SAND2011-0777C

Preparation of Nucleic Acid Libraries for Next Generation Sequencing with a Digital Microfluidic Hub for Unknown Pathogen Detection

Kamlesh D. Patel, PhD

Sandia National Laboratories

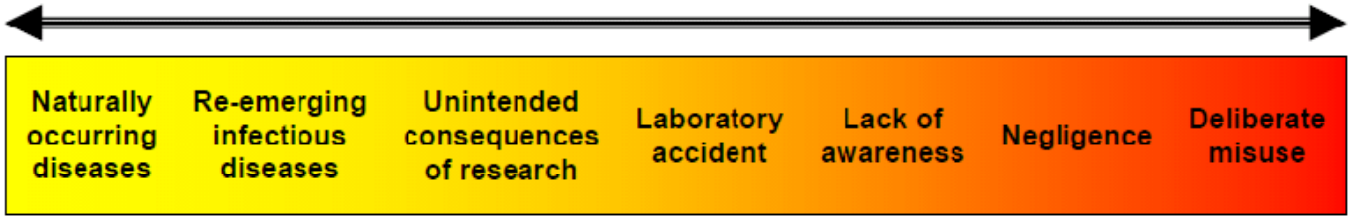
Livermore California

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Overview of my presentation

- **The emergence of unknown pathogens**
- **RapTOR (Rapid Threat Organism Recognition) Grand Challenge**
  - Approach to detecting unknown pathogens from clinical samples using next-gen sequencing
- **Automated Molecular Biology (AMB) Platform**
  - Preparing Nucleic Acid libraries (from complex mixtures)
    1. Digital Microfluidic Hub architecture
    2. Hydroxyapatite chromatography normalization
    3. Transposase-mediated NA prep and barcoding
- **Conclusions**

The risk of an infectious outbreak from an **unknown** agent could arises from a complex array of sources



Wilton Park, September 2009

Our mission is to provide solutions to the most challenging biological analysis problems that threaten our national security.

Sandia National Laboratories

Novel pathogens increasingly threaten national security & public health

- **Factors promoting pathogen emergence:**
  - Human encroachment on wildlife habitat
  - Increased population density, international travel & trade

- **Factors enabling pathogen engineering:**
  - Greater knowledge of pathogenicity & human biology
  - More, better, & cheaper tools for modification, synthesis, & evaluation of biological agents, including pathogens
  - Global dispersion of biological materials, knowledge, technology, & expertise

- **Infectious disease outbreaks & risk of bioterrorism are on the rise.**
  - Nature 451:990 '08
  - "Globalization, Biosecurity, & the Future of the Life Sciences", NRC '06

Do-it-yourself biology on rise

New breed of scientists using technology to experiment outside usual lab settings

By Julian Guthrie

In a kitchen in Saratoga, an electrical engineer is working with pure strain of E. coli purchased over the Internet in hopes of creating a handheld diagnostic tool to detect dangerous bacteria.

Out of a garage in Sacramento, a biocomputer is designing low-cost equipment to allow people to see and construct DNA.

From a studio in San Francisco, an artist is building houses from a medicinal fungus.

Across the Bay Area, and in other high-tech hotbeds, a revolution is under way. Citizen scientists — or biohackers, as they're being called — are taking biology out of academia and closed-door laboratories and bringing it into garages and

San Francisco Chronicle, Dec 20 2009

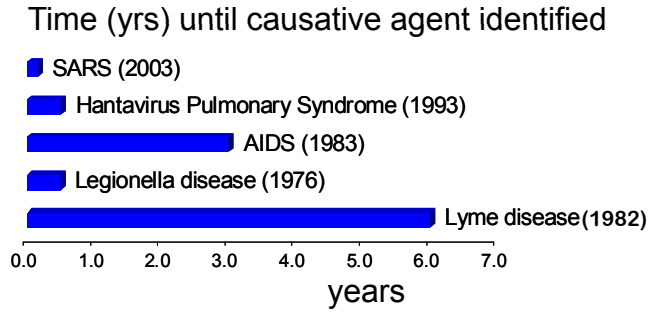
New York Times Feb 10, 2010

Sandia National Laboratories

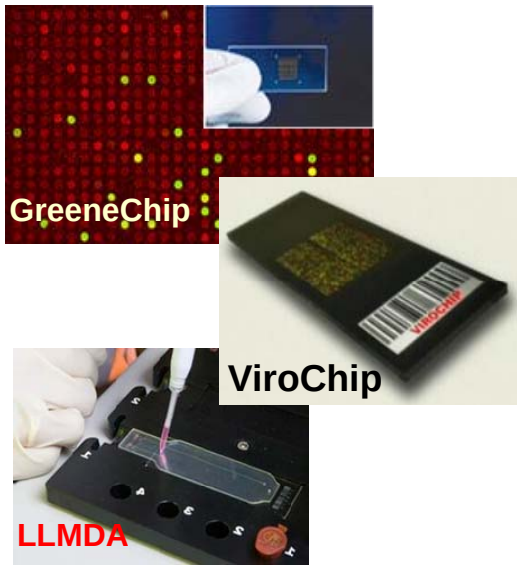
We need new tools for rapid identification & characterization of novel pathogens

- Outbreak dynamics are often measured in ***days to weeks***.

- Identification of a novel causative agent by conventional methods can take ***months to years***.



- Modern probe-based methods are fast, but are often confounded by novel pathogens.
  - Probes don't recognize deeply diverged targets
    - Target-poor pathogens may escape detection altogether
  - No probes for unanticipated features
    - Novel virulence factors, unusual insertions (*e.g.*, IL-4)
  - Unusual profiles can be difficult to interpret
    - Reflect pathogen's mix-and-match features?
    - Result from cross-hybridization?



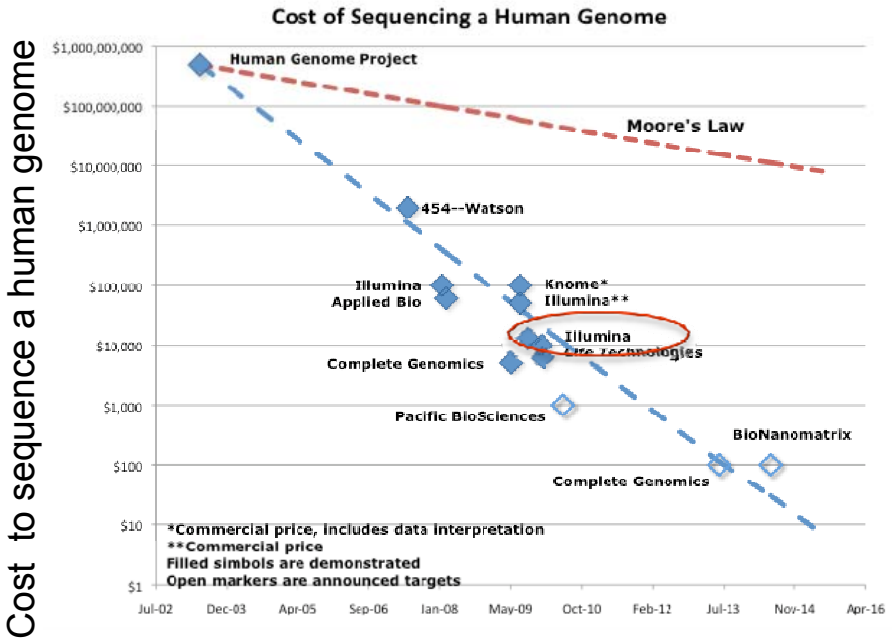
Sandia National Laboratories

Next Generation Sequencing is a transformational technology for pathogen characterization

