Discovery of biomarker candidates associated with the risk of short- and mid/long-term relapse after infliximab withdrawal in Crohn's patients: a proteomics-based study

Nicolas Pierre[1], Dominique Baiwir[2,*], Vân Anh Huynh-Thu[3,*], Gabriel Mazzucchelli[4], Nicolas Smargiasso[4], Edwin De Pauw[4], the GETAID (Groupe d'Etude Thérapeutique des Affections Inflammatoires du tube Digestif), Yoram Bouhnik[5], David Laharie[6], Jean-Frederic Colombel[7], Marie-Alice Meuwis[1,8,#], Edouard Louis[1,8,#]

[1]Laboratory of Translational Gastroenterology, GIGA-institute, University of Liège, Liège, Belgium; [2]GIGA Proteomics Facility, University of Liège, Liège, Belgium; [3]Department of Electrical Engineering and Computer Science, University of Liège, Liège, Belgium; [4]Laboratory of Mass Spectrometry, MolSys Research Unit, University of Liège, Liège, Belgium; [5]Gastroentérologie, Maladies Inflammatoires Chroniques de l'Intestin et Assistance Nutritive, Hôpital Beaujon, APHP-Université Paris VII, Clichy, France; [6]Service d'Hépato-Gastroentérologie, CHU Bordeaux, Bordeaux, France; [7]Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [8]Hepato-Gastroenterology and Digestive Oncology Department, Liège University Hospital CHU, Liège, Belgium. *,# *Contributed equally*

Corresponding author:

Nicolas Pierre

Address: Translational Gastroenterology, GIGA institute Bât. B34 Quartier Hôpital, avenue de l'Hôpital 11, 4000 Liège 1, Belgique

Tel: +32 4 3662538; Fax: +32 4 3667889; Email: nicolas.pierre@uliege.be

Word count: 4253

ABSTRACT

Objective A subset of Crohn's disease (CD) patients experiences mid/long-term remission after infliximab withdrawal. Biomarkers are needed to identify those patients.

Design New biomarkers of relapse were searched in the baseline serum of CD patients stopping infliximab when they were under combined therapy (antimetabolite and infliximab) and stable clinical remission (STORI cohort, n=102). From shotgun proteomics experiment (discovery step), biomarker candidates were identified and further targeted by selected reaction monitoring (verification step). The dataset was stratified to search for markers of short- (<6 months) or mid/long-term relapse (>6 months). The risk of relapse and the predicting capacity associated with biomarker candidates were evaluated using univariate Cox model and log-rank statistic, respectively. To test their complementary predicting capacity, biomarker candidates were systematically combined in pairs.

Results Distinct biomarker candidates were associated with the risk (hazard ratio: HR) of short- (15 proteins, $2.9<HR<16.1$, $p<0.05$) and mid/long-term (17 proteins, $2<HR<4.4$, $p<0.05$) relapse, they reflect different pathophysiological processes. In stratified and non-stratified datasets, novel marker combinations exhibited a high predicting capacity as shown by their higher Z-scores (FDR<0.001) than CRP and faecal calprotectin (current references in predicting relapse).

Conclusion We identified for the first time circulating biomarker candidates associated with the risk of mid/long-term relapse in CD patients stopping infliximab. We also highlight a sequence of pathophysiological processes leading to relapse, this could help to better understand the disease progression. Our findings may pave the way for a better non-invasive evaluation of the risk of relapse when contemplating anti-TNFα withdrawal in CD patients.

What is already known about this subject?

- After infliximab withdrawal, a subset of Crohn's disease (CD) patients experiences a mid/long-term remission.
- C-reactive protein (CRP) and faecal calprotectin are currently the most broadly used biomarkers to predict a short-term (<6 months) relapse after infliximab withdrawal in CD patients.
- Biomarkers predicting a mid/long-term (>6 months) relapse are needed.

What are the new findings?

- The risk of short- and mid/long-term relapse were associated with specific proteins reflecting distinct pathological processes.
- Novel and simple combinations of two biomarker candidates showed a high potential to predict the relapse when they were compared to CRP and faecal calprotectin.

How might it impact on clinical practice in the foreseeable future?

- By providing a mid/long-term perspective in predicting relapse, the new biomarker candidates could become an essential tool for clinicians when contemplating infliximab withdrawal in CD patients.

INTRODUCTION

In Crohn's disease (CD), the introduction of biologics directed against the tumour necrosis factor α (TNFα) noticeably improved the patient's quality of life by inducing and maintaining remission and by reducing hospitalisations and surgical resections [1–3]. This relative success leads to a new challenge for clinicians: when to stop anti-TNFα in patients experiencing long-standing clinical remission[4–6]? In properly identified cases, stopping anti-TNFα would be beneficial for both patients and healthcare system. Indeed, long-term exposure to anti-TNFα has been associated with serious adverse events (e.g., infections, lymphoma, melanoma and non-melanoma skin cancer)[7–10] and a loss of response[11,12]. On the other hand, chronic treatment with anti-TNFα remains costly and its long-term cost-effectiveness is still uncertain[13,14]. For instance, a Dutch study reported that anti-TNFα treatment represents the main part (64%) of the healthcare costs related to the management of CD patients[15]. In this context, intensive research has been conducted to find predictors of relapse[16]. The search of biomarkers predicting the relapse after anti-TNFα withdrawal has been mainly conducted by measuring classical parameters used in routine practice (e.g., C-reactive protein: CRP, faecal calprotectin, haemoglobin, platelet count, leukocyte count)[16]. Although such approach is simple and can be easily translated to clinical practice, it does not allow for the discovery of new biomarkers showing interest in predicting relapse. Currently, CRP and faecal calprotectin are recognised by experts as the best biomarkers to evaluate the risk of relapse after stopping biologics[6]. However, these predictors seem to be essentially associated with the risk of short-term relapse (<6 months)[17,18] while a median time to relapse ranging from 4.8 to 16.4 months after infliximab withdrawal has been reported[19]. Thus, identification of biomarkers predicting a longer-term relapse (>6 months) is an unmet clinical need. From a biological point of view, CRP and faecal calprotectin seem to testify of an ongoing inflammation related to an advanced state of the pathophysiological process leading to

relapse. They most probably do not reflect the upstream defect of the immune system. This fundamental part of the pathophysiological process could be captured by new biomarkers, i.e., proteins that would predict a mid/long-term relapse.

In the present study, we performed the classical workflow dedicated to the search of novel protein biomarkers[20]. In a shotgun proteomics experiment (discovery step), we identified biomarker candidates which were further targeted (verification step) by selected reaction monitoring (SRM). The two experiments (discovery and verification) were conducted on baseline serum of patients included in the study of infliximab diSconTinuation in CrOhn's disease patients in stable Remission on combined therapy with Immunosuppressors (STORI)[21]. This study was specifically designed to investigate predictors of relapse after infliximab withdrawal in CD patients.

To test the hypothesis of different biomarkers for the short- and mid/long-term relapse, we also stratified the dataset according to the time to relapse, either short- (<6 months) or mid/long-term (>6 months) relapse. The predicting capacity of the studied markers was compared to CRP and faecal calprotectin.

METHODS

Subjects and samples

STORI is a prospective and multicentre study which aimed to determine the risk factors of relapse in CD patients after stopping the infliximab treatment[21]. The included CD patients (102 out of the initial 115 patients in STORI; Table 1) were on a corticosteroid-free clinical remission, defined by a CD activity index (CDAI)<150 for at least 6 months and they were treated with a combined therapy (infliximab and antimetabolites) for at least one year. At baseline, these patients stopped infliximab and continued antimetabolite. To investigate predictive markers of relapse, we analysed the blood samples taken at baseline. After blood

clotting and centrifugation, sera were stored at -80 C° until analysis. During the follow-up, a relapse was defined by a CDAI>250 or a CDAI between 150 and 250 with a 70-point increase from baseline over two consecutive weeks. The study protocol and documents were approved by the French Ethics Committee- Hôpital Saint-Louis (CPP 2005/14) and the AFSSAPS (0809/ALV/EG05). The investigational review board at each of the participating centers approved the protocol. All patients gave their written informed consent before screening. Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Biomarker discovery strategy

Online supplementary methods.

Clinical measurement of faecal calprotectin and CRP

Faecal calprotectin and CRP were measured using respectively the PhiCal® ELISA (Lysaker, Norway) and the high-sensitivity turbidimetric immunoassay (hsCRP) as previously described[22]. CRP was measured in all samples (n=102) and faecal calprotectin was measured on 77 available samples.

Statistical analysis

The dataset was stratified to search for markers of short- (<6 months) or mid/long-term relapse (>6 months). Since the short- and the mid/long-term relapses are not defined in the literature, this stratification strategy was somewhat arbitrary. However, this choice was supported by data showing, as early as 4-6 months before the relapse, an elevation of CRP and faecal calprotectin, the markers already associated with the prediction of short-term relapse[17,18,23]. In the dataset corresponding to the short-term relapse (<6 months), the

mid/long-term relapsers (n=29) were classified as non-relapsers and the non-relapser follow-ups were censored at 6 months. Thus, this dataset was composed of 102 patients (15 relapsers: R and 87 non-relapsers: NR). The dataset corresponding to the mid/long-term relapse (>6 months) was obtained by eliminating short-term relapsers (n=15) and non-relapsers with a follow-up inferior to 6 months (n=14). Thus, this dataset was composed of 73 patients (29 R and 44 NR). All statistical analyses were performed in the stratified and non-stratified datasets.

