ORIGINAL ARTICLE

Correlations for untargeted GC × GC-HRTOF-MS metabolomics of colorectal cancer

Nicolas Di Giovanni[1] · Marie-Alice Meuwis[2] · Edouard Louis[2] · Jean-François Focant[1]

Abstract

Introduction Modern comprehensive instrumentations provide an unprecedented coverage of complex matrices in the form of high-dimensional, information rich data sets.

Objectives In addition to the usual biomarker research that focuses on the detection of the studied condition, we aimed to define a proper strategy to conduct a correlation analysis on an untargeted colorectal cancer case study with a data set of 102 variables corresponding to metabolites obtained from serum samples analyzed with comprehensive two-dimensional gas chromatography coupled to high-resolution time-of-flight mass spectrometry (GC×GC-HRTOF-MS). Indeed, the strength of association existing between the metabolites contains potentially valuable information about the molecular mechanisms involved and the underlying metabolic network associated to a global perturbation, at no additional analytical effort.

Methods Following Anscombe's quartet, we took particular attention to four main aspects. First, the presence of non-linear relationships through the comparison of parametric and non-parametric correlation coefficients: Pearson's r, Spearman's rho, Kendall's tau and Goodman–Kruskal's gamma. Second, the visual control of the detected associations through scatterplots and their associated regressions and angles. Third, the effect and handling of atypical samples and values. Fourth, the role of the precision of the data on the attribution of the ranks through the presence of ties.

Results Kendall's tau was found the method of choice for the data set at hand. Its application highlighted 17 correlations significantly altered in the active state of colorectal cancer (CRC) in comparison to matched healthy controls (HC), from which 10 were specific to this state in comparison to the remission one (R-CRC) investigated on distinct patients. 15 metabolites involved in the correlations of interest, on the 25 unique ones obtained, were annotated (Metabolomics Standards Initiative level 2).

Conclusions The metabolites highlighted could be used to better understand the pathology. The systematic investigation of the methodological aspects that we expose allows to implement correlation analysis to various fields and many specific cases.

Keywords Metabolic correlations · GC×GC · Comprehensive gas chromatography · Untargeted metabolomics · QC system · Colorectal cancer

✉ Jean-François Focant
  jf.focant@uliege.be

  Nicolas Di Giovanni
  n.digiovanni@uliege.be

  Marie-Alice Meuwis
  marie-alice.meuwis@uliege.be

  Edouard Louis
  edouard.louis@uliege.be

[1] Department of Chemistry, Organic and Biological Analytical Chemistry Group, Quartier Agora, University of Liège, Allée du Six Août,B6c, B-4000 Liège, Sart Tilman, Belgium

[2] GIGA Institute, Translational Gastroenterology and CHU de Liège, Hepato-Gastroenterology and Digestive Oncology, Quartier Hôpital, University of Liège, Avenue de L'Hôpital 13, B34-35, B-4000 Liège, Belgium

🠚 Springer

1 Introduction

Colorectal cancer (CRC) is the third deadliest (9.2%, 0.86 million deaths) and the fourth most diagnosed (6.1% of all cancers, 1.1 million new cases in 2018) cancer worldwide (Bray et al., 2018). Because of its increase in countries in transition, and despite the stabilization and decrease in mortality observed in developed countries where the rates of incidence are the highest, it is expected to reach 2.2 million new cases and 1.1 million deaths by 2030 (Arnold et al., 2017). The internationally recognized diagnosis relies on the analysis of a tissue sample by a pathologist obtained by an invasive method of clinical examination. Metabolic profiling (Lin et al., 2019; Zhang et al., 2017), the untargeted analysis of the metabolites in a sample, is widely performed in translational research (Collino et al., 2013; Mamas et al., 2011). Since it aims to measure as many compounds as possible, it is mostly a discovery, hypothesis generating approach (Kell & Oliver, 2004). Modern instrumentations, particularly comprehensive ones such as comprehensive two-dimensional gas chromatography (GC × GC) (Zanella et al., 2021), generate high-dimensional, complex data sets that present a considerable challenge in terms of their interpretation (Mendes et al., 2005; Weckwerth & Morgenthal, 2005) and therefore require adapted statistical tools to extract as much chemical information as possible to be translated into biologically relevant knowledge (Tebani et al., 2018). Centered on diagnosis capability, biomarker research provides limited biological knowledge even when the marker metabolites are linked to their metabolic pathways, an approach that has become more frequent over the years (Farshidfar et al., 2012; Mal et al., 2012; Tan et al., 2013). On the other hand, the correlations or strength of association existing between the metabolites, calculated as the statistical dependence between them, are rarely considered. This is despite the publication more than 15 years ago of pioneer studies (Camacho et al., 2005; Mendes et al., 2005; Steuer et al., 2003), the facility to perform such an analysis through available workflows (Xia & Wishart, 2016), and the potentially valuable information they contain about the specific metabolic changes induced in the underlying metabolic network associated to a biological process (Alonso et al., 2015; Kotze et al., 2013; Steuer, 2006). Indeed, the correlation analysis can be seen as a fingerprint of the enzymatic and regulatory reaction network and a measure of their alterations between the biological groups compared in the biomarker discovery study (Siska & Kechris, 2017). It has the potential to improve the understanding of the various molecular mechanisms involved in a phenomenon -for example the occurrence, progression, remission and recurrence of a disease- through the generation of interesting hypothesis (Wilcox & Rousselet, 2018). Such analysis is particularly interesting in cases where the concentrations of the metabolites are not strongly altered and when the effect measured is a combination of many low impact factors that are thus hard to unveil individually (Tebani et al., 2018). In addition, they make possible to complete the metabolic pathways as they are already known as well as to discover and to model associations outside them (Kotze et al., 2013; Weckwerth & Fiehn, 2002). Finally, they require no additional analytical effort since they are purely data processing. On the other hand, they demand strong metabolic knowledge for their interpretation (Camacho et al., 2005; Steuer, 2006). This study aims to investigate the methodological aspects of correlation analysis, based on Anscombe's quartet (Anscombe, 1973), i.e. to develop strategies in order to properly detect and visually control the significant correlations and variations of correlations through a case study where a metabolomics data set of colorectal cancer was obtained through serum samples from four subgroups of patients analyzed with comprehensive two-dimensional gas chromatography (GC × GC) coupled to time-of-flight mass spectrometry (TOF–MS).

2 Results and discussion

2.1 Parametric coefficient. Presence of atypical samples

Regarding the samples, no outlier (consistently out of the limits defined by the 95% ellipses or clustered at a high dissimilarity (HCA)) was observed with PCA, HCA and PLS plots in the log-transformed data, but 6 samples were dissimilar (regularly out of or close to the limits) from the group they belong to: CRC18, HC23, R-CRC38, R-CRC46, R-HC56 and R-HC58 (see SI S-1A). In the raw data, we detected two outliers: CRC18 and HC37, and 4 dissimilar points: CRC17, R-CRC38, R-HC56 and R-HC65. Overall, the best homogenization of the samples in each subgroup was obtained with the log transformation, particularly when looking at the HCA plots.

Based on the robust z-scores and the relative ranks, one outlier: HC37, and 4 dissimilar points: CRC14, CRC18, HC32 and R-CRC38 were detected in the log-transformed data (SI S-1B). Similarly, 6 outliers were observed in the raw data: CRC18, HC37, R-CRC38, RCRC-39, R-CRC40 and R-HC57, along with 4 dissimilar points: CRC9, CRC17, R-HC55 and R-HC56. Overall, CRC 18 and R-CRC38 were consistently found as dissimilar points in the log-transformed data while CRC18 and HC37 were consensus outliers and CRC 17, R-CRC38 and R-HC56 were found consensus dissimilar points in the raw data. Again, as expected, the

log transformation made the data more Gaussian and less sensitive to the atypical samples. And this procedure showed efficient to highlight the main atypical samples.

2.2 Parametric coefficient. Presence of atypical values

Many atypical values were observed in all four groups. Indeed, in the 29 selected correlations, we found 36 variables that had at least one outlier and 45 variables that had at least one outlier or one dissimilar value in the raw data, for a total of 55 outliers and 69 dissimilar values. After the log transformation, 19 variables had a least one outlier and 42 had at least one outlier or one dissimilar value, for a total of 24 outliers and 53 dissimilar values. Globally, for each type of data, 27 correlations had at least one outlier or one dissimilar value while 25 (raw data) and 14 (log-transformed data) had at least one outlier. This is likely because the 29 correlations were selected precisely for their differences of values between the coefficients, which is probably due to the presence of outliers. In addition, we observed that the atypical values in the raw and log-transformed data were often high and low ones, respectively (SI S-2A). This is likely due to the application of the log transformation, which reduces the right skewness of a distribution by moving it globally to the left, to (almost) symmetrical data. As for the atypical samples, we also confirmed on our specific data set that the log-transformed data were less prone to atypical values, especially to outliers. From a methodological point of view, looking at the detection methods, in almost half the cases, univariate (boxplots and Grubbs and Dixon tests) and multivariate (2D scatterplots) tools were in agreement. However, in agreement with previous observations (Pernet et al., 2013), bivariate visualization was found useful to take into account the alignment of a value with the other points of a specific correlation as well as to evaluate properly the reality of the correlation (Anscombe, 1973; Wilcox, 2004) (SI S-2B). Indeed, the alignment was found responsible for

the different status (normal value, dissimilar one or outlier) of a single value according to the correlation involved. As the Dixon test is limited to the detection of one outlier, it logically underestimated the presence of outliers. Therefore, it is overall recommended to use Grubbs test and boxplots, along with scatterplots that allow to visualize the correlations by considering both variables simultaneously, and to pay particular attention to the non-aligned values. Finaly, the percentages of outliers and dissimilar values in the selected correlations designated the CRC sample 18 as an outlier and the HC sample 19 as a dissimilar sample in the raw data, with respectively 38 and 12% of outlier values and 45 and 33% of dissimilar values. In the log-transformed data, they highlighted the CRC sample 18 as a dissimilar sample with 21% and 29% of outlier and dissimilar values.

