Disparities in the gut metabolome of post-operative Hirschsprung’s disease patients

Vera Plekhova1, Ellen De Paepe1, Katrien Van Renterghem2, Myriam Van Winckel3, Lieselot Y. Hemeryck1,5 & Lynn Vanhaecke1,4,5

Hirschsprung’s disease (HD) is a congenital structural abnormality of the colon seen in approximately 1 to 5000 live births. Despite surgical correction shortly after presentation, up to 60% of patients will express long-term gastrointestinal complaints, including potentially life-threatening Hirschsprung-associated enterocolitis (HAEC). In this study fecal samples from postoperative HD patients (n = 38) and their healthy siblings (n = 21) were analysed using high-resolution liquid chromatography—mass spectrometry aiming to further unravel the nature of the chronic gastrointestinal disturbances. Furthermore, within the patient group, we compared the faecal metabolome between patients with and without a history of HAEC as well as those diagnosed with short or long aganglionic segment. Targeted analysis identified several individual metabolites characteristic for all HD patients as well as those with a history of HAEC and long segment HD. Moreover, multivariate models based on untargeted data established statistically significant (p < 0.05) differences in comprehensive faecal metabolome in the patients’ cohort as a whole and in patients with a history of HAEC. Pathway analysis revealed the most impact on amino sugar, lysine, sialic acid, hyaluronan and heparan sulphate metabolism in HD, as well as impaired tyrosine metabolism in HAEC group. Those changes imply disruption of intestinal mucosal barrier due to glycosaminoglycan breakdown and dysbiosis as major metabolic changes in patients’ group and should be further explored for potential diagnostic or treatment targets.
To date, several studies have focused on the long-term post-surgical complications of HD and factors that might explain HAEC etiology and incidents. Increasing evidence points towards microbiome alterations as a cause or contributing factor in the development of enterocolitis in HD patients. Moreover, a distinctly different gut microbial pattern has been demonstrated in HAEC patients, even after symptom remission, suggesting that enterocolitis may either be caused by or result in a disruption of the patient’s uniquely adapted intestinal microbiome.

In addition to direct evidence of microbial shifts resulting from intestinal dysbiosis, Demehri et al. demonstrated suppressed microbial short-chain fatty acid (SCFA) formation in HD patients with a history of enterocolitis. SCFAs, mainly butyrate, propionate and acetate, are crucial for normal colonocyte function and a decrease in their concentration has been implicated in the development of inflammatory bowel disease (IBD) In HD patients and in particular patients with a long aganglionic segment have been shown to have a higher risk of developing IBD later in life. Several studies reported decreased mucin production and turnover in HD patients as a potential predisposing factor for the development of HAEC. Indeed, decreased production and turnover of mucus in the protective intestinal mucus layer leads to an impaired intestinal barrier function and contributes to intestinal inflammation by promoting closer contact between host cells and the intestinal microbiota. Furthermore, several microbiome-derived inflammatory markers (including ASCA IgA and OmpC antibodies) were retrieved in serum of both Crohn’s disease and HD patients with HAEC episodes. None of those markers, however, allow to fully explain or predict HAEC episodes and long-term surgery outcomes in HD patients, pointing towards interference from multiple and some yet unknown factors in the manifestation of this pathology.

As a measure to account for the multifactorial nature of some complex gastrointestinal diseases, like e.g. colorectal cancer and IBD, a growing number of studies employ metabolomics for the investigation and comprehensive characterization of biochemical processes within a specific pathophysiological state. Analysis of the fecal metabolome, representing the total sum of small molecules produced in biochemical reactions by the human host and gut microbiota under the influence of genetic, environmental and dietary factors, allows to outline and broadly characterize any gut-related pathophysiological condition. As such, metabolomics does not only enhance our knowledge on the etiology and pathogenesis, but also represents a powerful tool for biomarker discovery, bidding new opportunities for e.g. the development of diagnostic tests. Additionally, non-invasive sampling of feces is more favorable to conventional invasive sampling of blood, particularly in young children.