For each protein, an optimal cut-off was identified as the one maximising the Youden's index. From these cut-offs, the classification of patients (below vs above the cut-off) was obtained for each protein in each dataset. Then, association of each protein with time to relapse was assessed with the univariate Cox model. This method generated for each protein the hazard ratio (HR) which can be interpreted as a measure of a relative risk of relapse. The reference (HR=1) was defined as the patients having a protein level under the cut-off. For a given protein "A", a HR>1 or a HR<1 means that increase or decrease of "A" is associated with a risk of relapse, respectively. The survival curves were estimated using the Kaplan-Meier method and they were compared with the log-rank test.

We performed the systematic combinations of two biomarker candidates with the "AND" or "OR" logical operators (n=2556 combinations per dataset and per logical operator). To control the false discovery rate (FDR) associated with this high number of tested combinations, log-rank p-values were adjusted with the Benjamini–Hochberg procedure. This correction was not applied to CRP and faecal calprotectin since these biomarkers were used as references. The proportional hazard assumption was checked using both the Schoenfeld's residuals test and graphically by verifying whether the survival curves crossed. As appropriate, proteins that did not respect the proportional hazard assumption were not presented. The strength of association between two variables was evaluated with the Spearman's rank correlation.

Survival analyses and volcano plots were generated using lifelines Python library[24] and ggplot2 R package, respectively. The p-values or FDR<0.05 were considered significant.

RESULTS

Study population and overview of the biomarker discovery strategy

The patients' clinical characteristics are presented in Table 1. We observed 44/102 relapsers (R) with a median (interquartile range: IQR) time to relapse of 6.7 (4.2-11.2) months and 58/102 non-relapsers (NR), with a median (IQR) time of follow up of 22.9 (6.6-30.1) months. The biomarker discovery strategy is presented in Figure 1. We used the asset of mass spectrometry-based proteomics to investigate a large number of proteins during the discovery step. We then developed an in-house targeted method using SRM, a mass spectrometry-based technology allowing an accurate and robust quantitative approach with a high degree of multiplexing and a broad dynamic range (around five orders of magnitude)[25,26]. Candidate biomarkers highlighted by the discovery study were finally assayed by our SRM method during a verification step. In the discovery step, 849 proteins were identified and quantified in the serum pools of relapser (n=6) and non-relapser (n=7) patients from the STORI cohort (online supplementary method 1). Among these proteins, 252 were considered candidate biomarkers since they were differentially abundant between relapsers and non-relapsers (online supplementary method 1, online supplementary table 1). After a technical development (online supplementary method 2), 91/252 biomarker candidates were included in the SRM method. During the verification step, the SRM method was applied to individual samples of the STORI cohort (n=102, online supplementary method 3). After inspection of the SRM data, 19 proteins were eliminated of the dataset (online supplementary method 4). The curated SRM dataset contained 72 biomarker candidates (online supplementary table 2). The quality controls of the SRM method are detailed in online supplementary method 4.

Common and distinct markers associated with the risk of short- and mid/long-term relapse

We first stratified the dataset to test whether the studied proteins were associated with the risk of short-term (<6 months) and/or mid/long-term (>6 months) relapse (see methods). The median (IQR) time to relapse was 3.6 (2.8-4.1) and 9.8 (6.7-12.5) months in the short- and mid/long-term relapse datasets, respectively.

We used faecal calprotectin and CRP as references since they are the most recognised biomarkers of relapse after anti-TNFα withdrawal in CD patients. The faecal calprotectin was added in the data analysis while CRP was a biomarker candidate included in our SRM method (online supplementary table 2). As shown in Figure 2, the measure of CRP by SRM was highly correlated (r=0.96) with the one performed by high-sensitivity turbidimetric immunoassay (gold standard). In addition to support the quality of our in-house developed SRM method, this result indicated that the measure of CRP by SRM can be used as an appropriate reference.

In the stratified and non-stratified datasets, we determined the risk of relapse associated with each protein through the hazard ratio (HR) generated by the univariate Cox model. As hypothesised, we identified proteins specifically associated with the risk of short-term (n=15) or mid/long-term (n=17) relapse (Figures 3, online supplementary table 3). We depicted the Log_2 HR of each protein in volcano plots (Figure 4). For a given protein, a positive or a negative Log_2 HR means a risk of relapse linked to an increase or a decrease level of this protein, respectively (see methods). Patients who experienced a short-term relapse were characterised by an innate immune response which seems to originate mainly from the liver. Indeed, the risk of short-term relapse was specifically associated with increased circulating levels of proteins that are mainly produced by the liver: acute-phase reactants (ceruloplasmin:

CP; haptoglobin-related protein: HPR; inter-alpha-trypsin inhibitor heavy chain H3: ITIH3; leucine-rich alpha-2-glycoprotein: LRG1 and serum amyloid P-component: APCS) and factors of the complement system (complement C3: C3; complement C4-B: C4B; complement C5: C5; complement factor H: CFH and complement factor H-related protein 2: CFHR2 and mannose-binding protein C: MBL2) (Figure 4A; online supplementary table 3). As expected, increased circulating level of CRP was associated with the risk of short-term relapse (Figure 4A). In the mid/long-term relapse dataset, the CRP variable did not respect the proportional hazard assumption required for the Cox model thus forbidding to depict this protein in the volcano plot (Figure 4B) since the associated statistic may not be valid. Herein, the non-respect of this assumption simply testifies that CRP cannot predict the mid/long-term relapse. Indeed, the survival curves showed the ability of CRP to predict short-term but not mid/long-term relapse (Figures 5A and 5B).

Compared to the specific markers of short-term relapse, the specific markers of mid/long-term relapse presented a distinct and heterogeneous biological meaning (Figures 4A and 4B; online supplementary table 3). A decrease of their circulating levels was, in most of the cases (13/17), associated with the risk of mid/long-term relapse. This may reflect a partial weakening of the immune defence as shown by the distribution of carboxypeptidase N subunit 2 (CPN2), histidine-rich glycoprotein (HRG), immunoglobulin heavy constant alpha 1 (IGHA1), immunoglobulin J chain (JCHAIN), immunoglobulin kappa variable 4-1 (IGKV4-1) and mannan-binding lectin serine protease 2 (MASP2). Some of the specific markers of mid/long-term relapse are involved in angiogenesis (HRG; serpin family F member 1: SERPINF1) and others seem to indicate an activation of leukocytes (neutrophil defensin: DEFA1; L-selectin: SELL). In addition to liver, some of them (5/17) mainly originate from immune cells: IGHA1, IGKV4-1, JCHAIN, SELL and DEFA1 (online supplementary table 3). Finally, some of the specific markers of mid/long-term relapse are involved in the

complement system. A decrease (CPN2 and MASP2) or an increase (complement component C6: C6; complement factor B: CFB) of their circulating levels was associated with the risk of mid/long-term relapse.

Ten proteins were associated with the risk of both short- and mid/long-term relapse (Figure 3). Some of those markers were more significantly associated with the risk of either short-term (alpha-1-acid glycoprotein 1: ORM1; beta-2-glycoprotein 1: APOH; beta-ala-his dipeptidase: CNDP1; haptoglobin: HP; complement component C9: C9 and serotransferrin: TF) or mid/long-term relapse (apolipoprotein A-I: APOA1) (Figures 4D and 4E; online supplementary table 4). The distribution of HP, ORM1 and TF underlined, once again, that short-term relapsers presented evident signs of an ongoing acute-phase response (Figure 4D). Such observations emphasise the coherence of the present dataset. The complement component C8 beta chain (C8B) and the plasminogen (PLG) presented a similar statistical association with the risk of short- and mid/long-term relapse (Figures 4D and 4E; online supplementary table 4). However, increased circulating level of C8B was associated with the risk of short- and mid/long-term relapse while increased and decreased circulating level of PLG were associated with the risk of short- and mid/long-term relapse, respectively (Figures 4D and 4E; online supplementary table 4). These examples show the different dynamics of markers before the relapse. In a subset of the cohort (n=77), faecal calprotectin was associated with the risk of short- and mid/long-term relapse (Figures 4D and 4E). In the non-stratified dataset, haptoglobin and faecal calprotectin showed similar statistical association with the risk of relapse (Figure 4F). As shown Figures 4C and 4F, if no stratification was applied, some relevant markers and key biological information would have been missed.

Comparison with CRP and faecal calprotectin highlights a high potential of novel marker combinations in predicting relapse

To evaluate the potential of markers in predicting relapse, we performed the systematic combinations of two of them with the "AND" or "OR" logical operators and, the combination presenting the best discriminatory ability (R vs NR, based on the log-rank Z-score) was compared to CRP and faecal calprotectin in each dataset (Figure 5). Novel combinations of markers showed a higher log-rank Z-score than CRP and faecal calprotectin in the short-term relapse dataset (7.62 vs 3.32 vs 2.85, respectively), mid/long-term relapse dataset (6.11 vs NA vs 3.66) and non-stratified dataset (5.99 vs 3.58 vs 3.97). Contrary to our marker combinations, CRP was not able to predict the mid/long-term relapse as shown by the crossing of the survival curves and the overlapping of their confidence intervals (Figures 5B and 5H). Remarkably, the selected cut-off of faecal calprotectin was higher in the short-term (193 µg/g) than the mid/long-term (85 µg/g) relapse dataset (Figures 5D and 5E). A high number of novel marker combinations presented a FDR<0.05 and a Z-score superior to CRP and faecal calprotectin (short-term relapse dataset: n=594; mid/long-term relapse dataset: n=176; non-stratified dataset: n=104; online supplementary table 5). Taken together, these results support that the newly identified markers exhibit a high potential in predicting relapse.