2.3 Effect of atypical samples and values on r and Z

Combining the results above, it was found that the samples CRC18 and HC37 were potential outliers in the raw data while CRC17, R-CRC38 and R-HC56 were dissimilar samples. In the log-transformed data, CRC18 and R-CRC38 were dissimilar samples. The effect of those atypical samples was assessed by removing them, which is equivalent to a skipped-correlation coefficient (McClelland, 2000; Pernet et al., 2013), and by comparing the values of r and Z (Z_{CRC} and Z_{R-CRC}) with and without them (Wilcox, 2004). Other options exist that were tested elsewhere and found less powerful and less efficient (Pernet et al., 2013), particularly with bivariate outliers, such as the winsorizing of the data (Hardin et al., 2007) or the use of alternative coefficients such as Tukey's biweight (Wilcox, 1994) or a percentage bend Pearson coefficient (Janse et al., 2021). As expected, the dissimilar samples were less influential than the outliers, especially for Z where the effect of atypical samples in two subgroups could combine (Tables 1 and 2). In addition, the dissimilar samples in the log-transformed data (CRC 18)

Table 1 Effect of the atypical samples on Pearson's r

r			Mean		Median	
			Absolute	%	Absolute	%
CRC	Raw	1 Outlier	0.21	221	0.15	58
HC	Raw	1 Outlier	0.16	313	0.08	41
CRC	Raw	1 Dissimilar	0.07	95	0.03	17
R-CRC	Raw	1 Dissimilar	0.13	197	0.06	28
R-HC	Raw	1 Dissimilar	0.04	64	0.02	10
CRC	Log	1 Dissimilar	0.09	158	0.06	23
R-CRC	Log	1 Dissimilar	0.09	169	0.06	22

Mean and median absolute and relative (in percentage of the correlation values with the outliers) differences of Pearson's correlation coefficients r in the four subgroups of samples, for all the 5151 correlations in the data set, when removing the atypical samples

Table 2 Effect of the atypical samples on the differential correlation coefficient Z_{CRC} and $Z_{R\text{-}CRC}$

Z	Mean		Median	
	Absolute	%	Absolute	%
CRC/HC Raw–2 Outliers	0.96	332	0.74	95
CRC/HC Raw–1 Dissimilar	0.22	102	0.10	16
R-CRC/R-HC Raw–2 Dissimilar	0.44	249	0.23	37
CRC/HC Log–1 Dissimilar	0.28	138	0.17	29
R-CRC/R-HC Log–1 Dissimilar	0.25	132	0.17	25

Mean and median absolute and relative (in percentage of the correlation values with the outliers) differences of the differential Z_{CRC} and $Z_{R\text{-}CRC}$, for all 5151 correlations in the data set, when removing the atypical samples

and in the raw data (R-CRC 1) produced similar absolute and relative differences, for both r and Z.

The atypical values also had a clear effect on Pearson's r and on the differential Z, especially with the raw data (Table 3). Again, and as expected, the raw data were more sensitive to the presence of atypical values than the log-transformed ones (Wilcox, 2004), but the two suffered from this phenomenon. Therefore, the log transformation revealed efficient but not sufficient to make the GC×GC-TOF–MS data fully Gaussian.

This was observed visually on the scatterplots where the deviant values biased the linear regressions and the corresponding angles between them (Anscombe, 1973;

Wilcox, 2004) (SI S-3A). When considering the angles, scaling the data prior to the calculation was reported to be of importance to perform a relevant analysis and to avoid misinterpretations (Camacho et al., 2005). Four types of scaling were tested, consisting in the multiplication of the slopes of the linear regressions by the ratio of the means, the medians, the maxima or the ranges of values of the subgroups considered (CRC and HC or R-CRC and R-HC). Logically since the correlations values depend on the range of observations (Giovanni et al., 2020), we found that the range method was the most consistent with the associations observed in the scatterplots as well as with the angles calculated on the ranks, taken as the reference since they measure the metabolites on the same scale and thus do not require any change (SI S-3B). The results obtained confirmed what was seen with the numerical calculations since with the raw data, even the dissimilar values were influential, while with the log-transformed data, only the clear outliers represented an issue (SI S-3C). Overall, and to conclude on this section, when using Pearson's r on GC×GC-TOF–MS, we found appropriate to use the log-transformed data along with a proper, uni- and multivariate outlier detection strategy, that includes a visual control through 2D graphical representations (Wilcox, 2004), in order to look for the presence of atypical samples and values. In our view, the same procedure could be applied to biomarker research as well, particularly if non-parametric statistical methods are not included in the biomarker selection process.

Table 3 Effect of the atypical values on Pearson's r and on the differential Z_{CRC} and $Z_{R\text{-}CRC}$

Absolute	r		Z	
Raw	Outliers	All atypical	Outliers	All atypical
Mean	0.21	0.28	1.34	1.62
Median	0.12	0.21	0.63	0.95
Log	Outliers	All atypical	Outliers	All atypical
Mean	0.06	0.13	0.40	0.74
Median	0.00	0.08	0.00	0.68
%	r		Z	
Raw	Outliers	All atypical	Outliers	All atypical
Mean	99	136	98	124
Median	31	64	60	74
Log	Outliers	All atypical	Outliers	All atypical
Mean	18	44	20	44
Median	0	17	0	29

Mean and median absolute (above) and relative (in percentage of the (variation of) correlation values with the outliers, below) differences between the correlation coefficients r and their variations Z_{CRC} and $Z_{R\text{-}CRC}$ with and without the outliers or all atypical values, for all 5151 correlations

Table 4 Defined thresholds for confident and probable ties and non-ties

	t-value	p-value* (%)	Hedges's g	RSD (%)
Confident tie	0.60	70	0.20	10
Probable (non-) tie	1.50	92	0.50	25
Confident non-tie	2.40	99	0.80	40

*p-values for sample sizes $17 \leq n \leq 19$

2.4 Non-parametric coefficients. Presence of ties observed through analytical precision

To begin with, thresholds were defined to distinguish between ties from non-ties. The three ways selected to assess the presence of ties agreed that a distance under 10% of the combined RSD of two consecutive values, which corresponds to a very small effect size (< 0.2), a t-value < 0.6 and a corresponding probability of rejecting the tie hypothesis (p-value) of less than 70%, could be seen as a confident tie (Table 4). On the contrary, a distance over 40% of the combined RSD, which corresponds to a large effect size (> 0.8), a t-value > 2.4 and a corresponding p-value $> 99\%$, was a confident non-tie. In between, the decision was less clear but it appeared that under 25% of the RSD, equivalent to a medium effect size (0.5), a t-value of 1.5 and a p-value of 90%, the points were probably tied, while over those values they were likely to be not. The thresholds were confirmed visually through scatterplots of the raw data against their respective ranks, constructed for 8 metabolites manually selected as representatively distributed over the data set. To take the uncertainty of measure into account, error bars were drawn that were equal to combined the RSD in the QC samples (Fig. 1). In addition, the signal axis was log-transformed when the data range exceeded 10, which was often the case. As observed by Anscombe (Anscombe, 1973), this allowed the lowest signals and their error bars to be visible in comparison to the highest ones. Overall, the thresholds made sense visually for our specific data set but were also consistent with the more standard notions of effect size, t-values and p-values, which tends to show that the procedure and the criteria are generalizable to other studies, while the thresholds values would likely have to be adapted, particularly as a function

Fig. 1 Visualization of the rank attribution in the form of the scatterplot (right) of the values and their uncertainties according to their ranks, along with the corresponding relative distances (table, left). In green, the relative distances were large enough ($> 40\%$ of the combined RSD) for confident rank attributions (confident non-ties); in red, the relative distances were small enough ($< 10\%$; confident ties); in orange, the cases were intermediate (between 10 and 25%, probable non-ties; between 25 and 40%, probable ties)
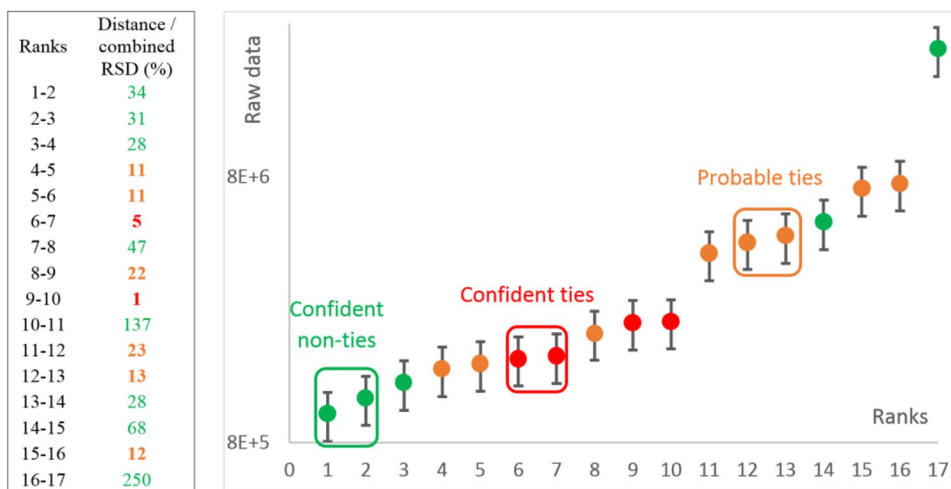
Table 5 Ties, groups of ties and numbers of points left in the data set

Number	Groups of ties		Ties		Points left	
	Confident	Probable	Confident	Probable	Confident	Probable
Mean	3.5	3.8	9.0	14.0	13.0	8.3
Median	3	4	9	14	13	9
%	Confident	Probable	Confident	Probable	Confident	Probable
Mean	19	21	49	75	70	45
Median	16	22	49	76	70	49

Mean and median aggregated numbers (above) and percentages (below) of groups of ties, ties and numbers of points left when considering the confident and probable ties

of the sample size and the dynamic range of the detector (SI S-4).

The application of the thresholds to the entire data set showed numerous ties and groups of ties (Table 5), with frequent 'chains' of ties difficult to interpret. The resultant number of points left to calculate the correlations was reduced by 30% and more than 50% with the confident and probable ties, respectively. As a result, the precision of the data appeared to be low in comparison to the distances between consecutive values to get confident rank attributions. Despite the care taken in the measurements and the partial correction of the peak volumes for the analytical variation measured in the QC samples through a locally estimated scatterplot smoothing (LOESS) procedure (Metsämuuronen, 2021a). This led to what we regard as an excessive attribution of the ranks that could impair the efficiency of the non-parametric correlation coefficients if this issue is not raised, which, again, depends on the specific case at hand through both the sample size and the dynamic range of the analytical method employed. On the other hand, the frequency of such a case in the GC×GC-TOFMS data set analyzed here, and the consecutive reduction of the number of points left to evaluate the correlations could also be a problem, which tends to favor the consideration of the confident ties against the probable ones.