- 30x coverage of two human genomes in a single run under $10,000 per sample
- 150-200 Gb for 2 x 100 bp run (8-days)
- Equivalent to sequencing 1.2 to 1.6 NCBI Genbanks per run

http://www.illumina.com/systems/hiseq_2000.ilmn
ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt

http://www.illumina.com/systems/hiseq_2000.ilmn



http://www.flickr.com/photos/doe_jgi/3876606040

…but DNA sample prep is primarily a benchtop process

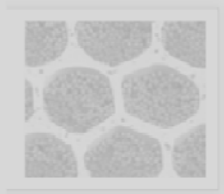
Sandia National Laboratories

Brute-force NGS of clinical samples can enable discovery of novel pathogens

NGS is currently a late step in pathogen identification pipelines

- Example from Ian Lipkin (Columbia), 11-09:

Deplete non-informative NA to improve efficiency of NGS analysis



Shotgun NGS

- $5000 *per* assay
- 1 week

Disease	Sample	Novel Agent Detected	Total Reads	Hits on Agent	Reference
Merkel cell carcinoma	tumors	"Merkel cell polyomavirus"	395,734	2 (**0.00005%**)	Science 319:1096 '08
organ transplant related fatality	serum & organs	"Dandenong" arenavirus	103,632	14 (**0.014%**)	N Engl J Med 358:991 '08
pediatric gastroenteritis	feces	"human klassevirus "	937,935	849 (**0.09%**)	Virol J 6:82 '09
pediatric influenza-like illness	nasopharyngeal swabs	"human enterovirus type 109"	20,825,810	119 (**0.0006%**)	J Virol 84:9047 '10

The Challenge: Develop a new approach to rapidly characterize unknown bioagents

Rapid Threat Organism Recognition (RapTOR) system

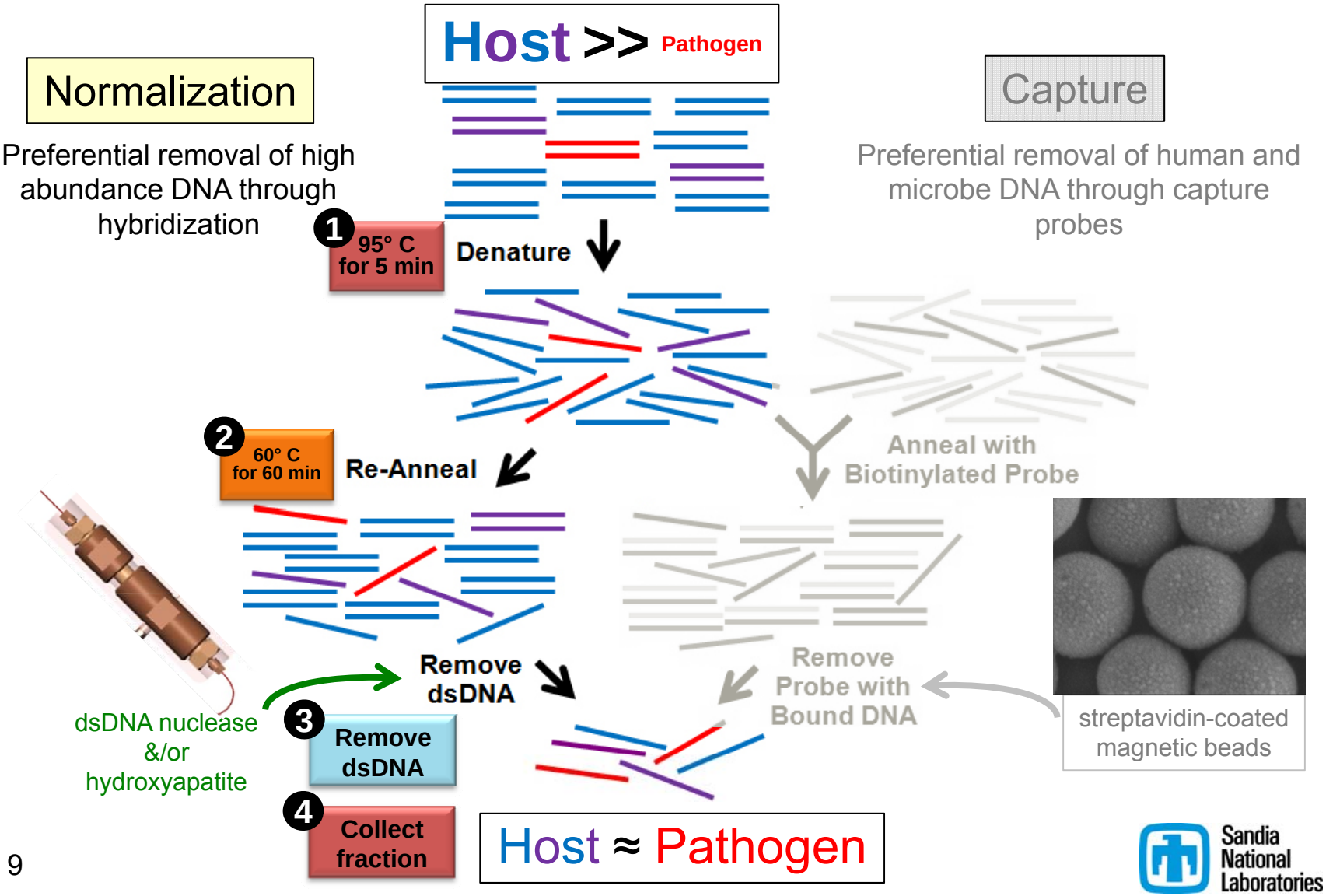
- **Goal:** Efficient analysis of pathogen nucleic acids (NA) in clinical samples *via* targeted Next Generation Sequencing (NGS)
- **Key advance:** Automated microfluidic hardware to enable molecular suppression and preparation to improve signal-to-background (pathogen-to-host) NA ratio in samples

Drivers for our approach

- Identify and characterize unknown pathogens in a timeframe compatible with rapid disease outbreak detection and response
- No *a priori* knowledge of a pathogen or culturing of organism; (deep genomic sequencing)
- Automate the nucleic acid processing and integrate data analysis and knowledge discovery system inclusion at federal and state-wide laboratories

