Title: Metabolomic pathway activity with genomic single nucleotide polymorphisms associated with colorectal cancer recurrence and five year overall survival

Running Head: Prognostic metabolomics in colorectal cancer

Authors: Christina A. Fleming[a,b], Helen M. Mohan[c], Donal P. O'Leary[d], Mark Corrigan[b], H. Paul Redmond[b,e]

[a] Department of Colorectal Surgery, Cork University Hospital, Cork

[b] Department of Academic Surgery, Cork University Hospital, Cork

[c] Department of Colorectal Surgery, Peter MacCallum Cancer Centre, Melbourne, Australia

[d] Department of Surgical Oncology, Bon Secours Cork Cancer Centre, Cork, Ireland

[e] Department of Surgical Oncology, Cork University Hospital, Cork, Ireland

Corresponding Author: Dr Christina A Fleming

Department of Academic Surgery,

Cork University Hospital,

Cork, Ireland

T: +353214922000

christina.fleming49@gmail.com

ORCID ID: 0000-0003-2326-2655

Funding: No funding was obtained to perform this work.

Conflict of Interest: The authors have no conflict of interest to declare

Previous Communication: Awarded the British Association of Surgical Oncology (BASO) -Association of Cancer Surgeons (ACS) Proffered Paper Prize, November 2020

Abstract

Purpose: Metabolomic analysis in colorectal cancer (CRC) is an emerging research area with both prognostic and therapeutic targeting potential. We aimed to identify metabolomic pathway activity prognostic for CRC recurrence and overall survival and cross-reference such metabolomic data with prognostic genomic single nucleotide polymorphisms (SNPs).

Methods: A systematic search of Pubmed, Embase and Cochrane Library was performed for studies reporting prognostic metabolomic pathway activity in CRC in keeping with PRISMA guidelines. The QUADOMICS tool was used to assess study quality**.** MetaboAnalyst software, (version4.0) was used to map metabolites that were associated with recurrence and survival in CRC to recognised metabolic pathways and identify genomic SNPs associated with CRC prognosis, referencing the following databases: Human Metabolome Database (HMDB), the Small Molecule Pathway Database (SMPDB),PubChem and Kyoto Encyclopaedia of Genes and Genomes (KEGG) Pathway Database.

Results: Nine studies met the inclusion criteria, reporting on 1117 patients. Increased metabolic activity in the urea cycle (p=0.002, FDR=0.198) ammonia recycling (p=0.004, FDR=0.359) and glycine and serine metabolism (p=0.004, FDR=0.374) were prognostic of CRC recurrence. Increased activity in aspartate metabolism (p<0.001, FDR=0.079) and ammonia recycling (p=0.004, FDR=0.345) were prognostic of survival. Eight resulting SNPs were prognostic for CRC recurrence (rs2194980, rs1392880, rs2567397, rs715, rs169712, rs2300701, rs313408, rs7018169) and three for survival (rs2194980, rs169712, rs12106698) of which two overlapped with recurrence (rs2194980, rs169712).

Conclusions: With a caveat on study heterogeneity, specific metabolites and metabolic pathway activity appear evident in the setting of poor prognostic colorectal cancers and such metabolic signatures are associated with specific genomic SNPs.

Keywords: metabolomics; metabolism; colorectal cancer; genomic SNPs; cancer outcomes

Introduction

Metabolomics may be defined as the comprehensive study of the metabolome, a collection of all end-points of metabolism in an organism or biologic system at a given point in time and is the end point of analysis in the omics-cascade[1]. In its simplest form the omics-cascade proceeds from genomics to transcriptomics, proteomics and metabolomics however multi-level cross talk can also occur in any direction[1]. Specific environmental factors impact transition along the omics cascade and metabolomics therefore most accurately represents the posttranscriptional cellular phenotype compared to upstream omics within the omics-cascade. Cancer cells undergo metabolic transformation to sustain fast cell growth and proliferation[2]. This transformation results in different metabolic phenotypes in cancer cells compared with their control counterparts[3]. Identifying such differential metabolites has potential in developing diagnostic and prognostic markers in cancer and identifying therapeutic targets[4,5].

Specific metabolomic patterns have been identified in colorectal cancer that are distinct from healthy controls, are stage-specific and can differentiate metastatic states[6–9]. Studies have also proposed that the metabolome is a useful prognostic biomarker in CRC and can be associated with recurrence and survival outcomes[10]. Prior to development of metastatic disease, circulating tumour cells become ‘active’ from dormant states and undergo migration, adhesion and invasion with neoangiogenesis to form established metastases[11,12]. These processes require energy and thus produce metabolites which creates a metabolomic pattern that may be different in patients who develop recurrence and those disease- free[13].

Multilevel cross-talk occurs within the omics-cascade and genomic alterations can modify multiple levels of the biologic system[14,15]. For example, genomic

alterations can modify transcription factors which regulate enzymatic signalling and lead to metabolome modifications[3,16,17]. Cross-talk can occur top-down (from genomics), bottom up (from metabolomics) or middle-out (starting with transcriptomics or proteomics). The metabolome always remains as the final result and thus most closely represents the post-transcriptional cellular phenotype.

In this study we aimed to perform a systematic review of published metabolomic research to identify metabolites prognostic for CRC recurrence and poor survival and cumulatively map these metabolites to metabolic pathway activity. We also aimed to identify genomic single-nucleotide polymorphisms (SNPs) related to these metabolic findings that were prognostic for poor CRC outcomes as a marker of omics-crosstalk.

Materials and Methods

Search strategy

A systematic review of published literature reporting on the prognostic benefits of metabolomics for five-year CRC recurrence (local or distant) and/or survival was performed in keeping with PRISMA guidelines[18]. The following electronic databases were searched: PubMed, Embase and Cochrane Library. To identify relevant studies the following terms were used: (metabolomics OR metabonomics OR metabolites OR metabolome) AND (colon OR rectal OR colorectal) AND (cancer). The most recent search was performed on 9th May 2020.

Inclusion and exclusion criteria

Two authors independently examined the title and abstract of citations and the full texts of potentially eligible articles were obtained. The reference lists of retrieved

papers were further screened for additional eligible publications. All studies including adult (>18 years old) human patients were searched. No language or temporal restrictions were applied. Studies including data on specific metabolites as prognostic indicators in colorectal cancer were included. Studies reporting on metabolites from serum, plasma and tissue samples were included (as these are readily available samples in routine clinical practice). Studies reporting on faecal samples were excluded as faecal sampling is not routinely sourced or analysed in routine clinical practice. Further exclusion criteria involved the following: studies without original data such as reviews, letters, position statements and editorials; studies relating to metabolomic methodological analysis and therapeutic pharmacokinetics; studies reporting on metabolomics in cancers other than colorectal cancer; studies describing the use of metabolomics for risk of cancer development, diagnosis or screening for colorectal cancer.

