A Metabolomic Study of Cervical Dystonia

Chang Liu, MPH[a], **Laura Scorr, MD**[b], **Gamze Kilic-Berkmen, PhD**[b], **Adam Cotton, BA**[b], **Stewart A. Factor, DO**[b], **Alan Freeman, MD**[b], **ViLinh Tran, MS**[c], **Ken Liu, PhD**[c], **Karan Uppal, PhD**[d], **Dean Jones, PhD**[d], **H. A. Jinnah, MD PhD**[b,e], **Yan V. Sun, PhD**[a,f]

[a]Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, Georgia, USA

[b]Jean and Paul Amos Parkinson's Disease and Movement Disorder Clinic Department of Neurology, Emory University School of Medicine, Atlanta, Georgia, USA

[c]Clinical Biomarkers Laboratory, Department of Pulmonary, Allergy, Critical Care and Sleep Medicine, Emory University School of Medicine, Atlanta, Georgia, USA

[d]Department of Pulmonary, Allergy, Critical Care and Sleep Medicine, Emory University School of Medicine, Atlanta, Georgia, USA

[e]Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, USA

[f]Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, Georgia, USA

Abstract

Introduction: Cervical dystonia is the most common of the adult-onset focal dystonias. Most cases are idiopathic. The current view is that cervical dystonia may be caused by some combination of genetic and environmental factors. Genetic contributions have been studied extensively, but there are few studies of other factors. We conducted an exploratory metabolomics analysis of cervical dystonia to identify potentially abnormal metabolites or altered biological pathways.

Methods: Plasma samples from 100 cases with idiopathic cervical dystonia and 100 controls were compared using liquid chromatography coupled with mass spectrometry-based metabolomics.

Results: A total of 7,346 metabolic features remained after quality control, and up to 289 demonstrated significant differences between cases and controls, depending on statistical criteria chosen. Pathway analysis revealed 9 biological processes to be significantly associated at $p<0.05$, 5 pathways were related to carbohydrate metabolism, 3 pathways were related to lipid metabolism.

Conclusion: This is the first large scale metabolomics study for any type of dystonia. The results may provide potential novel insights into the biology of cervical dystonia.

Introduction

Cervical dystonia (CD) is characterized by excessive involuntary contractions of muscles of the neck and shoulder leading to abnormal head postures, involuntary movements including tremor, and pain [1]. It is considered to be a rare disorder with an overall incidence of 1.2 per 100,000 person years [2], and crude prevalence of 29.5 per 100,000 population [3]. It typically begins in adults between 40–60 years of age, with females being affected more often than males [4]. For the majority of cases, a cause cannot be found, even after extensive diagnostic testing. Much of the recent research regarding potential causes has focused on genetics, as ~15% of patients with CD have a family member affected with dystonia [4]. However, the genes currently known to cause dystonia account for no more than 4% of all cases [1]. These observations suggest that other, as yet undiscovered, genes may play a role.

Non-genetic factors may also have a place in pathogenesis. Rare cases of CD may be caused by focal lesions of the nervous system [5], exposure to medications such as neuroleptics [6], infection [7], or autoimmune mechanisms [8]. A few small studies have methodically addressed other potential environmental contributions. A survey study of 184 cases of all types of idiopathic dystonia in Australia suggested cigarette smoking to be a risk factor [9]. A case control study of 67 cases of CD suggested a role for trauma to the head and neck [10].

The metabolome is the global collection of all small molecules in the body. It reflects the combined systemic effects of genetics, lifestyle, environmental exposure and secondary biological responses. Metabolomics, or the study of global metabolism, is an emerging discipline that has the potential to transform the study of the mechanism of disease development. High-resolution chemical profiling involves the use of liquid chromatography to separate analytes and high-resolution mass spectrometry to accurately measure mass and abundance [11]. To our knowledge, there are no published metabolomic studies for any of the common adult-onset focal dystonias. The current study describes results of a blood-based metabolomics study of 100 subjects with CD and 100 healthy controls.

Methods

Study Population

The study was approved by the Emory Institutional Review Board, and all cases and controls provided written informed consent to participate. Subjects with a diagnosis of idiopathic CD were recruited from the Emory University Movement disorders clinic. Dystonia in other body regions was allowed (segmental and multifocal patterns), but involvement of neck muscles had to be the primary problem. Subjects were excluded if they had CD secondary to a known cause such as focal lesions, Parkinsonism or neuroleptic exposure. Control subjects were recruited from the same clinic. The majority of the controls were spouses or family members of the subjects with CD. Precise matching of cases and controls was not conducted.

We also collected information known to influence the metabolome. Subjects were asked if they use nicotine products currently or in the past or if they had a history of metabolic diseases (e.g., hypertension, hyperlipidemia, heart disease, or diabetes). Additionally, data about subjects' body mass index (BMI) and diet were also collected.

