Methods for Virus Detection and Discovery

W. Ian Lipkin, MD
Director, Center for Infection and Immunity
John Snow Professor of Epidemiology
Mailman School of Public Health
Vagelos College of Physicians & Surgeons
Columbia University

9 January 2024

Tools for Precision Medicine and Public Health

Capture sequencing

rapid, sensitive, inexpensive, straightforward platform for discovery, surveillance, and differential diagnosis

Microarrays and Phage Display

assays for exposure that can elucidate causes of outbreaks and provide early evidence of cross species transmission

HIV/AIDS

Two years from disease emergence to identification of the causative agent

CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

Morbidity and Mortality Weekly Report (MMWR)

Weekly | June 5, 1981 | 30(21);1-3

***Pneumocystis* Pneumonia --- Los Angeles**

In the period October 1980-May 1981, 5 young men, all active homosexuals, were treated for biopsy-confirmed *Pneumocystis carinii* pneumonia at 3 different hospitals in Los Angeles, California. Two of the patients died. All 5 patients had laboratory-confirmed previous or current cytomegalovirus (CMV) infection and candida mucosal infection.

Weekly | December 10, 1982 | 31(48);625-9

Epidemiologic Notes and Reports Possible Transfusion-Associated Acquired Immune Deficiency Syndrome (AIDS) -- California

CDC has received a report of a 20-month old infant the from San Francisco area who developed unexplained cellular immunodeficiency and opportunistic infection. This occurred after multiple transfusions, including a transfusion of platelets derived from the blood of a male subsequently found to have the acquired immune deficiency syndrome (AIDS).

Science
AAAS

20 May 1983 | Vol 220, Issue 4599 | pp. 868-871 | DOI: 10.1126/science.6189183

BARRÉ-SINOUSSI F, CHERMANN JC, … , MONTAGNIER L

Isolation of T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)

4 May 1984 | Vol 220, Issue 4648 | pp. 500-503 | DOI: 10.1126/science.6200936

GALLO RC, SALAHUDDIN SZ, … , MARKHAM PD

Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS

454 Pyrosequencing: Dandenong Virus

Three weeks to identification of causative agent in 3 transplant recipients linked a single donor

nature **Rapid sequencer puts virus in the frame for deaths** Heidi Ledford | 02 May 2007 | doi: 10.1038/447012b

"Although this sort of sequencing has been used to identify viruses in the past, the 454 technology cuts down on time and effort."
Anthony Fauci

nature biotechnology **The development and impact of 454 sequencing**
Jonathan M Rothberg & John H Leamon | 09 October 2008 | doi: 10.1038/nbt1485

The 454 Sequencer has dramatically increased the volume of sequencing conducted by the scientific community and expanded the range of problems that can be addressed by the direct readouts of DNA sequence. Key breakthroughs in the development of the 454 sequencing platform included higher throughput, simplified all *in vitro* sample preparation and the miniaturization of sequencing chemistries, enabling massively parallel sequencing reactions to be carried out at a scale and cost not previously possible. Together with other recently released next-generation technologies, the 454 platform has started to democratize sequencing, providing individual laboratories with access to capacities that rival those previously found only at a handful of large sequencing centers…In longer term, the principles established by the 454 sequencing might reduce cost further, potentially enabling personalized genomics.

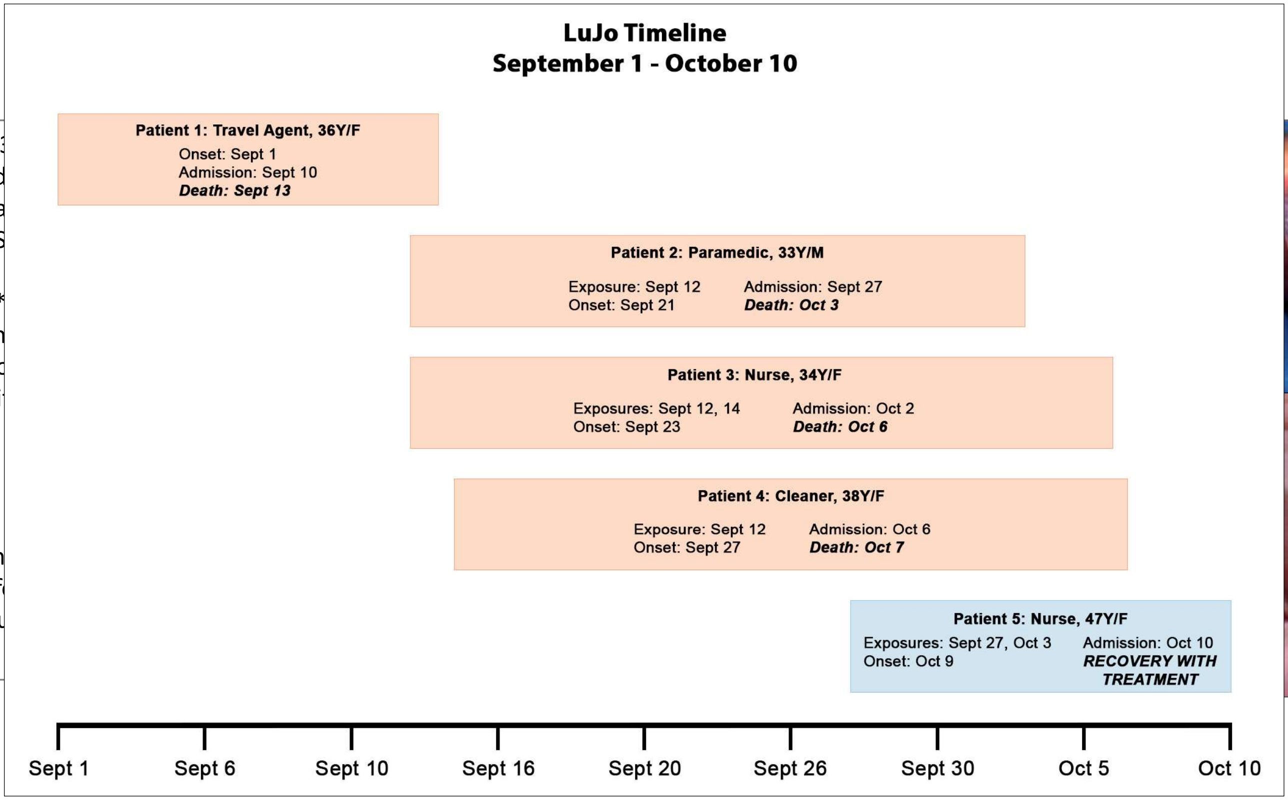
454 Pyrosequencing: LuJo Virus

80% mortality; one week from report to identification of causative agent

Date: Fri, 10 Oct 2008 17:00:3
From: ProMED-mail <promed
Subject: PRO/AH/EDR> Undia
UNDIAGNOSED FATALITIES - S

****

On 12 Sep 2008, an office em
underwent medical evacuatio
died in a Johannesburg hospi
for the index case during her
Johannesburg where he died
in South Africa died on 5 Oct
in South Africa at the Special
(NICD) of the National Health
series of viral haemorrhagic f
identify the pathogen contin
performed.



LuJo Timeline
September 1 - October 10

Patient 1: Travel Agent, 36Y/F
Onset: Sept 1
Admission: Sept 10
Death: Sept 13

Patient 2: Paramedic, 33Y/M
Exposure: Sept 12
Onset: Sept 21
Admission: Sept 27
Death: Oct 3

Patient 3: Nurse, 34Y/F
Exposures: Sept 12, 14
Onset: Sept 23
Admission: Oct 2
Death: Oct 6

Patient 4: Cleaner, 38Y/F
Exposure: Sept 12
Onset: Sept 27
Admission: Oct 6
Death: Oct 7

Patient 5: Nurse, 47Y/F
Exposures: Sept 27, Oct 3
Onset: Oct 9
Admission: Oct 10
RECOVERY WITH TREATMENT

Capture Sequencing: VirCapSeq-VERT

Faster, cheaper, 1,000-fold more sensitive than standard next-gen sequencing (NGS)

H1N1 INFLUENZA is one of many viruses snared by a single new test.

Trawling for Viruses

A new method identifies every virus in a given sample with near-perfect accuracy

When doctors want to identify the virus behind an infection, they usually turn to the polymerase chain reaction (PCR), a method for "amplifying" scattered bits of DNA into a sample large enough to study. To use PCR, however, a physician must know what kind of virus to look for, and that involves guesswork.

