Association network analysis identifies enzymatic components of gut microbiota that significantly differ between colorectal cancer patients and healthy controls

Dongmei Ai [1,2*], Hongfei Pan[2], Xiaoxin Li[2], Min Wu[2], Li C. Xia[3*]

[1] Basic Experimental Center for Natural Science, University of Science and Technology Beijing, Xueyuan Road, Haidian District, Beijing, 100083, China

[2] School of Mathematics and Physics, University of Science and Technology Beijing, Xueyuan Road, Haidian District, Beijing, 100083, China

[3] Department of Medicine, Stanford University School of Medicine, 269 Campus Dr., Stanford, CA 94305, USA

Corresponding Author:

Dongmei Ai

No.30, Xueyuan Road, Haidian District, 100083, Beijing, China

aidongmei@ustb.edu.cn

Li C. Xia

269 Campus Dr., Stanford, CA 94305, USA

l.c.xia@stanford.edu

24 **Abstract**

25   The human gut microbiota plays a major role in maintaining human health and was recently

26 recognized as a promising target for disease prevention and treatment. Many diseases are

27 traceable to microbiota dysbiosis, implicating altered gut microbial ecosystems, or, in many

28 cases, disrupted microbial enzymes carrying out essential physio-biochemical reactions. Thus,

29 the changes of essential microbial enzyme levels may predict for human disorders. With the

30 rapid development of high-throughput sequencing technologies, metagenomics analysis has

31 emerged as an important method to explore the microbial communities in the human body, as

32 well as their functionalities. In this study, we analyzed 156 gut metagenomics samples from

33 patients with colorectal cancer (CRC) and adenoma, as well as that from healthy controls. We

34 estimated the abundance of microbial enzymes using the HMP Unified Metabolic Analysis

35 Network method (HUMAnN2) and identified the differentially abundant enzymes between

36 CRCs and controls. We constructed enzymatic association networks using the extended local

37 similarity analysis (ELSA) algorithm. We identified CRC-associated enzymic changes by

38 analyzing the topological features of the enzymatic association networks, including the clustering

39 coefficient, the betweenness centrality, and the closeness centrality of network nodes. The

40 network topology of enzymatic association network exhibited difference between the healthy and

41 the CRC environments. The ABC (ATP binding cassette) transporter and small subunit

42 ribosomal protein S19 enzymes, had the highest clustering coefficient in the healthy enzymatic

43 networks. In the contrast, the Adenosylhomocysteinase enzyme had the highest clustering

44 coefficient in the CRC enzymatic networks. These enzymic and metabolic differences may serve

45 as risk predictors for colorectal cancers and are worthy of further research.

46

Introduction

The human body harbors hundreds of trillions of microorganisms (Huttenhower et al. 2012; Turnbaugh et al. 2007). These microbial communities, comprising of bacteria, fungi, archaea, and viruses, are highly complex. These microbes colonize internal and external surfaces, such as mouth, esophagus, stomach, colon, respiratory tract, genitourinary tract and skin (Grice et al. 2009). In particular, the disruption of gut microbiota has been linked to a number of gastrointestinal diseases (Cotter et al. 2013). With the rapid development of high-throughput sequencing technologies, the metagenomics approach allowed us to extensively investigate the human microbiota in their naturally occurring state. These findings will ultimately lead to a better understanding of the gut microbiota in relation to the onset of gastrointestinal (GI) diseases (Sung et al. 2016). The colonic microbiota also might promote colorectal cancer by eliciting host responses (Cho & Blaser 2012).

Gut microbiota was known to play an active role in gut homeostasis for many years. The gut microbial composition and metabolic activity can modify the host's susceptibility to diseases via diverse pathologies. For instance, the disruption of gut microbial communities   was implicated with colorectal cancer (CRC) (Cipe et al. 2015). In particular, the alterations in the composition, distribution, and metabolism of colon microbiota were shown to disrupt normal homeostasis and lead to the onset of inflammation, dysplasia, and eventually cancer (Abreu & Peek Jr 2014; Nistal et al. 2015). More recently, many specific species whose abundance were positively associated with CRC have been identified. Those included *Streptococcus bovis, Helicobacter pylori, Bacteroides fragilis, Enterococcus faecalis, Clostridium septicum, Fusobacterium* spp., and *Escherichia coli* (Boulangé et al. 2016; Gagnière et al. 2016).