Untargeted High-Resolution Metabolic Profiling

We conducted high-resolution metabolic profiling using liquid chromatography with ultra-high resolution mass spectrometry (LC-HRMS; Thermo Scientific Fusion) following established protocols and published methods [12]. Plasma specimens were collected and stored at −80°C before thawing for liquid-chromatography mass spectrometry (LC/MS). Samples were treated with acetonitrile containing internal isotopic standard mix, centrifuged for 10 minutes at 4°C to remove protein, and then maintained in a Dionex Ultimate 3000 autosampler at 4°C until analysis. NIST 1950 [13] was analyzed at the beginning and end of the entire analysis procedure. Two replicated pooled human plasma samples were analyzed at the beginning, middle, and end of each batch of 40 study samples for batch effect correction and normalization. Samples were analyzed in triplicate using C18 chromatography (Higgins endcapped C18 2.1 cm $\times$ 5 cm $\times$ 3 μm column) the following gradient program: initial 0.5 min period of 60% A (LCMS grade water), 35% B (LCMS grade acetonitrile, 5% C (10 mM ammonium acetate), followed by linear increase to 0% A, 95% B, 5% C at 1.5 min which was held for an additional 3 min. Finally, mobile phase flow rate was held at 0.4 mL/min for 1.5 min, and then increased to 0.5 mL/min. MS was operated in negative ionization mode using electrospray ionization, mass resolution of 120,000 and data collected from 85 to 1275 m/z over 5 minutes) [12]. Source tune settings included capillary temperature, sheath gas, auxiliary gas, sweep gas and spray voltage settings of 300 °C, 45 (arbitrary units), 25 (arbitrary units), 1 (arbitrary units) and + 3.5 kV for negative mode. S-Lens RF level was maintained at 45. Feature extraction and alignment was performed in apLCMS [14], with quality control and feature filtering using xMSanalyzer [15]. Features are defined by their accurate mass *m/z*, retention time and ion abundance. Features with median coefficient of variation within technical replicates 75%, and samples with Pearson correlation coefficient 0.7 within triplicates were retained for downstream analysis. Batch effect was corrected using ComBat [16].

Feature Annotation and Identification

Feature annotation was performed using xMSannotator [17] with the Human Metabolome Database [18]. Annotations with confidence score of 3 indicates high confidence level, score of 2 indicates medium confidence level. Only annotations with medium confidence or above were used. Furthermore, metabolite identification and confirmation were performed using an in-house library of metabolites that have been previously confirmed using comparison of retention time and MS/MS with authentic standards and comparison of MS/MS fragments with online spectral databases. The library was matched to the metabolic features that significantly associated with CD, allowing an m/z difference of 5 ppm and a retention time difference of 20 seconds. Feature identification levels were assigned according to the published criteria [19]. At level 1, the proposed structure was confirmed via appropriate measurement of a reference standard with MS, MS/MS and retention time matching. At level 2, the proposed structure was confirmed by MS/MS and matched with online databases or in-silico predicted spectra. At level 3, the proposed structure was confirmed by MS/MS at the chemical class level, but no evidence for a specific metabolite. At level 4, the proposed structure was computationally annotated using xMSannotator with confidence level medium or above [17]. At level 5, the proposed structure was matched accurately by mass.

Statistical Analysis

Subject characteristics were reported as mean ± standard deviation for continuous variables and frequency (percentage) for categorical variables. The differences between cases and controls were compared using two sample t-tests for continuous variables and Chi-square or Fisher's exact tests for categorical variables where appropriate.

The metabolic feature intensities were summarized as median values of the non-zero readings across triplicates. Features were retained for analysis if there were less than 20% zero readings across CD samples or control samples. Intensities were log 2 (intensity+1) transformed, mean centered and scaled by standard deviation.

Logistic regression was utilized for the metabolome-wide association study (MWAS). The association was examined between the binary response variable of CD and each of the metabolic features, including age, gender, diet restrictions (e.g., veganism and low-carb diet), and BMI as covariates. We also additionally adjusted for ethnicity (Black, White, other) and smoking status (current smoker, past smoker, never smoker) for the selected features of interest to compare the estimates. The Benjamini-Hochberg false discovery rate (FDR) [20] method was used to correct for multiple comparisons. Analyses were performed using R version 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria) and SAS statistical software version 9.4 (SAS Institute Inc, Cary, NC).

Pathway Analysis

Features associated with CD that had p<0.05 in MWAS were chosen as target features that entered pathway analysis in Mummichog (version 2.1.1) [21], which is a software designed for high throughput untargeted metabolomics data. This analysis allows us to explore the functional activities directly from the untargeted features, bypassing metabolite

identification. All features entered MWAS were utilized as the feature reference pool. Pathways enriched with p<0.05 were considered as significant.