The effect of the ties on the correlations was moderate in absolute values (0.03 mean and median variations of the coefficients due to the confident ties, 0.05–0.1 for the probable ones), especially in comparison to the changes induced in Pearson's r by the atypical values, but it was important in relative terms. Indeed, the median and mean variations were respectively around 10 and 40% when considering the confident ties and around 30 and 120% when considering the probable ones. The differential correlation coefficient Z_{CRC} was influenced by the ties in a similar manner, with median and mean variations both around 0.05–0.1 in absolute values and respectively around 20 and 80% in relative terms for the confident ties. For the probable ones, the median and mean variations were both around 0.3 and 0.7 in absolute values, representing around 40 and 200% median and mean

relative variations. Logically, the significant variations of correlations were also influenced by the ties: their number remained the same (but their identity changed) in the case of Spearman and it was increased with Goodman–Kruskal as well as with Kendall, but in a lesser extent. Detailed results for this entire section are given in SI S-5. The higher inflation of Goodman–Kruskal due to the ties could be explained by the fact that it explicitly takes them into account, through an adjustment of the sample size, while Kendall's tau leaves them aside from the correlation calculation. However, it is also argued that this phenomenon could be due to its implicit directional nature that automatically uses the narrower variable as the independent one (Metsämuuronen, 2021b; Walker et al., 2003). The detailed visual evaluation of the 8 selected correlations confirmed the effect of the ties but it was found somewhat limited, particularly on the regressions and with the confident ties, as illustrated in Fig. 2.

Regarding the coefficients, it comes from our results, detailed in SI S5 and S-6, that with no ties or the only confident ties, Spearman's rho was the most different coefficient of the three (the highest, with a mean rank of 1.1, in Table S17, panels A and C; see also the left part of Table S11 for the absolute values). However, when the probable ties were taken into account, Goodman–Kruskal's gamma caught up and Kendall's tau became the coefficient with the most different values (with the lowest mean rank of 2.9, panel C of Table S17). After transforming Kendall's tau and Goodman–Kruskal's gamma to be directly comparable to Pearson's r and Spearman's rho (Siqueira Santos et al., 2014) (right part of Table S11 and panels B and D of Table S17), they were much closer to Spearman at first (no tie) and then their inflation due to the ties, in addition to separate them from Spearman as well as from each other, made them the highest coefficients. Again, this affected particularly Goodman–Kruskal's gamma that became the most different coefficient (mean ranks of 1.1 and 1.0 with the confident and probable ties, panel D of Table S17). Those observations were confirmed by the differential correlation coefficient Z_{CRC} (Tables S14 for the absolute values and S18 for the differences between coefficients). Overall, Goodman–Kruskal
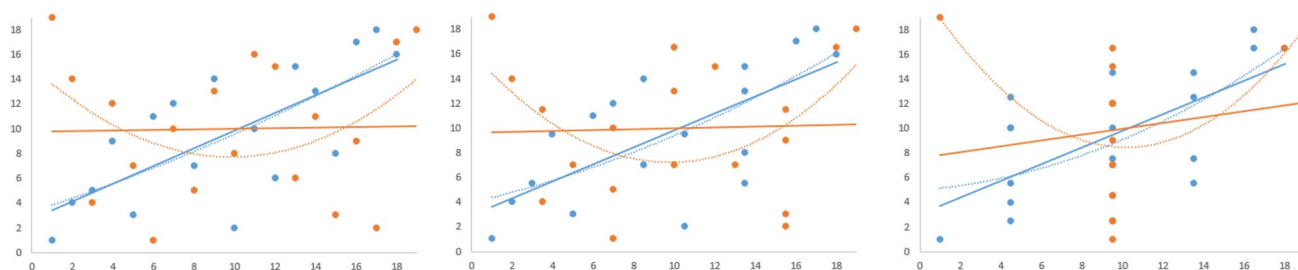

Fig. 2 Visual representation of the correlation between metabolites 78 and 83 with no consideration of the ties (left) and the consideration of the confident (middle) and probable ties (right). The HC samples and regressions are in orange and the CRC samples and regressions are in blue. The full lines represent the linear regressions, the dashed lines represent the polynomial ones

was found the most liberal coefficient, and Spearman the most conservative, both by quite far, especially with the ties. Globally, the mean and median correlation coefficients r and differential Z_{CRC} were relatively low, with respective values of 0.2—0.3 (up to 0.4 for Goodman–Kruskal after transformation and the consideration of the ties, Table S11) and 0.7 (up to 1 and more for Goodman–Kruskal when considering the ties, Table S14). The 75th percentiles were moderate, respectively around 0.4 (up to 0.6 for G-K) and a bit more than 1 (up to 1.8 for G- K). While the comparison between the non-parametric coefficients will be continued in the next section, it already comes from this investigation of the ties and their effects that, in order to take them into account without risking to artificially change the results, it seems a good compromise to calculate those coefficients with only the confident ties. To conclude on this section, when using a non-parametric correlation coefficient, it appears important to consider the presence of possible ties that could bias the results. To do so, the proposed strategy based on the definition and application of specific thresholds for three parameters able to highlight such ties, and the evaluation of their effects on the data and the various coefficients worked well on our GC×GC-TOF–MS data set and should be generalizable, given that the results are properly adjusted to the specific case considered.

2.5 Determination of the appropriate coefficient

To find the appropriate coefficient to perform the correlation analysis, here on our untargeted metabolomics GC×GC-TOFMS data set, the methodology consisted in comparing Pearson's r calculated on the log-transformed data, without the outlier samples and values previously determined, to the non-parametric coefficients applied with the consideration of the confident ties. We first assessed the aggregated (5151 correlations) absolute values for both the correlation coefficients and their differential Z_{CRC}. The first ones and their respective ranks were quite similar for the four coefficients. Spearman's rho and Goodman–Kruskal's gamma were the most different coefficients with respectively the lowest and highest values (0.28 and 0.34, mean ranks of 1.5 and 3.2; Table 6). The aggregated absolute differences between the coefficients, however, showed that the parametric one, Pearson's r, was the most different. Thus, the effect of the confident ties to separate the non-parametric measures from each other was not sufficient in comparison to the difference between parametric and non-parametric coefficients. Kendall's tau, despite being very close to Goodman–Kruskal's gamma, was the most central coefficient, with no maximum and only 11% of minimum values.

Those observations were confirmed, and even amplified, in the aggregated differential correlations Z_{CRC}, particularly regarding the difference between Pearson's r and the non-parametric coefficients (Table 7). The number of significant variations of correlations led to the same conclusions, whatever the threshold used, with a ranking of the coefficients from the more liberal to the more conservative: Goodman–Kruskal's gamma > Kendall's tau > Pearson's r > Spearman's rho (Table 8). Regarding the thresholds for significance, the 0.05 and 0.01 p-values, used in the absence

Table 6 Comparison of the correlation coefficients

Distribution	P	G-K	K	Sp
0.25	0.13	0.15	0.14	0.12
Mean	0.30	0.34	0.32	0.28
Median	0.27	0.31	0.29	0.25
0.75	0.44	0.50	0.47	0.42
Max	1.00	1.00	0.99	0.98
Mean rank	2.7	1.5	2.6	3.2
Median rank	3	1	2	3
% of Max	35	60	0	5
% of Min	43	0	11	45

Distribution	P/G-K	P/K	P/Sp	G-K/K	G-K/Sp	K/Sp
0.25	0.04	0.03	0.03	0.01	0.02	0.02
Mean	0.10	0.09	0.08	0.02	0.06	0.04
Median	0.08	0.07	0.06	0.01	0.05	0.04
0.75	0.14	0.13	0.12	0.02	0.08	0.06
Max	0.86	0.87	0.91	0.12	0.25	0.18

Aggregated absolute values, ranks and absolute differences between the correlation coefficients, for all 5151 correlations. Goodman–Kruskal's gamma (G-K) and Kendall's tau (K) were transformed to be directly comparable to Pearson's r (P) and Spearman's rho (Sp) (Siqueira Santos et al., 2014)

Table 7 Comparison of the differential correlation coefficients Z_{CRC}

Distribution	P		G-K		K		Sp
0.25	0.31		0.36		0.34		0.29
Mean	0.80		0.92		0.87		0.76
Median	0.66		0.77		0.73		0.64
0.75	1.16		1.33		1.26		1.09
Max	8.64		5.13		4.77		4.18
Distribution of the diff	**P**		**G-K**		**K**		**Sp**
Mean rank	2.7		2.5		2.4		2.4
Median rank	3		3		2		2
Number of Max	1836		1547		394		1374
% of Max	36		30		8		27
Number of Min	2559		1436		256		900
% of Min	50		28		5		17
Distribution of the diff	**P/G-K**	**P/K**	**P/Sp**	**G-K/K**	**G-K/Sp**	**K/Sp**
0.25	0.19	0.17	0.15	0.02	0.08	0.06
Mean	0.48	0.45	0.41	0.06	0.21	0.16
Median	0.39	0.36	0.33	0.04	0.17	0.13
0.75	0.68	0.64	0.57	0.08	0.29	0.23
Max	10.63	10.42	10.16	1.05	1.55	0.86

Aggregated absolute values, ranks and absolute differences between the differential correlation coefficients, for all 5151 correlations. Pearson's r (P), Goodman–Kruskal's gamma (G-K), Kendall's tau (K) and Spearman's rho (Sp)

Table 8 Comparison of the significant differential correlation coefficients Z_{CRC}

Threshold uncorrected α	P	G-K	K	Sp
5.10^{-2}	284	492	397	221
10^{-2}	59	139	104	47
10^{-3}	9	38	17	5
10^{-4}	1	7	3	1
Bonferroni	1	3	1	0
BY	1	1	0	0

Numbers of significant variations of differential correlations coefficients Z between CRC and R-CRC samples, depending on the threshold considered, for all 5151 correlations in the data set. Pearson's r (P), Goodman–Kruskal's gamma (G-K), Kendall's tau (K) and Spearman's rho (Sp)

of appropriate direct thresholds for the correlation coefficients (Rousselet & Pernet, 2012), led to dozens and even hundreds (with 0.05 and/or ties) of significant results. However, when Bonferroni and Benjamini-Yekutieli corrections for multiple testing were introduced (0.05 p-value threshold) (Pernet et al., 2013; Schwarzkopf et al., 2012), very few results were left for further exploitation (Table 8). Given the exploratory nature of the procedure, and the fact that the true significance can only be achieved through replication (Chok, 2010; Wilcox & Rousselet, 2018), we found most

appropriate to compromise between the quality (necessary with low sample sizes (Camacho et al., 2005)) and the quantity of the results (Pernet et al., 2013) through an uncorrected p-value threshold of 0.001.

Given the recognized capacity of non-parametric correlation coefficients to be more robust to atypical observations (and to provide higher statistical power in their presence (Hazra & Gogtay, 2016; Winter et al., 2016)), to non-linearity and to non-normal, sparse or small data sets (Walker et al., 2003), they were expected to give a more accurate assessment of the differential correlations.

