

MChip, a low density microarray, differentiates among seasonal human H1N1, North American swine H1N1, and the 2009 pandemic H1N1

Gary L. Heil,^a Troy McCarthy,^b Kyoung-Jin Yoon,^c Siyuan Liu,^c Magdi D. Saad,^d Catherine B. Smith,^e Julie A. Houck,^f Erica D. Dawson,^g Kathy L. Rowlen,^g Gregory C. Gray^a

^aEmerging Pathogens Institute and College of Public Health and Health Professions, The University of Florida, Gainesville, FL, USA. ^bCenter for Emerging Infectious Diseases, The University of Iowa, Coralville, IA, USA. ^cIowa State University College of Veterinary Medicine, Ames, IA, USA. ^dUS Naval Medical Research Unit No. 3, Cairo, Egypt. ^eThe Centers for Disease Control and Prevention, Atlanta, GA, USA. ^fDepartment of Chemistry, University of Colorado, Boulder, CO, USA. ^gInDevR Inc., Boulder, CO, USA.

Correspondence: Gary L. Heil, Department of Environmental and Global Health, Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611.

E-mail: glheil@phhp.ufl.edu

Accepted 15 August 2010. Published online 12 October 2010.

Background The MChip uses data from the hybridization of amplified viral RNA to 15 distinct oligonucleotides that target the influenza A matrix (M) gene segment. An artificial neural network (ANN) automates the interpretation of subtle differences in fluorescence intensity patterns from the microarray. The complete process from clinical specimen to identification including amplification of viral RNA can be completed in <8 hours for under US\$10.

Objectives The work presented here represents an effort to expand and test the capabilities of the MChip to differentiate influenza A/H1N1 of various species origin.

Methods The MChip ANN was trained to recognize fluorescence image patterns of a variety of known influenza A viruses, including examples of human H1N1, human H3N2, swine H1N1, 2009 pandemic influenza A H1N1, and a wide variety of avian, equine, canine, and swine influenza viruses. Robustness of the MChip ANN was evaluated using 296 blinded isolates.

Results Training of the ANN was expanded by the addition of 71 well-characterized influenza A isolates and yielded relatively high accuracy (little misclassification) in distinguishing unique H1N1 strains: nine human A/H1N1 (88.9% correct), 35 human A/H3N2 (97.1% correct), 31 North American swine A/H1N1 (80.6% correct), 14 2009 pandemic A/H1N1 (87.7% correct), and 23 negative samples (91.3% correct). Genetic diversity among the swine H1N1 isolates may have contributed to the lower success rate for these viruses.

Conclusions The current study demonstrates the MChip has the capability to differentiate the genetic variations among influenza viruses with appropriate ANN training. Further selective enrichment of the ANN will improve its ability to rapidly and reliably characterize influenza viruses of unknown origin.

Keywords Influenza A Virus, H1N1 Subtype.

Please cite this paper as: Heil *et al.* (2010) MChip, a low density microarray, differentiates among seasonal human H1N1, North American swine H1N1, and the 2009 pandemic H1N1. *Influenza and Other Respiratory Viruses* 4(6), 411–416.

The MChip was developed with the goal of designing a fast, low-cost diagnostic and surveillance tool capable of differentiating among multiple subtypes of influenza A.¹ Prior to MChip development, the classical process of distinguishing influenza strains required a tedious series of viral culture and serologic assays to distinguish the antigenic subtype of the major influenza A antigens, hemagglutinin (HA), and neuraminidase (NA). This process is not commonly used in outpatient settings, because it requires several days to first culture the virus and several more to determine the subtype. Even then, this system is complicated by cross-reactivity among the subtyping antisera. The

development of methods for subtype-specific reverse transcription-PCR (RT-PCR) and subsequent sequencing of all eight influenza gene segments has allowed more accurate determination of virus subtype by inference from the gene sequence.^{2–4} However, RT-PCR's ability to determine specific subtype is limited by the availability of conserved subtype-specific sequences in the genetic targets and the cognate subtype-specific primers. The development of a RT-PCR method to amplify all eight influenza gene segments has facilitated more rapid characterization of full viral genomes from which subtype can be determined by inference from the genetic sequence.⁵ More recently,

real-time RT-PCR assays have come into use for the determination of suspect subtype.⁶ These methods are limited to the detection of only a few subtypes of viruses, therefore the sequencing of additional gene segments is usually required to detect newly emerging viruses, as was the case with the 2009 pandemic H1N1 outbreak. Even though the currently available molecular tools have greatly reduced the time to subtype influenza A viruses, it can still take 2–3 days for confirmation of newly emerged viruses.

