UCSF
UC San Francisco Previously Published Works

Title
A tale of two approaches

Permalink
https://escholarship.org/uc/item/0tv2h4hh

Journal
Current Opinion in Neurology, 28(3)

ISSN
1350-7540

Authors
Schubert, Ryan D
Wilson, Michael R

Publication Date
2015-06-01

DOI
10.1097/wco.0000000000000198

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

A tale of two approaches: how metagenomics and proteomics are shaping the future of encephalitis diagnostics

Ryan D. Schubert and Michael R. Wilson

Purpose of review
We highlight how metagenomics and proteomics-based approaches are being applied to the problem of diagnosis in idiopathic encephalitis.

Recent findings
Low cost, high-throughput next-generation sequencing platforms have enabled unbiased sequencing of biological samples. Rapid sequence-based computational algorithms then determine the source of all the nonhost (e.g., pathogen-derived) nucleic acids in a sample. This approach recently identified a case of neuroleptospirosis, resulting in a patient's dramatic clinical improvement with intravenous penicillin. Metagenomics also enabled the discovery of a neuroinvasive astrovirus in several patients. With regard to autoimmune encephalitis, advances in high throughput and efficient phage display of human peptides resulted in the discovery of autoantibodies against tripartite motif family members in a patient with paraneoplastic encephalitis. A complementary assay using ribosomes to display full-length human proteins identified additional autoantibody targets.

Summary
Metagenomics and proteomics represent promising avenues of research to improve upon the diagnostic yield of current assays for infectious and autoimmune encephalitis, respectively.

Keywords
16S rRNA, encephalitis, metagenomics, next-generation sequencing, phage immunoprecipitation sequencing, parallel analysis of translated open reading frame, proteomics

INTRODUCTION

In the United States, there are approximately 20 000 cases of encephalitis annually. Inpatient costs alone total $2 billion, and the significant morbidity and mortality associated with the condition makes it almost certain that the outpatient and rehabilitative costs for these patients are significant as well. A major driver of high healthcare costs and poor outcomes in encephalitis cases is the fact that more than 50% remain undiagnosed despite often exhaustive, expensive, and invasive diagnostic testing [1]. The present review highlights recent progress in two fields, metagenomics and proteomics, that will likely improve the diagnosis rate in cases of encephalitis.

Diagnosis of infectious cases of encephalitis was revolutionized in the 1990s with the development of PCR-based assays by Saiki *et al.* [2]. The ability to rapidly and sensitively diagnosis herpes simplex virus (HSV) encephalitis with PCR coupled with early administration of acyclovir halved mortality in patients treated promptly [3]. Similarly, Josep Dalmau's discovery of an autoimmune basis for many causes of encephalitis has revolutionized our understanding of this disease, showing that cases presumed to be infectious may not be [4]. There is now growing recognition that patients with idiopathic encephalitis may have treatable conditions if rapidly and accurately diagnosed.

Department of Neurology, University of California San Francisco, San Francisco, California, USA

Correspondence to Michael R. Wilson, MD, UCSF Multiple Sclerosis and Neuroinflammation Center, 1500 Owens Street, Suite 320, San Francisco, CA 94158, USA. Tel: +415 353 2069; e-mail: michael.wilson@ucsf.edu

Curr Opin Neurol 2015, 28:283–287
DOI:10.1097/WCO.0000000000000198

1350-7540 Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

KEY POINTS

- Dramatic progress has been made recently in the speed, cost, and output of new gene sequencers, referred to as next-generation sequencing (NGS).
- NGS has enabled the application of metagenomics to the discovery of pathogens in infectious encephalitis cases, with significant diagnostic, public health monitoring, and treatment implications.
- Proteomics is the study of proteins recovered directly from the environment.
- Proteomics-based techniques are particularly useful for the identification and further study of novel autoantigens in autoimmune encephalitis.