Results

Among the 200 subjects, 100 were CD cases and 100 were controls (Table 1). The proportion of females was significantly higher among cases than controls (78% vs 55%, p<0.001). Cases were older compared to controls at enrollment (62.3 vs 57.5, p=0.011), and had lower BMI (26.1 vs 28.4, p=0.002). All of these variables were adjusted in the statistical model as covariates.

A total of 7,346 metabolic features passed quality control and entered MWAS. With the adjustment of age, gender, diet restrictions and BMI, the MWAS revealed metabolomic associations for 289 features at p<0.05, 35 features at p<0.01, and 1 feature at p<0.001. None of these features were significant at FDR cutoff 0.2 (Figure 1 & Figure 2). Of the 289 features with p<0.05, 64 features were annotated based on Human Metabolome Database (Supplemental Table 1). Table 2 presents the 6 annotated features out of the 35 with p value < 0.01. The 289 features associated with CD at p<0.05 were further confirmed using the in-house library of confirmed metabolites. A total of 5 features were successfully matched (Table 3). Among those, higher level of uracil was associated with 36% lower odds of CD (m/z 111.0199, time 13.4s, OR 0.64, 95% CI 0.46–0.87, p=0.005). The estimates for the 5 features were similar after additionally controlling for ethnicity and smoking (Supplemental Table 2).

Pathway analysis identified 9 significant pathways at p<0.05 that are enriched based on the 289 CD-associated metabolic features, including prostaglandin formation from dihomo gama-linoleic acid (p<0.001), pentose phosphate pathway (p=0.005), arachidonic acid metabolism (p=0.013), leukotriene metabolism (p=0.015), starch and sucrose metabolism (p=0.030), valine, leucine and isoleucine degradation (p=0.031), galactose metabolism (p=0.031), fructose and mannose metabolism (p=0.047) and propanoate metabolism (p=0.047) (Table 4).

Discussion

This report describes the first large-scale metabolomics study for CD, the most common form of the idiopathic adult-onset dystonia. A total of 7,346 metabolites were measured, and significant differences could be detected between 100 CD and 100 control samples for individual metabolites. Pathway analysis revealed 9 biological processes to be associated at p<0.05. These results provide novel insights into the biological basis for CD.

An interesting metabolite that was identified as potentially abnormal in CD was the oxidized lipid 2-hydroxydecanoate. This metabolite was also identified in patients with Parkinson disease associated with mutations in *parkin* [22]. Dystonia is a well-known feature of this genetic subtype of Parkinson disease, although the dystonia is mostly in the limbs [23, 24]. Docosahexaenoic acid was also potentially abnormal in CD. This metabolite is an omega-3 fatty acid which is essential for brain development and normal function, and many studies have implied that dystonia is a disorder of neural development or maladaptive plasticity [25].

Among the biological pathways that were significantly associated with CD (Table 4), five were related to carbohydrate metabolism, suggesting a potential link with energy production. Three pathways were related to lipid metabolism, which produces prostaglandins and other metabolites known to mediate inflammatory responses [26]. Involvement of these pathways may suggest a link to an immune-related process in dystonia. Several case series have reported CD to be associated with autoimmune diseases such as systemic lupus erythematosus, myasthenia gravis and Sjogrens syndrome [27–29]. CD can also occur as a delayed consequence of various infections, due to immune system cross-reactivity between the infectious agent and host targets [30]. Thus, the immune link suggested by our metabolomics findings may be worthy of further investigation.

Although this study has revealed several intriguing novel potential biological leads for future studies, the results must be interpreted with caution. Numerous variables are known to affect the metabolome in human blood samples, and it is challenging to control for all of them. For instance, treatments for CD such as botulinum toxin injections may affect the metabolome. Prior botulinun toxin treatment is difficult to incorporate in the statistical model due to varying products, doses and intervals. To address this, we minimized impact of botulinum toxin by collecting blood samples 12 weeks following prior botulinum toxin treatments when the effects are expected to have worn off. Although this study was based on a relatively large number of cases and controls all collected in a standardized manner at a single center, a replication study is still warranted. Because of the large number of measurements made, correction for multiple comparisons is needed to increase confidence in the findings. On the other hand, because this study was exploratory, the application of overly conservative statistical criteria is not helpful either, because important novel leads may be inappropriately overlooked. We therefore report findings at different statistical thresholds. Despite the fact that our balanced composition of cases and controls achieved the optimal statistical power given a fixed sample size of 200, future studies of larger sample sizes may be needed to detect smaller effect sizes. Further, there is a possibility that some CD cases were attributable to genetic causes. Future studies would be needed to assess the contributions of specific genes. Additionally, due to the cross-sectional study design, no temporality between the different metabolomic profiles and CD can be inferred. A longitudinal cohort study with follow-up would be needed to better understand differences in onset and progression. Finally, using blood samples to infer biological mechanisms for a brain disorder carries significant limitations, given that relevant biomarkers may not cross the blood-brain barrier. Further investigations of cerebral spinal fluid or brain samples may be needed in future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Reference