DISCUSSION

In the present study, we found for the first time biomarker candidates associated with the risk of mid/long-term relapse (>6 months) in CD patients after stopping infliximab. Given that for those patients the median time to relapse ranges from 4.8 to 16.4 months[19], our findings meet a clinical need. Actually, no study clearly evaluated the time-windows in which the biomarkers have a capacity to predict a relapse. By stratifying our dataset according to time to relapse, we evaluated the prediction time-windows of each studied markers including the current reference biomarkers of relapse CRP and faecal calprotectin. As expected, we clearly demonstrated that CRP can predict a relapse if and only if it occurs within 6 months after anti-

TNFα withdrawal. This lack of mid/long-term perspective constitutes a real issue when the question of stopping anti-TNFα arises.

In the present study, high level of faecal calprotectin was associated with the risk of both short- and mid/long-term relapse. It probably makes sense that a marker of mucosal neutrophil infiltration as faecal calprotectin reflects the whole pathological process leading to relapse. Indeed, neutrophil infiltration is closely linked to the development of Crohn's lesions and it is an early feature of the disease[27,28]. In line with our results, low level of faecal calprotectin (<56 µg/g) predicted stable remission (median follow-up of 11 months) in IBD patients with mucosal healing[29]. In our work, we found a cut-off value higher for predicting short-term (193 µg/g) than mid/long-term (85 µg/g) relapse. This is in agreement with another study[18] and it seems to be coherent with the disease progression. Such an observation also brings a new piece in the delicate debate around the optimal cut-off value of the faecal calprotectin and it could partially explain the difficulty to find a unique cut-off[30].

Although faecal calprotectin is a very effective predictor of CD relapse, the use of faecal samples is not well accepted by patients and it is subject to practical and methodological limitations[30–32]. Herein, we reported that serum haptoglobin could predict the relapse with a similar performance as faecal calprotectin in the non-stratified dataset (as it is usually evaluated). The potential of haptoglobin in predicting relapse needs to be confirmed in an independent study. Compared to faecal calprotectin and endoscopy, circulating biomarkers have the advantage to provide an objective, non-invasive and well-accepted test to evaluate the risk of relapse. By comparing our results with CRP and faecal calprotectin, we showed that novel and simple combinations of two circulating proteins presented a high capacity to predict the relapse.

Mucosal healing is thought to be a favourable prognostic factor[33]. An endoscopic activity (CDEIS >0) has been identified as a risk factor for relapse in CD patients stopping

infliximab[21]. However, this parameter presented a lower prognostic value than biomarkers (CRP and calprotectin) and ~30% of patients with mucosal healing (CDEIS equal to zero) relapsed within two years after cessation of the treatment[21]. Mucosal healing can also be defined histologically but currently there is no validated procedure and score to evaluate residual histologic activity in CD[34]. Thus, biomarkers seem to present an added value over invasive methods in terms of disease course prediction and evaluation of the degree of remission.

Albeit descriptive, our data deserve to be discussed in the context of CD pathophysiology. As hypothesised, our data clearly support that short- and mid/long-term markers of relapse reflect distinct parts of the pathophysiological process leading to relapse. In a coherent manner, we found that the short-term relapsers were characterised by an innate immune response of the liver as demonstrated by the association between the short-term risk of relapse and the increased circulating levels of acute-phase reactants and complement factors. Like others[17,35], we found an association between the risk of relapse and the increased circulating level of the acute-phase reactant ORM1 in CD patients. According to our results, acute-phase reactants and complement factors mainly reflect an advanced state of the pathophysiological process leading to clinical relapse. In this work, we also explored upstream events by studying markers of mid/long-term relapse. In contrast to short-term relapsers, patients who experienced a longer-term relapse were not characterised by an innate immune response of the liver but they seemed to present a partial weakening of the immune system. This temporal information is in favour of the debated theory according to which a partial immunodeficiency would be a primary defect in CD[36]. Further investigations are needed to challenge such hypothesis, this is beyond the scope of the present study. In accordance with the partial weakening of the immune system, our most striking finding is certainly the association of the decreased circulating levels of JCHAIN and IGHA1 with the

risk of mid/long-term relapse. In line with this result, a decrease of JCHAIN incorporation in IgA has been reported in the mucosa of Crohn's patients compared to healthy subjects[37]. The secretory IgA constitutes the first line defence of the epithelium against pathogens. On the other hand, JCHAIN is produced by glandular and mucosal plasma cells where it plays a central role in secretory immunity. Its incorporation in IgA dimers and IgM pentamers positively regulates their secretion by epithelial cells and allows to limit the inflammatory process[38]. Thus, decreased circulating level of JCHAIN and IGHA1 could testify from a specific defect (immune exclusion) of the intestinal mucosa that could advertise from early pathological process leading to a remote relapse.

Together with the partial weakening of the immune system, our results suggest the presence of a low-grade immune activation long time before the relapse. This hypothesis is supported by faecal calprotectin, serum DEFA1 and SELL, three markers of leukocytes whose increases were associated with the risk of mid/long-term relapse in our study.

Our data also support that modulation of angiogenesis could be an early predictor of relapse. Indeed, proteins involved in angiogenesis (HRG, SERPINF1 and PLG) were associated with a risk of mid/long-term relapse. SERPINF1 and PLG (via its Kringle fragments such as angiostatin) are recognised as potent inhibitors of angiogenesis while HRG has complex pro- and anti-angiogenic properties[39–41]. These results could testify from the well-established role of angiogenesis in CD pathophysiology[42].

The present study demonstrates that integration of simple biological concepts in data analysis can bring out fundamental information. Indeed, an analysis without stratification, as it is conventionally done, leads to miss key information of the dataset. To take into account the dynamic of the disease, we stratified the dataset according to time to relapse. We acknowledge that such approach decreases the statistical power since it unbalances the cohort (number of relapsers vs number of non-relapsers) and it reduces the sample size (mid/long-term relapse

dataset). However, as others, we reported that the use of prior knowledge in data mining is able to reveal critical information missed by hypothesis-free approach exclusively based on statistical data analysis[43]. For instance, principal component analysis did not allow to separate short- and mid/long-term relapsers (data not shown). Thanks to our data mining strategy, we brought out relevant relations between CD pathophysiology and markers.

In the field of biomarkers, SRM is seen as a promising technology since immunoassay has a limited capacity of multiplexing and its development is time and cost consuming. Indeed, SRM allows accurate, precise and robust measurements, all with a high degree of multiplexing (50-100 analytes) and a broad dynamic range (around five orders of magnitude)[25,26]. Such performance makes SRM attractive for the verification step in which multiple candidate biomarkers must be assayed with a high degree of confidence. However, application of this technology in proteomics is criticised and its contribution to biomarker discovery remains limited[44]. Experts in the field pointed out the need of methodological guaranties and transparency that would allow to generate more reliable and reproducible results[44]. Taking into account these recommendations, we provided all the information needed to transfer our SRM method to another laboratory and we enabled public access to raw data and analysed files supporting our findings (online supplementary methods). We also presented evidences showing the stability of the instrumental set-up, the high precision of the measurements, the rightness of the normalisation strategy, the correctness of the target identities and the quantotypic properties of the chosen peptides (online supplementary method 4). This clearly emphasises the high quality of the technical development and therefore the reliability of the present data. In the verification step of the biomarkers discovery workflow, the use of heavy peptides is recommended to prove target identities[44]. The absence of heavy peptides is an analytical limitation of our study. However, the identity of the studied markers

is well supported by a deep verification of the data (online supplementary method 4) and by coherent biological effects.

The main limitation of our study is certainly the absence of an independent cohort to evaluate the predictive performance of the newly identified markers. This prevents to know the generalisation capacity of our results and, at this stage, our study is not able to identify the winning combinations of biomarkers. However, the present study did not aim to validate a particular combination of biomarkers. Our objective was to show the capacity of our method to bring out relevant and new markers in predicting relapse. Others studies are needed to determine the predictive performance of the newly identified markers.

Other limitation of our study could be the influence of drugs on our results. Indeed, biologic and immunosuppressive treatments (e.g., anti-TNFα and anti-metabolite), administered until the blood sample, could have impacted the circulating level of the identified markers. Hence, the biological functions highlighted may not really correspond to the natural history of the disease but rather to the evolution of the pathological process under the influence of the administered treatment.

In conclusion, we found for the first time biomarker candidates associated with the risk of mid/long-term relapse after anti-TNFα withdrawal in CD patients. Novel and simple combinations of two proteins showed a high potential to predict the relapse when they were compared to CRP and faecal calprotectin, the current references in the field. Our finding could constitute a crucial progress to identify a subpopulation of CD patients who could stop anti-TNFα with an optimal risk-benefit ratio. We also demonstrated that proteins associated with the risk of short- or mid/long-term relapse reflect distinct pathophysiological processes. This observation allows to better understand the chronology of biological events leading to the relapse.

Acknowledgments

We thank Charlotte Massot, Samira Azarzar, Nancy Rosière, Lisette Trzpiot, and Haingo Nanou Tanteliarisoa for their precious technical assistance. We also thank the people of the GIGA Proteomics Facility for sharing their expertise. The following GETAID Centers (investigators) participated in this study: Amiens (J. L. Dupas), Bordeaux (D. Laharie), Caen (J.M. Reimund), Clichy-Beaujon (Y. Bouhnik), Colombes-L Mourier (P. Jouet), Gent (M. De Vos), Liège (J. Belaiche, E. Louis), Lille (J.-F. Colombel, G. Vernier-Massouille), Lyon (S. Nancey), Marseille (J. C. Grimaud), Montpellier (M. Veyrac), Nantes (A. Boureille, M. Flamant), Paris-HEGP (R. Jian), Paris-Lariboisière (P. Marteau), Paris-St Louis (M. Lemann, M. Allez), Rouen (G. Savoye), Strasbourg (B. Duclos), and Tours (L. Picon).