This was confirmed with the 29 selected correlations, through their numerical comparison and their visual control. It came out that all coefficients, used taking into account the outliers or the ties, gave similar results -with Pearson's r confirmed as the most different (Fig. 3)- and seemed to correctly represent numerically the visual distributions and correlations. However, Goodman–Kruskal's gamma tended to inflate a bit the correlations, on the contrary to Pearson's r and Spearman's rho who appeared sometimes to underestimate them (Table 9). Based on those observations, as well as on the previous analysis conducted on all correlations where it was found liberal but not as much as Goodman–Kruskal's gamma, Kendall's tau seemed the method of choice to perform an untargeted exploratory correlation study on our GC×GC-TOF MS metabolomics data. This is in agreement with the usual

Fig. 3 Scatterplot of the differential correlations Z_{CRC} of the 29 representative correlations selected for the four coefficients tested. Pearson's r values are in blue, Spearman's rho in green, Kendall's tau in yellow, and Goodman–Kruskal's gamma in orange
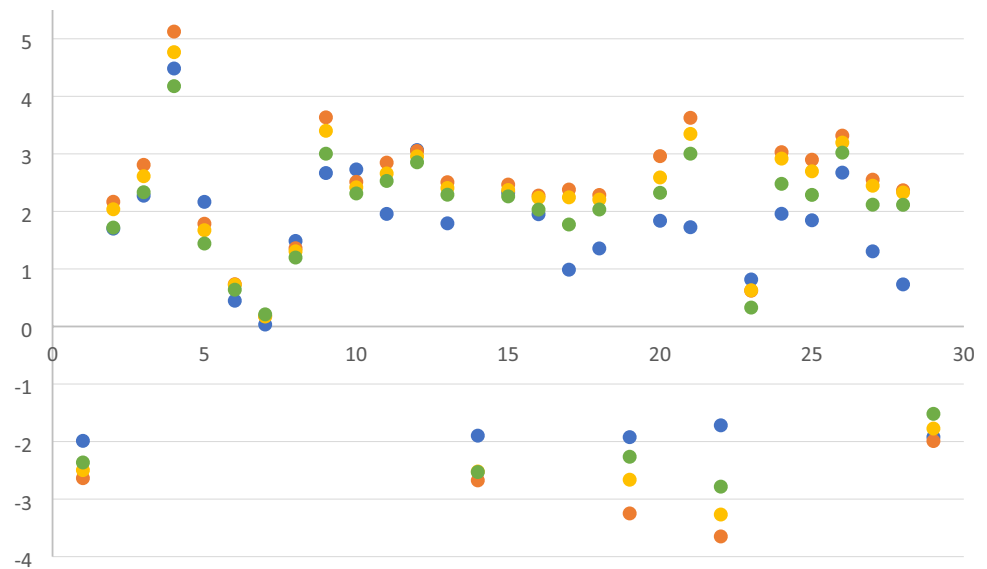
preference for robust algorithms in the case of atypical data, particularly with small sample sizes and unknown distributions (Armstrong, 2019; Chok, 2010; Dunn et al., 2013; Rousselet & Pernet, 2012; Siska & Kechris, 2017). However, this is in disagreement with the observation that, because of the way it handles them, Goodman–Kruskal's gamma is generally preferred to Spearman or Kendall for data sets with many tied ranks (Walker et al., 2003). This is possibly because we decided to use only the confident ties in the calculation of the correlations. In addition to the strategy that consisted in comparing the coefficients in multiple ways (aggregated values, significant results and representative cases) and revealed appropriate, the scatterplots were found very useful from a methodological point of view, confirming previous recommendations (Anscombe, 1973; Schwarzkopf et al., 2012), despite the small data sizes available leading to incomplete distributions. The use of regressions helped a lot to determine the tendencies in the data (SI S-7). Especially with moderate correlations (between 0.3 and 0.6), of which the significance was more difficult to evaluate, as well as when comparing correlations with ranges and scales that differed between the subgroups studied. Most frequently, we observed quite clear linear behaviors that were well modeled by linear regressions, which made the polynomial ones superfluous. However, the polynomial approximation, limited to the second order to avoid overfitting, could be compared to the linear regression to inform, through its curvature, about the linearity of the data. It was also able to model alternatives trends present in the subgroups of the data, again through its curvature as well as through the length of the curve on each side of it. In addition to the log-transformed data used in the calculations, the graphs included the raw data for comparison purpose. The main observation was

a logical increase in linearity of the distributions through the reduction of the distances between the points.

2.6 Specific significant variations of correlations associated to colorectal cancer

Using the 0.001 p-value threshold chosen above led to the selection of 17 correlations significantly altered in the "active" state of colorectal cancer (CRC samples) in comparison to the gender and age matched healthy controls (HC samples; Table 10). Among those, 10 were found specific to the active state in comparison to the CRC samples in remission (R-CRC).

Using the mass spectra, the linear retention indices (LRI) and the exact mass led to the confident annotation (MSI level 2) of 15 molecules on the 25 unique ones involved in the correlations of interest (Table 11), providing 3 correlations with two annotated metabolites, 11 with one metabolite annotated and 3 with no metabolites annotated. All were found to have biological functions. Since the annotation of a correlation requires both metabolites to be annotated, the issue of identifying the metabolites of interest is even more crucial than in biomarker research (Bingol et al., 2016; Chaleckis et al., 2019; Giovanni et al., 2020b; Nash & Dunn, 2019) and therefore constitutes a major bottleneck. This is illustrated in the network visualization, where the two main metabolic hubs highlighted (metabolites 45 and 77 in the data set) could not be reliably annotated (Fig. 4). Therefore, there was limited interest to perform a specific network analysis (Weckwerth & Morgenthal, 2005) (such as hubs, connectivity, centrality parameters; Fig. 4).

Most publications about the CRC molecular aspects studied the subtypes of CRC as well as the CRC-initiating or tumorigenesis cellular pathways (Dai et al., 2019;

Table 9 Differential correlations Z_{CRC} of the 29 selected correlations for the four coefficients tested

Differential correlations Z_{CRC} of the 29 selected correlations for the four coefficients tested.

Variable	ID number	P	G-K	K	Sp
1	10	-2.0	-2.6	-2.5	-2.4
1	69	1.7	2.2	2.0	1.7
5	68	2.3	2.8	2.6	2.3
5	78	4.5	5.1	4.8	4.2
6	25	2.2	1.8	1.7	1.4
7	48	0.4	0.7	0.7	0.6
7	62	0.0	0.2	0.2	0.2
7	69	1.5	1.4	1.3	1.2
14	51	2.7	3.6	3.4	3.0
14	57	2.7	2.5	2.4	2.3
14	85	2.0	2.8	2.7	2.5
18	77	3.1	3.1	3.0	2.9
19	77	1.8	2.5	2.4	2.3
20	22	-1.9	-2.7	-2.5	-2.5
20	68	2.3	2.5	2.4	2.3
22	74	2.0	2.3	2.2	2.0
22	82	1.0	2.4	2.2	1.8
25	77	1.4	2.3	2.2	2.0
28	93	-1.9	-3.2	-2.7	-2.3
29	98	1.8	3.0	2.6	2.3
54	55	1.7	3.6	3.3	3.0
58	90	-1.7	-3.6	-3.3	-2.8
62	95	0.8	0.6	0.6	0.3
76	96	2.0	3.0	2.9	2.5
76	98	1.8	2.9	2.7	2.3
77	101	2.7	3.3	3.2	3.0
78	83	1.3	2.6	2.4	2.1
88	96	0.7	2.4	2.3	2.1
89	91	-1.9	-2.0	-1.8	-1.5

The values which absolute difference with the nearest other coefficient is ≥ 0.5 and 1, the empirical thresholds chosen to represent the medium and large differences, are respectively in light grey and grey. Pearson's r (P), Goodman–Kruskal's gamma (G-K), Kendall's tau (K) and Spearman's rho (Sp)

Dienstmann et al., 2017; Vecchia & Sebastián, 2020; Wan et al., 2020; Wang et al., 2019; Warburg, 1956). However, alterations of the metabolic pathways have been observed in various matrices (serum, tissues, urine, feces and breath) that can be informatively compared to the ones observed in the study (Table 12 and in bold below). They include: cell energetic metabolism according to the Warburg effect (Qiu et al., 2009) (glycolysis and TCA cycle), urea cycle, structural maintenance through glycerol and ketone bodies, oxidative stress, cytochrome P450 activity and lipids, amino acids, fatty acids, bile acids and nucleotides metabolisms (Eylem et al., 2020; Monedeiro et al., 2020; Qiu et al., 2010; Serra et al., 2018; Seyfried & Shelton, 2010; Tan et al., 2013; Zhang et al., 2017; Zhu et al., 2014) (including multiple potential contributions of the microbiota (Wan et al., 2020)). The metabolic pathways highlighted here suffer from

Table 10 Differential correlations Z and associated p-values for the indirect comparison between CRC and R-CRC samples

Variable ID numbers		Z_{CRC}	p-value	Z_{R-CRC}	$Z_{CRC/R-CRC}$	p-value
5	**78**	4.8	0.000002	0.7	4.1	**0.00004**
45	**57**	−4.3	0.00002	−0.5	−3.8	**0.0001**
77	**98**	4.2	0.00003	0.6	3.6	**0.0003**
45	**90**	−3.8	0.0001	−0.1	−3.8	**0.0002**
7	77	3.8	0.0002	0.9	2.9	0.004
20	95	3.7	0.0002	0.9	2.8	0.005
8	**43**	3.6	0.0003	−0.6	4.2	**0.00003**
44	92	−3.5	0.0004	−1.4	−2.1	0.033
3	43	−3.5	0.0004	−1.1	−2.4	0.014
14	51	3.4	0.0007	1.9	1.5	0.14
54	**55**	3.3	0.0008	−0.9	4.3	**0.00002**
45	**70**	−3.3	0.0008	−0.1	−3.2	**0.001**
6	14	3.3	0.0009	0.7	2.6	0.009
45	**51**	−3.3	0.0009	−0.1	−3.2	**0.001**
29	77	3.3	0.0010	0.6	2.7	0.008
56	**98**	3.3	0.0010	−1.1	4.4	**0.00001**
18	**71**	3.3	0.0010	−2.9	6.2	**< 0.00001**

The differential Z between CRC and R-CRC, $Z_{CRC/R-CRC,}$ is the difference between the two differential Z obtained individually against the gender and age matched healthy controls, Z_{CRC} and $\mathbf{Z}_{R-CRC}$. In bold, the variations of correlations specific to CRC

two weaknesses. First, they are only very weakly enriched, except for aminoacyl-tRNA biosynthesis. Second, they are rather non-specific and seem to be only indirect consequences of colorectal cancer.