Quality assessment

Assessment of quality of screened studies was performed using *QUADOMICS*, an adoption of the quality assessment tool QUADAS (**QU**ality **A**ssessment tool for **D**iagnostic **A**ccuracy **S**tudies). The QUADAS tool comprises four domains: patient selection, index test, reference standard, and flow and timing of conduct of study[19]. Each domain is assessed in terms of risk of bias with the first three domains also assessed in terms of applicability. Signalling questions are used to help judge risk of bias. The QUADOMICS tool is a modified version that assesses the methodological quality of studies utilising omics-based techniques[20]. Compared to QUADAS, four new items are incorporated to QUADOMICS related to pre-analytical conditions and methods to avoid overfitting of data, resulting in a total of 16 questions[20]. Each of

the 16 included questions (one of which is two part) of quality assessment were answered as ‘yes, no, unclear, not applied’ for each suitable study. The quality of each study was subsequently expressed as a percentage of applied criteria scored positively[20]. A threshold quality percentage was not applied for inclusion (Appendix 1).

Data extraction

Two authors independently performed data extraction. Following confirmation of included studies that met the required criteria the following study data was extracted: first author, year and country of publication; study design and number of patients studied; demographic and clinicopathological data of included patients (age, gender, body mass index (BMI) or weight category); biological specimen utilised and timing of sampling (tissue, serum, plasma); method of metabolome extraction and analytic platform; follow-up, five-year recurrence and survival data. In general terms metabolomic analysis was performed on extracted samples by either liquid or gas chromatography whereby metabolites are separated using either gas or liquid media followed by detection and or quantification using mass spectrometry (MS) which identifies metabolites based on mass to charge ration or nuclear magnetic resonance (NMR) spectrometry which uses magnetic field changes to determine the structure of organic molecules. Specific metabolites significantly associated with recurrence or survival were recorded by name and whether increased, decreased or bi-directionally changed.

Data analysis

MetaboAnalyst, version 4.0 software (http://www.metaboanalyst.ca) was used to map all metabolites that were identified as associated with recurrence and survival in colorectal cancer to recognised metabolic pathways. MetaboAnalyst is a comprehensive web-based tool suite designed to perform metabolomic data analyses, visualisation and functional interpretation[21]. MetaboAnalyst performs in-house mapping of common metabolites to a wide variety of database identifiers including Human Metabolome Database (HMDB), the Small Molecule Pathway Database (SMPDB), PubChem (National Institute of Health) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) Pathway Database prior to performing functional analysis.

Metabolite set enrichment analysis (MSEA) was performed to identify subtle but consistent biologically meaningful patterns of metabolic pathway activity indicative of CRC prognosis. Metabolites significantly associated with prognosis were included in analysis with independent analyses performed for recurrence and survival. Significance was accepted as area under the curve (AUC) >0.5 and an associated p value <0.05 on diagnostic accuracy testing. MSEA consisted of four steps: 1) metabolite data input using an over representation analysis; 2) data processing of standardised compounds from database cross-referencing; 3) data analysis using metabolic pathway associated metabolite sets from 99 separate libraries; 4) reporting of results using *Holm* p <0.05 to assess for statistical significance (adjusts p value to counteract the problem of multiple comparisons) and a false discovery rate (FDR) <0.40 to further adjust for multiple comparisons and type I errors. FDR<0.40 was the most liberal limit to set FDR however with significant heterogeneity in study design and data analytic platforms it was deemed necessary a priori. Pathway enrichment analysis with pathway topology analysis was then

performed to identify the most active metabolites in recurrence and survival. Again, results were interpreted using *Holm* p <0.05 and FDR <0.40.

A second MSEA was performed in a similar fashion using MetaboAnalyst to identify associations between included metabolites and genomic single-nucleotide polymorphisms (*SNPs*) i.e. a substitution of a single nucleotide that occurs at a specific position in the genome. The same four steps of MSEA (as outlined above) were followed. SNPs that identified significant association through MSEA were then automatically searched through the database of short genetic variations (dbSNP, National Institute of Health) to identify their chromosomal position, associated genes and previously reported clinical relevance in cancer and non-cancer conditions[22]. This was performed to identify metabolomic-genomic cross-talk associations that may also be prognostic in colorectal cancer.

Results

Search results and quality assessment

The results of the systematic literature search is outlined in the PRISMA diagram in Figure 1. Following exclusion of duplicate records a total of 669 references were screened by title and abstract. Subsequently, 129 full texts were reviewed of which nine studies were included in final analysis reporting on 1117 CRC patients[7,9], [23–29]. Reasons for manuscript exclusion are clearly outlined in Figure 1. Publication dates ranged from 2012-2020. Individual study weight ranged from 2.3-20.3% of overall patients numbers. Table 1 summaries the quality of included studies using the modified diagnostic accuracy quality assessment tool ‘QUADOMICS’. Overall, studies were of high quality with scores ranging from 83.8-95.6% (median score 86.5%). All nine studies scored highly in assessment of biospecimen sampling

and analysis as well as analytical process and reporting of positive and negative findings. However, few studies reported clearly on demographic and clinicopathological data in particular operative techniques (minimally invasive or open surgery) and type of operative performed where applicable. One element of QUADOMICS assesses whether all clinically relevant data that would be available in clinical practice was available under study conditions. Jimenez et al was the only study to comprehensively report on clinical data[28].

Clinical characteristics

Characteristics of included studies are summarised in Table 1. Of the 1117 CRC patients studied, 52.6% (n=588) were male and 47.2% (n=529) female. Median age in included studies ranged from 57-73.9 years. Three studies reported body mass index (BMI, kg/m^2). Zaimenko et al reported on a cohort with a median BMI of 25[24], Wang et al, a mean BMI 23.0-23.5[9] and in the cohort studied by Bertini et al 45% were described as normal weight, 33% overweight and 15.1% obese[7]. Four studies (reporting on n=537 patients) described findings in all patients from stage I-IV CRC and two studies (n=223) reported on Stage I-III. Single studies reported on Stage II-IV (n=112), Stage II only (n=92) and one Stage IV only (n=153). All patients were followed-up for five years and cancer recurrence and survival data is summarised in Table 1.

Sampling and Metabolomic Analytic Techniques

Four studies reported on the metabolome of tumour tissue, one study reported on a combination of plasma and tumour tissue and three studies blood-based metabolome (one analysing plasma and two analysing serum). Metabolite detection

was performed using a variety of methods including: ultra-performance liquid chromatography with electrospray ionization quadrupole time-of-flight mass spectrometry (UPLC-QTOF-MS); liquid chromatography-mass spectrometry (LC-MS); hydrophilic interaction liquid chromatography solid phase extraction (LC-HILIC-SPE); high resolution magic angel spinning nuclear magnetic resonance and mass spectrometry (HR-MAS-NMR) and gas chromatography with time of flight mass spectrometry (GC-TOF-MS).

Multiple data processing analytic platforms and statistical packages of varying versions were also used to normalise and report on spectral outputs, the most common being ProteoWizard, Amix, TopSpin, R CAMERA package. Patient follow-up was reported as five-years in eight studies with follow-up duration not clearly defined in one study. Three studies reported on recurrence alone, four on survival alone and two on both recurrence and survival.