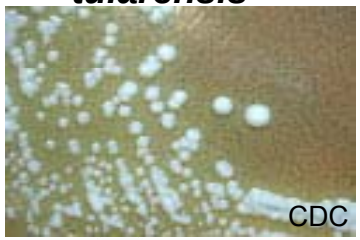
Sandia National Laboratories

We are focusing on complementary suppression methods for depletion of host NA

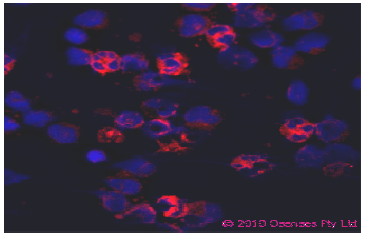


HAC normalization strategy to enhance pathogen detection with next gen sequencing

Francisella tularensis



Human PBMCs



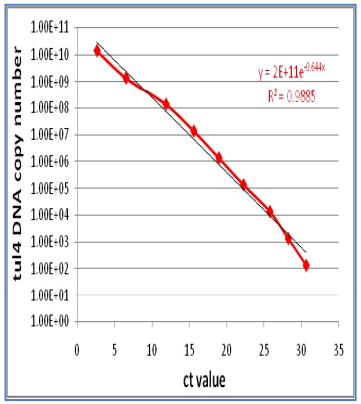
Combined total RNA at ratios of 1:100 and 1:10000 In duplicate

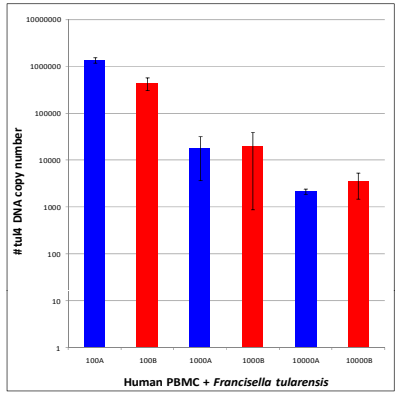
SMARTER® cDNA synthesis

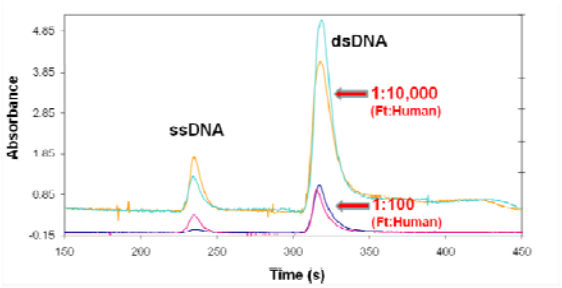
Standalone HAC Normalization module

Benchtop Library Prep for Illumina NGS

Transcript Analysis with Databases





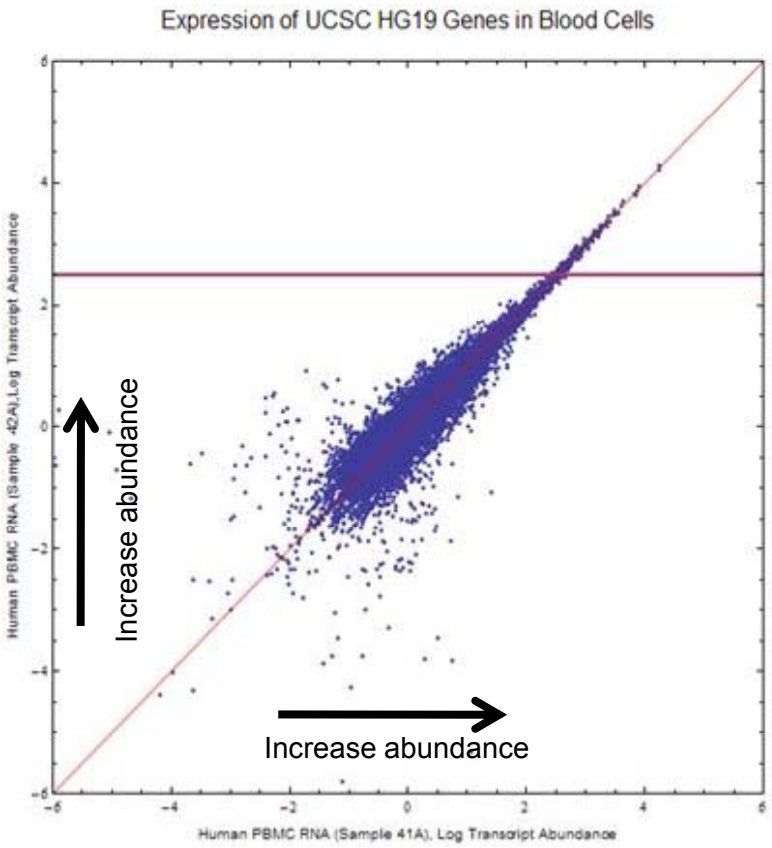


7 M reads/sample 51 bp SE read

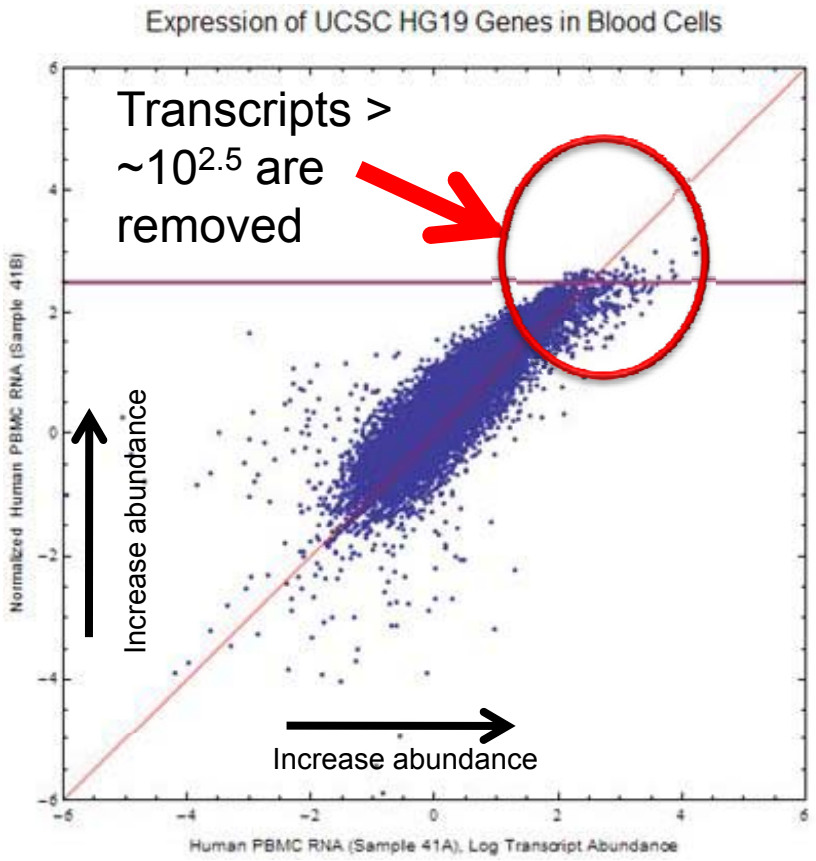
Geneious Bioinformatics platform and Sandia-developed pipeline