[1]. LeDoux MS, Vemula SR, Xiao J, Thompson MM, Perlmutter JS, Wright LJ, Jinnah HA, Rosen AR, Hedera P, Comella CL, Weissbach A, Junker J, Jankovic J, Barbano RL, Reich SG, Rodriguez RL, Berman BD, Chouinard S, Severt L, Agarwal P, Stover NP, Clinical and genetic features of cervical dystonia in a large multicenter cohort, Neurol Genet 2(3) (2016) e69. [PubMed: 27123488]

[2]. Claypool DW, Duane DD, Ilstrup DM, Melton LJ 3rd, Epidemiology and outcome of cervical dystonia (spasmodic torticollis) in Rochester, Minnesota, Mov Disord 10(5) (1995) 608–14. [PubMed: 8552113]

[3]. Nutt JG, Muenter MD, Aronson A, Kurland LT, Melton LJ 3rd, Epidemiology of focal and generalized dystonia in Rochester, Minnesota, Mov Disord 3(3) (1988) 188–94. [PubMed: 3264051]

[4]. Jinnah HA, Berardelli A, Comella C, Defazio G, Delong MR, Factor S, Galpern WR, Hallett M, Ludlow CL, Perlmutter JS, Rosen AR, Dystonia Coalition I, The focal dystonias: current views and challenges for future research, Mov Disord 28(7) (2013) 926–43. [PubMed: 23893450]

[5]. Corp DT, Joutsa J, Darby RR, Delnooz CCS, van de Warrenburg BPC, Cooke D, Prudente CN, Ren J, Reich MM, Batla A, Bhatia KP, Jinnah HA, Liu H, Fox MD, Network localization of cervical dystonia based on causal brain lesions, Brain 142(6) (2019) 1660–1674. [PubMed: 31099831]

[6]. Factor SA, Burkhard PR, Caroff S, Friedman JH, Marras C, Tinazzi M, Comella CL, Recent developments in drug-induced movement disorders: a mixed picture, Lancet Neurol 18(9) (2019) 880–890. [PubMed: 31279747]

[7]. Bredenkamp JK, Maceri DR, Inflammatory torticollis in children, Arch Otolaryngol Head Neck Surg 116(3) (1990) 310–3. [PubMed: 2306349]

[8]. Balint B, Vincent A, Meinck HM, Irani SR, Bhatia KP, Movement disorders with neuronal antibodies: syndromic approach, genetic parallels and pathophysiology, Brain 141(1) (2018) 13–36. [PubMed: 29053777]

[9]. Newman JR, Boyle RS, O'Sullivan JD, Silburn PA, Mellick GD, Risk factors for idiopathic dystonia in Queensland, Australia, J Clin Neurosci 21(12) (2014) 2145–9. [PubMed: 25082409]

[10]. Molloy A, Kimmich O, Williams L, Butler JS, Byrne N, Molloy F, Moore H, Healy DG, Lynch T, Edwards MJ, Walsh C, Reilly RB, O'Riordan S, Hutchinson M, An evaluation of the role of

environmental factors in the disease penetrance of cervical dystonia, J Neurol Neurosurg Psychiatry 86(3) (2015) 331–5. [PubMed: 24963124]

[11]. Patti GJ, Yanes O, Siuzdak G, Innovation: Metabolomics: the apogee of the omics trilogy, Nat Rev Mol Cell Biol 13(4) (2012) 263–9. [PubMed: 22436749]

[12]. Walker DI, Lane KJ, Liu K, Uppal K, Patton AP, Durant JL, Jones DP, Brugge D, Pennell KD, Metabolomic assessment of exposure to near-highway ultrafine particles, J Expo Sci Environ Epidemiol 29(4) (2019) 469–483. [PubMed: 30518795]

[13]. Simon-Manso Y, Lowenthal MS, Kilpatrick LE, Sampson ML, Telu KH, Rudnick PA, Mallard WG, Bearden DW, Schock TB, Tchekhovskoi DV, Blonder N, Yan X, Liang Y, Zheng Y, Wallace WE, Neta P, Phinney KW, Remaley AT, Stein SE, Metabolite profiling of a NIST Standard Reference Material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources, Anal Chem 85(24) (2013) 11725–31. [PubMed: 24147600]

[14]. Yu T, Park Y, Johnson JM, Jones DP, apLCMS--adaptive processing of high-resolution LC/MS data, Bioinformatics 25(15) (2009) 1930–6. [PubMed: 19414529]

