Published in final edited form as:

Nature. 2012 October 4; 490(7418): 44–46. doi:10.1038/490044a.

Resident risks

JULIA OH and JULIA A. SEGRE

National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892-4442, USA

JULIA OH: julia.oh@nih.gov; JULIA A. SEGRE: jsegre@nhgri.nih.gov

Abstract

An innovative method for probing the genomes of the vast community of microorganisms that inhabit the human gut provides an alternative approach to identifying risk factors for type 2 diabetes.

The interplay between genetics and the environment in controlling our health remains an open area of investigation. Technological advances in genome sequencing and analysis have empowered the search for the genetic determinants of human traits and diseases, but genetic variants that directly cause disease have so far been identified for only a small proportion of common diseases. For many other conditions, variants that increase disease risk are known, but the incompleteness of these associations suggests that environmental factors significantly contribute to the likelihood of development of these diseases.

A candidate source of environmental influence on disease development that is gaining increasing attention is the human microbiome — the trillions of microorganisms that inhabit our bodies. Our primary genome is in continual interaction with the combined genomes of these residents, which makes this ‘metagenome’ a fascinating part-environmental, part-genetic factor in human biology. On page 55 of this issue, Qin *et al.*¹ present an innovative method for searching the microbial metagenome for disease risk factors, using as their example type 2 diabetes, a disease that is reaching epidemic proportions in many regions across the world.

Heritability estimates for the common forms of type 2 diabetes range between 30 and 60%, suggesting that it is a complex disorder with multiple interacting genetic and environmental components. Genome-wide association studies (GWAS), which are used to investigate individuals’ DNA sequences for genetic variants associated with disease, have identified variants that contribute to the development of type 2 diabetes, but these only partially explain the variation in individual risk^{2,3}. Other significant risk factors for the disease, such as obesity and diet, are also associated with an altered gut microbiome⁴, suggesting that this disease is a good candidate in which to search for risk factors associated with the human microbial metagenome.

Qin *et al.* surveyed the gut metagenomes (derived from stool samples) of 345 Han Chinese individuals with and without type 2 diabetes using a method called shotgun sequencing, which produces random fragments of DNA sequence from the complex mix of microbial genomes. These fragments are then assembled into larger, gene-length fragments on the basis of sequence overlap. The authors compared these larger fragments with reference genome sequences and gene-function databases to determine the proportions of different bacterial species present in the samples, and consequently the relative representation of various microbial functional and metabolic attributes. They then used these ‘markers’ in a similar way to how variations at specific nucleotides are used in GWAS to identify blocks of

DNA sequences, and carried out a ‘metagenome-wide association study’ (MGWAS) to identify different features of the metagenomes in people with and without the disease.

The MGWAS identified more than 42,000 gene markers associated with type 2 diabetes — a massive and unwieldy number of hits. To reduce the complexity of this gene set and to remove redundant information, Qin *et al.* created metagenomic linkage groups (MLGs) of genes that showed similar profiles in either the patient or the control samples, in terms of taxonomic assignment or relative abundance of the markers (Fig. 1). This approach addressed two major problems with current metagenomic analyses. First, analyses so far have depended on comparisons with available microbial reference sequences to assess the species and functional composition of the metagenome. Such studies are limited by the fact that many detected sequences do not map to these references, so uncharacterized or rare microbes may go undetected. But by using MLGs, rare gene regions can be associated with a larger MLG, and therefore taken into account. Second, associations based on taxonomy alone can be misleading, as gene transfer between species or variation in gene content within a species can complicate taxonomic assignments.

One of the associations with type 2 diabetes identified by Qin and colleagues was a depletion of butyrate-producing bacteria. Butyrate is a molecule used as an energy source by cells in the gut lining, but it has also been suggested to reduce inflammatory responses in the colon⁵. Consequently, a decrease in butyrate may contribute to the increased intestinal inflammation implicated in insulin resistance⁵ and type 2 diabetes. Qin *et al.* also saw an increase in colonization by opportunistic pathogens and an enrichment in microbial genes related to oxidative-stress resistance in people with the disease; both of these changes may further contribute to the inflammatory environment.