Prognostic metabolites for colorectal cancer recurrence

Five studies reported on metabolites prognostic for CRC recurrence[7,9,24,27,29]. Available reported recurrence rates are summarised in Table 1. Pathway results from metabolite set enrichment analysis (MSEA) for CRC recurrence is summarised in Figure 2. Eight metabolic pathways were identified as significantly active in patients who developed CRC recurrence: urea cycle [holm p=0.002]; ammonia recycling [holm p=0.0037]; glycine and serine metabolism [holm p=0.0039]; glutathione metabolism [holm p=0.001]; alanine metabolism [holm p=0.0219]; glutamate metabolism [holm p=0.038]; taurine and hypotaurine metabolism [holm p=0.04]; glucose-alanine cycle [holm p=0.0495]. However, only three pathways

demonstrated significance and an FDR<0.4: urea cycle (FDR=0.198); ammonia recycling (FDR=0.359); glycine and serine metabolism (FDR=0.374).

Appendix two summarises pathway enrichment analysis for specific metabolites prognostic for CRC recurrence identified through pathway analysis. These metabolites were particularly significant in ammonia recycling (p=0.022) with predominant flux in the following metabolites: L-aspartic acid, L-histidine, L-serine, pyruvic acid and glycine. Activity in homocysteine degradation and ß-alanine metabolism approached significance (p=0.075, p=0.091 respectively). In homocysteine degradation, flux in L-serine and L-cysteine were most significantly observed and in ß-alanine metabolism, flux in uracil, ß-alanine, L-aspartic acid and L-histidine concentrations were observed.

Prognostic metabolites for overall survival

Six studies reported prognostic metabolites for CRC survival[23–28]. Available reported survival rates are summarised in Table 1. Metabolite set enrichment analysis (MSEA) for CRC survival is summarised in Figure 2. Eight metabolic pathways were significantly active in the setting of worse CRC survival: aspartate metabolism [holm p<0.001]; ammonia recycling [holm p=0.0036]; glutathione metabolism [holm p=0.0045]; glutamate metabolism [holm p=0.0023]; ß-alanine metabolism [holm p=0.0256]; arginine and proline metabolism [holm p=0.0304]; purine metabolism [holm p=0.035]; malate-aspartate shuttle [holm p=0.043]. However, only two pathways demonstrated significance and an FDR<0.4: aspartate metabolism (FDR=0.08) and ammonia recycling (FDR=0.345).

Appendix two summarises pathway enrichment analysis for specific metabolites associated with poor CRC survival. These were particularly active in

aspartate metabolism (p=0.027) with predominant flux in the following metabolites: L-aspartic acid, acetylglycine, L-asparagine and ß-alanine. Activity in ammonia and ß-alanine metabolic pathways approached significance (p=0.052, p=0.058 respectively). In ammonia recycling, flux in L-aspartic acid, L-histidine, glycine, L-serine and pyruvic acid concentrations were most active and in ß-alanine metabolism flux in uracil ß-alanine, L-aspartic acid and L-histidine concentrations were most active.

Identification of metabolite associated SNPs

In figure three, we outline that eleven single-nucleotide polymorphisms (SNPs) demonstrated significant association with prognostic activity in identified metabolic pathways: n=6 for recurrence only [rs1392880 (holm p=0.0037, FDR<0.001); rs2567397/ Rs7664292 (holm p=0.0037, FDR<0.001); rs715 (holm p=0.0037, FDR<0.001); rs2300701 (holm p=0.0181, FDR=0.002); rs313408 (holm p=0.0181, FDR=0.002); rs7015169 (holm p=0.0181, FDR=0.002);], n=2 for recurrence and survival [rs2194980 (holm p<0.001, FDR<0.001); rs169712 (holm p=0.001, FDR5<0.001)] and n=1 for survival only [rs12106696 (holm p=0.016, FDR=0.005)]. Table 2 summarises relevant genomic, clinical and cancer data related to each SNP that was identified as significantly prognostic for CRC outcomes. Overall, four SNPs were of unknown clinical significance and seven SNPs had no previously reported relevance in cancer. Three SNPs prognostic for CRC recurrence (rs2300701, rs313408, rs7018169) had previously shown clinical association in cancer and one SNP prognostic for CRC survival (rs12106696).

Discussion

Recent advances in biomolecular techniques and bioinformatic technology have allowed unravelling of metabolic pathway information with real clinical application[30]. In this review we identified activity in specific oncometabolites and activity in specific metabolic pathways that are associated with CRC recurrence and five-year survival through cross-reference of synthesised published data and publicly-available metabolite databases. Previous reviews have focused on CRC diagnosis and screening biomarker development and comparison of CRC metabolomic data to healthy controls or differentiating patients with cancer and benign colorectal polyps[10,29].

Identifying metabolic pathways and metabolite clusters that are prognostic for cancer recurrence and survival is an important biomarker discovery development[10,31,32]. Early identification of poor prognosis in CRC cancer allows us to personalise surveillance and administer systemic treatment to optimise outcome. Identifying recurrence or risk of recurrence and poor outcome at the earliest opportunity offers the best potential for outcome[33,34]. Currently, carcinoembryonic antigen (CEA) is the surveillance biomarker of choice in CRC and is the only biomarker recommended in screening guidelines, however its sensitivity is low[35], [36]. There remains an opportunity to identify more robust markers to identify poor outcomes in CRC.

We identified specific metabolites and metabolic pathways that were representative of poor prognosis in CRC. Poor survival is usually a result of biologically aggressive disease that is metastatic at diagnosis or recurs at a later stage thus there is obvious cross-over in metabolic signatures of recurrence and survival [37,38]. Previously reported cellular metabolic signatures support our

findings[8,39–41]. Global activity in ammonia recycling was prognostic for both cancer recurrence and survival. Flux in its specific metabolites L-aspartic acid, pyruvic acid, L-histidine, L-serine and glycine were also prognostic for cancer recurrence. Urea cycle activity and ammonia recycling are interlinking metabolic pathways[39]. Ammonia has previously been considered as a 'toxic' by-product of upregulated cellular metabolism due to metabolic demand of tumour cells[40]. Such ammonia must be exported from cells and is usually cleared through urea cycle activity in the liver[40,41]. At a cellular level, ammonia in itself is an abundant by-product of cellular metabolism, however the biological effects of ammonia production and recycling in cancer is not fully understood[40,42]. As previously outlined, it is proposed that ammonia is not just a waste product but rather is recycled into central amino acid metabolism to maximise nitrogen utilisation [42]. In this way, cancer cells primarily utilise ammonia through reductive amination, catalysed by glutamate dehydrogenase (GDH) and secondary reactions enable other amino acids to directly acquire nitrogen, namely proline and aspartate [41,42]. Metabolic recycling of ammonia has been demonstrated to accelerate proliferation of breast cancer cells in vivo and thus the potential for cell survival and migration[43]. In mouse models of breast cancer, ammonia accumulated in the tumour microenvironment was directly used to generate amino acids through GDH activity in a cyclical nature and thus is a fundamental nitrogen source that supports tumour development and growth[43]. Significantly higher levels of ammonia have been observed in plasma samples from gastric, breast and thyroid cancer patients compared to healthy controls[44].