INFECTIOUS ENCEPHALITIS

The history of encephalitis diagnostics is rich in molecular techniques. In addition to virus-specific PCRs, PCR amplification of 16S rRNA and 28S rRNA is well established for identifying bacterial and fungal pathogens, respectively [5]. Multiplex PCR assays have also been developed to rapidly test for the presence of multiple neuroinvasive pathogens [6]. In one example called MassTag PCR, conventional PCR primers are labeled with molecules of known masses called Masscodes. If DNA matching any of the primers is present in a sample, it will be amplified. After this step, Masscode-labeled primers remain bound to the sample. Unbound primers are washed away, and the Masscodes remaining on the bound primers are identified with mass spectrometry (MS). MassTag PCR has been validated and is now commercialized as FilmArray, a test that screens a sample for 17 meningitis and encephalitis causing pathogens [7]. Although FilmArray is low cost, efficient, and highly specific, the primary limitation is the requirement for sequence-specific primers and thus knowledge of each microbe's genomic sequence prior to testing. Although the FilmArray is a multiplex assay, it still functions fundamentally as a traditional candidate-based diagnostic regime that combines separate assays, each developed to identify individual pathogens of interest.

Metagenomics-based approaches radically depart from traditional candidate-based infectious disease diagnostics in the sense that they simultaneously interrogate all of the genetic material (i.e., host and nonhost RNA and DNA) recovered from a biologic sample in an unbiased fashion. Metagenomics is made possible by the dramatic progress that has been made in the speed, cost, and output of new gene sequencers. Collectively, current sequencing technologies are referred to as next-generation sequencing (NGS), and several platforms are available commercially [8▪]. The wealth of data generated by NGS have necessitated the development of powerful sequence-based computational algorithms to cope with the problem of ever enlarging datasets. In response to this demand, the field of bioinformatics has expanded in parallel, allowing for rapid analysis of large and complex datasets. Given the enormity of the data generated in a typical sequencing experiment, a cloud-compatible bioinformatics pipeline has been developed to leverage existing filter and alignment algorithms to improve processing speeds [9▪]. Instead of days to weeks, metagenomic data can now be accurately processed in clinically actionable timeframes of minutes to hours.

For patients with encephalitis, metagenomic data are usually obtained from cerebrospinal fluid (CSF) or brain biopsy tissue. First the raw genetic sequence data, typically 10–20 million individual sequences, each 50–200 nucleotides long, are filtered to remove low-quality sequences and any sequences that align to the human genome. This step typically removes 98–99% of the raw sequences. Thus, the remaining 1–2% of the sequence fragments, typically numbering in the thousands or hundreds of thousands, are likely to be nonhuman in origin. Their provenance is determined by searching for similarities to all the genomic data contained in GenBank, the NIH genetic sequence database curated by the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov). GenBank now contains over one trillion sequences from all manner of living organisms (http://www.ncbi.nlm.nih.gov/genbank/statistics). By searching GenBank, one is able to interrogate a patient sample for the presence of any type of infectious agent (e.g., fungi, viruses, bacteria, and parasites) except for prions as the latter lack nucleic acids. In addition, the search algorithms are flexible enough that they register both exact matches and sequences that may only be 20% similar to a previously deposited sequence [10]. Thus, there is the ability to identify novel pathogens that may only have a distant phylogenetic relationship to a known microbe. This latter fact is a critical advance for improving public health surveillance against the wide range of novel microbes, many of which are neuroinvasive viruses [11,12].

NGS is an important step forward from the traditional candidate-based approach to infectious disease diagnostics in which each individual diagnostic test (e.g., antibody test, pathogen-specific PCR assay) is tailored to a particular infection suspected by the physician based on a patient's medical history, risk factors, and physical exam.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

The unbiased approach offered by NGS has three potential advantages: identification of an entirely novel microbe for which a traditional candidate-based test does not exist, identification of a known microbe that is not known to cause a particular patient's disease phenotype, and identification of a microbe known to cause a patient's disease phenotype that is nevertheless rarely tested for because it has such a low pretest probability of being the etiologic agent. If the sensitivity of NGS-based diagnostics proves to be high, then an NGS-based assay may have a fourth advantage to the degree that a negative result would have a high negative predictive value that a patient has an active, ongoing central nervous system (CNS) infection. Knowledge of the latter is becoming more and more critical as there is increasing pressure on physicians to empirically prescribe immunosuppressive medication for encephalitis patients in whom multiple infectious disease diagnostics have come back negative.