However, these earlier works have primarily analyzed the species abundance and diversity differences among communities to reveal the associations between gut microbiota and CRC (Halfvarson et al. 2017; Sobhani et al. 2011). It was only recently, researchers begun to study microbes at the functional level. The development of metabolomics has markedly facilitated such endeavors. This is made possible because the microbial abundance change will disrupt microbial enzymes that carry out essential physio-biochemical reactions. Such enzymes, as profiled by metabolomics, thus may predict for human disorders. For examples, in the metabolic disorders, such as IBD or obesity, the changes in gut microbiota cause a state of imbalance in metabolic activity and lead to the diseases (Chang et al. 2015; Tsuda et al. 2010; Weir et al. 2013).

Identifying such systematic enzymatic changes requires a high resolution analysis of

82   microbial metabolic networks. Two primary methods are available to construct enzymatic
83   networks: the constraint-based method (Borenstein 2012) and the topological theory-based
84   method. The constraint-based method requires establishing a series of ecological constraint
85   conditions, which requires iterative approximation of unknown parameters in turn. In contrast,
86   the topological theory-based method constructs the network directly by pulling the annotated
87   metabolic and biochemical reaction pathways from databases and removing redundant
88   information from these pathway data as necessary.
89
90   Alternatively, in this paper, we demonstrated an approach to construct association networks
91   by computing the local association coefficients of enzymes involved in biochemical reactions
92   and constructing the enzymatic networks using these statistically significant local associations as
93   edges and nodes. Using such derived enzymatic networks, we studied the differentially presented
94   microbial enzymes and metabolic activities using topological analysis. Applying our approach to
95   a large CRC dataset, we identified that the relative abundance of the enzymatic components of
96   *Bifidobacterium, Clostridium, Fusobacterium nucleatum, Porphyromonas gingivalis,* and
97   *Eubacterium*, were significantly higher in the adenoma and colorectal cancer patients, as
98   compared to the healthy controls.
99

Materials and methods

Cohort, fecal samples and metagenomic sequencing

102   The metagenomic dataset (Zeller et al. 2014) used in this study consists of 156 fecal samples of
103   randomly selected volunteers recruited from the Henri Mondor Hospital (Creteil, France),
104   including 61 healthy people, 42 patients with colorectal adenoma, and 53 colorectal cancer
105   patients. The dataset were downloaded from the EBI database, where the detailed data description
106   can be found: https://www.ebi.ac.uk/ena/data/view/PRJEB6070.

Estimation of enzyme abundance

108   Raw metagenomics data were translated to functional enzymes using the Kyoto Encyclopedia of
109   Genes and Genomes (KEGG) Ortholog database and the HMP Unified Metabolic Analysis
110   Network 2 (HUMAnN2) software, which is a pipeline for efficient and accurate profiling of
111   abundance in microbial pathways from a community based on metagenomic or
112   metatranscriptomic sequencing data (Abubucker et al. 2012). We computed the difference of
113   relative abundances for 2400 microbial enzymes involved in biochemical reactions between the
114   CRC patients and the healthy controls. The results were standardized to eliminate potential bias
115   in sequencing batches, and to remove any batch effects in enzyme abundances either specific to

116   CRC patients or healthy controls.

117   HUMAnN2 uses an in-house script to normalize the abundance of enzymes. The script

118   provides a method called 'total sum scaling' (TSS) normalization. The relative abundance is then

119   calculated as follows:

120
$$Ra_i = \frac{a_i}{\sum_j a_j}$$

121   where Ra_i is the relative abundance of the i-th enzyme and a_i is its absolute abundance.

122

123   **Enzymatic association network construction**

124   We performed enzymatic association network construction using the Extended Local Similarity

125   Analysis (ELSA) algorithm. ELSA was designed and developed by Xia et al. (2011) which uses

126   a dynamic programming algorithm to effectively discover potential local and time-delayed

127   associations of time-series and cross-sectional data. The dynamic programming algorithm can

128   find all potential global, local and time-delayed associations between two series and identify the

129   association with the highest similarity score as the maximum association between them. We used

130   a false discovery rate Q-value cutoff of 0.05 to assess the statistical significance. The resulting

131   enzymatic association networks representing statistically significant associations within the three

132   environments were visualized by the Cytoscape (Shannon et al. 2003) software, in which nodes

133   with blue borders in the three association networks represent enzymes enriched in the healthy

134   controls, while yellow nodes represent enzymes enriched in colorectal cancer patients.