Sandia National Laboratories

HAC normalization removes abundant human RNA transcripts



Control vs. Control Transcripts



HAC Normalized vs. Control

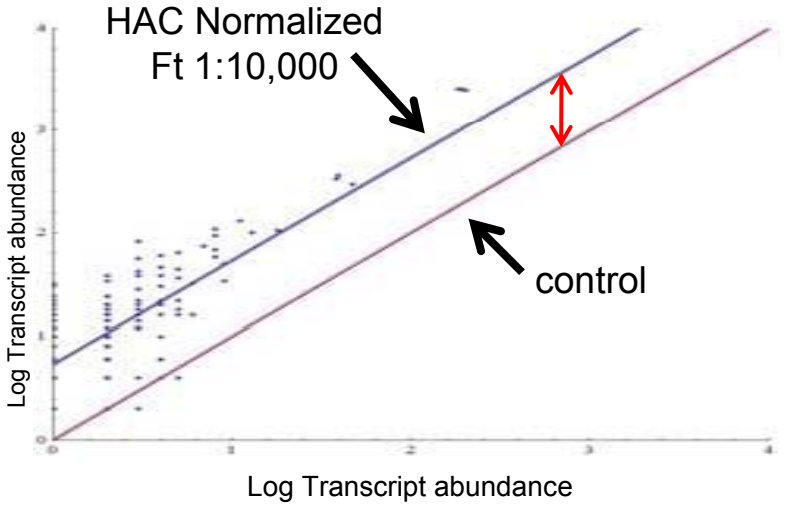
Sandia National Laboratories

HAC normalization enriches bacterial pathogen transcripts

- **Total** percent of sequence hits mapped to the *Francisella tularensis* genome were increased
  - 2.7-fold (Ft 1:100)
  - 7.3-fold (Ft 1:10,000)

- Enhanced the **detection** of Ft transcripts for 1:10,000 ratio
  - Sample 1: 5-fold
  - Sample 2: 3-fold

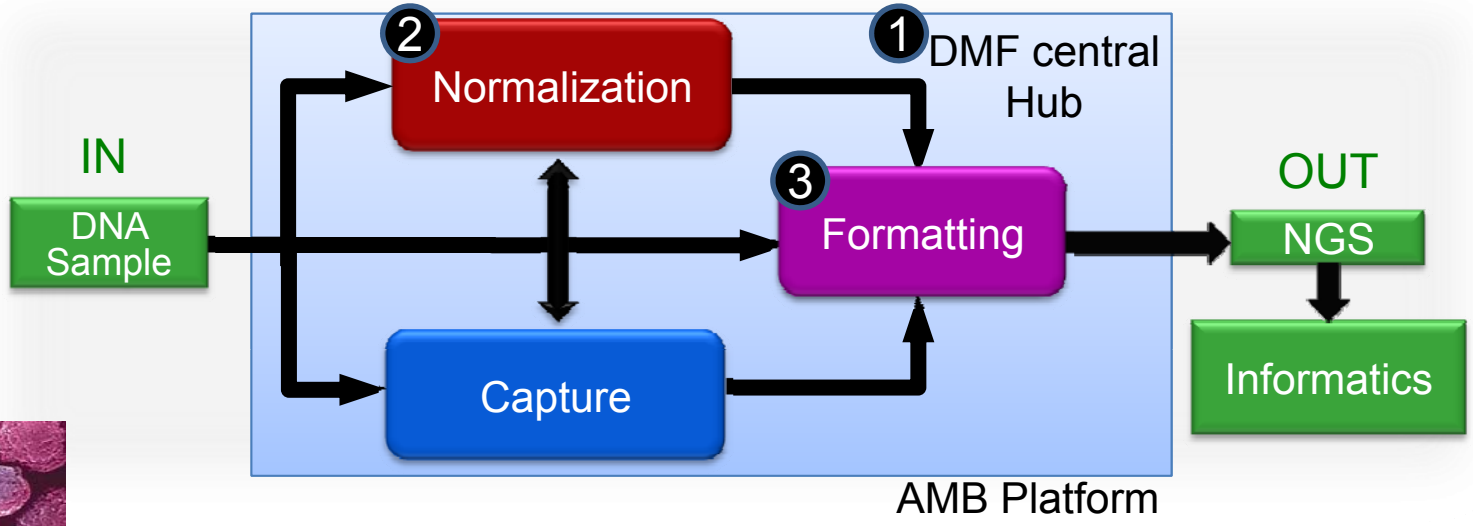
- More detail analysis is underway
  - 65 bp paired-end run barcoded samples
  - Triplicates w/ internal controls
  - Multiple rounds of normalization
  - Direct comparison to enzymatic methods

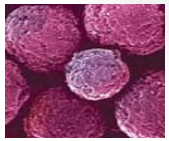


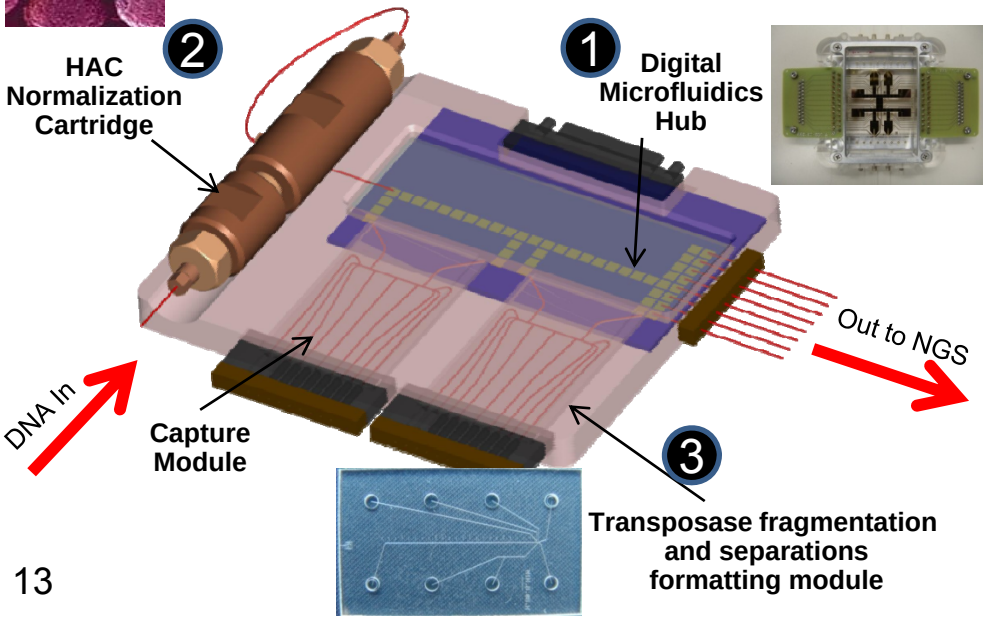
Misera, Langevin, Schoeniger 2010

Sandia National Laboratories

AMB platform integrates suppression & formatting modules within a flexible architecture





