

ARTICLE

Received 23 May 2014 | Accepted 20 Jan 2015 | Published 28 Apr 2015

DOI: 10.1038/ncomms7342

# Fat, fibre and cancer risk in African Americans and rural Africans

Stephen J.D. O'Keefe<sup>1</sup>, Jia V. Li<sup>2</sup>, Leo Lahti<sup>3,4</sup>, Junhai Ou<sup>1</sup>, Franck Carbonero<sup>5,†</sup>, Khaled Mohammed<sup>1</sup>, Joram M. Posma<sup>2</sup>, James Kinross<sup>2</sup>, Elaine Wahl<sup>1</sup>, Elizabeth Ruder<sup>6</sup>, Kishore Vippera<sup>1</sup>, Vasudevan Naidoo<sup>7</sup>, Lungile Mtshali<sup>7</sup>, Sebastian Tims<sup>3</sup>, Philippe G.B. Puylaert<sup>3</sup>, James DeLany<sup>8</sup>, Alyssa Krasinskas<sup>9</sup>, Ann C. Benefiel<sup>5</sup>, Hatem O. Kaseb<sup>1</sup>, Keith Newton<sup>7</sup>, Jeremy K. Nicholson<sup>2</sup>, Willem M. de Vos<sup>3,4,10</sup>, H. Rex Gaskins<sup>5</sup> & Erwin G. Zoetendal<sup>3</sup>

Rates of colon cancer are much higher in African Americans (65:100,000) than in rural South Africans (<5:100,000). The higher rates are associated with higher animal protein and fat, and lower fibre consumption, higher colonic secondary bile acids, lower colonic short-chain fatty acid quantities and higher mucosal proliferative biomarkers of cancer risk in otherwise healthy middle-aged volunteers. Here we investigate further the role of fat and fibre in this association. We performed 2-week food exchanges in subjects from the same populations, where African Americans were fed a high-fibre, low-fat African-style diet and rural Africans a high-fat, low-fibre western-style diet, under close supervision. In comparison with their usual diets, the food changes resulted in remarkable reciprocal changes in mucosal biomarkers of cancer risk and in aspects of the microbiota and metabolome known to affect cancer risk, best illustrated by increased saccharolytic fermentation and butyrogenesis, and suppressed secondary bile acid synthesis in the African Americans.

<sup>1</sup> Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA.

<sup>2</sup> Department of Surgery and Cancer and Centre for Digestive and Gut Health, Institution of Global Health Innovation, Imperial College, London SW7 2AZ, UK.

<sup>3</sup> Laboratory of Microbiology, Wageningen University, Wageningen 6703 HB, The Netherlands. <sup>4</sup> Department of Veterinary Bioscience, University of Helsinki, Helsinki, Finland. <sup>5</sup> University of Illinois at Urbana-Champaign, Champaign, Illinois 61801, USA. <sup>6</sup> Division of Sports Medicine and Nutrition, School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. <sup>7</sup> University of KwaZulu-Natal, Durban, South Africa. <sup>8</sup> Division of Endocrinology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. <sup>9</sup> Division of Pathology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. <sup>10</sup> RPU Immunobiology, Department of Bacteriology and Immunology, University of Helsinki, Helsinki 00014, Finland. † Present address: Department of Food Science, University of Arkansas, Fayetteville, Arizona 72704, USA. Correspondence and requests for materials should be addressed to S.J.D.O'.K. (email: sjokeefe@pitt.edu).

In the west, colon cancer is the second leading cause of cancer death. It afflicts ~150,000 Americans, 250,000 Europeans and 1 million people worldwide annually. Nearly one third will die<sup>1</sup>. Colonoscopy has permitted early detection and recent studies have shown this to be associated with a reduction in mortality rates, but overall impact has been small, particularly among African Americans who shoulder the greatest burden of the disease in the United States. Colon cancer is one of the westernized diseases, with an incidence in Americans of African descent of 65:100,000 compared with <5:100,000 in rural Africans. Migrant studies, such as those in Japanese Hawaiians, have demonstrated that it only takes one generation for the immigrant population to assume the colon cancer incidence of the host western population<sup>2</sup>. Although the change in diet is most probably responsible for this<sup>3</sup>, migration changes many other aspects of the environment. For example, cigarettes, chemicals, infections and antibiotics might be equally responsible for the change in colon cancer risk. To focus on the importance of diet in African Americans and to explore the hypothesis that colon cancer risk is determined by the influence of the diet on the microbiota to produce anti- or pro-neoplastic metabolites, we have performed a series of investigations between African Americans and rural South Africans<sup>4,5</sup>. First, we observed that their diets were fundamentally different in preparation, cooking and composition. Animal protein and fat intake was two to three times higher in Americans, whereas carbohydrate and fibre, chiefly in the form of resistant starch, were higher in Africans. On colonoscopy, African Americans had more polyps and higher rates of mucosal proliferation, measured by Ki67 epithelial cell staining, confirming its potential use as a biomarker of cancer risk<sup>6</sup>. These differences were shown to be associated with profound differences in the microbiota (Americans dominated by genus *Bacteroides* and Africans by the genus *Prevotella*) and metabolic phenotype. Notable differences included higher levels of starch degraders, carbohydrate fermenters and butyrate producers and their metabolites in Africans, and higher levels of potentially pathogenic proteobacteria (*Escherichia* and *Acinetobacter*) and bile acid deconjugators and their products in Americans<sup>5</sup>. As there is extensive experimental evidence that the products of fibre fermentation, in particular butyrate, are anti-inflammatory and anti-neoplastic<sup>7-9</sup>, and that the products of bacterial bile acid conjugation, secondary bile acids, are carcinogenic<sup>10</sup>, these findings suggested two potential mechanisms for diet-associated cancer risk: the protective effect of dietary fibre in increasing butyrogenesis and the promotional effect of dietary fat on stimulating bile acid synthesis by the liver.

We know from the results of two recently published human studies that modifications of the fibre and fat content of the diets

have a major impact on the colonic microbiota within a few days<sup>11,12</sup>. Here we provide critical information on the functional significance of these microbiota changes on the colonic mucosa by switching the fat and fibre proportions of the diet in our two high and low colon cancer risk study populations. Employing the study design outlined on Table 1, we showed the anticipated increase in saccharolytic fermentation and butyrogenesis, and suppression of secondary bile acid synthesis produced by switching African Americans to a high-fibre, low-fat diet for 2 weeks, were associated with a significant reduction in colonic mucosal inflammation and proliferation biomarkers of cancer risk. In stark contrast, the diet switch in rural Africans to a high-fat, low-fibre diet resulted in reverse changes in all these parameters.

## Results

### Changes in the colonic mucosa following diet switch.

Measurements made in the home environment (HE) confirmed our preceding studies that the dietary intakes were quite different, with animal protein and fat intake being two to three times higher in Americans and fibre intake, chiefly in the form of resistant starch, being higher in Africans<sup>4</sup> (Supplementary Fig. 1 and Supplementary Tables 1–3). On colonoscopy, adenomatous polyps were found and removed in nine Americans and no Africans. Measurements of mucosal epithelial proliferation rates by Ki67 staining of biopsies from ascending, transverse and descending colon, all showed significantly higher levels in African Americans than in Africans (Fig. 1 and Supplementary Fig. 2), again supporting its use as a biomarker of cancer risk. The dietary switch was well tolerated by both populations and body weights were all maintained with 2 kg. Remarkably, the switch decreased proliferative rates in African Americans to levels lower than those of Africans at baseline, whereas rates increased in Africans to levels greater than African Americans at baseline (Fig. 1). In tandem, histological (Supplementary Tables 5a–d) and immunohistochemical markers of inflammation (Fig. 1b,c) used, namely CD3<sup>+</sup> intraepithelial lymphocytes and CD68<sup>+</sup> lamina propria macrophages, increased on switching Africans to the high-fat diet and decreased in Americans given the high-fibre diet, thus paralleling the measured changes in mucosal proliferation.

### Changes in colonic microbial metabolism following diet switch.

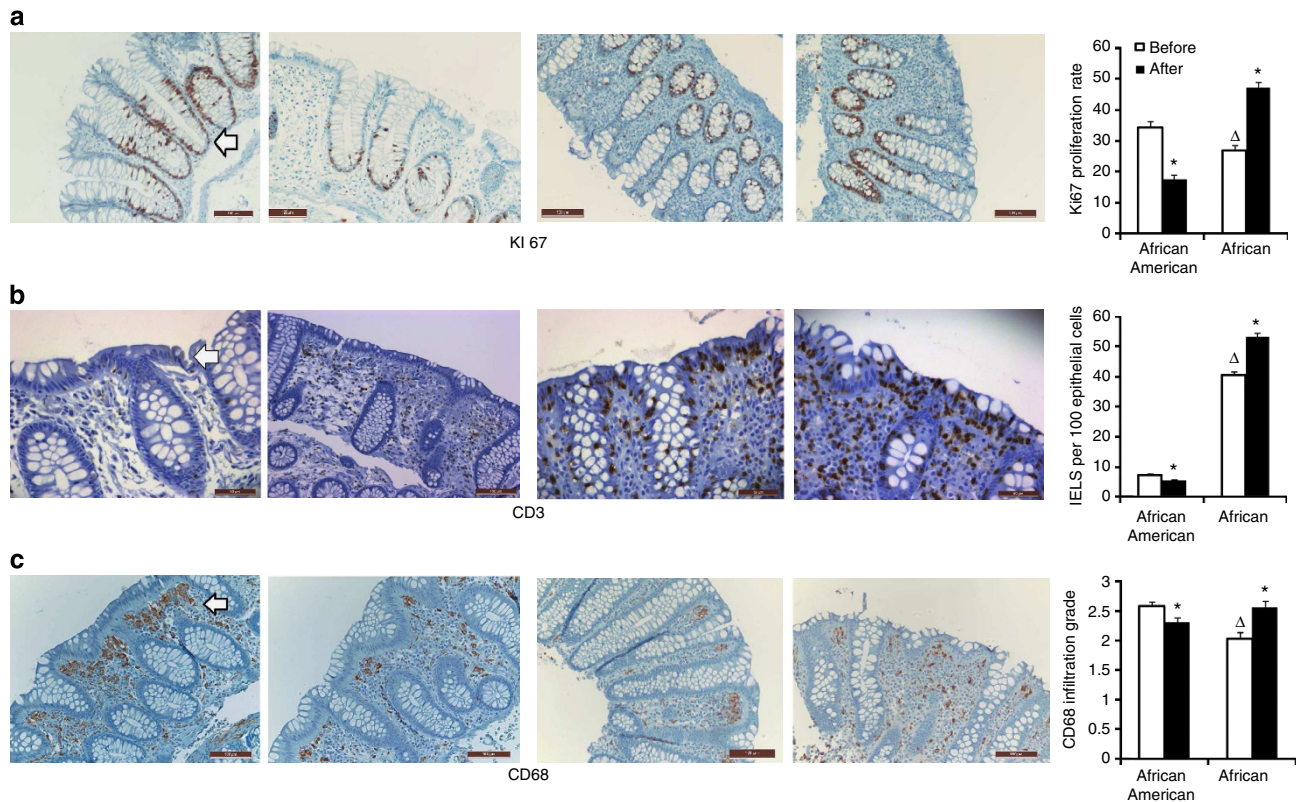
The reciprocal changes in mucosal proliferation and inflammation between Africans and Americans were associated with reciprocal changes in specific microbes and their metabolites that support our hypothesis that the mechanism whereby diet

**Table 1 | Study design, time schedule and sampling.**

| Day                                     | HE study<br>(at home) |      |      | DI<br>(in-house)        |      |      |
|---|-----------------------|------|------|-------------------------|------|------|
|   | 0                     | 7    | 14   | 15                      | 22   | 29   |
| Period                                  | ED 1                  | HE 1 | HE 2 | DI 1                    | DI 2 | ED 2 |
| Fecal samples                           | X                     | X    | X    | X                       | X    | X    |
| Colonic evacuates                       | X                     |      |      |                         |      | X    |
| Breath H <sub>2</sub> , CH <sub>4</sub> | X                     | X    | X    | X                       | X    | X    |
| Bloods                                  | X                     |      |      |                         |      | X    |
| Mucosal biopsies                        | X                     |      |      |                         |      | X    |
| Dietary observations                    |                       | XXX  |      | X X X X X X X X X X X X |      |      |

DI, dietary intervention periods; ED 1, initial endoscopy days; ED 2, final endoscopy days; HE, home environment.