In this work, we present metabolomics data from stool samples obtained from post-operative HD patients and their healthy siblings. Additionally, within the patients’ group we investigated metabolic shifts in children with and without a history of HAEC as well as those diagnosed with short or long aganglionic segment in search for disease-specific biochemical alterations that can aid to unravel the pathophysiology of HAEC and identify novel diagnostic or prognostic biomarkers.

Results
Comprehensive metabolomics analysis of fecal extracts produced 27,906 positively and 14,854 negatively charged metabolite features reflecting the individuals’ unique fingerprints. Accurate evaluation of specified targeted components (291 standards) allowed the identification of several metabolites responsible for biochemical differences within the investigated groups. Targeted and untargeted data were integrated to elucidate the main pathways underlying the metabolic differences in the investigated groups.

Metabolic differences between post-operative HD patients and their healthy siblings. Targeted metabolite profiling. Of the 142 metabolites retrieved and identified, 9 were significantly different between the HD patient group and their healthy siblings (Fig. 1). Decreased concentrations were observed for piperolic acid (p < 0.001), N-acetylneuraminic acid (p = 0.005), N-acetylgalactosamine (p = 0.017), glutamic acid (p = 0.030), isoleucine acid (p = 0.024), octan-3-one (p = 0.023), fructose (p = 0.027) and serine (p = 0.043). Dopamine concentrations (p = 0.034) were increased in HD patients as opposed to control group.

Untargeted metabolic fingerprinting. The principal component analysis (PCA) revealed the presence of four outliers identified by 95% Hotelling’s T2 criterion and residual distance of the observation to its projection. All four outliers belonged to patients with total colonic aganglionosis and were excluded from further evaluation in this multivariate model. Subsequently, a valid OPLS-DA regression model was constructed, enabling supervised group separation (Fig. 2). Among features contributing the most to group separation, it was possible to list 10 components (VIP > 1.5) that were increased in HD patients as opposed to control group.

Pathway analysis. Pathway analysis was based on 1870 unique metabolites of interest (CV-ANOVA p-value < 0.05). Subsequent pathway visualization (Fig. 3) included both empirically identified untargeted features (Supplementary materials Table S3) as well as significant metabolites from targeted analysis. With adjustment for multiple comparisons, the main differences between the patient’s and sibling’s fecal metabolic fingerprint due to diet did not interfere with disease classification.

Pathway analysis. Pathway analysis was based on 1870 unique metabolites of interest (CV-ANOVA p-value < 0.05). Subsequent pathway visualization (Fig. 3) included both empirically identified untargeted features (Supplementary materials Table S3) as well as significant metabolites from targeted analysis. With adjustment for multiple comparisons, the main differences between the patient’s and sibling’s fecal metabolic fingerprint due to diet did not interfere with disease classification.
prints could be attributed to several catabolic pathways. The most affected was the lysine pathway (p = 0.024). Amino sugar metabolism was significantly (p = 0.047) downregulated in HD patients, as supported by the targeted data where lower glutamic acid, N-acetylglucosamine, N-acetylneuraminic acid and fructose concentrations were measured in the patient group. Also, in HD patients, pathways associated with glycosaminoglycan (GAG) degradation, including hyaluronan (p = 0.049) and heparan sulfate degradation pathway (p = 0.049) were upregulated.

**Metabolic differences between HD patients with and without a history of HAEC.**

**Targeted metabolite profiling.** Within the HD patient cohort, patients who previously endured HAEC episodes showed increased levels of trans-4-hydroxy-l-proline (p = 0.010) and 4-methyl-3-penten-2-one (p = 0.049) as well as a decrease in ethyl pentyl ketone (p = 0.046) compared to patients without a history of HAEC (Fig. 4).