Author Contributions

NP, M-AM, YB, DL, J-FC and EL designed the experiments. The GETAID provided the samples and the clinical information. M-AM performed the discovery study. M-AM, DB and NP developed the SRM method. NP and M-AM performed the sample preparation. NP, M-AM, DB, NS, GM, and ED-P managed the injection of the proteomic experiments and provided advises for the raw data analysis. V-A H-T and NP performed the statistical analysis. NP, M-AM, V-A H-T and EL interpreted the data. NP, M-AM, and EL wrote the paper.

Competing interests

None declared

Funding

This work was supported by the Walloon Region and the Fond européen de développement régional [FEDER] [Grant portfolio 246099-510388]. Financial support for the STORI trial

was provided by the SNFGE and the association François Aupetit. M-AM received an ECCO research grant (2012) for the discovery study. Additional support was provided by internal funding from ULiège and CHU de Liège.

REFERENCES

1 Colombel J, Sandborn WJ, Rutgeerts P, *et al.* Adalimumab for Maintenance of Clinical Response and Remission in Patients With Crohn's Disease: The CHARM Trial. *Gastroenterology* 2007;**132**:52–65. doi:10.1053/J.GASTRO.2006.11.041

2 Hanauer SB, Feagan BG, Lichtenstein GR, *et al.* Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. *Lancet (London, England)* 2002;**359**:1541–9. doi:10.1016/S0140-6736(02)08512-4

3 Costa J, Magro F, Caldeira D, *et al.* Infliximab reduces hospitalizations and surgery interventions in patients with inflammatory bowel disease: a systematic review and meta-analysis. *Inflamm Bowel Dis* 2013;**19**:2098–110. doi:10.1097/MIB.0b013e31829936c2

4 Louis E. Stopping Biologics in IBD-What Is the Evidence? *Inflamm Bowel Dis* 2018;**24**:725–31. doi:10.1093/ibd/izx098

5 Louis E, Belaiche J, Reenaers C. Anti-TNF and Crohn's disease: when should we stop? *Curr Drug Targets* 2010;**11**:148–51.

6 Pittet V, Froehlich F, Maillard MH, *et al.* When do we dare to stop biological or immunomodulatory therapy for Crohn's disease? Results of a multidisciplinary European expert panel. *J Crohns Colitis* 2013;**7**:820–6. doi:10.1016/j.crohns.2013.04.013

7 Cottone M, Criscuoli V. Infliximab to treat Crohn's disease: an update. *Clin Exp Gastroenterol* 2011;**4**:227–38. doi:10.2147/CEG.S6440

8 Burmester GR, Panaccione R, Gordon KB, *et al.* Adalimumab: long-term safety in 23 458 patients from global clinical trials in rheumatoid arthritis, juvenile idiopathic arthritis, ankylosing spondylitis, psoriatic arthritis, psoriasis and Crohn's disease. *Ann Rheum Dis* 2013;**72**:517–24. doi:10.1136/annrheumdis-2011-201244

9 Siegel CA, Marden SM, Persing SM, *et al.* Risk of lymphoma associated with combination anti-tumor necrosis factor and immunomodulator therapy for the treatment of Crohn's disease: a meta-analysis. *Clin Gastroenterol Hepatol* 2009;**7**:874–81. doi:10.1016/j.cgh.2009.01.004

10 Long MD, Martin CF, Pipkin CA, *et al.* Risk of melanoma and nonmelanoma skin cancer among patients with inflammatory bowel disease. *Gastroenterology* 2012;**143**:390-399.e1. doi:10.1053/j.gastro.2012.05.004

11 Karmiris K, Paintaud G, Noman M, *et al.* Influence of trough serum levels and immunogenicity on long-term outcome of adalimumab therapy in Crohn's disease. *Gastroenterology* 2009;**137**:1628–40. doi:10.1053/j.gastro.2009.07.062

12 Nanda KS, Cheifetz AS, Moss AC. Impact of antibodies to infliximab on clinical outcomes and serum infliximab levels in patients with inflammatory bowel disease (IBD): a meta-analysis. *Am J Gastroenterol* 2013;**108**:40–7; quiz 48. doi:10.1038/ajg.2012.363

13 Bodger K, Kikuchi T, Hughes D. Cost-effectiveness of biological therapy for Crohn's disease: Markov cohort analyses incorporating United Kingdom patient-level cost data. *Aliment Pharmacol Ther* 2009;**30**:265–74. doi:10.1111/j.1365-2036.2009.04033.x

14 Bolin K, Hertervig E, Louis E. The Cost-effectiveness of Biological Therapy Cycles in the Management of Crohn's Disease. *J Crohns Colitis* 2019;**13**:1323–33. doi:10.1093/ecco-jcc/jjz063

15 van der Valk ME, Mangen M-JJ, Leenders M, *et al.* Healthcare costs of inflammatory

bowel disease have shifted from hospitalisation and surgery towards anti-TNFalpha therapy: results from the COIN study. *Gut* 2014;**63**:72–9. doi:10.1136/gutjnl-2012-303376

16 Torres J, Boyapati RK, Kennedy NA, *et al.* Systematic Review of Effects of Withdrawal of Immunomodulators or Biologic Agents From Patients With Inflammatory Bowel Disease. *Gastroenterology* 2015;**149**:1716–30. doi:10.1053/j.gastro.2015.08.055

17 Consigny Y, Modigliani R, Colombel J-F, *et al.* A simple biological score for predicting low risk of short-term relapse in Crohn's disease. *Inflamm Bowel Dis* 2006;**12**:551–7. doi:10.1097/01.ibd.0000225334.60990.5b

18 Molander P, Farkkila M, Ristimaki A, *et al.* Does fecal calprotectin predict short-term relapse after stopping TNFalpha-blocking agents in inflammatory bowel disease patients in deep remission? *J Crohns Colitis* 2015;**9**:33–40. doi:10.1016/j.crohns.2014.06.012

19 Chapman TP, Gomes CF, Louis E, *et al.* De-escalation of immunomodulator and biological therapy in inflammatory bowel disease. *Lancet Gastroenterol Hepatol* 2020;**5**:63–79. doi:10.1016/S2468-1253(19)30186-4

20 Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* 2006;**24**:971–83. doi:10.1038/nbt1235

21 Louis E, Mary J, Vernier–Massouille G, *et al.* Maintenance of Remission Among Patients With Crohn's Disease on Antimetabolite Therapy After Infliximab Therapy Is Stopped. *Gastroenterology* 2012;**142**:63-70.e5. doi:10.1053/J.GASTRO.2011.09.034

22 Meuwis M-A, Vernier-Massouille G, Grimaud JC, *et al.* Serum calprotectin as a biomarker for Crohn's disease. *J Crohns Colitis* 2013;**7**:e678-83.

doi:10.1016/j.crohns.2013.06.008

23 de Suray N, Salleron J, Vernier-Massouille G, *et al.* 864 Close Monitoring of CRP and Fecal Calprotectin is Able to Predict Clinical Relapse in Patients With Crohn's Disease in Remission After Infliximab Withdrawal. a Sub-Analysis of the Stori Study. *Gastroenterology* 2012;**142**:S-149. doi:https://doi.org/10.1016/S0016-5085(12)60560-5

24 Davidson-Pilon C, Kalderstam J, Jacobson N, *et al.* CamDavidsonPilon/lifelines: v0.24.8. Published Online First: 18 May 2020. doi:10.5281/ZENODO.3833188

25 Addona TA, Abbatiello SE, Schilling B, *et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* 2009;**27**:633–41. doi:10.1038/nbt.1546

26 Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* 2012;**9**:555–66. doi:10.1038/nmeth.2015

27 Geboes K. Histopathology of Crohn ' s Disease and Ulcerative Colitis. *Inflamm Bowel Dis* 2003;**18**:255–76. doi:10.1038/ncpgasthep0528

28 Zhou GX, Liu ZJ. Potential roles of neutrophils in regulating intestinal mucosal inflammation of inflammatory bowel disease. *J Dig Dis* 2017;**18**:495–503. doi:10.1111/1751-2980.12540

29 Mooiweer E, Severs M, Schipper MEI, *et al.* Low fecal calprotectin predicts sustained clinical remission in inflammatory bowel disease patients: a plea for deep remission. *J Crohns Colitis* 2015;**9**:50–5. doi:10.1093/ecco-jcc/jju003

30 Brookes MJ, Whitehead S, Gaya DR, *et al.* Practical guidance on the use of faecal calprotectin. *Frontline Gastroenterol* 2018;**9**:87–91. doi:10.1136/flgastro-2016-100762

31 Louis E. Fecal calprotectin: towards a standardized use for inflammatory bowel disease

management in routine practice. J. Crohns. Colitis. 2015;**9**:1–3. doi:10.1093/ecco-jcc/jju012

32 Buisson A, Gonzalez F, Poullenot F, *et al.* Comparative Acceptability and Perceived Clinical Utility of Monitoring Tools: A Nationwide Survey of Patients with Inflammatory Bowel Disease. *Inflamm Bowel Dis* 2017;**23**:1425–33. doi:10.1097/MIB.0000000000001140