[15]. Uppal K, Soltow QA, Strobel FH, Pittard WS, Gernert KM, Yu T, Jones DP, xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data, BMC Bioinformatics 14 (2013) 15. [PubMed: 23323971]

[16]. Johnson WE, Li C, Rabinovic A, Adjusting batch effects in microarray expression data using empirical Bayes methods, Biostatistics 8(1) (2007) 118–27. [PubMed: 16632515]

[17]. Uppal K, Walker DI, Jones DP, xMSannotator: An R Package for Network-Based Annotation of High-Resolution Metabolomics Data, Anal Chem 89(2) (2017) 1063–1067. [PubMed: 27977166]

[18]. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, Macinnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L, HMDB: the Human Metabolome Database, Nucleic Acids Res 35(Database issue) (2007) D521–6. [PubMed: 17202168]

[19]. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, Hollender J, Identifying small molecules via high resolution mass spectrometry: communicating confidence, Environ Sci Technol 48(4) (2014) 2097–8. [PubMed: 24476540]

[20]. Benjamini Y, Hochberg Y, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society. Series B (Methodological) 57(1) (1995) 289–300.

[21]. Li S, Park Y, Duraisingham S, Strobel FH, Khan N, Soltow QA, Jones DP, Pulendran B, Predicting network activity from high throughput metabolomics, PLoS Comput Biol 9(7) (2013) e1003123. [PubMed: 23861661]

[22]. Okuzumi A, Hatano T, Ueno SI, Ogawa T, Saiki S, Mori A, Koinuma T, Oji Y, Ishikawa KI, Fujimaki M, Sato S, Ramamoorthy S, Mohney RP, Hattori N, Metabolomics-based identification of metabolic alterations in PARK2, Ann Clin Transl Neurol 6(3) (2019) 525–536. [PubMed: 30911576]

[23]. de Schipper LJ, Boon AJW, Munts AG, Foot Drop Dystonia Resulting from parkin (PARK2) Mutation, Mov Disord Clin Pract 2(3) (2015) 292–294. [PubMed: 30838231]

[24]. Potulska-Chromik A, Hoffman-Zacharska D, Lukawska M, Kostera-Pruszczyk A, Dopa-responsive dystonia or early-onset Parkinson disease - Genotype-phenotype correlation, Neurol Neurochir Pol 51(1) (2017) 1–6. [PubMed: 27667361]

[25]. Quartarone A, Hallett M, Emerging concepts in the physiological basis of dystonia, Mov Disord 28(7) (2013) 958–67. [PubMed: 23893452]

[26]. Wang X, Lin H, Gu Y, Multiple roles of dihomo-gamma-linolenic acid against proliferation diseases, Lipids Health Dis 11 (2012) 25. [PubMed: 22333072]

[27]. Rajagopalan N, Humphrey PR, Bucknall RC, Torticollis and blepharospasm in systemic lupus erythematosus, Mov Disord 4(4) (1989) 345–8. [PubMed: 2811894]

[28]. van den Berg JS, Horstink MW, van den Hoogen FH, Oyen WJ, Dystonia; a central nervous system presentation of Sjogren's syndrome, Mov Disord 14(2) (1999) 374–5. [PubMed: 10091641]

[29]. Hinduja A, Chokroverty S, Hanna P, Grewal RP, Dystonia with superimposed myasthenia gravis: An experiment in nature, Mov Disord 23(11) (2008) 1626–7.

[30]. Green C, Riley DE, Treatment of dystonia in striatal necrosis caused by Mycoplasma pneumoniae, Pediatr Neurol 26(4) (2002) 318–20. [PubMed: 11992764]

