Abstract

Recently, the field of proteomics has rapidly expanded in its application towards clinical research with objectives ranging from elucidating disease pathogenesis to discovering clinical biomarkers. As proteins govern and/or reflect underlying cellular processes, the study of proteomics provides an attractive avenue for research as it allows for the rapid identification of protein profiles in a biological sample. Inflammatory bowel disease (IBD) encompasses several heterogeneous and chronic conditions of the gastrointestinal tract. Proteomic technology provides a powerful means of addressing major challenges in IBD today, especially for identifying biomarkers to improve its diagnosis and management. This review will examine the current state of IBD proteomics research and its use in biomarker research. Furthermore, we also discuss the challenges of translating proteomic research into clinically relevant tools. The potential application of this growing field is enormous and is likely to provide significant insights towards improving our future understanding and management of IBD.

Key words: Proteomics; Inflammatory bowel disease; Biomarkers; Molecular diagnostic techniques; Mass spectrometry

Core tip: Proteomic methods provide a powerful tool that can be applied to the discovery of disease markers, allowing for rapid identification and quantification of proteins. Inflammatory bowel disease (IBD) currently faces many challenges, ranging from the elucidation of its pathophysiology to the accurate diagnosis in patients. Proteomics has been widely employed in many disease in the search of biomarkers, particularly cancer proteins. It has great potential to improve both our understanding and clinical management of IBD. Our review summarises the current application of proteomics to IBD and discusses challenges relating to translation into clinical practice.
INTRODUCTION

Inflammatory bowel disease (IBD) encompasses a group of conditions characterised by chronic gastrointestinal inflammation, with the two major subtypes being Crohn’s disease (CD) and ulcerative colitis (UC). Differentiating between subtypes of IBD sometimes has a degree of uncertainty due to overlapping clinical and pathological features\(^\text{(1)}\). Despite clinical evaluation, radiological, endoscopic and histopathological testing by expert physicians, up to 20% of IBD cases are classified as "indeterminate colitis" or "IBD undifferentiated\(^\text{(2,3)}\). However, accurate classification of IBD is essential as response to medication, surgical indications and prognosis can vary between UC and CD\(^\text{(4)}\). The field of proteomics is a rapidly expanding area of research that has been employed in many diseases such as cancer\(^\text{(5,6)}\), exploring everything from understanding disease pathways to discovering diagnostic markers\(^\text{(7-9)}\). This review examines the current state of biomarkers in IBD, with particular reference to the application of proteomics.

CURRENT BIOMARKERS IN IBD

Biomarkers are measureable substances that can objectively evaluate either physiological processes or therapeutic outcomes\(^\text{(10)}\) and could potentially play a pivotal role in IBD as cheap and non-invasive alternatives to endoscopy\(^\text{(11)}\). Different biomarkers could be beneficial across all aspects of IBD (Illustrated in Figure 1)\(^\text{(12)}\). The major commercially available biomarkers are summarised below based on their application in Table 1. Whilst some of these biomarkers demonstrate high diagnostic accuracy, they are currently unable to replace endoscopy entirely and limited only to being adjuncts\(^\text{(11,13)}\). Therefore, there is a prevailing need for the development of additional non-invasive biomarkers that are sufficiently sensitive and specific in the diagnosis and prognosis of IBD.

PROTEOMICS

The term "proteome" was initially defined as the total protein complement encoded by a given genome\(^\text{(14)}\) but now also encompasses any isoforms, post-translational modifications, interactions and effectively anything "post-genomic"\(^\text{(15)}\). The study of proteomics involves large scale detection, identification and characterisation of proteins, making it highly promising for biomarker discovery across many diseases\(^\text{(16)}\). The most common method applied is a combination of two-dimensional electrophoresis (2-DE) and mass-spectrometry. 2-DE provides a powerful tool isolating proteins that differ in abundance between cases and controls\(^\text{(17)}\). Mass spectrometry can then identify proteins utilising techniques such as “surface enhanced laser desorption/ionisation time-of-flight” (SELDI-TOF) or “matrix-assisted laser desorption/ionisation time-of-flight” (MALDI-TOF).