33 Liverani E, Scaioli E, Digby RJ, *et al.* How to predict clinical relapse in inflammatory bowel disease patients. *World J Gastroenterol* 2016;**22**:1017–33. doi:10.3748/wjg.v22.i3.1017

34 Novak G, Parker CE, Pai RK, *et al.* Histologic scoring indices for evaluation of disease activity in Crohn's disease. *Cochrane database Syst Rev* 2017;**7**:CD012351–CD012351. doi:10.1002/14651858.CD012351.pub2

35 Brignola C, Campieri M, Bazzocchi G, *et al.* A laboratory index for predicting relapse in asymptomatic patients with Crohn's disease. *Gastroenterology* 1986;**91**:1490–4. doi:10.1016/0016-5085(86)90206-4

36 Marks DJB, Rahman FZ, Sewell GW, *et al.* Crohn's disease: an immune deficiency state. *Clin Rev Allergy Immunol* 2010;**38**:20–31. doi:10.1007/s12016-009-8133-2

37 Kett K, Brandtzaeg P, Fausa O. J-chain expression is more prominent in immunoglobulin A2 than in immunoglobulin A1 colonic immunocytes and is decreased in both subclasses associated with inflammatory bowel disease. *Gastroenterology* 1988;**94**:1419–25. doi:10.1016/0016-5085(88)90681-6

38 Johansen FE, Braathen R, Brandtzaeg P. Role of J chain in secretory immunoglobulin formation. *Scand J Immunol* 2000;**52**:240–8. doi:10.1046/j.1365-3083.2000.00790.x

39 Dawson DW, Volpert O V, Gillis P, *et al.* Pigment epithelium-derived factor: a potent inhibitor of angiogenesis. *Science* 1999;**285**:245–8. doi:10.1126/science.285.5425.245

40 Nguyen TMB, Subramanian I V, Kelekar A, *et al.* Kringle 5 of human plasminogen, an angiogenesis inhibitor, induces both autophagy and apoptotic death in endothelial cells. *Blood* 2007;**109**:4793–802. doi:10.1182/blood-2006-11-059352

41 Poon IKH, Patel KK, Davis DS, *et al.* Histidine-rich glycoprotein: the Swiss Army knife of mammalian plasma. *Blood* 2011;**117**:2093–101. doi:10.1182/blood-2010-09-303842

42 Alkim C, Alkim H, Koksal AR, *et al.* Angiogenesis in Inflammatory Bowel Disease. *Int J Inflam* 2015;**2015**:970890. doi:10.1155/2015/970890

43 Imhann F, Vich Vila A, Bonder MJ, *et al.* Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut* 2018;**67**:108–19. doi:10.1136/gutjnl-2016-312135

44 Carr SA, Abbatiello SE, Ackermann BL, *et al.* Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. Mol. Cell. Proteomics. 2014;**13**:907–17. doi:10.1074/mcp.M113.036095

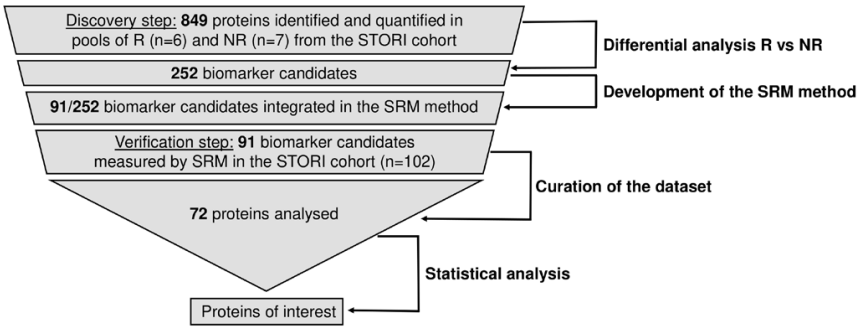


Figure 1. Overview of the biomarker discovery strategy. NR: non-relapsers; R: relapsers; SRM: selected reaction monitoring.

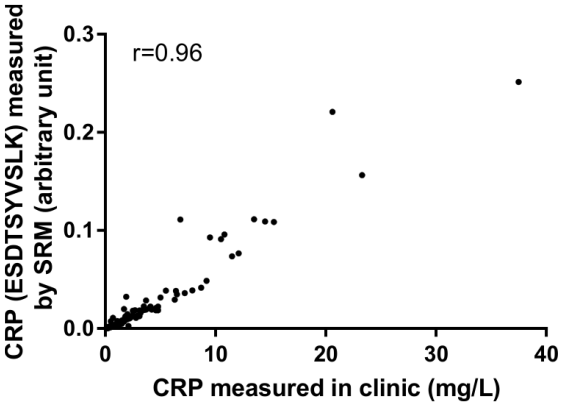


Figure 2. Correlation between CRP measured by SRM and CRP measured in clinic. The strength of association between CRP measured by SRM and CRP measured in clinic was evaluated using Spearman's rank correlation (n=102).

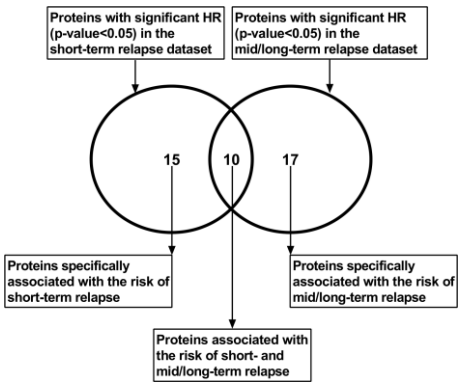
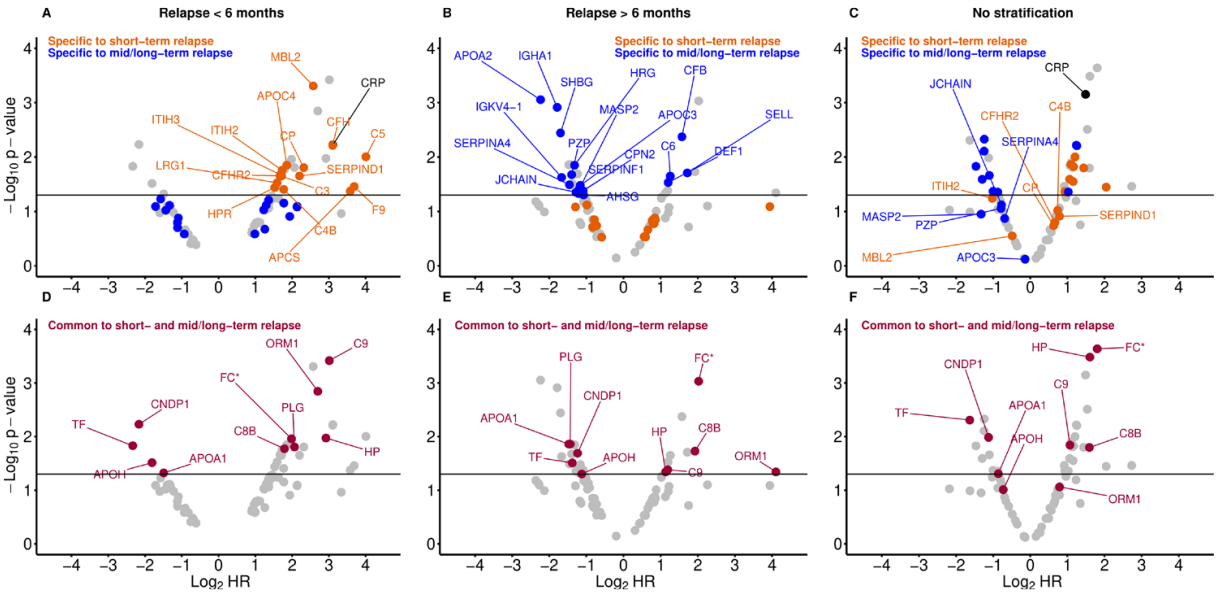


Figure 3. Selection of the markers associated with the risk of short- and/or mid/long-term relapse. Venn diagram showing the strategy to select markers associated with the risk of short- and/or mid/long-term relapse. HR: hazard ratio.



601

602 **Figure 4. Common and distinct markers associated with the risk of short- and mid/long-**
603 **term relapse.** In the stratified (relapse <6 months or >6 months) and non-stratified datasets,
604 volcano plots illustrated the distinct (A, B, C) and the common (D, E, F) markers associated
605 with the risk of short- and mid/long-term relapse. For each protein, the Log_2 hazard ratio (HR)
606 obtained from the univariate Cox model was plotted against its degree of significance ($-Log_{10}$
607 p-value). The horizontal line indicates the significance threshold (p-value=0.05). *Analysis of
608 faecal calprotectin (FC) was based on a subset of the cohort (n=77). AHSG: Alpha-2-HS-
609 glycoprotein; APCS: Serum amyloid P-component; APOA1: Apolipoprotein A-I; APOA2:
610 Apolipoprotein A-II; APOC3: Apolipoprotein C-III; APOC4: Apolipoprotein C-IV; APOH:
611 Beta-2-glycoprotein 1; C3: Complement C3; C4B: Complement C4-B; C5: Complement C5;
612 C6: Complement component C6; C8B: Complement component C8 beta chain; C9:
613 Complement component C9; CFB: Complement factor B; CFH: Complement factor H;
614 CFHR2: Complement factor H-related protein 2; CNDP1: Beta-Ala-His dipeptidase; CP:
615 Ceruloplasmin; CPN2: Carboxypeptidase N subunit 2; CRP: C-reactive protein; DEFA1:
616 Neutrophil defensin 1; F9: Coagulation factor IX; HP: Haptoglobin; HPR: Haptoglobin-
617 related protein; HRG: Histidine-rich glycoprotein; IGHA1: Immunoglobulin heavy constant
618 alpha 1; IGKV4: Immunoglobulin kappa variable 4-1; ITIH2: Inter-alpha-trypsin inhibitor

heavy chain H2; ITIH3: Inter-alpha-trypsin inhibitor heavy chain H3; JCHAIN: Immunoglobulin J chain; LRG1: Leucine-rich alpha-2-glycoprotein; MASP2: Mannan-binding lectin serine protease 2; MBL2: Mannose-binding protein C; ORM1: Alpha-1-acid glycoprotein 1; PLG: Plasminogen; PZP: Pregnancy zone protein; SELL: L-selectin; SERPINA4: Kallistatin; SERPIND1: Heparin cofactor 2; SERPINF1: Serpin family F member 1; SHBG: Sex hormone-binding globulin; TF: Serotransferrin.