2.7 Limitations

Just like the candidate biomarkers, the correlations of interest are only primary results. They require to be properly confirmed through a targeted quantitative analysis of the variations observed, followed by a proper validation conducted on an independent, larger set of patients, to see if they translate to different groups of patients with the same significant distributions. After that, further biological investigation would have to be performed, that ideally would lead to a better understanding of the metabolic effects of the disease process and its remission state detectable in serum samples. In this perspective, a powerful but complex approach would be to integrate other omics data obtained on the specific biological samples analysed for this study.

Indeed, besides the risk of overfitting and the annotation of the metabolites mentioned above, another recurrent limitation in metabolomics that was highlighted here is the analytical stability. Indeed, the generation of quality data requires high accuracy (Weckwerth & Morgenthal, 2005) and therefore suffers from the intrinsic noise present in high-dimensional data sets (Dias & Koal, 2016). In addition to its effect on the non-parametric coefficient investigated above, it reduced the number of signals available to

discover candidate biomarkers and correlations of interest. Missing values, besides the problem they can be for correlation calculation by producing incomplete profiles difficult to process (an issue that had no particular consequence in this work), are another important cause of the loss of potentially interesting signals. Here, despite the care given to the chromatography, the implementation of an external QC system, coupled to a LOESS partial correction, and the replacement of all missing values by the half of the lowest signal measured for the specific metabolite, the metabolic coverage was decreased by the selection of only the stable metabolites (RSD in the QC samples under 30% (Kotze et al., 2013)) and the ones present in at least half the samples of any class. Therefore, through the various preprocessing steps (summarized in SI S-8), we went from 646 quality features in the chromatographic template to 102 variables corresponding to metabolites in the final data set, confirming a previous study performed on Crohn's disease samples where 524 quality features became 183 high-quality metabolites. If it obviously downgrades any biomarker research where it linearly decreases the number of potential candidate biomarkers, it harms even more a correlation analysis where every additional variable can have potentially interesting links with every other and where, therefore, the number of correlations available is proportional to the number of metabolites at the power 2. However, such a selection seems necessary in order to get confident results, as it was shown how even lower analytical variations have a clear

Table 11 Annotation of the metabolites involved in the correlations of interest

Data set	Metabolite	Molecular formula	Match factor	Probability	Theoretical LRI	Mesured LRI	Delta LRI	Exact mass		Number of criteria met
3	3-Hydroxybutyric acid	$C_4H_8O_3$	**793**	**49.5**	1167	1164	**3**	**0.8**	1.4	3
5	Serine	$C_3H_7NO_3$	**821**	**85.0**	1368–1388	1372	**6**	**0.5**	**0.0**	3
8	Threonine	$C_4H_9NO_3$	**877**	**91.8**	1375–1400	1399	**12**	**0.2**	**0.3**	3
14	Aminomalonic acid	$C_4H_4NO_4$	**819**	**96.0**	1485	1484	**1**	**0.1**	**0.0**	3
18	Malic acid	$C_4H_6O_5$	**727**	47.9	1478	1503	**25**	**0.3**		3
29	Phenylalanine	$C_9H_{11}NO_2$	**704**	**53.5**	1636	1631	**5**	1.7	2.3	2*
44	Glycylglycine	$C_4H_8N_2O_3$	**731**	4.0	1818	1824	**6**	**0.2**		3
54	Quinic acid	$C_7H_{12}O_6$	**661**	**69.4**	1851–1918	1897	**13**	**0.6**	**0.7**	3
56	Tyrosine	$C_9H_{11}NO_3$	**876**	**80.3**	1959	1957	**2**	**0.7**		3
57	Pentadecanoic acid	$C_{15}H_{30}O_2$	**658**	**90.2**	1950	1951	**1**	**0.3**		3
70	Uric acid	$C_5H_4N_4O_3$	**745**	**60.7**	2136	2140	**4**	4.0	4.4	2*
71	Heptadecanoic acid	$C_{17}H_{34}O_2$	**647**	**76.6**	2146	2147	**1**	2.8	3.3	2*
78	Uridine	$C_9H_{12}N_2O_6$	**771**	**89.5**	2375	2381	**6**	**0.1**		3
95	Erythritol / Threitol	$C_4H_{10}O_4$	**810**	43.9	1510	1525	**15**	**0.0**	**0.4**	3
98	Norleucine	$C_9H_{12}NO_2$	**791**	11.0	1313	1312—1328	**7**	**0.2**		3

The annotation was made through full mass spectrum (library match and probability of correct annotation), LRIs (delta between the measured and theoretical values) and exact mass (mass error in ppm). In bold, the values that met the criteria defined

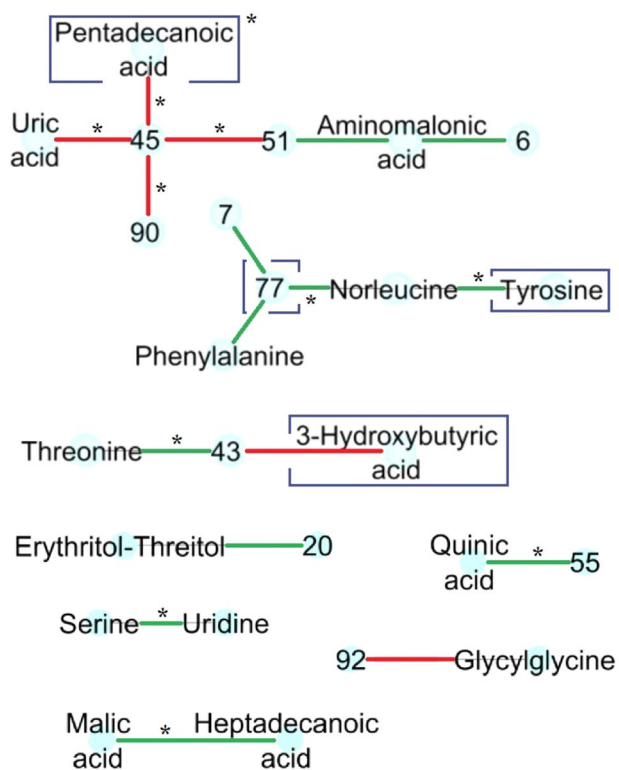*Indicates that the exact mass criterion was made less strict, up to 5 ppm

Fig. 4 Network representation of the correlations positively (green) and negatively (red) altered between the CRC and their matched HC samples. The candidate biomarkers of CRC vs HC previously highlighted (Rodriguez-Salas et al., 2017) are framed in blue. The correlations that were found specific to the active state of colorectal cancer (CRC samples) in comparison to the remission state (R-CRC samples) are indicated by an asterisk

the number of significant variations of correlations was limited with the threshold applied, it could nevertheless be sufficient to provide interesting metabolic insights and associations with colorectal cancer alterations and regulations. Because of those limitations, only a small number of annotated correlations of interest were obtained and thus a very small network. The last, but not the least, limitation of our study is the small sample size we used ($17 \leq n \leq 19$), due to clinical constraint (Bujang & Baharum, 2016).

Among the several potential effects mentioned in SI S-9, the reduced statistical power and the inflation of the correlation values are the most dangerous for us. Here, we investigated not only correlations but also variations of correlation (Z_{CRC} or Z_{R-CRC}) and differences in the variations of correlation between different states of the pathology ($Z_{CRC}/_{R-CRC}$). This makes difficult to evaluate the power as well as the metabolic effect sizes of interest and their significance, particularly prior to the study. Indeed, the size of the variations of interest not only depends on the specific metabolites involved, like the simple correlations, but also on both the initial and final correlation values r, as investigated for Pearson's r by Bujang and Baharum, 2016 (Hulley et al., 2013). Power calculation applied to our study shows that with $n \geq 17$, the minimal correlation value detectable with a power of 80% at a significance level of 0.05 is around 0.6 (Yarkoni, 2009), which is already a large correlation. Regarding the variations of correlations, we observed that the significant ones always involved a low or medium correlation and a very high one (in absolute value). While the very high correlations are much likely inflated by the low sample size (Rakotomalala, 2005), the statistical power can nevertheless be estimated by calculating the sample size necessary to detect the low or medium values. With a significance threshold (type I error) of 0.05 the power to detect the variations of correlations significant at 0.05, 0.01, 0.001 and 0.0001 are respectively of 32, 38, 63 and 90%.

effect on the non-parametric correlation estimations. This drawback is shared by all analytical platforms, in various extents, and no single instrumentation is able to provide a complete and fully stable coverage of a complex sample. Because of the exploratory nature of the analysis, if

Table 12 Metabolic pathways most altered when considering the highlighted candidate biomarkers and correlations of interest

CB/Corr	Pathway name	Match status
CB + Corr	Aminoacyl-tRNA biosynthesis	5/48
CB	Alanine, aspartate and glutamate metabolism	3/28
CB + Corr*	Phenylalanine, tyrosine and tryptophan biosynthesis	2/4
CB + Corr*	Phenylalanine metabolism	2/10
CB + Corr*	Butanoate metabolism	2/15
CB	Glycolysis/Gluconeogenesis	2/26
CB	Glyoxylate and dicarboxylate metabolism	2/32
CB + Corr	Purine metabolism	2/65

CB + Corr* refers to the case where the same metabolite(s) was (were) highlighted in the biomarker research and the correlation analysis; CB + Corr means that different metabolites were highlighted in the two processes that are involved in the same pathway. The match status is the number of metabolites highlighted through the biomarker research and the correlation analysis over the total number of metabolites present in the pathway

3 Materials and methods

3.1 Samples

Patient recruitment and serum sample intakes were performed at the university hospital of Liège, through the Bibliothèque hospitalière universitaire de Liège (BHUL), Belgium. The sample intake, processing and storage procedures were standardized and followed our biobanking guidelines developed for proteomics studies and utilized for clinical trials (Metsämuuronen, 2021a). We analyzed serum samples from patients with colorectal cancer (CRC; n = 18) and with colorectal cancer in remission (R-CRC; n = 17) as well as their respective controls (HC and R-HC) matched for gender and age (n = 19 and n = 17). Based on endoscopic examination confirmed on surgical resection specimen, patients with confirmed cancer lesions (primary adenocarcinoma according to the Tumor Node Metastasis (TNM) staging system, CRC group) or with a previous history and no evidence of remaining disease of cancer (R-CRC group) were included. The ("healthy") control groups consisted of patients screened by colonoscopy and diagnosed negative for CRC (no visible lesion or hyperplastic polyp) as well as for any other pathology affecting the bowel (as inflammatory bowel disease, diverticulitis…) or any other known cancer at the time of sample intake. The internal QC samples consisted in 30 µL aliquots of the study samples, made through one freeze–thaw cycle.