Both these technique involve fragmentation of proteins into peptides, determining their mass-to-charge ratio based on their “time-of-flight” within an electric field and comparing their peptide mass signatures to a database of known proteins to identify the original protein.

Although mass spectrometry is not inherently quantitative, many methods have been developed to achieve accurate quantitative data\(^\text{(17,18)}\). The crux of selecting candidate biomarkers in proteomic studies rely detecting differences in abundances between cases and controls; therefore quantitative proteomics is an essential aspect. Multiple reaction monitoring (MRM) is a quantitative technique that achieves absolute quantitation and has a relatively high sensitivity when detecting peptides in low abundance, suitting it towards application in proteomic biomarker studies\(^\text{(19)}\).

APPLYING PROTEOMICS TO IBD

The process leading up to clinical implementation of a novel proteomic biomarker can be divided into three major stages of a pipeline: Discovery, verification and validation, which all vary in both aim and study design (Figure 2)\(^\text{(20)}\). At present, the application of proteomics in IBD (and many other diseases) remains largely in its infancy in the initial discovery phase. This stage involves the rapid analysis of entire protein profiles within a target sample (e.g., plasma from an IBD patient), to screen for proteins that have relative differences in abundance compared to control samples\(^\text{(21)}\). The main disadvantage however, is that these discovery experiments do not provide absolute quantification and are labour intensive (and therefore typically have small sample sizes). The “verification” and "validation" stages addresses these issues by confirming the presence of and quantifying candidate markers in larger populations to assess their value in clinical usage.

Biomarker discovery studies

Proteomic studies involving IBD biomarkers have been divided into those relating to diagnosis and those pertaining to disease characteristics.

The most common approach towards biomarker discovery in proteomics involves assessing relative differences in proteins between cases and controls, for example, identifying which protein is differentially expressed between IBD patients and healthy controls. Furthermore, with the common objective of developing a clinically relevant assay, many groups have analysed
plasma/serum for candidate markers (summarised in Table 2).

In 2007, Meuwis et al.[22] reported a proteomic profile detected with SELDI-TOF MS that could discriminate active UC and CD with a high sensitivity and specificity, performing similarly or better than current ANCA and ASCA serology. From the protein spectra detected, platelet factor 4, myeloid related protein 8, fibrinopeptide A and haptoglobin α2 were considered diagnostically important.

Kanmura et al.[23] examined UC serum samples using SELDI-TOF MS and identified that human neutrophil peptide (HNP) 1-3 was differentially expressed. HNP 1-3 was confirmed by ELISA to differentiate active UC from inactive UC, all CD cases and controls, but not colorectal cancer. Similar studies using variants of mass spectrometry have yielded similar results where protein profiles could accurately distinguish between selected UC and CD cases.[24-27] A recent study by Vaiopoulou et al.[28] sought to investigate pediatric biomarkers for CD by comparing the proteomic profile between adult and pediatric CD patients. 3 proteins (ceruloplasmin, clusterin and apolipoprotein B-100) were shown to be significantly different between the two cohorts. Whilst the plasma proteome is the most comprehensive collection of proteins, potential biomarkers are more difficult to detect as they exist in