Recently, an NGS-based result was demonstrated to affect clinical outcomes in a case of neuroleptospirosis [13▪]. In the report, a young boy with severe combined immunodeficiency (SCID) presented with an idiopathic meningitis that progressed to severe meningoencephalitis over a period of 5 months and three hospitalizations. Exhaustive testing with traditional serologic and PCR-based assays and also a brain biopsy failed to detect a pathogen. Ultimately, empiric immunosuppression was tried, but the patient deteriorated and ultimately required a medically induced coma for status epilepticus and an extra-ventricular drain for hydrocephalus. Within 48 h of receipt, NGS identified leptospirosis in the patient's CSF. Given the patient's critically ill status and the time it would take to perform confirmatory leptospirosis testing in a clinical laboratory, the decision was made to act on the experimental results after a multidisciplinary deliberation between laboratory scientists, biomedical ethicists, the patient's physicians, and the patient's family. The empiric steroids were discontinued, and intravenous penicillin was initiated. The patient was discharged from the ICU 4 days later and was discharged from rehabilitation to home 2 weeks later, returning to school soon thereafter.

Astrovirus (AstV) is an example of a virus previously thought only to cause gastrointestinal disease in humans. However, an NGS-based approach identified it in an immunocompromised boy in Seattle, Washington, who died of idiopathic encephalitis [14]. Fulminant AstV encephalitis was also diagnosed by NGS in a patient who initially presented with bilateral sensorineural hearing loss [15▪]. The patient was a 42-year-old man with chronic lymphocytic leukemia who was immunosuppressed post allogeneic bone marrow transplantation. His neurologic condition progressed over several months to include additional cranial nerve deficits, stupor, and ultimately coma. MRI revealed injury to deep grey matter and brainstem structures, a pattern consistent with viral injury. A comprehensive clinical evaluation was negative for an etiologic diagnosis. A brain biopsy also demonstrated pathologic features consistent with viral infection, but no specific pathogen was identified. Acyclovir, intravenous immunoglobulin, antibiotics, and glucocorticoid therapy failed to result in clinical improvement. As the patient's condition progressed, CSF was sent to researchers for NGS analysis. Within 48 and 96 h of NGS respectively, the patient's CSF was negative for AstV, but 0.0012% of the sequences obtained from the patient's frontal lobe biopsy contained AstV reads. RT-PCR, Sanger sequencing, and *in situ* hybridization confirmed the presence of a neuroinvasive AstV strain, now named HAstV-VA1/HMO-C-UK1. Another case of HAstV-VA1/HMO-C-UK1 encephalitis was discovered in an 18-month-old patient with cartilage hair hypoplasia and associated immunodeficiency who was 6 weeks post-peripheral stem cell transplantation [16]. Within 1 week of testing, AstV reads were detected in 0.0002% of the brain biopsy reads. Retrospective testing of the patient's CSF, stool, and serum samples with a pan-astrovirus PCR test demonstrated the presence of the virus in those specimens. These results have led to the recognition that astrovirus, previously known only to cause gastrointestinal disease, is a cause of encephalitis in immunosuppressed posttransplant patients.

Drawbacks of NGS-based approaches include the significant noise-to-signal ratio in the sequencing data. In other words, the ability to amplify up all the genetic information in a sample is both a blessing and a curse as contaminants in the sample will be amplified as well, including skin contaminants and contaminants from lab reagents. Indeed, a contaminant from the latter source was recently misidentified as a novel cause of viral hepatitis [17]. This difficulty is tied to a more general caution for any genetics-based test used to identify a microbe: obtaining a sequence is not equivalent to fulfilling Koch's postulates, which require, in part, that one actually culture the organism. In the future, we may apply a modified version of Koch's postulates such as those proposed by Fredricks and Relman [18]. For now any microbial identifications made on an NGS platform will have to be validated with traditional diagnostics, most preferably by culture.

