Routledge
Taylor & Francis Group

EDITORIAL

A metagenomic approach to investigate the microbial causes of myalgic encephalomyelitis/chronic fatigue syndrome: moving beyond XMRV

Three years ago, a novel association between myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and the murine retrovirus XMRV was published.[1] Since then, 191 papers have been published on the subject (NCBI PubMed, accessed 6 November 2012), largely disproving the initial association, a trend confirmed by a recent multicentre blinded trial which definitively concluded that there is no association between ME/CFS and XMRV.[2] It is therefore time to revisit the investigation of ME/CFS aetiology. Metagenomics offers a promising new opportunity for hypothesis discovery in microbial associations with ME/CFS, and we describe herein the technical basis of this approach and its advantages in aetiological agent investigation.

Metagenomics: a brief primer

Metagenomics is the analysis of all nucleic acid recovered directly from a clinical or environmental sample. Using next-generation sequencing platforms, researchers can sequence the DNA of an entire sample, generating multiple short sequences known as "reads." Computational techniques can then be used to identify the microbes present in the sample (the "microbiome") and their relative abundance. Improvements in sequencing platforms have brought the cost of a typical metagenomics sample to under $200, resulting in an increasing number of metagenomics-based analyses in the literature. A complete microbial census of various body sites was performed on 250 human volunteers as part of the Human Microbiome Project,[3] while subsequent "metagenome-wide association studies" (MGWAS) have compared the microbiomes of healthy individuals to those with various conditions. Amongst other findings, these data have identified associations between inflammatory bowel disease and enterobacteriaceae,[4] colorectal carcinoma and fusobacterium,[5] and type two diabetes and butyrate-producing bacteria.[6]

Metagenomics for ME/CFS

Metagenomics offers an unbiased opportunity to investigate potential novel associations between microbes and ME/CFS. Unlike previous studies, which have examined the host immune response, response to antimicrobial treatment regimens, or

used PCR-based screening, a metagenomics protocol replaces the reductionist search for a specific agent with a more holistic discovery-oriented strategy capable of revealing associations with new candidate aetiological agents, including novel pathogens. Metagenomics also offers several other advantages relative to other experimental approaches, summarised in Table 1. These include technical advantages, such as the elimination of a culture step and the ability to detect low-abundance microorganisms. More general opportunities include the ability to investigate the role of microbial communities and/or functional networks in ME/CFS as opposed to an individual species.

The metagenomics approach is also unique in that even a negative result is of use. Presuming a study is performed with sufficient power and scientific rigour, metagenomics has the potential to detect any microbe in a sample; therefore the lack of observed associations between ME/CFS and microbial entities would provide strong, albeit not conclusive, evidence that the origins of ME/CFS lie in non-infectious causes. However, it must be noted that metagenomics is not able to prove that a disease was instigated by a microbe if it has since been removed from the body site of investigation. Nevertheless, even if the causal microbe is no longer present, it may have changed the composition of the microbiome, by altering the presence or relative abundance of other microbes. Such changes could be detected by metagenomics and be used to diagnose and, in principle, treat the symptoms of ME/CFS, even if not the initial cause.

Methodological considerations

As in genome-wide association studies, which attempt to identify human genetic markers associated with a given condition, cases and controls for MGWAS must be carefully chosen. When selecting cases, it is vital to use validated definitions of ME/CFS to enable comparisons between groups. Examples include the 1994 CDC (Fukuda) definition [7] or the more specific Canadian definition.[8] Further to this, some researchers have suggested that ME/CFS should be divided into subtypes to

Table 1.   Advantages of a metagenomic approach.

Advantage	Explanation
Non-targeted	Can detect an association between any microbe and CFS, even if not previously hypothesised
New species	New microbial species may be identified
Microbial culture not required	Allows potential associations of many more microbes with ME/CFS, since the majority of microorganisms are unculturable
Detection of microbes present at low levels	One run of the Illumina MiSeq "desktop" sequencer can produce over 10 million short DNA sequences, allowing identification of microbes present at very low levels
Functional associations	Analysis of microbial DNA allows genes to be identified and analysed at the functional level
Synergistic associations	Combinations of microbes working together may be associated with ME/CFS
Identification of mutations	Analysis of microbial DNA allows identification of mutations in the microbes present

aid detection of differences between groups.[9] Cases and controls should be closely matched to prevent population stratification (differences in marker frequencies due to ancestry or other fundamental differences in the populations being studied) and detailed clinical and epidemiological information should be routinely collected.

In the laboratory, case and control specimens should be handled in an identical manner to prevent microbial contamination giving a false positive association. Since the metagenome varies widely between body sites,[10] samples from each participant will be limited to the same type to allow for comparability. Previous studies, using other methodologies to investigate a microbial cause for ME/CFS, have investigated samples from the blood and gastrointestinal tract, both of which may be used as sites for metagenomic studies. In terms of the unit of investigation within a sample, previous research into ME/CFS and an infectious agent has traditionally focussed on viruses, however, it is possible that bacteria may play a role, perhaps through the brain–gut axis.[11] Therefore, it may be desirable to investigate all microbes, thus requiring detection of both RNA and DNA. This can be achieved by way of separate sequencing experiments for RNA and DNA, the former requiring DNase treatment and subsequent reverse transcription to convert RNA to cDNA prior to DNA sequencing.

