Genome Biology

RESEARCH HIGHLIGHT

Functional predictions from inference and observation in sequence-based inflammatory bowel disease research

Folker Meyer[1,*], William L Trimble[1], Eugene B Chang[2] and Kim M Handley[1,3]

See research article www.genomebiology.com/2012/13/9/R79

Abstract

Meta-omics approaches such as metagenomics, metatranscriptomics and metaproteogenomics have the potential to improve our understanding of how the human microbiome affects digestive health and disease.

Inflammatory bowel disease and gut microbiome research

The importance of understanding the microbial contribution to the emergence of inflammatory bowel disease (IBD) cannot be overstated. IBD disorders, such as ulcerative colitis or Crohn's disease, currently afflict an estimated 3.6 million people in Europe and the United States alone, and are becoming increasingly prevalent worldwide [1]. Although the etiology of IBD is unknown, the inflamed gastrointestinal tract in patients with IBD is characterized by an imbalance in associated gut microbiota (dysbiosis). A growing body of evidence indicates that gut dysbiosis may induce or exacerbate IBD, and that this may be linked to a genetic susceptibility in the host [2]. Owing to its prevalence and the likely role of bacteria in the disease, IBD provides a model system for studying the impact that microorganisms have on human health. Host-microbiome and intra-microbiome interactions are complex, addition or subtraction of individual organisms has been shown to induce or inhibit colitis in the gastrointestinal tract under specific conditions [3]; however, attempts to manipulate host-microbiome interactions have had varying outcomes, likely due to heterogeneity among individual hosts in terms of gut microbiota [2] and strain level differences of the gut microbiota.

*Correspondence: folker@anl.gov
[1]Institute for Genomics and Systems Biology, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA
Full list of author information is available at the end of the article

BioMed Central © 2012 BioMed Central Ltd

A large number of bacterial species have been cultivated (and many genomes sequenced) from the human gut in comparison with other environments; however, the number of isolates is estimated to represent only 20 to 56% (reports vary widely) of the total gut microbiome at the species level [4,5]. High-throughput cultivation techniques can generate personalized culture collections that capture over 50% of species-level diversity and substantial strain-level variation [5]. These collections offer the ability to test clonal behavior under defined conditions, or in the presence of specific bacteria. Isolation techniques further facilitate genomic studies of individual organisms, and are essential to improve our ability to meaningfully annotate genes. Culture-based methods, however, are unlikely to uncover the true diversity of community genotypes. In fact, the real genotypic diversity in the human microbiome is almost completely unknown. There is clearly a need for studies that use culture-independent meta-omics techniques to better define metabolic potential and activity at a strain level within microbial communities [6]. In the recent study by Sokol *et al.* [7], the authors investigate the changes in gastrointestinal microbial composition and metabolism in patients with IBD compared with healthy volunteers.

Approaches to studying human host-microbial interactions

With recent advances in sequencing technologies, metagenomic shotgun sequencing of the genomic DNA of complex mixtures of organisms has become a reality [8]. Several research groups are using random sequencing of community DNA to study the genomic potential of microbial communities as a way of understanding their potential contribution to human health and disease. Determining the genes or proteins expressed by these microorganisms using shotgun sequencing of messenger RNA (metatranscriptomics) or mass spectrometry-based shotgun analysis of peptides (metaproteogenomics) is the next logical step. All these methods

allow reconstruction of microbial community metabolism, with metatranscriptomics and metaproteogenomics giving greater insight into the actual active community metabolism.

These meta-omic techniques unlock access to specific strains, and the relative abundances of these strains that are normally present in the human gut or in gastrointestinal tracts affected by IBD. Such techniques have the power to reveal the full range of genetic variation and metabolic processes operating within a microbial community particular to individual hosts. In the future they will enable us to decipher the complex properties of microbial communities interacting with the human host cells.