The Xs corresponding to dietary evaluation indicate the frequency of dietary recordings: three days during the home environment when consuming their usual diets, and daily during the dietary intervention.



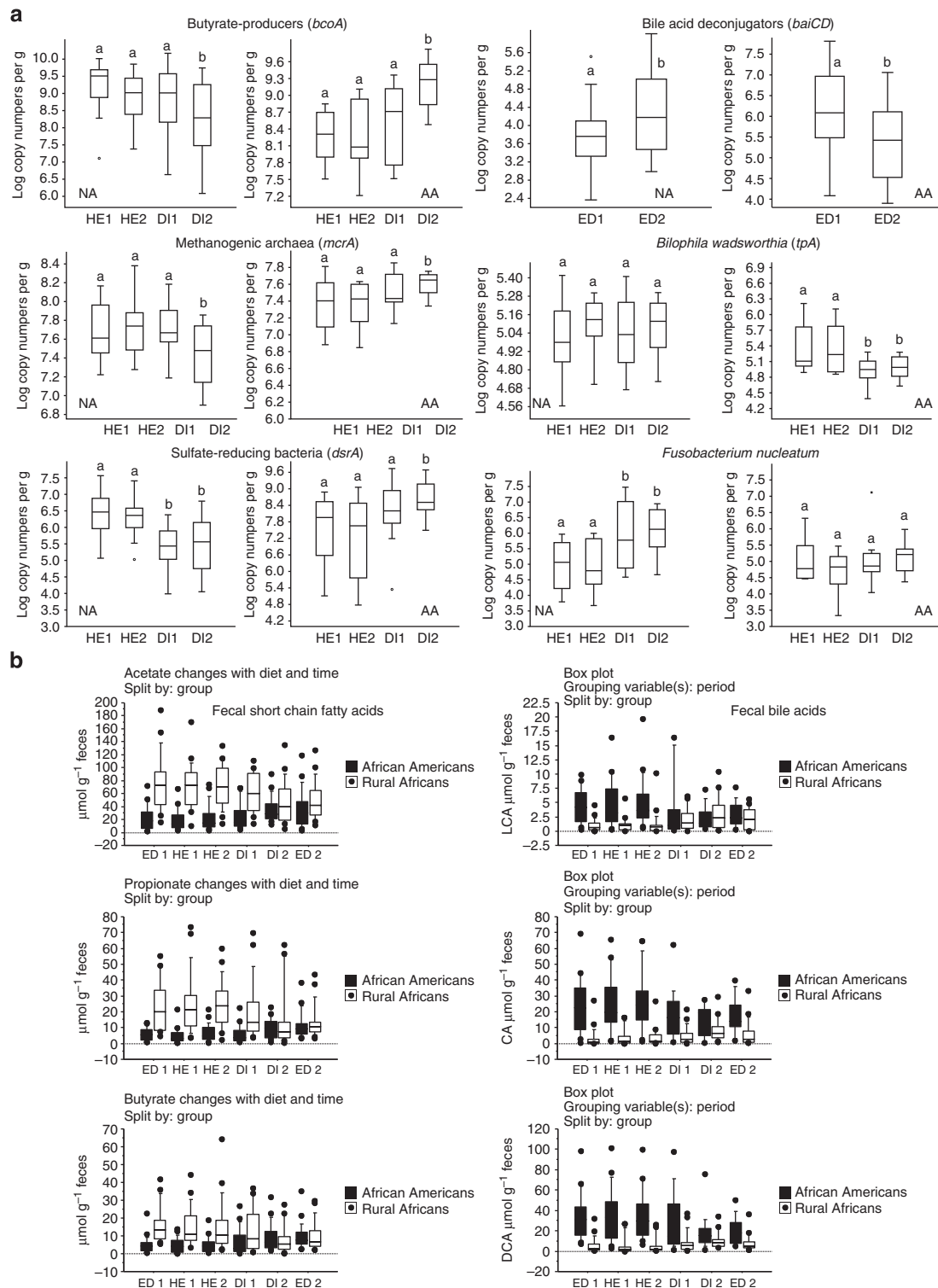
**Figure 1 | Colonic mucosal immunohistochemistry of proliferative and inflammatory biomarkers.** Immunohistochemical analysis of colonic mucosal biopsies taken at colonoscopy with their associated quantitative analysis on the right panel. **(a)** An illustration of the paired changes in Ki67 staining of epithelial crypt cells in one African American (the first and second panels) and one rural African (the third and 4th panels left to right) before and after dietary switch. **(b)** Paired changes in CD3<sup>+</sup> staining, similar to **a**. **(c)** Paired changes in CD68<sup>+</sup> macrophages in the lamina propria. The bar graphs on the far right summarize the group mean  $\pm$  s.e. results in 20 African Americans and 12 rural Africans. The two-tailed Mann-Whitney *U*-test was used for comparisons for non-paired samples and the Wilcoxon rank sum test for paired samples, with Bonferroni correction for multivariate comparisons. Triangles indicate significant ( $P < 0.05$ ) baseline differences and stars indicate significant changes induced by diet switch.

affects colon cancer risk is mediated by the microbiota. First, there was strong evidence that butyrogenesis was associated with lower mucosal proliferation, both in Africans consuming their usual high-fibre diet and in African Americans switched to a high-fibre African-style diet (Figs 1 and 2, and Supplementary Table 6). As a wide variety of microbes have the capacity to produce butyrate, we targeted our analysis on the measurement of microbial genes encoding the final enzyme in butyrogenesis, namely butyryl-coenzyme-A-CoA transferase<sup>13</sup>. On their usual diet, Africans had significantly greater abundances of butyrate synthesizing genes, and higher fecal butyrate concentrations (Fig. 2) and colonic evacuate quantities (Supplementary Table 7), whereas switching them to a western, low-fibre diet suppressed each of these parameters. Reciprocally, our measurements of butyrogenesis were lower in Americans consuming their usual low-fibre diet and increased after dietary switch to a high-fibre African-style diet. Impressively, ‘Africanization’ of the diet increased total quantities of butyrate in colonic evacuates 2.5 times, whereas ‘westernization’ reduced quantities by half.

Recognizing that butyrogenesis is a final common pathway of fibre fermentation, which can only be accomplished by multiple microbial metabolic interactions, we examined the simultaneous changes in the group of microbes that remove the major metabolic end product, hydrogen. If hydrogen is allowed to accumulate, it impairs the regeneration of oxidized pyridine nucleotides, which in turn inhibit the fermentation process and butyrogenesis<sup>14</sup>. The efficient removal of hydrogen by the hydrogenotrophic microbes, methanogens, which bind hydrogen

into non-toxic methane, sulfate-reducers, which incorporate hydrogen into hydrogen sulfide, and acetogens, which consume hydrogen to produce acetate and water, is therefore a prerequisite for enhanced butyrogenesis. Our observation that all these microbial groups were indeed in greater abundance, as were their metabolic products, both in Africans consuming their usual high-fibre diet and when Americans were switched to the high-fibre diet (Fig. 2) well illustrates the importance of microbial partnership in the maintenance of colonic health.

Flint and colleagues<sup>15</sup> have provided strong evidence that the most active butyrate producers, for example, *Roseburia* spp. and *Eubacterium rectale* and *Faecalibacterium prausnitzii* are contained within *Clostridium* groups IV and XIVa. As Africans and African Americans have very distinct microbiota composition at baseline (Supplementary Table 8), we performed a global analysis of the microbiota by 16S ribosomal RNA gene phylogenetic microarray (Human Intestinal Tract Chip (HITChip)) to examine whether these groups were more prevalent with high fibre consumption<sup>16</sup>. Remarkably, these groups were no more apparent in Africans or Americans eating a high-fibre diet. In fact, the compositional changes that were specifically associated to the diet switch were minor (ten genus-like groups for Africans and two for Americans; Fig. 3a). These observations suggest either dissociation between structure and function or a change in microbial interactions within the community<sup>17</sup>. On the other hand, we observed stronger co-occurrence patterns between the genus-level taxa in Africans consuming their usual high-fibre diet (Fig. 3b), which included



**Figure 2 | Targeted Analysis of microbial functional genes of specific interest and their metabolites.** Box plots with whiskers and outliers summarizing the results for the targeted functional quantitative PCR analysis of microbial genes of interest (**a**) and measurements of their metabolic products in fecal samples (**b**) in 20 African American (AA) and 20 rural Africans (NA) at sequential time points during the study. Group median and variability outside the upper and lower quartiles are shown. Measurements were taken before (Table 1: ED1, HE1 and HE2), during (DI1 and DI2) and after (ED2) 2 weeks of dietary change. With regard to the functional gene analysis, *bcoA* codes for the enzyme butyryl-CoA:acetate CoA-transferase, which is responsible for the last step in butyrate synthesis, *mcrA* for the enzyme responsible for methanogenesis, *baiCD* for the 7- $\alpha$ -dehydroxylating enzyme responsible for secondary bile acid production, and *tpA* for taurine:pyruvate aminotransferase. LCA, lithocholic acid; CA, cholic acid; DCA, deoxycholic acid. As data were not normally distributed, the Kruskal-Wallis test was used to evaluate the significance of the group median differences across the different time points shown on the box plots. In the upper panel, comparison with baseline ED1, 'a' signifies no significant change and 'b' a significant change. In the lower panel for the short-chain fatty acids acetate, propionate and butyrate, and the secondary bile acids LCA and DCA, and primary bile acid CA, there were no significant differences between time points ED1, HE1 and HE2, while ED2 was consistently different ( $P < 0.05$ , for details, see Supplementary Table 6). This, together with Supplementary Fig. 6, illustrates no significant effect from colonoscopy and a profound effect from the dietary change.

these potential butyrate producers, for example, *E. rectale* et rel. and *Clostridium symbiosum* et rel., and bacteria associated with complex carbohydrate utilization, for example, *Oscillospira guillermoidii* et rel. Reduction in fibre consumption led to opposite associations (Fig. 3b, bottom right). Reciprocally, high-fibre feeding in Americans was associated with a shift from correlations between *Bacteroides* and potential butyrate-producing groups (*Roseburia intestinalis* et rel. and *C. symbiosum* et rel.) towards stronger co-occurrence patterns including Firmicutes that are typically associated with complex carbohydrate fermentation (Fig. 3b, top right).

### Global changes in the metanome following diet switch.

We have described metanomics and the associated term, metabolomics, as the multiparametric metabolic responses of complex systems to perturbations through time<sup>18</sup>. Specifically, metanomics addresses such phenotypic changes at the level of small-molecule metabolites and maps these processes using appropriate analytical and statistical processes. Recognizing the extreme complexity of the microbiota and its ability to produce a wide array of other metabolites that could affect mucosal health, we performed compositional and pathway analysis<sup>19</sup> on fecal water and urine samples by <sup>1</sup>H NMR (Figs 4 and 5). Perhaps one of the most striking findings of this study was the dramatically higher quantities and diversity of metabolites in fecal water in Africans consuming their usual diet compared with African Americans (Fig. 4a). Although the higher levels of products of saccharolytic fermentation were expected from our targeted analysis, the higher quantities of proteolytic fermentation products were not. The only significantly more abundant fecal metabolite identified in the Americans at baseline, and in Africans after westernization of their diets, was choline. This is highly significant, as recent studies have shown that choline is extensively metabolized by the microbiota to trimethylamine<sup>20</sup> that is absorbed and metabolized by the liver to trimethylamine-*N*-oxide, which is strongly atherogenic<sup>21</sup>, providing yet another link between westernized diets, the microbiota and westernized diseases. Evidence that this metabolic pathway was indeed stimulated in Africans given the western diet was the observed increase in urine trimethylamine-*N*-oxide (Table 2). Reciprocal changes in urinary metabolites derived from microbial metabolism of green vegetables, for example, *N*-acetyl-*s*-methyl-*L*-cysteine sulfoxide, were also observed following diet switch, suggesting that some of the reduction in mucosal proliferation in Americans could also have been related to the effects of increased phytochemical consumption in the Africanized diets<sup>22</sup>. The urinary spectra obtained from African Americans following the change to a high-fibre diet also revealed a marked increase in an unrecognized metabolite (Fig. 4a, left upper panel), possibly a nicotinamide breakdown product, whose character and function have yet to be determined.