135

136   The ELSA algorithm has been extensively used in the association network analysis of

137   microbial ecological data with a good performance (Weiss et al. 2016). Studies using ELSA have

138   discovered symbiotic relationships among microbes and relationships between microbes and the

139   environment that could not be identified by conventional correlation methods. For instance,

140   Shade et al. found dynamic mixed associations between bacteria in lakes by using ELSA (Shade

141   et al. 2010). Ki et al. examined the relationship between bacterial community structure and odor

142   emissions during the degradation of soil and pig carcasses by using ELSA. This algorithm not

143   only finds the associations between delayed data, but also effectively calculates the pairwise

144   global associations between for multivariate series data (Ki et al. 2018).

145

146   **Network topological analysis**

147   Several topological measures were used in our network analysis, including the clustering

148   coefficient and the betweenness and closeness centralities. The clustering coefficient is a

149   measure of the degree to which nodes in a graph tend to cluster together. The clustering

150   coefficient of the nodes represents their proximity in the network. An increased clustering

151   coefficient correlates with the higher tightness of the cluster involving the node and its

152   neighbors, hence, the node's importance. This feature of clustering coefficient allows its use for

153   the identification of key enzymes. The clustering coefficient is defined as

154
$$C_v = \frac{2n}{k(k-1)} \quad ,$$

155   where n is the number of edges between all the k neighbors of node v.

156

157   The betweenness centrality is a topological measure that refers to the number of times a node

158   acts as a bridge along the shortest path between two other nodes. As the betweenness centrality

159   of a node increases, the frequency with which the node acts as a "mediator" between other nodes

160   also increases, indicating the importance of the node. In a enzymatic association network, a high

161   betweenness centrality indicates that a node plays an important linking role and is likely an

162   important disease-related enzyme.

163   We also used the closeness centrality measure, which is the inverse of the distance from the

164   node to all other nodes in the network. The more important nodes in a network generally have a

165   higher closeness centrality, because they tend to locate close to the center of the network from

166   the geometrical perspective. We identified the shortest paths within the network using the

167   Kruskal's algorithm as implemented in the R package *plyr*.

168

Statistical analysis

169

170   We tested for the statistically significant abundance difference of enzymes associated with

171   biochemical reactions of gut microbiota between the healthy controls, and adenoma or colorectal

172   cancer patients by the two-tailed Wilcoxon rank-sum test.

Results and Discussion

Analysis of disease-associated enzymes

We got 157 differentially abundant enzymes among healthy control, adenoma and CRC patients. We selected 13 enzymes with differentially node degree among healthy control, adenoma and CRC patients association network. In Figure 1, we identified 13 enzymes showing statistically significant changes in abundance levels. We conducted a detailed analysis of differential enzymes (IDs and names listed in Table 1) and associated microorganisms to better understand the relationships between microbial metabolic functions and human disease. The enzymes K02003 (ABC transport system ATP-binding protein, P= 5.22E-04), K06147 (ATP-binding cassette, subfamily B, bacterial, P= 9.80E-05), K02025(Multiple sugar transport system permease protein, P= 1.74E-06), K02965 (Small subunit ribosomal protein S19, P= 1.17E-04), K02470(DNA gyrase subunit B,P= 3.62E-04), K01624(Fructose-bisphosphate aldolase, P= 2.90E-04) , K02878 (Large subunit ribosomal protein L16, P= 3.45E-04), K04043(Molecular chaperone DnaK, P= 4.66E-03), K09157(Uncharacterized protein, P= 1.46E-05), K01854(UDP-galactopyranose mutase, P= 4.31E-06), K03091(RNA polymerase sporulation-specific sigma factor, P= 2.10E-03) were all significantly higher in healthy controls. The enzymes K01667(Tryptophanase, P= 4.57E-04) and K01251(Adenosylhomocysteinase, P= 2.37E-03) were all significantly higher in CRC patients. These difference trends were also observed between adenoma and CRC patients, however, with less significant P-values, suggesting that though adenoma is an intermediary stage in the CRC development, the enzymatic levels observed in adenoma patients resemble more the healthy controls' as compared to that of the CRC patients.

K02003 is the ABC (ATP binding cassette) transporter, and this protein showed a higher relative abundance in the healthy environment in comparison to either the adenoma or the colorectal environment. As one of the largest metabolic transport systems in humans, the ABC (ATP binding cassette) transporter system mainly transfers nutrients, biosynthetic precursors, trace metals, and vitamins, but it also transports lipids, drugs, and primary and secondary metabolites. It plays a major role in biosynthetic pathways. It was noted that low ATP-binding cassette protein subfamily (ABCB1,P-glycoprotein) protein levels may promote colorectal carcinogenesis (Andersen et al. 2013). ABCB1 protein levels were also found to be lower in CRC tissue as compared to well-differentiated tissue (De Iudicibus et al. 2008).