Figure 1 Potential application of biomarkers in inflammatory bowel disease in different stages of clinical management. When presenting clinically, one important use of biomarkers could be in the diagnosis of IBD, as well as differentiating subtypes (e.g., UC vs CD) and phenotypes (e.g., fistulising). Whilst not currently part of management, preclinical screening for IBD may be a possibility. Biomarkers can also be used to predict response to therapy and objectively measure therapeutic response and disease severity. Due to the relapsing and remitting course of IBD, monitoring is necessary for assessing relapse, adverse outcomes and complications (e.g., strictures, fistulas and colorectal cancer). Most of these aspects necessitate endoscopic procedures and would benefit from biomarker substitutes. CD: Crohn’s disease; UC: Ulcerative colitis; IBD: Inflammatory bowel disease.
significantly lower concentrations compared to other proteins such as albumin\(^{20,26,30}\). The alternative approach that has also become popular involves sampling "proximal fluid", as any biological material directly sampled from the site of disease is likely to contain greater concentrations of potential biomarkers relative to plasma\(^{20,30-32}\). Employing a similar rationale, direct sampling of diseased tissue in IBD (a far simpler task compared to other diseases due to routine endoscopic biopsies) has been utilised for proteomic experiments (Table 2). Shkoda \textit{et al.}\(^{33}\) reported the first proteomic study of intestinal tissue, identifying nine statistically significant proteins delineating inflamed IBD tissue from non-inflamed controls. Furthermore, 40 proteins were further detected between inflamed and non-inflamed UC tissue, although only two pairs of patient samples were analysed. Similarly, Han \textit{et al.}\(^{31}\) identified a large number of differentially expressed proteins (37 relevant for CD, 27 for UC and 11 associated with general IBD) that were seen as candidate biomarkers. M’koma and colleagues conducted two studies that identified spectral peaks representing unknown protein profiles and reported being able to accurately distinguish between the UC and CD using an algorithm\(^{35,36}\). These tissue findings however are likely to require validation in plasma samples as the aim involves develop a clinical assay such as a blood test.

A similar demand for objective biomarkers exists across all aspects of IBD patient management, as such these markers have been investigated in a number of studies (summarised in Table 3). Han \textit{et al.}\(^{35}\) identified 16 additional proteins that were expressed differently between active and inactive CD. Kanmura \textit{et al.}\(^{37}\) associated a higher level of HNP 1-3 with a positive response following induction of corticosteroid therapy, whilst non-responders had lower HNP 1-3 levels. Meuwis \textit{et al.}\(^{37}\) published a second report which identified a serum protein profile which correlated with infliximab response. Gazouli \textit{et al.}\(^{38}\) performed a similar study using MALDI-TOF MS, identifying 15 proteins that were differentially expressed amongst patients that responded differently to infliximab. They were however, unable to confirm the findings by Meuwis \textit{et al.}\(^{37}\).

Most recently, Wasinger \textit{et al.}\(^{39}\) reported a panel of protein markers that were progressed into the "validation" stage using MRM. Two proteins [phosphoprotein 24 (SPP24) and \(\alpha-1\) microglobulin], were reported to be able to differentiate IBD patients and health controls whilst guanylin and secretogranin-1 differentiated UC and CD. Furthermore, three of these proteins (secre-
togranin-1, SPP24 and α-1 microglobulin), were able to distinguish between active and quiescent disease in UC and CD.

An important consideration when investigating IBD biomarkers is that a single protein may not provide the clinical utility desired, but rather a panel of markers governed by a scoring index or algorithm \[40\]. An existing example is the Brignola score which predicts relapse risk in asymptomatic Crohn’s patients by measures erythrocyte sedimentation rate, white blood cell count, hemoglobin, albumin, alpha 2-globulin, serum iron, C-reactive protein, alpha 1-glycoprotein, and alpha 2-antitrypsin \[41\]. This has been hinted at in several IBD proteomic studies which differentiated UC and CD using protein profiles rather than discrete markers \[22,24\]. The role of multiple biomarkers is highlighted by OVA1, the first Food and Drug Administration approved proteomic panel of biomarkers, consisting of 5 markers as a multivariate index assay. This assay combines multiple variables in an algorithm that produces a single diagnostic result \[42\]. These markers were identified using SELDI \[43\] and predicts the probability of a malignancy in a woman undergoing surgery for an adnexal mass \[44\]. Similarly, Plevy et al \[45\] used a combined panel of 8 serological markers, 4 genetic markers and 5 inflammatory aimed at discriminating CD from UC. The utility of this test however still requires validation in a prospective cohort. Furthermore, as it was a North American cross-sectional study, this warrants additional investigation into its validity when considering factors such as stability of markers over time \[45\] and ethnical variations \[46\].

