**Table S1.**

<table>
<thead>
<tr>
<th>Human gene</th>
<th>Description</th>
<th>Disease or adverse effect</th>
<th>Sampling site</th>
<th>Impact on the microbiota</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MEFV</strong></td>
<td>The <em>MEFV</em> gene product, Pyrin, has been proposed to help control inflammation by interacting with the cytoskeleton in certain white blood cells.</td>
<td>Familial Mediterranean Fever</td>
<td>Faeces&lt;sup&gt;2&lt;/sup&gt;</td>
<td>A study of mutations in the <em>MEFV</em> gene revealed that distinct grouping of microbiomes was dependent on the allele carrier status of the host. FMF patients had lower microbial richness&lt;sup&gt;2&lt;/sup&gt;.</td>
</tr>
<tr>
<td><strong>APOA1</strong></td>
<td>The main protein component of High-Density Lipoprotein (HDL) in plasma promoting cholesterol efflux from the liver for excretion. Cholesterol is a component of bile acids, which restrict bacterial growth in the small intestine, and in turn, bacterial enzymatic activity processes them to secondary bile acids, that are re-absorbed&lt;sup&gt;3&lt;/sup&gt;.</td>
<td>Polymorphisms in the human <em>APOA1</em> gene have been associated with the risk of obesity and cardiovascular disease&lt;sup&gt;4&lt;/sup&gt;&lt;sup&gt;5&lt;/sup&gt;, and hyperlipidaemia&lt;sup&gt;6&lt;/sup&gt;.</td>
<td>Faeces&lt;sup&gt;2&lt;/sup&gt;</td>
<td>DGGE analysis indicated that the microbiotas of Apoa-I deficient mice had a different community structure from the wildtype mice&lt;sup&gt;6&lt;/sup&gt;.</td>
</tr>
<tr>
<td><strong>Leptin</strong></td>
<td>Some of the functions of leptin are to regulate appetite, energy expenditure and metabolism&lt;sup&gt;7&lt;/sup&gt;. Leptin also acts as a cytokine with impacts on immune cells, and reduced leptin levels have been shown to be associated with increased susceptibility to infection&lt;sup&gt;8&lt;/sup&gt;.</td>
<td>Polymorphisms of the leptin receptor (LEPR) have been associated with obesity and type-2 diabetes mellitus&lt;sup&gt;9&lt;/sup&gt;,&lt;sup&gt;10&lt;/sup&gt;. In a recent GWAS of 1,504 women of European ancestry, SNPs at the LEPR locus were observed to be significantly associated with plasma soluble leptin receptor (sOB-R) levels, which are inversely associated with diabetes risk factors&lt;sup&gt;11&lt;/sup&gt;.</td>
<td>Faeces</td>
<td>When comparing obese leptin deficient (ob/-) to heterozygous (ob+/-) or wildtype (+/+), obesity and type-2 diabetes mellitus&lt;sup&gt;9&lt;/sup&gt;,&lt;sup&gt;10&lt;/sup&gt;. In the fa/fa rat model, loss of the leptin receptor resulted in lower levels of total bacteria, and a different species composition (e.g., reduced levels of Bifidobacteria, greater levels of Halomonas)&lt;sup&gt;13&lt;/sup&gt;.</td>
</tr>
<tr>
<td><strong>MYD88</strong></td>
<td>An important signaling molecule that acts as one of the nodes in the information-gathering system of the innate immune system involved in sensing bacterial products. Most importantly, information from sensors of microbial products (e.g., TLRs) is routed through MyD88 in inflammation response pathways&lt;sup&gt;14&lt;/sup&gt;.</td>
<td>A loss of MyD88 compromises the innate immune response to pathogens&lt;sup&gt;14&lt;/sup&gt;.</td>
<td>Caeca&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Wen et al. analyzed the microbiotas using clone libraries and reported that 3 bacterial families (Lactobacillaceae, Rikenellaceae and Porphyromonadaceae) differed in abundance between MyD88-deficient mice and their wildtype counterparts&lt;sup&gt;15&lt;/sup&gt;,&lt;sup&gt;16&lt;/sup&gt;. Turnbaugh et al. also used clone libraries to compare wildtype to MyD88-KO and reported no UniFrac-based clustering by genotype. MyD88 is not required for a caecal bloom of Erysipelotrichaceae when mice are fed a high-fat diet&lt;sup&gt;15&lt;/sup&gt;.</td>
</tr>
</tbody>
</table>

<sup>© 2011 Macmillan Publishers Limited. All rights reserved.</sup>
**TLR**

TLR5 (Toll-Like Receptor) is expressed baso-laterally in gut epithelial cells, and recognizes flagellin, a highly conserved protein unit in the bacterial flagellum. TLR2 is known to bind to components of the cell wall in gram positive bacteria and to several products of gram negative bacteria. Vijay-Kumar et al. reported that a high proportion of mice lacking TLR5 developed metabolic syndrome in one group of related mice, whereas another group that had been derived prior tended to be colitic. Genetic variation at the TLR2 locus in humans is significantly associated with the observed variation in response to tumour necrosis factor (TNF)-blocking agents in patients with rheumatoid arthritis.

Loss of TLR5 resulted in an altered microbiota in the mice with metabolic syndrome compared to wildtype mice from the same derivation group: they displayed an altered overall community diversity and an altered abundance of dominant phylotypes, as well as a greater caecal bacterial load.