In another example, K01854 (UDP-galactopyranose mutase (UGM), an enzyme that

207 catalyzes chemical reactions) had a higher abundance in healthy samples compared to the other
208 two groups (Figure 1). This enzyme is specifically involved in galactose metabolism, as well as
209 amino sugar and nucleotide sugar metabolism, and it plays a major role in the production of
210 cellular energy and the modification of proteins and glycolipids. In an early study, Petry et al.
211 indicated the importance of galactose metabolism in humans. Galactose can be converted into
212 energy. UGM is one of the main enzymes involved in galactose metabolism. Galactose
213 metabolism is crucial for the health of both neonatal development and adults. Disruption of the
214 production of this enzyme will affect galactose metabolism and even cause transferase-
215 deficiency galactosemia in severe cases (Petry & Reichardt 1998). Brown et al. have shown the
216 presence of galactose metabolism disorders in colorectal cancer patients. Based on our metabolic
217 association network, we have shown that the abundance of K01854 is lower in the colorectal
218 cancer patients than the healthy controls, in agreement with previous findings (Brown et al.
219 2016).

220

221   We also found that UDP-galactopyranose mutase, a main enzyme involved in galactose
222 metabolism, was more abundant in healthy people. The result again indicated that the reduced
223 galactose metabolism was associated with the development of colorectal cancer. The main
224 enzymatic active genera of galactose metabolism are *Bacteroides* spp. and *Bifidobacterium* spp..
225 The *Bacteroides* spp. participates in polysaccharide degradation and the *Bifidobacterium* spp.
226 participates in galactose metabolism in the gut. Thus, both of these genera may exert protective
227 effects against developing colorectal cancer.

228

229   There are enzymes had significantly higher abundances in cancer patients. The most
230 prominent enzymes of the kind were K01251 (adenosylhomocysteinase) and K01667
231 (tryptophanase) (Figure 1). Therefore, the microbial species producing adenosylhomocysteinase,
232 an enzyme involved in tryptophan metabolism, may facilitate tumor progression as it presented a
233 significantly higher abundance in cancer patients. By using gas chromatography, Kim et al.
234 found adenosylhomocysteinase to be a protein marker for colorectal cancer (Kim et al. 2009; Yin
235 et al. 2013). Interestingly, our enzymatic association network also showed an enriched
236 abundance of K01667 (tryptophanase) in colorectal cancer samples. Tryptophanase is a basic
237 amino acid, and tryptophanase metabolism can help cancer escape immune surveillance.
238 Indoleamine 2,3-dioxygenase 1 (IDO1) is a tryptophan-catabolizing enzyme and the main
239 enzyme expressed in malignant inflamed gut. In addition, clinical data have shown that
240 tryptophan metabolism promotes tumor progression (Santhanam et al. 2016).

241

242 **Topological attributes of enzymatic association networks**

243 In Figure S1, we showed the global enzymatic association networks we constructed from the
244 ELSA analysis of the enzymic abundance values. Meanwhile, we exhibited the enzymatic
245 network topological attributes of thirteen enzymes in Table 2, including the clustering
246 coefficient, and the closeness and betweenness centralities. As we can see, in the order from
247 healthy, to adenoma and to colorectal cancer samples, there is a decreasing trend of node degree
248 among the enzymes enriched in healthy controls (nodes with blue circles). Here, the degree of a
249 node refers to the number of neighboring nodes that the node has. Such decreasing trend
250 suggested that these enzymes were highly cooperating in the metabolic processes in health
251 controls (Figure S1A), while the level of cooperation was reduced in adenoma (Figure S1B) and
252 further reduced to minimum in the colorectal cancer environment (Figure S1C). Notably, a
253 similar decreasing trend was observed in the clustering coefficient (Table 2) of healthy control
254 enriched enzymes, which suggested a gradual loss of importance of these healthy enzymes in a
255 pathogenic CRC gut microbial environment.

256

257 The enzyme K02003, an ABC (ATP binding cassette) transporter, had the highest clustering
258 coefficients in the healthy enzymatic network. This ABC transporter was also identified as
259 significantly more abundant in healthy samples by abundance level (see Figure 1). Both results
260 consistently suggested its status as a signature enzyme species for the healthy gut microbial
261 environment. Similarly, the clustering coefficient of enzyme K02965 (small subunit ribosomal
262 protein S19) was one in the healthy network; however, the coefficient was zero in the colorectal
263 cancer network, also suggesting that the enzyme has an important biochemical role in the healthy
264 people, which was no longer found in cancer patients.