Increased activity in the serine and glycine metabolic pathway was associated with CRC recurrence and flux in individual metabolites L-serine and glycine were also associated with recurrence following pathway analysis. Serine has previously

been described as a central element of cancer metabolism[45]. Cancer cells extensively use glycolysis to sustain anabolism (the Warburg effect), which is necessary for tumour growth[46,47]. Serine biosynthesis is a component of these glycolysis-diverting pathways. De novo synthesis of serine also provides precursors for a variety of biosynthetic pathways, and a pivotal aspect of the serine de novo synthesis is the conversion of serine to glycine by serine hydroxymethyltransferase (SHMT)[45,48]. Glycine is a major source of methyl groups for the one-carbon pools required for the biosynthesis of glutathione, protein, purines and DNA/histone methylation[48].

The tumour suppressor p53 has been associated with the capacity of cancer cells to deal with oxidative stress[49]. The idea that cancer cells reprogram their metabolism to counteract reactive oxygen species (ROS) production fits with the suggestion that aerobic glycolysis (the Warburg effect) is adopted to avoid the generation of metabolic ROS and thus support tumorigenesis and cell survival and are also suggested to support metastatic progression[50]. In vivo studies have also explored the effect of restricting serine and glycine metabolism for cancer intervention[45,49]. In the absence of p53, serine and glycine withdrawal had an even greater effect, suggesting an epistatic interaction between p53 and the availability of serine and glycine[49]. Aspartate was also associated with poor survival. Aspartate levels are quite low in circulation as most cells lack asparaginase activity required to convert asparagine to aspartate however it is abundant in cancer cells[51]. Thus circulating cancer cells undergoing metastatic processes may be associated increased aspartate levels[52].

Cross-referencing associations between identified metabolic signatures and publicly available SNP databases we identified 11 genomic SNPs that are potentially

associated with poor prognosis CRC. Most interesting is rs7018169, an intron variant SNV on chromosome eight at the location 8p13 that may contribute to alteration of the ST3GAL1 gene[53]. ST3GAL1 is a member of the sialyltransferase gene family and over production has been shown to promote cell migration and invasion due to a subsequent increased expression of MUC1 (mucin protein)[53]. MUC1 is associated with increased tumour invasive potential via epithelial-mesenchymal transition (EMT) which is suggested as a mechanism for tumour cell migration, implantation and metastatic development[54]. These interactions have also been associated with expression of cadherin and integrin which underpin the mechanism of peritoneal carcinomatosis development in ovarian cancer and poor prognosis[55].

There are a number of limitations to this review. While the quality of included studies was high in individual studies, methodology and timing of biological specimen sampling and analytical platforms and pre-processing data analysis were heterogenous across all studies. The main deficiency in study design was minimal reporting of clinical data. Understanding more about the biological entity of patients e.g. right or left sided colonic or rectal tumours, KRAS status, MMR (mismatch repair) status would prove more informative for clinical application. The data also does not provide context for other therapies i.e. chemotherapy or radiotherapy. In data analysis we utilised and FDR<0.4. While this is acceptable it is at the lower limit of acceptable but heterogeneity in study design required adjustment. This liberal FDR rate should be considered when considering the applicability of these findings. Finally, BMI or body composition data was only reported in three of nine studies and it is recognised that the metabolome is significantly influenced by body composition[52,53]. Dietary preferences were also not referenced.

In conclusion, we outline specific metabolite and metabolic pathway activity associated with CRC recurrence and survival and using these metabolic signatures we identified related, relevant genomic SNPs. These findings provide a platform for prognostic biomarker discovery and development in colorectal cancer.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements: all studies from which data was extracted are accurately referenced.

References

[1] W. B. Dunn, D. I. Broadhurst, H. J. Atherton, R. Goodacre, and J. L. Griffin, "Systems level studies of mammalian metabolomes: The roles of mass spectrometry and nuclear magnetic resonance spectroscopy," *Chem. Soc. Rev.*, vol. 40, no. 1, pp. 387–426, 2011.

[2] S. Romero-Garcia, J. S. Lopez-Gonzalez, J. L. B´ez-Viveros, D. Aguilar-Cazares, and H. Prado-Garcia, "Tumor cell metabolism," *Cancer Biol. Ther.*, vol. 12, no. 11, pp. 939–948, 2014.

[3] R. Janke, A. Dodson, and J. Rine, "Metabolism and epigenetics.," *Annu. Rev. Cell Dev. Biol.*, vol. 31, p. 473e496, 2015.

[4] S. Santhanama, D. M. Alvaradoa, and M. A. Ciorba, "Therapeutic targeting of inflammation and tryptophan metabolism in colon and gastrointestinal cancer," *Transl. Res.*, vol. 167, no. 1, pp. 67–69, 2016.

[5] I. A. Bettencourt and J. D. Powell, "Targeting Metabolism as a Novel Therapeutic Approach to Autoimmunity, Inflammation, and Transplantation," *J. Immunol.*, vol. 198, no. 3, pp. 999–1005, 2017.

[6] L. C. Phua *et al.*, "Non-invasive fecal metabonomic detection of colorectal cancer," *Cancer Biol Ther*. Apr;15(4):389–397, 2014.

[7] I. Bertini *et al.*, "Metabolomic NMR Fingerprinting to Identify and Predict Survival of Patients with Metastatic Colorectal Cancer," *Cancer Res.* vol. 72, no. 37, pp. 356–365, 2012.

[8] Z. Long *et al.*, "Metabolomic markers of colorectal tumor with different clinicopathological features," *Frony Oncol.* vol. 17, no.10, pp. 981, 2020.

[9] Z. Wang *et al.*, "Development of a Correlative Strategy To Discover Colorectal Tumor Tissue Derived Metabolite Biomarkers in Plasma Using Untargeted Metabolomics," *Anal Chem.* , vol 91, no. 3, pp 2401-2408, 2019.

[10] A. Zhang, H. Sun, G. Yan, P. Wang, Y. Han, and X. Wang, "Metabolomics in diagnosis and biomarker discovery of colorectal cancer," *Cancer Lett.*, vol. 345, no. 1, pp. 17–20, 2014.

[11] I. J. Fidler, "The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited," *Nat. Rev. Cancer*, vol. 3, no. 6, pp. 453–458, 2002.

[12] D. P. O'Leary, E. O'Leary, N. Foley, T. G. Cotter, J. H. Wang, and H. P. Redmond, "Effects of surgery on the cancer stem cell niche," *Eur. J. Surg. Oncol.*, vol. 42, no. 3, pp. 319–325, 2016.

[13] F. Zhang *et al.*, "Metabolomics for biomarker discovery in the diagnosis, prognosis, survival and recurrence of colorectal cancer: a systematic review," *Oncotarget*, vol. 8, no. 21, pp. 35460–35472, 2017.

[14] B. C. Yoo, K. H. Kim, S. M. Woo, and J. K. Myung, "Clinical multi-omics strategies for the effective cancer management," *J. Proteomics*, no. July, pp. 0–1, 2017.

[15] E. A. Vucic *et al.*, "Translating cancer ' omics ' to improved outcomes paradigm Translating cancer ' omics ' to improved outcomes," *Genome Reasearch*, pp. 188–195, 2012.