**Untargeted metabolic fingerprinting.** Good descriptive and predictive properties of the OPLS-DA model, i.e., R²Y of 0.844, Q² of 0.55, CV-ANOVA p-value of 2.73e⁻⁰⁵ and acceptable permutation test, demonstrates a significant difference in the comprehensive fecal metabolic fingerprint of patients with and without a history of HAEC (Fig. 5). In total, 19 discriminating features of interest could be listed for the enterocolitis group (Supplementary

---

**Figure 1.** Boxplots representing signal intensities of significantly (p < 0.05) altered metabolites in HD patients (green, n = 34) vs. healthy controls (blue, n = 21) for pipecolic acid (a), N-acetylneuraminic acid (b), N-acetylglucosamine (c), glutamic acid (d), isovaleric acid (e), octan-3-one (f), fructose (g), serine (h) and dopamine (i).
Figure 2. Fecal metabolome OPLS-DA scatter plot of HD patients (n = 34) vs. healthy siblings (n = 21). The validity of the OPLS-DA model was confirmed (R²(y(cum)) = 0.784, Q²(cum) = 0.505, CV-ANOVA p-value = 2.83e⁻⁶ and good permutation testing).

Amino sugar metabolism

Lysine metabolism

Hyaluronan, heparan and sialic acid metabolism

Figure 3. Overview of metabolic pathways underlying the differentiation of the fecal metabolome of HD patients (n = 34) and their healthy siblings (n = 21). The network demonstrates connected metabolic features (hexagons) with corresponding enzyme reaction (diamonds) within statistically significant (p < 0.05) pathways.
Table S4), revealing several dipeptides consisting of phenylalanine as common amino acid: i.e. glutaminylphenylalanine, phenylalanylvaline; as well as 6-hydroxymelatonin and 1-methylnicotinamide.

Pathway analysis. Pathway analysis was based on 1782 features with a CV-ANOV A p-value < 0.05. Among the affected pathways, a significant difference between patients based on their HAEC history was revealed for tyrosine metabolism (p = 0.0028).

Metabolic differences between HD patients with short and long aganglionic segment. Targeted metabolite profiling. In patients with long segment HD (including patients with TCA), 3-methyl-2-cyclohexen-1-one (p = 0.025) and 2-methyl-3-pentanone (p = 0.043) were significantly upregulated in contrast to patients diagnosed with short aganglionic segment (Fig. 6).
Untargeted metabolic fingerprinting. OPLS-DA models based on patient’s aganglionic segment length showed only borderline significant validation parameter values with average prediction abilities. The following parameters were obtained from the OPLS-DA model comparing short segment (n = 24) vs. long segment HD including total colonic aganglionosis (n = 14): $R^2Y(cum) = 0.997$, $Q^2(cum) = 0.452$, CV-ANOVA p-value = 0.014 and acceptable permutation test. However, when evaluating the PCA-X model for natural clustering and outlier detection as described under 2.1.2, it was found that the patients with total colonic aganglionosis (TCA) differed significantly from the other patients and the healthy cohort. Namely, three out of four TCA patients acted as outliers based on Hotteling’s 95% T2 confidence limit and 4 out of 4 observations were classified as outliers based on residual distance of the observation to its projection (DMod). Both the PCA-X score and Distance to model plots can be consulted in the Supplementary material (Fig. S1).

Further, an OPLS-DA model excluding TCA patients was constructed to evaluate differences with short (n = 24) and long (n = 10) segment aganglionosis. Despite exclusion of potential outliers, the performance of the model did not significantly improve, with $R^2Y(cum) = 0.849$, $Q^2(cum) = 0.326$, CV-ANOVA p-value = 0.026 and acceptable permutation test. As no fully validated untargeted models were constructed, no discriminating feature extraction was performed for pathway analysis.

Discussion
In this study, several gut microbiome derived metabolites and fermentation pathways, as a direct reflection of gut microbiome functionality, were established to be discriminating for the fecal metabolome of HD patients compared to their healthy siblings. Indeed, involvement of the gut microbiome in the pathogenesis of post-operative HD complications is in line with previously published reports. This can be viewed as a consequence of the (partial) colon resection, as the latter drastically changes the intestinal microenvironment required for maintaining a healthy microbiome. Lysine catabolism was upregulated in HD patients and has previously been associated with inadequate bacterial butyrate production in the large bowel, in which lysine represents alternative to carbohydrates as the main substrates for SCFA production. Since a drop in amino sugars (cfr. the reported decrease in N-acetylglucosamine and N-acetylneuraminic acid) has also been linked to impaired carbohydrate digestion resulting from dysbiosis, these findings points towards microbiome-related disruptions in HD patients. This is in line with previous reports on the distinctive microbiome and SCFA profile in HD patients as a potential contributing factor in the development of enterocolitis. A distinctly different gut microbial pattern has e.g. indeed been demonstrated in HAEC patients, even after symptom remission, suggesting that enterocolitis may either be caused by or result in a disruption of the patient’s uniquely adapted intestinal microbiome.