265

266 In the contrast, for the enzymes enriched in the colorectal cancer patients, they showed an
267 increasing connectivity from healthy, to adenoma and to colorectal cancer enzymatic networks.
268 Several such enzymes were uniquely presented in the CRC network. For instance, the K01251
269 enzyme (Adenosylhomocysteinase), which is a highly connected node in the CRC and adenoma
270 network, however, did not show up in the healthy network at all (Figure 1), suggesting its
271 functional role in the early stage of tumorigenesis. We also found that the betweenness centrality
272 of K01251 was higher in the colorectal cancer environment, and zero in the healthy controls (see
273 Table 2). These results suggest that K01251 may play a negative role in intestinal flora
274 metabolism causing its enrichment in the colorectal cancer patients.

275

276

Conclusion

278 We analyzed a large human gut metagenomics dataset by using the HUMAnN2 tool to estimate
279 the relative abundances of microbially produced enzymes in the healthy control, the adenoma
280 and the colorectal cancer patient samples, respectively. We identified differentially abundant
281 enzymes among these healthy, adenoma, and colorectal cancer patient groups. We constructed
282 enzymatic association networks representing the healthy, adenoma, and colorectal cancer
283 microbial environments using the ELSA algorithm, and analyzed the topological attributes of the
284 resulting networks.

285

286 The new enzymatic network analysis approach we took in this study addressed the issue in
287 previous studies that the enzymatic association networks were constructed only for individual
288 metabolic pathways involving a limited number of microorganisms. We were able to globally
289 integrated thousands of enzymes and hundreds of metabolic substrates carrying out biochemical
290 reactions. These microbial enzymatic networks represent the set of essential relationships
291 between enzymes of microbial origin. These relationships were indicators of underlying
292 biochemical reactions and production of metabolites. Among the methods available to study such
293 relationships between microbes and metabolites, only this systematic network-based study can
294 reveal these relationships between microbial communities and their host in high resolution.

295

296 The innovative point of our network construction is to use an extensively tested and validated
297 metric – local similarity score for building edges in the network. We know various noises and
298 biases are associated with large-scale network construction, which incur challenges in
299 subsequent topological analysis. And many constructed enzymatic networks are not specific
300 simply because of the presence of irrelevant metabolites. The use of an accurate and robust
301 association metrics helped us in truthfully representing these metabolic interactions.

302

303 Moreover, the topological attributes of the enzymatic network can carry important
304 information of underlying dynamics between the enzymes and the tumorigenesis process. For
305 example, the clustering coefficient is a positive indicator of an enzyme's cooperativity with other
306 enzymes. By systematically constructing the networks using the ELSA algorithm and data-
307 mining in node degree, clustering coefficient, betweenness and closeness centralities of the
308 enzymatic network under different tutorial stages, we were able to identify signature enzymic
309 species as well as the local shift in network structure of cooperating enzymes as potential cancer

310   risk markers, which merit further research.

311

312   **Acknowledgements**

313   Dongmei Ai thanks Professor Fengzhu Sun at the University of Southern California. Li.C.Xia.

314   thanks Dr. Nancy Zhang at the University of Pennsylvania and Dr. Hanlee Ji at Stanford

315   University for their support and helpful discussions.

316

317   **References**

318   Abreu MT, and Peek Jr RM. 2014. Gastrointestinal malignancy and the microbiome.
319   Gastroenterology 146:1534-1546. e1533.

320   Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL, Rodriguez-Mueller B, Zucker
321   J, Thiagarajan M, and Henrissat B. 2012. Metabolic reconstruction for metagenomic data
322   and its application to the human microbiome. PLoS computational biology 8:e1002358.

323   Andersen V, Vogel U, Godiksen S, Frenzel FB, Sæbø M, Hamfjord J, Kure E, and Vogel LK.
324   2013. Low ABCB1 gene expression is an early event in colorectal carcinogenesis. PloS
325   one 8:e72119.

326   Borenstein E. 2012. Computational systems biology and in silico modeling of the human
327   microbiome. Briefings in bioinformatics 13:769-780.

328   Boulangé CL, Neves AL, Chilloux J, Nicholson JK, and Dumas M-E. 2016. Impact of the gut
329   microbiota on inflammation, obesity, and metabolic disease. Genome medicine 8:42.