[16] Y. Liao, "Akt and glucose metabolism in cancer (Invited review).," *J. Metab. Nutr. Cancer*, vol. 1, no. 2, p. 80e88, 2014.

[17] L. Dang, D. White, S. Gross, and E. Al., “Cancer-associated IDH1 mutations produce 2-hydroxyglutarate.,” *Nature*, vol. 462, no. 7274, p. 739e744, 2009.
[18] Page, Matthew J et al. “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.” BMJ (Clinical research ed.) vol. 372 n71. 29 Mar. 2021,.” .
[19] J. B. Reitsma, M. M. G. Leeflang, J. A. C. Sterne, and P. M. M. Bossuyt, “Research and Reporting Methods Accuracy Studies,” no. 4, 2011.
[20] B. Lumbreras, M. Porta, S. Márquez, M. Pollán, L. A. Parker, and I. Hernández-Aguado, “QUADOMICS: An adaptation of the Quality Assessment of Diagnostic Accuracy Assessment (QUADAS) for the evaluation of the methodological quality of studies on the diagnostic accuracy of ’-omics’-based technologies,” *Clin. Biochem.*, vol. 41, no. 16–17, pp. 1316–1325, 2008.
[21] J. Chong *et al.*, “MetaboAnalyst 4 .0 : towards more transparent and integrative metabolomics analysis,” no. May, pp. 0–9, 2018.
[22] N. I. of Health, “NCBI Database of short genetic variations.” .
[23] Y. Cai *et al.*, “Sex Differences in Colon Cancer Metabolism Reveal A Novel Subphenotype,” *Sci Rep.*, vol. 17, no. 1 pp. 4905, 2020.
[24] I. Zaimenko *et al.*, “Non-invasive metastasis prognosis from plasma metabolites in stage II colorectal cancer patients : The DACHS study,” Int J Cancer, vol. 145, no. 1, pp. 221-231, 2019.
[25] S. W. De Vroome *et al.*, “Serum N -glycome alterations in colorectal cancer associate with survival,” *Oncotarget*, vol. 9, no. 55, pp. 30610–30623, 2018.
[26] B. Pacholczyk-sienicka and A. Fabia, “Prediction of Survival for Patients With Advanced Colorectal Cancer Using 1 H High-Resolution Magic Angle Spinning Nuclear MR Spectroscopy,” *J Magn Reson Imaging*, vol. 1674, pp. 1669–1674, 2015.
[27] Y. Qiu *et al.*, “A Distinct Metabolic Signature of Human Colorectal Cancer with Prognostic Potential,” *Clin Cancer Res.* vol. 20, no. 8, pp. 2136–2147, 2014.
[28] B. Jimenez *et al.*, “H HR-MAS NMR Spectroscopy of Tumor-Induced Local Metabolic ‘ Field-E ff ects ’ Enables Colorectal Cancer Staging and Prognostication,” *J Proteome Res.* vol 12, no. 2, pp. 959-968, 2013.
[29] F. Farshidfar *et al.*, “Serum metabolomic profile as a means to distinguish stage of colorectal cancer,” *Genome Med.*, vol. 4, no. 5, pp. 42, 2012.
[30] R. D. Beger *et al.*, “Metabolomics enables precision medicine: ‘A White Paper, Community Perspective,’” *Metabolomics*, vol. 12, no. 10, 2016.
[31] J. L. Spratlin, N. J. Serkova, and S. G. Eckhardt, “Clinical applications of metabolomics in oncology: A review,” *Clin. Cancer Res.*, vol. 15, no. 2, pp. 431–440, 2009.
[32] F. Zhang *et al.*, “Applications of metabolomics in cancer studies,” *Oncotarget*, vol. 345, no. 1, pp. 552–574, 2017.
[33] E. Morris *et al.*, “Surgical management and outcomes of colorectal cancer liver metastases.,” *Br. J. Surg.*, vol. 97, no. 7, p. 1110, 2010.
[34] V. Väyrynen *et al.*, “Incidence and management of patients with colorectal cancer and synchronous and metachronous colorectal metastases : a population-based study,” *BJS Open*, vol. 4, no. 4, pp. 685-692, 2020.
[35] B. Shinkins *et al.*, “Serum carcinoembryonic antigen trends for diagnosing colorectal cancer recurrence in the FACS randomized clinical trial,” *Br. J. Surg.*, vol. 105, no. 6, pp. 658–662, 2018.
[36] K. Leong, J. Hartley, and S. Karandikar, “Association of Coloproctology of Great Britain & Ireland (ACPGBI): Guidelines for the Management of Cancer of the Colon, Rectum and Anus (2017) – Follow Up, Lifestyle and Survivorship,” *Color. Dis.*, vol. 19, pp. 67–70, 2017.
[37] J. P. Ryuk *et al.*, “Predictive factors and the prognosis of recurrence of colorectal cancer within 2 years after curative resection,” *Ann. Surg. Treat. Res.*, vol. 86, no. 3, pp. 143–151, 2014.
[38] L. G. M. van der Geest, J. Lam-Boer, M. Koopman, C. Verhoef, M. A. G. Elferink, and J. H. W. de Wilt, “Nationwide trends in incidence, treatment and survival of colorectal cancer patients with synchronous metastases,” *Clin. Exp. Metastasis*, vol. 32, no. 5, pp. 457–465, 2015.
[39] R. Keshet, P. Szlosarek, A. Carracedo, and A. Erez, “Rewiring urea cycle metabolism in cancer to support anabolism,” *Nat. Rev. Cancer*, vol. 18, no. October, 2018.
[40] M. Adeva, G. Souto, N. Blanco, and C. Donapetry, “Ammonia metabolism in humans.,” *Metabolism.*, vol. 61, pp. 1495–1511, 2012.
[41] R. DeBerardinis and T. Cheng, “The diverse functions of glutamine in metabolism, cell biology and cancer.,” *Oncogene*, vol. 29, pp. 313–324.
[42] J. Coloff and E. Al., “Differential Glutamate Metabolism in Proliferating and Quiescent Mammary Epithelial Cells.,” *Cell Metab.*, vol. 23, pp. 867–880, 2016.
[43] M. C. Haigis, “Metabolic recycling of ammonia via glutamate dehydrogenase supports breast cancer biomass,” *Science*, vol. 358, no. 6365, pp. 941–946, 2018.
[44] Y. Gu *et al.*, “Perioperative dynamics and significance of amino acid profiles in patients with cancer,” *J. Transl. Med.*, vol. 13, no. 1, pp. 1–14, 2015.
[45] A. Antonov, I. Amelio, F. Cutruzzola, M. Agostini, and G. Melino, “Serine and glycine metabolism in cancer,” *Trends in Biochemical Sciences*, vol. 39, no. 4, pp. 191–198, 2014.
[46] M. Potter, E. Newport, and K. Morten, “The Warburg effect: 80 years on.,” *Biochem. Soc. Trans.*, vol. 44, no. 5, p. 1499e1505, 2016.
[47] O. Warburg, F. Wind, and E. Negelein, “The metabolism of tumors in the body.,” *J. Gen. Physiol.*, vol. 8,

