RESEARCH ARTICLE

Exploration of Serum Proteomic Profiling and Diagnostic Model That Differentiate Crohn's Disease and Intestinal Tuberculosis

Fenming Zhang[1], Chengfu Xu[1], Longgui Ning[1], Fengling Hu[1], Guodong Shan[1], Hongtan Chen[1], Ming Yang[1], Wenguo Chen[1], Jiekai Yu[2], Guoqiang Xu[1]*

1 Department of Gastroenterology, The First Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang Province, China, **2** Department of Tumor Research Institute, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang Province, China

* xuguoqi@mail.hz.zj.cn

Abstract

Aim

To explore the diagnostic models of Crohn's disease (CD), Intestinal tuberculosis (ITB) and the differential diagnostic model between CD and ITB by analyzing serum proteome profiles.

Methods

Serum proteome profiles from 30 CD patients, 21 ITB patients and 30 healthy controls (HCs) were analyzed by using weak cationic magnetic beads combined with MALDI-TOF-MS technique to detect the differentially expressed proteins of serum samples. Three groups were made and compared accordingly: group of CD patients and HCs, group of ITB patients and HCs, group of CD patients and ITB patients. Wilcoxon rank sum test was used to screen the ten most differentiated protein peaks ($P < 0.05$). Genetic algorithm combining with support vector machine (SVM) was utilized to establish the optimal diagnostic models for CD, ITB and the optimal differential diagnostic model between CD and ITB. The predictive effects of these models were evaluated by Leave one out (LOO) cross validation method.

Results

There were 236 protein peaks differently expressed between group of CD patients and HCs, 305 protein peaks differently expressed between group of ITB patients and HCs, 332 protein peaks differently expressed between group of CD patients and ITB patients. Ten most differentially expressed peaks were screened out between three groups respectively ($P < 0.05$) to establish diagnostic models and differential diagnostic model. A diagnostic model comprising of four protein peaks (M/Z 4964, 3029, 2833, 2900) can well distinguish CD patients and HCs, with a specificity and sensitivity of 96.7% and 96.7% respectively. A diagnostic model comprising four protein peaks (M/Z 3030, 2105, 2545, 4210) can well distinguish ITB patients and HCs, with a specificity and sensitivity of 93.3% and 95.2% respectively. A differential diagnostic model comprising three potential biomarkers protein peaks (M/Z 4267,

4223, 1541) can well distinguish CD patients and ITB patients, with a specificity and sensitivity of 76.2% and 80.0% respectively. Among the eleven protein peaks from the diagnostic models and differential diagnostic model, two have been successfully purified and identified, Those two peaks were M/Z 2900 from the diagnostic model between CD and HCs and M/Z 1541 from the differential diagnostic model between CD and ITB. M/Z 2900 was identified as appetite peptide, M/Z 1541 was identified as Lysyl oxidase-like 2 (LOXL-2).

Conclusion

The differently expressed protein peaks analyzed by serum proteome with weak cationic magnetic beads combined MALDI-TOF-MS technique can effectively distinguish CD patients and HCs, ITB patients and HCs, CD patients and ITB patients. The diagnostic model between CD patients and HCs consisting of four protein peaks (M/Z 4964, 3029, 2833, 2900), the diagnostic model between ITB patients and HCs comprising four protein peaks (M/Z 3030, 2105, 2545, 4210) and the differential diagnostic model between CD patients and ITB patients comprising three protein peaks (M/Z 4267, 4223, 1541) had high specificity and sensitivity and can contribute to diagnoses of CD, ITB and the differential diagnosis between CD and ITB. Two proteins from the diagnostic model of CD and the differential diagnostic model between CD and ITB were identified. Further experiments are required using a larger cohort of samples.

Introduction

Crohn's disease (CD) and Intestinal tuberculosis (ITB) are chronic inflammatory intestinal diseases[1–4], it is difficult to differentiate them due to their high similarity in clinical manifestations, morphological features and histological characteristics[1–4]. Both CD and ITB have symptoms like abdominal pain, diarrhea, fever and weight loss, leading to low specificity of clinical manifestations. It is believed that appearence of longitudinal ulcers, aphthous ulcers and a cobblestone-like appearance are seen in typical CD with an endoscopy, and linear or circular ulcers are more inclined to the diagnosis of ITB[5–7], however, in clinical practice, those findings are often atypical and with low positive rates. Both CD and ITB show granulomatous inflammation pathologically, though caseous granulomas is regarded as the gold standard for the diagnosis of ITB[8], but due to low positive rate[9,10], the differential diagnosis between CD and ITB becomes really difficult[11,12]. However, the treatment, prognosis and natural history of these two entities are quite different. So, it is necessary to discover new specific biomarkers which can differentiate CD from ITB.

Serum proteomic fingerprint, which aims to analyze the proteome signature of each patient, although technically difficult, could be a realistic approach to understand the biological processes involved in inflammatory processes[13,14]. It is a new type of proteomics technology developed in recent years, aiming to explore disease-related proteins, study protein expression or change after modification, look for diagnostic markers and establish diagnostic models[15,16]. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) is an important proteomic technology. Many protein biomarkers of certain diseases have been indicated by using MALDI-TOF MS to analyze the serum proteome. Magnetic beads have large surface area and can capture more small molecular peptides and proteins[17,18]. The combination of MALDI-TOF-MS and magnetic beads can take advantages of both and therefore detect more low molecular weight proteins in serum. Few research used

this technology to detect potentially useful markers that can help differentiating ITB and CD. In this study, we used weak cationic magnetic beads combined with MALDI-TOF-MS technique to analyze serum proteome of CD patients, ITB patients and healthy controls (HCs). Based on the genetic algorithm combined with support vector machine (SVM) model, we selected the best diagnostic models and differential diagnostic models which were validated with Leave one out (LOO) method, expecting to provide new ideas about the early diagnoses of CD, ITB and the differential diagnosis between CD and ITB at serum protein level.

Materials and Methods

Patients

Patients with CD and ITB at the Department of Gastroenterology, the First Affiliated Hospital of Zhejiang University School of Medicine were enrolled from December 2011 to December 2014. The diagnosis of CD was made by gastroenterologists specialized in Inflammatory Bowel Disease (IBD), based on the European Crohn's and Colitis Organization guideline which is a combined evaluation of clinical, endoscopic and histological features[19]. The diagnostic criteria of ITB is a combination of clinical manifestations, endoscopic presentation, histological features, micro-biological tests results and response to anti-tuberculous treatment[20,21]. This study was approved by the Ethics committee of the First Affiliated Hospital of Zhejiang University School of Medicine (Reference Number: 2014–303) and informed consent was obtained from all patients. It was an observational study and verbal informed consent was received from every participant. All patients were in inpatients department and undergone routine blood examination during hospital stay, so when the nursing staff performed blood drawing for every patient for the first time, there was a specialist consulting with patients' willingness of using their specimens in this study, we collected venous blood from those who were willing to use their serum specimens in our research. Subjects without any familial disposition for IBD, daily medication, or any known diseases and did healthy check-up in this hospital at the corresponding period were selected as HCs. This study did no harm to the participants and the procedures were approved by the Ethics Committee of Zhejiang University School of Medicine. As shown in Table 1, totally, 30 CD patients (male/female 16/14, mean age 36.3±13.2, aged 21–61) and 21 ITB patients (male/female 11/10, mean age 42.1±14.6, aged 25–70) were enrolled. All the patients were newly diagnosed cases, CD patients had not received infliximab therapy or Acetazolamide (AZA)/ 6-Mercaptopurine (6-MP)/ Methotrexate (MTX) and ITB patients had not received anti-TB treatment. HCs were composed of 30 persons (male/female 15/15, mean age 38.5±13.9, aged 24–66). There was no significant difference in either age ($P = 0.218$) or gender ratio ($P = 0.102$) between three groups.

Collection and preservation of serum samples

Blood sample of about 2mL venous blood were collected in vacuum tubes without anticoagulant and centrifuged for 6 min, with a rotate speed of 3600rpm. Then, the serum on the upper

Table 1. Gender and age of the participants

Group	Male/female	Average age
CD	16/14	36.3±13.2 (21–61)
ITB	11/10	42.1±14.6 (25–70)
HCs	15/15	38.5±13.9 (24–66)

CD: Crohn's disease. ITB: Intestinal tuberculosis. HCs: Healthy Controls.

doi:10.1371/journal.pone.0167109.t001

layer was suck up and put into the eppendorf tubes and was preserved at −80˚C for further analysis. Blood sample from only those patients were included when the diagnosis of CD or ITB was confirmed and the follow up was complete. There was no hemolysis of blood specimens, no upper layer grease and lower layer cells were suck up. The above-mentioned steps were finished within 2 hours and all the samples were freeze and thawed no more than twice.