Current human microbiome studies (for example, the Missouri Adolescent Female Twin study, MetaHit and the Human Microbiome Project) use different sequencing techniques and post-sequencing data transformation strategies, leading to potentially different results, and more importantly to a situation in which results cannot be compared without great efforts being invested in normalization. With sequencing and analysis technologies advancing quickly (such as the new memory reduction method [9]), our ability to reconstruct microbial community genomic compositions and metabolic activity is also improving.

Beyond microbial DNA, mRNA and proteomics, studying metabolites will lead to increased understanding of microbial and microbe-host interactions by supplying increased functional resolution [10]. Complementary human gene expression studies will also be necessary to advance our understanding of host contribution and response and to improve our emerging *in silico* model of IBD.

The gap in diversity between current experiments and sequence databases and annotations

The recent study by Sokol *et al.* [7] uses a wealth of sample material collected from a long running (four year) prospective cohort study to answer questions related to microbiome function associated with IBD. Using a large sample size (27 healthy volunteers and 196 patients with ulcerative colitis/Crohn's disease) and geographical limitation, Sokol *et al.* [7] reconfirm findings from a number of earlier studies [4] that identified specific decreases and increases in the abundance of Firmicutes and Enterobacteriaceae in affected gastrointestinal tracts. The study design allowed the authors to examine the effect of sampling location and age on the measured 16S rDNA taxonomy. By comparing mucosal and luminal samples, the authors also account for variations in the gut microbial community that occur as a function of biogeography.

Ambitiously, the authors of this study [7] chose partial-length 16S amplicon sequencing and a bioinformatics projection approach to characterize microbial community function. They use a novel mapping procedure that relies on 1,200 genome-derived metabolic 'models' from the KEGG database to produce reconstructions of microbial community function across the phylogenetic tree. Of all environments in which to attempt a projection from 16S data to function, the gastrointestinal environment is probably the best candidate, as genome databases are heavily biased toward human pathogens or symbionts.

There is significant uncertainty in projecting from a single gene representation onto a comparatively small collection of reference genomes and then on to metabolism. A direct observation of potential function (metagenome) or expressed function (metatranscriptome) would have been less risky; however, it is often difficult to obtain sufficient quantities of DNA from metagenomic shotgun sequencing to perform such analyses. Furthermore, the study [7] does yield results consistent with findings from previous research on the role of sulfate-reducing bacteria and Proteobacteria. It also confirms existing findings on decreasing carbohydrate metabolism and amino acid biosynthesis in favor of nutrient transport and uptake.

The method used [7] is novel in that it primarily uses a bioinformatics approach to circumvent the formidable challenges that currently exist in defining functional profiles of complex microbiomes (metagenomes, metabolomes and metatranscriptomes), using available genome information of representative microbial taxa. Current approaches are mired in the technical and bioinformatic challenges associated with analyzing large datasets.

Using a 16S-based phylogeny to infer function, however, is highly speculative. Without higher taxonomic resolution (and, realistically, the resolution used in this study allows determination of genera) and clear evidence linking taxonomy to reference genome sequences, readers are left to question the accuracy of the results. Of course, the authors [7] exploit the assumption that taxonomically similar bacteria tend to have functionally similar traits, even though this method is limited by the fact that gene function and pathogenic attributes can vary significantly even within species. Projections and interpretations made by the study are restricted to a predefined space including only well-characterized, cultured genera found in genomic and KEGG pathway databases. The study highlights the limitations of bioinformatically interpolated data because functional inferences made from genomic data are potentially misleading when taken out of physiological context. Factors such as substrate availability, variation in host microbiome composition, regional host factors, genetics, and

other confounding clinical metadata probably affect the expression profiles of the gut microbiome.

The questions that can be asked using these data are also necessarily limited. Because these samples [7] were collected after the initiation of IBD, the microbiota found during active (or even quiescent) disease might not be representative of those that have a role in increasing risk and triggering IBD. The authors [7] recognize this limitation, and we agree with them that the interpretation of data has to be focused on consequential changes in gut microbiota that may have a role in sustaining immune activation and the inflammatory response. In this regard, microbes that can survive in a hostile inflammatory milieu and promote a chronic inflammatory state can establish selective conditions that favor their fitness over other commensal microbiota found in the healthy bowel.