pp. 519–530, 1927.
[48] N. Shyh-Chang and E. Al., “Influence of threonine metabolism on S-adenosylmethionine and histone methylation,” *Science*, *339*, pp. 222–22638, 2013.
[49] O. D. Maddocks and E. Al., “Serine starvation induces stress and p53- dependent metabolic remodelling in cancer cells.,” *Nature*, vol. 493, pp. 542–546, 2013.
[50] T. Han *et al.*, “How does cancer cell metabolism affect tumor migration and invasion?,” *Cell Adhes. Migr.*, vol. 7, no. 5, pp. 395–403, 2013.
[51] J. Mayers and M. Vander Heiden, “Famine versus feast: understanding the metabolism of tumors in vivo.,” *Trends Biochem. Sci.*, vol. 40, pp. 130–140, 2015.
[52] T. Batool, E. Makky, M. Jalal, and M. Yusoff, “A Comprehensive Review on l-Asparaginase and Its Applications.,” *Appl. Biochem. Biotechnol.*, vol. 178, pp. 900–923, 2016.
[53] X. Wu, J. Zhao, Y. Ruan, L. Sun, C. Xu, and H. Jiang, “Sialyltransferase ST3GAL1 promotes cell migration, invasion, and TGF-β1-induced EMT and confers paclitaxel resistance in ovarian cancer,” *Cell Death Dis.*, vol. 9, no. 11, 2018.
[54] L. Zhang and E. Al, “Effects of Kras activation and Pten deletion alone or in combination on MUC1 biology and epithelial-to-mesenchymal transition in ovarian cancer.,” *Oncogene*, vol. 35, pp. 5010–5020, 2016.
[55] M. Canel, A. Serrels, M. Frame, and V. Brunton, “E-cadherin-integrin crosstalk in cancer invasion and metastasis.,” *J. Cell. Sci.*, vol. 126, pp. 393–401, 2013.
[56] S. Moore, C. Matthews, J. Sampson, and E. Al, “Human metabolic correlates of body mass index.,” *Metabolomics*, vol. 10, no. 2, pp. 259–269, 2014.
[57] D. B. Liesenfeld *et al.*, “Metabolomics and transcriptomics identify pathway differences between visceral and subcutaneous adipose tissue in colorectal cancer patients : the ColoCare study,” *Am J Clin Nutr*, vol. 102, no. 2, pp. 433-443, 2015.

Tables and Figures

Table 1: Summary of study characteristics of included studies. SD=standard deviation. CRC=colorectal cancer. D=day. UPLC-QTOF-MS=ultra-performance liquid chromatography with electrospray ionization quadrupole time-of-flight mass spectrometry. LC-MS=liquid chromatography-mass spectrometry. LC-HILIC-SPE=hydrophilic interaction liquid chromatography solid phase extraction. HR-MAS-NMR=high resolution magic angel spinning nuclear magnetic resonance and mass spectrometry. GC-TOF-MS=gas chromatography with time of flight mass spectrometry.

AUTHOR	YEAR	COUNTRY	CRC Cohort (N)	Age	Gender	BMI	CRC Stage	Biospecimen	Metabolite Detection	Pre-Processing Analysis
Caf[23]	2020	USA	197	Mean 63.3-73.9 SD 5.8-16.1	Male 54.7% Female 45.3%	Not reported	I 23.9% II 43.7% III 32.4%	Tissue - Tumour and normal colon	UPLC-QTOF-MS	ProteoWizard MS Converter v3.06150
Zaimenko[24]	2019	Germany	92	Median 73 Range 45-87	Male 54% Female 46%	25	All stage II	Plasma – single pre-operative sample	LC-MS	Bruker Data Analysis; centWave
Wang[9]	2019	China	73	Median 60.2 SD 9.4	Male 55% Female 45%	Mean 23.0-23.5 SD 3.2-2.8	I 17.8% (n=13) II 39.7% (n=29) III 37% (n=27) IV 5.5% (n=4)	Tissue - surgical specimen Plasma - pre-op D1 and post-op D7	LC-MS	ProteoWizard MS Converter v 3.06150
de Vroome[25]	2018	Netherlands	185	Mean 63-66.8 SD 10.6-13.4	Male 54.7% Female 45.3%	Not reported	Is 2.7% (n=5) I 18.4% (n=34) II 36.8% (n=68) III 28.1% (n=52) IV 14% (n=26)	Serum – pre-operatively and >45 days post-operative	LC-HILIC-SPE	MALDI-TOF-MS
Pacholczyk[26]	2015	Poland	52	Mean 61 Range 34-86	Male 44% Female 56%	Not reported	I 19.2% (n=10) II 26.9% (n=14) III 30.8% (n=16) IV 23.1% (n=12)	Tissue - surgical specimen	HR-MAS-NMR	TopSpin 2.1
Qiu[27]	2014	USA/China	227	Median 57-61 Range 31-86	Male 53% Female 47%	Not reported	I 13.2% (n=30) II 32.2% (n=73) III 5.4% (n=103) IV 9.2% (n=21)	Tissue – surgical specimen	GC-TOF-MS	ChromaTOF V4.22, Leco Co.
Jimenez[28]	2013	UK	26	Median 72 Range 26-87	Male 38% Female 62%	Not reported	I 7.7% (n=2) II 38.5% (n=10) III 53.8% (n=14)	Tissue –surgical specimen	HR-MAS-NMR; GC-MS	TopSpin 2.2 Amix v3.9.9
Bertini[7]	2012	Italy	153	Median 63 Range 46-87	Male 60.2% Female 39.8%	Underweight 6% Normal 45.3% Overweight 33% Obese 15.1%	All stage IV	Serum - pretreatment	^{1}H-NMR	TopSpin V2.1 Amix v3.8.4
Farshidfar[29]	2012	Canada	112	Median 63-72 SD +/- 11-13	Male 59.8% Female 40.2%	Not reported	II 18.8% (n=21) III 18.8% (n=21) IV 62.4% (n=70)	Serum - pretreatment	^{1}H NMR; GC-MS	Metabolite Detector v2.06
			1117							

Table 2: Summary of genomic, clinical and cancer data relevant to identified prognostic single-nucleotide polymorphisms (SNPs) in colorectal cancer recurrence and survival. Rs=reference SNP. SNV=single nucleotide variant. Chr=chromosome. LOC=locus. SRD5A2= Steroid 5 alpha-reductase 2. DYNC2H1= dynein cytoplasmic-2-heavy chain 1. ST3GAL1= ST3 -beta-galactoside alpha-2,3-sialyltransferase 1. ZNF385D=zinc finger protein 385D.