This past September a team of Columbia University researchers described a new method that could eliminate that guesswork. The technique, which has the unfortunate name of "virome capture sequencing platform for vertebrate viruses," or VirCapSeq-VERT, can find every virus in a given drop of saliva, tissue or spinal fluid with near-perfect accuracy. The method makes it possible to simultaneously analyze 21 samples in less than 48 hours at an estimated cost of just $200 per sample. It can also detect novel or mutated viruses, so long as they are at least 40 percent identical to known ones. "When someone goes into an emergency room and winds up having all kinds of tests run, it costs thousands of dollars," says W. Ian Lipkin, John Snow Professor of Epidemiology at Columbia University's Mailman School of Public Health. "This method is very inexpensive and allows us to personalize medicine by telling you exactly what you have."

To develop the technique, Lipkin and his colleagues first created a database of more than 1,000 vertebrate viruses. Then they synthesized genetic probes to match every strain of every virus—two million of them, each a strand of DNA 25 to 50 nanometers long. When a probe encounters a matching virus, it binds to it. To extract those viruses, laboratory workers add magnetic beads measuring one to three microns in diameter to the mix; a chemical linker binds the beads to the genetic probes and the viruses they

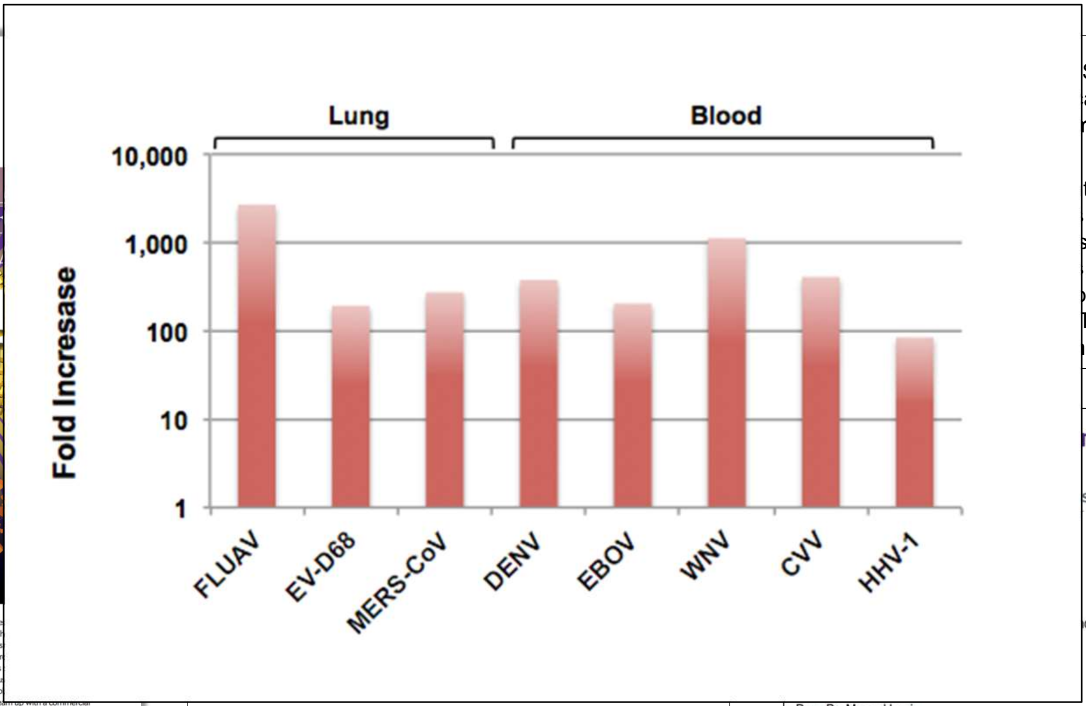
have captured. Re
tube containing th
stand, which pulls
walls. After resear
probe-bead-virus
sequence the virus
false positives. Lip
now looking to te
provider that can distribute the technology to hospitals and clinics around the world. They are also planning on adding probes for all known infectious bacteria and fungi.

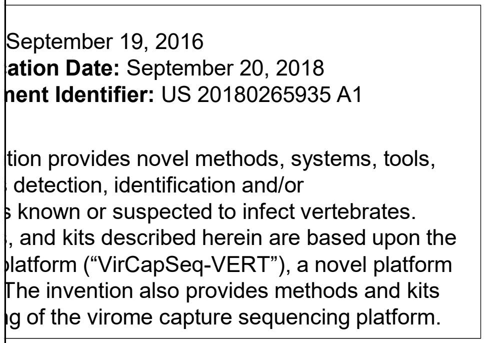
—R.N.

SCIENCE SOURCE (virus)

34 Scientific American, December 2015

© 2015 Scientific American

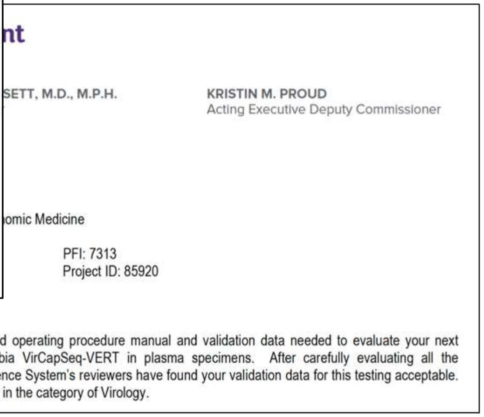




September 19, 2016
ation Date: September 20, 2018
nent Identifier: US 20180265935 A1

tion provides novel methods, systems, tools,
detection, identification and/or
known or suspected to infect vertebrates.
, and kits described herein are based upon the
platform ("VirCapSeq-VERT"), a novel platform
The invention also provides methods and kits
g of the virome capture sequencing platform.

Columbia University freely provides the underlying intellectual property for public health.



nt

SETT, M.D., M.P.H.

KRISTIN M. PROUD
Acting Executive Deputy Commissioner

omic Medicine

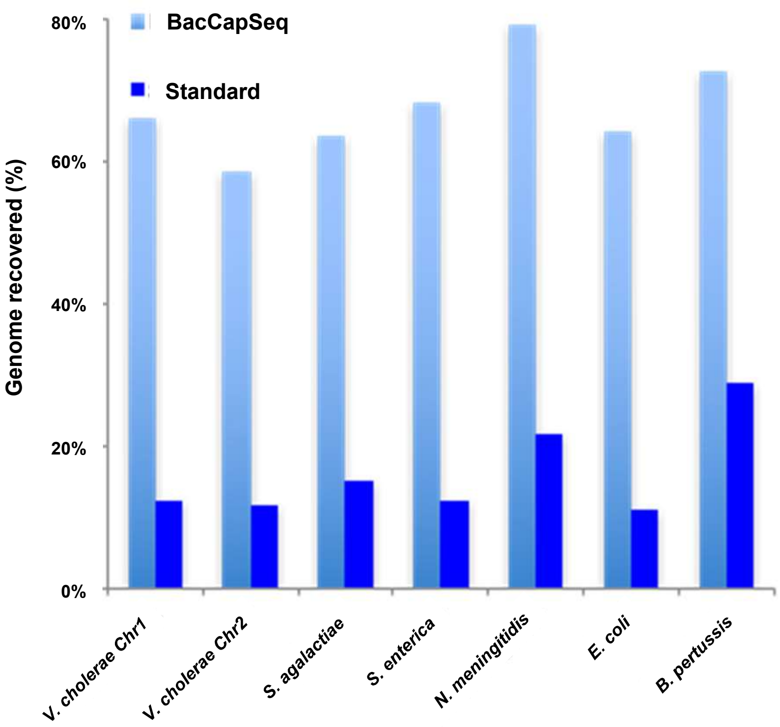
PFI: 7313
Project ID: 85920

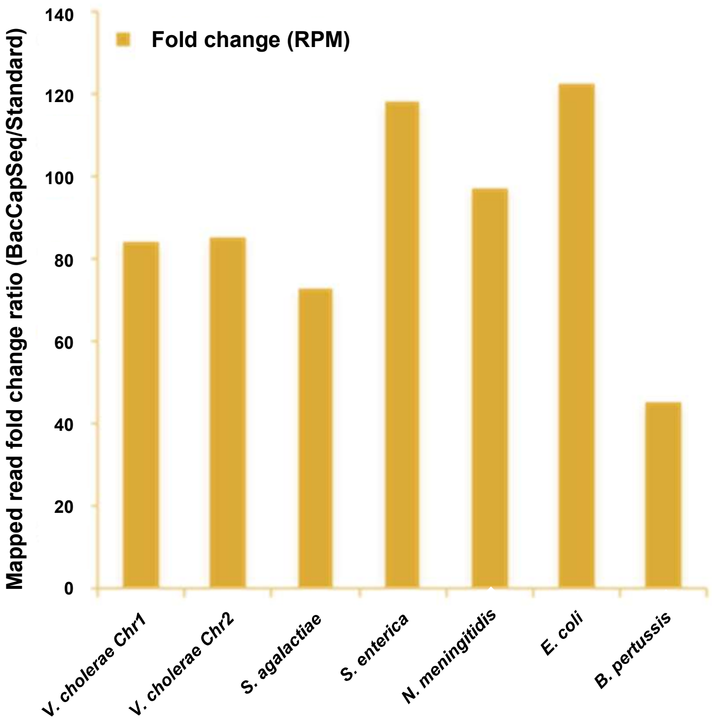
Dear Dr. Mansukhani:

Thank you for submitting portions of your standard operating procedure manual and validation data needed to evaluate your next generation sequencing-based method for Columbia VirCapSeq-VERT in plasma specimens. After carefully evaluating all the information provided, the Clinical Laboratory Reference System's reviewers have found your validation data for this testing acceptable. You may offer this testing under your current permit in the category of Virology.