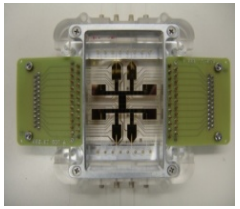


Gen 1 requirements:
- Normalization + Formatting
- Handle ng quantities of gDNA and cDNA
- 10-100 fold suppression
- Semi-automated

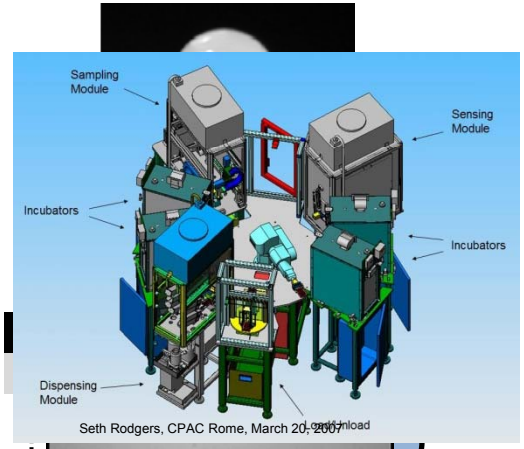
Gen 2 requirements:
- Orthogonal normalization (Norm + capture)
- ≥1000 fold suppression
- Faster than benchscale processing
- Fully automated

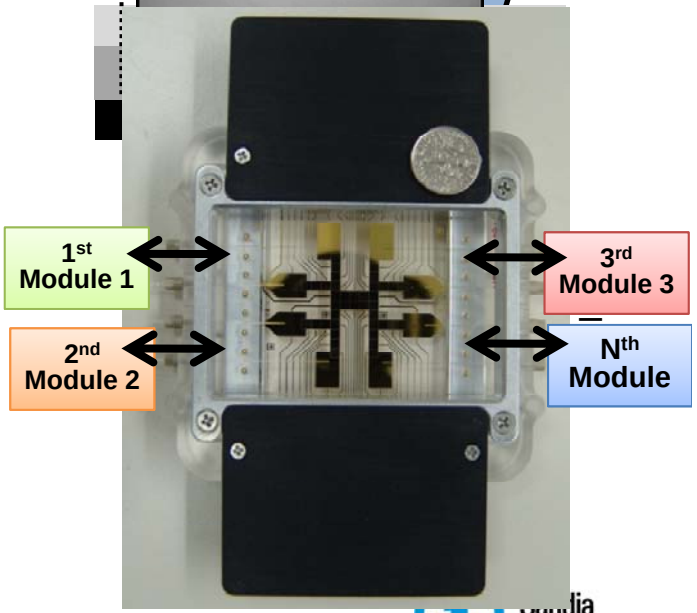
Sandia National Laboratories

Core architecture of the AMB platform for NA processing is the Digital Microfluidic Hub



- Droplets are ideal micro-cargo containers
  - Operated “digital” fashion (virtual tubes or microreactors)
  - Nanoliter to microliter in volume
  - Merge, mix, split (virtual pipetting)
- Based on principles of electrowetting-on-dielectric (EWOD) and dielectrophoresis
  - voltage is applied to electrode pads in an addressable 2-D array on glass substrates.
- Pollack and Fair at Duke University
  - Oil/water system –PCR in DMF (Hua et al. *Anal Chem* 2010)
- Aaron Wheeler’s group at U. of Toronto
  - Air/water system—cell-based microculturing (Bluovak et al, Lab Chip 2009)
- DMF as a central hub for interfacing multiple lab-on-a-chip sample processing modules through droplets
  - Goal and objective is to couple
    - Flexibility and spatial manipulations of droplets
    - Modularity and temporal resolution of continuous-flow microchannel devices
    - Overcome world-to-chip interface difficulties
      - Sample volume mismatch & timing

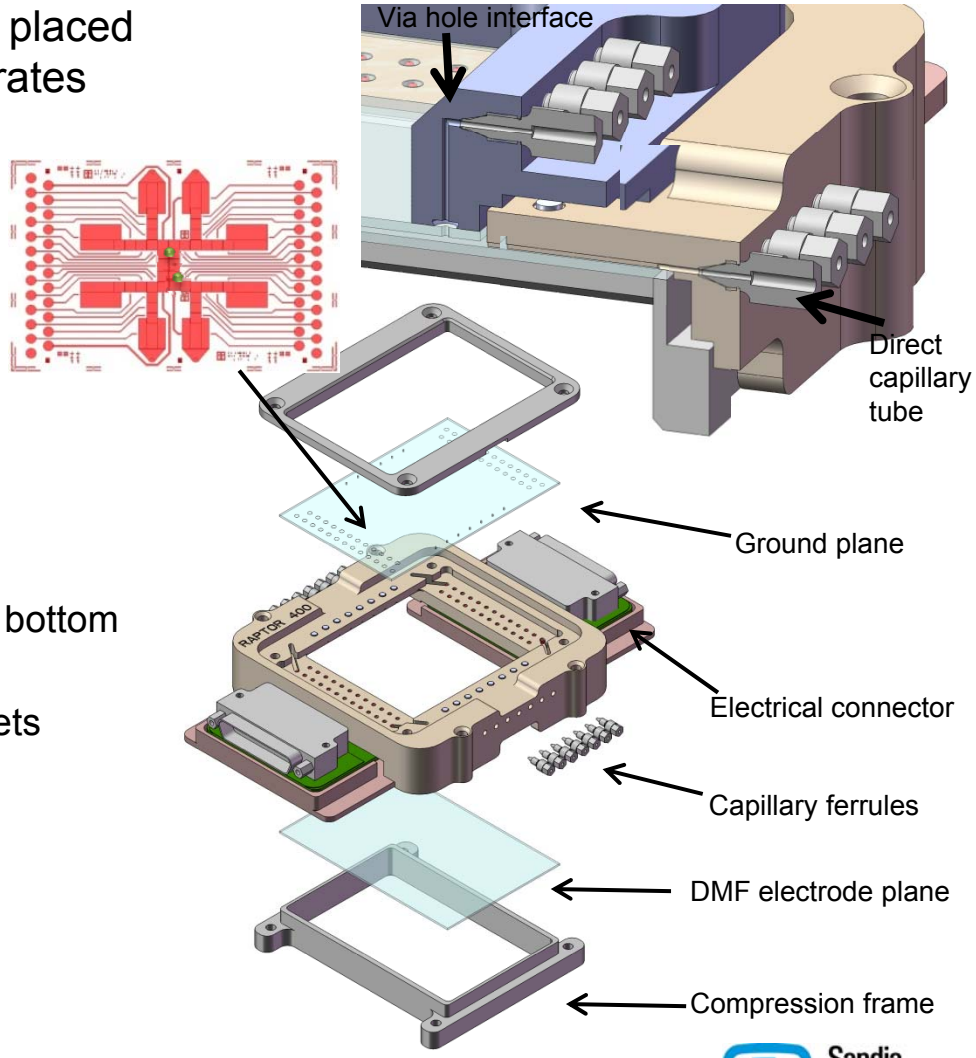




DMF hub platform uses capillary tubes for reliable droplet to module interface

- Teflon-coated, 150 um o.d capillaries placed in gap between the DMF glass substrates
- Top-side through via holes