Separation of serum protein

Preparation of serum sample. 5μL serum sample was took from the −80˚C freezer and put into the sample tube which containing magnetic beads and magnetic beads in combination with buffer solution, after suction and blending, the sample tube was preserved indoor temperature.

Preparation of weak cationic magnetic beads. Firstly, 10μL magnetic beads in combination with buffer solution was put into the 200μL sample tube, then, 10μL magnetic beads and 5μL serum sample were added blended for suction and blending. Secondly, a magnetic separator was put into the sample tube and the suspended liquid was sucked after separation, then, 100μL magnetic bead cleaning buffer was put into the sample tube and the suspended liquid was sucked after separation. Repeat the above steps twice to ensure that the suspended liquid was completely sucked. Lastly, 5μL magnetic bead elution buffer and the magnetic separator were added, after complete separation of the magnetic beads and suspended liquid, the suspended fluid sample was moved into a new 10μL tube, and 5μL magnetic beads stable buffer was added for suction and blending.

Standard substance sample application. The standard substance was dissolved at the room temperature and mixed with 1μL standard substance and 10μL matrix. Then, 1μL mixed liquor was put at the standard position of Anchorchip target and dried at room temperature. The Anchorchip target was put into the mass spectrometer (microflex MALDI-TOF-MS, German), using LP-Clinprot method. Then, the protein peaks with a relative molecular mass was collected at a range of Dalton 1000–13000. The same sample was with the same crystallization point, 5–10 points at every crystallization point were collected and the protein map was generated. The average molecular weight deviation was less than 0.0001.

Serum sample application. 1μL serum sample disposed with the magnetic bead was put at the standard position of Anchorchip target and dried at room temperature. Anchorchip target was put into the mass spectrometer, choosing linear model, the protein peak with a relative molecular mass at a range of Dalton 800–10000 was collected. The same sample was with the same crystallization point, 8 points were collected, accumulating to 400. The average molecular weight deviation is less than 0.0001.

Data processing

Firstly, we uploaded the original mass spectrogram to the server and removed noise caused by mass spectrometer detector using wavelet transform. After removing the noise, we corrected the baseline of mass spectrometry. Then, we corrected the molecule weight value of the whole spectrogram, found the protein peaks with local extremum method and filtered protein peaks with signal-to-noise ratio, thus, the ratio of each protein peak appeared in every sample. Lastly, we homogenized the intensity.

Mass spectra analysis and establishment of diagnostic models

Preliminary statistical analysis was performed using Biomarker Wizard 3.1 software (calculation of $\overline{X}$, S and P value), the corresponding protein peak intensity of each M/Z was showed as

$\overline{X} \pm S$. Differentially expressed ($P < 0.05$) protein peaks were screened out with Wilcoxon rank-sum test. Genetic algorithm combining with SVM was utilized to establish the optimal diagnostic models for CD, ITB and the optimal differential diagnostic model between CD and ITB. Due to the limited number of samples, the separation of the samples into independent and training test sets was not possible, so the evaluation of the classification was limited to LOO cross validation to calculate sensitivity, specificity and accuracy. One number of spectra was removed from the training set, a classifier was generated using the remaining instances, and the performance of this classifier was evaluated by applying it to the left out instances and comparing classifier and true labels. The protein peak combination which showed the highest cut-off value predicted by SVM was regarded as the final candidate biomarkers. $P < 0.05$ was considered as statistically significant.

Identification of the biomakers

The sequences of differential peptides were identified by using ultrafleXtreme MALDI-TOF/TOF directly. After calibrated by peptide standard (Bruker Daltonic, Germany), we used the reflection mode to detect the isotopic peaks, and turned to the lift model to detect the MS/MS fragment. The BioTools and mascot software was used to search the database.

Results

Serum protein fingerprint

Serum protein peaks combined to weak cationic magnetic beads were detected by MALDI-TOF-MS. The serum protein peaks of 30 CD patients, 21 ITB patients and 30 HCs were shown in Fig 1.

Screening of differentially expressed protein peaks

There were 236 differentially expressed ($P < 0.05$) serum protein peaks between 30 CD patients and 30 HCs, 305 differentially expressed serum protein peaks between 21 ITB and 30 HCs, 332 differentially expressed serum protein peaks between 30 CD patients and 21 ITB patients found by mass spectrometer. Ten most differentially expressed protein peaks were further chosen to establish diagnostic and differential diagnostic models, as shown in Tables 2–4. Among the ten most differentially expressed protein peaks between 30 CD patients and
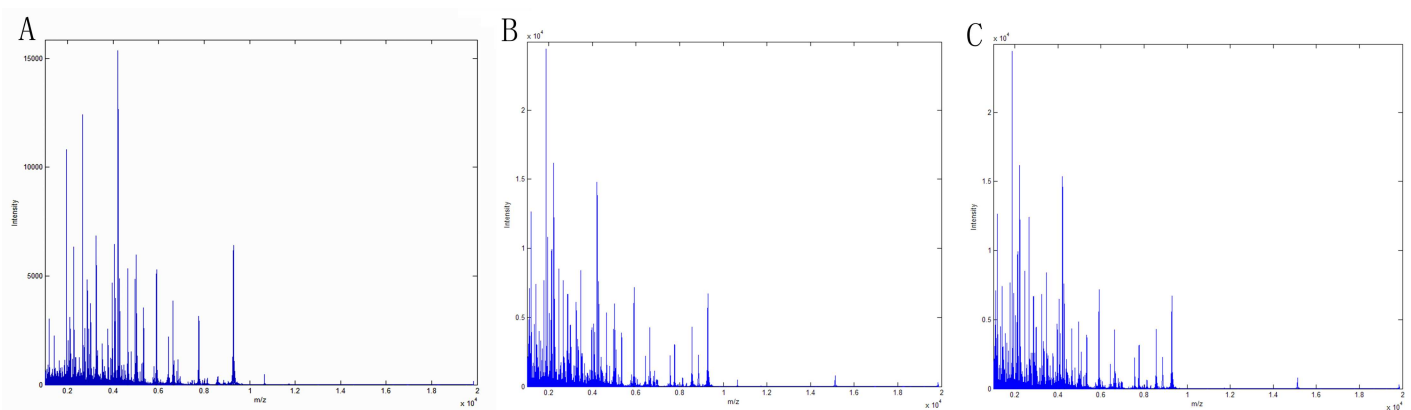

Fig 1. Overlay protein map of 3 groups. A Crohn's disease patients vs Healthy controls. B Intestinal tuberculosis patients vs Healthy controls. C Crohn's disease patients vs Intestinal tuberculosis patients

Table 2. Ten differentially expressed serum protein peaks of CD patients and HCs ($\overline{X}\pm S$)

M/Z	$\overline{X}\pm S$		P
	CD(N = 30)	HCs(N = 30)	
4976	369.99±252.56	91.56±49.67	$1.31*10^{-8}$
4964	1585.34±1255.30	446.25±225.48	$7.70*10^{-8}$
3029	321.63±228.67	747.92±327.02	$7.04*10^{-7}$
4988	108.34±51.78	42.37±28.54	$8.20*10^{-7}$
3315	325.63±180.45	666.65±275.90	$3.83*10^{-6}$
2833	135.66±81.02	50.74±43.03	$1.02*10^{-5}$
6630	1019.33±436.15	1746.38±721.76	$1.53*10^{-5}$
4069	299.74±151.92	160.36±71.69	$2.77*10^{-5}$
2900	399.23±183.07	207.46±114.64	$3.59*10^{-5}$
5065	232.41±251.78	505.92±262.69	$3.59*10^{-5}$

CD: Crohn's disease. HCs: Healthy Controls.

doi:10.1371/journal.pone.0167109.t002

30 HCs, six protein peaks (M/Z 4976, 4964, 4988, 2833, 2900, 4069) were highly expressed, four protein peaks (M/Z 3029, 3315, 6630, 5065) were lowly expressed (Table 2). Among the ten most differentially expressed protein peaks between 21 ITB patients and 30 HCs, one protein peak (M/Z 2545) was highly expressed, nine protein peaks (M/Z 3030, 2105, 5065, 1945, 1680, 3275, 4210, 6379, 4248) were lowly expressed (Table 3). Among the ten most differentially expressed protein peaks between 30 CD patients and 21 ITB patients, nine protein peaks (M/Z 4210, 4195, 4267, 2105, 2933, 4223, 6380, 2952, 4248) were highly expressed, one protein peak (M/Z 1541) was lowly expressed (Table 4).