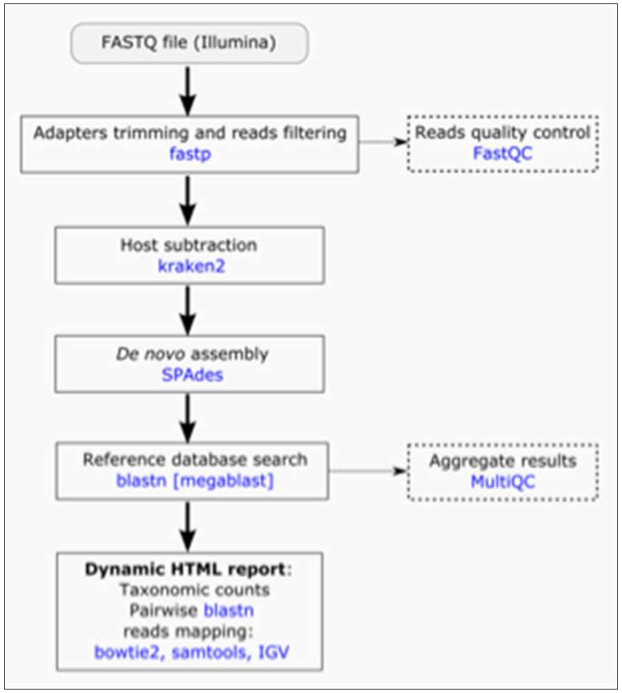
Capture Sequencing: BacCapSeq

Rapid Detection of Bacterial and Antimicrobial Resistance Elements





Rapid Identification of Microbes (RIM)



FASTQ file (Illumina)

Adapters trimming and reads filtering
fastp

Reads quality control
FastQC

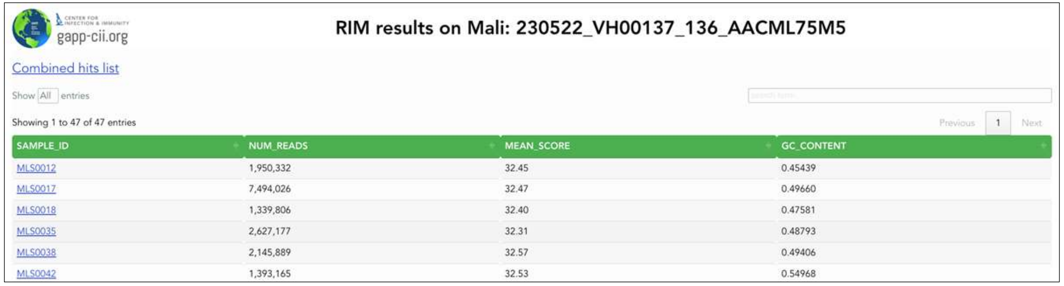
Host subtraction
kraken2

De novo assembly
SPAdes

Reference database search
blastn [megablast]

Aggregate results
MultiQC

Dynamic HTML report:
Taxonomic counts
Pairwise blastn
reads mapping:
bowtie2, samtools, IGV



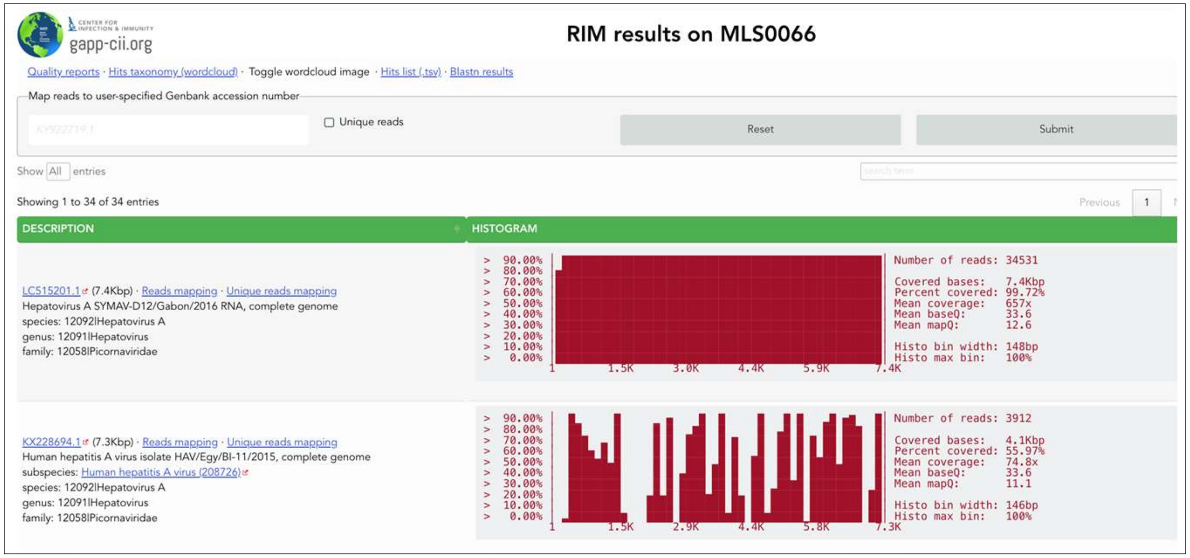
RIM results on Mali: 230522_VH00137_136_AACML75M5

Combined hits list

Show All entries

Showing 1 to 47 of 47 entries

SAMPLE_ID	NUM_READS	MEAN_SCORE	GC_CONTENT
MLS0012	1,950,332	32.45	0.45439
MLS0017	7,494,026	32.47	0.49660
MLS0018	1,339,806	32.40	0.47581
MLS0035	2,627,177	32.31	0.48793
MLS0038	2,145,889	32.57	0.49406
MLS0042	1,393,165	32.53	0.54968



RIM results on MLS0066

Quality reports · Hits taxonomy (wordcloud) · Toggle wordcloud image · Hits list (.tsv) · Blastn results

Map reads to user-specified Genbank accession number

Show All entries

Showing 1 to 34 of 34 entries

DESCRIPTION	HISTOGRAM
LC515201.1 (7.4Kbp) · Reads mapping · Unique reads mapping Hepatovirus A SYMAV-D12/Gabon/2016 RNA, complete genome species: 12092|Hepatovirus A genus: 12091|Hepatovirus family: 12058|Picornaviridae	Number of reads: 34531 Covered bases: 7.4Kbp Percent covered: 99.72% Mean coverage: 657x Mean baseQ: 33.6 Mean mapQ: 12.6 Histo bin width: 148bp Histo max bin: 100%
KX228694.1 (7.3Kbp) · Reads mapping · Unique reads mapping Human hepatitis A virus isolate HAV/Egy/BI-11/2015, complete genome subspecies: Human hepatitis A virus (208726) species: 12092|Hepatovirus A genus: 12091|Hepatovirus family: 12058|Picornaviridae	Number of reads: 3912 Covered bases: 4.1Kbp Percent covered: 55.97% Mean coverage: 74.8x Mean baseQ: 33.6 Mean mapQ: 11.1 Histo bin width: 146bp Histo max bin: 100%