- **Custom-fabricated manifold**
  - plug-and-play installation of the DMF substrates
  - electrical contact for 46 actuation pads
  - Up to 14 capillary interface ports
- **DMF substrates**
  - ITO-coated glass for complete top and bottom optical access
  - 185-um gap size microliter-sized droplets using ~2x2 mm pads
- **Programmable interface**
  - Manual keyboard keys
  - Script routines

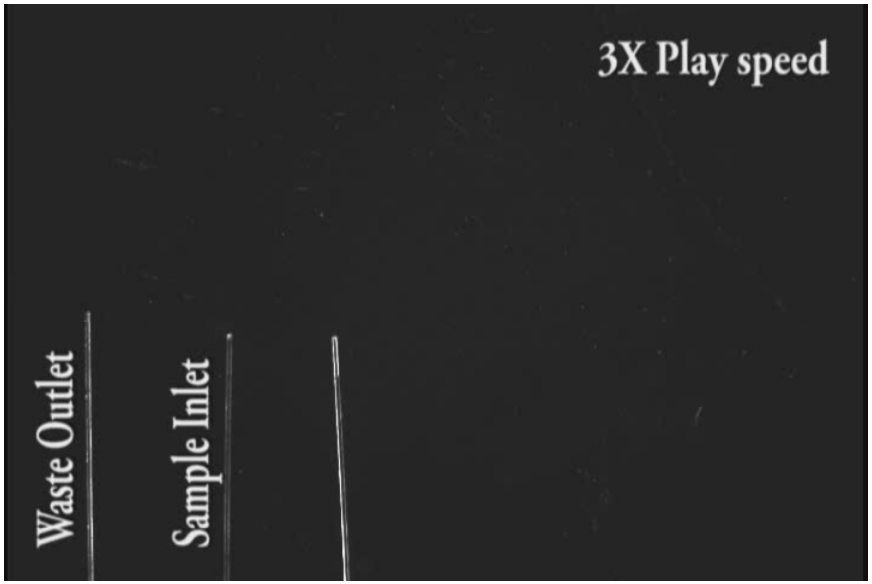


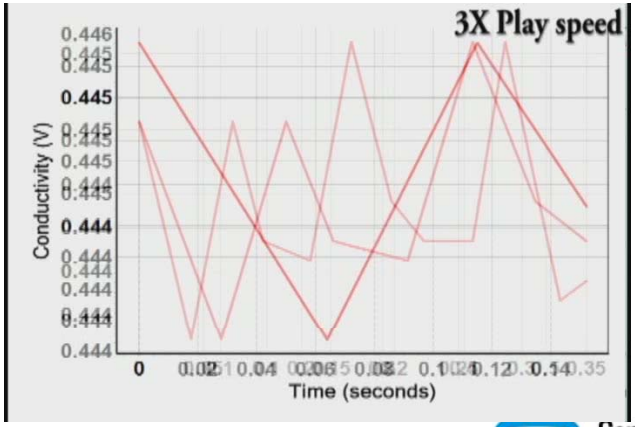
Bartsch & Renzi, 2010

Sandia National Laboratories

Example of a capillary to DMF interface for discritizing eluting peaks

- Teflon-coated, 150 um o.d capillaries placed in gap between the DMF glass substrates
- Dark-field with epifluorescence made possible by transparent DMF substrates
- Contactless conductivity sensor sheathed around the capillary to detect conductivity changes to enable switching and fraction collection
- 4 µL Fluorescein in 130 mM Na_3PO_4
- Flow rate = XX µL/hour



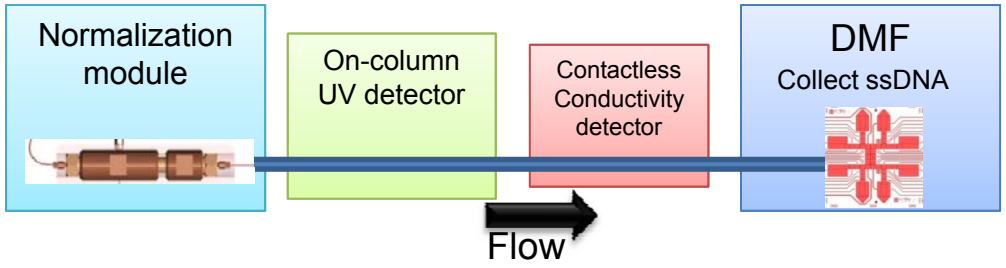


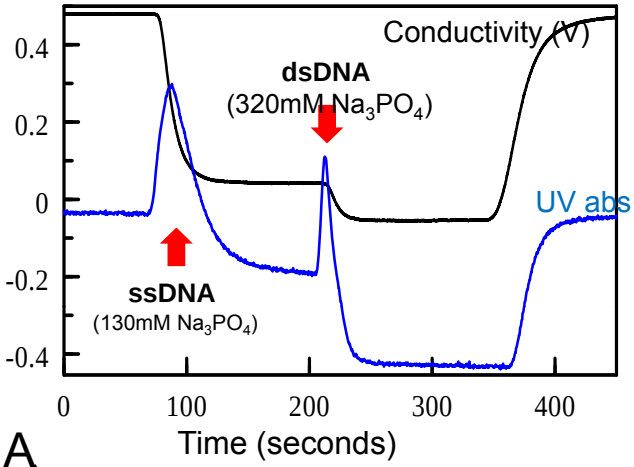
Kim & Patel, 2010

Sandia National Laboratories

Demonstration of HAC normalization module integration w/ DMF using a capillary interface

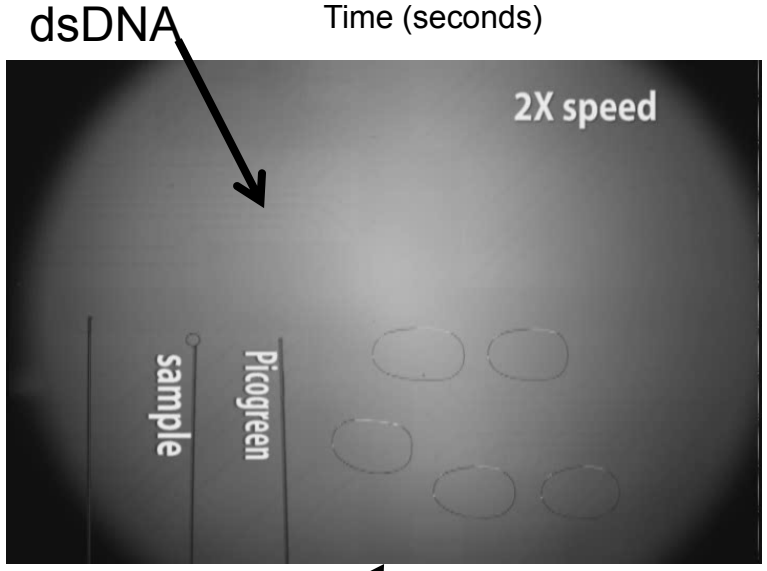
- **Normalization**
  - Removal of high abundance DNA through selective hybridization
  - Denature, anneal, separate, collect