3.2 Data acquisition and processing

The clinical data of the patients as well as the chemicals used and all the steps leading to the data set used in this study (sample preparation, GC×GC-HRTOF-MS analysis, data (pre)processing, annotation of the compounds of interest) have been described in Giovanni et al., (2020); Rodriguez-Salas et al., (2017) and a summary can be found in the Supplementary Materials (SI S-10). All calculations and plots were made in Microsoft® Excel® except for the correlation coefficients that were calculated using the Excel® add-in Tanagra (Xia et al., 2015) and the HCA, PCA, PLS and sPLS plots that were constructed using the web-resource MetaboAnalyst (Shannon et al., 2003). The correlation networks were drawn in Cytoscape® software (Tufte, 2001).

3.3 Correlations analysis

In a previous study, we looked for specific candidate biomarkers of the 'active state' of colorectal cancer (CRC) in comparison to healthy controls (HC) and patients in remission (R-CRC) (Rodriguez-Salas et al., 2017). As a result, in

the data set of 102 metabolites, 24 metabolites were found significantly altered and able to discriminate the CRC and HC samples: Receiver Operating Characteristic (ROC) area under the curve (AUC) of 0.86, sensitivity and specificity of 0.72 and 0.78. Ten of those were found to have signals close to healthy levels also in the R-CRC samples and were therefore potentially specific to the CRC samples analysed. Here, in order to go further, take advantage of their above-mentioned benefits and complement the previous study, we aimed to investigate in a similar manner the correlations between metabolites, on the same data set. To do so, we first looked for the most appropriate way to calculate them, given the type and structure of the data at hand. Following Anscombe's quartet (Anscombe, 1973; Ark & Aert, 2015), we identified four main points to take into account: the presence of non-linear relationships, the role and handling of atypical samples and values, the effect of the analytical and biological noise in the data and the need for a visual control of the detected associations. Therefore, four types of correlation parameters were examined: Pearson's r (parametric) as well as Spearman's rho (Pearson's r on the ranks), Kendall's tau and Goodman–Kruskal's gamma (non-parametric). The two later belong to the same family of coefficients, Kendall's tau being the parent measure to which the others reduce in the absence of ties, and can be interpreted as the ranked pairs in agreement between two variables (Gorrie, 2020; Metsämuuronen, 2021b). The other coefficient of the family, Somers' D, was contemplated at first but finally left aside since it treats the variables asymmetrically, with a dependent variable and an independent one, producing two results for each correlation. To be useful, it would require to completely annotate the molecules and to determine their biological hierarchy prior to the correlation analysis, which was not the case in this exploratory work. For the same reason, partial correlations, which were also considered, were finally left aside. Indeed, without the proper prior biological knowledge, the determination of the influences between the metabolites would be purely mathematical, which is particularly dangerous with low sample sizes and non-negligible analytical errors. Besides, the use of partial correlations would also have limited the use of the visual control tools that are an important part of the protocol proposed (Anscombe's quartet (Anscombe, 1973)).

3.4 Parametric coefficient. Presence and effect of atypical samples and values

For Pearson's r, we assessed the influence of the data structure on the results from both the raw and log-transformed data (Armstrong, 2019; Weckwerth & Morgenthal, 2005), the one that was used in the biomarker research (Rodriguez-Salas et al., 2017). Both atypical samples and values can be a serious practical problem for correlation calculation

(Wilcox, 2004; Wilcox & Rousselet, 2018) and produce misleading results (Anscombe, 1973; Pernet et al., 2013). The appropriateness of the Pearson's correlation depends more on the presence of outliers in the data than on shape of the empirical or theoretical distribution (Winter et al., 2016). The atypical samples were investigated through hierarchical clustering analysis (HCA), principal component analysis (PCA) and partial least squares (PLS) plots constructed on all metabolites as well as through the aggregation of several parameters: mean and median ranks, mean and median robust z-scores, proportion of robust z-scores > 1, 2 and 3.5 (Grubbs & Beck, 1972; Leys et al., 2013; Miller & Miller, 2010). The atypical values were studied through robust versions of Grubb's and Dixon's tests (Dekking et al., 2005; Rorabacher, 1991), as well as boxplots and scatterplots (Anscombe, 1973). Two types of atypical samples and values were distinguished, with the aim to differenciate the clear outliers (to be reconsidered for inclusion in the data set) from the dissimilar samples. The *outliers* were defined as the samples and values well outside their respective defined thresholds (95% ellipses for the PCA, PLS and sparse partial least squares (sPLS) plots; high dissimilarity for HCA, 95% probability values for Grubbs and Dixon's tests; 1.5 IQR limit (Iglewicz et al., 1993) for the boxplots; 1, 2 and 3.5 (Dunn et al., 2011; Grubbs & Beck, 1972) for the z-scores (see SI S-11 for details)) and that were not aligned, monotonically, in the scatterplots. The *dissimilar* samples and values were defined as the ones inside but close to the thresholds and distant in the scatterplots so that they could affect the calculation of the correlations. We then investigated the effects (Wilcox, 2004) of those atypical samples and values on the correlation coefficients as well as on the variations of correlations Z_{CRC} and Z_{R-CRC} defined as the difference between the correlation coefficients adjusted for the global variance (Rodriguez-Salas et al., 2017) (SI S-12), between the serum samples from patients with colorectal cancer (CRC) or from patients in remission (R-CRC) and their matched "healthy" controls (HC and R-HC), for all 5151 correlations existing between the 102 variables corresponding to metabolites of the data set.

3.5 Non-parametric coefficients. Presence and effect of ties

For the non-parametric coefficients, the frequent proximity of consecutive values in the GC×GC-TOF–MS data set relatively to the uncertainty of measure, evaluated as the residual standard deviation in the QC samples (RSD), led us to take the influence of the precision of the data on the assignment of the ranks into account. Indeed, the QC system we implemented only partially corrects for the analytical variations that affect the data and since the selection of the stable metabolites allows their residual RSD to go up to

30%, which is the conventional threshold used for GC–MS (Borenstein et al., 2009). Therefore, the attribution of the ranks can be somewhat misleading since it sets equal distances between the values whatever their differences in terms of signal measured in the data set and whatever their uncertainty. This could lead to false positive and negative results when looking for the significant (variations of) correlations. To solve this issue, we first investigated the presence of ties through three different aspects, for each pair of consecutive values. First, the ratio between their absolute difference and their combined residual standard deviation measured on the QC samples (Borenstein et al., 2009; Metsämuuronen, 2021a) (see SI S-13 for the formulas). Second, the Hedges'g effect size (the corrected Cohen's d for sample sizes < 20) (Rodriguez-Salas et al., 2017; Student., 1908). Third, the one-tailed t-tests (Broadhurst & Kell, 2006) corresponding to the probability of correct ranking. Then, the effect of the ties on the correlation coefficients and on the variations of correlation Z_{CRC}, between CRC and HC groups of samples, was investigated, after proper transformation for Kendall's tau and Goodman–Kruskal's gamma (Rodriguez-Salas et al., 2017; Rousselet & Pernet, 2012; Siqueira Santos et al., 2014; Walker et al., 2003).

3.6 Selection and application of the appropriate correlation coefficient

After that, considering the atypical values and the ties and knowing that no single method is optimal in all situations (Pernet et al., 2013; Wilcox & Rousselet, 2018), the four coefficients were compared in three different ways in order to determine the method of choice to perform correlation analysis on the untargeted metabolic GC×GC TOF–MS data at hand. First, through their aggregated absolute values, differences and ranks for all 5151 correlations present in the data set. Second, through the visual monitoring of the scatterplots established for 29 correlations selected to be representative of the data set (detailed in Table 9). Third, by looking at the significant variations of correlation Z_{CRC} using multiple thresholds (p-value obtained under the assumption of normality of Z_{CRC}, Bonferroni correction (Benjamini & Yekutieli, 2001) and Benjamini-Yekutely false discovery rate (FDR) procedure (Hyötyläinen, 2010) applied because of the possible dependency of the metabolic signals measured).

Finally, the selected method was applied with the aim to determine the significantly and specifically altered correlations in the active state of colorectal cancer (CRC samples), in the same way that it was done for the specific candidate biomarkers of CRC in a previous study (Rodriguez-Salas et al., 2017). To do so, the significant variations of correlations Z_{CRC} between the CRC and their age and gender matched healthy controls (HC) were highlighted. Then,

their specificity to the active state in regard to the remission state (R-CRC) measured on distinct samples was assessed by comparing them to the variations of correlation Z_{R-CRC} measured between the R-CRC and their distinct age and gender matched healthy controls (R-HC). This comparison, indirect to avoid any bias due to the fact that the CRC and R-CRC samples were not matched, was done through the absolute difference between the two respective Z_{CRC} and Z_{R-CRC} (Kotze et al., 2013) that led to $Z_{CRC/R-CRC}$ as well as through the associated p-values obtained under the assumption of normality. The metabolites involved in the correlations of interest were identified using full mass spectra, exact mass and linear retention indices (LRI) (Bingol et al., 2016; Wishart et al., 2018), according to a procedure described in a previous publication (Rodriguez-Salas et al., 2017) that uses the following acceptance thresholds: match factor of $> 700^{23}$ or > 600 with a probability of $> 50\%$, $\Delta LRI < 25^{80}$ and mass error < 1 ppm for any specific fragment. Then, the biological plausibility of the annotated candidates was assessed with the Human Metabolome Database (HMDB (Kanehisa et al., 2017)) and the Kyoto Encyclopaedia of Genes and Genomes (KEGG (Motulsky et al., 2019)) in order to avoid analytical artefacts. After that, a literature review of the metabolic pathways potentially altered by colorectal cancer was conducted to see if, among the annotated candidates, some had already been associated to and would therefore be likely to play a role in CRC. Finally, a correlation network was drawn with the significantly (and specifically) altered variations along with the significantly (and specifically) altered metabolites (candidate biomarkers) already highlighted (Rodriguez-Salas et al., 2017). Importantly, further biological interpretation and exploration of the results are beyond the scope of the present paper that focuses on the methodological aspects of correlation analysis. Ideally, they should include a targeted confirmation of the variations observed, conducted on an independent set of patients of a much larger size. A powerful approach would be to integrate other omics data obtained on these specific biological samples.