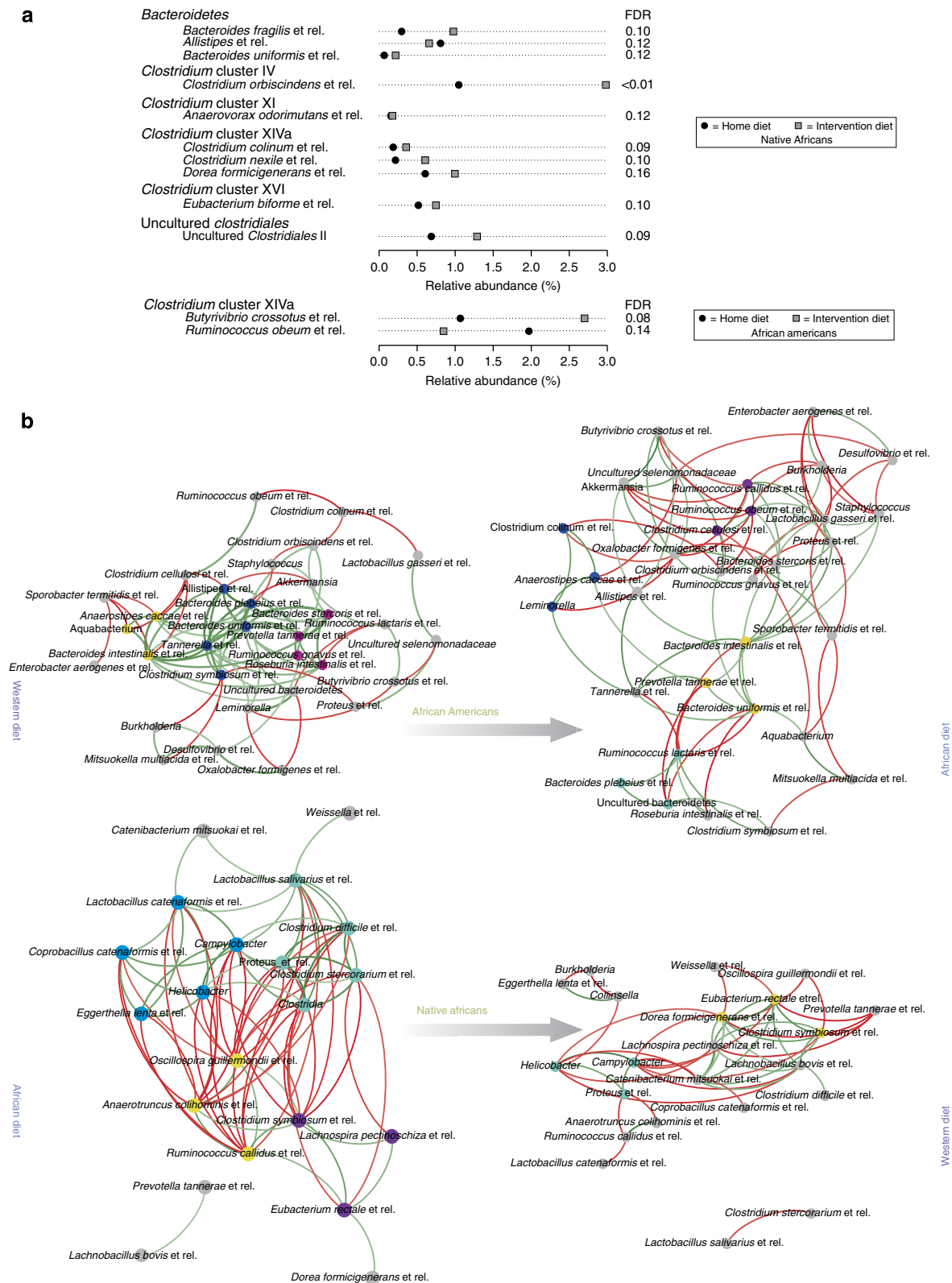
The dietary intervention caused, in tandem with the microbiota changes, substantial modifications in the metabolic phenotypes of both groups (Table 2, Fig. 4a), with Africans demonstrating less metabolic variation and Americans becoming more diverse. This is probably best appreciated from the metabolic pathway analysis (Fig. 5b,c) of stool and urine samples from both groups, with colour coding identifying the significant baseline differences, based on integration of the enzymes of bacterial groups identified by HITChip microarray<sup>16</sup>, host metabolic enzymes, and fecal and urinary metabolites by <sup>1</sup>H NMR associated with reactions from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database. The lower two panels illustrate the significant shifts in metabolic pathways identified in fecal samples after dietary switch, showing mirror image changes reflecting the above, most

notably involving reductions in short-chain fatty acids and increases in glycosylated proteins and choline metabolism in Africans, and the unexpected increases in amino acid metabolism in Americans. Consistent with the results of our targeted analyses, the effects on saccharolytic fermentation were reciprocal, with increases in acetate and lactate in Africans and reductions in Africans. Unlike the results obtained from targeted gas chromatography shown of Fig. 2, the increases in butyrate and propionate measured by global <sup>1</sup>H NMR spectroscopy did not achieve statistical significance, probably possibly due to sample preparation differences between GC and NMR. Again, contrary to expectation, fecal concentrations of colonic branched-chain amino acids, which are likely to be derived from proteolytic fermentation, decreased in Africans given the high-protein western diet for unclear reasons.

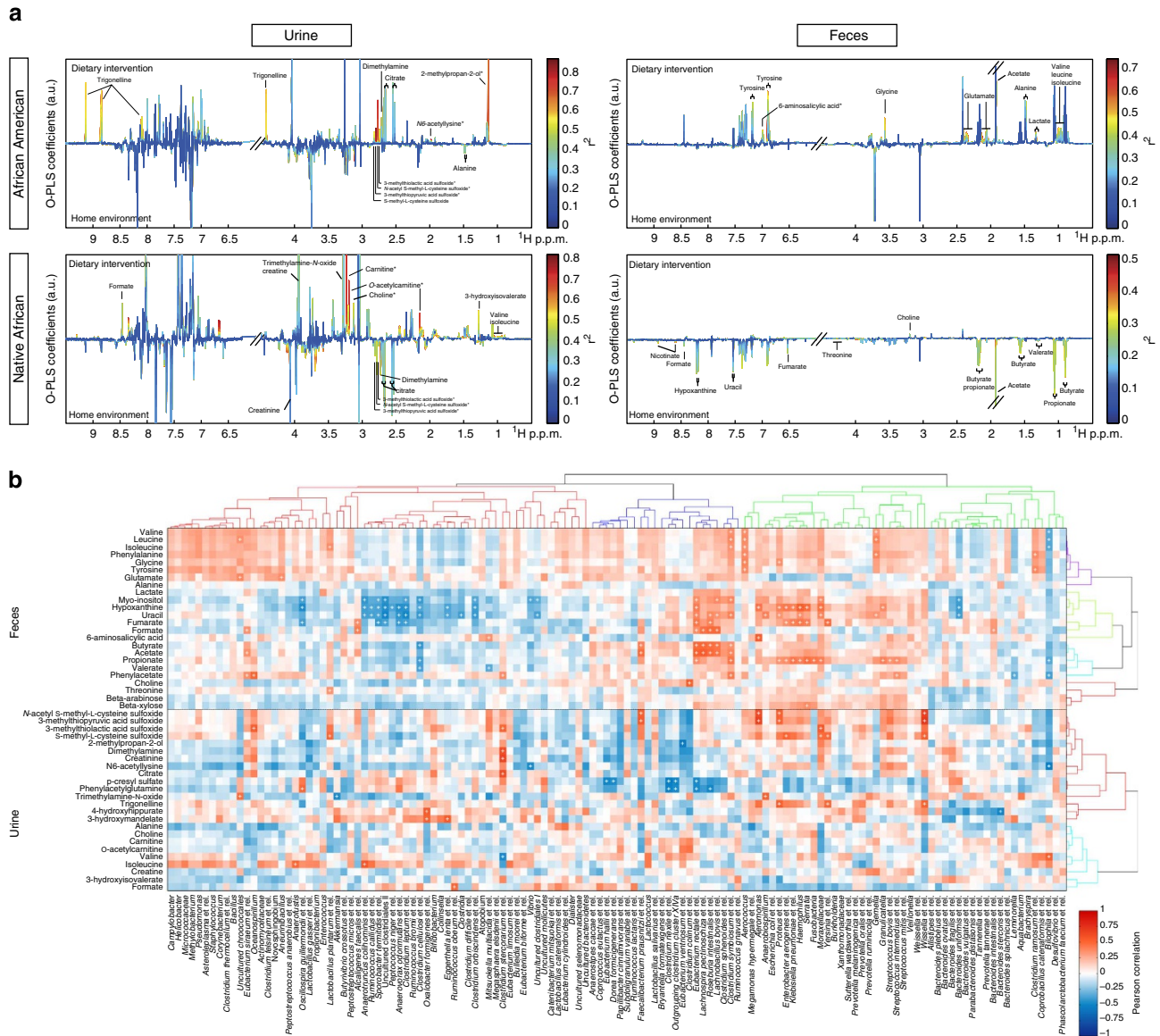
**Global microbe–metabolite associations.** To investigate this, we performed a partial correlation analysis between genus-level microbes and urine and fecal metabolites (Fig. 4b), adjusted for nationality and diet. Although we cannot biologically explain all associations (for example, propionate correlates with enterobacteriaceae and *Streptococcus*, which are not known to produce propionate), one striking feature was that the metabolite butyrate was exclusively associated with microbial groups that are known to contain butyrate producers, well exemplified by *R. intestinalis* et rel, *E. rectale* et rel and *C. symbiosum* et rel, which are shown in the co-occurrence networks in Fig. 3b, to be enhanced when high-fibre diets were consumed by both population groups.

**High-fat, low-fibre diets and toxic microbial metabolites.** It is unlikely that butyrogenesis can alone explain diet-associated cancer risk. Our preceding studies had shown that the higher-fat diet of African Americans was associated with a greater abundance of fecal microbial genes that encode for one of the key enzymes responsible for deconjugation of bile acids (the *baiCD* gene encodes for the 7 $\alpha$ -dehydroxylating enzyme<sup>23</sup>) as well as their metabolic products, secondary bile acids<sup>5</sup>. Secondary bile acids have been shown to be carcinogenic in experimental models and multiple human studies have shown that their fecal levels are associated with cancer risk<sup>10,24,25</sup>. Furthermore, there is substantial experimental evidence that the carcinogenic potential of these metabolites is potentiated by colonic butyrate deficiency<sup>26–29</sup>. Once again, we show that there were reciprocal changes in both *baiCD* gene abundance and the secondary bile acids, lithocholic and deoxycholic acid, after diet switch, with levels increasing in Africans given the high-fat, western-style diet and reductions in Americans fed the low-fat, traditional African diet (Fig. 2, lower panel). Although the use of these functional genes may not cover all the enzymatic pathways of interest, they do provide legitimate tracers for them<sup>30</sup>. Remarkably, Africanization reduced colonic evacuate secondary bile acids by 70% and westernization increased them by 400% (Supplementary Table 7).