Establishment of diagnostic models of CD, ITB and differential diagnostic model between CD and ITB

Ten most differentially expressed protein peaks of the three groups were chosed respectively to establish the optimal diagnostic models for CD, ITB and the optimal differential diagnostic model to distinguish CD from ITB using genetic algorithm combining with SVM. The predictive effects of these models were evaluated by LOO cross validation method. The protein peaks combination which showed the highest cut-off value predicted by SVM was regarded as the final candidate biomarkers combination (Figs 2–4).

Table 3. Ten differentially expressed serum protein peaks of ITB patients and HCs ($\overline{X}\pm S$)

M/Z	$\overline{X}\pm S$		P
	ITB(N = 21)	HCs(N = 30)	
3030	254.30±214.62	828.02±352.65	$2.03*10^{-7}$
2105	439.37±425.78	1325.04±443.44	$3.75*10^{-7}$
2545	355.37±258.72	100.56±81.06	$1.01*10^{-6}$
5065	117.29±182.40	544.69±303.70	$1.97*10^{-6}$
1945	1128.58±1453.16	3689.98±1842.34	$4.17*10^{-6}$
1680	56.62±41.68	162.02±81.39	$1.03*10^{-5}$
3275	261.96±252.91	999.05±774.40	$1.03*10^{-5}$
4210	3333.29±3034.13	7918.54±2663.17	$1.03*10^{-5}$
6379	30.20±41.31	112.59±64.51	$1.22*10^{-5}$
4248	185.86±170.31	419.39±177.54	$2.24*10^{-5}$

ITB: Intestinal tuberculosis. HCs: Healthy Controls.

doi:10.1371/journal.pone.0167109.t003

Table 4. Ten differentially expressed serum protein peaks of CD patients and ITB patients ($\overline{x}\pm S$)

M/Z	$\overline{x}\pm S$		P
	CD(N = 30)	ITB(N = 21)	
4210	8342.18±2548.39	3291.11±3046.65	$3.46*10^{-6}$
4195	558.47±159.18	239.90±265.59	$5.00*10^{-6}$
4267	853.94±489.67	335.87±234.22	$6.57*10^{-6}$
2105	1185.54±417.36	433.09±428.49	$7.86*10^{-6}$
2933	734.23±572.21	237.39±533.94	$7.86*10^{-6}$
4223	787.90±329.38	326.99±266.34	$1.22*10^{-5}$
6380	108.64±72.41	30.12±41.40	$1.73*10^{-5}$
2952	681.89±416.57	276.08±425.83	$6.59*10^{-5}$
4248	382.41±173.16	178.09±172.40	$7.15*10^{-5}$
1541	70.37±56.20	153.44±70.42	$9.83*10^{-5}$

CD: Crohn's disease. ITB: Intestinal tuberculosis.

doi:10.1371/journal.pone.0167109.t004

The discriminant results by SVM of the three groups were obviously showed on the scatter diagram (Figs 5–7). We found a diagnostic model consisting of four protein peaks (M/Z 4964, 3029, 2833, 2900) can well distinguish CD patients and HCs. Among those four protein peaks, three protein peaks (M/Z 4964, 2833, 2900) were highly expressed and one protein peak (M/Z 3029) was lowly expressed in CD patients compared with that in HCs (Table 2). When using this diagnostic model to detect 30 CD patients and 30 HCs, 1 CD patient was misjudged as HC, and 1 HC was misjudged as CD patient, with a specificity and sensitivity of 96.7% and 96.7% respectively (Fig 5), indicating this diagnostic model had a high accuracy. A diagnostic model consisting of four protein peaks (M/Z 3030, 2105, 2545, 4210) can well distinguish ITB patients and HCs. Among those four protein peaks, one protein peak (M/Z 2545) was highly expressed and three protein peaks (M/Z 3030, 2105, 4210) were lowly expressed in ITB patients compared with that in HCs (Table 3). When using this diagnostic model to detect 21 ITB patients and 30 HCs, 1 ITB patient was misjudged as HC and 2 HCs were misjudged as ITB patients, with a specificity and sensitivity of 93.3% and 95.2% respectively (Fig 6), indicating this diagnostic model had a high accuracy. A differential diagnostic model consisting of three protein peaks (M/Z 4267, 4223, 1541) can well distinguish CD patients from ITB patients. Among the three protein peaks, two protein peaks (M/Z 4267, 4223) were highly expressed and one protein peak (M/Z 1541) was lowly expressed in CD patients compared with that in ITB patients (Table 4). When using this diagnostic model to detect 30 CD patients and 21 ITB patients, 6 CD patients were misjudged as ITB patients, and 5 ITB patients were misjudged as CD patients, with a specificity and sensitivity of 76.2% and 80% respectively (Fig 7), indicating this differential diagnostic model had a relatively high accuracy. The specificity, sensitivity, positive predictive value and negative predictive value of the diagnostic models of CD, ITB and the differential diagnostic between CD and ITB were shown in Table 5.

Identification of the biomarkers

Among the eleven protein peaks of the diagnostic models and differential diagnostic model, two have been successfully purified and identified. They two peaks were M/Z 2900 from the diagnostic model between CD and HCs and M/Z 1541 from the differential diagnostic model between CD and ITB. Table 6 showed the Mascot Search Results of the three protein peaks. M/Z 2900 was identified as appetite peptide, M/Z 1541 was identified as LOXL-2.
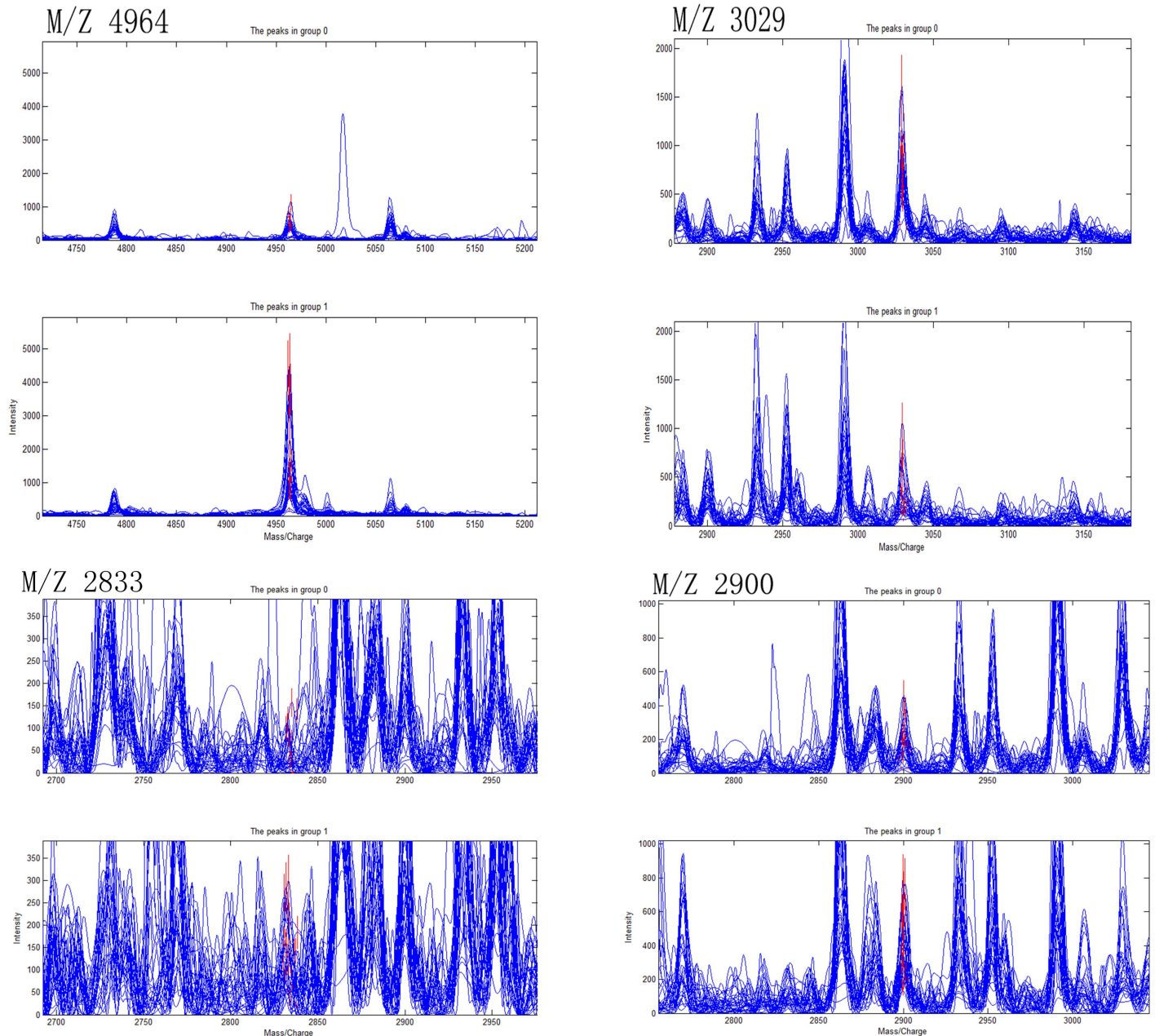
Fig 2. Four selected peaks (M/Z 4964, 3029, 2833, 2900) that compose of the SVM model for the diagnose of Crohn's disease. Group 1 representatives Crohn's disease patients, group 0 representatives Healthy controls.