Equally important is the involvement of dysbiosis in altering the intestinal mucus layer. Several detected alterations point towards an increased degradation of major mucus components such as GAGs, regardless of patient’s HAEC history. This mucus depletion has been investigated in light of several other gastro-intestinal diseases, such as ulcerative colitis, and is known to be a factor supporting and promoting gut inflammation. Moreover, several previous studies have reported decreased mucin production and turnover specifically in HD patients with a possible link to developing enterocolitis. Here we find upregulated hyaluronan and heparan sulfate degradation in HD patients, involved in GAG turnover. GAGs together with mucopolysaccharides are essential components of the protective intestinal

Figure 6. Boxplots representing signal intensities of significantly (p < 0.05) altered metabolites in patient groups diagnosed with short aganglionic segment (green, n = 24) and with either long aganglionic segments or total colonic aganglionosis (blue, n = 14) for 3-methyl-2-cyclohexene-1-one (a) and 2-methyl-3-pentanon (b).
mucus layer. In the healthy colon, simple sugars derived from GAGs provide binding sites and nutrition for commensal microbiota. However, in case of mucus barrier dysfunction or damage, the supplementary nutrient source provided by degradation of GAGs will contribute to bacterial overgrowth. In addition, the degradation of GAGs is accompanied by the release of small hyaluronan fragments, known to stimulate leukocyte adhesion and inflammatory cytokine production. When overdriven it has been associated with chronic gut inflammation. Additional evidence supporting the theory of increased microbial mucus degradation in HD patients could be found in significantly altered sialic or N-acetylneuraminic acid metabolism, since many commensal bacteria produce sialidases to harvest sialic acid at the terminal regions of mucin molecules. Mucins, as well as GAGs, are essential for a normal activity of the local immune system and epithelial barrier integrity. Additionally, the mucus layer provides a natural habitat for the residing gut microbiota and has been suggested to impact on colon-specific microbiome formation. Loss of the mucus barrier stimulates passive intestinal absorption and allows closer contact between host immune cells and bacterial antigens. An increased rate of mucin degradation and a decrease in epithelial barrier function are therefore additional predisposing factors for maintaining an intestinal inflammatory state. In other words, metabolic and enzymatic shifts introduced by a dysregulated gut microbiome induce increased mucin deterioration, which in turn creates favorable circumstances for further development of dysbiosis. In HD patients, colonic dysbiosis has been appointed as one of the major drivers of the development of enterocolitis. However, alterations in microbial composition can also be responsible for more chronic HD complications, such as constipation and abdominal distention. Future research is warranted to confirm that the observed metabolic alterations are coinciding the hypothesized shifts in microbiome composition and would thus benefit from an integrative omics approach.

Among evidence of dysbiosis and alterations in mucus turnover in HD patients in this study, HAE patients presented additional metabolic changes including increased tyrosine catabolism, a decreased Trans-4-hydroxy-L-proline (Hyp) degradation and alterations in specific volatile compounds (ethyl pentyl ketone and 4-methyl-3-penten-2-one). Hyp is the common product of anaerobic proline transformation utilized by the gut microbiota, primarily by Clostridium spp. Clostridia are widely abundant commensal species that have been shown to populate specific regions of the gastrointestinal tract, mainly the ascending colon, where they act as major butyrate-producing species. Increased Hyp concentrations may imply a decrease in Hyp-utilizing microbiota and can explain lower SCFAs concentrations in HAEC patients, as previously demonstrated. Further, 4-methyl-3-penten-2-one—increased in patients with a history of HAEC—and ethyl pentyl ketone—decreased in patients with a history of HAEC—are volatile compounds that can originate from the microbiome, but also from nutrition (host metabolite). In pediatric celiac disease patients, 4-methyl-3-penten-2-one and ethyl pentyl ketone were shown to be augmented in urine and saliva, and associated with shifts in the gut microbiome. Interestingly, similarities of HAEC biomarkers with pediatric celiac disease markers hint towards similarities in pathophysiology and warrant further investigation into potential diagnostic and/or management correlations, furthermore, providing promising predictive, diagnostic and/or prognostic marker candidates.