330   Brown DG, Rao S, Weir TL, O'Malia J, Bazan M, Brown RJ, and Ryan EP. 2016. Metabolomics
331   and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and
332   stool. Cancer & metabolism 4:11.

333   Chang C-H, Qiu J, O'Sullivan D, Buck MD, Noguchi T, Curtis JD, Chen Q, Gindin M, Gubin
334   MM, and van der Windt GJ. 2015. Metabolic competition in the tumor microenvironment
335   is a driver of cancer progression. Cell 162:1229-1241.

336   Cho I, and Blaser MJ. 2012. The human microbiome: at the interface of health and disease. Nature
337   Reviews Genetics 13:260.

338   Cipe G, Idiz UO, Firat D, and Bektasoglu H. 2015. Relationship between intestinal microbiota and
339   colorectal cancer. World J Gastrointest Oncol 7:233-240.

340   Cotter PD, Ross RP, and Hill C. 2013. Bacteriocins—a viable alternative to antibiotics? Nature
341   Reviews Microbiology 11:95.

342   De Iudicibus S, De Pellegrin A, Stocco G, Bartoli F, Bussani R, and Decorti G. 2008. ABCB1
343   gene polymorphisms and expression of P-glycoprotein and long-term prognosis in
344   colorectal cancer. Anticancer research 28:3921-3928.

345   Gagnière J, Raisch J, Veziant J, Barnich N, Bonnet R, Buc E, Bringer M-A, Pezet D, and Bonnet
346   M. 2016. Gut microbiota imbalance and colorectal cancer. World journal of
347   gastroenterology 22:501.

348   Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, Bouffard GG, Blakesley RW,

349   Murray PR, and Green ED. 2009. Topographical and temporal diversity of the human skin
350   microbiome. science 324:1190-1192.
351   Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, Bramer LM, D'Amato
352   M, Bonfiglio F, McDonald D, and Gonzalez A. 2017. Dynamics of the human gut
353   microbiome in inflammatory bowel disease. Nature microbiology 2:17004.
354   Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl
355   AM, FitzGerald MG, and Fulton RS. 2012. Structure, function and diversity of the healthy
356   human microbiome. Nature 486:207.
357   Ki B-M, Ryu HW, and Cho K-S. 2018. Extended local similarity analysis (eLSA) reveals unique
358   associations between bacterial community structure and odor emission during pig carcasses
359   decomposition. Journal of Environmental Science and Health, Part A:1-10.
360   Kim H-J, Kang HJ, Lee H, Lee S-T, Yu M-H, Kim H, and Lee C. 2009. Identification of S100A8
361   and S100A9 as serological markers for colorectal cancer. Journal of proteome research
362   8:1368-1379.
363   Nistal E, Fernández-Fernández N, Vivas S, and Olcoz JL. 2015. Factors determining colorectal
364   cancer: the role of the intestinal microbiota. Frontiers in oncology 5:220.
365   Petry KG, and Reichardt JK. 1998. The fundamental importance of human galactose metabolism:
366   lessons from genetics and biochemistry. Trends in Genetics 14:98-102.
367   Santhanam S, Alvarado DM, and Ciorba MA. 2016. Therapeutic targeting of inflammation and
368   tryptophan metabolism in colon and gastrointestinal cancer. Translational Research
369   167:67-79.
370   Shade A, Chiu CY, and McMahon K. 2010. Differential bacterial dynamics promote emergent
371   community robustness to lake mixing: an epilimnion to hypolimnion transplant
372   experiment. Environmental microbiology 12:455-466.
373   Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and
374   Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular
375   interaction networks. Genome research 13:2498-2504.
376   Sobhani I, Tap J, Roudot-Thoraval F, Roperch JP, Letulle S, Langella P, Corthier G, Van Nhieu
377   JT, and Furet JP. 2011. Microbial dysbiosis in colorectal cancer (CRC) patients. PloS one
378   6:e16393.
379   Sung J, Hale V, Merkel AC, Kim P-J, and Chia N. 2016. Metabolic modeling with Big Data and
380   the gut microbiome. Applied & translational genomics 10:10-15.
381   Tsuda H, Ochiai K, Suzuki N, and Otsuka K. 2010. Butyrate, a bacterial metabolite, induces
382   apoptosis and autophagic cell death in gingival epithelial cells. Journal of periodontal
383   research 45:626-634.
384   Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, and Gordon JI. 2007. The
385   human microbiome project. nature 449:804.
386   Weir TL, Manter DK, Sheflin AM, Barnett BA, Heuberger AL, and Ryan EP. 2013. Stool
387   microbiome and metabolome differences between colorectal cancer patients and healthy
388   adults. PloS one 8:e70803.
389   Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, Xia LC, Xu ZZ, Ursell L,