SNP	Variant	Chromosome Gene	Relevance to Cancer	Other Characteristics
RECURRENCE				
rs2194980*	SNV Intron variant	Chr 12 LOC102723639	Not reported	Developmental delay
rs1392880	SNV	Chr 4 Unknown	Not reported	Clinical significance unknown
rs2567397	SNV	Chr 4	Not reported	Clinical significance unknown
rs715	SNV	Chr 2 CPS1	Not reported	Congenital hyperammonaemia Raised BMI Coronary disease Smoking T2DM Enterocolitis
rs169712*	SNV Intron variant	Chr 15 LOC105370991 LOC105370990	Not reported	Clinical significance unknown
rs2300701	SNV Intron variant	Chr 2 SRD5A2	Gleason score Prostate cancer stage Endometrial cancer	Hypogonadism
rs313408	SNV Intron variant	11 DYNC2H1	Diffuse-type gastric cancer	Primary cilium dysfunction
rs7018169	SNV Intron variant	Chr 8 ST3GAL1	Promotes cell migration and invasion Ovarian cancer Breast cancer	Meningioma
SURVIVAL				
rs2194980*	SNV Intron variant	Chr 12 LOC102723639	Not reported	Developmental delay
rs169712*	SNV Intron variant	Chr 15 LOC105370991 LOC105370990	Not reported	Clinical significance unknown
rs12106696	SNV Intron variant	Chr 3 ZNF385D	Liver, prostate, cervical and ovarian cancer	Dyslexia

Figure 1: PRISMA flowchart for systematic review of published literature identifying metabolites prognostic for colorectal cancer recurrence and survival

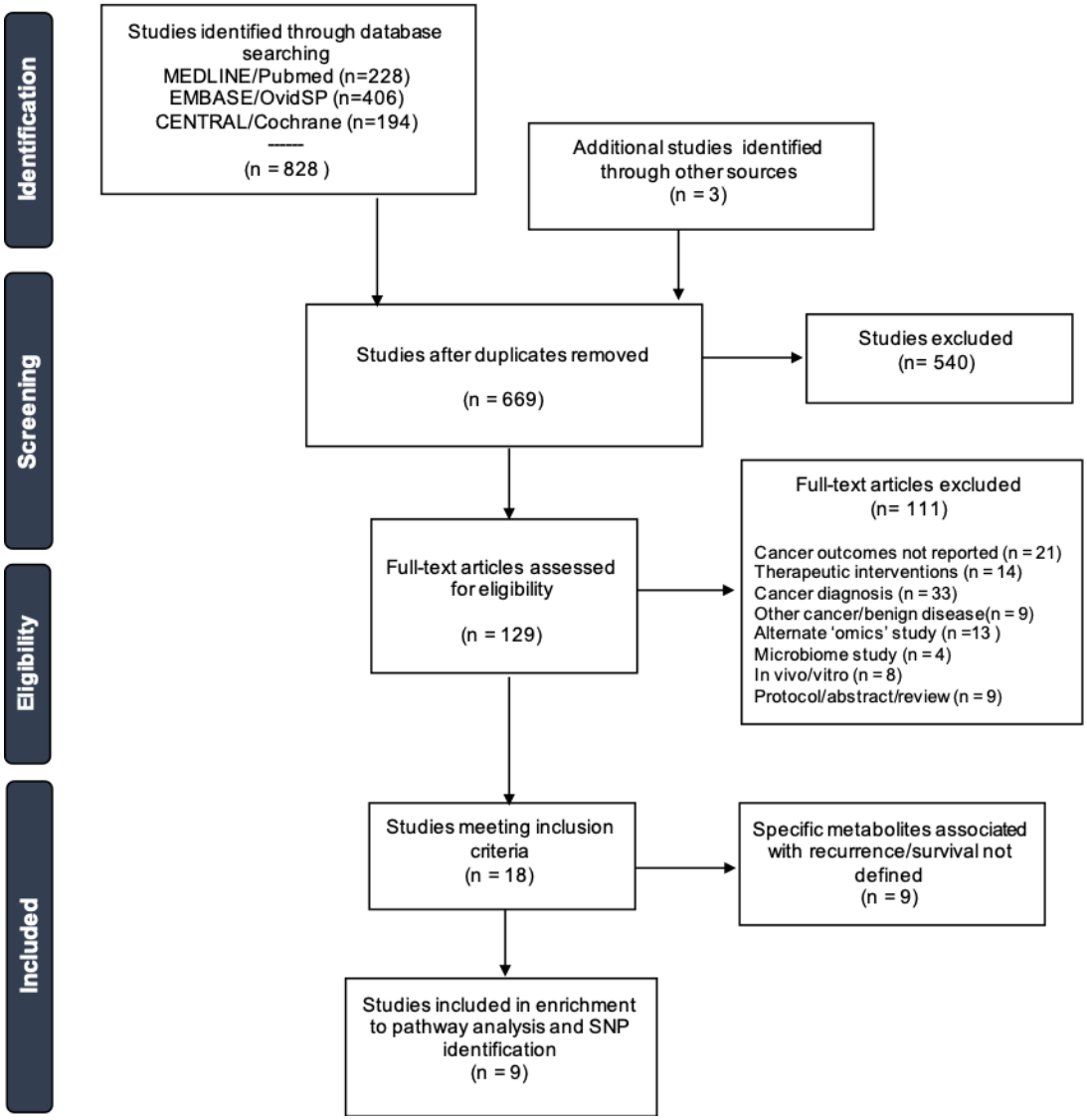


Figure 2: Metabolite set enrichment analysis (MSEA) of metabolites associated colorectal cancer (A) recurrence and (B) survival. Bar colours indicate difference in significance and bar lengths indicate different fold enrichment (number of fold greater than occurrence by chance). FDR=false discovery rate.