Among the affected pathways, a significant difference between patients based on their HAEC history was only revealed for tyrosine metabolism. Aromatic amino acids such as tyrosine have been shown to act as signaling molecules within the microbiome-host cross-talk. Moreover, phosphorylation of tyrosine by tyrosine kinases is involved in multiple aspects of cell metabolism regulation. Tyrosine kinase inhibitors have demonstrated potential to subdue intestinal inflammation by reducing cytokine expression in response to bacterial stimulation and represent promising therapeutical targets for colitis patients. Although the same process of gut-microbiome tyrosine dependent interaction can be suggested for HAEC patients, the exact mechanism and consequences are yet to be determined.

In patients with long segment HD (including patients with TCA), 3-methyl-2-cyclohexen-1-one (p = 0.025) and 2-methyl-3-pentanone (p = 0.043) were found to be upregulated. Both metabolites represent volatile organic compounds commonly detected in fecal samples. In literature, an increase in 2-methyl-3-pentanone has been reported in patients with Campylobacter jejuni infections, while 3-methyl-2-cyclohexen-1-one has been observed in patients with Clostridium difficile infections. Although more data are needed to determine the significance of this increase in long segment HD patients, it is noteworthy that microbial infections or dysbiosis can be again assigned as probable cause. Indeed, a significant disparity in the comprehensive metabolic fingerprint of long and short segment HD was observed, as unsupervised data exploration based on PCA-X clearly separated TCA patients as outlying observations in the entire cohort. Those findings imply that the fecal metabolome of TCA patients differs markedly not only from healthy subjects, but also from patients diagnosed with short and long aganglionic segments. Although we are unable to draw conclusions about the nature of those changes due to the small sample size of the TCA group (n = 4), these preliminary data can act as proof-of-concept for further research focusing on TCA patients.

Material and methods

Subjects and samples. Participants. Participants included 38 post-surgery HD patients and 21 healthy siblings recruited at the Ghent University Hospital (Prof. Van Winckel and Dr. Van Renterghem). Enrolled patients were diagnosed with a short aganglionic segment (n = 24), long segment (n = 10), and total colonic aganglionic (n = 4) that underwent the same surgical treatment (colonic pull through procedure). Fecal samples were collected at the earliest one year after surgery to ensure adequate recovery. Within the HD cohort, 19 patients (50%) reported to be diagnosed with at least one episode of HAEC over the course of the disease. A more detailed overview of the investigated cohort is included in supplementary material Table S1. This study was approved by the Ghent University Hospital ethical committee (EC 2015/1193) and executed in accordance with relevant guidelines and regulations. The parental informed consent (EC 2015/1426) was obtained from each participant in accordance with the Declaration of Helsinki.
Sample handling. Fecal samples were obtained through spontaneous defecation or routine enema in those unable to spontaneously defecate and stored at −20 °C until pick-up (up to maximum two weeks). Following cooled transport (dry ice) to the lab, samples were briefly stored at −80 °C, thereafter lyophilized and homogenized in preparation of a more long-term storage at −80 °C awaiting analysis.

Metabolomics. Standards and reagents. Solvents for extraction and chromatographic separation were of LC–MS purity grade and purchased from WVR International (Merck) and Thermo Fisher Scientific. Ultrapure water was produced in-house using a Sartorius arium® 611 Ultrapure Type 1 Water System (Thermo Fisher Scientific). Chemical standards (a total of 291 metabolites and D-valine-d8 as internal standard) were obtained from diverse suppliers as reported previously. Standards and reagents.