Much like matrix-specific real-time PCR protocols, the MChip takes advantage of genetic variations in the matrix (M) gene that may have become conserved among members of specific subtype or host species. However, rather than requiring knowledge of these specific markers like real-time PCR assays, the MChip relies on the differential hybridization of viral RNA to an array of 15 oligonucleotides consisting of sequences designed to be either broadly reactive against all influenza A subtypes or reactive against only a restricted subset of subtypes (Figure 1).

The MChip was the first single gene microarray to indirectly provide subtype information for influenza A when it was reported to be able to distinguish between human influenza A/H3N2, human influenza A/H1N1 and

A/H5N1 in clinical isolates.^{1,7} Since then, further studies have demonstrated that the MChip system can be a powerful tool for rapid and reliable detection of genotypes associated with drug resistance as demonstrated by its ability to detect the presence distinct point mutations in the M-gene associated with resistance to the antiviral adamantane.⁸ Comparison of the MChip assay to viral culture, rapid immunoassay kits, and RT-PCR demonstrated performance characteristics similar to other popular diagnostic techniques, while providing the additional benefit of rapid turnaround times and the ability to differentiate influenza A subtypes.⁹ Overall, these studies suggested that the MChip system may have the ability to distinguish a wider variety of influenza viruses, provided it has adequate training.

The goal of the current work was to further test the capabilities of the MChip by expanding the training of the ANN to encompass additional subtypes of influenza viruses through the addition of unique isolates to those isolates previously used in training for subtype differentiation. This work describes the doubling of the number of images from unique influenza A isolates to include 374 images from characterized isolates in addition to 52 negative samples. Of these 426 images, 78 (plus 20 of the 52 negative samples) were used to train the neural network, while 32 were used as validation examples during automated training. The remaining 264 images (plus the remaining 32 images of negative samples) were used to evaluate the robustness of the MChip ANN.

Methods

Sources of viruses used in this study

In addition to images obtained from isolates in the collection of our laboratory, MChip images for this study were obtained from large collections of well-characterized influenza A isolates at the Veterinary Diagnostic Laboratory of Iowa State University (Ames, IA, USA) and the US Naval Medical Research Unit No. 3 (Cairo, Egypt). Additional isolates and confirmed clinical specimens used by the authors to create microarray images were obtained through agreements with various collaborators. The majority of viruses used in this study were of cultured isolates, including the 2009 pandemic H1N1 prototypical strain A/California/04/2009 (H1N1). The remainder of 2009 pandemic H1N1 viruses used, however, were clinical specimens that had been confirmed by the University of Iowa Iowa State public health laboratory to contain 2009 pandemic H1N1 virus.

Microarray image acquisition and processing

The following standardized methods were used by all three data acquisition sites to process viral isolates and acquire