An area that has yet to be addressed relates to the influence of IBD medications on protein abundance levels. Schreiber et al \[47\] reported the possible link between high dose 5-aminosalicylic acid (5-ASA) and modulated urinary protein concentrations. However, other groups have suggested that these urinary proteins reflect renal extra-intestinal manifestations rather than 5-ASA toxicity \[48,49\]. Derici et al \[50\] identified an association between similar urinary proteins and disease activity in UC, however none of these have been conclusive. Similarly, Mishima et al \[51\] detected elevated plasma levels of osteopontin in IBD patients, whilst Lorenzen et al \[52\] suggested a possible association between increased urinary osteopontin expression and steroid induced nephrotic syndrome. Whilst the relation between medications and their effect on protein expression is currently unclear, there are a number of implications in the context of biomarker discovery. Depending on the clinical question, the influence of medications would require strict experimental design and patient selection to avoid confounders. Additionally, biomarkers predicting or identifying adverse drug

---

**Figure 2** Biomarker “Pipeline” indicating the various stages from biomarker discovery to clinical application \[20\]. The number of candidate proteins (rough estimate of numbers indicated in figure) is narrowed down significantly in each step, selecting only the best candidates for further assessment and characterisation in a larger sample. The methodology also varies between the different phases. The early discovery phase uses low throughput methods such as 2-DE and mass spectrometry to screen large numbers of proteins in a low number of samples. Verification and validation require much more accurate quantitative methods as candidate proteins are narrowed down from the discovery phase and are assessed for their clinical utility in a large target population. This requires higher throughput methods such as MRM and immunoassays such as ELISA. CD: Crohn’s disease; UC: Ulcerative colitis; IBD: Inflammatory bowel disease; MRM: Multiple reaction monitoring.

---

Unbiased discovery of candidate proteins - “hypothesis free” driven experiments where hypotheses are formed based on the results

Low throughput and number of samples (10-100)

Techniques: 2D, mass spectrometry

Intermediate step to assess viability of select protein candidates

Moderate throughput and number of samples (100+)

Techniques: MRM, Western blot, ELISA

Characterising specificity and sensitivity using an optimised assay to select candidates for commercialisation

High throughput and number of samples from target population (1000+)

Techniques: ELISA, immunoblotting
Table 2  Proteomic studies for discovering diagnostic inflammatory bowel disease biomarkers