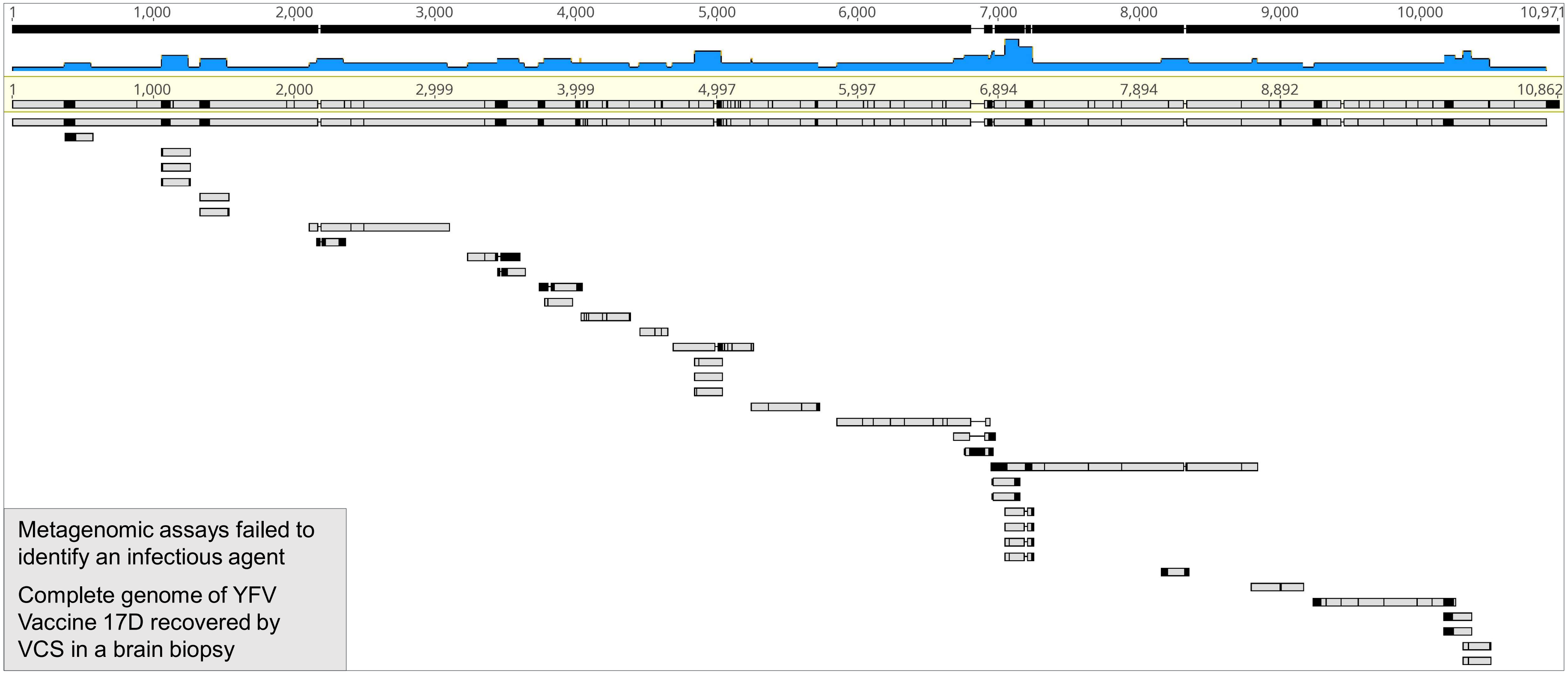
Sample Receipt to Pathogen Identification in <8 hours

Instrument	# of Samples	# of Reads	Read Length	Runtime
Illumina NextSeq 1000/2000	100	1.2 billion	150 nt	11h
Illumina NextSeq 500	40	400 million	150 nt	12h
Illumina MiSeq	3	25 million	300 nt	5h
Illumina MiniSeq	3	25 million	150 nt	4h
Illumina iSeq	1	4 million	150 nt	9.5h
Oxford Nanopore MinION	1	5 million	1000 nt	8h

Workflow	Unbiased	Capture
Extraction	1h	1h
Library preparation	6h	6h
Hybridization	n/a	1h
Sequencing on MiniSeq	4h	4h
Bioinformatics analysis	8h	2h
Total turn around time	19h	14h
Sensitivity	1x	100-1000x

Rapid Detection of YFV Vaccine Sequences in Brain of Patient With Rapid Onset Dementia Using VirCapSeq-VERT

43YO man, progressive weakness, cognitive decline, hypogammaglobulinemia

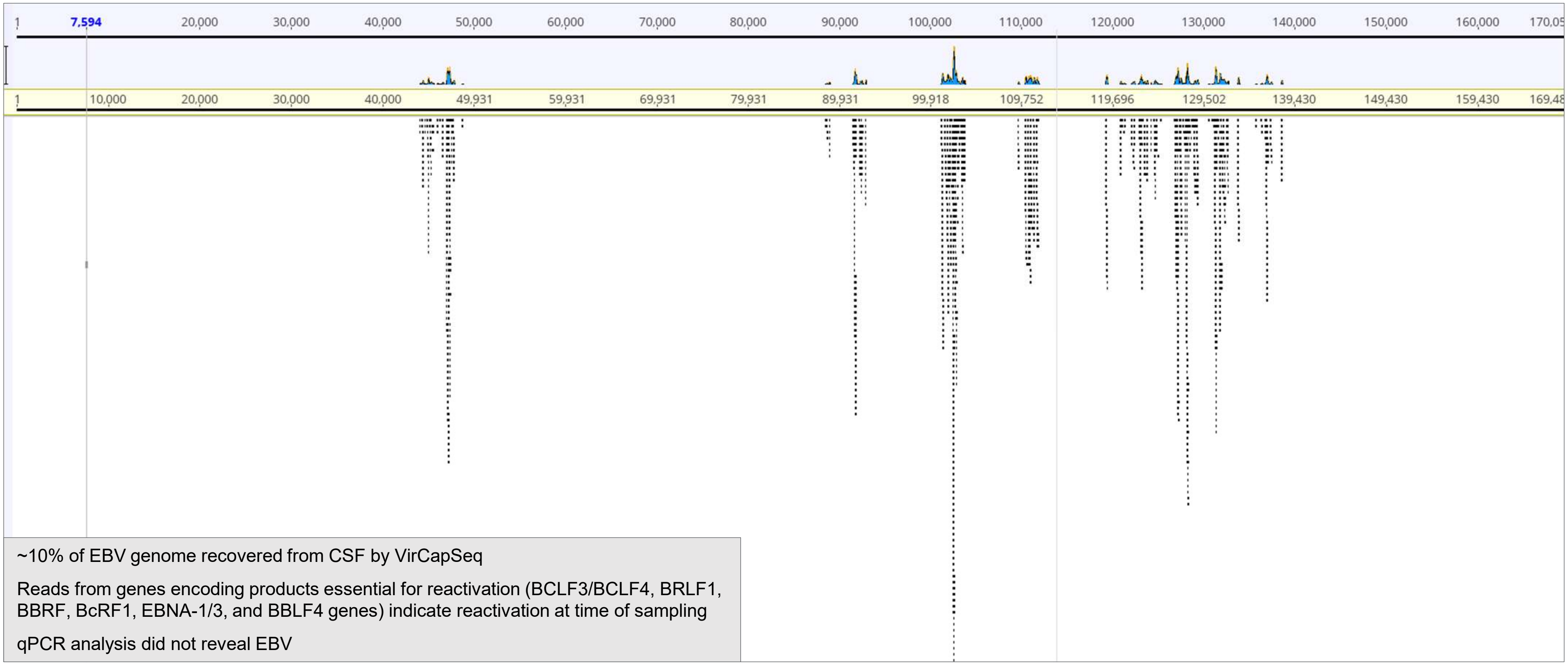


Metagenomic assays failed to identify an infectious agent

Complete genome of YFV Vaccine 17D recovered by VCS in a brain biopsy

Detection of EBV in CSF of a Patient with Headache and Cognitive Dysfunction Using VirCapSeq-VERT