4 Conclusions

This study confirmed the interest of conducting correlation analysis in addition to the more frequent biomarker research in order to increase the metabolic information obtained about the changes induced by the phenomenon investigated. To do so, the issues raised by Anscombe's quartet are of primary importance. Therefore, a global strategy has been proposed to cover them, which methodological aspects were systematically examined in order to allow the interested reader to implement it to its own specific case. The strategy includes descriptive visual tools such as scatterplots that were found necessary (Anscombe, 1973; Hardin et al., 2007;

Schwarzkopf et al., 2012) to study the relationships between metabolites, either with parametric or non-parametric coefficients, in order to overcome the limitations of purely technical calculations, namely to assimilate the data in all their complexity [86], to detect the trends and irregularities and to control the summarized statistical values (Armstrong, 2019). If Kendall's tau was found the most appropriate coefficient to use with our own data obtained from serum samples analyzed with GC×GC-TOFMS, the other ones tested were efficient as well to translate numerically the visual tendencies of associations and should, in our view, be considered.

Authors Contribution Conceptualization, NDG, MAM, EL and JFF; methodology, NDG; formal analysis, NDG; investigation, NDG; data curation, NDG; writing—original draft preparation, NDG; writing—review and editing, MAM, EL and JFF; supervision, MAM, EL and JFF; project administration (patients enrolment, inclusion and sample collection, ethics), MAM, EL and JFF; funding acquisition, EL, MAM and JFF. All authors have read and agreed to the published version of the manuscript.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the University of Liège (B707201213737).

Informed consent statement Informed consent was obtained from all subjects involved in the study.

References

Alonso, A., Marsal, S., & Julià, A. (2015). Analytical methods in untargeted metabolomics: State of the Art in 2015. *Front. Bioeng. Biotechnol., 3*, 23. https://doi.org/10.3389/fbioe.2015.00023

Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician, 27*(1), 17–21. https://doi.org/10.2307/2682899

Armstrong, R. A. (2019). Should Pearson's correlation coefficient be avoided? *Ophthalmic and Physiological Optics, 39*(5), 316–327. https://doi.org/10.1111/opo.12636

Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2017). Global patterns and trends in colorectal

cancer incidence and mortality. *Gut, 66*(4), 683. https://doi.org/10.1136/gutjnl-2015-310912

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics, 29*(4), 1165–1188.

Bingol, K., Bruschweiler-Li, L., Li, D., Zhang, B., Xie, M., & Brüschweiler, R. (2016). Emerging new strategies for successful metabolite identification in metabolomics. *Bioanalysis, 8*(6), 557–573. https://doi.org/10.4155/bio-2015-0004

Borenstein, M.; Hedges, L. V; Higgins, J. P. T.; Rothstein, H. R. *Introduction to Meta-Analysis*; 2009. https://doi.org/10.1002/9780470743386.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer Journal for Clinicians, 68*(6), 394–424. https://doi.org/10.3322/caac.21492

Broadhurst, D. I., & Kell, D. B. (2006). Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics, 2*(4), 171–196. https://doi.org/10.1007/s11306-006-0037-z

Bujang, M. A., & Baharum, N. (2016). Sample size guideline for correlation analysis. *World Journal of Social Science Research.* https://doi.org/10.22158/wjssr.v3n1p37

Camacho, D., de la Fuente, A., & Mendes, P. (2005). The origin of correlations in metabolomics data. *Metabolomics, 1*(1), 53–63. https://doi.org/10.1007/s11306-005-1107-3

Chaleckis, R., Meister, I., Zhang, P., & Wheelock, C. E. (2019). Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics. *Current Opinion in Biotechnology, 55*, 44–50. https://doi.org/10.1016/j.copbio.2018.07.010

Chok, N. S. Earson's Versus Spearman's and Kendall's Correlation Coefficients for Continuous Data, University of Pittsburgh, 2010.

Collino, S., Martin, F. P. J., & Rezzi, S. (2013). Clinical metabolomics paves the way towards future healthcare strategies. *British Journal of Clinical Pharmacology, 75*(3), 619–629. https://doi.org/10.1111/j.1365-2125.2012.04216.x

Dai, Z., Zhang, J., Wu, Q., Chen, J., Liu, J., Wang, L., Chen, C., Xu, J., Zhang, H., Shi, C., Li, Z., Fang, H., Lin, C., Tang, D., & Wang, D. (2019). The role of microbiota in the development of colorectal cancer. *International Journal of Cancer, 145*(8), 2032–2041. https://doi.org/10.1002/ijc.32017

de Siqueira Santos, S., Takahashi, D. Y., Nakata, A., & Fujita, A. (2014). A Comparative Study of Statistical Methods Used to Identify Dependencies between Gene Expression Signals. *Briefings in Bioinformatics, 15*(6), 906–918. https://doi.org/10.1093/bib/bbt051

de Winter, J. C. F., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. *Psychological Methods.* https://doi.org/10.1037/met0000079

Dekking, F. M.; Kraaikamp, C.; Lopuhaä, H. P.; Meester, L. E. *A Modern Introduction to Probability and Statistics*, 1st ed.; Springer-Verlag London, 2005. https://doi.org/10.1007/1-84628-168-7.

Di Giovanni, N., Meuwis, M.-A., Louis, E., & Focant, J.-F. (2020). Untargeted serum metabolic profiling by comprehensive two-dimensional gas chromatography–high-resolution time-of-flight mass spectrometry. *Journal of Proteome Research, 19*(3), 1013–1028. https://doi.org/10.1021/acs.jproteome.9b00535

Di Giovanni, N., Meuwis, M.-A., Louis, E., & Focant, J.-F. (2020b). Specificity of metabolic colorectal cancer biomarkers in serum through effect size. *Metabolomics.* https://doi.org/10.1007/s11306-020-01707-w

Dias, D. A., & Koal, T. (2016). Progress in metabolomics standardisation and its significance in future clinical laboratory medicine. *EJIFCC, 27*(4), 331–343.

Dienstmann, R., Vermeulen, L., Guinney, J., Kopetz, S., Tejpar, S., & Tabernero, J. (2017). Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nature Reviews Cancer, 17*(2), 79–92. https://doi.org/10.1038/nrc.2016.126

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-Mcintyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A., Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B., & Goodacre, R. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols, 6*(7), 1060–1083. https://doi.org/10.1038/nprot.2011.335

Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., Hankemeier, T., Goodacre, R., Neumann, S., Kopka, J., & Viant, M. R. (2013). Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics, 9*(1), 44–66. https://doi.org/10.1007/s11306-012-0434-4

Eylem, C. C., Yilmaz, M., Derkus, B., Nemutlu, E., Camci, C. B., Yilmaz, E., Turkoglu, M. A., Aytac, B., Ozyurt, N., & Emregul, E. (2020). Untargeted Multi-Omic Analysis of Colorectal Cancer-Specific Exosomes Reveals Joint Pathways of Colorectal Cancer in Both Clinical Samples and Cell Culture. *Cancer Letters, 469*, 186–194. https://doi.org/10.1016/j.canlet.2019.10.038

Farshidfar, F., Weljie, A. M., Kopciuk, K., Buie, W. D., MacLean, A., Dixon, E., Sutherland, F. R., Molckovsky, A., Vogel, H. J., & Bathe, O. F. (2012). Serum metabolomic profile as a means to distinguish stage of colorectal cancer. *Genome Med., 4*(5), 42. https://doi.org/10.1186/gm341

Gorrie, C. Three ways to detect outliers http://colingorrie.github.io/outlier-detection.html (accessed 2020 -05 -19).

Grubbs, F. E., & Beck, G. (1972). Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics, 14*(4), 847–854. https://doi.org/10.1080/00401706.1972.10488981

Hardin, J., Mitani, A., Hicks, L., & VanKoten, B. (2007). A robust measure of correlation between two genes on a microarray. *BMC Bioinformatics, 8*(1), 220. https://doi.org/10.1186/1471-2105-8-220

Hazra, A., & Gogtay, N. (2016). Biostatistics series module 6: correlation and linear regression. *Indian Journal of Dermatology, 61*(6), 593–601. https://doi.org/10.4103/0019-5154.193662

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., & Newman, T. B. (2013). *Designing Clinical Research: An Epidemiologic Approach* (4th ed.). Lippincott Williams & Wilkins.

Hyötyläinen, T. (2010). Analytical methodologies utilized in the search for chronic disease biomarkers. *Bioanalysis, 2*(5), 919–923. https://doi.org/10.4155/bio.10.38

Iglewicz, B.; Hoaglin David C. (David Caster), 1944-. *How to Detect and Handle Outliers*; Milwaukee, Wis. : ASQC Quality Press, 1993.

Janse, R. J., Hoekstra, T., Jager, K. J., Zoccali, C., Tripepi, G., Dekker, F. W., & van Diepen, M. (2021). Conducting correlation analysis: important limitations and pitfalls. *Clinical Kidney Journal, 14*(11), 2332–2337. https://doi.org/10.1093/ckj/sfab085

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways diseases and drugs. *Nucleic Acids Research, 45*(D1), D353–D361. https://doi.org/10.1093/nar/gkw1092

Kell, D. B., & Oliver, S. G. (2004). Here Is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *BioEssays, 26*(1), 99–105. https://doi.org/10.1002/bies.10385

Kotze, H. L., Armitage, E. G., Sharkey, K. J., Allwood, J. W., Dunn, W. B., Williams, K. J., & Goodacre, R. (2013). A novel untargeted metabolomics correlation-based network analysis incorporating human metabolic reconstructions. *BMC Systems Biology, 7*(1), 107. https://doi.org/10.1186/1752-0509-7-107

La Vecchia, S., & Sebastián, C. (2020). Metabolic pathways regulating colorectal cancer initiation and progression. *Seminars in Cell & Developmental Biology, 98*, 63–70. https://doi.org/10.1016/j.semcdb.2019.05.018

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Lin, Y., Ma, C., Bezabeh, T., Wang, Z., Liang, J., Huang, Y., Zhao, J., Liu, X., Ye, W., Tang, W., Ouyang, T., & Wu, R. (2019). 1H NMR-based metabolomics reveal overlapping discriminatory metabolites and metabolic pathway disturbances between colorectal tumor tissues and fecal samples. *International of Journal of Cancer.* https://doi.org/10.1002/ijc.32190

Mal, M., Koh, P. K., Cheah, P. Y., & Chan, E. C. Y. (2012). Metabotyping of human colorectal cancer using two-dimensional gas chromatography mass spectrometry. *Analytical and Bioanalytical Chemistry, 403*(2), 483–493. https://doi.org/10.1007/s00216-012-5870-5

Mamas, M., Dunn, W. B., Neyses, L., & Goodacre, R. (2011). The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Archives of Toxicology, 85*(1), 5–17. https://doi.org/10.1007/s00204-010-0609-6

McClelland, G. H. (2000). Nasty data: Unruly, ill-mannered observations can ruin your analysis. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology; Cambridge University Press.* Cambridge.