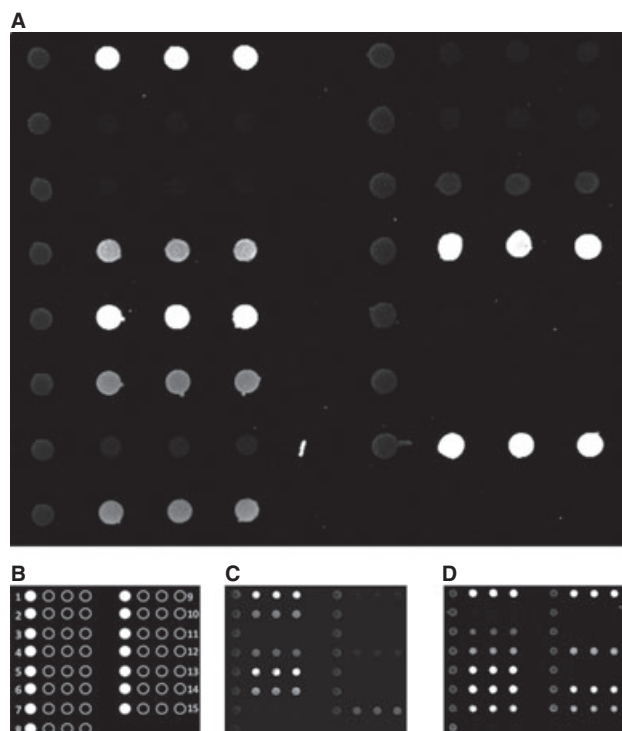


Figure 1. (A) Microarray fluorescence image representative of a typical human A/H1N1 influenza virus. (B) MChip microarray layout with positive control sequences (closed circles) and capture sequences labeled numerically (open circles) (C) Example of swine A/H1N1 image and (D) Typical Pandemic A/H1N1 image. Lighter shading represents greater fluorescence intensity.

MChip images for each of the viruses added to the MChip ANN database during the course of this study.

RNA extraction

Viral RNA was extracted from isolates or clinical specimens using the Qiagen RNeasy kit following the manufactures instructions (Qiagen Inc., Valencia, CA, USA).

RNA amplification

Reverse transcription was performed on the extracted viral RNA as described by Mehlman *et al.*¹⁰ The method of Zou⁴ was used to PCR amplify the first strand cDNA. The 5' primer included a T7 promoter site that facilitated T7 polymerase-mediated run-off transcription. Following transcription, the amplified RNA was immediately fragmented as described by Dawson *et al.*¹ Hybridization was performed immediately after fragmentation as described by Townsend *et al.*¹¹ and carried out in a humidified environment at room temperature for 1 hour.

Imaging and analysis

Slides were scanned using a Molecular Devices GenePix Personal 4100 microarray scanner and analyzed using GenePix Pro 6 microarray image analysis software (Molecular Devices, Sunnyvale, CA, USA). Hybridization efficiency was assessed by visual examination of the positive control spots in each slide. The intensity value for each spot was extracted, and the mean signal for each triplicate set was calculated.

To normalize all the images for analysis by the neural network and account for intensity differences attributed to the RNA concentration used on each slide, relative signal values were used rather than absolute intensities. For each microarray, the oligonucleotide producing the highest mean fluorescence intensity was used to define the maximum intensity and normalized to 100. Subsequently, the mean intensities of all other oligonucleotides were appro-

priately scaled relative to 100. This treatment of the data was not expected to introduce any bias, as relative intensity among the different oligonucleotides on a single chip are solely a function of the hybridization thermodynamics of the RNA attributed to the sequence of the M-gene for that particular isolate to the oligonucleotide sequences on the MChip. The use of multiple isolates (≥ 20) for training against a single subtype output also lessened the likelihood that this treatment of the intensity data would bias outcomes and maximized the ability of the MChip to recognize the broadest range of acceptable representations of a specific trained output.

Training and validation of the neural network

To assist in image analysis, the relative fluorescence intensities of microarray images from a number of characterized viruses were entered into an ANN that learns to recognize fluorescence intensity patterns associated with known isolates. During the automated learning process, additional known viruses were used to challenge the training of the ANN to validate the learning before it was used to characterize query examples. EasyNN-Plus (<http://www.easynn.com>), a commercially available software package, was used to develop the ANN for automated image interpretation as previously described by Dawson *et al.*¹ The ANN used for this study had 16 inputs, seven outputs and had a hidden layer that consisted of nine nodes. The inputs consisted of the mean intensity data for each of the 15 oligonucleotides and the signal to background ratio (S/B) for each slide as the 16th input. The 16th input was included to help differentiate negative samples from influenza A-positive samples, as the S/B for a negative slide should be very low. The six output categories were Human A/H1N1, human A/H3N2, swine A/H1N1, 2009 pandemic A/H1N1, and generic FluA (i.e., influenza A positive, but not belonging to the other output categories, and negative). Although