Sample size is also an important factor in the design of metagenomic studies. It may be beyond the scope of small institutions to carry out statistically well-powered studies; however, successful metagenomics analyses can begin with pilot studies, after which interesting findings are followed up with increased sample size or sequencing. Apart from the number of samples, it is important to consider the depth at which we sequence each sample, i.e., the number of DNA sequence reads generated per sample. To detect microbes present at very low levels, a large number of reads are required. For example, viral RNA diluted at a ratio of 1:10,000,000 (viral:human) has been detected from within 22,000,000 read pairs.[12] However, it has been suggested that only tens of thousands of reads are required to detect approximately 250 functional categories within a metagenomic sample.[13] Although exploratory experiments can begin with sequencing on a desktop sequencer, such as the Illumina MiSeq, which generates 30 million sequences per run, the HiSeq will be more useful for detecting low copy number microorganisms, as it generates 200 million sequences in each of eight lanes. Finally, it must be noted that analysis of metagenomic data is computationally demanding and requires unique bioinformatics expertise. It is necessary to perform the analysis in a staged approach in which human DNA is first removed, after which the remaining DNA sequence is classified into species or functional groups. Identification of DNA sequence data is generally performed using one of two approaches: a mapping approach, where reads are aligned to a reference DNA database, or *de-novo* assembly without a reference. Each approach has its advantages and disadvantages, and the results of a study may vary greatly depending on which approach is taken.

Future considerations

Although metagenomics may reveal correlations between microbes and ME/CFS, correlation is not causation and further validation experiments are required to conclusively determine aetiology. A modified version of Koch's postulates has been developed for use in molecular studies. These state that a candidate DNA sequence should be present in most cases and sites of disease, few sequences should be present in those without

disease, and sequences should diminish in frequency with clearing of disease.[14] However, the complex nature of ME/CFS, in which environmental, genetic, or other host factors also play a part, means that even these postulates may not apply. Therefore demonstration of causation may require alternative measures, for example, alleviation of ME/CFS symptoms upon removal of a candidate microbe in a controlled trial.

It is time to put the controversy surrounding ME/CFS and XMRV behind us and refocus our efforts on finding what does cause ME/CFS rather than what does not. Metagenomics has the potential to reveal novel associations between ME/CFS and infectious agents. Given the current controversy surrounding the best form of treatment for ME/CFS, as illustrated by the response to the 2011 PACE trial,[15],[16] evidence for or against a microbial cause would be an important step towards developing a potentially more effective treatment regimen for these under-served patients.

Acknowledgements

References

[1] Lombardi VC, Ruscetti FW, Das Gupta J, Pfost MA, Hagen KS, Peterson DL, et al. Detection of an infectious retrovirus, XMRV, in blood cells of patients with chronic fatigue syndrome. Science. 2009;326:585–589.

[2] Alter HJ, Mikovits JA, Switzer WM, Ruscetti FW, Lo SC, Klimas N, et al. A multicenter blinded analysis indicates no association between chronic fatigue syndrome/myalgic encephalomyelitis and either xenotropic murine leukemia virus-related virus or polytropic murine leukemia virus. mBio. 2012;3.

[3] Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, et al. The NIH Human Microbiome Project. Genome Res. 2009;19:2317–2323.

[4] Garrett WS, Gallini CA, Yatsunenko T, Michaud M, DuBois A, Delaney ML, et al. Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. Cell Host Microbe. 2010;8:292–300.

[5] Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. 2012;22:299–306.

[6] Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490:55–60.

[7] Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A. The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. Ann Intern Med. 1994;121:953–959.

[8] Jason LA, Evans M, Porter N, Brown M, Brown A, Huntnell A, et al. The development of a revised Canadian myalgic encephalomyelitis chronic fatigue syndrome case definition. Am J Biochem Biotechnol. 2010;6:120–135.

[9] Jason LA, Corradi K, Torres-Harding S, Taylor RR, King C. Chronic fatigue syndrome: the need for subtypes. Neuropsychol Rev. 2005;15:29–58.

[10] Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. PLoS Comput Biol. 2012;8:e1002606.

[11] Collins SM, Surette M, Bercik P. The interplay between the intestinal microbiota and the brain. Nat Rev Microbiol. 2012;10(11):735–742.

[12] Moore RA, Warren RL, Freeman JD, Gustavsen JA, Chenard C, Friedman JM, et al. The sensitivity of massively parallel sequencing for detecting candidate infectious agents associated with human tissue. PLoS One. 2011;6(5):e19838.

[13] Hamady M, Knight R. Microbial community profiling for human microbiome projects: tools, techniques, and challenges. Genome Res. 2009;19:1141–1152.

[14] Fredericks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. Clin Microbiol Rev. 1996;9:18–33.

[15] Patients' power and PACE. Lancet. 2011;377:1808.
[16] White PD, Goldsmith KA, Johnson AL, Potts L, Walwyn R, DeCesare JC, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. Lancet. 2011;377:823–836.

Ruth R. Miller
School of Population and Public Health, University of British Columbia, Vancouver, Canada
Email: ruth.miller@bccdc.ca

Jennifer L. Gardy
School of Population and Public Health, University of British Columbia, Vancouver, Canada and British Columbia Centre for Disease Control, Vancouver, Canada

Patrick Tang
Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, Canada and British Columbia Public Health Microbiology and Reference Laboratory, Vancouver, Canada

David M. Patrick
School of Population and Public Health, University of British Columbia, Vancouver, Canada and British Columbia Centre for Disease Control, Vancouver, Canada