- ssDNA labeled with cy3 (100 bp)= **low abundance**
  - 160 mM Na_3PO_4 buffer
- dsDNA: DNA ladder >100 bp = **high abundance**
  - 330 mM Na_3PO_4 buffer

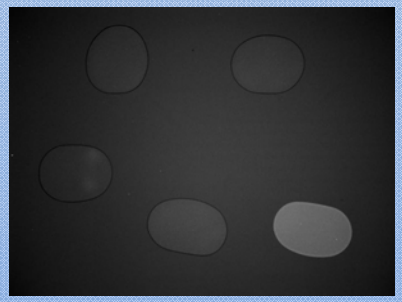
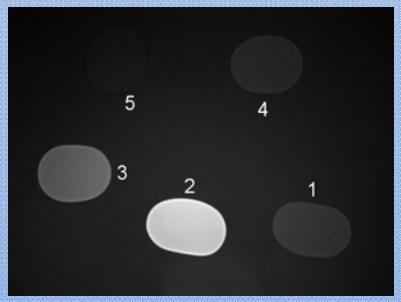
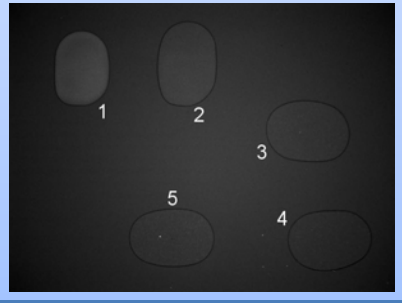
- Second downstream capillary dispenses 100 nL of picogreen for on-demand labeling on the DMF device
- Conditions:
  - 2 µL HAC column cartridge
  - Flow rate ~XX µL/hour
  - Step gradient 10mM: 160mM: 320mM Na_3PO_4

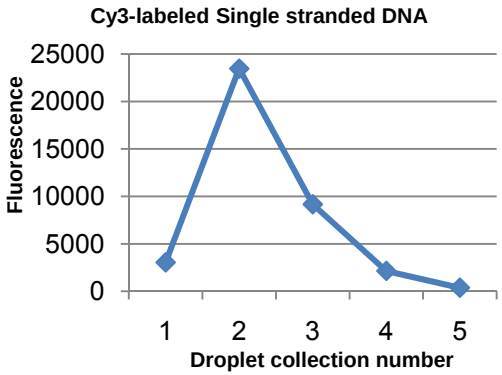


Kim & VanderNoot, 2010

Sandia National Laboratories

Demonstration of HAC normalization module integration w/ DMF using a capillary interface

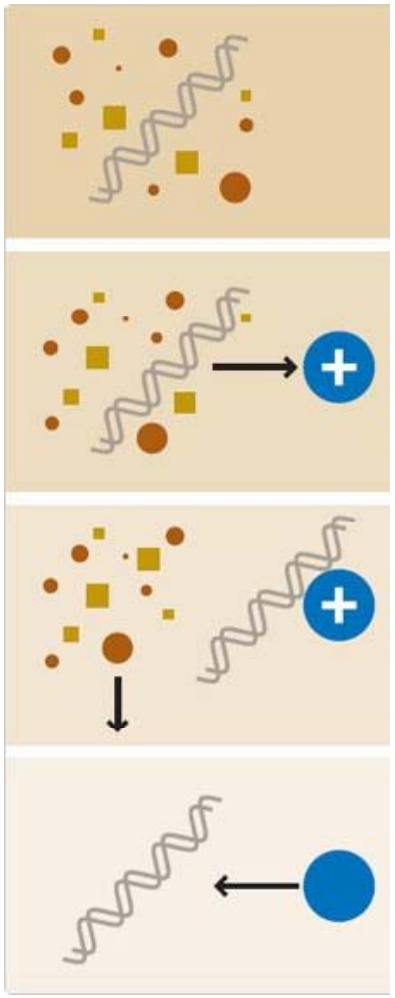
	Picogreen	Cy3
ssDNA fraction		
dsDNA fraction		



ssDNA eluting peak is discritized into droplets on the DMF platform

Sandia National Laboratories

Magnetic beads assay performed on DMF captures DNA effectively



DNA + Contaminants

DNA with contaminants
Ex) Excessive salts, enzymes

DNA binding

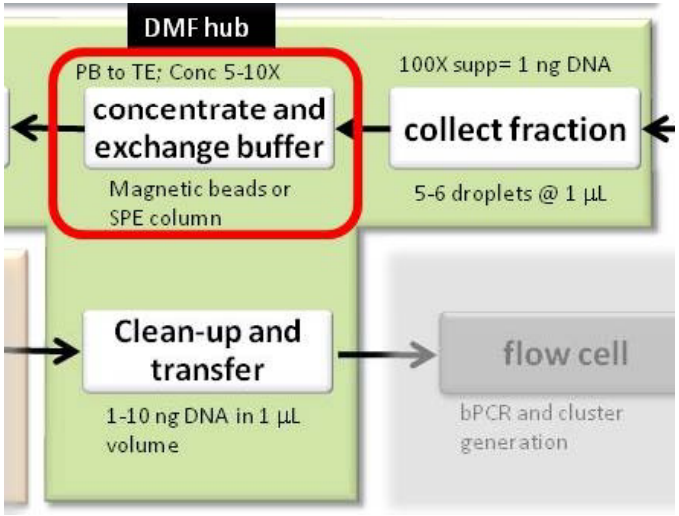
Purification buffer lowers pH
Beads added take positive charges & bind DNA