Highlights

- The first large-scale metabolomics study of cervical dystonia

- Altered biological pathways may relate to carbohydrate and lipid metabolism

- Novel insights into the disease mechanism of cervical dystonia
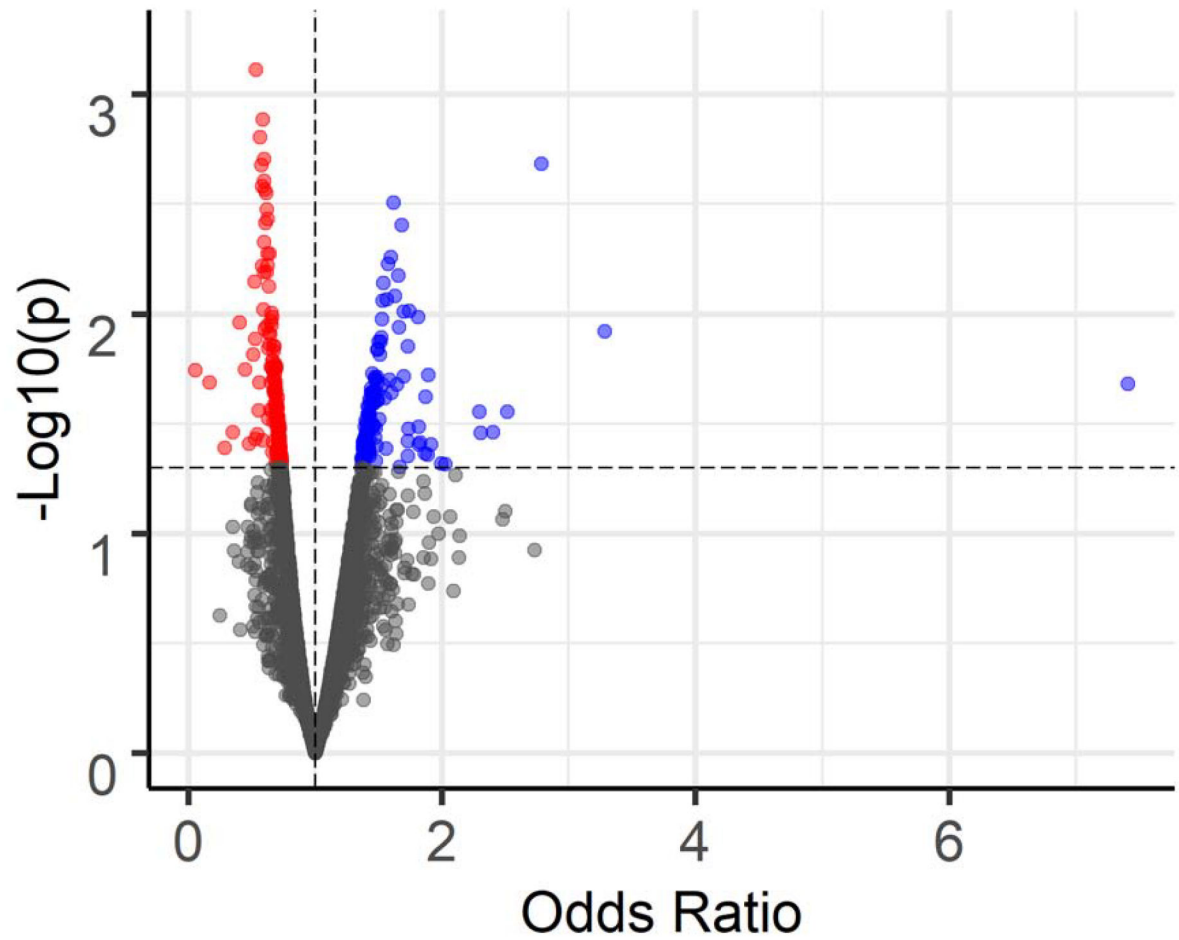
Figure 1.
Volcano plot of the metabolome-wide association study of CD. Total of 7,346 features. Horizontal dashed line indicates p=0.05. 157 features with p<0.05 and OR<1 colored in red; 132 features with p<0.05 and OR>1 colored in blue.
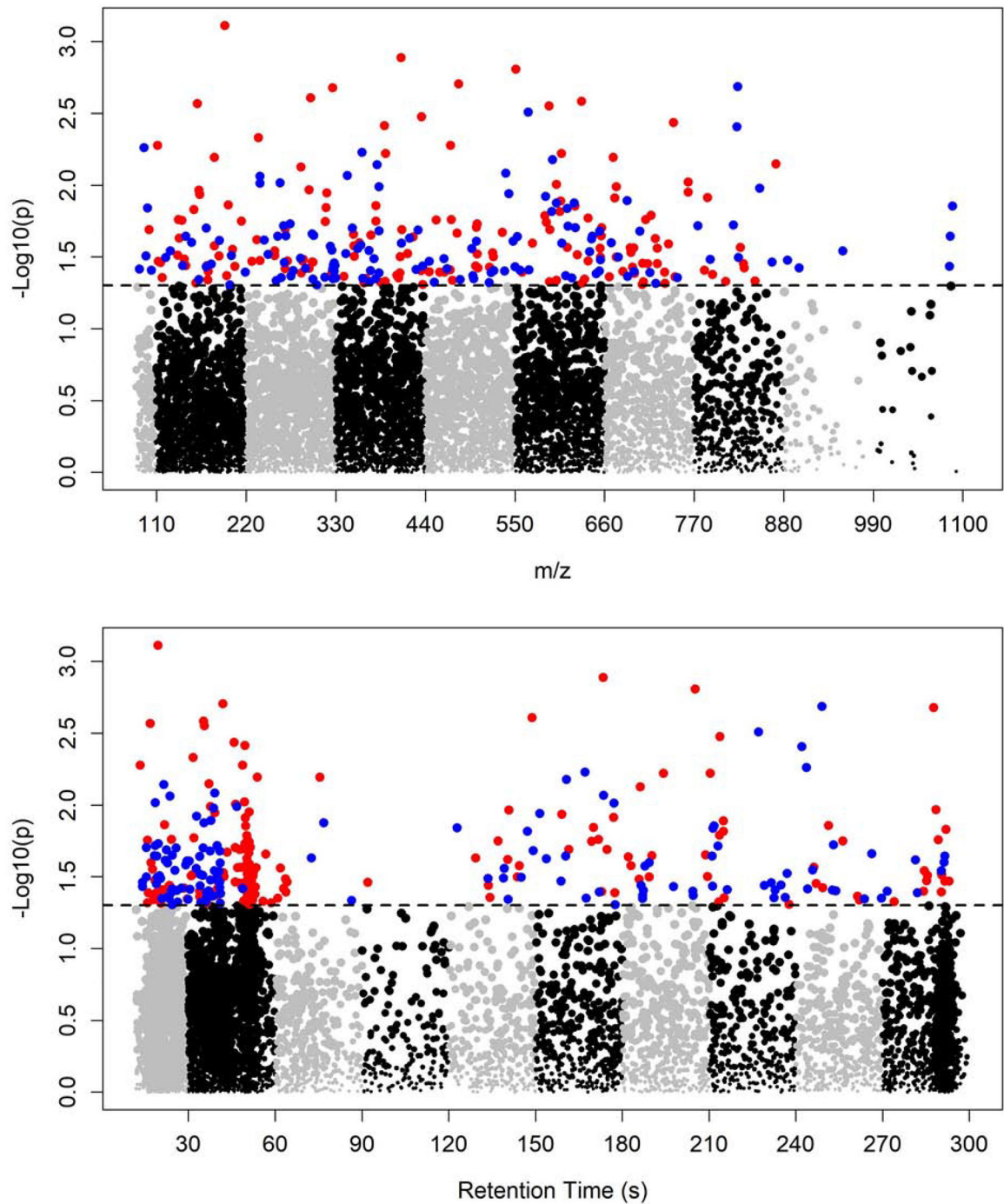
Figure 2A & 2B.
Manhattan plot of the metabolome-wide association study of CD. Total of 7,346 features.
Dashed line indicates p=0.05. 157 features with p<0.05 and OR<1 colored in red; 132
features with p<0.05 and OR>1 colored in blue.