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Bio-sample</th>
<th>Sample size</th>
<th>Proteomic technique</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meuwis et al [24]</td>
<td>Serum</td>
<td>CD: 30</td>
<td>SELDI-TOF</td>
<td>4 candidate proteins selected for high diagnostic value; PF4, MRPS, FIBA, Hps 2. PF4 and Hps 2 were also confirmed and correlated using ELISA and immunoblotting.</td>
</tr>
<tr>
<td>Kanamura et al [23]</td>
<td>Blood</td>
<td>CD: 22</td>
<td>SELDI-TOF</td>
<td>Plasma concentrations of HNPI, 2 and 3 were significantly higher in active UC compared to inactive UC, CD and control patients</td>
</tr>
<tr>
<td>Nanni et al [27]</td>
<td>Blood</td>
<td>Healthy controls: 48</td>
<td>Liquid chromatography quadrupole-TOF</td>
<td>Exopeptidase activity may distinguish CD from UC. Protein signature of 12 mass: Charge peaks could classify CD with approximately equal 95% sensitivity/specificity</td>
</tr>
<tr>
<td>Sumramanian et al [25]</td>
<td>Serum</td>
<td>Healthy controls: 48</td>
<td>MALDI-TOF</td>
<td>4 proteins identified as clinically useful</td>
</tr>
<tr>
<td>Nanni et al [23]</td>
<td>Serum</td>
<td>Healthy controls: 48</td>
<td>Solid-phase extraction</td>
<td>20 protein signals could be used to accurately classify IBD patients</td>
</tr>
<tr>
<td>Vaiopoulou et al [24]</td>
<td>Serum</td>
<td>CD: 24 (12 adults, 12 children)</td>
<td>2-DE</td>
<td>Clusterin was found to be overexpressed in adult CD. Ceruloplasmin and apolipoprotein B-100 was over-expressed in children</td>
</tr>
<tr>
<td>Han et al [24]</td>
<td>Intestinal biopsy</td>
<td>CD: 3</td>
<td>Liquid chromatography quadrupole-TOF</td>
<td>Increased in UC: TTBK2, SYNE2, SUCLG2, POSTN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UC: 4</td>
<td>MALDI-TOF</td>
<td>Up-regulated in CD: ANXA2, EPX, LAP3, RDX</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inflammatory polyps: 2</td>
<td>MALDI-TOF</td>
<td>Up-regulated in IBD: S100A8, MPO, DEFA1B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Normal colon: 3</td>
<td></td>
<td>Up-regulated in CD (P &lt; 0.05 AND &gt; 2x increase): PRG2, LCPI, PSM1</td>
</tr>
<tr>
<td>M’koma et al [35]</td>
<td>Colon samples</td>
<td>CD: 27</td>
<td>Histology directed</td>
<td>5 mz peaks were identified and cross-validated for the differentiation of UC and CD</td>
</tr>
<tr>
<td>Seeley et al [26]</td>
<td>Colon samples</td>
<td>CD: 26</td>
<td>Histology directed</td>
<td>Using a support vector machine and 25 mz peaks, UC and CD cases were predicted in 93.3% and 60.4% respectively.</td>
</tr>
<tr>
<td>Wasinger et al [26]</td>
<td>Serum</td>
<td>UC: 27</td>
<td>MRM</td>
<td>A lower spectral accuracy cut-off increased sensitivity SPP24 differentiated IBD patients from healthy controls alpha-1-microglobulin distinguished patients with UC in remission from healthy controls</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UC: 56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Controls: 14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RA controls: 12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CD: Crohn’s disease; UC: Ulcerative colitis; IBD: Inflammatory bowel disease; MRM: Multiple reaction monitoring; PF4: Platelet factor 4; MRPS: Myeloid related protein 8; FIBA: Fibrinopeptide A; Hps2: Haptoglobin a2.

Reactions introduces an additional area of research as IBD often requires lifelong medical therapy.

It is clear that proteomics could play a potentially significant role towards improving the clinical management of IBD. Despite this, the value of these studies and their findings remain unknown and require validation in future studies.

FUTURE CONSIDERATIONS FOR IBD PROTEOMICS

Current limitations

Despite significant advancements in discovery-phase technologies and protocols, the rate at which new diagnostic protein assays are being introduced remains static, averaging 1.5/year [29]. The stagnation occurs at the verification stage, effectively obstructing any progression towards the development of a clinical assay [33,54]. This is clearly evident by the inundation of IBD discovery-phase experiments published in the recent decade with little to no candidate proteins undergoing validation.