390   and Alm EJ. 2016. Correlation detection strategies in microbial data sets vary widely in
391   sensitivity and precision. The ISME journal 10:1669.
392 Xia LC, Steele JA, Cram JA, Cardon ZG, Simmons SL, Vallino JJ, Fuhrman JA, and Sun F. 2011.
393   Extended local similarity analysis (eLSA) of microbial community and other time series
394   data with replicates. BMC systems biology: BioMed Central. p S15.
395 Yin H-R, Zhang L, Xie L-Q, Huang L-Y, Xu Y, Cai S-J, Yang P-Y, and Lu H-J. 2013. Hyperplex-
396   MRM: a hybrid multiple reaction monitoring method using mTRAQ/iTRAQ labeling for
397   multiplex absolute quantification of human colorectal cancer biomarker. Journal of
398   proteome research 12:3912-3919.
399 Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, and

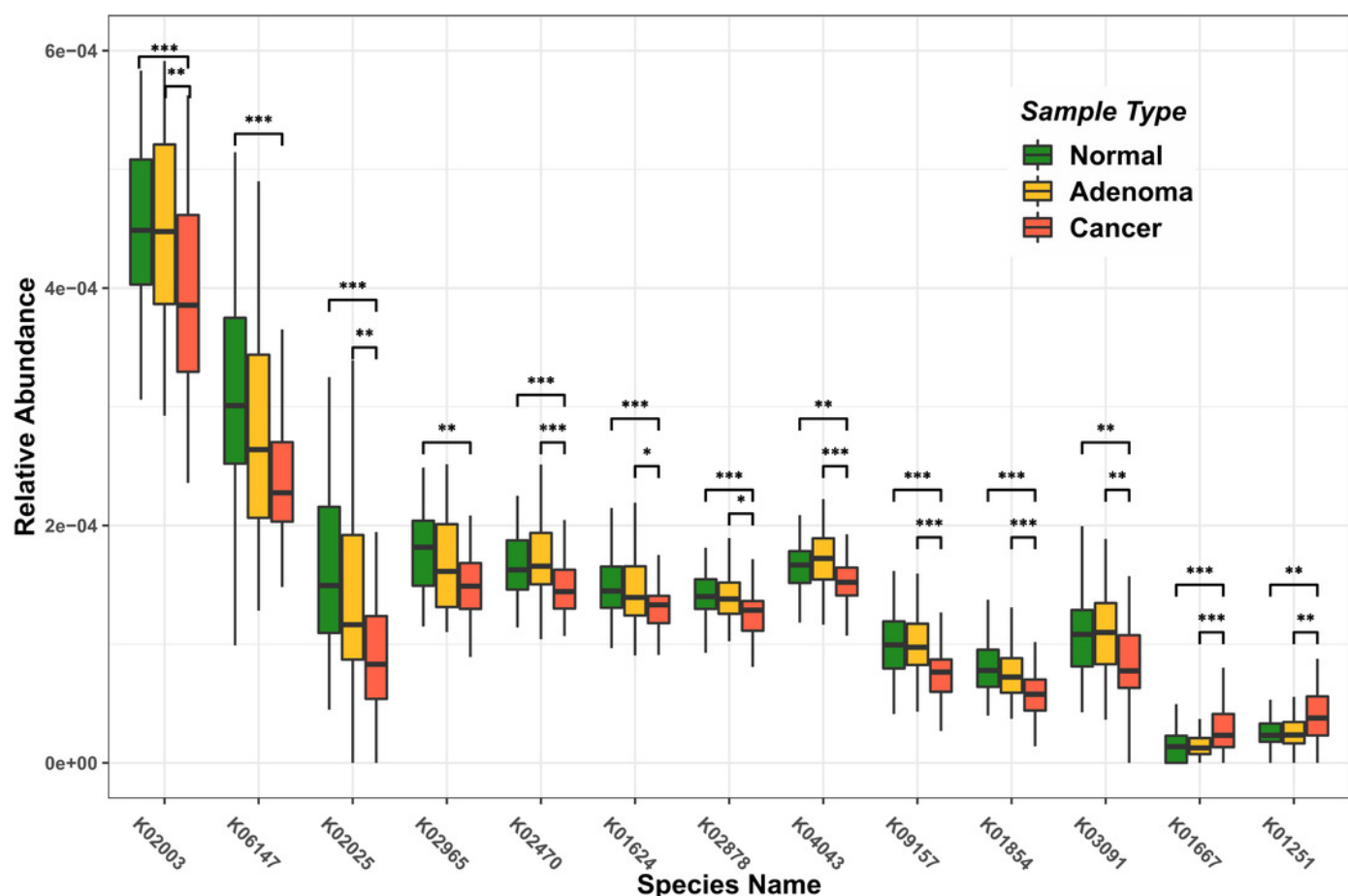
400   Habermann N. 2014. Potential of fecal microbiota for early‐stage detection of colorectal