AUTOIMMUNE ENCEPHALITIS

The discovery of antigens in idiopathic autoimmune encephalitis cases is also rich in molecular

1350-7540 Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

techniques, particularly proteomics. Traditionally, the targets of autoantibodies have been identified by performing immunoprecipitation of the patient's CSF together with rodent brain tissue and subsequently attempting to identify any rodent neural proteins that co-precipitated with the patient sample via MS [19]. This approach has been successful, with a multitude of new autoantibodies having been discovered in the past decade [20]. Nonetheless, cases remain in which clinical suspicion for humoral autoimmunity is high despite a negative result on the rodent brain tissue immunoprecipitation.

Several proteomics methods have been developed to further characterize antibody–antigen interactions that might underly humoral autoimmune encephalitis cases. Protein display technologies allow the screening of human cDNA expression libraries using bacteriophage, yeast, or mammalian vectors. Sera or CSF from patients is evaluated for the presence of antibodies binding to these expressed libraries. The antigen is subsequently identified by MS or NGS. Protein microarrays are another high-throughput technique in which human proteins are purified and individually spotted on a glass array for subsequent screening against a patient sample.

One novel proteomic approach is phage immunoprecipitation sequencing (PhIP-Seq). PhIP-Seq can be used to identify interactions between an individual's antibody repertoire and individual peptides in a library of over 400 000 overlapping 36mer peptides [21]. The authors named the library the T7 peptidome phage display library (T7-Pep). T7-Pep is a synthetic representation of the entire human proteome, constructed with cDNA derived from the 24 239 known human genomic open reading frames (ORFs). The oligonucleotide fragments that constitute the T7-Pep library are fused to sequences coding for the T7 bacteriophage surface protein, allowing them to be packaged for display. Once displayed on the surface of the phage, T7-Pep is immunoprecipitated with the patient's serum antibody repertoire, and candidate autoantigens are subsequently identified with DNA sequencing. The linking of peptide structure to genetic sequence in display techniques such as T7-Pep allows a researcher to purify and express immunoprecipitated autoantigens, facilitating functional studies. PhIP-Seq has revealed novel candidate autoantigens in CNS paraneoplastic diseases in which commercially available serologic testing had been negative [21]. Loss of posttranslational modifications, loss of tertiary and quarternary structures, and nonphysiologic epitopes are limitations of PhIP-Seq.

Another powerful in-vitro method for identifying autoantigens is the protein interaction discovery using parallel analysis of translated ORFs (PLATOs). PLATO utilizes a collection of 15 483 human cDNAs representing the human ORFeome [22]. From these cDNAs, corresponding mRNAs are transcribed *in vitro*. In-vitro translation of their corresponding proteins is then performed. However, because the mRNA transcript has been engineered to lack a stop codon, the mRNA and its translated protein remain tethered to the ribosome in a technique called ribosome display. After exposing this protein library displayed on ribosomes to an autoantibody-containing sample, candidate binding partners are eluted, and the mRNA still attached to the ribosome is sequenced to determine the protein's identity. Again, this technique links phenotype with genotype in a single screening assay. Limitations are similar to PhIP-Seq, but also include the low translation efficiency for some human proteins. A relative strength is that instead of screening against small peptide fragments as in the T7-Pep method, whole proteins with more of their tertiary folding structure are produced. Therefore, PLATO increases the probability that antigenic epitopes are screened and detected.