**NOD2**

NOD2 is an intracellular pattern recognition receptor that recognizes muramyl dipeptide, a peptidoglycan constituent. Mutations in the NOD2 gene are among the strongest genetic risk factors for the ileal Crohn’s disease in humans. Intestinal tissue composition (assessed by UniFrac) of intestinal tissues for a subset of Crohn’s disease and ulcerative colitis patients were significantly correlated with NOD2 genotype. In humans, NOD2 composite genotype is significantly associated with shifts in the microbial community composition (quantified with RT-PCR).

Compared to wildtype, NOD2-deficient mice have been shown to harbor a greater load of commensal bacteria belonging to both the Firmicutes and Bacteroidetes. Compared to wildtype, NOD2-deficient mice have been shown to harbor a greater load of commensal bacteria belonging to both the Firmicutes and Bacteroidetes (quantified with RT-PCR).

**Defensin genes**

Antimicrobial peptides produced by the host. Importantly for mucosal defense. Secreted into the crypts of the intestine to eliminate bacteria that may otherwise induce inflammation and compromise the epithelial barrier. Fellermann et al. reported an association of β-defensin copy number with Crohn’s disease. Both β- and α-defensin expression levels have been shown to be reduced in some Crohn’s disease patients, and Crohn’s disease is associated with an altered microbiota.

Mice deficient for matrilysin (MMP7) cannot cleave and activate α-defensin: their microbiotas have a reduced percentage of Bacteroidetes compared to wildtype mice. Transgenic mice expressing α-defensin 5 (HD-5) have a higher proportion of Firmicutes. Defensins also appear to reduce the levels of SFBs, which recruit IL-17 producing T cells and stimulate IgA production.
| **RELMB** | A cytokine expressed in the GI tract that has been implicated in innate immunity\(^{31,32}\). It has also been shown to regulate expression of RegIII-γ, an antimicrobial peptide\(^{33}\). RELMB-deficient mice remain relatively lean while fed freely a high-fat diet that renders wildtype mice obese\(^{34}\). | Unknown | Faeces\(^{34}\) | The RELMB (-/-) genotype modestly but significantly impacted the abundance of 15 Bacteroidetes lineages, 1 Proteobacteria lineage and 15 Firmicutes lineages. Much of the difference between the 2 genotypes owed to changes in abundances of relatively low-level lineages\(^{34}\). |
| **IgA locus** | An antibody that has an important role in mucosal immunity. | Humans lacking IgA have higher incidence of inflammatory bowel diseases\(^{35}\). | Small intestine\(^{36}\) | Mice deficient in IgA harbor an increased abundance of the SFBs\(^{36}\). Mice that are unable to export IgA, shed more *Salmonella typhimurium* when infected compared to wildtype mice\(^{36}\). Weight loss has been associated with a decrease in IgA-coated bacteria in humans\(^{37}\). |
| **HLA genes** | HLA genes of the major histocompatibility complex (MHC) genomic region encode cell-surface antigen-presenting proteins. The class II HLAs present antigens to helper T-cell lymphocytes (CD4+) that direct the differentiation of antibody-producing B cells. | In humans, the risk of Coeliac's disease (CD) is strongly linked with HLA gene variation: Class II HLA genotypes DQ2 and DQ8 have a major role in predisposing individuals to the development of CD, and inheritance of specific HLA-DQ genotypes explains 40% of the genetic predisposition in CD\(^{38}\). | Faeces\(^{40-42}\) | Mouse studies have noted differences in faecal microbiotas using fingerprinting techniques, but the strength of the cohort effect was generally underestimated at that time, and may have confounded the results\(^{41,42}\). De Palma *et al.* compared faecal microbiotas of infants with low, medium and high CD risk (based on HLA types). The faecal bacterial analysis (FISH-based) showed small but significant differences between the risk groups in the abundances of common gut bacteria such as the *Bacteroides-Prevotella* group\(^{40}\). |
| **IFN** | Interferons (IFNs) have roles in a variety of immune functions, and mice deficient in IFN signaling pathways are highly susceptible to microbial infections. | IFNs fall into three classes; Thompson *et al.* studied two types of mice with different deficiencies in IFN signaling pathways: STAT1 KO mice that lack the signal transducer and activator of transcription 1 (STAT1) essential for the signalling of type I (IFN-α/β) and type II IFNs (IFN-γ), and interferon regulatory factor 9 (IRF9) primarily involved in type I IFN signaling. | Faeces\(^{43}\) | DGGE analysis of faecal samples collected daily over several weeks revealed that mice deficient in IRF9 had greater variability in microbial composition (daily, and over 5-days intervals) than controls and STAT1-KO mice\(^{43}\). |
References


4. Chen, E.S. et al. APOA1/A5 variants and haplotypes as a risk factor for obesity and better lipid profiles in a Brazilian Elderly Cohort. Lipids 45, 511-7 (2010).


38. Wiburg, O.L. et al. N I n t e r e n t s e c r e t o r y a n t i b o d i e s p r o t e c t a g a i n s t n a t u r a l S a l m o n e l l a t y p h i m u r i n i n f e c t i o n. J. Exp. Med. 203, 21-6 (2006).