However, before the results of this study can be accepted as providing generalizable risk factors for type 2 diabetes, further investigations will be necessary. First, obesity was not a factor in this study, which — given the link between type 2 diabetes and obesity — limits its applicability. Similarly, diet, which is known to have a strong influence on the composition of the gut microbiota⁶, must be considered in future research. Butyrate is a by-product of the metabolism of plant-based foods, but this raises a ‘chicken-and-egg’ question of whether individuals with type 2 diabetes have a diet lacking in these foods to begin with, leading to differences in the composition of their gut microbiomes, or whether it is the microbiota that contributes to disease development. To clarify some of these questions, it would be of great interest for future investigations to cross ethnic lines and to follow patients on controlled diets over extended periods. Furthermore, it is important to note that this study was associative rather than causal, and that controlled prospective studies are needed to better distinguish a metagenomic contribution from potentially confounding factors such as diet and obesity.

Beyond its contribution to our understanding of type 2 diabetes, this study introduced a novel approach — the application of traditional elements of GWAS to metagenomic data, and the introduction of MLGs to analyse the association of metagenomic variants with a disease. This method will be applicable to a wide range of diseases or human traits that are likely to be influenced by the microbiome. The potential interactions between human genetic variation and metagenomic variation are staggeringly complex, and we are only beginning to understand how the human genotype influences the characteristics of its resident microbiota, and how, in turn, the microbiome can influence human health.

Finally, although MGWASs have great potential to account for some of the ‘missing heritability’⁷ in diseases such as type 2 diabetes, this same concept also leads to questions about whether the microbiome, and its effect on human traits, is itself heritable. Infants

inherit, or converge on, similar microbiota to that of their parents, through breast milk, physical contact and shared living. But the gut microbiome is also amazingly plastic, responding rapidly to such factors as dietary changes and antibiotic treatment. And so our growing understanding of the metagenome may in some cases further blur, rather than define, the boundaries between genetic and environmental influence on human traits and disease.

References

1. Qin J, et al. *Nature*. 2012; 490:55–60. [PubMed: 23023125]
2. Morris AP, et al. *Nature Genet*. 2012; 44:981–990. [PubMed: 22885922]
3. Frayling TM. *Nature Rev Genet*. 2007; 8:657–662. [PubMed: 17703236]
4. Turnbaugh PJ, et al. *Nature*. 2009; 457:480–484. [PubMed: 19043404]
5. Roelofsen H, Priebe M, Vonk R. *Benefic Microbes*. 2010; 1:433–437.
6. Turnbaugh PJ, et al. *Sci Transl Med*. 2009; 1:6ra14.
7. Zuk O, Hechter E, Sunyaev SR, Lander ES. *Proc Natl Acad Sci USA*. 2012; 109:1193–1198. [PubMed: 22223662]

Figure 1. Mining the metagenome

Qin *et al.*¹ studied the metagenome of the human gut — the combined genomes of the trillions of its resident microorganisms — in the search for microbial species and gene functions that are associated with type 2 diabetes. First, they identified microbial gene markers and determined the abundance of these markers in stool samples from a group of patients with type 2 diabetes and from a control group of people without the disease. The authors then clustered these markers according to their relative abundance, and assigned markers that had similar abundance profiles within either the patient or the control group to a metagenomic linkage group (MLG). In this example, MLG1 is over-represented in the patient group, MLG2 is similarly represented in patient and control groups and MLG3 is over-represented in the control group. The use of MLGs reduced the number of associations

for subsequent analysis, minimized redundancy and allowed rare gene markers to be taken into account.

\$watermark-text

\$watermark-text

\$watermark-text