doi:10.1371/journal.pone.0167109.g002

Discussion

Proteomic analysis, as a new comprehensive technology used for the identification of biomarkers, has been widely used in many fields since 1994. There are more than 20000 kinds of proteins in serum, which can reflect the serological changes of different stages for different diseases[22–25]. Using proteomics technology to screening specific proteins can provide new ideas and methods for the early diagnosis, treatment and prognosis of diseases. MALDI--TOF-MS technology is a newly developing kind of proteomics research methods in recent
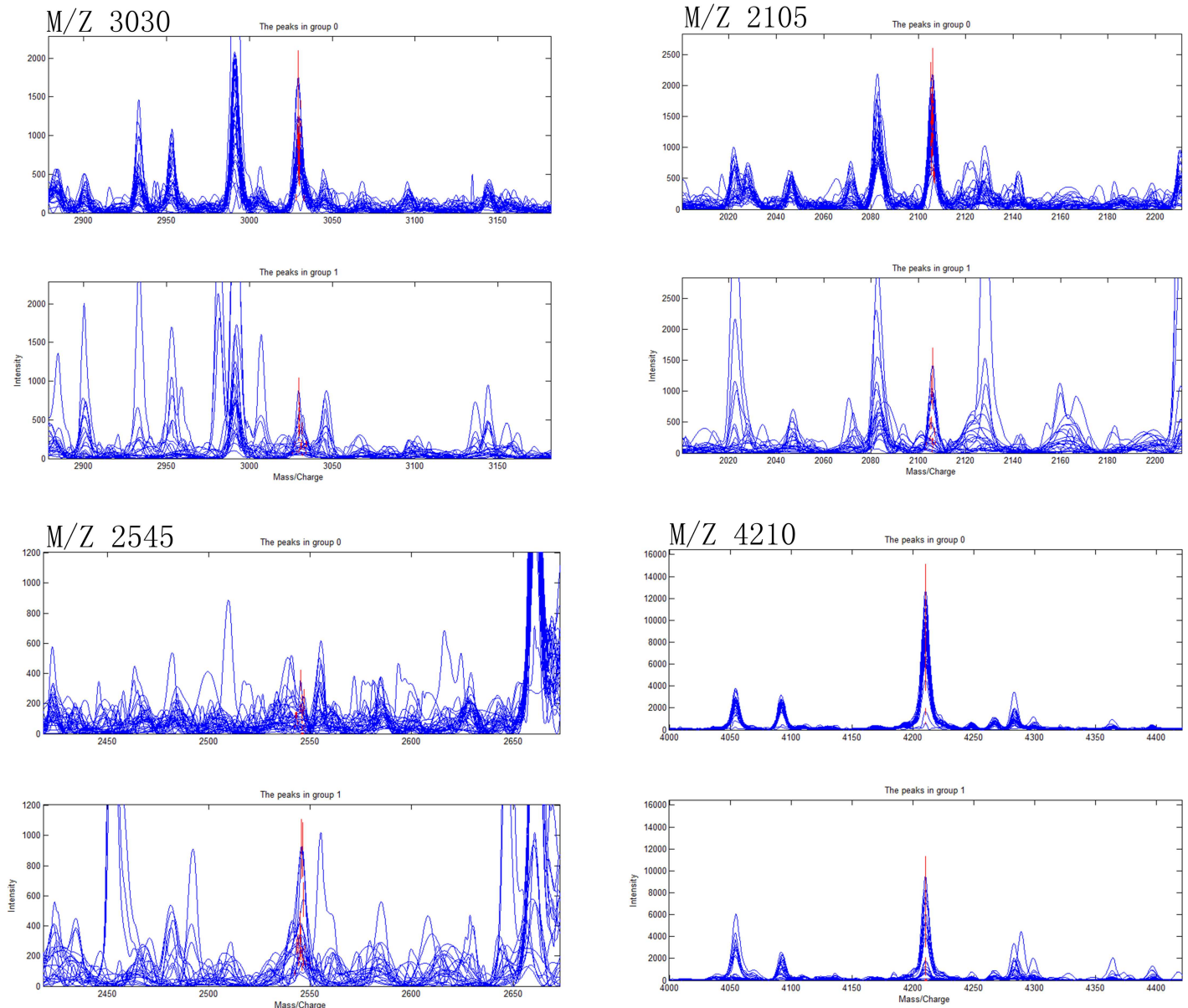
Fig 3. Four selected peaks (M/Z 3030, 2105, 2545, 4210) that compose of the SVM model for the diagnose of Intestinal tuberculosis. Group 1 representatives Intestinal tuberculosis patients, group 0 representatives Healthy controls.

doi:10.1371/journal.pone.0167109.g003

years, with a theory that the laser radiates the crystals comprised of sample and matrix, leading to sample ionization, ionized sample fly accelerated through pipes with the effect of electric field, by detecting the flight time to the detector, M/Z is measured. This method, with advantages of low sample consumption, easy operation, high resolution, good repeatability and high sensitivity, can detect and differentiate polypeptides containing diagnostic message and obtain polypeptide biological information with high sensitivity[26–28]. When detecting low abundance proteins with small relative molecular mass, the effects are more apparent. Compared with surface-Enhanced Laser Desorption/ Ionization-Time of Fight-Mass Spectrometry (SELDI-TOF-MS) technology, the difference of MALDI-TOF-MS lies in that it replace chip with weak cationic magnetic beads which has a better binding force to separate protein[29].

M/Z 4267

M/Z 4223

M/Z 1541
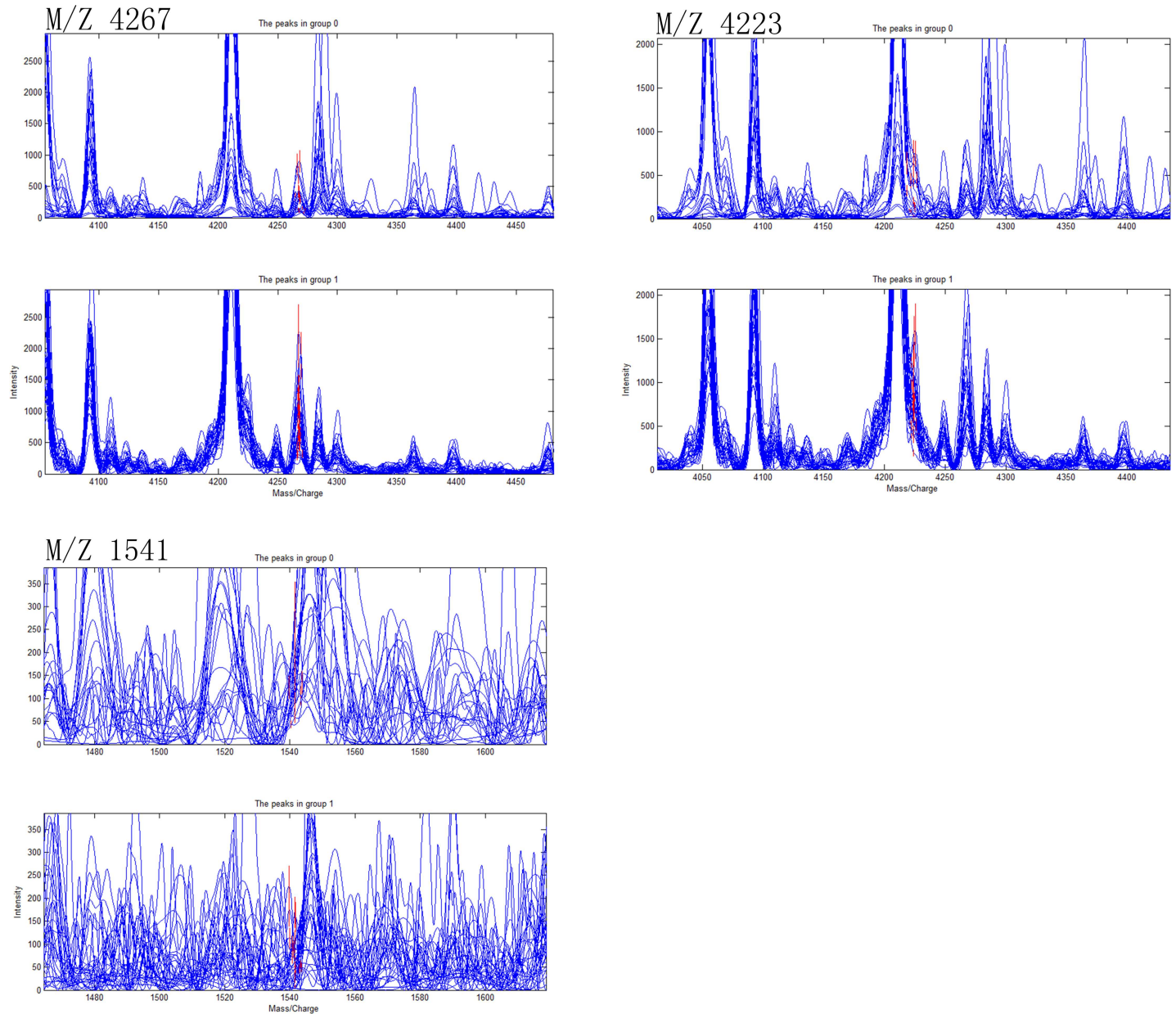
Fig 4. Three selected peaks (M/Z 4267, 4223, 1541) that compose of the SVM model for the differential diagnose between Crohn's disease and Intestinal tuberculosis. Group 1 representatives Crohn's disease patients, group 0 representatives Intestinal tuberculosis patients.

