Broad-Based Molecular Methods for Viral Discovery, Diagnosis

Molecular methods are instrumental in detecting pathogens that were not associated with human epidemics or animal disease outbreaks

William Check

A disease outbreak in 2007 on Yap, an island in Micronesia, “looked a lot like Dengue,” says Robert Lanciotti of the Arbovirus Diseases Branch of the Centers for Disease Control and Prevention (CDC) in Fort Collins, Colo. However, a pan-flaviviruses array-based analysis of samples revealed that it was not dengue but a Zika virus. “The only prominent thing about Zika is that it is at the end of the list of viruses in the flavivirus family. It had never before been associated with a human epidemic,” he says. “That’s a good justification for these broad-based chips.”

Since 1967, microbiologists and other infectious disease specialists identified more than 40 new pathogens, more than half of them viruses, including HIV, Lassa, Marburg, Ebola, Bornavirus, hepatitis C, Sin Nombre, and Nipah. Moreover, new disease-causing strains of earlier identified viruses were uncovered, including parvovirus B19, adenovirus B14, novel rhinoviruses, and the coronavirus that causes SARS. And several known viruses were recognized for playing new roles, such as epidemic Zika, chikungunya in U.S. travelers returning from India, and West Nile virus, first in New York City and now throughout much of North America.

This litany shows “a real need for broad-based diagnostic assays,” Lanciotti says. “As new diseases emerge and move globally, we face the challenge of testing for many viruses at once. Our lab gets samples from all over the world. We want to be able to detect pathogens that may never before have been associated with a human epidemic.”

Molecular Techniques Help in Responding to Novel Diagnostic Challenges

Responding to this challenge, scientists have devised broad-spectrum molecular detection techniques, including pan-virus microarrays that can theoretically recognize members of every known virus genus, and are proving critical for identifying novel pathogens.

W. Ian Lipkin of the Columbia University Mailman School of Public Health and its College of Physicians & Surgeons in New York, N.Y., began developing such tools during the mid-1980s. For example, he and his collaborators “used only molecular methods to identify” the Borna disease virus in 1990, he says. Back then it required “two years to do what we can now do in a week.”

In 1999, Lipkin and his collaborators helped in identifying West Nile Virus, the cause of an encephalitis outbreak that ap-

Summary

- New analytic approaches prove critical as microbiologists and other infectious disease specialists continue to identify new pathogens, half of which prove to be viruses.
- Pan-virus microarrays theoretically can recognize members of every known viral genus, and are proving critical for identifying novel pathogens.
- Although details differ, the designs behind Virochip and GreeneChip, which is part of a staged strategy, are based on the same principles.
- Molecular methods do not replace classical methods, but complement them.
peared in New York City but subsequently spread throughout much of North America. Their efforts depended on a precursor of today’s molecular methods, domain-specific differential display, which uses degenerate primer sets to hybridize to viral genes representing large taxonomic groups.

Several years later, a pan-virus array called Virochip proved key for identifying the virus responsible for causing a widespread outbreak of severe acute respiratory syndrome (SARS). Joseph DeRisi of the Howard Hughes Medical Institute at the University of California, San Francisco, says that he had just finished designing the second version of Virochip when the SARS outbreak began. CDC officials sent him material from a SARS patient. “The signature on the chip was absolutely clear for coronavirus,” DeRisi says. “It contained a mix of different elements across coronaviruses and an element shared with some astroviruses. We were reasonably certain we had a divergent coronavirus, which was consistent with electron microscope photos from CDC.” Identification was confirmed by sequencing 1 kb of DNA recovered from a single spot on the chip. “The array proved to be a massive purification device,” he adds. “A student literally scraped DNA off the array with a tungsten needle and we sequenced what was on the end of that needle.”

Adapting Molecular Techniques To Uncover other Novel Viruses

Investigators continue using advanced molecular techniques to solve riddles about novel infectious agents in human and veterinary medicine. In 2007, evidence from high-throughput metagenomic sequencing implicated the Israel Acute Paralysis Virus as perhaps responsible for the economically important condition called colony collapse disorder that occurs among domestic honeybees, according to Lipkin and his collaborators. Separately, DeRisi and his collaborators reported a few months ago that Virochip data had provided a “compelling” lead for divergent avian bornaviruses as the etiologic agent of proventricular dilation disease, which threatens parrots, parakeets, and other psittacine birds worldwide.

“Microarrays offer huge theoretical advantages in clinical and diagnostic virology,” says David Schnurr from the California Department of Public Health Viral and Rickettsial Disease Laboratory in Richmond, who collaborates with DeRisi. “Head-to-head, PCR is a bit more sensitive. But with microarrays you can run many tests at the same time.”