Summary and future directions

In summary, this study [7] uses 16S rRNA gene data to estimate microbiome function in the gastrointestinal tracts of patients with IBD. The results require verification for two reasons. First, there is a lack of strain-resolved information. Second, as the authors themselves state in their closing sentence, techniques such as metatranscriptomics or metabolomics are necessary to better characterize microbiome function.

Despite the speculative nature of the Sokol *et al.* study [7], it will be interesting to observe how their functional inferences compare with studies using more direct genomic approaches to assess the role of microbiome metabolism in gastrointestinal tract inflammatory disease. In our opinion, this study will incentivize others to bring higher resolution tools to bear on the problem. While doing so, these researchers can enhance our understanding of microbiome function in disease if they carefully consider the advantages, disadvantages and predictive power of each method (Figure 3 in [6]).

One important condition for arriving at a meta-omics-based predictive model for IBD will be the presence of high quality functional annotations for reference genomes, which are necessary for building metabolic and regulatory models. It will be important to study physical structure and localization of microbial communities within the gut (for example, placing organisms accurately between human epithelial cells and the lumen), intra-community interactions, and host responses to, and influence on community composition and function. Strain-resolved meta-omics techniques will allow characterization of the microbial component of IBD, and assist in developing an accurate model of disease onset and maintenance.

Author details
[1]Institute for Genomics and Systems Biology, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA. [2]The University of Chicago Medicine, 5841 S. Maryland Avenue, MC 6084, Chicago, IL 60637, USA. [3]Computation Institute, University of Chicago, Searle Chemistry Laboratory, 5735 South Ellis Avenue, Chicago, IL 60637, USA.

Published: 26 September 2012

References

1. Loftus EV Jr: **Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences.** *Gastroenterol* 2004, **126:**1504-1517.
2. Fava F, Danese S: **Intestinal microbiota in inflammatory bowel disease: Friend or foe?** *World J Gastroenterol* 2011, **17:**557-566.
3. Waidmann M, Bechtold O, Frick J-S, Lehr H-A, Schubert S, Dobrindt U, Löffler J, Bohn E, Autenrieth IB: **Bacteroides vulgatus protects against Escherichia coli-induced colitis in gnotobiotic Interleukin-2-deficient mice.** *Gastroenterol* 2003, **125:**162-177.
4. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA: **Diversity of the human intestinal microbial flora.** *Science* 2005, **308:**1635-1638.
5. Goodman AL, Kallstrom G, Faith JJ, Reyes A, Moore A, Dantas G, Gordon JI: **Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice.** *Proc Nat Acad Sci U S A* 2012. doi: 10.1073/pnas.1121464109.
6. Lepage P, Leclerc MC, Joossens M, Mondot S, Blottière HM, Raes J, Ehrlich D, Doè J: **A metagenomics insight into our gut's microbiome.** *Gut* 2012. doi:10.1136/gutjnl-2011-301805.
7. Sokol H, Tickle TL, Morgan XC, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C: **Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment.** *Genome Biol* 2012, in press.
8. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF: **Community structure and metabolism through reconstruction of microbial genomes from the environment.** *Nature* 2004, **428:**37-43.
9. Pell J, Hintze A, Canino-Koning R, Howe A, Tiedje JM, Brown CT: **Scaling metagenome sequence assembly with probabilistic de Bruijn graphs.** *Proc Nat Acad Sci U S A* 2012. doi:10.1073/pnas.1121464109.
10. Tang J: **Microbial metabolomics.** *Curr Genomics* 2011, **12:**391-403.

doi:10.1186/gb-2012-13-9-169
Cite this article as: Meyer F, *et al.*: Functional predictions from inference and observation in sequence-based inflammatory bowel disease research. *Genome Biology* 2012, **13:**169.