doi:10.1371/journal.pone.0167109.g004

Borrowed from the calculation model of the natural selection of Darwin's biological theory and the biological evolution process of genetic mechanism, genetic algorithm simulates the natural evolution process and searches the optimal solution method through mechanism of natural selection, heredity and mutation, which is very effective for combinatorial optimization problems[30,31]. SVM method, based on limited sample information, aims to seek the best compromise between the complexity of the model (accuracy of the specific training samples) and the learning ability (the ability to identify any sample without error) to get the best generalization ability.
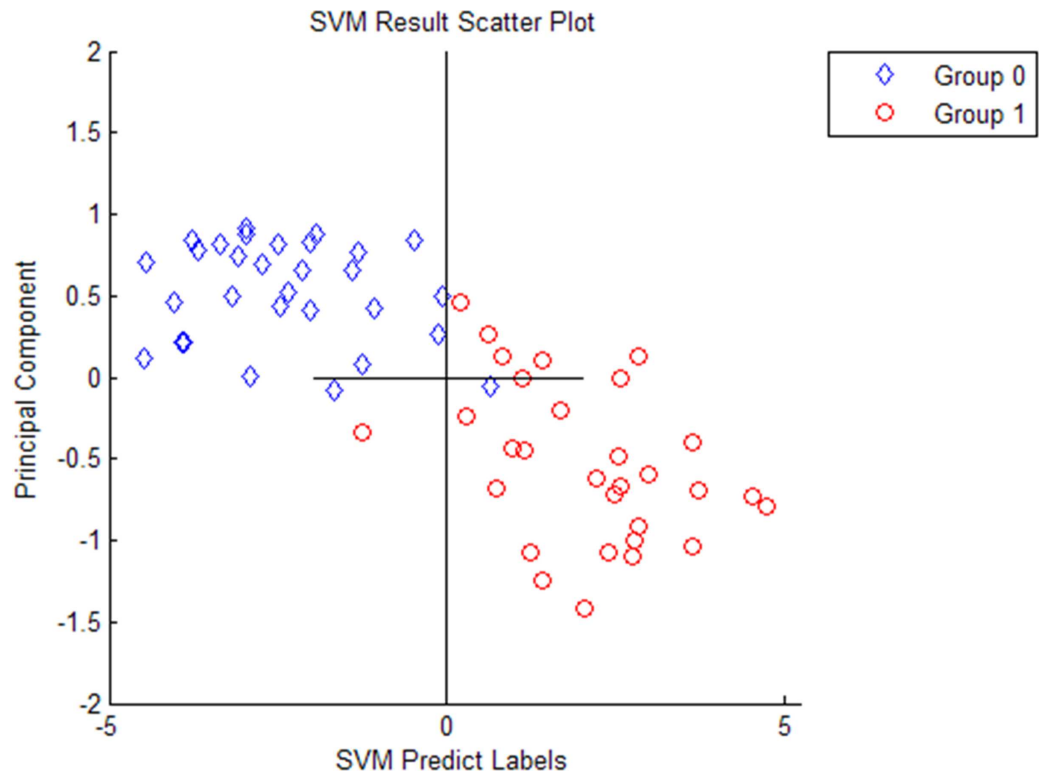
Fig 5. The discriminant results by SVM of Crohn's disease patients vs Healthy controls. Group 1 represents Crohn's disease patients, group 0 represents Healthy controls. Every hollow point indicates one sample.

This study adopted the strategy of combination of filtering method which uses Wilcoxon rank sum test and model dependent method, filtered out protein peaks without significant difference and made model dependent screening for protein peaks with significant difference. The benefits of this strategy are that it overcomes the defect of a relatively low model accuracy which established by filtering method and overcome shortcomings like large amount of calculation and time-consuming which exists in model dependent method. Another problem is how to verify the accuracy of the models involving sample classification, that is, how many samples are divided into training set to build the model, how many samples are divided into test set to verify the model and evaluate the accuracy of the model. There are three methods to validate model accuracy: dichotomy, 10 times crossover method and LOO method. Here, we used LOO method.

The differential diagnosis between CD and ITB is difficult. Firstly, they have a lot in common in endoscopic and pathological manifestations, with granulomatous inflammation as the common pathological features. Features like submucosal inflammation, aphthous ulcer, annular ulcer, expansion of ileocecal junction, granuloma and granulation tissue hyperplasia are hard to differentiate these two diseases. Therefore, the pathological diagnoses for both of them are huge challenges even for experienced pathologists. Secondly, diagnostic criterial for ITB contain acid fast stain and caseous necrosis are often difficult to be found, with a positive rate of only 18.33% and the positive rate of cultivation for mycobacterium tuberculosis is less than 20%[32–34]. However, the treatment for CD and ITB are widely divergent, especially glucocorticoid which is an important treatment for CD may cause focal diffusion, if was used mistakenly for ITB. The prognosis for this two diseases are totally different, CD, instead of being cured, is
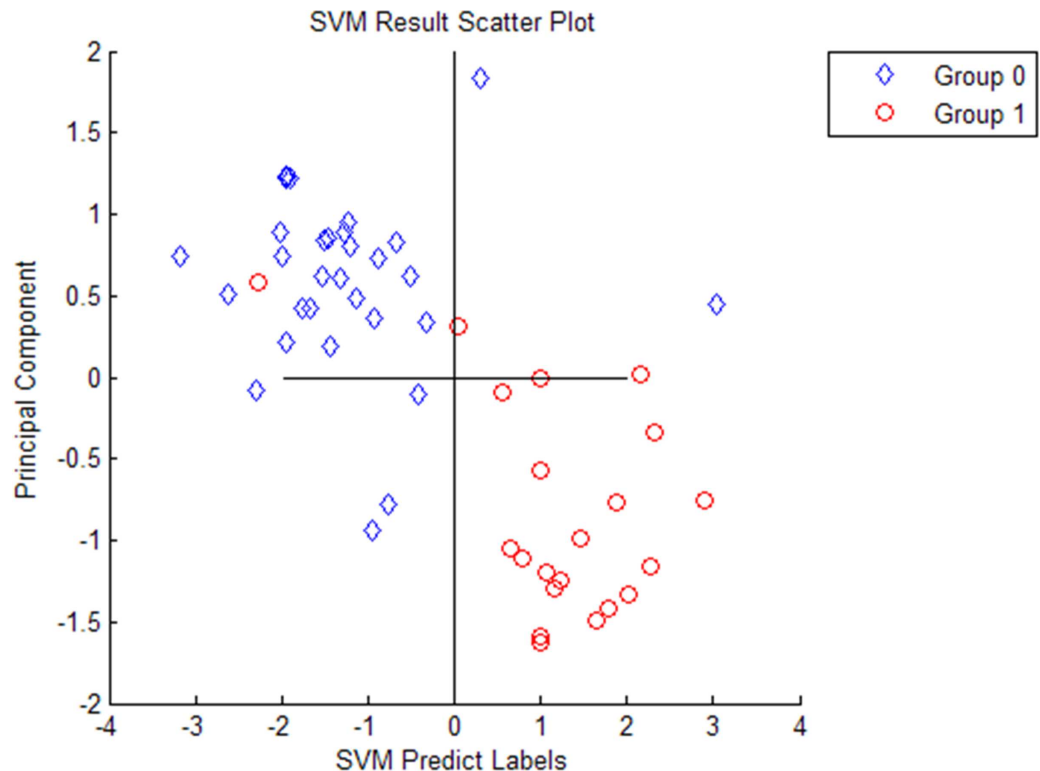
Fig 6. The discriminant results by SVM of Intestinal tuberculosis patients vs Healthy controls. Group 1 represents Intestinal tuberculosis patients, group 0 represents Healthy controls. Every hollow point indicates one sample.

doi:10.1371/journal.pone.0167109.g006

usually controlled in a stable state and easy to recur. But for ITB, once with early confirmed diagnosis and timely treatment, it can be completely cured. Although there are some studies at home and abroad have found that regression mathematical model of clinical indicators[35], anus rectum disease[36], tuberculosis, anal fistula, parenteral performance[37] can help differentiating CD and ITB, however, no specific serological biomarkers can differentiate these two diseases yet. Lokesh A, et al performed proteomics analysis of mucosal biopsies from patients with five CD and five ITB patients, they chose six differentially expressed proteins for validation using immunohistochemistry in a larger cohort of patients, but none of them was differentially expressed in patients with ITB and CD, and they did not establish a diagnostic models for CD, ITB or the differential diagnostic model between CD and ITB.