Dietary fat may influence colon cancer risk through other pathways as illustrated by a recent study in an *Il10*<sup>-/-</sup> mouse model, which showed that a high-fat diet cultivated an increase in the abundance of *Biophila wadsworthia*, a member of the *Desulfovibrionaceae* family, which generates hydrogen sulfide via taurine respiration, leading to acute inflammation<sup>31</sup>. Exogenous hydrogen sulfide is a potent genotoxin *in vitro*<sup>32</sup>. Finally, although no strong evidence has yet been found to indict a specific microbe in colonic carcinogenesis, a recent study found that *Fusobacterium nucleatum* was enriched in human colon cancer tissue, and suggested that the microbe might promote colorectal neoplasia progression through recruitment



**Figure 3 | The impact of diet switch on microbiota composition and co-occurrence networks.** (a) An overview of significantly altered genus-like bacterial groups (FDR < 0.2; linear model) following the 14-day diet change. Average fraction of total HITChip signal (proxy for relative abundance) over the samples for each comparison. Black and grey squares indicate the average abundances at home- and intervention diet, respectively. (b) The microbiota genus-like group co-occurrence networks of Africans and African Americans under both dietary regimens for genus-like groups. Groups with correlation differences > 1 before and after the diet and at least  $|r| > 0.5$  are shown. Positive correlations are indicated with green lines; negative correlations with red lines. Identically coloured nodes indicate network modules with three or more co-occurring groups. Grey nodes represent modules of less than three co-occurring groups or those not represented by a module. The sample sizes are 20 African Americans and 17 rural (Native) Africans.

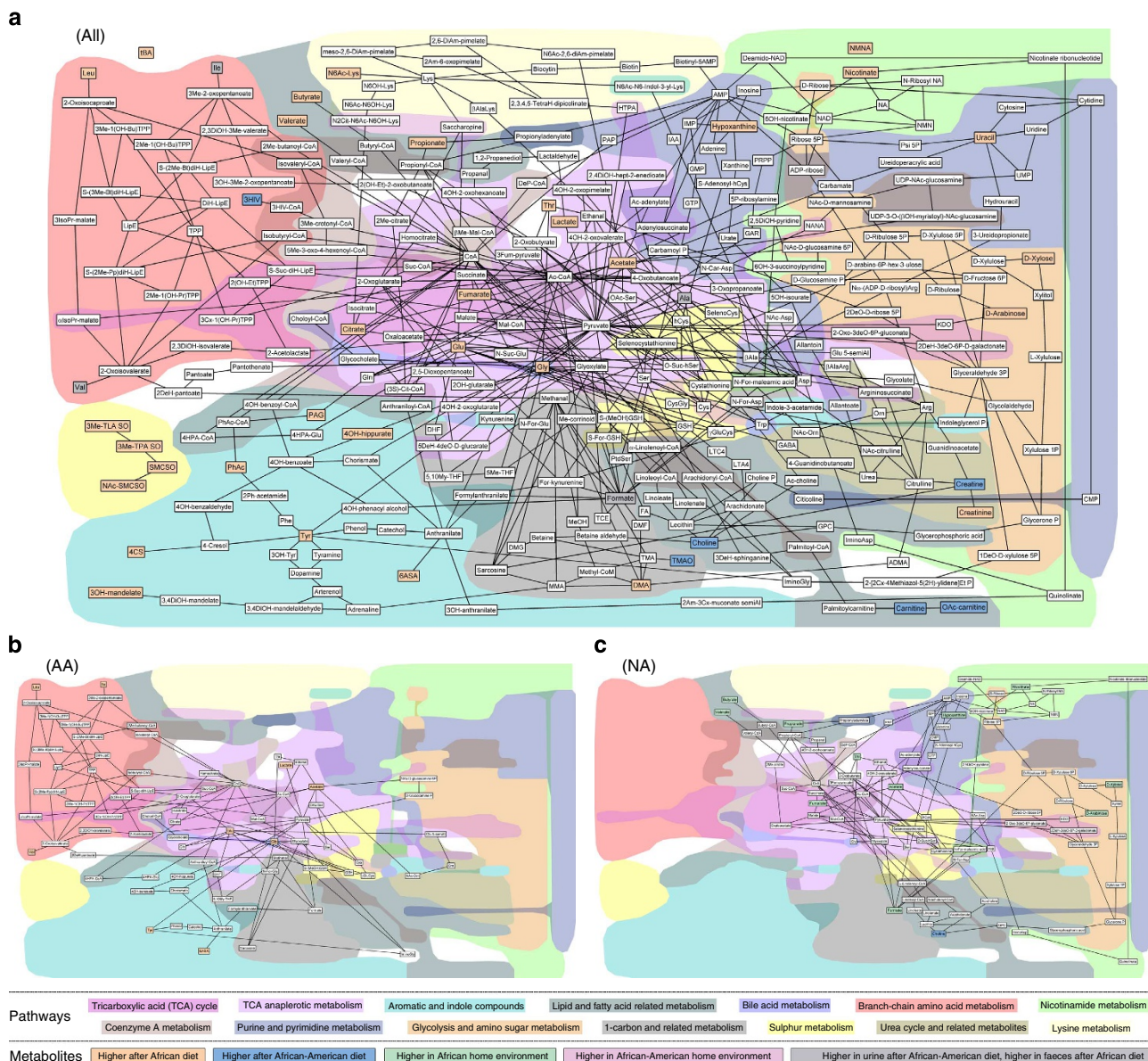


**Figure 4 | The impact of diet switch on fecal and urinary metabolites and their correlation with fecal microbiota.** (a) <sup>1</sup>H NMR fecal and urinary spectra obtained from African Americans (AA) and native Africans (NA) during the home environment period (HE) and post dietary intervention (DI). Metabolites distinguishing the two groups are labelled either above the line (higher after dietary intervention) or below (higher in home environment), and the colour represents the correlation coefficient ( $r^2$ ) of the metabolite with the groups. (b) Two-way clustered heatmap of partial correlation between phylotypes assigned to genus-level taxa whose representation in the fecal microbiota (x axis) of African American (AA) and rural Africans (NA), and relative metabolite concentrations in urine and feces (y axis). P-values from correlation were adjusted for multiple testing using the Benjamini-Hochberg FDR. Given a large number of tests performed, a cutoff value of 0.3 was used to define the significance denoted by '+'. This level follows the Food and Drug Administration guidelines for bioanalytical biomarkers and was chosen because of the relatively low number of samples it allows for less false negatives. Fecal metabolites (upper compartment) demonstrated a higher number of significant correlations with fecal microbiota compared with urinary metabolites (lower compartment). *E. rectale et rel.*, *Lachnospira pectinoschiza et rel.*, *R. intestinalis et rel.* and *Lactobacillus bovis et rel.* correlate strongly with butyrate, acetate and formate. Adjustments were made for country (binary vector with 0 = AA and 1 = NA) and diet/stage (0 = HE 2 and 1 = DI 2, see Table 2) in calculating the partial correlations).

of proinflammatory tumour-infiltrating immune cells<sup>33</sup>. Alternatively, members of our group have hypothesized that fusobacteria may promote genotoxicity by their ability to convert cysteine to hydrogen sulfide<sup>34</sup>. Figure 2 (upper panel) shows that the presence and activity of both of these microbes may have played a role in the changes in mucosal biomarkers: the high-fibre, low-fat dietary intervention in African Americans was associated with a significant decrease of *B. wadsworthia*, whereas the low-fibre, high-fat intervention in Africans was associated with an increase of *F. nucleatum*.

**Discussion**

We have shown in individuals from high-risk and from low-risk cancer populations that changes in the food content of fibre and fat had remarkable effects on their colonic microbiota and metabolome within 2 weeks, and, critically, that these changes were associated with significant changes in mucosal inflammation and proliferation. Although we cannot claim from our results that these changes in mucosa will result in changes in the development of cancer, there is good experimental evidence that increased epithelial proliferation predicts neoplastic change, because it



**Figure 5 | The impact of diet switch on fecal and urinary metabolic reaction networks.** (a-c) Metabolic reaction networks of metabolites found differentially expressed between different dietary comparisons, created using the MetaboNetworks software<sup>19</sup> using information from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) on the reactions and enzymes within the biological system. The network shows links between metabolites if the reaction entry in KEGG indicates a main reactant pair and the reaction is either mediated by (1) an enzyme linked to human genes, (2) an enzyme linked to genes from identified bacterial groups using the HITchip or (3) it is part of a spontaneous process. The colour of the metabolites indicate whether the metabolite is found in higher concentrations in the urine/faeces of individuals consuming an African diet (orange), higher in individuals consuming an African American diet (blue), higher in the African American home environment (light green), higher in the African home environment (green) or whether the metabolites is not significantly associated with any comparison but is part of the metabolic network (white). Two metabolites (alanine and formate) are found with opposite associations in feces and urine (Table 2), these are therefore represented with a grey background colour. A table with full names for each of the abbreviated metabolite names can be found in Supplementary Table 11. The background shading indicates different interconnected pathways. In the top part of the legend, the colour for each pathway, as represented in the figure, is shown. The bottom part of the legend, with black boxes around text box, indicates the association of each metabolite. (a) Global network of changes ('All') observed in fecal water and urine, and in any comparison of the different groups. (b) Network of fecal metabolites found differentially expressed in African Americans before and after the dietary switch ('AA'). (c) Network of fecal metabolites found differentially expressed in Africans before and after the dietary switch ('NA').

increases the risk of development of DNA mutations due to the higher rate of exposure of sensitive proliferating cells to luminal carcinogens. The evidence for the use of proliferation biomarkers in humans is based on the known increase in the percentage of epithelial cells engaged in DNA synthesis (measured as 'proliferation index') in all premalignant conditions of the gastrointestinal tract, as well as the observed expansion of the

proliferative zone, causing a shift of the proliferating cells from the base to the surface of the crypt<sup>35-37</sup>. There is also substantial evidence that chronic inflammation increases cancer risk<sup>38</sup> and it has been observed in human studies that colon cancer is increased fivefold in patients with chronic ulcerative colitis<sup>38</sup> and reduced 50% by anti-inflammatory drugs<sup>39</sup>. In keeping with this, it was noteworthy that our two biomarkers on mucosal inflammation,



