, Hasegawa K, Kochav SM, et al. Application of proteomics profiling for biomarker discovery in hypertrophic cardiomyopathy. J Cardiovasc Transl Res 2019;12:569–579.
35. Giudice V, Biancotto A, Wu Z, et al. Aptamer-based proteomics of serum and plasma in acquired aplastic anemia. Exp Hematol 2018;68:38–50.
36. Qiao Z, Pan X, Parlayan C, et al. Proteomic study of hepatocellular carcinoma using a novel modified aptamer-based array (SOMAscanTM) platform. Biochim Biophys Acta Proteins Proteom 2017;1865:434–443.

Received November 8, 2021. Accepted April 22, 2022.

Correspondence:
Address correspondence to: Towia A. Libermann, PhD, Beth Israel Deaconess Medical Center, 99 Brookline Avenue, Boston, Massachusetts 02115. e-mail: tliberma@bidmc.harvard.edu.

Acknowledgment:
The authors thank Emma X. Li and Catherine Pursley for contributing to sample collection and management of the biorepository. Figure 1 was taken by Kenan Sauder, MD.

Authors' Contributions:
Debora Curci: Writing – original draft, Writing – review & editing. Simon T. Dillon: Investigation, performed research, analysis and/or interpretation of data. Xuesong Gu: Investigation, performed research, analysis and/or interpretation of data. Harland Winter: Conception and design of the study; generation, collection, assembly, analysis, and/or interpretation of data; drafting or revision of the manuscript; approval of the final version of the manuscript. Towia A. Libermann: Conception and design of the study; generation, assembly, analysis, and/or interpretation of data; drafting or revision of the manuscript; approval of the final version of the manuscript.

Conflicts of Interest:
The authors disclose no conflicts.

Funding:
Harland Winter, MD, has received grant support from Janssen, Nutricia, Nestle, Pediatric IBD Foundation, Women's Wellness Foundation, QOL, Autism Research Foundation, Scipher, Abbvie, and Crohn's & Colitis Foundation. He has served as a Consultant to Finch Therapeutics and as a member of Data Safety Monitoring Boards for Pfizer and Janssen. Dr Winter received philanthropic support from Martin Schlaff and the Diane and Dorothy Brooks Foundation. The funding sources had no role in the design, conduct, or reporting of this study.

Ethical Statement:
The corresponding author, on behalf of all authors, jointly and severally, certifies that their institution has approved the protocol for any investigation involving humans or animals and that all experimentation was conducted in conformity with ethical and humane principles of research.

Data Transparency Statement:
The data, analytic methods, and study materials used to support the findings of this study are available from the corresponding author upon request.