“Even with multiplex PCR you are limited to maybe 20 analytes,” Schnurr continues. “Microarrays are pretty much unlimited in number. And they can be designed to detect any agent we know as well as novel agents as long as they are somewhat related to those we know.” He notes that microarrays also offer advantages in surveillance for bioterrorism agents.

In the Zika investigation, Lanciotti’s group at CDC in Colorado used a high-density array that has 12,000 oligonucleotide spots and can detect all arthropod-borne viruses, especially the alphavirus and flavivirus families. To make the array, the investigators chose an extended region of 800 bases that is common to all 70 viruses in those families and overlapping 35-mers were made that “walk the entire region,” says Lanciotti. “The chip has so much redundancy that, if a new flavivirus emerges, there is a high probability it will show some relationship to other flaviviruses on the chip.”

However, because high-density arrays have so much overlap, many spots will light up, even for a novel virus. “You can’t just look at the chip and say, ‘Oh, that’s West Nile virus,’” Lanciotti says. “You have to use a complicated software package to analyze the results.”

Lipkin concurs. “We use algorithms and statistics to establish significant values,” he says. “Our reports display a prototypic virus for the relevant class so we can visualize the location of each probe printed for that particular prototypic agent. And the hybridization result shows which probes perform. So if you have a chimeric virus, either deliberate or natural, the display will recognize it.”

Same Principles but Nuanced Differences Behind Designs for Virochip, GreeneChip

The designs behind Virochip and GreeneChip, which is Lipkin’s version of a pan-virus array, are based on the same principles. Both use 70-mer probes because viruses mutate often, and longer oligomers tolerate more mismatches. Moreover, both arrays, as well as high-throughput sequencing, use random amplification and
labeling. However, random priming reduces sensitivity by several orders of magnitude relative to PCR, according to Lipkin. “Random amplification works well in situations where a large amount of target is present and background is low,” he says. These conditions apply to fluids such as plasma, urine, sputum, and culture supernatants, but not to tissues. “Working with tissue, much of what you amplify will be background material representing the host,” he says. Digesting DNA with DNAase improves matters, but sensitivity is another challenge still to face. “In the next few months we anticipate a significant improvement,” Lipkin says. “I think we’re close.”

While those two chips are based on similar principles, details of their molecular design differ. Virochip contains the most highly conserved 70-mer sequences from every fully sequenced reference viral genome in GenBank. “We leveraged evolution to our advantage,” DeRisi says. Large numbers of viruses within each family share certain evolutionarily conserved regions of the genome. “If we overrepresent conserved stretches of viral families,” he explains, “we maximize the probability that when we have an unknown agent from that family it will hybridize to the array.”

To augment this advantage, DeRisi adds in a taxonomic approach. Virochip contains not only highly conserved elements from viral families, it also has conserved elements at the genus and species level and elements specific to subspecies. “So Virochip encompasses evolutionarily conserved segments as well as the specificity inherent in individual viral entities,” DeRisi says. Virochip is now in its fourth iteration. Version 3 had about 22,000 viral elements, while Virochip 4 was condensed to 16,000 elements. “We use different versions for different purposes,” he says. “It is not a one-chip-fits-all situation.”

GreenChip Is Part of a Staged Strategy for Identifying Novel Viruses

The GreeneChip is used as part of a staged strategy for pathogen surveillance and discovery, according to Lipkin. At the first level is MassTag, a multiplex PCR platform with panels for viruses that infect either the respiratory, enteric, or central nervous systems; hemorrhagic fever viruses; and viruses for tickborne diseases. Reporters are detected by mass spectrometry.

When MassTag fails or when the candidate list exceeds 30 targets, GreeneChip is brought into play (Fig 1). “Where the first two methods fail or we are looking at complex microflora, we use high-throughput sequencing methods,” Lipkin says. “High-throughput sequencing is unbiased and allows an opportunity to consider the entire tree of life.” Each step entails increased cost and time. For instance, MassTag takes six hours, GreeneChip takes 14 hours, and sequencing takes one week.

Viral probes used for the GreeneChip were designed using a database that integrates the International Committee on Taxonomy of Viruses Database (ICTVdB) and GenBank using the Protein Families (Pfam) database of alignments and Hidden Markov Models to address 405,543 viral sequences representing all 2,011 vertebrate virus species.
“Probes were based first and foremost on coding sequences,” Lipkin says. “Two-thirds of probes were designed using the nucleic acid sequences underlying Pfam amino acid sequences.” The remainder of the probes were designed based on coding sequences not represented in Pfam or from noncoding sequences. “Redundancy is a key element in chip design,” Lipkin says. “For any sequence more than 400 nucleotides long, we demand a minimum of four probes; for sequences 200 to 400 nucleotides long, we demand a minimum of two probes.”