Using PLATO, Zhu *et al.* [22] discovered several additional autoantibodies not detected by PhIP-Seq. In a single patient with clinical features suggestive of paraneoplastic disease but with negative serologic testing for known autoantigens, PhIP-Seq identified an interaction between the patient's CSF antibodies and the tripartite motif family (TRIM) members TRIM9 and TRIM67. TRIM proteins are induced by type 1 interferon in innate immune responses and tend to localize to cytoplasmic or nuclear structures. Yet interestingly TRIM9 has been associated with the synaptosomal-associated protein 25 kDa, and through this interaction may regulate synaptic vesicle exocytosis [23]. One could hypothesize that blockade of synaptic vesicle exocytosis could interfere with neurotransmission, thus providing a plausible mechanism of action for these autoantibodies. Using PLATO, several additional TRIM protein family members were identified as targets of the very same patient's CSF antibody repertoire. Again, the authors speculate that the ability of ribosome display in PLATO to produce longer peptides that incorporate tertiary folding structures enhanced the technique's sensitivity. As with a microbe identified by an NGS-based diagnostic, additional evidence always needs to be marshaled to demonstrate that a newly discovered antibody actually plays a pathogenic role and is not merely an epiphenomenon.

CONCLUSION

The urgency for better encephalitis diagnostics is underscored by a growing recognition that severe

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

and seemingly treatment-refractory cases of encephalitis may be treatable with available therapeutics. This recognition has increased pressure on physicians to consider empiric immunosuppression in encephalitis cases for which a pathogen is not readily identified. However, as the neuroleptospirosis case discussed above illustrates, immunosuppression can be life threatening if an infectious cause of encephalitis is missed. Moreover, world health experts and epidemiologists need low-cost, unbiased, high-throughput techniques to quickly identify and monitor novel pathogens, many of which have neuroinvasive potential. Metagenomic and proteomic techniques are being developed and improved upon at a rapid pace. Once the sensitivity and specificity of these approaches are firmly validated and the testing becomes clinically available, there is the potential for their application to lower the rate of undiagnosed encephalitis cases, improve patient outcomes, and enhance public health surveillance efforts.

Acknowledgements

M.R.W. thanks Professors Joseph DeRisi and Stephen Hauser for their mentorship and support.

Financial support and sponsorship

M.R.W. thanks the American Brain Foundation Clinical Research Training Fellowship and the UCSF Clinical and Translational Science Institute KL2 Career Development Program for their support.

Conflicts of interest

There are no conflicts of interest.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- ■ of special interest
- ■■ of outstanding interest

1. Vora NM, Holman RC, Mehal JM, *et al.* Burden of encephalitis-associated hospitalizations in the United States, 1998–2010. Neurology 2014; 82:443–451.
2. Saiki RK, Gelfand DH, Stoffel S, *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 1988; 239:487–491.
3. Whitley RJ, Alford CA, Hirsch MS, *et al.* Vidarabine versus acyclovir therapy in herpes simplex encephalitis. N Engl J Med 1986; 314:144–149.
4. Gable MS, Sheriff H, Dalmau J, *et al.* The frequency of autoimmune N-methyl-D-aspartate receptor encephalitis surpasses that of individual viral etiologies in young individuals enrolled in the California Encephalitis Project. Clin Infect Dis 2012; 54:899–904.
5. Schuurman T, de Boer RF, Kooistra-Smid AM, van Zwet AA. Prospective study of use of PCR amplification and sequencing of 16S ribosomal DNA from cerebrospinal fluid for diagnosis of bacterial meningitis in a clinical setting. J Clin Microbiol 2004; 42:734–740.
6. Hsu CC, Tokarz R, Briese T, *et al.* Use of staged molecular analysis to determine causes of unexplained central nervous system infections. Emerg Infect Dis 2013; 19:1470–1477.
7. Poritz MA, Blaschke AJ, Byington CL, *et al.* FilmArray, an automated nested multiplex PCR system for multipathogen detection: development and application to respiratory tract infection. PLoS One 2011; 6:e26047.
8. ■ Frey KG, Herrera-Galeano JE, Redden CL, *et al.* Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. BMC Genom 2014; 15:96.