All hospital based viral/bacterial/fungal cultures and PCR tests were negative

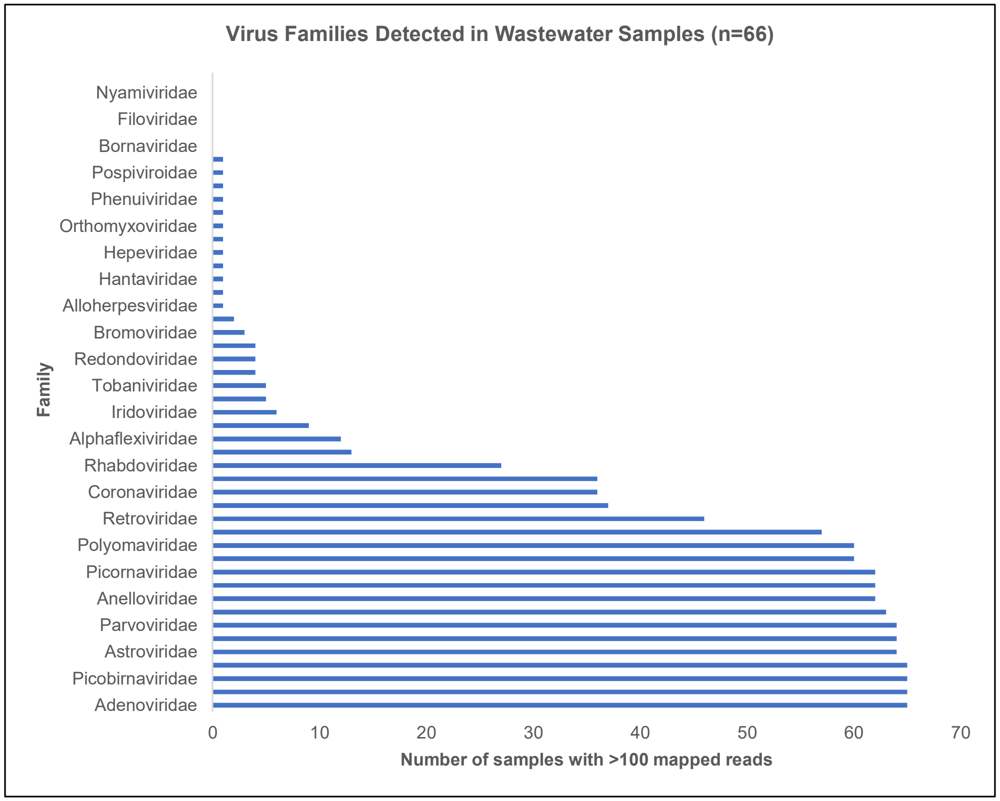


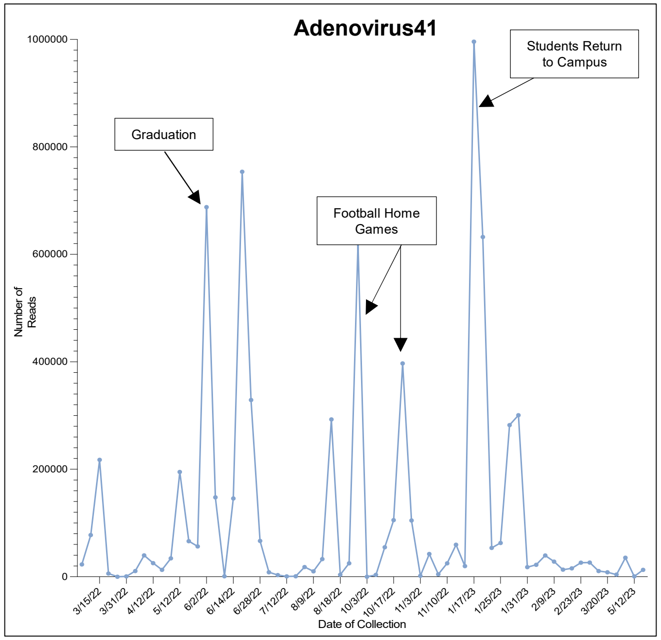
~10% of EBV genome recovered from CSF by VirCapSeq

Reads from genes encoding products essential for reactivation (BCLF3/BCLF4, BRLF1, BBRF, BcRF1, EBNA-1/3, and BBLF4 genes) indicate reactivation at time of sampling

qPCR analysis did not reveal EBV

Detection of Adenovirus in US Air Force Academy Wastewater Using VirCapSeq-VERT





- Wastewater collected Feb 2022-May 2023
- Major events on campus with visitors illustrate spikes in Adenovirus 41

Investigation of Acute Encephalitis Syndrome (AES) (India) Using VirCapSeq-VERT and BacCapSeq

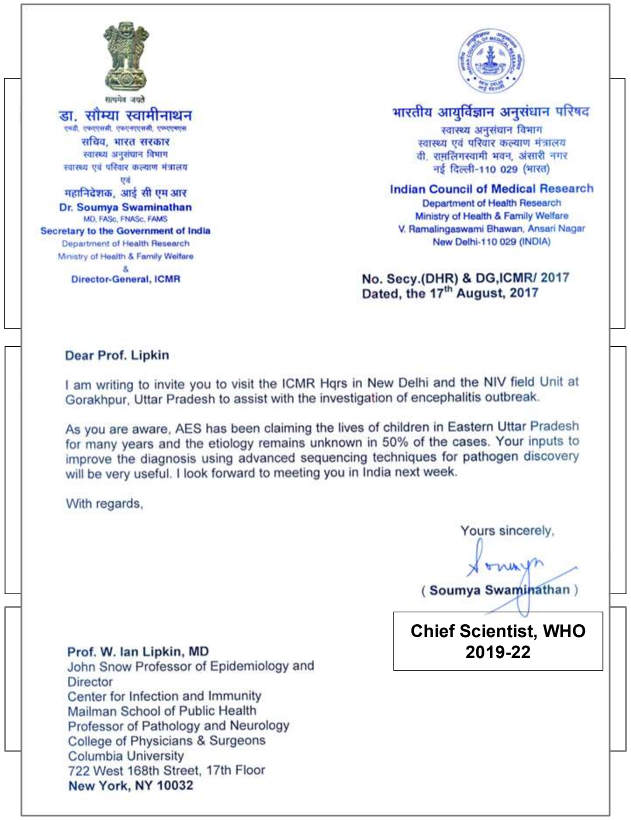

Uttar Pradesh


AES hospital
under construction


Musahar children


Waiting room


Bed shortage



सत्यमेव जयते
डा. सौम्या स्वामीनाथन
एमडी, एफएएससी, एफएनएएससी, एफएएमएस
सचिव, भारत सरकार
स्वास्थ्य अनुसंधान विभाग
स्वास्थ्य एवं परिवार कल्याण मंत्रालय
एवं
महानिदेशक, आई सी एम आर
Dr. Soumya Swaminathan
MD, FASc, FNASc, FAMS
Secretary to the Government of India
Department of Health Research
Ministry of Health & Family Welfare
&
Director-General, ICMR

भारतीय आयुर्विज्ञान अनुसंधान परिषद
स्वास्थ्य अनुसंधान विभाग
स्वास्थ्य एवं परिवार कल्याण मंत्रालय
वी. रामलिंगस्वामी भवन, अंसारी नगर
नई दिल्ली-110 029 (भारत)
Indian Council of Medical Research
Department of Health Research
Ministry of Health & Family Welfare
V. Ramalingaswami Bhawan, Ansari Nagar
New Delhi-110 029 (INDIA)

No. Secy.(DHR) & DG,ICMR/ 2017
Dated, the 17th August, 2017

Dear Prof. Lipkin

I am writing to invite you to visit the ICMR Hqrs in New Delhi and the NIV field Unit at Gorakhpur, Uttar Pradesh to assist with the investigation of encephalitis outbreak.

As you are aware, AES has been claiming the lives of children in Eastern Uttar Pradesh for many years and the etiology remains unknown in 50% of the cases. Your inputs to improve the diagnosis using advanced sequencing techniques for pathogen discovery will be very useful. I look forward to meeting you in India next week.

With regards,

Yours sincerely,

Soumya

(Soumya Swaminathan)

Chief Scientist, WHO
2019-22

Prof. W. Ian Lipkin, MD
John Snow Professor of Epidemiology and
Director
Center for Infection and Immunity
Mailman School of Public Health
Professor of Pathology and Neurology
College of Physicians & Surgeons
Columbia University
722 West 168th Street, 17th Floor
New York, NY 10032

Pathogen Discovery in Musahar Children with Encephalitis in Uttar Pradesh Using VirCapSeq-VERT and BacCapSeq

Agents detected in CSF by capture sequencing; >33% treatable with antivirals/antibiotics

	Total	**2015**	**2016**	**2017**
Total AES Samples	543	153	148	242
Fatal	222	75	77	67
Non-Fatal	321	77	61	175

535 samples (CSF = 532; brain autopsies = 3) negative in standard NGS

Known Pathogenic Agents	**n = Cases Positive**	**% Positivity**
Japanese Encephalitis Virus	22/543	4.0%
Enteroviruses	14/543	2.5%
Herpesviruses	21/543	3.8%
Orientia tsutsuganmushi	73/543	13.5%
Rickettsia sp.	86/543	15.8%
Co-infection	42/543	7.7%
Other agents*	126/543	23.2%
Total	**342/543**	**63.0%**