**Table 2 | Fecal and urinary metabolites.**

| (a) Fecal metabolites | Chemical shift ( <sup>1</sup> H p.p.m.) | HE African American (N=15) versus African (N=19)  |        |        | DI African American (N=14) versus African (N=20)  |        |        | African American HE (N=15) versus DI (N=14)       |        |        | African HE (N=19) versus DI (N=20)                |        |       |
|-----------------------|---|---|--------|--------|---|--------|--------|---|--------|--------|---|--------|-------|
|                       |   | Q <sup>2</sup> Y = 0.59; R <sup>2</sup> X = 29.8% |        |        | Q <sup>2</sup> Y = 0.74; R <sup>2</sup> X = 26.8% |        |        | Q <sup>2</sup> Y = 0.66; R <sup>2</sup> X = 37.2% |        |        | Q <sup>2</sup> Y = 0.37; R <sup>2</sup> X = 30.7% |        |       |
|                       |   | r   | P      | q      | r   | P      | q      | r   | P      | q      | r   | P      | q     |
| Valine                | 0.99                                    | 0.77*   | <0.001 | 0.002  |   |        |        | 0.82†   | <0.001 | 0.003  |   |        |       |
| Leucine               | 0.97                                    | 0.75*   | <0.001 | 0.003  |   |        |        | 0.75†   | 0.001  | 0.02   |   |        |       |
| Isoleucine            | 1.01                                    | 0.78*   | <0.001 | 0.001  |   |        |        | 0.77†   | <0.001 | 0.01   |   |        |       |
| Glutamate             | 2.36                                    |   |        |        | 0.82†   | <0.001 | <0.001 | 0.84†   | <0.001 | <0.001 |   |        |       |
| Alanine               | 1.48                                    | 0.7*  | 0.003  | 0.007  |   |        |        | 0.79†   | <0.001 | 0.009  |   |        |       |
| Glycine               | 3.57                                    | 0.72*   | 0.002  | 0.005  |   |        |        | 0.82†   | <0.001 | 0.002  |   |        |       |
| Tyrosine              | 6.91                                    | 0.74*   | <0.001 | 0.003  |   |        |        | 0.81†   | <0.001 | 0.001  |   |        |       |
| Butyrate              | 0.89                                    | 0.67*   | 0.007  | 0.01   |   |        |        |   |        |        | 0.71*   | <0.001 | 0.006 |
| Propionate            | 1.06                                    | 0.75*   | <0.001 | 0.002  |   |        |        |   |        |        | 0.71*   | <0.001 | 0.006 |
| Valerate              | 1.29                                    | 0.73*   | 0.001  | 0.004  |   |        |        |   |        |        | 0.66*   | 0.005  | 0.01  |
| Acetate               | 1.93                                    | 0.74*   | <0.001 | 0.003  |   |        |        | 0.7†  | 0.03   | 0.006  | 0.75*   | <0.001 | 0.005 |
| Lactate               | 1.34                                    | 0.75*   | <0.001 | 0.002  |   |        |        | 0.77†   | <0.001 | 0.009  |   |        |       |
| Formate               | 8.45                                    |   |        |        |   |        |        |   |        |        | 0.69*   | 0.002  | 0.009 |
| Choline               | 3.20                                    | 0.78§   | <0.001 | 0.001  |   |        |        |   |        |        | 0.66  | 0.006  | 0.01  |
| Uracil                | 5.81                                    | 0.78*   | <0.001 | 0.001  |   |        |        |   |        |        |   |        |       |
| Hypoxanthine          | 8.21                                    | 0.74*   | <0.001 | 0.003  |   |        |        |   |        |        | 0.69*   | 0.002  | 0.008 |
| 6-Aminosalicylic acid | 7.01                                    |   |        |        | 0.85†   | <0.001 | <0.001 | 0.77†   | <0.001 | 0.005  |   |        |       |
| Fumarate              | 6.53                                    |   |        |        |   |        |        |   |        |        | 0.69*   | 0.002  | 0.008 |
| Phenylacetate         | 3.54                                    | 0.81*   | <0.001 | <0.001 |   |        |        |   |        |        |   |        |       |
| β-Xylose              | 4.58                                    |   |        |        |   |        |        |   |        |        | 0.72*   | <0.001 | 0.006 |
| β-Arabinose           | 4.53                                    |   |        |        |   |        |        |   |        |        | 0.74*   | <0.001 | 0.005 |
| Nicotinate            | 8.62                                    |   |        |        |   |        |        |   |        |        | 0.71*   | <0.001 | 0.005 |
| Threonine             | 4.26                                    |   |        |        |   |        |        |   |        |        | 0.77*   | <0.001 | 0.002 |

| (b) Urinary metabolites                 | Chemical shift ( <sup>1</sup> H p.p.m.) | HE African American (N=8) versus African (N=12)   |        |      | DI African American (N=7) versus African (N=12)   |        |        | African American HE (N=8) versus DI (N=7)         |        |       | African HE (N=12) versus DI (N=12)                |        |        |
|---|---|---|--------|------|---|--------|--------|---|--------|-------|---|--------|--------|
|   |   | Q <sup>2</sup> Y = 0.41; R <sup>2</sup> X = 26.6% |        |      | Q <sup>2</sup> Y = 0.86; R <sup>2</sup> X = 36.4% |        |        | Q <sup>2</sup> Y = 0.51; R <sup>2</sup> X = 20.8% |        |       | Q <sup>2</sup> Y = 0.83; R <sup>2</sup> X = 27.8% |        |        |
|   |   | r   | P      | q    | r   | P      | q      | r   | P      | q     | r   | P      | q      |
| N-acetyl s-methyl-L-cysteine sulfoxide† | 2.78                                    | 0.8*  | 0.003  | 0.16 | 0.93†   | <0.001 | <0.001 | 0.93†   | <0.001 | 0.03  | 0.8*  | <0.001 | 0.007  |
| 3-Methylthiopyruvic acid sulfoxide†     | 2.81                                    | 0.81*   | 0.002  | 0.14 | 0.94†   | <0.001 | <0.001 | 0.95†   | <0.001 | 0.009 | 0.77*   | <0.001 | 0.008  |
| 3-Methylthiolactic acid sulfoxide†      | 2.76                                    |   |        |      | 0.88†   | <0.001 | <0.001 | 0.83†   | 0.002  | 0.12  | 0.79*   | 0.001  | 0.01   |
| S-methyl-L-cysteine sulfoxide           | 2.84                                    |   |        |      |   |        |        | 0.84†   | 0.003  | 0.13  |   |        |        |
| p-Cresyl sulfate                        | 2.35                                    | 0.7*  | 0.03   | 0.28 |   |        |        |   |        |       |   |        |        |
| Phenylacetylglutamine                   | 2.27                                    | 0.71*   | 0.02   | 0.26 |   |        |        |   |        |       |   |        |        |
| 3-Hydroxyisovalerate                    | 1.27                                    | 0.78§   | 0.005  | 0.18 |   |        |        |   |        |       | 0.85  | <0.001 | 0.004  |
| 2-Methylpropan-2-ol‡                    | 1.14                                    |   |        |      | 0.91†   | <0.001 | <0.001 | 0.89†   | <0.001 | 0.07  |   |        |        |
| Alanine                                 | 1.49                                    |   |        |      | 0.8   | 0.003  | 0.02   | 0.75§   | 0.03   | 0.33  |   |        |        |
| Valine                                  | 0.99                                    |   |        |      |   |        |        |   |        |       | 0.85  | <0.001 | 0.003  |
| Isoleucine                              | 0.93                                    |   |        |      |   |        |        |   |        |       | 0.88  | <0.001 | <0.001 |
| Carnitine                               | 3.23                                    |   |        |      | 0.88  | <0.001 | 0.001  |   |        |       | 0.91  | <0.001 | <0.001 |
| O-acetylcarnitine†                      | 2.15                                    |   |        |      | 0.91  | <0.001 | <0.001 |   |        |       | 0.91  | <0.001 | <0.001 |
| N6-acetyllysine‡                        | 1.99                                    |   |        |      |   |        |        | 0.9†  | <0.001 | 0.01  |   |        |        |
| Citrate                                 | 2.53                                    |   |        |      | 0.85†   | <0.001 | 0.003  | 0.71†   | 0.04   | 0.39  | 0.71*   | 0.006  | 0.04   |
| Creatine                                | 3.94                                    |   |        |      |   |        |        |   |        |       | 0.73  | 0.004  | 0.03   |
| Creatinine                              | 4.07                                    |   |        |      | 0.82†   | 0.001  | 0.01   |   |        |       | 0.77*   | <0.001 | 0.007  |
| Trigonelline                            | 4.44                                    |   |        |      | 0.9†  | <0.001 | <0.001 | 0.86†   | 0.002  | 0.12  |   |        |        |
| 4-Hydroxyhippurate                      | 6.97                                    |   |        |      | 0.81†   | 0.003  | 0.02   |   |        |       |   |        |        |
| 3-Hydroxymandelate                      | 6.92                                    |   |        |      | 0.87†   | <0.001 | 0.002  |   |        |       |   |        |        |
| Dimethylamine                           | 2.73                                    | 0.88*   | <0.001 | 0.07 |   |        |        | 0.79†   | 0.01   | 0.25  | 0.78*   | 0.001  | 0.01   |
| Trimethylamine-N-oxide                  | 3.27                                    |   |        |      |   |        |        |   |        |       | 0.76  | 0.02   | 0.07   |
| Choline                                 | 3.23                                    |   |        |      |   |        |        |   |        |       | 0.84  | <0.001 | 0.004  |
| Formate                                 | 8.46                                    |   |        |      |   |        |        |   |        |       | 0.79  | 0.002  | 0.01   |
| N-methyl-2-pyridone-5-carboxamide‡      | 6.66                                    |   |        |      |   |        |        |   |        |       | 0.93  | <0.001 | <0.001 |

DI, dietary intervention; HE, home environment.  
 Parts (a) and (b) show significantly changed fecal and urinary metabolites in each pairwise comparison (for example, African American versus African during HE, African American at HE versus DI). Representative chemical shift of each metabolite is provided. The r-values are Pearson's correlation coefficient from the correlation of the metabolite concentrations with classification (for example, African Americans versus Native Africans). P- and q-values represent the significance of the metabolite changes and false discovery rate-adjusted P-values<sup>76</sup>, respectively.  
 \*Higher in HE: Africans.  
 †Higher in dietary switch: African Americans.  
 ‡Tentative assignment.  
 §Higher levels of metabolites in HE: African Americans.  
 ||Higher in dietary switch: Africans.

CD3<sup>+</sup> intraepithelial lymphocytes and CD68<sup>+</sup> lamina propria macrophages responded in the same direction as our Ki67 proliferative biomarker. CD3<sup>+</sup> intraepithelial lymphocytes provide a measure of T-cell activation and, therefore, the pivotal balance between immune activation and quiescence in response to luminal antigens, such as bacteria<sup>40,41</sup>, whereas CD68<sup>+</sup> lamina propria macrophages play an active role in the persistence of chronic inflammation<sup>42</sup> and neoplastic progression<sup>43,44</sup>. Interestingly, the microbial metabolite, butyrate, was shown recently to reduce the responsiveness of lamina propria macrophages to commensal bacteria<sup>45</sup>, suggesting that the dietary change in Americans may have resulted in a decrease in both macrophage numbers and their activity.

Our study revealed two unanticipated colonic mucosal findings in rural Africans consuming their usual baseline diets. First, despite the significantly lower epithelial proliferation and population colon cancer rates, mucosal inflammation detected endoscopically, histologically and by immunohistochemistry was more common and pronounced compared with high-risk African Americans (Fig. 1). Second, despite the lower protein intake and lower plasma amino acid levels (Supplementary Table 9), proteolytic fermentation was also increased, releasing metabolites that are known to be inflammatory and pro-carcinogenic<sup>46</sup>. The explanation for these findings is unclear, but may be linked to the high prevalence of chronic parasitic infestations in rural Africans causing inflammation and exudation of mucinous proteins. Of note, schistosomes were incidentally identified in two subjects and a tapeworm segment in another (Supplementary Fig. 3). Intestinal protozoa have been hypothesized to stimulate immunosurveillance against colon cancer<sup>47</sup>. Clearly, further focused investigations of these paradoxical findings are needed, but possible explanations involve the complex associations between inflammation, immunosurveillance and carcinogenesis, and the ability of butyrogenesis to arrest the progression of inflammation to neoplasia. In murine models, lymphocytes can suppress carcinogenesis as part of the process of immunosurveillance, and in patients with tumours, specific intra-tumour lymphocyte markers for Th1 polarization, and cytotoxic and memory T cells have been shown to suppress growth and metastasis, a process termed immunoeediting<sup>48,49</sup>. Furthermore, T-cell infiltration may suppress or enhance tumorigenesis, depending on the type, density and location of the immune cells. There is substantial experimental evidence, recently reviewed by us<sup>7,8</sup>, that the major products of saccharolytic fermentation, acetate, propionate and butyrate have potent anti-neoplastic properties<sup>50</sup>, but butyrate is unique in its maintenance of colonic mucosal health and defense. First, it is the preferred energy source for colonocytes<sup>51</sup>. Second, it has a remarkably wide variety of antineoplastic properties, best exemplified by its function as a histone deacetylase inhibitor<sup>52</sup>, its capacity to downregulate the key canonical *Wnt*-signalling pathway linked to colonic carcinogenesis<sup>53</sup> and its ability to reduce the burden of carcinogens, such as bile acids<sup>28,29,54</sup> and red meat products<sup>26</sup>. Further studies are needed to explore the intriguing possibility that increased saccharolytic fermentation and butyrogenesis may counter the recognized proliferative properties of T-helper type 17 T cells<sup>55</sup>.