The present study is the first ever proteomics analysis of serum biopsies from patients with CD and ITB and it has revealed several promising biomarker candidates for differentiation between these two diseases. We firstly analyzed the serum protein profiling of 30 CD patients, 21 ITB patients and 30 HCs by using MALDI-TOF-MS technology, and found that there were 236 differently expressed protein peaks between CD patients and HCs, 305 differently expressed protein peaks between ITB patients and HCs, 332 differently expressed protein peaks between CD patients and ITB patients, indicating that the serum protein profiling between CD and ITB, CD and HCs, ITB and HCs have a lot of differences. The mass spectra analysis revealed a large number of potentially interesting protein peaks with significant P value. We chose ten most differentially expressed protein peaks of each group, and used genetic algorithm combining with SVM to establish a diagnostic model of CD (combination of four protein peaks of M/Z 4964, 3029, 2833, 2900, with a specificity and sensitivity of 96.7%
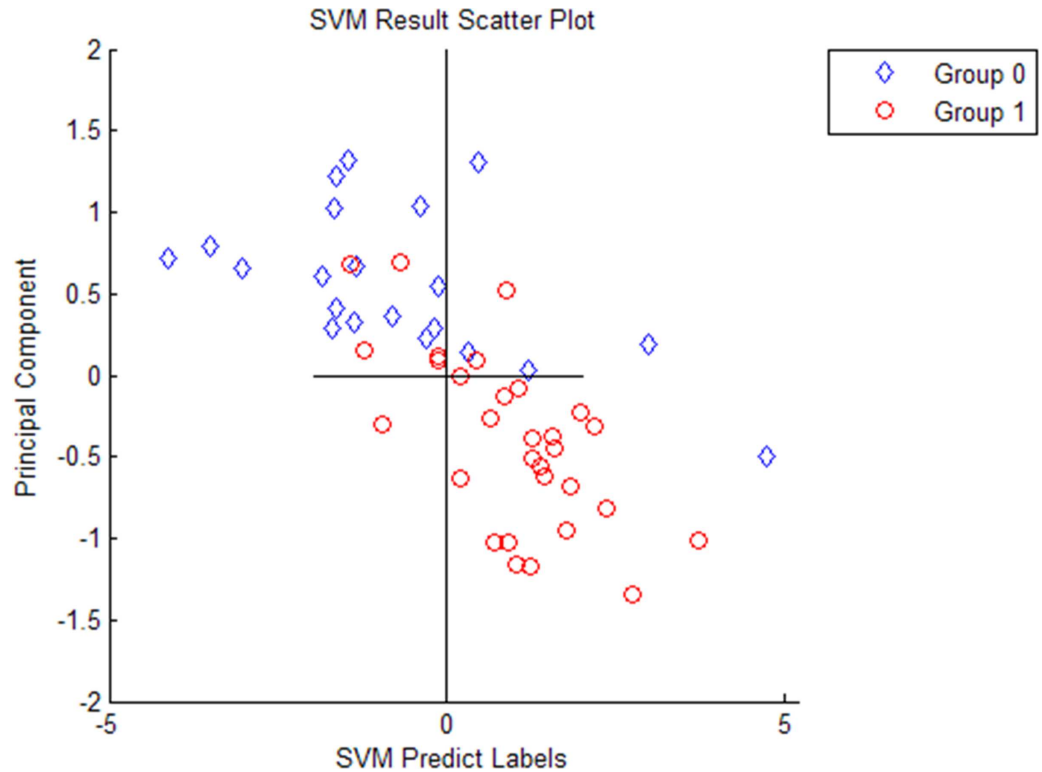
Fig 7. The discriminant results by SVM of Crohn's disease patients vs Intestinal tuberculosis patients.
Group 1 represents Crohn's disease patients, group 0 represents Intestinal tuberculosis patients. Every hollow
point indicates one sample.

doi:10.1371/journal.pone.0167109.g007

and 96.7% respectively), a diagnostic model of ITB (combination of four protein peaks of M/Z 3030, 2105, 2545, 4210, with a specificity and sensitivity of 93.3% and 95.2% respectively), a differential diagnostic model distinguishing CD from ITB (combination of three protein peaks of M/Z 4267, 4223, 1541, with a specificity and sensitivity of 76.2% and 80.0% respectively), demonstrating that the diagnostic models consisting of several markers had great value in the diagnosis of CD, ITB and differential diagnosis between them. As these two entities are complex and heterogeneous, combination of several biomarkers may be more relevant for their diagnosis and differential diagnosis between them.

We achieved purification and identification of two relevant biomarkers ($P < 0.05$): Appetite peptide and LOXL-2. They are all known proteins and were reported in many studies. Dixit V. D, et al found that appetite peptide could control immune cell activation and inflammation by inhibiting the expression of proinflammatory anorectic cytokines such as IL-1β, IL-6, and TNF-α[38]. Studies also showed that appetite peptide could inhibit FGF-2-mediated

Table 5. Validation results of the diagnostic models

Groups	N	Sensitivity(%)	Specificity(%)	Positive predictive value(%)	Negative predictive value(%)
CD/HCs	30/30	96.7	96.7	96.7	96.7
ITB/HCs	21/30	95.2	93.3	90.9	96.6
CD/ITB	30/21	80.0	76.2	82.3	72.7

CD: Crohn's disease. ITB: Intestinal tuberculosis. HCs: Healthy Controls.

doi:10.1371/journal.pone.0167109.t005

Table 6. Mascot Search Results of the identified protein peaks

SwissProt Accession No	P	Protein Name	Peptide
GHRL_HUMAN	$3.7*10^{-2}$	Appetite peptide	N.SPSSTGSGNTEHSCSS.Q
LOXL2_HUMAN	$3.6*10^{-2}$	LOXL-2	T.LAACTSNGWGVTDCK.H

Lysyl oxidase-like 2: LOXL-2

doi:10.1371/journal.pone.0167109.t006

angiogenesis in vitro and in vivo and was involved in protecting cardiomyocytes from apoptosis[39]. Mazzocchi G elaborated that appetite peptide enhanced the growth of cultured human adrenal zona glomerulosa cells by exerting MAPK-mediated proliferogenic and antiapoptotic effects[40]. A recent research showed that enhanced enteroendocrine cells responses may directly and adversely affect appetite in CD patients through increased gut-brain signaling by reducing appetite peptide[41], but another study discovered that CD itself has no significant influence on circulating appetite peptide levels in the outpatients[42]. However, we found that appetite peptide distribution was higher in CD patients when compared with healthy controls, but due to the limited number of samples, we did not validated appetite peptide in a larger sample. Next, this step should be performed and the definite mechanism behind need to be explored. LOXL-2 was suggested to be specifically involved in cell adhesion and senescence in earlier study and recently, it was proved as a role in regulating angiogenesis through collagen IV scaffolding[43] and to participated in the mechanisms in many tumors and carcinoma, like the effect on proliferation of hepatocellular carcinoma, mediating transition and proliferative growth of dormant tumor cells, and drive lung cancer invasion and metastasis[44–46]. However, there was no report of the role of LOXL2 in the pathogenesis of CD or ITB, and we found that LOXL2 distribution was lower in CD patients when compared with ITB, but due to the limited number of samples, we did not validated LOXL2 in a larger sample and the mechanisms warrant further investigation in the future.