**Other viruses, or low coverage, or plant, or insect origin viruses*

Agnostic Serology: Peptide Microarrays

A granular agnostic approach for detecting footprints of infection

- 75 mm x 26 mm glass slides
- Non-redundant 12-mer peptides that tile the length of a proteome with 11-residue overlap

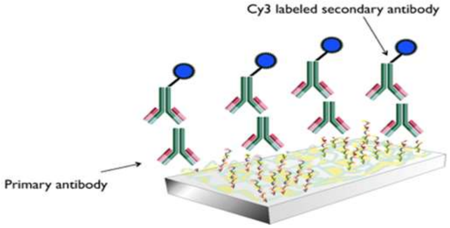
MAAKVADKASVGIAKGIKE **"Targeted Protein Sequence"**

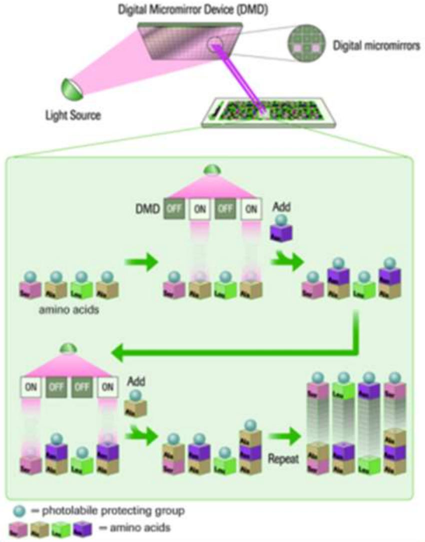
MAAKVADKASVK
AAKVADKASVKG
AKVADKASVKGIA
KVADKASVKGIAK
VADKASVKGIAKG
ADKASVKGIAKGI
DKASVKGIAKGIK
KASVKGIAKGIKE

Continuous overlapping peptides printed in array design

Available designs and peptide density/subarray

	1-plex	3-plex	12-plex
Features in low density array	3M	1M subarray	172K subarray
Features in high density array	6.1M	2.1M subarray	380K subarray





Zika

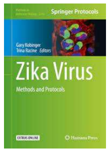
Discovery of NS2B ZIKV peptide that enabled development of an ELISA

Diagnosis of Zika Virus Infection by Peptide Array and Enzyme-Linked Immunosorbent Assay

Mishra N, Caciula A, Price A,…Lipkin WI

Research Article | Published 06 September 2018 | doi: 10.1128/mbio.00095-18

A high-density microarray comprising nonredundant 12-mer peptides that tile, with one-residue overlap, the proteomes of Zika, dengue, yellow fever, West Nile, Ilheus, Oropouche, and chikungunya viruses. Serological analysis enabled discovery of a ZIKV NS2B 20-residue peptide that had high sensitivity (96.0%) and specificity (95.9%) versus natural infection.



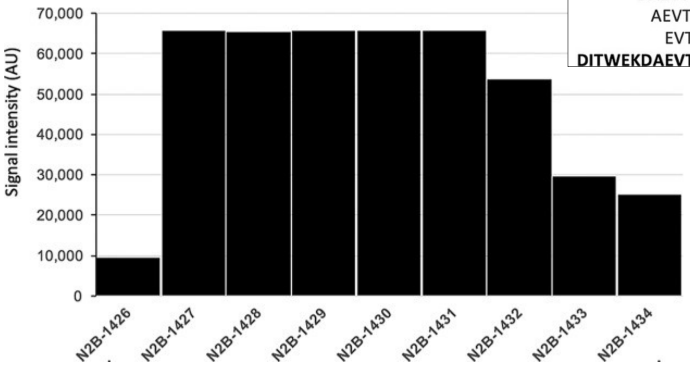
Zika Virus Peptide ELISA (ZIKV-NS2B-Concat ELISA) for Detection of IgG Antibodies to Zika Virus Infection

Mishra N, Thakkar R, Ng J, Lipkin WI

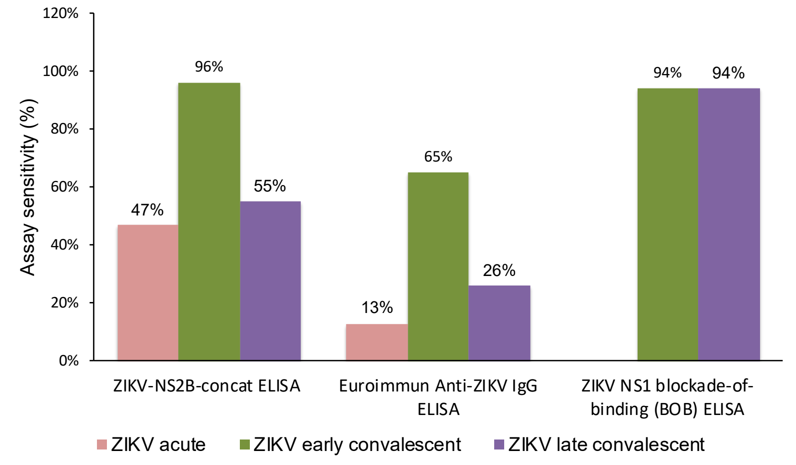
Protocol | Published 05 May 2020 | doi: 10.1007/978-1-0716-0581-3_10

An affordable ZIKV NS2B biotinylated peptide ELISA was built and compared with peptide array:
~47% sensitivity in ZIKV acute patients (2-3 weeks post-infection)
~96% sensitivity in ZIKV early convalescent (1-6 months post-infection)
~55% sensitivity in late convalescent (>6 months post-infection)

DITWEKDAEVTG	NS2b-1426
ITWEKDAEVTGN	NS2b-1427
TWEKDAEVTGNS	NS2b-1428
WEKDAEVTGNSP	NS2b-1429
EKDAEVTGNSPR	NS2b-1430
KDAEVTGNSPRL	NS2b-1431
DAEVTGNSPRLD	NS2b-1432
AEVTGNSPRLDV	NS2b-1433
EVTGNSPRLDVA	NS2b-1434
DITWEKDAEVTGNSPRLDVA	**NS2b-1426-1434**



Transfer reactive peptides to simple peptide ELISA



Tick-Borne Diseases

SCIENTIFIC REPORTS

A multiplex serologic platform for diagnosis of tick-borne diseases

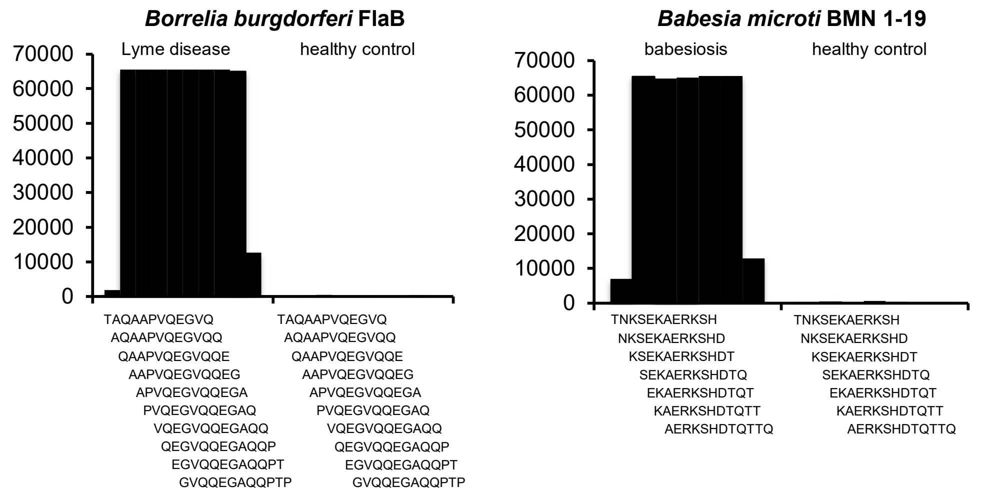
Tokarz R, Mishra N, Tagliaferro T, … Lipkin WI

Published 16 February 2018 | doi: 10.1038/s41598-018-21349-2

Agent and antigens	Number of 12-mer peptides
Anaplasma phagocytophilum MSP2, MSP4, MSP5, P55, P62, Omp1N	16,787
Babesia microti SA-1, BMN1 (-2, -3, -4-5-6, -7, -8, -10, -11, -12-13, -17, -20, -21), GPI 12, AMA1; Hsp70, DnaK, Bmp32	11,333
Borrelia burgdorferi OspA, OspB, OspC, OspD, VlsE, DbpA and B, BmpA-BmpD, P100, OppA, OppA2, RevA, P66, LA7, BBK07, BBK32, BBK50, FlaA, FlaB, FliL, FlgE, DnaK, BBA04, BBA36, BBA57, BBA64, BBA65, BBA66, BBA68, BBA69, BBA73, BBA74, BBI38, BBI42, BBE31, OspE, OspF, Erp (all paralogs), Mlp (all paralogs), Bdr (all paralogs), BBO03 (all paralogs)	91,338
Borrelia miyamotoi GlpQ, FhbA, ipA, P66, OppA2, FlgG, FlaB, FliL, VLP (1, A1, A2, C1, C2, C3, D1, D2, D3, D4, D5, D5S, D6S, D6, D7S, D8, D9, D10, 3 S, A2S, 4 S 15/16, 18), VSP (1,2, 3, 4, 6)	23,946
Ehrlichia chaffeensis P156, P120, P28/omp-1, Gp47, VLPT, SP-related protein	4,156
Rickettsia rickettsii OmpA, OmpB, OmpW, Porin 4, adr1, adr2	5,855
Heartland virus N, Gn, Gc, L	4,153
Powassan virus polyprotein	7,688
Long Island tick rhabdovirus N, P, M, G, L	3,949

Table 1. Composition of the TBD-Serochip.