Table 1.

Baseline characteristics.

Variable	Level	Control N=100	CD N=100	P-value[*]
Female (%)		55 (55)	78 (78)	**<.001**
Ethnicity	Black (%)	10 (10)	4 (4)	0.179
	White (%)	88 (88)	95 (95)	
	Other (%)	2 (2)	1 (1)	
Age in year (sd)		57.5 (14.6)	62.3 (11.9)	**0.011**
BMI in kg/m^2 (sd)		28.4 (6.0)	26.1 (4.4)	**0.002**
Weight in lb (sd)		181.4 (48.4)	162.2 (34.7)	**0.001**
Height in inch (sd)		67.1 (4.0)	66.2 (3.8)	0.074
Dietary restrictions (%)		9 (9)	15 (15)	0.277
Smoking	Current smoker (%)	7 (7)	7 (7)	0.912
	Past smoker (%)	27 (27)	30 (30)	
	Never smoker (%)	64 (64)	62 (62)	

[*] Two-sample t-tests used for continuous variables; Chi-square tests or Fisher's exact tests used for categorical variables where appropriate.

Table 2.

Metabolic features matched to possible annotations significantly associated with CD in MWAS at p<0.01.

m/z	Time(s)	Odds Ratio	95% CI	P-value	FDR	Confidence*	Potential Matching Annotations	HMDB** ID	Adduct
							Salicyluric acid	HMDB00840	
							Dopaquinone	HMDB01229	
							4-Carboxyphenylglycine	HMDB02016	
							Alpha-Hydroxyhippuric acid	HMDB02404	
							Leucodopachrome	HMDB04067	
194.046	19.5	0.53	0.37, 0.77	0.001	0.904	2	3-Hydroxyhippuric acid	HMDB06116	M-H
							4-Hydroxyhippuric acid	HMDB13678	
							2-Hydroxy-67-dimethoxybenzoxazole	HMDB30726	
							2-Hydroxy-7-methoxy-2H-14-benzoxazin-3 (4H)-one	HMDB38055	
							L-Dopaquinone	HMDB39119	
							N-acetyl-5-aminosalicylic acid	HMDB60602	
551.359	205.3	0.56	0.40, 0.80	0.002	0.904	2	Cyclopassifloic acid E	HMDB36298	M-H
							4-Hydroxy-L-glutamic acid	HMDB02273	
							M- Ketoglutaric acid oxime	HMDB02467	
160.025	16.9	0.6	0.43, 0.84	0.003	0.904	2	Erucin	HMDB33764	M-H
							M- Thiocyanato-4-(methylthio)butane	HMDB38446	
							N-Formyl-L-aspartate	HMDB60495	
111.02	13.4	0.633	0.46, 0.87	0.005	0.904	2	Uracil	HMDB00300	M-H
							4-Carboxypyrazole	HMDB60760	
287.072	186.2	0.63	0.45, 0.88	0.008	0.904	2	Hydroxyanigorufone	HMDB34449	M-H
							Irenolone	HMDB40716	
							Geranyl 3-methylbutanoate	HMDB29350	
							Linalyl isovalerate	HMDB30428	
237.186	177.2	1.7	1.13, 2.53	0.01	0.904	2	(3S4S6R7S)-110-Bisaboladiene-34-diol	HMDB31383	M-H
							beta-Kessyl alcohol	HMDB31384	
							alpha-Terpinyl pentanoate	HMDB32054	