Sample analysis and quality assurance. Polar metabolome extraction, chromatographic separation and mass spectrometric analysis followed protocols described previously. Chromatographic separation was performed using an Accela UHPLC installation (Thermo Fisher Scientific) equipped with an Acquity HSS-T3 C18 column (1.8 μm, 150 mm × 2.1 mm, Waters) and VanGuard precolumn (1.8 μm, 5 mm × 2.1 mm, Waters) as reported previously. Mass spectrometric detection was carried out on a hybrid quadrupole-Orbitrap high-resolution mass spectrometer (Q-Exactive™, Thermo Fisher Scientific) equipped with a heated electrospray ionization source (HESI-II).

In light of targeted analysis, a mixture of 291 analytical standards was prepared in a pooled matrix extract and analyzed alongside the samples, under the same instrumental conditions. Signal linearity was ensured by dilution series. To assure instrumental precision (mass deviation < 5 ppm), the instrument was calibrated in both polarity modes according to manufacturer recommendations. Pooled quality control (QC) samples consisting of equal aliquots of all sample extracts were analyzed alongside test specimens at equal intervals to monitor stability of instrument performance. For the same purpose, the internal standard D-valine-d8 was added to all extracted samples at a concentration of 100 ng mL⁻¹.

Data analysis. Analysis of targeted metabolites. Targeted data processing, including peak detection and comparison with the reference standard solution mix, was performed using Xcalibur™ 3.0 (Thermo Fisher Scientific). Compound intensities were normalized to the intensity of the reference internal standard (D-valine-d8).

Partial least squares discriminant analysis (OPLS-DA) was employed for supervised group separation. Validation of OPLS-DA models was based on the total variation explained by the model (R²Y) close to 1, model predictive properties (Q²) above 0.5, CV-ANOVA p-value < 0.05 and valid permutation testing (n = 100). For valid models, features contributing the most to group separation were selected based on a variable importance (VIP) score > 1.5, a correlation (pcorr) > 0.5 and covariance (p) > 0.02 derived from the S-plot and pareto-scaling to achieve approximate normal distribution. Further, principal component analysis (PCA) was used for exploration of inherent sample clustering and outlier detection. Subsequently, orthogonal partial least squares discriminant analysis (OPLS-DA) was employed for supervised group separation. Validation of OPLS-DA models was based on the total variation explained by the model (R²Y) close to 1, model predictive properties (Q²) above 0.5, CV-ANOVA p-value < 0.05 and valid permutation testing (n = 100). For valid models, features contributing the most to group separation were selected based on a variable importance (VIP) score > 1.5, a correlation (pcorr) > 0.5 and covariance (p) > 0.02 derived from the S-plot and Jack-knifed confidence intervals not including zero. Tentative identification was performed by HMDB database search, based on accurate m/z with a mass deviation tolerance of 5 ppm. Putative identities were cross-checked with the features' ¹²C/¹³C isotope ratio, corresponding to Tier level 3 confidence level of identification.

Combined pathway analysis. Pathway analysis was performed on the MetaboAnalyst 4.0 platform applying the “MS peaks to Pathways” tool based on the mummichog algorithm. Mass deviation tolerance was set at 5 ppm. Significance of metabolite identification was determined as a p-value of permutation test between any other metabolites in the reference database. Pathway significance was determined by enrichment analysis with Holm adjustment for multiple comparisons. Pathway visualization was performed on the MetScape 3 platform, based on the combined targeted and untargeted data.

Data availability
The experimental output and anonymized patient data will be made freely available upon publication.