Epitope Identification


***Borrelia burgdorferi* FlaB**

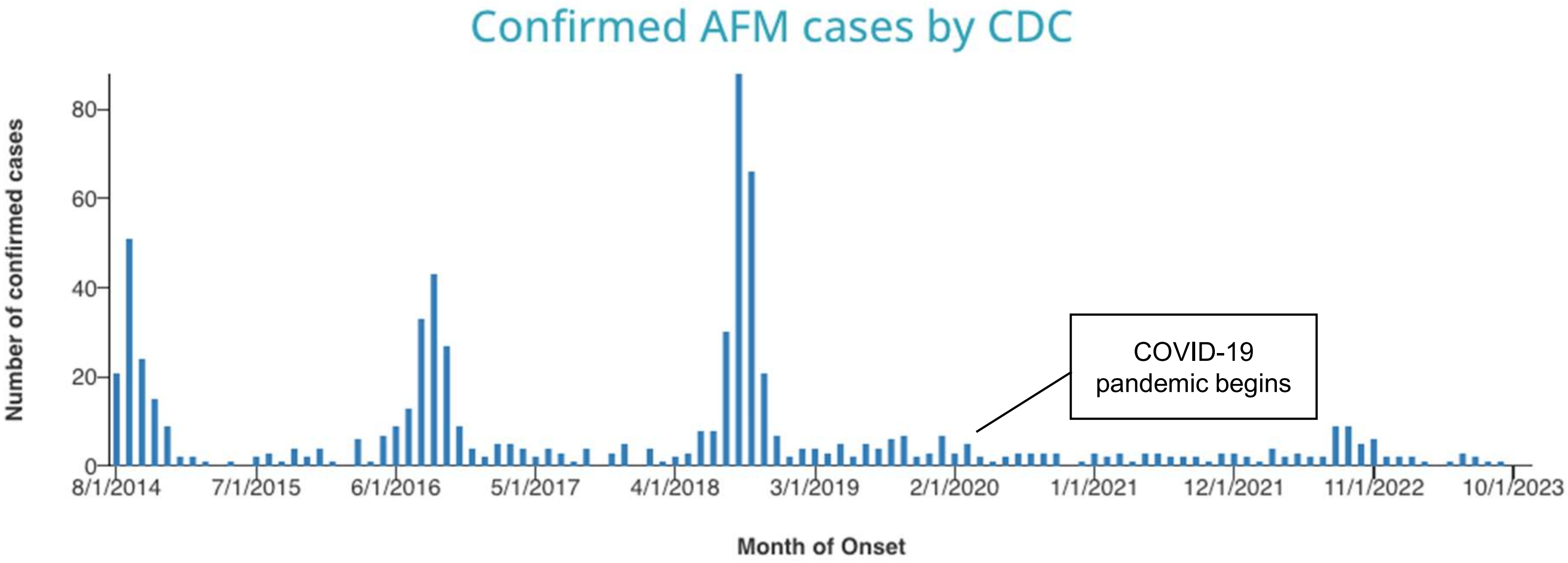
***Babesia microti* BMN 1-19**

Tick-borne disease	Tick-borne agent	Number of discriminatory peptides identified
Lyme disease	*Borrelia burgdorferi*	2
Borrelia miyamotoi disease	*Borrelia miyamotoi*	5
Human granulocytic anaplasmosis	*Anaplasma phagocytophilum*	2
Human monocytic ehrlichiosis	*Ehrlichia chaffeensis*	5
Babesiosis	*Babesia microti*	3

Minimum number of specific and discriminatory peptides required for accurate detection of selected tick-borne agents.

Acute Flaccid Myelitis (AFM)

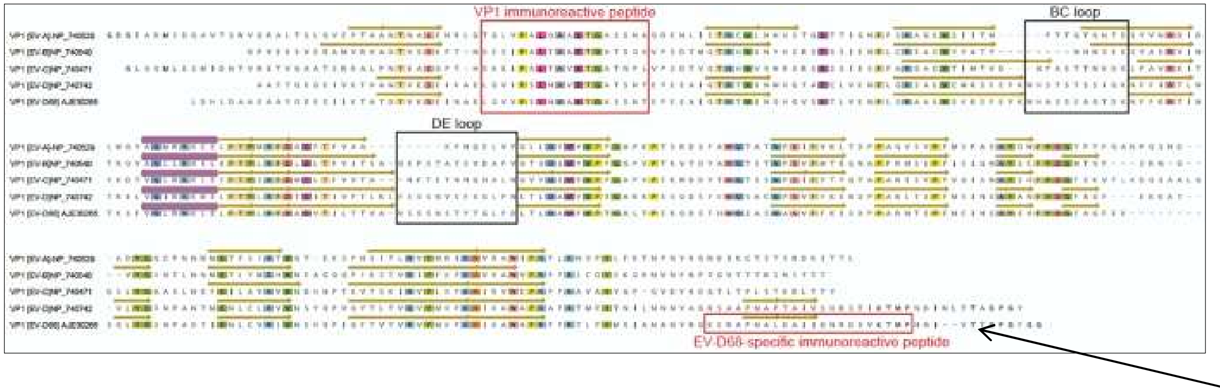
An example of pathogen discovery through serology

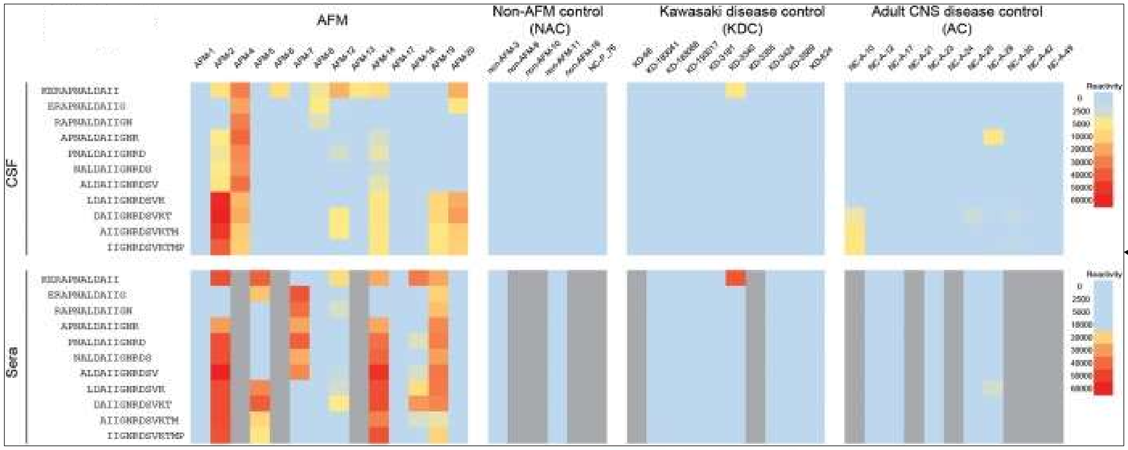


AFM case definition:

- An illness with onset of acute flaccid limb weakness
- MRI showing spinal cord lesion largely restricted to gray matter and spanning one or more vertebral segments
- 99% have a history of febrile illness within 4 weeks of onset of neurological disease (respiratory 75%, GI tract 38%)

Enterovirus D68 Antibodies in CSF of Patients with AFM





mBio **Antibodies to Enteroviruses in Cerebrospinal Fluid of Patients with Acute Flaccid Myelitis**