This study is an exploratory research to achieve preliminary proteomic experiments on a panel of CD and ITB cases and HCs in order to obtain general diagnostic tools and differential diagnostic tool as well as general information concerning pathophysiology of these diseases. It has some limitations. Firstly, the sample size was relatively small and we selected both clinically active and inactive CD patients, ignoring the fact that important discriminating biomarkers, among which acute phase reactant proteins, might appear significant only in clinically active diseases. It also limits the evaluation of classification to LOO cross validation performance without the independent validation of the test set. In the next step, studies of larger numbers of patients are needed to validate these results on the independent sample cohorts, to analyze between subgroups of active and inactive CD patients and compare from two aspects: serum specimens and tissue samples for proteomics analysis to improve the credibility and specificity of the results. Secondly, validation of those diagnostic models and differential diagnostic model is necessary in larger sample size to confirm the value of these models in the diagnosis of CD, ITB and the differential diagnosis between CD and ITB. Thirdly, purification and identification of relevant biomarkers to make clear the corresponding peptides and proteins is indispensable to understand the etiology and pathogenesis of those diseases. Besides, although protein fingerprint has the advantages such as fast and high throughput, easy to standardization, it do has uncertainty of sample preparation method, protein separation, data acquisition and analysis method, which influences the results of mass spectrum fingerprint classification[47]. So efforts from clinical experts, personnel of mass spectrum technology research and development projects and bioinformatics experts are necessary to make MALDI-TOF-MS technique get widely used and long-term development.

To sum up, this study, using MALDI-TOF-MF technology, screened out differentially expressed protein peaks from serum samples between CD patients and HCs, ITB patients and HCs, CD patients and ITB patients. The diagnostic model of CD, IBD and the differential diagnostic model between CD and ITB established by genetic algorithm and SVM model has high specificity and sensitivity. Researching those protein peaks in-depth will contribute to further understanding of the etiology and pathogenesis of CD, will provide new methods for the diagnoses of CD and the differential diagnosis between CD and ITB.

Conclusion

The differently expressed protein peaks analyzed by serum proteome with weak cationic magnetic beads combined MALDI-TOF-MS technique can effectively distinguish CD patients, ITB patients and healthy controls. By genetic algorithm combining with SVM, the diagnostic model between CD and HCs consisting of four protein peaks (M/Z 4964, 3029, 2833, 2900), the diagnostic model between ITB and HCs comprising four protein peaks (M/Z 3030, 2105, 2545, 4210) and the differential diagnostic model between CD and ITB comprising three potential biomarkers (M/Z 4267, 4223, 1541) had high specificity and sensitivity to the diagnoses of CD, ITB and the differential diagnosis between CD and ITB.

Supporting Information

S1 File. Serum proteomic profiling between CD and HCs. The biomarkers principal component scatter plot, the gel figure and peak figure of the ten most differentially expressed peaks and the results by using genetic algorithm combining with SVM.
(PDF)

S2 File. Serum proteomic profiling between ITB and HCs. The biomarkers principal component scatter plot, the gel figure and peak figure of the ten most differentially expressed peaks and the results by using genetic algorithm combining with SVM.
(PDF)

S3 File. Serum proteomic profiling between CD and ITB. The biomarkers principal component scatter plot, the gel figure and peak figure of the ten most differentially expressed peaks and the results by using genetic algorithm combining with SVM.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: GQX.

Data curation: JKY HTC.

Formal analysis: GDS HTC.

Funding acquisition: GQX.

Investigation: FLH MY.

Methodology: CFX JKY.

Project administration: GQX CFX.

Resources: FMZ LGN.

Software: JKY LGN.

Supervision: CFX GDS.

Validation: JKY WGC.

Visualization: MY.

Writing – original draft: FMZ.

Writing – review & editing: FLH WGC.

References

1. Epstein D, Watermeyer G, Kirsch R. Review article: the diagnosis and management of Crohn's disease in populations with high-risk rates for tuberculosis. Aliment Pharmacol Ther 2007; 25: 1373–1388. doi: 10.1111/j.1365-2036.2007.03332.x PMID: 17539977

2. Almadi MA, Ghosh S, Aljebreen AM. Differentiating intestinal tuberculosis from Crohn's disease: a diagnostic challenge. Am J Gastroenterol 2009; 104: 1003–1012. doi: 10.1038/ajg.2008.162 PMID: 19240705

3. Makharia GK, Srivastava S, Das P, Goswami P, Singh U, Tripathi M, et al. Clinical, endoscopic, and histological differentiations between Crohn's disease and intestinal tuberculosis. Am. J. Gastroenterol 2010; 105: 642–51. doi: 10.1038/ajg.2009.585 PMID: 20087333

4. Wong WM, Lai KC, Yiu WC, Wong BC, Chan FL, Lai CL. Intestinal tuberculosis mimicking fistulizing Crohn's disease. J Gastroenterol Hepatol 2007; 22: 137–139. doi: 10.1111/j.1440-1746.2006.03311.x PMID: 17201898

5. Leighton JA, Shen B, Baron TH, Adler DG, Davila R, Egan JV, et al. ASGE guideline: endoscopy in the diagnosis and treatment of inflammatory disease. Gastrointest Endosc 2006; 63: 558–565. doi: 10.1016/j.gie.2006.02.005 PMID: 16564852

6. Epstein D, Watermeyer G, Kirsch R. Review article: the diagnosis and management of Crohn's disease in populations with high-risk rates for tuberculosis. Aliment Pharmacol Ther 2007; 25: 1373–1388. doi: 10.1111/j.1365-2036.2007.03332.x PMID: 17539977

7. Venkatesh PG, Navaneethan U. Mimickers of intestinal tuberculosis: could this be Crohn's disease? An unsolved enigma. Saudi J Gastroenterol 2011; 17: 95–96. doi: 10.4103/1319-3767.77236 PMID: 21372344

8. Sailer J, Peloschek P, Schober E, Schima W, Reinisch W, Vogelsang H, et al. Diagnostic value of CT enteroclysis compared with conventional enteroclysis in patients with Crohn's disease. AJR Am J Roentgenol 2005; 185: 1575–1581. doi: 10.2214/AJR.04.1534 PMID: 16304016

9. Lee YJ, Yang SK, Myung SJ, Byeon JS, Park IG, Kim JS, et a1. The usefulness of colonoscopic biopsy in the diagnosis of intestinal tuberculosis and pattern of concomitant extra-intestinal tuberculosis. Korean J Gastroenterol 2004; 44: 153–159. PMID: 15385724

10. Gilinsky NH, Marks IN, Kottler RE, Price SK. Abdominal tuberculosis. A 10-year review. S Afr Med J 1983; 64: 849–57. PMID: 6635879

11. Chatzicostas C, Koutroubakis IE, Tzardi M, Roussomoustakaki M, Prassopoulos P, Kouroumalis EA. Colonic tuberculosis mimicking Crohn's disease: case report. BMC Gastroenterol 2002; 2:10. doi: 10.1186/1471-230X-2-10 PMID: 12019037

12. Petroianni A, Mugnaini L, Laurendi G, Giousuè S, Schininà V, Bibbolino C, Bisetti A, et al. Abdominal tuberculosis mimicking Crohn's disease: a difficult diagnosis. Report of a case. Panminerva Med 2002; 44: 155–158. PMID: 12032437

13. Felley-Bosco E, Andre M. Proteomics and chronic inflammatory bowel diseases. Pathol Res Pract 2004; 200: 129–33. doi: 10.1016/j.prp.2004.02.002 PMID: 15237921

14. Xiao Z, Prieto D, Conrads TP, Veenstra TD, Issaq HJ. Proteomic patterns: their potential for disease diagnosis. Mol Cell Endocrinol 2005; 230: 95–106. doi: 10.1016/j.mce.2004.10.010 PMID: 15664456

15. Cravatt BF, Simon GM, Yates JRr. The biological impact of mass-spectrometry-based proteomics. Nature 2007; 450: 991–1000. doi: 10.1038/nature06525 PMID: 18075578

16. Hanash SM, Pitteri SJ, Faca VM. Mining the plasma proteome for cancer biomarkers. Nature 2008; 452: 571–579. doi: 10.1038/nature06916 PMID: 18385731

17. Villanueva J, Philip J, Entenberg D, Chaparro CA, Tanwar MK, Holland EC, et al. Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. Anal Chem 2004; 76: 1560–70. doi: 10.1021/ac0352171 PMID: 15018552

18. Whiteaker JR, Zhao L, Zhang HY, Feng LC, Piening BD, Anderson L, et al. Antibody-based enrichment of peptides on magnetic beads for mass-spectrometry-based quantification of serum biomarkers. Anal Biochem 2007; 362: 44–54. doi: 10.1016/j.ab.2006.12.023 PMID: 17241609

19. Van Assche G, Dignass A, Panes J, Beaugerie L, Karagiannis J, Allez M, et al. European Crohn's and Colitis Organisation (ECCO). The second European evidence-based Consensus on the diagnosis and management of Crohn's disease: Definitions and diagnosis. J Crohns Colitis 2010; 4: 7–27. doi: 10.1016/j.crohns.2009.12.003 PMID: 21122488

20. Hoss-Schewietzek S, Alles JU, Muhrer KH. Tuberculosis of the large intestine-a differential diagnosis of inflammatory and neoplastic colon tumors Chirurg 1995; 66(7): 731–4.