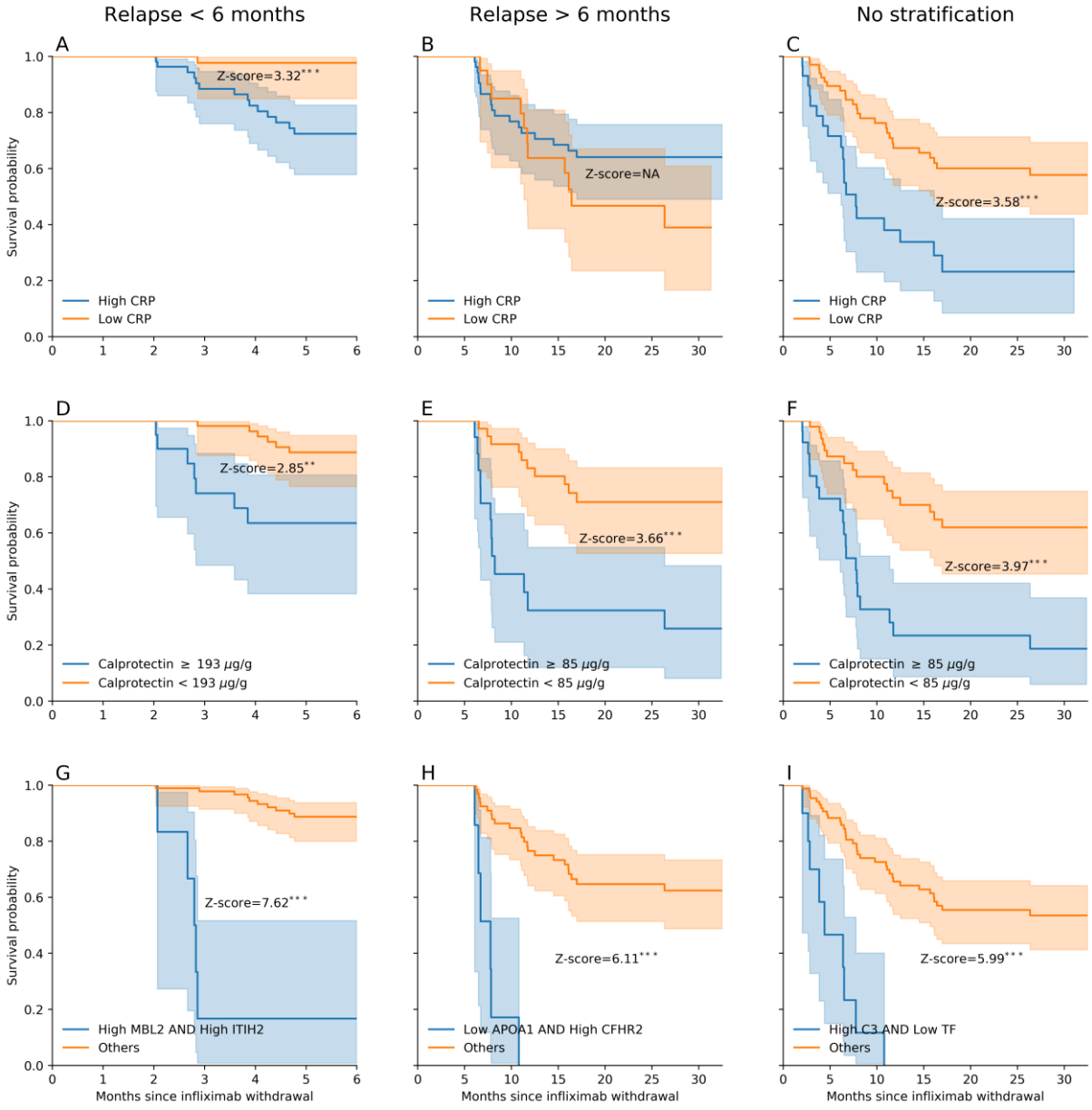


Figure 5. Comparison with CRP and faecal calprotectin highlights a high potential of novel marker combinations in predicting relapse. In the stratified (relapse <6 months or >6

months) and non-stratified datasets, the Kaplan-Meier survival curves were estimated for CRP (A, B, C) and faecal calprotectin (D, E, F) according to their optimal cut-offs. For each dataset, the systematic combinations of two markers were performed with the "AND" or "OR" logical operators (n=2556 combinations per dataset and per logical operator). For the combinations of markers presenting the best discriminatory ability (R vs NR, based on the log-rank Z-score), the Kaplan-Meier survival curves were estimated in the stratified and non-stratified datasets (G, H, I). "High" and "low" mean a circulating level of the protein above or under the cut-off, respectively. Survival curves were compared with the log-rank test. To control the false discovery rate (FDR) associated with the high number of tested combinations, log-rank p-values were adjusted with the Benjamini–Hochberg procedure (G, H, I). This correction was not applied to CRP and faecal calprotectin since these biomarkers were used as references. Analysis of faecal calprotectin was based on a subset of the cohort (n=77). **p-value or FDR<0.01; ***p-value or FDR<0.001. APOA1: Apolipoprotein A-I; CFHR2: Complement factor H-related protein 2; C3: Complement C3; ITIH2: Inter-alpha-trypsin inhibitor heavy chain H2; MBL2: Mannose-binding protein C; TF: Serotransferrin.

Table 1. Patients' characteristics	
Male, n (%)	47 (46)
Age, median years (IQR)	31 (25-39)
Disease duration, median years (IQR)	7 (4-12)
Active smoker, n (%)	38 (37)
CDAI (IQR)	35.7 (14.5-59.5)
Intestinal stricture before infliximab initiation, n (%)	11 (11)
Intestinal stricture at infliximab initiation, n (%)	5 (5)
Disease site	
Ileal, n (%)	13 (13)
Colonic, n (%)	31 (30)
Ileocolonic, n (%)	57 (56)
Upper gastrointestinal tract, n (%)	8 (8)
Perianal lesions, n (%)	36 (35)
Treatment history	
Methotrexate, n (%)	18 (18)
Azathioprine/mercaptopurine, n (%)	84 (82)
Duration of antimetabolite treatment, median years (IQR)	2.6 (1.7-4.5)
Duration of infliximab treatment, median years (IQR)	2.2 (1.6-3.1)
Previous surgical resection, n (%)	21 (21)

Endoscopy	
CDEIS (IQR)	0.6 (0.0-2.8)
CDEIS=0, n (%)	39 (38)
Remaining ulcers, n (%)	31 (30)
Biologic variables	
Haemoglobin level, g/L, median (IQR)	136 (127-144)
Haematocrit, %, median (IQR)	40 (37-43)
Leukocyte count, 10^9/L, median (IQR)	6.0 (4.9-7.3)
Platelet count, 10^9/L, median (IQR)	261 (225-310)
hsCRP level, mg/L, median (IQR)	1.9 (0.8-4.2)
Infliximab trough level, mg/L, median (IQR)	3.7 (1.9-7.9)
Faecal calprotectin, µg/g, n=77, median (IQR)	48.3 (29.7-193.1)

CDAI: Crohn's disease activity index; CDEIS: Crohn's disease endoscopic index of severity; hsCRP: high-sensitivity CRP; IQR: interquartile range

SUPPLEMENTARY METHODS

Supplementary method 1: biomarkers discovery

Sample preparation

An identical volume of each serum samples (n=102) were mixed in 7 and 6 pools of non-relapser (NR) and relapser (R) patients, respectively. These sample pools were depleted of the 20 most abundant proteins of human plasma (ProteoPrep™ 20 Plasma kit, Sigma, USA) as previously described[1]. The protein concentration was determined using the RCDC Protein Assay Kit (BioRad, USA) according to the manufacturer's instructions. Proteins (20 µg) were precipitated using the 2D-clean up assay (GE Healthcare, USA) according to the manufacturer's instructions. Then, samples were reduced, alkylated and digested as previously described[2]. Finally, 3.5 µg of the protein digests were purified on ZipTip C18 (Merck, USA), dried in a vacuum centrifuge and stored at -20 C°. The dried samples were reconstituted using 100 mM ammonium formate adjusted to pH 10. Each sample was spiked with a commercial mixture of protein digest standard, the MassPREP™ Digestion Standard Mixture (MPDSmix, Waters, USA). The MPDSmix contains four non-human proteins digested: yeast alcohol dehydrogenase 1 (ADH1, P00330), bovine serum albumin (BSA, P02769), yeast enolase 1 (ENO1, P00924), rabbit glycogen phosphorylase b (GPB, P00489).