One common criticism of many proteomic studies is the lack of strict experimental design, resulting in questionable results that cannot be reproduced; in particular, a small sample size and insufficient statistical power biomarker discovery [40,55]. This issue holds true across the aforementioned studies in IBD as out of 19 bio-sample based discovery experiments, 8 studies used ≤ 6 cases and controls [33,34,38,56-59]. In an effort to address this issue, Skates et al [55] designed a statistical model that estimates the statistical power of discovery and verification studies in tissue and plasma. Statistical power is estimated using 5 parameters: Biospecimen used (serum/tissue/proximal fluid), number of candidate proteins selected during discovery, number of cases/controls, percentage of cases where the biomarker is expressed and the difference in standard deviation between the biomarker signal in cases compared to controls. In addition, biomarkers typically occur in low abundance and may randomly exceed machine
sensitivity limits, resulting in artificial differences between samples. This combined with inherent biological variations between patient samples further emphasises the importance of achieving sufficient statistical power\(^\text{[60,61]}\). It has already been noted that the concentration of candidate IBD biomarkers may be more concentrated in the intestinal tissue compared to serum, potentially reducing the chance of false discoveries. This highlights one advantage of analysing tissue samples over serum, although it is unknown which would yield better results\(^\text{[20]}\). Significant efforts have been made to address such limitations including: Recent requirements on reporting, inclusion of standards, and superior methods. These all aim to improve accuracy and reliability and will all contribute to translatable proteomic markers for disease\(^\text{[62,63]}\).

Hanash et al\(^\text{[30]}\) identified a number of confounding factors that could contribute to variations and false discoveries when identifying potential biomarkers. Patient factors include genetic variations, metabolic state, acute phase reactants and non-specific changes such as cell death. The use of model systems such as cell cultures and animal models, provides an alternative approach that could control for confounding environmental and genetic factors\(^\text{[20,30]}\). At least 66 different animal models of IBD exist, however these may not accurately reflect the true pathophysiology of IBD. Differences in methodology that could produce artificial differences include: Sample collection and preparation, improper characterisation and randomisation, and sample/statistical analysis. Zhang et al\(^\text{[64]}\) hypothesize that many are likely site specific, suggesting that "multisite sampling" may suffice in the absence of careful prospective sample collection and randomization. This would theoretically reduce the impact of these factors and improve the likelihood of clinically useful candidate biomarkers being detected\(^\text{[64]}\).

The issues highlighted above demonstrate the requirement for standardisation of protocols in large-scale proteomics experiments or at least stringent experimental design to increase the chances of discovering valid biomarkers.

### Towards verification and validation

The process of validation differs significantly from the initial discovery stage as candidate proteins are tested in thousands of samples. This phase uses reliable high throughput methods (e.g., immunoassays) in order to evaluate the biomarker’s utility in the target population. Unfortunately this phase requires significant financial investment and produces a major barrier to validating the numerous proteins identified as “candidate biomarkers”\(^\text{[20,53,54]}\). Consequently, many potential markers are identified in the literature but require further investigation.

The gap between the inherent inaccuracies of the...
discovery phase and the prohibitive cost of validation gave rise to the notion of an intermediate “verification” stage, aimed towards bridging this gap. This is achieved by quantification of selected candidate biomarkers in a larger sample that better represents the target population[56]. Ideally this is performed using reliable and established immunoassays, however, commercial antibodies are unavailable for the majority of target proteins, especially novel candidate markers. Assays must then be developed specifically for testing of the biomarker, an extremely costly endeavour when considering the large numbers of biomarkers[54]. Mass spectrometry can be further utilised here through quantitative techniques. Methods such as MRM have emerged as viable alternatives towards cost-effectively triaging proteins of interest for further validation[20,53,86,105] and has been published for a number of biomarkers in other diseases[66-68].

CONCLUSION

The field and application of proteomics has expanded greatly in recent years and could have profound implications on the clinical diagnosis and management of IBD through the discovery of novel biomarkers. Many groups have already begun the “discovery” process and have identified many potential candidates. Although the transition into clinical validation is challenging, the tremendous potential of proteomics has garnered great interest and success in other diseases and further investigations into IBD proteomics should certainly be pursued.

REFERENCES


42 Zhang Z. An In Vitro Diagnostic Multivariate Index Assay (IVDMA) for Ovarian Cancer: Harvesting the Power of Multiple Biomarkers. Rev Olotst Gynecol 2012; 5: 35-41 [PMID: 22582125]


53 Makawita S, Diamandis EP. The bottleneck in the cancer biomarker


S- Editor: Qiu S L- Editor: A E- Editor: Jiao XK