m/z	Time(s)	Odds Ratio	95% CI	P-value	FDR	Confidence*	Potential Matching Annotations	HMDB** ID	Adduct
							Citronellyl trans-2-methyl-2-butenoate	HMDB32209	
							Geranyl 2-methylbutyrate	HMDB32291	
							Isobornyl 2-methylbutyrate	HMDB32343	
							Guaidiol	HMDB33228	
							Calamendiol	HMDB34673	
							Daucol	HMDB35287	
							Germacrenone	HMDB35887	
							3beta-7-Drimene-311-diol	HMDB36032	
							Tricyclohumuladiol	HMDB36730	
							169-Farnesatriene-311-diol	HMDB36880	
							2-Propenyl cyclohexanehexanoate	HMDB37195	
							210-Bisaboladiene-14-diol	HMDB37279	
							5(1–10)-Abeo-112-patchoulanediol	HMDB37284	
							Curcumadiol	HMDB38136	
							alpha-Bisabolol oxide C	HMDB38138	
							alpha-Bisabolol oxide A	HMDB38196	
							alpha-Bisabolol oxide B	HMDB38197	
							Bornyl valerate	HMDB38247	
							Bornyl isovalerate	HMDB38248	
							Isobornyl isovalerate	HMDB38250	
							Methyl (ZZ)-58-tetradecadienoate	HMDB39130	
							(2E6E9xi)-Farnesol	HMDB39745	
							Terpenyl isovalerate	HMDB40273	
							Geranyl valerate	HMDB40281	
							alpha-Kessyl alcohol	HMDB41589	

*
Confidence: confidence score of the annotation, based on R package xMSannotator.

High confidence: score = 3. Medium confidence: score = 2. Low confidence: score < 2.

**
HMDB: Human Metabolome Database.

Table 3.

Confirmed features significantly associated with CD in MWAS at p<0.05. Logistic regression models adjusted for age, gender, diet restrictions and BMI.

m/z	Time(s)	Odds Ratio	95% CI	P-value	Chemical Name	KEGG[*] ID	HMDB[**] ID	Identification Level[***]
111.0199	13.4	0.64	0.46, 0.87	0.005	Uracil	C00106	HMDB00300	1
187.1338	22.1	1.43	1.05, 1.94	0.024	10-Hydroxydecanoate	C02774	N/A	1
177.0407	19	1.41	1.02, 1.93	0.036	D-Gulonic Acid Gamma-Lactone	C01040	HMDB03466	1
327.2333	234.9	1.40	1.02, 1.91	0.036	Docosahexaenoic Acid	C06429	HMDB02183	1
173.0569	17.9	0.73	0.54, 0.99	0.046	N-Alpha-Acetyl-L-Asparagine	CA1214	HMDB06028	1

[*]
KEGG: Kyoto Encyclopedia of Genes and Genomes.

[**]
HMDB: Human Metabolome Database.

[***]
Identification Level: adapted from Schymanski et al. –

Level 1: confirmed via appropriate measurement of a reference standard with MS, MS/MS and retention time matching.

Level 2: confirmed by MS/MS and matched with online databases or in-silico predicted spectra.

Level 3: confirmed by MS/MS at the chemical class level, but no evidence for a specific metabolite.

Level 4: computationally annotated using xMSannotator with confidence level medium or above.

Level 5: matched accurately by mass.

Table 4.

Significant pathways based on the 289 features associated with CD at p<0.05 in MWAS.

Pathways	Overlap size	Pathway size	P-value
Prostaglandin formation from dihomo gama-linoleic acid	3	5	<.001
Pentose phosphate pathway	4	22	0.005
Arachidonic acid metabolism	3	17	0.013
Leukotriene metabolism	3	18	0.015
Starch and Sucrose Metabolism	2	10	0.030
Valine, leucine and isoleucine degradation	3	23	0.031
Galactose metabolism	3	23	0.031
Fructose and mannose metabolism	2	14	0.047
Propanoate metabolism	2	14	0.047