663 To check the quality of the instrumental set-up, two different mix (MPDSmix 1 and
664 MPDSmix 2) were spiked in the R and NR pools.
665
666 *Liquid chromatography and mass spectrometry*
667 Samples (2.5 µg) were injected into the 2D-nanoAquity UPLC chromatography (Waters)
668 coupled online with the Q-TOF Synapt HDMS^{TM} G2 system (Waters) using ion mobility as
669 supplementary separation. The mass spectrometer was equipped with a nano-electrospray
670 source operated in positive ion mode. The capillary voltage and the source temperature were
671 set to 2.9 kV and 100 °C, respectively. The first dimension of UPLC separation was
672 performed at pH 10 on a X-Bridge BEH C18 column (5 µm, 300 µm × 50 mm, Waters). The
673 samples were loaded at 2 µL.min^{-1} in a solution of 20 mM ammonium formate adjusted to pH
674 10. Then, samples were eluted in five steps (10.8, 14.0, 16.7, 20.4 and 65.0% of acetonitrile).
675 After a ten times dilution to pH 3, each eluted fraction was desalted on the trap column
676 Symmetry C18 (5 µm, 180 µm × 20 mm, Waters) and subsequently separated (second
677 dimension of UPLC) on an analytical column BEH C18 (1.7 µm, 75 µm × 250 mm, Waters).
678 The flow rate was constant (250 nL.min^{-1}) and a linear gradient of solvent A (0.1% formic
679 acid in water) and B (0.1% formic acid in acetonitrile) was applied: 0 min, A/B: 99/1%; 1
680 min, A/B: 99/1%; 80 min, A/B: 65/35%; 90 min, A/B: 1/99% and hold for 5 minutes then
681 reconditioning of the column until 120 minutes. The data acquisition was performed in the 50-
682 1500 *m/z* range, a scan time of 0.6 s and a collision energy voltages set in independent
683 alternative scanning mode. The ion-mobility spectrometry (IMS) cell pressure was set to 2.49
684 mbar, the variable IMS wave velocity ranged from 850 to 1 200 m.s^{-1} and the wave height
685 was 40 V. A lock mass correction was applied using [Glu1]-Fibrinopeptide B ([M^{+}2H]$^{2+}$:
686 785.84206 *m/z*).
687

Raw data processing

To maximize the number of proteins identified and quantified, we combined results from ProteinLynx Global SERVER™ (PLGS) 3.01 (Waters) and Progenesis QI 2.1 (Waters). The processing parameters were set as follows: the MS TOF resolution and the chromatographic peak width were set to automatic, the low and elevated energy detection threshold were set to respectively 15 and 150 counts, the identification intensity threshold was 500 counts and the lock mass window at 785.84206 *m/z* was set to 0.3 Da. Proteins were searched in the Uniprot-human database (reviewed entries, release 2014_10) enriched with the 4 proteins of the MPDSmix. The search parameters were set as follows: carbamidomethylation (C) as fixed modification, oxidation (M) and phosphorylation (S, T, Y) as variable modifications, maximum 2 miscleavages, minimum 3 fragments ion matches per peptide, minimum 7 fragments ion matches per protein, minimum 2 peptides per protein and a false discovery rate (FDR) of 4% for protein identification. All the raw data and results of protein identifications have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD019008[3].

Differential analysis

For the differential analysis (R vs NR), we used three methods from distinct softwares: PLGS 3.01, Progenesis QI 2.1 and Perseus 1.5.31[4]. As others, we used different statistical strategies to deal with the common problem of comparing a large quantity of variables in a small number of group[5]. Each statistical method being differently sensitive to the dataset structure (e.g., variance, type of distribution, effect size), using a unique statistical approach risks to eliminate some true biomarker candidates while a discovery step intends to minimize the number of false negative (assuming a high rate of false positive). Furthermore, identification of proteins with PLGS and Progenesis QI is associated with distinct

quantification and normalisation strategies[6,7]. At a statistical level, the differential analysis (R vs NR) is based on the one-factor ANOVA in Progenesis QI while a pairwise comparison based on a Bayesian approach is used in PLGS. In Perseus, two-tailed t-tests were applied to compare R vs NR. For all these statistical comparisons, the risk of type 1 error was set to 5%. Then, proteins exhibiting a differential abundance (R vs NR) in one of the three statistical methods were combined to yield a list of biomarker candidates (n=252, online supplementary table 1).

Supplementary method 2: development of the SRM method

The discovery study and the SRM experiments were performed on two distinct separation systems, respectively: 2D and 1D nanoUPLC. For the biomarker candidates assayed in SRM, the retention time of the precursors was obtained via a 1D separation of digested serum followed by shotgun proteomics.

Sample preparation, liquid chromatography and mass spectrometry

An identical volume of each serum samples (n=102) were mixed in one pool named super pool (SP). The SP was treated as described above excepted that no depletion was performed. Prior to injection, the SP sample was solubilized in 0.01% formic acid aqueous solution and MPDSmix was spiked. This sample was injected into a 1D Acquity UPLC M-Class (Waters) chromatography coupled online with a Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ mass spectrometer (Thermo Fisher Scientific, USA), equipped with a nano-electrospray source operated in positive ion mode. The spray voltage and the temperature of the heated capillary were set to 2.3 kV and 270 °C, respectively. Sample were loaded at 20 $\mu L.min^{-1}$ on the trap column Symmetry C18 (5 µm, 180 µm × 20 mm, Waters) in 100% solvent A (0.1% formic acid in water) during 3 minutes. Then, peptides were separated through the analytical

column HSS T3 C18 (1.8 µm, 75 µm × 250 mm, Waters). The flow rate was constant (600 $nL.min^{-1}$) and the following linear gradient of A and B (0.1% formic acid in acetonitrile) was applied: 0 min, A/B: 98/2%; 5 min, A/B: 93/7%; 135 min, A/B: 70/30%; 150 min, A/B: 60/40%. The total run time was 180 min: 150 min for the linear gradient and 30 min for cleaning and re-equilibration steps. The mass spectrometer method consisted of one full MS scan followed by full MS/MS scans of the 12 most intense ions. The parameters for MS spectrum acquisition were: mass range from 400 to 1 750 *m/z* with R=70 000 (defined at *m/z* 200), automatic gain control (AGC) target of 1×10^6 and a maximum injection time of 50 ms. The parameters for MS/MS spectrum acquisition were: 2.0 *m/z* isolation window with R=17 500 (defined at *m/z* 200), a stepped normalized collision energy of 25.0, AGC target of 1×10^5 and a maximum injection time of 50 ms. Raw data were recorded with Xcalibur Exactive Serie 2.8.1.2806 (Thermo Fisher Scientific).

Raw data processing

Identifications of proteins were obtained using Proteome Discoverer 2.1.0.81 (Thermo Fisher Scientific). Proteins were searched in the Uniprot-human database (reviewed entries, release 2014_10) enriched with the 4 proteins of the MPDSmix. The search parameters were set as follows: trypsin as digestion enzyme, carbamidomethylation (C) as fixed modification, oxidation (M) and phosphorylation (S, T, Y) as variable modifications, maximum 2 miscleavages, peptide length between 6 and 144 amino acids, precursor mass tolerance of 5 ppm, fragment mass tolerance of 0.02 Da, a minimal number of peptides per protein equal to 2 and a FDR of 1% at peptide and PSM level.

Selection of the targeted peptides

For each biomarker candidate, the peptides targeted by the SRM method were selected with the help of empirical and theoretical information: 1) MS/MS identification and retention time obtained in the shotgun experiments (see above); 2) guidelines for the selection of quantotypic peptides[8]; 3) PeptidePicker web interface[9]; 4) data generated during preliminary SRM experiments performed with the 1D chromatography. The final SRM method included 91 biomarker candidates, 208 peptides and 832 fragments. This method was exported from Skyline 4.2[10] to MassLynx^{TM} 4.1 SCN 843 (Waters) using scheduled detection with 4 minutes elution time windows for each targeted peptide.

Supplementary method 3: verification of biomarker candidates using selected reaction monitoring (SRM)

Sample preparation

Serum protein concentrations of the 102 samples were determined using the BCA protein assay (Thermo Fisher Scientific) according to the manufacturer's instructions. The yeast enolase 1 (ENO1, P00924, Sigma, USA) was spiked in each sample and served as internal standard. The serum proteins (20 µg) and the spiked ENO1 (93 ng) were digested with the Trypsin/Lys-C Mix Mass Spec Grade (Promega, USA) according to the manufacturer's instructions. Then, 3.5 µg of the resulting peptide mixtures were purified on ZipTip C18 (Thermo Fisher Scientific), dried in a vacuum centrifuge and stored at -20 C°. Before injection, dried samples (3.5 µg) were reconstituted in 41.6 µL of 0.1% formic acid and 5.2 µL of a solution containing a synthetic stable isotope-labelled peptide of ENO1 (Thermo Fisher Scientific): VNQIGTLSESI**K** (^{15}N- and ^{13}C-labeled lysine residue, 50 fmol.µL^{-1}) solubilised in 5% acetonitrile. The quantity of VNQIGTLSESI**K** injected was 50 fmol. This peptide was used to check the instrumental set-up stability across sample injections.