401   cancer. Molecular systems biology 10:766.

402

Figure 1

Enzymes with significant differences of relative abundance in healthy people, adenoma patients, and colorectal cancer patients.

Green bar is the relative abundance of enzymes of healthy people. Orange bar is the relative abundance of enzymes of adenoma patients. Red bar is the relative abundance of enzymes of colorectal cancer patients. *** represent $p<0.001$, ** represent $p<0.01$, * represent $p<0.05$.

Enzyme annotations

ID	Name of Enzyme
K02003	ABC transport system ATP-binding protein
K06147	ATP-binding cassette, subfamily B, bacterial
K02025	Multiple sugar transport system permease protein
K02965	Small subunit ribosomal protein S19
K02470	DNA gyrase subunit B
K01624	Fructose-bisphosphate aldolase
K02878	Large subunit ribosomal protein L16
K04043	Molecular chaperone DnaK
K09157	Uncharacterized protein
K01854	UDP-galactopyranose mutase
K03091	RNA polymerase sporulation-specific sigma factor
K01667	Tryptophanase
K01251	Adenosylhomocysteinase

Topological attributes of 13 nodes of metabolic association networks in the healthy, adenoma, and colorectal cancer environments

The data range in the table is between [0,1].

ID	Clustering Coefficient			Closeness Centrality			Betweenness Centrality		
	Healthy	Adenoma of Gut	CRC	Healthy	Adenoma of Gut	CRC	Healthy	Adenoma of Gut	CRC
K09157	7.42E-01	6.70E-01	4.72E-01	5.66E-01	4.62E-01	4.14E-01	8.76E-03	4.58E-03	4.62E-03
K06147	7.79E-01	7.17E-01	5.49E-01	6.09E-01	6.10E-01	5.03E-01	3.75E-03	5.71E-03	8.28E-03
K04043	7.81E-01	7.33E-01	5.00E-01	5.72E-01	5.12E-01	4.17E-01	3.98E-03	9.93E-03	1.73E-03
K03091	5.85E-01	6.65E-01	7.20E-01	4.97E-01	5.31E-01	4.95E-01	4.32E-03	3.07E-03	6.63E-03
K02965	1.00E+00	7.56E-01	0.00E+00	4.49E-01	4.38E-01	3.00E-01	0.00E+00	7.65E-05	0.00E+00
K02878	9.62E-01	6.98E-01	0.00E+00	4.61E-01	5.30E-01	3.63E-01	9.23E-06	2.89E-03	2.53E-04
K02470	7.31E-01	7.18E-01	6.43E-01	5.97E-01	5.35E-01	4.70E-01	8.81E-03	3.79E-03	3.20E-03
K02025	8.17E-01	6.12E-01	4.83E-01	6.06E-01	6.31E-01	5.47E-01	4.36E-03	1.03E-02	1.27E-02
K02003	8.81E-01	7.67E-01	8.02E-01	5.37E-01	5.67E-01	4.52E-01	1.20E-03	3.68E-03	7.40E-04
K01854	8.76E-01	8.07E-01	6.19E-01	5.31E-01	5.59E-01	4.45E-01	2.72E-03	1.65E-03	1.84E-03
K01667	7.33E-01	8.67E-01	5.16E-01	4.72E-01	3.65E-01	4.06E-01	2.82E-04	5.77E-05	3.82E-03
K01624	9.82E-01	7.93E-01	7.33E-01	5.02E-01	5.15E-01	3.85E-01	2.55E-05	1.82E-03	7.60E-04
K01251	0.00E+00	0.00E+00	5.90E-01	0.00E+00	4.38E-01	4.03E-01	0.00E+00	3.69E-04	2.32E-03