Table 1. Summary of results from training the artificial neural network (ANN) upon challenge with 217 query influenza A virus isolates and negative slides

Challenge isolates	Trained ANN output classifications and results for query examples <i>n</i> (%)					
	Human H1N1	Human H3N2	Swine H1N1	2009 pandemic H1N1	Influenza A Positive	Influenza A Negative
Human H1N1, <i>n</i> = 9	8 (88.9%)	1 (11.1%)	0	0	0	0
Human H3N2, <i>n</i> = 35	0	34 (97.1%)	0	0	0	1 (2.9%)
Swine H1N1, <i>n</i> = 31	0	0	25 (80.6%)	4 (12.9%)	2 (6.5%)	0
2009 H1N1, <i>n</i> = 14	0	0	1 (7.1%)	12 (87.7%)	1 (7.1%)	0
Other, <i>n</i> = 105	1 (<1.0%)	1 (<1.0%)	2 (1.9%)	0	100 (95.2%)	1 (<1.0%)
Negative, <i>n</i> = 23	0	0	1 (4.3%)	0	1 (4.3%)	21 (91.3%)

sufficient swine A/H3N2 isolates were available to train a swine A/H3N2 output during these studies, insufficient isolates were available to provide query examples for testing, and thus we could not assess MChip performance for identifying this subtype. However, by training swine H3N2 as a possible output, we were able to assess whether the ANN could differentiate other isolates that did not belong in this subtype output. In the ANN database, the outputs of the known examples used in the training set were entered as 1 or 0 designating true or false. The images of 110 known samples and 20 negatives were used to train the ANN, of these 130 images, 32 were used as validating examples in which known samples were used to automatically test the neural network performance during learning. Learning was considered complete only when all 32 validating examples were correctly assigned and the output score was within 5% of the expected value, which was set to 1. The trained neural network was used to determine the subtype of the 296 blinded query examples. Subtype assignment was based on a threshold output score of 0.75.

Results

Additional training of the ANN database yielded relatively high MChip system accuracy in distinguishing unique H1N1 strains (see Table 1): nine human seasonal H1N1 strains (88.9% correct), 35 human seasonal H3N2 (97.1% correct), 31 North American swine H1N1 (80.6% correct), 14 2009 pandemic influenza A H1N1 (87.7% correct), and 23 negative samples (91.3% correct) (shaded boxes in Table 1). The range of human H1N1 and swine H1N1 included in this analysis is summarized in Table 2. For the training of the ANN for the 2009 pandemic H1N1, we made the assumption that the ANN will not differentiate 2009 pandemic H1N1 from other Eurasian H1N1; unfortunately, we did not have sufficient isolates of Eurasian H1N1 to test this hypothesis. In addition to the viruses expected to be classified into one of the specific lineages, 105 influenza A isolates including a wide variety of subtypes of low pathogenic avian influenza, highly pathogenic avian influenza (H5N1), equine (H3N8 and H7N7), and canine (H3N8) influenzas as well as non-H1N1 or non-H3N2 influenza isolates from swine (H3N1, H1N2, and H2N3) were used to challenge the training of the ANN. The ANN performed very well at differentiating these isolates, with very few being misclassified as one of the trained subtypes. Overall, the ANN produced a very low rate of false negative (<1.0%) or false classifications (<9.0%). About 25% of our diverse collection of challenge viruses (79 isolates) were unclassified by the ANN. This result was often attributed to low signal intensity likely due to low viral RNA input, which could easily be improved. These

Table 2. Isolate-specific output classifications for Swine H1N1 and Human H1N1 viruses reported in Table 1