Because the probe density required to meet these coverage criteria cannot be met with arrays spotted in house, Lipkin has them printed commercially. “We need hundreds of thousands of probes for some applications,” he says. To deal with other, nonviral pathogens, an additional panmicrobial database (GreenePmdB) was established. In it the GreeneVrdB is supplemented with ribosomal RNA sequence data for fungi, bacteria, and parasites from the Ribosomal Database Project.

**Molecular Tools with Pluses and Minuses Raise Plenty of Questions**

Lipkin mentions several caveats in relying on molecular tools to identify viruses. “All these techniques give you candidates,” he emphasizes. Those candidates need to be further evaluated by using other methods, including PCR, direct dideoxy sequencing, and biological correlations. “It is important to appreciate that molecular methods do not replace classical methods, they complement them,” he adds. “We still grow viruses whenever we can. It is far easier to get a complete sequence from material propagated in culture.” Immunohistochemistry can also help. Lipkin points out that the new paramyxovirus Nipah was identified in part because it binds to antibodies against hendaviruses.

Another factor to keep in mind when identifying new viruses is the increasing connections between veterinary and human medicine. During the onset of the West Nile virus outbreak in North America, for instance, the impact of this virus on crows in the wild and on other birds in zoos provided important clues that could have shortened the discovery time if they had been recognized sooner as connecting with nearby outbreaks involving humans. Lipkin and other investigators advocate closer interactions between professionals who care for humans and those who care for animals.

DeRisi, Lipkin, and their collaborators also have used molecular methods to pursue viruses in other situations. In 2006, DeRisi and colleagues reported using Virochip to find a novel gammaretrovirus in prostate tumors of patients homozygous for a mutation in the gene for ribonuclease L. Last year they reported “unexpected diversity among human coronaviruses and human rhinoviruses” infecting the respiratory tract of humans. “Previously there were thought to be only two rhinovirus clades, A and B,” DeRisi says. “We found—as others have—much more than this.” Such diversity is likely due to the rapid mutation and evolution of RNA viruses, he says. Further, the study demonstrated “the chip’s ability to pick up these very divergent rhinoviruses exactly as predicted.”

The chip is useful in clinical settings, helping in diagnosing infections among patients whose conditions prove unconventional. In one case, a previously healthy young woman with a respiratory tract infection grew worse on antibiotics and was admitted to the hospital. Conventional diagnostic tests were negative. The woman progressed to respiratory failure and was intubated. At this point, specimens were submitted for Virochip analysis under a research protocol. “The result was surprising,” DeRisi says. Virochip detected parainfluenza 4 (PIV4), which was confirmed by nucleotide sequencing and by seroconversion during the course of the illness. “PIV4 is fairly understudied,” DeRisi says. “When we did that work, the whole genome hadn’t been fully sequenced. In medical lore it causes mild self-limiting disease.” He speculates that this may have been a more-virulent strain.

Molecular methods also are good for tracking viruses for epidemiological purposes, according to Schnurr from the California Department of Public Health. “Each winter we do surveillance for influenza,” he says. In one flu season, from about 300 specimens, they could not identify any pathogen in 10 samples. However, with Virochip, they identified viruses in three of those cases, one of which proved to be a new human picornavirus. After sequencing it, DeRisi’s group made primers and found closely related coronaviruses in other respiratory specimens as well as in stool specimens. “So this is a new family of human picornaviruses, and our only entry was via microarray,” he says.
Lipkin’s group also uses their staged molecular strategy for epidemiological studies—in one instance, finding a novel clade of rhinoviruses. In late 2004, the New York State Department of Health recognized an increased incidence of influenza-like illnesses. However, both molecular tests and culture techniques came up negative. Of 79 samples analyzed by MassTag PCR, a pathogen was found in 26, with rhinoviruses in a large proportion of those samples, approximately half of which belonged to a new genetic clade. Subsequent studies revealed that the novel genotype caused severe respiratory tract infections in children in Germany as well as community outbreaks and pediatric respiratory disease nearly worldwide.

Lipkin and his collaborators recently used “unbiased” high-throughput sequencing to identify the arenavirus that led three patients to die of febrile illness shortly after receiving organ transplants from a single donor. The viral sequence data were confirmed by PCR and immunohistochemical and serologic analyses.

Other challenges remain. For instance, current methods probably will fail to identify some viruses, particular those that are unrelated to the ones now in databanks, according to Lipkin. Analysis at the protein level might reveal commonalities with known viruses, such as happened with astroviruses. “But if they are completely novel, we might only see something that looks like a kinase domain or a transmembrane domain,” he says. “They might be picked up by protein family, but not recognized as viral.” Put another way, there is more analytic development work to be done.

SUGGESTED READING