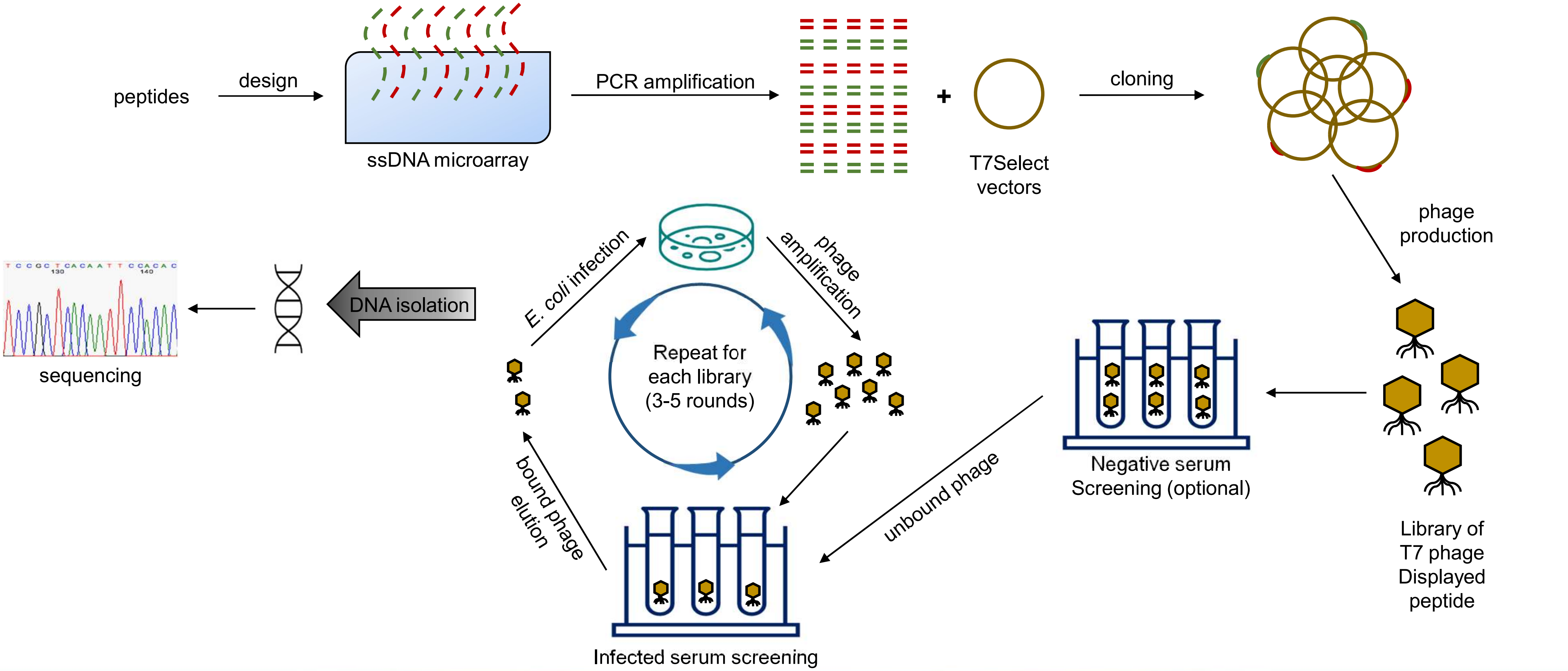
Nischay Mishra, Terry Fei Fan Ng, Rachel L. Marine, Komal Jain, James Ng, Riddhi Thakkar, Adrian Caciula, Adam Price, Joel A. Garcia, Jane C. Burns, Kiran T. Thakur, Kimbell L. Hetzler, Janell A. Routh, Jennifer L. Konopka-Anstadt, W. Allan Nix, Rafal Tokarz, Thomas Briese, M. Steven Oberste, W. Ian Lipkin

Research Article | Published 13 August 2019 | doi: 10.1128/mbio.01903-19

Immunoreactivity against an EV-D68-specific 22-aa VP1 capsid peptide in patients with AFM, non-AFM controls (NAC), Kawasaki disease controls (KDC), and adult CNS disease controls (AC).

Identification of an immunoreactive peptide sequence region in VP1 protein of reference sequence entries for EV-A, EV-B, EV-C, and EV-D

Multiplex Serology Using Phage Display



peptides

design

ssDNA microarray

PCR amplification

+

T7Select vectors

cloning

phage production

Library of T7 phage Displayed peptide

Negative serum Screening (optional)

unbound phage

Infected serum screening

bound phage elution

E. coli infection

phage amplification

Repeat for each library (3-5 rounds)

DNA isolation

sequencing

Concordance Between Peptide Array and Granular Phage Display

Opportunities to discriminate between infections with related viruses and find evidence of reactivation

🟢 Peptide Array
🟠 Phage Display

Reaction of Herpesvirus in ME/CFS Patients

Virus	Protein	CASE1	CASE2	CASE3	CASE4	CASE5	CASE6	CASE7	CASE8
HHV4	BZLF1	🟢			🟢				
HHV4	C M protease					🟢		🟢	🟢 🟠
HSV1	E- Protein	🟢	🟢 🟠	🟢 🟠	🟢 🟠			🟢 🟠	
HHV-6A	Helicase	🟢	🟢 🟠		🟢 🟠		🟢 🟠		
HHV3	Large Tegument Protein	🟢 🟠	🟢 🟠	🟢 🟠	🟠	🟢 🟠	🟢 🟠	🟠	🟠
HHV3	UL32			🟢	🟢		🟢		🟠

Convalescent COVID-19 Patients

Virus	Protein	Sample1	Sample2	Sample3	Sample4	Sample5	Sample6
Coronavirus	NP	🟠	🟢 🟠	🟢 🟠	🟢 🟠	🟢	
		🟢 🟠	🟢 🟠	🟢 🟠	🟢 🟠	🟢 🟠	🟢
	GP	🟢	🟢 🟠	🟢	🟢	🟢	🟢
				🟢 🟠	🟢 🟠	🟢	🟢 🟠
		🟢 🟠	🟢 🟠	🟢 🟠	🟢 🟠	🟢 🟠	🟢
		🟠	🟢 🟠	🟠	🟠		🟠
		🟠	🟢 🟠	🟠	🟠	🟠	
			🟢			🟢	🟢

GAPP

Mission in support of the IHR 2005

- Build capacity for infectious disease surveillance in low- and middle-income countries
- Catalyze regional and global collaboration
- Foster a culture of trust and respect by through recognition of contributions

Virtual (2 weeks)

Education

Ethics and biosafety
Experimental design
Collection, processing, and storage
Data management
Bioinformatics and data analysis

In-person NYC (3 weeks)

Train

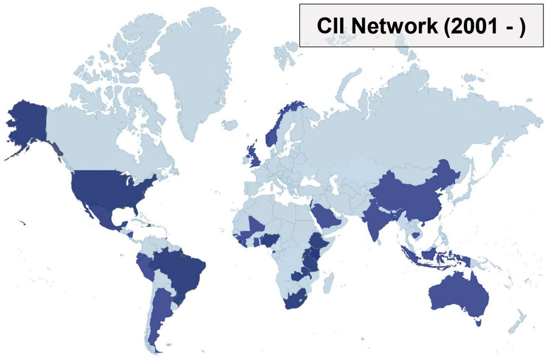
Sequencing
Serology
Data analysis

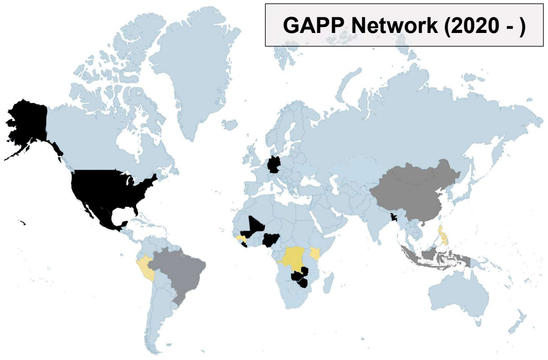
In-country

Facilitate

Needs assessments
Salary support/retention
Supplies
Bioinformatics

GAPP: Capacity Building





Black: active; grey: inactive; yellow: anticipated

Training Cohorts:

2021: Liberia

2022: Mali, Zambia

2023: Bangladesh, Germany, Mali, Mexico, Nigeria, US Air Force, Zambia

2024 (anticipated): Democratic Republic of the Congo, Guinea, Kenya, Peru, the Philippines, Taiwan, Zimbabwe

Sequencing and Serology

- Supplies (volume purchasing)
- Computational tools and databases

Mexico and Germany Cohorts at CII, *February 2023*

In-country training at CHAZ (Zambia), *August 2023*