Swine H1N1 isolates	ANN result
A/Swine/North Carolina/7498/2007 (H1N1)	X
unknown reassortant Swine H1N1	X
A/Swine/Virginia/16624/2007 (H1N1)	Misclassified as pandemic H1N1
A/Swine/North Carolina/35774/2007 (H1N1)	X
A/Swine/North Carolina/36260/2007 (H1N1)	X
A/Swine/North Carolina/37822/2007 (H1N1)	X
A/Swine/North Carolina/38614/2007 (H1N1)	X
A/Swine/Ohio/33431/2006 (SwH3N2)	X
A/Swine/England/117316/86	Misclassified as flu A
A/Swine/North Carolina/31430/2006 (H1N1)	X
A/Swine/North Carolina/30777/2006 (H1N1)	X
A/Swine/Wisconsin/238/97 (H1N1)	Misclassified as flu A
A/Swine/Iowa/11583/2007 (H1N1)	Misclassified as flu A
Unknown variant Swine H1N1	X
A/Swine/Iowa/34516/2006 (H1N1)	Misclassified as flu A
A/Swine/Iowa/17948/2007 (H1N1)	X
A/Swine/Iowa/18983/2007 (H1N1)	X
A/Swine/Iowa/35725/2007 (H1N1)	X
A/Swine/Iowa/37613/2007 (H1N1)	X
A/Swine/Iowa/38705/2007 (H1N1)	X
A/Swine/Oklahoma/8680/2006 (H1N1)	X
A/Swine/Iowa/18981/2007 (H1N1)	Misclassified as pandemic H1N1
A/Swine/North Carolina/40688/2006 (H1N1)	X
A/Swine/North Carolina/44504/2006 (H1N1)	X
A/Swine/Iowa/15891/2007 (H1N1)	X
A/Swine/North Carolina/37531/2007 (H1N1)	X
A/Swine/North Carolina/37825/2007 (H1N1)	X
A/Swine/Iowa/45043/2007 (H1N1)	X
A/Swine/Ohio/3129/2007 (H1N1)	X
A/Swine/Virginia/4488/2007 (H1N1)	X
A/Swine/Illinois/14913/2008 (H1N1)	X
Human H1N1 isolates	ANN Results
Unnamed H1N1 clinical specimen Colorado 2006	Misclassified as human H3N2
A/Bangkok/1544/2004 (H1N1)	X
A/Taiwan/1571/2004 (H1N1)	X
A/Singapore/5/2004 (H1N1)	X
A/Krasnoyarsk/51/2005 (H1N1)	X
A/Idaho/1/2005 (H1N1)	X
A/Singapore/23/2004 (H1N1)	X
Unnamed H1N1 clinical specimen Colorado 2006	X
A/Egypt/NAMRU3-2000909671/2001 (H1N1)	X

X indicates a correct classification by the MChip ANN.

isolates were not included in the aforementioned summary, thus only 217 isolates of the 296 blinded isolates could be included in this analysis.

Discussion

The lower success rate observed for swine H1N1 may be attributable to a significant amount of genetic variation among the matrix gene segments of the swine H1N1 viruses available for this (Personal Communication with Dr Kyoung-Jin Yoon). The genetic diversity of the M-gene segment among swine H1N1 was significant enough to be visually noticeable in scanned array images as apparent variations in the intensity pattern (Figure 2). As the number of viruses from each of the known genetic clusters of swine H1N1 viruses was insufficient to train output classifications for each cluster, we attempted to train the ANN with a broad diversity of isolates that represent the genetic variation as a single-output classification. In this instance rather than creating a “catchall” output for swine A/H1N1, it appears that the genetic diversity may have contributed ambiguity to the ANN training resulting in a higher rate of misclassification of swine H1N1 among the challenge isolates. This ambiguity may have also contributed to the number of swine H1N1 isolates that could not be classified by the ANN. Some misclassifications can be attributed to low signal to background (S/B) ratios (<20), which appears to be the case for all misclassified 2009 pandemic A/H1N1 as well as a number of misclassified or unassigned swine

A/H1N1. A few misclassifications can be attributed to errors in chip printing or processing during hybridization. Uniqueness of some isolates may have been a contributing factor in the failure of the ANN to assign any classification for a substantial number of isolates as was the case for some of the more unique avian subtypes used to query the ANN, as they were underrepresented in the training examples. The results from the challenge of the ANN with a large number of isolates from outside the trained subtypes suggests a potential use of this system as an early warning system for detecting the emergence of novel influenza A viruses among sentinel populations.