21. Logan V. S. Anorectal tuberculosis. Proc R Soc Med 1969; 62: 1227–30. PMID: 5363105

22. Taleahmad S, Mirzaei M, Parker LM, Hassani SN, Mollamohammadi S, Sharifi-Zarchi A, et al. Proteome Analysis of Ground State Pluripotency. Sci Rep 2015; 16: 17985.

23. Okano T, Seike M, Kuribayashi H, Soeno C, Ishii T, Kida K, et al. Identification of haptoglobin peptide as a novel serum biomarker for lung squamous cell carcinoma by serum proteome and peptidome profiling. Int J Oncol 2016; 48: 945–52. doi: 10.3892/ijo.2016.3330 PMID: 26783151

24. Wasinger VC, Yau Y, Duo X, Zeng M, Campbell B, Shin S, et al. Low Mass Blood Peptides Discriminative of Inflammatory Bowel Disease (IBD) Severity: A Quantitative Proteomic Perspective. Mol Cell Proteomics 2016; 15: 256–65. doi: 10.1074/mcp.M115.055095 PMID: 26530476

25. Lu K, Knutson CG, Wishnok JS, Fox JG, Tannenbaum SR. Serum metabolomics in a Helicobacter hepaticus mouse model of inflammatory bowel disease reveal important changes in the microbiome, serum peptides, and intermediary metabolism. J Proteome Res 2012; 11: 4916–26. doi: 10.1021/pr300429x PMID: 22957933

26. Wong MY, Yu KO, Poon TC, Ang IL, Law MK, Chan KY, et al. A magnetic bead-based serum proteomic fingerprinting method for parallel analytical analysis and micropreparative purification. Electrophoresis 2010; 31: 1721–1730. doi: 10.1002/elps.200900571 PMID: 20414880

27. Giebel R, Worden C, Rust SM, Kleinheinz GT, Robbins M, Sandrin TR. Microbial fingerprinting using marx-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) applications and challenges. Adv Appl Microbiol 2010; 71: 149–184. doi: 10.1016/S0065-2164(10)71006-6 PMID: 20378054

28. Sparbier K, Wenzel T, Dihazi H, Blaschke S, Müller GA, Deelder A, et al. Immuno-MALDI-TOF-MS: new prospectives for clinical applications of mass spectrometry. Proteomics 2009; 9: 1442–1450. doi: 10.1002/pmic.200800616 PMID: 19235164

29. Fung ET, Wright GL Jr, Dalmasso EA. Proteomic strategies for biomarker identification: progress and challenges. Curt Opin Mol Ther 2000; 2(6): 643–650.

30. Song J, Xue G, Kang Y. A Novel Method for Optimum Global Positioning System Satellite Selection Based on a Modified Genetic Algorithm. PLoS One 2016; 11(3): e0150005. doi: 10.1371/journal.pone.0150005 PMID: 26943638

31. Albert J. A Hybrid of the Chemical Master Equation and the Gillespie Algorithm for Efficient Stochastic Simulations of Sub-Networks. PLoS One 2016; 11(3):e0149909. doi: 10.1371/journal.pone.0149909 PMID: 26930199

32. Amarapurkar DN, Patel ND, Amarapurkar AD, Agal S, Baigal R, Gupte P. Tissue polymerase chain reaction in diagnosis of intestinal tuberculosis and Crohn's disease. J Assoc Physicians India 2004; 52: 863–7. PMID: 15906835

33. Gan HT, Chen YQ, Ouyang Q, Bu H, Yang XY. Differentiation between intestinal tuberculosis and Crohn's disease in endoscopicbiopsy specimens by polymerase chain reaction. Am J Gastroenterol 2002; 97: 1446–1451. doi: 10.1111/j.1572-0241.2002.05686.x PMID: 12094863

34. Alvares JF, Devarbhavi H, Makhija P, Rao S, Kottoor R. Clinical, colonoscopic, and histological profile of colonic tuberculosis in a tertiary hospital. Endoscopy 2005; 37: 351–6. doi: 10.1055/s-2005-861116 PMID: 15824946

35. Liu XW, Li XF, Zhou YY. Screening differential indicators for Crohn's disease and tuberculosis using logistic regression analysis. World Chinese Journal of Digestology 2010; 18: 621–627.

36. Mao R, Liao WD, He Y, Ouyang CH, Zhu ZH, Yu C, et al. Analyses of colonoscopic findings in the differential diagnosis between intestinal tuberculosis and Crohn's disease. Endoscopy 2006; 38: 592–597. doi: 10.1055/s-2006-924996 PMID: 16673312

37. Kirsch R, Pentecost M, Hall Pde M, Epstein DP, Watermeyer G, Friederich PW, et al. Role of colonoscopic biopsy in distinguishing between Crohn's disease and intestinal tuberculosis. J Clin Pathol 2006; 59: 840–844. doi: 10.1136/jcp.2005.032383 PMID: 16873564

38. Schulthess B, Brodner K, Bloemberg GV, Zbinden R, Böttger EC, Hombach M. Identification of Gram-positive cocci using MALDI-TOF-MS: comparison of different preparation methods and implementation of a practical algorithm for routine diagnostics. J Clin Microbiol 2013; 51(6): 1834–1840. doi: 10.1128/JCM.02654-12 PMID: 23554198

39. Conconi M.T, Nico B, Guidolin D, Baiguera S, Spinazzi R, Rebuffat P, et al. Ghrelin inhibits FGF-2-mediated angiogenesis in vitro and in vivo. Peptides 2004; 25: 2179–2185. doi: 10.1016/j.peptides.2004.08.011 PMID: 15572208

40. Mazzocchi G, Neri G, Rucinski M, Rebuffat P, Spinazzi R, Malendowicz L.K, et al. Ghrelin enhances the growth of cultured human adrenal zona glomerulosa cells by exerting MAPK-mediated proliferogenic and antiapoptotic effects Peptides 2004; 25: 1269–1277.

41. Nishi Y, Isomoto H, Ueno H, Ohnita K, Wen CY, Takeshima F, et al. Plasma leptin and ghrelin concentrations in patients with Crohn's disease. World J Gastroenterol. 2005; 11(46): 7314–7. doi: 10.3748/wjg.v11.i46.7314 PMID: 16437634

42. Dixit V.D, Schaffer E.M, Pyle R.S, Collins G.D, Sakthivel S.K, Palaniappan R, et al. Ghrelin inhibits leptin- and activation-induced proinflammatory cytokine expression by human monocytes and T cells. J Clin Invest 2004; 114: 57–66. doi: 10.1172/JCI21134 PMID: 15232612

43. Bignon M, Pichol-Thievend C, Hardouin J, Malbouyres M, Brechot N, Nasciutti L, et al. Lysyl oxidase-like protein-2 regulates sprouting angiogenesis and type IV collagen assembly in the endothelial basement membrane. Blood 2011; 118: 3979–3989. doi: 10.1182/blood-2010-10-313296 PMID: 21835952

44. Weidenfeld K, Schif-Zuck S, Abu-Tayeh H, Kang K, Kessler O, Weissmann M, et al. Dormant tumor cells expressing LOXL2 acquire a stem-like phenotype mediating their transition to proliferative growth. Oncotarget 2016.

45. Wu L, Zhang Y, Zhu Y, Cong Q, Xiang Y, Fu L, et al. The effect of LOXL2 in hepatocellular carcinoma. Mol Med Rep 2016; 14(3): 1923–32. doi: 10.3892/mmr.2016.5474 PMID: 27430160

46. Peng DH, Ungewiss C, Tong P, Byers LA, Wang J, Canales JR4, et al. ZEB1 induces LOXL2-mediated collagen stabilization and deposition in the extracellular matrix to drive lung cancer invasion and metastasis. Oncogene 2016 [Epub ahead of print].

47. Schulthess B, Brodner K, Bloemberg GV, Zbinden R, Böttger EC, Hombach M. Identification of Gram-positive cocci using MALDI-TOF-MS: comparison of different preparation methods and implementation of a practical algorithm for routine diagnostics. J Clin Microbiol 2013; 51(6): 1834–1840. doi: 10.1128/JCM.02654-12 PMID: 23554198