Although we have focused on the potential for a high-fibre and high-fat diet to modify metabolic pathways that have been shown experimentally to have an impact on mucosal biomarkers of cancer risk, it must be stressed that a change in one aspect of an otherwise isocaloric diet will inevitably result in a change in another. Consequently, the associated change in animal protein or digestible carbohydrate might also be responsible for the observed mucosal changes. Be that as it may, our results provide fairly substantial evidence for a powerful influence of butyrate on

mucosal health and raise hope that an increase in fibre consumption, together with a moderation in fat intake, may reduce the disproportionately high incidence of colon cancer in African Americans and indeed in most westernized communities. Clearly, these suggestions now need to be confirmed by longitudinal studies. With regard to fibre, total quantities may be critical, as fibre supplementation at  $\approx 35$  g per day has generally failed to reduce polyp recurrence in clinical trials<sup>56</sup> and recent experimental studies demonstrate that a threshold effect of butyrate availability determines its inhibitory effect on proliferation<sup>52</sup>. Furthermore, human studies by Burkitt<sup>57</sup> concluded that total fibre intake needs to exceed 50 g per day (as in our study) to prevent colon cancer. Our results also raise serious concern that the progressive westernization of African communities may reduce butyrogenesis, allowing their high basal state of chronic inflammation to progress through proliferation to neoplasia<sup>38</sup>, culminating in the emergence of colon cancer as a major health issue<sup>58</sup>.

## Methods

**Study design.** First, the study was submitted to the University of Pittsburgh and University of KwaZulu-Natal Institutional Review Boards for approval. We employed a unique design (Table 1 and Supplementary Note 1), where 20 healthy middle-aged African Americans and 20 rural Africans (Supplementary Tables 10: Demographics) from the same communities previously studied<sup>5</sup> were studied first for 2 weeks in their own HE, eating their usual food (HE study), and then again in-house while they were fed the intervention diet for 2 weeks (dietary intervention study). Consequently, each subject served as his/her own control, which is important given the known wide individual variation in colonic microbiota composition. We chose intervention diets that were at the same time palatable and contained reverse quantities of fibre and fat, such that African Americans would be given 'African style' foods increasing their average fibre intake from 14 to 55 g per day and reducing their fat from 35% to 16% of total calories, whereas Africans were given a 'western-style' diet reducing their fibre from 66 to 12 g per day and increasing their fat from 16% to 52% (Supplementary Tables 1–3). Cognizant of the problems of compliance to acute dietary change and the accuracy of dietary recall to estimate actual intakes within the community, we elected to perform all the dietary intervention studies in-house, where meals could be prepared and given under close supervision. With African Americans, participants were housed the University of Pittsburgh Clinical Translational Research Center, and with rural Africans we employed in a rural lodging facility, close to their homes, with full kitchen facilities. Body weights were maintained within 2 kg by adjusting food quantities while keeping the overall macronutrient composition the same. The sampling schedule is given on Table 1, showing that fresh fecal samples were taken at three intervals during the HE study and again three times after the diet switch. Colonoscopy was performed to identify latent disease, polyps, or cancer, to obtain 3 h colonic evacuates for analysis and biopsies for biomarkers of cancer risk<sup>4</sup> at the beginning of the HE study and end of the dietary intervention study. Full details of the menus, cooking methods and total dietary compositions are given under Supplementary Tables 2, 3 and 4a–c, and Supplementary Note 2).

**Subjects and recruitment.** Age- and sex-matched healthy volunteers, with the age range 50–65 years, were randomly selected from the African American population in the Pittsburgh region of Pennsylvania and from the rural native South Africans from the rural KwaZulu region. We collaborated with Dr Stephen Thomas, Director of Minority Studies at the University of Pittsburgh School of Public Health, to recruit healthy African American volunteers from the Pittsburgh region, and also advertised the study with the approval of our Institutional Review Board in public areas. In South Africa, volunteers were recruited through advertisements placed in public community centres, for example, post offices, town halls, civic centres and through the iZulu Community Health Center. Appropriate compensation (as advised by Minority studies and the iZulu Community and ratified by the University of Pittsburgh and KwaZulu-Natal) for time and testing was paid to volunteers for participation.

**Screening.** Informed signed consent was taken from each participant. All African volunteers could understand English, but a nurse-translator participated in the consent process to ensure proper understanding of the details of the research procedures. Screening was performed in Pittsburgh at the Clinical Translational Research Center and in Africa at Ngwelezana Hospital outpatient clinic, Empanjeni, KwaZulu-Natal, South Africa. A detailed medical history was first taken. With rural Africans, a local bilingual nurse acted as an interpreter. A 20-ml blood sample was taken for full blood count, ESR, electrolytes and urea, albumin, alkaline phosphatase, AST and bilirubin. If the results were normal and if they satisfied the eligibility criteria, they were invited to participate in the study.

**Subject eligibility.** Details are given under Supplementary Note 1. Inclusion criteria were as follows: healthy volunteers, from GI standpoint between 40 and 65 years (age at which colon cancer screening/colonoscopy is recommended in this population) and body mass index between 18 and 35 kg m<sup>-2</sup> (Supplementary Discussion). Exclusion criteria were as follows: participants pre-colonoscopy were ineligible if they had a history of familial adenomatous polyposis, hereditary non-polyposis colorectal cancer, inflammatory bowel disease or invasive cancer within 5 years before enrollment (h/o adenomatous polyps acceptable). In addition, ineligible participants were individuals with known renal, hepatic or bleeding disorders; previous GI surgery resulting in disturbed gut function due to loss of bowel or altered anatomy, or any form of chronic GI disease resulting in disturbed gut function, diarrhoea and malabsorption; and individuals with antibiotic use within the past 12 weeks (Supplementary Discussion), current steroids use or with diabetes. Exclusion criteria post colonoscopy were detection of previously unrecognized ulceration (with depth and > 0.5 cm), stricture, severe inflammation and polyps > 1 cm diameter or cancer.

**Fecal, colonic and mucosal sampling and colonoscopy.** To synchronize the measurements of the microbiota, the metabolome and the colonic mucosa, the sampling was tied to the preparation for and conduct of a colonoscopy, as previously described<sup>4</sup>, at baseline while on their usual diet (ED1, Table 1), and then again at the conclusion of the dietary change (ED2, Table 1). In this, fresh fecal samples are collected before colonic evacuation and immediately frozen at -80 °C and the total colonic contents are collected for 3 h during evacuation with a simple polyethylene glycol solution. In this, 2 l of polyethylene glycol (60 g l<sup>-1</sup>, average molecular weight 3350 g/mol) solution was consumed rapidly as possible over 30 min. We avoided using the commercial preparation 'Golytely', as it contains sodium sulfate, which is known to disturb the microbiota composition. Our experience had shown that the quality of bowel preparation with this technique was similar to the more conventional overnight bowel washout. The mucosa health status was assessed by visualization and biopsy. Polyps, when encountered, were removed per standard practice and biopsies were taken from normal mucosa from the proximal (caecum/ascending colon), mid (transverse) and distal (sigmoid) colon at 25 cm from the anal verge. Mucosal samples for immunohistochemistry were collected in formalized saline and for gene expression in RNAlater (Qiagen, Germantown, MD) before being stored at -80 °C.

**Measurement of mucosal biomarkers.** *Histology.* Colonic mucosal biopsies were obtained by colonoscopy before and after dietary switch from three different sites (ascending, transverse and descending) and stored in 10% buffered formalin. Later, the biopsy samples were embedded in paraffin and 5-mm sections were cut and stained with either routine haematoxylin and eosin (H&E) or immunohistochemical stains (see below). The histologic findings on the H&E-stained sections were evaluated by one blinded, experienced gastroenterological histopathologist (AK). Measurements focused on the numbers of inflammatory cells and eosinophils in the lamina propria and the numbers of intraepithelial lymphocytes. Scoring of the H&E-stained section was done as shown in Supplementary Table 5a. Any pathologic finding, such as the presence of parasitic organisms, was recorded.

**Immunohistochemistry.** Slides for CD3 and Ki67 staining were deparaffinized at 60 °C for 2 h. To inhibit endogenous peroxidase, the slides were pre-treated using 3% hydrogen peroxide/methanol at room temperature (RT) for 10 min, followed by antigen retrieval with 0.2% pepsin solution (P7012, Sigma, 3050 Spruce St, St Louis, MO, USA) at 37 °C for 10 min. Serum Free Protein Block (X0909, Dako, 6392 Via Real, Carpinteria, CA 93013, USA) was used at RT for 10 min. The slides were drained and incubated with primary antibodies CD3 and Ki67 (A0452, 1:100, rabbit polyclonal, Dako; MIB-1, 1:100, mouse monoclonal, Dako), respectively, at RT for 1 h. Secondary detection was applied using Immpress universal antibody Polymer detection kit (MP-7500, Vector Labs, 30 Ingold Road, Burlingame, CA 94010, USA) at RT for 30 min. The slides were stained with DAB substrate kit (SK-4100, Vector Labs) for 10 min and counter stained using Shandon Hematoxylin (6765015, Thermo Scientific, 81 Wyman St, Waltham, MA 02451, USA).

Slide staining for CD68 was done on a Ventana Benchmark Ultra slide stainer. Deparaffinized slides were pretreated using the ultra CC1 (950-224, Ventana, 1910 Innovation Park Dr, Tucson, AZ 85755, USA) for 24 min for antigen retrieval. Slides were incubated with the primary antibody CD68 (M087601, 1:100, mouse monoclonal, Clone PG-M1, Dako,) for 32 min RT. The Optiview DAB kit (760-700, Ventana) was used for secondary detection.

**Quantification of immunohistochemical staining.** Counting of the proportions of positive staining cells using light microscopy at × 400 magnification was performed by a single investigator (KM), under blinded conditions. To assess inter-observer variability, 40 slides were randomly selected and recounted by a second senior pathologist (AK), showing a concurrence of 88% for Ki67+ and 80% for CD68+ densities.

**Ki67.** The proportion of Ki67-positive staining cells were counted in well-oriented crypts (average/slide 8, range 4–14). Ki67 proliferation rate was defined as the

number of Ki67+ cells divided by the total number of crypt cells and were expressed as percentage. The differences were found to be the same in the total crypt and in the upper crypt; hence, only the total crypt proportions are reported here.

**CD3.** Only CD3+ staining intraepithelial lymphocytes were counted in a representative area of at least 300 epithelial cells. The density of intraepithelial lymphocytes was expressed as an index of number of CD3-positive lymphocytes per 100 epithelial cells.

**CD68.** The number of CD68-positive cells (macrophages) within the lamina propria were counted and graded on a scale from 1 to 3: grade 1 (none/rare), grade 2 (scattered superficial collections) and grade 3 (strong, diffuse or band-like infiltrate in the superficial lamina propria) as shown in Fig. 1c.

**Targeted analysis of fecal and colonic microbes and metabolites of special interest.** Details of the materials and methods used for the collection, preparation and analysis of fecal samples for targeted analysis of microbes of special interest (real-time quantitative PCR) and their metabolites (Agilent Technologies 6890N Network GC System with a flame-ionization detector for short-chain fatty acids and Shimadzu HPLC–mass spectrometry for quantification using electrospray ionization in negative ion mode by monitoring the (M–H)<sup>-</sup> ion for bile acids) in African Americans and rural Africans have been previously published<sup>5</sup>. Justification for the use of the *BcoA* functional gene for butyrate production is that Flint's group in Aberdeen have demonstrated that although there are a number of different metabolic pathways that use different enzymes that culminate in butyrate synthesis, butyryl-CoA:acetate CoA-transferase, the product of the *BcoA* gene, is responsible for the last step in butyrate synthesis in the vast majority of intestinal butyrate producers<sup>13</sup>. Similarly, there are other microbial enzymes that participate in bile acid deconjugation, but Wells *et al.*<sup>23</sup> demonstrated that there was a good correlation between human fecal bacterial dehydroxylating activity measured by fecal dilution assay and their PCR assay for the *baiCD* gene, which encodes the key enzyme responsible for the bile acid 7 $\alpha$ -dehydroxylation pathway.

**Plasma amino acids.** Fasting blood concentrations were measured in extracted plasma by reverse-phase C-18 precolumn derivatization HPLC (AccQ•Tag Ultra derivatization, Waters, Milford, MA) as previously described<sup>5</sup>.