Mendes, P., Camacho, D., & de la Fuente, A. G. (2005). Modelling and simulation for metabolomics data analysis. *Biochemical Society Transactions, 33*(Pt 6), 1427–1429.

Metsämuuronen, J. (2021a). Goodman-Kruskal gamma and dimension-corrected gamma in educational measurement settings. *International Journal of Research and Method in Education, 7*, 95–118. https://doi.org/10.12973/ijem.7.1.95

Metsämuuronen, J. (2021b). Directional nature of goodman-Kruskal gamma and some consequences: identity of Goodman-Kruskal gamma and somers delta, and their connection to Jonckheere-Terpstra test statistic. *Behaviormetrika, 48*(2), 283–307. https://doi.org/10.1007/s41237-021-00138-8

Miller, J. N.; Miller, J. C. *Statistics and Chemometrics for Analytical Chemistry*, 6th ed.; Prentice Hall, Ed.; 2010.

Monedeiro, F., Monedeiro-Milanowski, M., Ligor, T., & Buszewski, B. (2020). A review of GC-based analysis of non-invasive biomarkers of colorectal cancer and related pathways. *Journal of Clinical Medicine.* https://doi.org/10.3390/jcm9103191

Motulsky, H. J.; Vanthemsche, M. (Trad. . *Biostatistique*, 3e édition.; Deboeck, Ed.; Bruxelles, 2019.

Nash, W. J., & Dunn, W. B. (2019). From mass to metabolite in human untargeted metabolomics: recent advances in annotation of metabolites applying liquid chromatography-mass spectrometry data. *TrAC Trends in Analytical Chemistry, 120*, 115324. https://doi.org/10.1016/j.trac.2018.11.022

Pernet, C.; Wilcox, R.; Rousselet, G. Robust Correlation Analyses: False Positive and Power Validation Using a New Open Source Matlab Toolbox. *Frontiers in Psychology.* **2013**, p 606.[26] Schober, P.; Boer, C.; Schwarte, L. A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth. Analg.* **2018**, *126* (5).

Qiu, Y., Cai, G., Su, M., Chen, T., Liu, Y., Xu, Y., Ni, Y., Zhao, A., Cai, S., Xu, L. X., & Jia, W. (2010). Urinary metabonomic study on colorectal cancer. *Journal of Proteome Research, 9*(3), 1627–1634. https://doi.org/10.1021/pr901081y

Qiu, Y., Cai, G., Su, M., Chen, T., Zheng, X., Xu, Y., Ni, Y., Zhao, A., Xu, L. X., Cai, S., & Jia, W. (2009). Serum metabolite profiling of human colorectal cancer using GC−TOFMS and UPLC−QTOFMS. *Journal of Proteome Research, 8*(10), 4844–4850. https://doi.org/10.1021/pr9004162

Rakotomalala, R. TANAGRA : Une Plate-Forme d'expérimentation Pour La Fouille de Données. *Revised Modules* 2005, 70–85.

Rodriguez-Salas, N., Dominguez, G., Barderas, R., Mendiola, M., García-Albéniz, X., Maurel, J., & Batlle, J. F. (2017). Clinical relevance of colorectal cancer molecular subtypes. *Critical Reviews in Oncology Hematology, 109*, 9–19. https://doi.org/10.1016/j.critrevonc.2016.11.007

Rorabacher, D. B. (1991). Statistical treatment for rejection of deviant values: critical values of dixon's "Q" parameter and related subrange ratios at the 95% confidence level. *Analytical Chemistry, 63*(2), 139–146. https://doi.org/10.1021/ac00002a010

Rousselet, G., & Pernet, C. (2012). Improving standards in brain-behavior correlation analyses. *Frontiers in Human Neuroscience.* https://doi.org/10.3389/fnhum.2012.00119

Schwarzkopf, D., de Haas, B., & Rees, G. (2012). Better Ways to Improve Standards in Brain-Behavior Correlation Analysis. *Frontiers in Human Neuroscience., 6*, 200.

Serra, A., Coretto, P., Fratello, M., & Tagliaferri, R. (2018). Robust and Sparse Correlation Matrix Estimation for the Analysis of High-Dimensional Genomics Data. *Bioinformatics, 34*(4), 625–634. https://doi.org/10.1093/bioinformatics/btx642

Seyfried, T. N., & Shelton, L. M. (2010). Cancer as a metabolic disease. *Nutrition & Metabolism (london), 7*(1), 7. https://doi.org/10.1186/1743-7075-7-7

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research, 13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Siska, C., & Kechris, K. (2017). Differential correlation for sequencing data. *BMC Research Notes, 10*(1), 54. https://doi.org/10.1186/s13104-016-2331-9

Steuer, R. (2006). Review: on the analysis and interpretation of correlations in metabolomic data. *Briefings in Bioinformatics, 7*(2), 151–158. https://doi.org/10.1093/bib/bbl009

Steuer, R., Kurths, J., Fiehn, O., & Weckwerth, W. (2003). Observing and interpreting correlations in metabolomic networks. *Bioinformatics, 19*(8), 1019–1026. https://doi.org/10.1093/bioinformatics/btg120

Student. The Probable Error of a Mean. *Biometrika* **1908**, *6* (1), 1–25. https://doi.org/10.2307/2331554.

Tan, B., Qiu, Y., Zou, X., Chen, T., Xie, G., Cheng, Y., Dong, T., Zhao, L., Feng, B., Hu, X., Xu, L. X., Zhao, A., Zhang, M., Cai, G., Cai, S., Zhou, Z., Zheng, M., Zhang, Y., & Jia, W. (2013). Metabonomics identifies serum metabolite markers of colorectal cancer. *Journal of Proteome Research, 12*(6), 3000–3009. https://doi.org/10.1021/pr400337b

Tebani, A., Afonso, C., & Bekri, S. (2018). Advances in metabolome information retrieval: turning chemistry into biology. Part ii: Biological information recovery. *Journal of Inherited Metabolic Disease, 41*(3), 393–406. https://doi.org/10.1007/s10545-017-0080-0

Tufte, E. R. (2001). *The Visual Display of Quantitative Information.* Graphics Press.

van der Ark, L. A., & van Aert, R. C. M. (2015). Comparing confidence intervals for Goodman and Kruskal's gamma coefficient. *Journal of Statistical Computation and Simulation, 85*(12), 2491–2505. https://doi.org/10.1080/00949655.2014.932791

Walker, D.; A Walker, D. JMASM9: Converting Kendall's Tau For Correlational Or Meta-Analytic Analyses. *J. Mod. Appl. Stat. Methods Copyr.* **2003**, *2*, 525–530. https://doi.org/10.22237/jmasm/1067646360.

Wan, M.-L., Wang, Y., Zeng, Z., Deng, B., Zhu, B.-S., Cao, T., Li, Y.-K., Xiao, J., Han, Q., & Wu, Q. (2020). Colorectal cancer (crc) as a multifactorial disease and its causal correlations with multiple

signaling pathways. Bioscience Reports. https://doi.org/10.1042/BSR20200265

Wang, G., Yu, Y., Wang, Y.-Z., Wang, J.-J., Guan, R., Sun, Y., Shi, F., Gao, J., & Fu, X.-L. (2019). Role of SCFAs in gut microbiome and glycolysis for colorectal cancer therapy. *Journal of Cellular Physiology, 234*(10), 17023–17049. https://doi.org/10.1002/jcp.28436

Warburg, O. (1956). On the Origin of Cancer Cells. *Science, 123*(3191), 309–314. https://doi.org/10.1126/science.123.3191.309

Weckwerth, W., & Fiehn, O. (2002). Can we discover novel pathways using metabolomic analysis? *Current Opinion in Biotechnology, 13*(2), 156–160. https://doi.org/10.1016/S0958-1669(02)00299-9

Weckwerth, W., & Morgenthal, K. (2005). Metabolomics: from pattern recognition to biological interpretation. *Drug Discovery Today, 10*(22), 1551–1558. https://doi.org/10.1016/S1359-6446(05)03609-3

Wilcox, R. R. (1994). The percentage bend correlation coefficient. *Psychometrika, 59*(4), 601–616. https://doi.org/10.1007/BF02294395

Wilcox, R. (2004). Inferences Based on a Skipped Correlation Coefficient. *Journal of Applied Statistics, 31*(2), 131–143. https://doi.org/10.1080/0266476032000148821

Wilcox, R. R., & Rousselet, G. A. (2018). A Guide to Robust Statistical Methods in Neuroscience. *Current Protocols in Neuroscience*. https://doi.org/10.1002/cpns.41

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., … Scalbert, A. (2018). HMDB 40: the human metabolome database for 2018. *Nucleic Acids Research, 46*(D1), D608–D617. https://doi.org/10.1093/nar/gkx1089

Xia, J., Sinelnikov, I. V., Han, B., & Wishart, D. S. (2015). MetaboAnalyst 3.0–making metabolomics more meaningful. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv380

Xia, J., & Wishart, D. S. (2016). Using MetaboAnalyst 30 for comprehensive metabolomics data analysis. *Current Protocols in Bioinformatics*. https://doi.org/10.1002/cpbi.11

Yarkoni, T. (2009). Big correlations in little studies: inflated fmri correlations reflect low statistical power—commentary on Vul et al. (2009). *Perspectives on Psychological Science, 4*(3), 294–298. https://doi.org/10.1111/j.1745-6924.2009.01127.x

Zanella, D., Focant, J.-F., & Franchina, F. A. (2021). 30th anniversary of comprehensive two-dimensional gas chromatography: latest advances. *Anal. Sci. Adv., 2*(3–4), 213–224. https://doi.org/10.1002/ansa.202000142

Zhang, F., Zhang, Y., Zhao, W., Deng, K., Wang, Z., Yang, C., Ma, L., Openkova, M. S., Hou, Y., & Li, K. (2017). Metabolomics for biomarker discovery in the diagnosis, prognosis, survival and recurrence of colorectal cancer: a systematic review. *Oncotarget, 8*(21), 35460–35472. https://doi.org/10.18632/oncotarget.16727

Zhu, J., Djukovic, D., Deng, L., Gu, H., Himmati, F., Chiorean, E. G., & Raftery, D. (2014). Colorectal cancer detection using targeted serum metabolic profiling. *Journal of Proteome Research, 13*(9), 4120–4130. https://doi.org/10.1021/pr500494u