Conclusions

The results presented here for the current MChip ANN training demonstrates that the MChip has the capability to differentiate genetic variations among influenza A viruses when appropriate training examples are available. Further selective enrichment of the ANN training will likely improve the system’s ability to rapidly and reliably characterize influenza isolates and specimens of unknown origin. One could argue that perhaps the MChip is too sensitive to minor genetic variation in the M-gene as noted in the difficulty it demonstrated in handling the level of genetic diver-

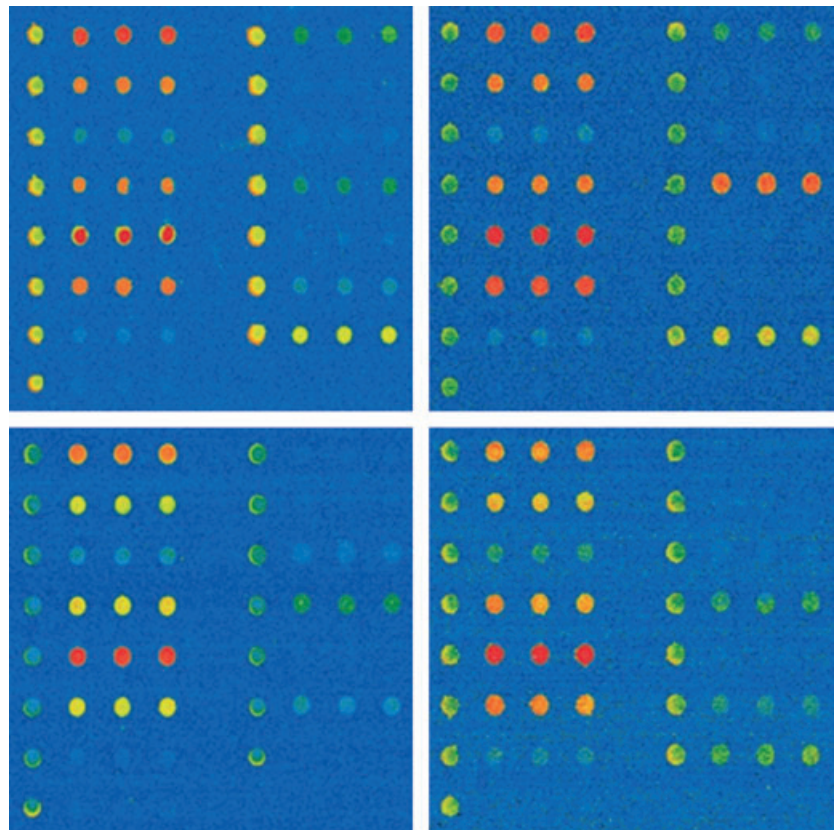


Figure 2. Images representing some of the variability of fluorescence intensity seen among swine A/H1N1 influenza isolates. Slide images are displayed in a color enhanced format to make variations in fluorescence intensities easier to distinguish. Red color represents the highest fluorescent intensity, while blue represents the lowest. Note the most significant difference in sequences 9, 12, and 15 (see Figure 1 for array layout and sequence numbering).

sity expected among the swine A/H1N1 isolates. However, these data also speak of the power of the method to detect subtle genetic changes that may be associated with the phenotypes of emerging viruses. While the MChip appears to possess the potential to be used as a method of detecting emerging M-gene lineages, it would not eliminate the need for sequencing, culture or HA subtyping of emerging strains that may have acquired other gene segments through reassortment.

The two swine H1N1 isolates that were misclassified as 2009 pandemic H1N1 (see Table 2) were of particular interest. While sequence data for the M-gene of these isolates were not available at submission time, differential RT-PCR¹² did confirm that their M-genes were of the North American lineage (data not shown). In addition, visual examination of the chip images from these two misclassified viruses suggests that they were more similar to other North American swine H1N1 isolates than to the 2009 pandemic H1N1 virus. Thus, it is believed that the misclassification of these isolates is the result of noise introduced by the variability of the North American swine H1N1 M-gene lineages used for training the ANN and not the existence of precursors of the 2009 pandemic H1N1 among North American swine.