**Statistical analysis.** Statistical analysis of the group differences in continuous variables was conducted using SPSS 16.0 (SPSS Inc.). The significance of group differences for normally distributed data was assessed with unpaired and paired Student's *t*-tests. The non-parametric data were analysed with a Mann–Whitney *U*-test or Kruskal–Wallis one-way analysis of variance (ANOVA) by ranks for unpaired data and Wilcoxon signed-rank tests for paired data. The significance of the association was evaluated with Spearman's rank correlation test. A level of *P* < 0.05 was accepted as statistically significant. Data are presented as means ± s.e. Complex microbiota and metabolome data were analysed by several multivariate ordinations detailed below (principal component analyses and non-metric multidimensional scaling), Kruskal–Wallis independent tests and multivariate ANOVA with Bonferroni correction.

**Global analysis of the microbiota composition and diversity.** We chose the HITChip phylogenetic microarray for the global profiling of microbiota composition. It has been demonstrated that the HITChip analysis of fecal samples provides highly concordant results concerning the microbiota composition when compared with 16S rRNA gene or metagenome sequencing<sup>59–62</sup>, as it allows deep profiling of phylotypes at high resolution, down to < 0.1% relative abundance, corresponding to a duplicated set of 100,000 pyrosequencing reads per sample with very high reproducibility (> 98%)<sup>59</sup> and at considerably lower cost.

**HITChip analysis.** DNA was isolated from fecal or colonic samples and subsequently used for phylogenetic profiling of the intestinal microbiota using the HITChip phylogenetic microarray<sup>16</sup>. Standardized quality control was maintained through our library of a duplicated set of 3,631 probes targeting the 16S rRNA gene sequences of over 1,000 intestinal bacterial phylotypes. Briefly, the full-length 16S rRNA genes were amplified, and PCR products were transcribed *in vitro* into RNA, labelled with Cy3 and Cy5, and fragmented. Hybridizations were performed in duplicate and data were extracted from microarray-scanned images using Agilent Feature Extraction software, version 10.7.3.1 (<http://www.agilent.com>). Array normalization was performed as previously described, using a set of custom R scripts (<http://r-project.org>) and stored in a custom MySQL database (<http://www.mysql.com>). Duplicate hybridizations with a Pearson correlation > 98% were considered for further analysis. Microbiota profiles were generated by pre-processing the probe-level measurements with minimum–maximum normalization and the frozen Robust Probabilistic Averaging probe summarization<sup>62,63</sup> into three phylogenetic levels: level 1, defined as order-like 16S rRNA gene sequence groups; level 2, defined as genus-like 16S rRNA gene sequence groups (sequence similarity > 90%); and level 3, phylotype-like 16S rRNA gene sequence groups (sequence

similarity > 98%). In the present work, we primarily focus on the genus-level (level 2) variation. Significance of the differences between the time points (Fig. 3a) were estimated based on a (paired) linear model for microarrays (limma) with the threshold of false discovery rate (FDR) < 0.2 estimated based on Benjamini–Hochberg procedure and a minimum fold change of 25% (0.1 at the Log10 scale)<sup>64</sup>.

**Microbial co-occurrence network analysis.** We constructed the co-occurrence networks between the 130 genus-like bacterial groups based on their logarithmic abundances (HITChip log10 signal) within each treatment group (African Americans and Native Africans; before and after the dietary intervention). For a robust correlation analysis, we applied the SparCC algorithm using 20 iterations and with 50 bootstrap data sets for significance testing, followed by  $q$ -value<sup>65</sup> correction of the pseudo  $P$ -values from the bootstrap analysis. We focused on the significant ( $|r| > 0.5$ ;  $q < 0.01$ ) correlations between the genus-like groups where a qualitative change (change of sign) in the correlation was observed following the dietary intervention. To ensure robust analysis, we included only the correlations that changed drastically in the intervention, with a difference of > 1 between the correlation values before and after intervention (so that the correlations would change for instance from  $-0.5$  to  $0.5$ , or higher). This provided a list of genus-like bacteria that had significant changes in their mutual correlation networks following the dietary intervention in either rural African or African American group. To simplify interpretation, we clustered the associated genus-like groups into coherent network modules with complete-linkage hierarchical clustering based on the SparCC correlations. We defined a module as a connected sub-network where the correlations between all genus-like groups within the module are  $r > 0.5$ . Networks were visualized using the network visualization and exploration platform Gephi<sup>66</sup>. The resulting modules are highlighted in Fig. 3b.

**Global analysis of the metabolome: sample preparation for NMR spectroscopic analysis.** All urinary samples were thawed at RT and vortexed for 10 s. A total of 400  $\mu$ l of urinary sample was thoroughly mixed with 250  $\mu$ l of 0.2 M sodium phosphate buffer containing 20% D<sub>2</sub>O, pH 7.4, 0.01% 3-(trimethylsilyl)-[2,2,3,3-<sup>2</sup>H<sub>4</sub>]propionic acid sodium salt and 3 mM sodium azide (NaN<sub>3</sub>). The mixture was subsequently centrifuged at 10,000g for 10 min and 600  $\mu$ l of supernatant was transferred into an NMR tube with an outer diameter of 5 mm. Approximately 200 mg of wet fecal sample was mixed with 600  $\mu$ l of H<sub>2</sub>O (HPLC grade) in a 1.5-ml Eppendorf tube. A cycle of 30-s vortexing, 5-min sonicating at 4 °C and 30-s vortexing was then carried out on the mixture, followed by centrifuging at 10,000g for 10 min. A total of 400  $\mu$ l of supernatant was added into a 1.5-ml Eppendorf tube containing 250  $\mu$ l of aforementioned sodium phosphate buffer, vortexed and spun at 10,000g for 10 min. An amount of 600  $\mu$ l of supernatant was put into an NMR tube.

**<sup>1</sup>H NMR spectroscopy.** <sup>1</sup>H NMR spectra of urinary and fecal water extract samples were acquired using a Bruker 600 MHz spectrometer (Bruker, Rheinstetten, Germany) at the operating <sup>1</sup>H frequency of 600.13 MHz at a temperature of 300 K. An NMR pulse sequence (recycle delay -90°-t<sub>1</sub>-90°-t<sub>m</sub>-90°-acquisition) and standard parameters were used to obtain standard one-dimensional <sup>1</sup>H NMR spectral data as described in Beckonert *et al.*<sup>67</sup>

**Multivariate statistical data analysis.** <sup>1</sup>H NMR spectra of urine and fecal extracts were automatically phased, referenced to 3-(trimethylsilyl)-[2,2,3,3-<sup>2</sup>H<sub>4</sub>]propionic acid sodium salt at <sup>1</sup>H  $\delta$  0.00 and baseline-corrected using an in-house developed MATLAB script (Dr Tim Ebbels, Imperial College London). The processed NMR spectra (<sup>1</sup>H  $\delta$  0 to 10) were imported to MATLAB (R2012a, MathWorks) and digitized into 20k data points with a resolution of 0.0005 p.p.m. using an in-house developed script. The water peak region in urine (4.7–5.14 p.p.m.) and fecal water spectra (4.7–5.18 p.p.m.) were removed to minimize the effect of the disordered baseline caused by water suppression. In addition, regions (<sup>1</sup>H 0–0.3) in urine spectra and regions (<sup>1</sup>H 0–0.25) in fecal water spectra containing only noise were therefore removed, together with the urea signal in urine (<sup>1</sup>H 5.48–6.24). Because of the heavy peak shifting in urinary spectra, a recursive segment-wise peak alignment method was applied, to improve metabolic biomarker recovery<sup>68</sup>. Probabilistic quotient normalization was subsequently performed on the resulting data sets, to account for dilution of complex biological mixtures<sup>69</sup>. Principal component analysis and orthogonal partial least-squares discriminant analysis was carried out with a unit variance scaling method in SIMCA (P + 13.0) and MATLAB software.  $P$ -values of metabolites were calculated using ANOVA and FDR was also calculated for each metabolite marker<sup>70</sup>. For the two-way clustered heatmap (Fig. 4b) between fecal microbiota and metabolite concentrations in urine and feces, the partial correlation was calculated for each phylotype and metabolite. The correlation was adjusted for country and dietary intervention to account for baseline differences. A pFDR cutoff value of 0.3 was used to define the significance (denoted by '+'), to account for the different numbers of samples in each data set. For each data set (microbiota, fecal metabolites and urinary metabolites), the data were clustered using hierarchical cluster analyses with complete linkage. The optimal number of clusters was chosen as the splitting that maximized the modularity using the correlation as weight of each edge in the network.

**Metabolic reaction network generation.** Using the freely available MetaboNetworks software<sup>19</sup> reactions that occur spontaneously or by means of enzymes linked to human and/or bacterial genomes were identified in the KEGG. All complete genome sequences listed in KEGG that could be associated to the bacterial groups represented on the HITChip were included in creating the database, resulting in a total of 817 bacterial genomes included in the database. Next, the software was used to construct a shortest-path metabolic reaction network between all (fecal and urinary) metabolites significantly different in any of the different dietary comparisons. From this 'global network', two sub-graphs were generated for each of the two dietary switches using significantly associated fecal metabolites. To highlight the differential expression of the metabolites associated with specific pathways, the background shading was added to the graphs to indicate the different interconnecting pathways (Fig. 5a–c). The abbreviations and full names of these metabolites can be found in Supplementary Table 11.

## References

- Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **127**, 2893–2917 (2010).
- Le Marchand, L. K. L. Cancer among Japanese migrants to Hawaii: gene-environment interactions. *Rev. Epidemiol. Sant. Public.* **40**, 425–430 (1992).
- Norat, T., Aune, D., Chan, D. & Romaguera, D. Fruits and vegetables: updating the epidemiologic evidence for the WCRF/AICR lifestyle recommendations for cancer prevention. *Cancer Treat. Res.* **159**, 35–50 (2014).
- O'Keefe, S. J. *et al.* Why do African Americans get more colon cancer than Native Africans? *J. Nutr.* **137**(1 Suppl): 175S–182S (2007).
- Ou, J. *et al.* Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. *Am. J. Clin. Nutr.* **98**, 111–120 (2013).
- Ponz de Leon, M. *et al.* Pattern of epithelial cell proliferation in colorectal mucosa of normal subjects and of patients with adenomatous polyps or cancer of the large bowel. *Cancer Res.* **48**, 4121–4126 (1988).
- Greer, J. B. & O'Keefe, S. J. Microbial induction of immunity, inflammation, and cancer. *Front. Physiol.* **1**, 168 (2011).
- Vipperla, K. & O'Keefe, S. J. The microbiota and its metabolites in colonic mucosal health and cancer risk. *Nutr. Clin. Pract.* **27**, 624–635 (2012).
- Smith, P. M. *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* **341**, 569–573 (2013).
- Bernstein, C. *et al.* Carcinogenicity of deoxycholate, a secondary bile acid. *Arch. Toxicol.* **85**, 863–871 (2011).
- David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
- Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
- Louis, P., Young, P., Holtrop, G. & Flint, H. J. Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. *Environ. Microbiol.* **12**, 304–314 (2010).
- Gibson, G. R., Macfarlane, G. T. & Cummings, J. H. Sulphate reducing bacteria and hydrogen metabolism in the human large intestine. *Gut* **34**, 437–439 (1993).
- Barcenilla, A. *et al.* Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl. Environ. Microbiol.* **66**, 1654–1661 (2000).
- Rajilic-Stojanovic, M. *et al.* Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ. Microbiol.* **11**, 1736–1751 (2009).
- Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
- Nicholson, J. K. & Lindon, J. C. Systems biology: metabolomics. *Nature* **455**, 1054–1056 (2008).
- Posma, J. M., Robinette, S. L., Holmes, E. & Nicholson, J. K. MetaboNetworks, an interactive Matlab-based toolbox for creating, customizing and exploring sub-networks from KEGG. *Bioinformatics* **30**, 893–895 (2014).
- Craciun, S. & Balskus, E. P. Microbial conversion of choline to trimethylamine requires a glycol radical enzyme. *Proc. Natl Acad. Sci. USA* **109**, 21307–21312 (2012).
- Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
- Liu, R. H. Dietary bioactive compounds and their health implications. *J. Food Sci.* **78**(Suppl 1): A18–A25 (2013).
- Wells, J. E. & Hylemon, P. B. Identification and characterization of a bile acid 7 $\alpha$ -dehydroxylating operon in *Clostridium* sp. strain TO-931, a highly active 7 $\alpha$ -dehydroxylating strain isolated from human feces. *Appl. Environ. Microbiol.* **66**, 1107–1113 (2000).
- de Kok, T. M. *et al.* Bile acid concentrations, cytotoxicity, and pH of fecal water from patients with colorectal adenomas. *Digest. Dis. Sci.* **44**, 2218–2225 (1999).
- Bernstein, H., Bernstein, C., Payne, C. M., Dvorakova, K. & Garewal, H. Bile acids as carcinogens in human gastrointestinal cancers. *Mutat. Res.* **589**, 47–65 (2005).
- Winter, J. *et al.* Inhibition by resistant starch of red meat-induced promutagenic adducts in mouse colon. *Cancer Prev. Res.* **4**, 1920–1928 (2011).