Received: 12 May 2021; Accepted: 19 July 2021
Published online: 09 August 2021

References
1. Whitehouse, E. R. & Kernohan, J. W. Myenteric plexus in congenital megacolon: Study of Eleven Cases. Arch. Intern. Med. 82, 75–111 (1948).
2. Bolande, R. P. The neurocristopathies A unifying concept of disease arising in neural crest maldevelopment. Hum. Pathol. 5, 409–429 (1974).
44. Pierre, J. E. et al. Intestinal dysbiose and bacterial enteroinvasion in a murine model of Hirschsprung’s disease. *J. Pediatr. Surg.* **49**, 1242–1251 (2014).
45. Cao, H. et al. Dysbiosis contributes to chronic constipation development via regulation of serotonin transporter in the intestine. *Sci. Rep.* **7**, 10322 (2017).
46. Huang, Y. Y., Martínez-Del Campo, A. & Balskus, E. P. Anaerobic 4-hydroxyproline utilization: Discovery of a new glycyl radical enzyme in the human gut microbiome uncovers a widespread microbial metabolic activity. *Gut Microbes* **9**, 437–451 (2018).
47. Nava, G. M., Friedrichsen, H. J. & Stappenbeck, T. S. Spatial organization of intestinal microbiota in the mouse ascending colon. *ISME J.* **5**, 627–638 (2011).
48. Pryde, S. E., Duncan, S. H., Hold, G. L., Stewart, C. S. & Flint, H. J. The microbiology of butyrate formation in the human colon. *FEBS Lett.* **217**, 133–139 (2002).
49. Francavilla, R. et al. Salvatory microbiota and metabolome associated with celiac disease. *Appl. Environ. Microbiol.* **80**, 3416–3425 (2014).
50. Di Cagno, R. et al. Different fecal microorganisms and volatile organic compounds in treated and untreated children with celiac disease. *Appl. Environ. Microbiol.* **75**, 3963–3971 (2009).
51. Liu, Y., Hou, Y., Wang, G., Zheng, X. & Hao, H. Gut microbial metabolites of aromatic amino acids as signals in host-microbe interplay. *Trends Endocrinol. Metab.* **31**, 818–834 (2020).
52. Wu, P., Nielsen, T. E. & Clausen, M. H. Small-molecule kinase inhibitors: An analysis of FDA-approved drugs. *Drug Discov. Today* **21**, 5–10 (2016).
53. Garner, C. E. et al. Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. *FASEB J.* **21**, 1675–1688 (2007).
54. De Spiegeleer, M. et al. Impact of storage conditions on the human stool metabolome and lipidome: Preserving the most accurate fingerprint. *Anal. Chem. Acta.* **1108**, 79–88 (2020).
55. Vanden Bussche, J., Marzaroti, M., Laukens, D. & Vanhaecke, L. Validated high resolution mass spectrometry-based approach for metabolomic fingerprinting of the human gut phenotype. *Anal. Chem.* **87**, 10927–10934 (2015).
56. Ulaszewska, M. M. et al. Nutrimetabolomics: An integrative action for metabolic analyses in human nutritional studies. *Mol. Nutr. Food Res.* **63**, 1800384 (2019).
57. Rocke, D. M. & Durbin, B. Approximate variance-stabilizing transformations for gene-expression microarray data. *Bioinformatics* **8**, 966–972 (2003).
58. Szymańska, E., Saccenti, E., Smilde, A. K. & Westerhuis, J. A. Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **8**(Suppl 1), 3–16 (2012).
59. de Souza, L. P., Naake, T., Tohge, T. & Fernie, A. R. From chromatogram to analyte to metabolite. How to pick horses for courses from the massive web resources for mass spectral plant metabolomics. *GigaScience* **6**, 1–20 (2017).
60. Summer, L. W. et al. Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**, 211–221 (2007).
61. Li, S. et al. Predicting network activity from high throughput metabolomics. *PLoS Comput. Biol.* **9**, e1003123 (2013).

**Acknowledgements**

The authors gratefully acknowledge all who have contributed to this study; first and foremost, all participating patients, their brothers and sisters, as well as their parents. In addition, the authors would also like to explicitly thank Saskia De Groote, Inge Vandenbroucke, Eva Pape, Mieke Naessen, Joke Goedgebuer, Beata Pomian and Dirk Stockx for their most invaluable contribution to this work. The Laboratory of Chemical Analysis is part of the Ghent University MSsmall Expertise Centre.

**Author contributions**

All authors read and approved the final version of the manuscript. V.P. contributed to Formal analysis, Investigation, Visualization and Writing—original draft; E.D.P. performed Investigation and Writing—review & editing; K.V.R. and M.V.W. contributed to study Conceptualization, Methodology, Recourses and Writing—review and editing; L.Y.H contributed to Conceptualization, Investigation, Methodology, Project administration, Supervision and Writing—review and editing; L.V. Conceptualization, Project administration, Supervision and Writing—review & editing.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-95589-0.

**Correspondence** and requests for materials should be addressed to L.V.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021