Liquid chromatography and mass spectrometry

Samples (0.675 µg) were injected into the 1D nanoAcquity UPLC M-class chromatography coupled online with a triple quadrupole mass spectrometer Xevo^{TM} TQ-S (Waters) equipped with a nano-electrospray source operated in positive ion mode. The temperature of the source and the capillary voltage were set to 100 °C and 2.9 kV, respectively. The separation of peptides in the 1D nanoAcquity UPLC system was performed as described in the Supplementary method 2. The SRM transitions were selected with a mass window of 0.75 Da in Q1 and Q3. In the Q2, the collision gas pressure was 3.5e-3 (argon gas flow set at 0.15 mL.min^{-1}) and peptides were fragmented with a predicted collision energy[11]. The mass spectrometer operated in a scheduled mode with a retention time window of 4 min. The dwell time and the cycle time varied from 3 to 8 ms and 0.45 to 1.35 s, respectively. Given that the minimum peak width was ~30 s, the number of points per peak was ~22 in the worst case while experts recommend a minimum of 8[8].

Raw data processing

The raw files were imported in Skyline 4.2[10] and the SRM traces were manually integrated. Addition of the two most intense fragments was used for the quantification of peptide. The signal of each targeted peptide was normalised with the mean intensity of three peptides (NVNDVIAPAFVK, TFAEALR, VNQIGTLSESIK) of the internal standard (ENO1 from yeast). To determine protein quantification, the normalised intensities of peptides belonging to the same protein was summed.

Supplementary method 4: quality controls of the SRM method

Intra-day precision

The precision of the SRM method was determined with a repeatability test. A serum pool of NR patients (n=7) and a serum pool of R patients (n=7) were processed in technical triplicates to determine the coefficient of variation (CV) of each peptide in biological conditions representative of R and NR patients. In the biomarkers verification step, experts recommend a high-to-moderate precision (CV<20-35%)[12]. In the present study, we followed this guideline by eliminating peptides with a CV>30% (n=17) in either the pool of relapsers or non-relapsers. The 154 peptides finally analysed were measured with a mean CV (± SD) of 6.0% (4.1) and 7.1% (5.8) in the pool of relapsers and non-relapsers, respectively. In this regard, our intra-day precision shows similar performance to high-quality SRM data[5,13] and ELISA (enzyme-linked immunosorbent assay).

Normalisation method

To correct for differences in sample amount injected, we normalised the data with an internal standard (ENO1 from yeast, online supplementary method 3) spiked in every sample. In the absence of established guideline, such a strategy needs to be evaluated. If this internal standard do reflect the sample amount, normalisation should correct for differences in injected quantity and thus it should ameliorate the precision. In the repeatability test, the average CV without normalisation and with normalisation were respectively: 8.9 vs 6.0% for the relapser pool; 10.9 vs 7.1% for the non-relapser pool. These results clearly indicate that our normalisation method is appropriate.

Inter-day precision

Since we divided the sample preparation over a five days period and that samples were injected over a period of three weeks, we controlled the inter-day variability associated with this procedure. In each day of sample processing, the same sample (SP) was treated in parallel

of the others and this control was injected every 25 samples (5 times in all). Regarding this quality control, the mean CV (±SD) for the 154 peptides was 14.5% (6.3). Therefore, the intra- and inter-day precision were under 20% thus respecting the criteria established for a verification step of the biomarker discovery pipeline[12].

Run-to-run variation

In SRM, the run-to-run variation needs to be carefully evaluated[8,12], especially in our case where samples were injected during three weeks. To this end, we spiked a heavy peptide (VNQIGTLSESI**K**) in all samples just before injection on the LC-MS system (online supplementary method 3). The CV of this internal standard across all samples was 9.8%. To control the stability of the instrumental set-up in the whole SRM method, we injected every 24 samples (5 times in all) a zip-tip replicate of the SP sample. Regarding this quality control, we obtained a mean CV (±SD) of 10.0% (7.6) for the 154 peptides analysed. Altogether, these results indicate that our instrumental set-up showed stable performances.

Guaranties on target identity and quantotypic properties

In SRM, guaranties are needed to establish the identity of the target[12]. In the present study, the selected transitions were confidently identified by experimental MS/MS data from the discovery study (online supplementary method 1). Furthermore, co-elution of fragments in SRM was systematically observed for the 154 analysed peptides. Even if a peptide is correctly identified and measured it does not mean that its level is stoichiometrically equivalent to the protein level (quantotypic). Although no guideline exists to evaluate empirically the quantotypic properties of a peptide, this assumption should be tested to ensure a reliable inference on protein abundance[14]. The measure of different quantotypic peptides of a same protein must converge in term of quantification. Thus, checking this postulate is a way to

evaluate the quantotypic properties of the selected peptides[14]. As already proposed[14], we performed correlation matrices between the signal intensity of peptides belonging to the same protein across the 112 samples injected. If this correlation was weak ($r<0.5$), peptides were eliminated of the analysis (n=8). Among the 72 proteins finally analysed (without the internal standard ENO1), 51 were measured with 2 (25 proteins) or 3 (26 proteins) peptides presenting signal intensities significantly and strongly correlated (mean ± SD coefficient of correlation: 0.83 ± 0.11) across the 112 samples injected. On the other hand, 21 proteins were measured with 1 peptide thus preventing to apply the correlation matrices described above. However, if really quantotypic, one peptide can be used to quantify a protein. This assertion is supported by the high correlation between CRP measured with one peptide (ESDTSYVSLK) in our SRM method and CRP measured in clinic (Figure 2). In addition, the studied biomarkers showed variations (R vs NR patients) in agreement with the literature and the pathophysiological processes involved in CD (see results and discussion). Taken together, these empirical controls sustain that the selected peptides present quantotypic properties.

Curation of the dataset

According to quality controls (see above) and verification of the dataset, some peptides of the SRM method were excluded from the analysis: 1) peptides measured with a CV>30% (n=17, see above); 2) peptides with low quality signal, i.e., 1 or several transitions not detected (n=5), too low intensity (n=2), two peaks (n=2), a retention time outside the window (n=1), partial co-elution of fragments (n=1); 3) peptides belonging to proteins present in erythrocytes (n=11) and impacted by haemolysis of some samples; 4) peptides not belonging to the three "best flyers" (most intense) when four peptides per protein were measured (n=9); 5) peptides of a same protein presenting signal intensities weakly correlated ($r<0.5$, n=8, see above) across the 112 samples injected; 6) peptides from potential contaminants (e.g., keratin, n=3).

The curated SRM method targeted 73 proteins (including the internal standard ENO1), 154 peptides and 616 fragments (online supplementary table 2).

SRM data availability

All the SRM raw files and the Skyline files (including those of the quality controls) are available on Panorama with the dataset identifier PXD019434 (https://panoramaweb.org/atQBcH.url)[15].

REFERENCES

1 Baiwir D, Mazzucchelli G, Smargiasso N, *et al.* Comparison of serum fractionation methods by data independent label-free proteomics. *EuPA Open Proteomics* 2015;**9**:14–22. doi:10.1016/J.EUPROT.2015.07.009

2 Fleron M, Greffe Y, Musmeci D, *et al.* Novel post-digest isotope coded protein labeling method for phospho- and glycoproteome analysis. *J Proteomics* 2010;**73**:1986–2005. doi:10.1016/J.JPROT.2010.06.003

3 Vizcaíno JA, Csordas A, Del-Toro N, *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* 2016;**44**:D447–56. doi:10.1093/nar/gkv1145

4 Tyanova S, Temu T, Sinitcyn P, *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 2016;**13**:731–40. doi:10.1038/nmeth.3901

5 Juste C, Kreil DP, Beauvallet C, *et al.* Bacterial protein signals are associated with Crohn's disease. *Gut* 2014;**63**:1566 LP – 1577. doi:10.1136/gutjnl-2012-303786

6 Silva JC, Denny R, Dorschel CA, *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem* 2005;**77**:2187–200. doi:10.1021/ac048455k

7 Di Luca A, Henry M, Meleady P, *et al.* Label-free LC-MS analysis of HER2+ breast

cancer cell line response to HER2 inhibitor treatment. *Daru* 2015;**23**:40. doi:10.1186/s40199-015-0120-y

8 Lange V, Picotti P, Domon B, *et al.* Selected reaction monitoring for quantitative proteomics: A tutorial. Mol. Syst. Biol. 2008;**4**. doi:10.1038/msb.2008.61

9 Mohammed Y, Domanski D, Jackson AM, *et al.* PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. *J Proteomics* 2014;**106**:151–61. doi:10.1016/j.jprot.2014.04.018

10 MacLean B, Tomazela DM, Shulman N, *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010;**26**:966–8. doi:10.1093/bioinformatics/btq054

11 Maclean B, Tomazela DM, Abbatiello SE, *et al.* Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Anal Chem* 2010;**82**:10116–24. doi:10.1021/ac102179j

12 Carr SA, Abbatiello SE, Ackermann BL, *et al.* Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. Mol. Cell. Proteomics. 2014;**13**:907–17. doi:10.1074/mcp.M113.036095

13 Ludwig C, Claassen M, Schmidt A, *et al.* Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry. *Mol Cell Proteomics* 2012;**11**:M111.013987. doi:10.1074/mcp.M111.013987

14 Worboys JD, Sinclair J, Yuan Y, *et al.* Systematic evaluation of quantotypic peptides for targeted analysis of the human kinome. *Nat Methods* 2014;**11**:1041–4. doi:10.1038/nmeth.3072

15 Sharma V, Eckels J, Taylor GK, *et al.* Panorama: a targeted proteomics knowledge base. *J Proteome Res* 2014;**13**:4205–10. doi:10.1021/pr5006636

936

937

938

939