27. Toden, S., Bird, A. R., Topping, D. L. & Conlon, M. A. Resistant starch prevents colonic DNA damage induced by high dietary cooked red meat or casein in rats. *Cancer Biol. Ther.* **5**, 267–272 (2006).
28. van Munster, I. P., Tangerman, A. & Nagengast, F. M. Effect of resistant starch on colonic fermentation, bile acid metabolism, and mucosal proliferation. *Dig. Dis. Sci.* **39**, 834–842 (1994).
29. Hylla, S. *et al.* Effects of resistant starch on the colon in healthy volunteers: possible implications for cancer prevention. *Am. J. Clin. Nutr.* **67**, 136–142 (1998).
30. Wells, J. E., Williams, K. B., Whitehead, T. R., Heuman, D. M. & Hylemon, P. B. Development and application of a polymerase chain reaction assay for the detection and enumeration of bile acid 7 $\alpha$ -dehydroxylating bacteria in human feces. *Clin. Chim. Acta* **331**, 127–134 (2003).
31. Devkota, S. *et al.* Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in IL10 $^{-/-}$  mice. *Nature* **487**, 104–108 (2012).
32. Attene-Ramos, M. S., Wagner, E. D., Plewa, M. J. & Gaskins, H. R. Evidence that hydrogen sulfide is a genotoxic agent. *Mol. Cancer Res.* **4**, 9–14 (2006).
33. Kostic, A. D. *et al.* Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res.* **22**, 292–298 (2012).
34. Carbonero, F., Benefiel, A. C., Alizadeh-Ghamsari, A. H. & Gaskins, H. R. Microbial pathways in colonic sulfur metabolism and links with health and disease. *Front. Physiol.* **3**, 448 (2012).
35. Terpstra, O. T., van Blankenstein, M., Dees, J. & Eilers, G. A. Abnormal pattern of cell proliferation in the entire colonic mucosa of patients with colon adenoma or cancer. *Gastroenterology* **92**, 704–708 (1987).
36. Lipkin, M. *et al.* Classification and risk assessment of individuals with familial polyposis, Gardner's syndrome, and familial non-polyposis colon cancer from [3H]thymidine labeling patterns in colonic epithelial cells. *Cancer Res.* **44**, 4201–4207 (1984).
37. Yu, C. C. & Filipe, M. I. Update on proliferation-associated antibodies applicable to formalin-fixed paraffin-embedded tissue and their clinical applications. *Histochem. J.* **25**, 843–853 (1993).
38. Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860–867 (2002).
39. Baron, J. A. in *Cancer Prevention II* (Springer 181, 223–229, 2009).
40. Montufar-Solis, D., Garza, T. & Klein, J. R. T-cell activation in the intestinal mucosa. *Immunol. Rev.* **215**, 189–201 (2007).
41. Kunisawa, J., Takahashi, I. & Kiyono, H. Intraepithelial lymphocytes: their shared and divergent immunological behaviors in the small and large intestine. *Immunol. Rev.* **215**, 136–153 (2007).
42. Caprioli, F. *et al.* Reduction of CD68 $^{+}$  macrophages and decreased IL-17 expression in intestinal mucosa of patients with inflammatory bowel disease strongly correlate with endoscopic response and mucosal healing following infliximab therapy. *Inflamm. Bowel Dis.* **19**, 729–739 (2013).
43. Smith, K., Bui, T. D., Poulosom, R., Kaklamanis, L., Williams, G. & Harris, A. L. Up-regulation of macrophage wnt gene expression in adenoma-carcinoma progression of human colorectal cancer. *Br. J. Cancer* **81**, 496–502 (1999).
44. Tanaka, S. *et al.* Monocyte chemoattractant protein 1 and macrophage cyclooxygenase 2 expression in colonic adenoma. *Gut* **55**, 54–61 (2006).
45. Chang, P. V., Hao, L., Offermanns, S. & Medzhitov, R. The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proc. Natl Acad. Sci. USA* **111**, 2247–2252 (2014).
46. Windey, K., De Preter, V. & Verbeke, K. Relevance of protein fermentation to gut health. *Mol. Nutr. Food Res.* **56**, 184–196 (2012).
47. Juckett, D. A., Aylsworth, C. F. & Quensen, J. M. Intestinal protozoa are hypothesized to stimulate immunosurveillance against colon cancer. *Med. Hypotheses* **71**, 104–110 (2008).
48. Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313**, 1960–1964 (2006).
49. Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoeediting: from immunosurveillance to tumor escape. *Nat. Immunol.* **3**, 991–998 (2002).
50. Waldecker, M., Kautenburger, T., Daumann, H., Busch, C. & Schrenk, D. Inhibition of histone-deacetylase activity by short-chain fatty acids and some polyphenol metabolites formed in the colon. *J. Nutr. Biochem.* **19**, 587–593 (2008).
51. Roediger, W. E. Utilization of nutrients by isolated epithelial cells of the rat colon. *Gastroenterology* **83**, 424–429 (1982).
52. Donohoe, D. R., Collins, L. B., Wali, A., Bigler, R., Sun, W. & Bultman, S. J. The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. *Mol. Cell* **48**, 612–626 (2012).
53. Bordonaro, M., Lazarova, D. L. & Sartorelli, A. C. Butyrate and Wnt signaling: a possible solution to the puzzle of dietary fiber and colon cancer risk? *Cell Cycle* **7**, 1178–1183 (2008).
54. Rosignoli, P., Fabiani, R., De Bartolomeo, A., Fuccelli, R., Pelli, M. A. & Morozzi, G. Genotoxic effect of bile acids on human normal and tumour colon cells and protection by dietary antioxidants and butyrate. *Eur. J. Nutr.* **47**, 301–309 (2008).
55. Wu, S. *et al.* A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat. Med.* **15**, 1016–1022 (2009).
56. Asano, T. K. & McLeod, R. S. Dietary fibre for the prevention of colorectal adenomas and carcinomas (Review). *The Cochrane Library, Issue 4* (2008).
57. Burkitt, D. Diseases of the alimentary tract and western diets. *Pathol. Microbiol.* **39**, 177–186 (1971).
58. Jemal, A. *et al.* Cancer burden in Africa and opportunities for prevention. *Cancer* **118**, 4372–4384 (2012).
59. Claesson, M. J. *et al.* Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS ONE* **4**, e6669 (2009).
60. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541–546 (2013).
61. Van den Bogert, B., De Vos, W. M., Zoetendal, E. G. & Kleerebezem, M. Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples. *Appl. Environ. Microbiol.* **77**, 2071–2080 (2010).
62. Lahti, L., Elo, L. L., Aittokallio, T. & Kaski, S. Probabilistic analysis of probe reliability in differential gene expression studies with short oligonucleotide arrays. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **8**, 217–225 (2011).
63. Lahti, L., Torrente, A., Elo, L. L., Brazma, A. & Rung, J. A fully scalable online pre-processing algorithm for short oligonucleotide microarray atlases. *Nucleic Acids Res.* **41**, e110 (2013).
64. Smyth, G. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. (eds Gentleman, R., Carey, V., Dudoit, S. & R. Irizarry, WH) 397–420 (Springer, 2005).
65. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
66. Bastian, M., Heymann, S. & Jacomy, M. in *Proceedings of the 3<sup>rd</sup> International AAAI Conference on Weblogs and Social Media*, 361–362 (San Jose, CA, USA, 2009).
67. Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2**, 2692–2703 (2007).
68. Veselkov, K. A. *et al.* Recursive segment-wise peak alignment of biological (1)H NMR spectra for improved metabolic biomarker recovery. *Anal. Chem.* **81**, 56–66 (2009).
69. Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabolomics. *Anal. Chem.* **78**, 4281–4290 (2006).
70. Storey, J. D. A direct approach to false discovery rates. *J. R. Stat. Soc. B* **64**, 479–498 (2002).

## Acknowledgements

We thank Chadd and Kate Bain, and the iZulu Lodge and Projects team in KwaZulu-Natal, for help with the recruitment and housing of the African volunteers, and Dr Iain Thirsk and his surgical endoscopy staff at Ngwelezana Hospital, Empangeni, KwaZulu-Natal, for providing the endoscopic facilities for the study, and Faye Brouard for help with the rural dietary assessments. We acknowledge the help of Dr Robert Branch and the staff of the University of Pittsburgh Clinical Translation and Research Center for help with the dietary switch studies on African Americans. We thank Ms Kayellen Umeakunne for help with the design of the intervention diets. Primary funding for the study was provided by a grant from the National Institutes of Health R01 CA135379 (O'Keefe) and CTRC support from UL1 RR024153 and UL1TR000005. The research was also supported by the National Institute for Health Research Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. We acknowledge the Academy of Medical Sciences, which funded part of the metabolic profiling work. Additional funding for the microbiota studies was provided by the Spirozo Award of the Netherlands Organization (de Vos) for Scientific Research, the ERC Advanced Grant 250172 (Microbes Inside) of the European Research Council and the Academy of Finland (grant 141140 to W.M.dV. and grant 256950 to L.L.). The University of Pittsburgh Genomics and Proteomics Core Laboratory (GPCL) produced the mucosal gene expression microarray data.

## Author contributions

S.J.D.O.K. designed and conducted the human studies, supervised the analysis and evaluation the results, prepared, edited and revised the manuscript. E.G.Z. assisted with study design, managed the global microbiota analysis and interpretation, and assisted with preparation of the manuscript. H.R.G. assisted with study design and execution of the human studies, managed the targeted microbiota analysis and interpretation, and assisted with preparation of the manuscript. J.K. was responsible for writing the text relating to the  $^1\text{H}$  NMR analysis and for interpreting the spectral data. J.N. provided strategic oversight for the metabolomic analytical strategy and was responsible for data interpretation. J.L. performed the  $^1\text{H}$  NMR analysis and the multivariate analysis. J.O.

was responsible for targeted analysis of metabolites, data management, sample distribution and manuscript preparation. L.L. and S.T. conducted the microbiota data analysis and interpretation, and assisted with preparation of the manuscript. P.P. sample coordination and microbiota phylogenetic profiling. W.M.D.V. managed, interpreted and supported microbiota analysis. K.M. and H.K. performed the mucosal immunohistochemistry. E.W. (USA), E.R. (USA) and F.B. (S Africa) performed the dietary analysis, meal design and production. K.A.N. and V.N. helped organize the South African studies and endoscopies. K.V. was responsible for the screening, recruitment and management of African Americans dietary switch studies. F.C. performed the targeted microbiota analysis. L.M. supervised the recruitment of African subjects and helped with the dietary exchange studies. A.K. supervised the histological assessments of biopsy samples. A.C.B. coordinated the targeted microbiota analyses. J.D. supervised the targeted metabolite analysis and helped with manuscript preparation. J.P. carried out the metabolic reaction network analysis and interpretation.

### Additional information

**Accession codes:** Microbiome and metabolome data have been deposited in the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.1mn1n>.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** O'Keefe, S. J. D. *et al.* Fat, fiber and cancer risk in African Americans and rural Africans. *Nat. Commun.* 6:6342 doi: 10.1038/ncomms7342 (2015).