Acknowledgements

The authors thank the following groups or individuals who kindly contributed prototype viruses, confirmed positive clinical specimens, or provided unique influenza A isolates for MChip imaging in collaboration with Dr Gray and colleagues first at the University of Iowa and now at the University of Florida: Juergen A. Richt, Kansas State University, Manhattan, Kansas; Michael Pentella, University of Iowa Hygienic Laboratory Iowa City, IA; Kristien Van Reeth, Ghent University, Merelbeke, Belgium; Richard Webby, St. Jude's Children's Hospital Memphis, TN; Dennis Senne, USDA/NADC/APHIS Ames, Iowa; Alexander I. Klimov, WHO Collaborating Center for Influenza, Centers for Disease Control and Prevention (CDC), Atlanta, GA; Chris Olsen, University of Wisconsin, Madison, WI, and the American Type Culture Collection, Manassas, VA. This work was supported by multiple grants from the US Department of Defense Global Emerging Infections Surveillance and Response Program (Drs Gray, Blair and Putnam

coprincipal investigators) and a grant from the National Institute of Allergy and Infectious Diseases (R01 AI068803-Dr Gray).

The authors thank Mark G. Lebeck and Sharon F. Setterquist for their technical contributions to this research while working at the University of Iowa's Center for Emerging Infectious Diseases and Ehab A. Saad at U.S. Naval Medical Research Unit 3 for his technical contributions to this work.

References

- 1 Dawson ED, Moore CL, Dankbar DM, Mehlmann M *et al.* MChip: a tool for influenza surveillance. *Anal Chem* 2006; 78:7610–7615.
- 2 Fouchier RAM, Bestebroer TM, Herfst S, Van der Kemp L, Rimmelzwaan GF, Osterhaus A. Detection of influenza A viruses from different species by PCR amplification of conserved sequences in the matrix gene. *J Clin Microbiol* 2000; 38:4096–4101.
- 3 Klimov AI, Cox NJ. PCR restriction analysis of genome composition and stability of cold-adapted reassortant live influenza vaccines. *J Virol Methods* 1995; 52:41–49.
- 4 Zou SM. A practical approach to genetic screening for influenza virus variants. *J Clin Microbiol* 1997; 35:2623–2627.
- 5 Hoffmann E, Stech J, Guan Y, Webster RG, Perez DR. Universal primer set for the full-length amplification of all influenza A viruses. *Arch Virol* 2001; 146:2275–2289.
- 6 Spackman E, Senne DA, Myers TJ *et al.* Development of a real-time reverse transcriptase PCR assay for type A influenza virus and the avian H5 and H7 hemagglutinin subtypes. *J Clin Microbiol* 2002; 40:3256–3260.
- 7 Dawson ED, Moore CL, Dankbar DM, Mehlmann M *et al.* Identification of A/H5N1 influenza viruses using a single gene diagnostic microarray. *Anal Chem* 2007; 79:378–384.
- 8 Townsend MB, Smagala JA, Dawson ED *et al.* Detection of adamantane-resistant influenza on a microarray. *J Clin Virol* 2008; 42:117–123.
- 9 Mehlmann M, Bonner AB, Williams JV *et al.* Comparison of the MChip to viral culture, reverse transcription-PCR, and the QuickVue influenza A plus B test for rapid diagnosis of influenza. *J Clin Microbiol* 2007; 45:1234–1237.
- 10 Mehlmann M, Townsend MB, Stears RL, Kuchta RD, Rowlen KL. Optimization of fragmentation conditions for microarray analysis of viral RNA. *Anal Biochem* 2005; 347:316–323.
- 11 Townsend MB, Dawson ED, Mehlmann M *et al.* Experimental evaluation of the FluChip diagnostic microarray for influenza virus surveillance. *J Clin Microbiol* 2006; 44:2863–2871.
- 12 Harmon K, Bower L, Kim W-I, Pantella M, Yoon K-J. A matrix-gene based multiplex real-time RT-PCR for detection and differentiation of 2009 pandemic H1N1 and other influenza A viruses in North America. *Influenza and Other Respiratory Viruses* 2010 in press.