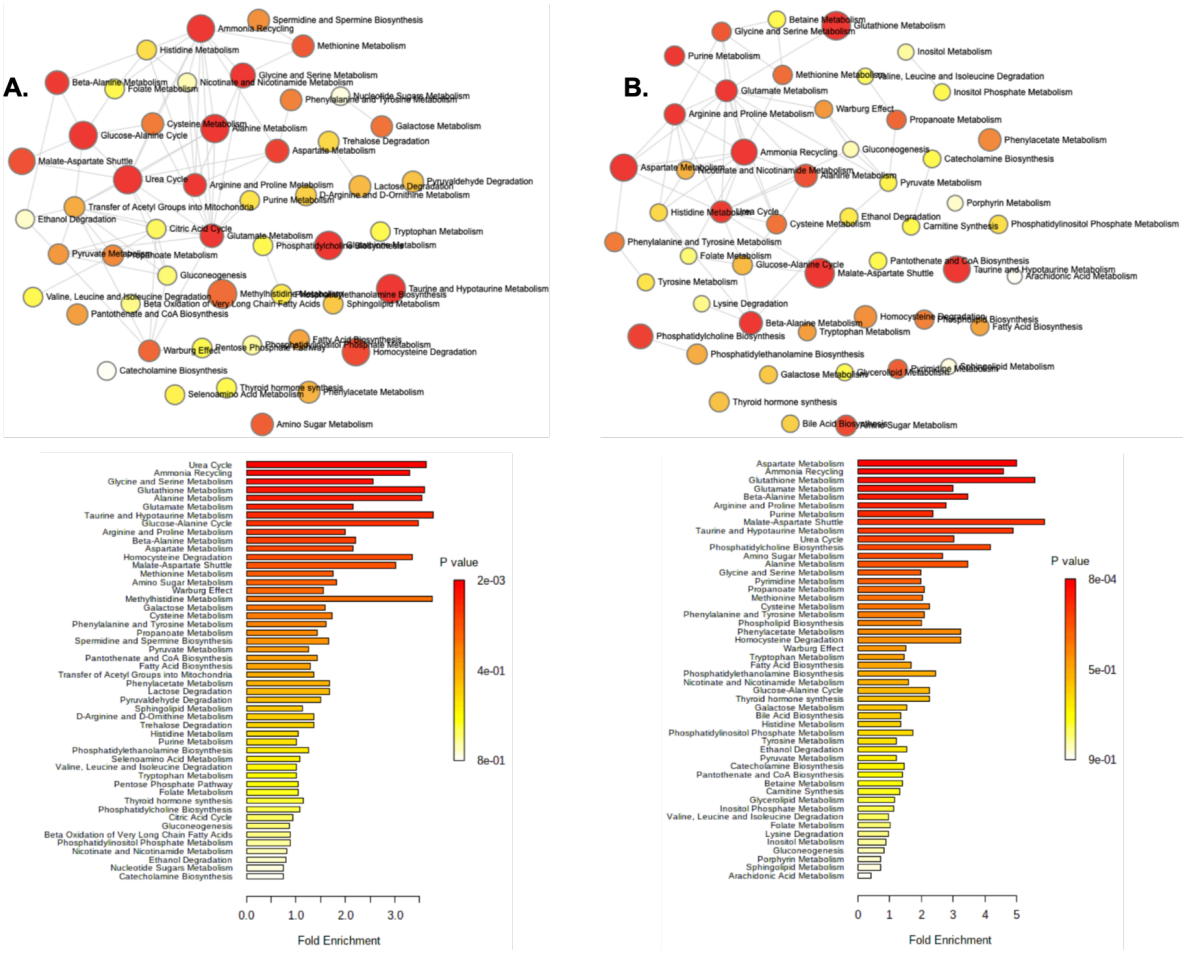
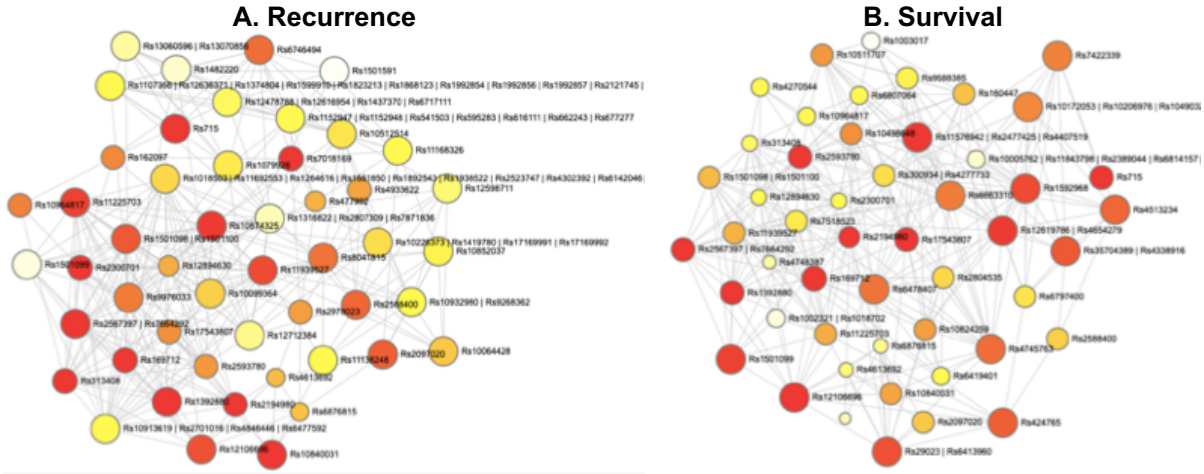


Figure 3: Single nucleotide polymorphisms identified as prognostic in colorectal cancer based on metabolic signatures. p-value is encoded by intensity of red colour of circles and size of the circles represent the diversity of genomic interactions based on the number of single nucleotide polymorphisms (SNPs) identified. Connecting lines indicate metabolite-related SNP relationships. rs=reference SNP.



Recurrence				
SNP	Hits	p	Holm p	FDR
Rs2194980	8	1.88E-8	**0.0001**	8.27E-5
Rs1392880	4	8.4E-7	**0.0037**	9.26E-4
Rs2567397 | Rs7664292	4	8.4E-7	**0.0037**	9.26E-4
Rs715	4	8.4E-7	**0.0037**	9.26E-4
Rs169712	5	4.1E-6	**0.0181**	0.002
Rs2300701	5	4.1E-	**0.0181**	0.002
Rs313408	5	4.1E-	**0.0181**	0.002
Rs7018169	5	4.1E-	**0.0181**	0.002

Survival				
SNP	Hits	p	Holm p	FDR
Rs2194980	8	4.13E-8	**1.82E-4**	1.82E-4
Rs169712	5	2.67E-7	**0.001**	5.88E-4
Rs12106696	3	3.7E-6	**0.016**	0.005
Rs1392880	4	1.47E-5	0.065	0.008
Rs17543807	4	1.47E-5	0.065	0.008
Rs2567397 | Rs7664292	4	1.47E-5	0.065	0.008
Rs2593780	4	1.47E-5	0.065	0.008
Rs715	4	1.47E-5	0.065	0.008

Appendix 1: QUADOMICS tool used for quality assessment

QUADOMICS

	STUDY ID:	Yes	No	Unclear	Not Applied
1	Were selection criteria clearly described?				
2	Was the spectrum of patients representative of patients who will receive the test in practice?				
3	Was the type of sample fully described?				
4	Were the procedures and timing of biological sample collection with respect to clinical factors described with enough detail?				
4.1	Clinical and physiological factors				
4.2	Diagnostic and treatment procedures				
5	Were handling and pre-analytical procedures reported in sufficient detail and similar for the whole sample? And, if differences in procedures were reported, was their effect on the results assessed?				
6	Is the time period between the reference standard and the index test short enough to reasonably guarantee that the target condition did not change between the two tests?				
7	Is the reference standard likely to correctly classify the target condition?				
8	Did the whole sample or a random selection of the sample receive verification using a reference standard of diagnosis?				
9	Did patients receive the same reference standard regardless of the result of the index test?				
10	Was the execution of the index test described in sufficient detail to permit replication of the test?				
11	Was the execution of the reference standard described in sufficient detail to permit its replication?				
12	Were the index test results interpreted without knowledge of the results of the reference standard?				
13	Were the reference standard results interpreted without knowledge of the results of the index test?				
14	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?				
15	Were uninterpretable/intermediate test results reported?				
16	Is it likely that the presence of overfitting was avoided?				
	TOTAL SCORE: %				

Appendix 2: Metabolites associated colorectal cancer recurrence and survival (pathway enrichment analysis). p-value is encoded by intensity of red colour of circles (representing pathways) and size of the circles represent the diversity of the pathway based on the number of metabolites contained in the pathway. SMPDB=Small Molecule Pathway Database (SMPDB). KEGG=Kyoto Encyclopaedia of Genes and Genomes (KEGG).