Metagenomic sequencing uses next-generation sequencers. In this article, three of the major sequencing platforms (Roche 454 Titanium, Ion Torrent PGM, and Illumina MiSeq) are compared head-to-head to identify the sensitivity and specificity of each. The authors reveal trade-offs in sequencing speed, breadth, and depth of coverage, and length of individual reads. The conclusion is that NGS is ideally suited to clinical and public health applications, in which pathogen identification is the sole focus. The subsequent characterization of pathogen sequences, although, would require additional effort.

9. ■ Naccache SN, Federman S, Veeraraghavan N, *et al.* A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. Genome Res 2014; 24:1180–1192.

The processing of one cerebrospinal fluid sample with NGS generates between 10 and 20 million sequences. GenBank, used as a reference library for sequence alignment, contains over one trillion entries. With these numbers, accurate analysis with conventional bioinformatics tools can take days even with state-of-the-art computer processing. A computational pipeline called sequence-based ultra-rapid pathogen identification (SURPI) was developed that leverages two aligners, SNAP and RAPSearch, to reduce analysis time even under the most stringent circumstances to 16 h for divergent viruses.

10. Stenglein MD, Sanders C, Kistler AL, *et al.* Identification, characterization, and in vitro culture of highly divergent arena viruses from boa constrictors and annulated tree boas: candidate etiological agents for snake inclusion body disease. MBio 2012; 3:e00180-12.
11. Tyler KL. Emerging viral infections of the central nervous system: part 1. Arch Neurol 2009; 66:939–948.
12. Chan BK, Wilson T, Fischer KF, Kriesel JD. Deep sequencing to identify the causes of viral encephalitis. PLoS One 2014; 9:e93993.
13. ■ Wilson MR, Naccache SN, Samayoa E, *et al.* Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med 2014; 370:2408–2417.

From start to finish, NGS can yield actionable information within 48 h. In this case report, a patient with neuroleptospirosis was diagnosed by NGS. The diagnosis was later confirmed by the CDC. High-dose penicillin G returned the patient to his functional baseline.

14. Quan PL, Wagner TA, Briese T, *et al.* Astrovirus encephalitis in boy with X-linked agammaglobulinemia. Emerg Infect Dis 2010; 16:918–925.
15. ■ Naccache SN, Peggs KS, Mattes FM, *et al.* Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. Clin Infect Dis 2015; 60:919–923.

NGS can diagnose novel causes of encephalitis because it can sequence all genetic material in a sample, unlike traditional candidate-based approaches that rely on prior knowledge to develop PCR primers or serologic assays. Astrovirus, previously known as a cause of gastrointestinal disease in humans has now been implicated in cases of undiagnosed encephalitis.

16. Brown JR, Morfopoulou S, Hubb J, *et al.* Astrovirus VA1/HMO-C: an increasingly recognized neurotropic pathogen in immunocompromised patients. Clin Infect Dis 2015; 60:881–888.
17. Naccache SN, Greninger AL, Lee D, *et al.* The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. J Virol 2013; 87:11966–11977.
18. Fredricks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. Clin Microbiol Rev 1996; 9:18–33.
19. Faix PH, Burg MA, Gonzales M, *et al.* Phage display of cDNA libraries: enrichment of cDNA expression using open reading frame selection. Biotechniques 2004; 36:1018–1022; 1024, 1026–1029.
20. Shams'ili S, Grefkens J, de Leeuw B, *et al.* Paraneoplastic cerebellar degeneration associated with antineuronal antibodies: analysis of 50 patients. Brain 2003; 126:1409–1418.
21. Larman HB, Zhao Z, Laserson U, *et al.* Autoantigen discovery with a synthetic human peptidome. Nat Biotechnol 2011; 29:535–541.
22. Zhu J, Larman HB, Gao G, *et al.* Protein interaction discovery using parallel analysis of translated ORFs (PLATO). Nat Biotechnol 2013; 31:331–334.
23. Li Y, Chin LS, Weigel C, Li L. Spring, a novel RING finger protein that regulates synaptic vesicle exocytosis. J Biol Chem 2001; 276:40824–40833.

1350-7540 Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.