Contaminants Removal

Collect beads pellet
Wash & collect beads pellet

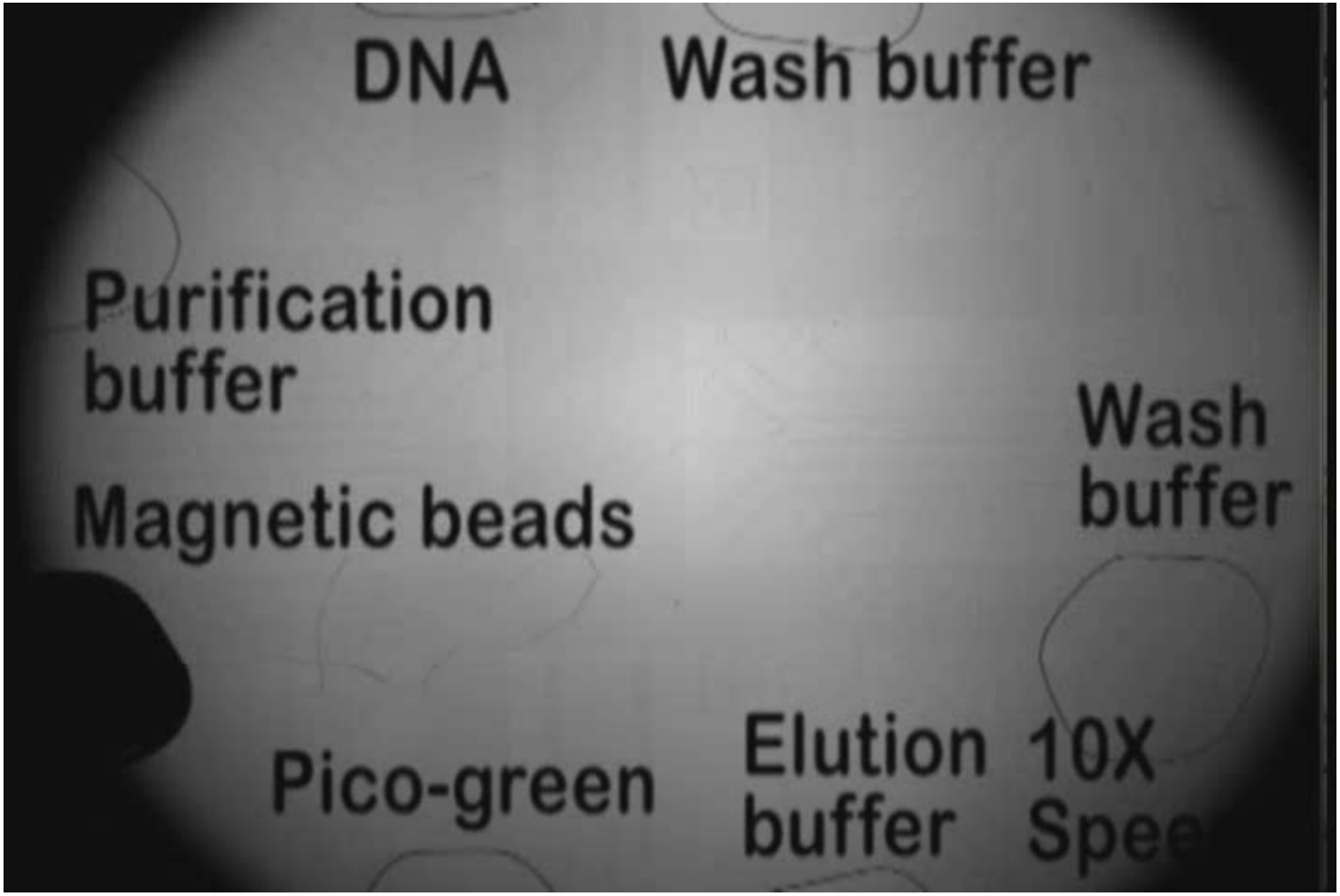
DNA Elution

Elution buffer raises pH
Collect elution buffer only



Sandia National Laboratories

Magnetic beads assay performed on DMF captures DNA effectively

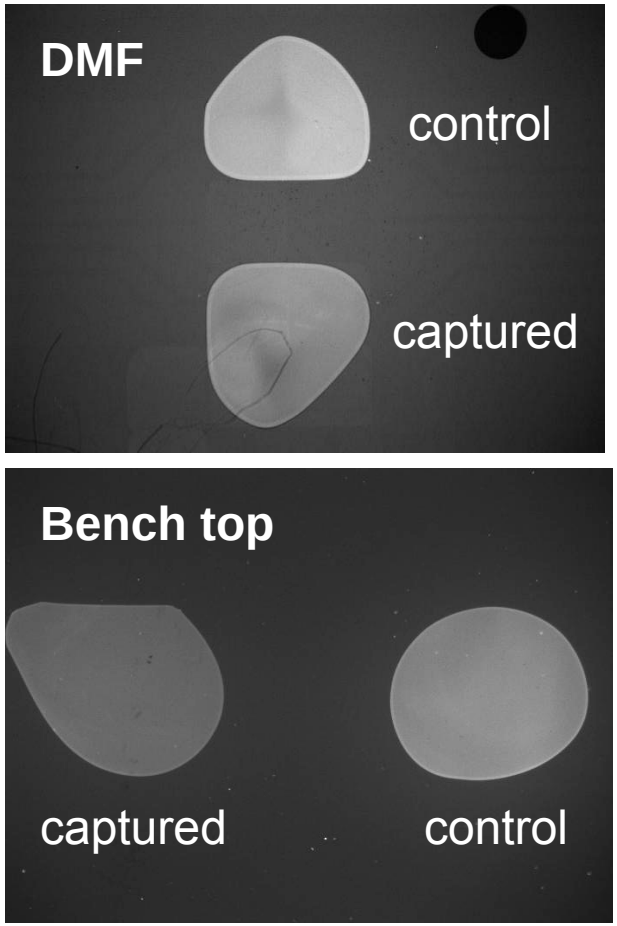


Kim, 2010

Sandia National Laboratories

Magnetic beads assay performed on DMF captures DNA effectively

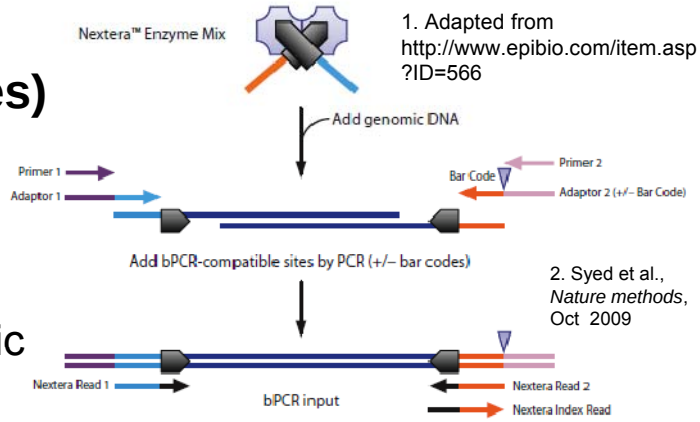
- Assay on DMF hub shows comparable capturing efficiency of DNA to Bench top process
- Reproducibility and efficiency studies are under way
- **Near future goals:**
  - Improve efficiency with capillary interface mixing
  - Adapt magnetic SPRI-beads methods for DNA sizing
  - Use biotinylated magnetic beads for exome capture probes for orthogonal suppression technique



Sandia National Laboratories

Transposase-mediated fragmentation and ligation simplifies library preparation

- **Simultaneously, randomly fragments and ligates Illumina adapters DNA (w/ barcodes)**
  - Based on *cut-and-paste* transposition
  - 5 minute reaction @ 55 deg C
  - Scalable to pg levels
  - Reaction is amendable to DMF and microfluidic separation platform



1. Adapted from http://www.epibio.com/item.asp?ID=566

2. Syed et al., *Nature methods*, Oct 2009

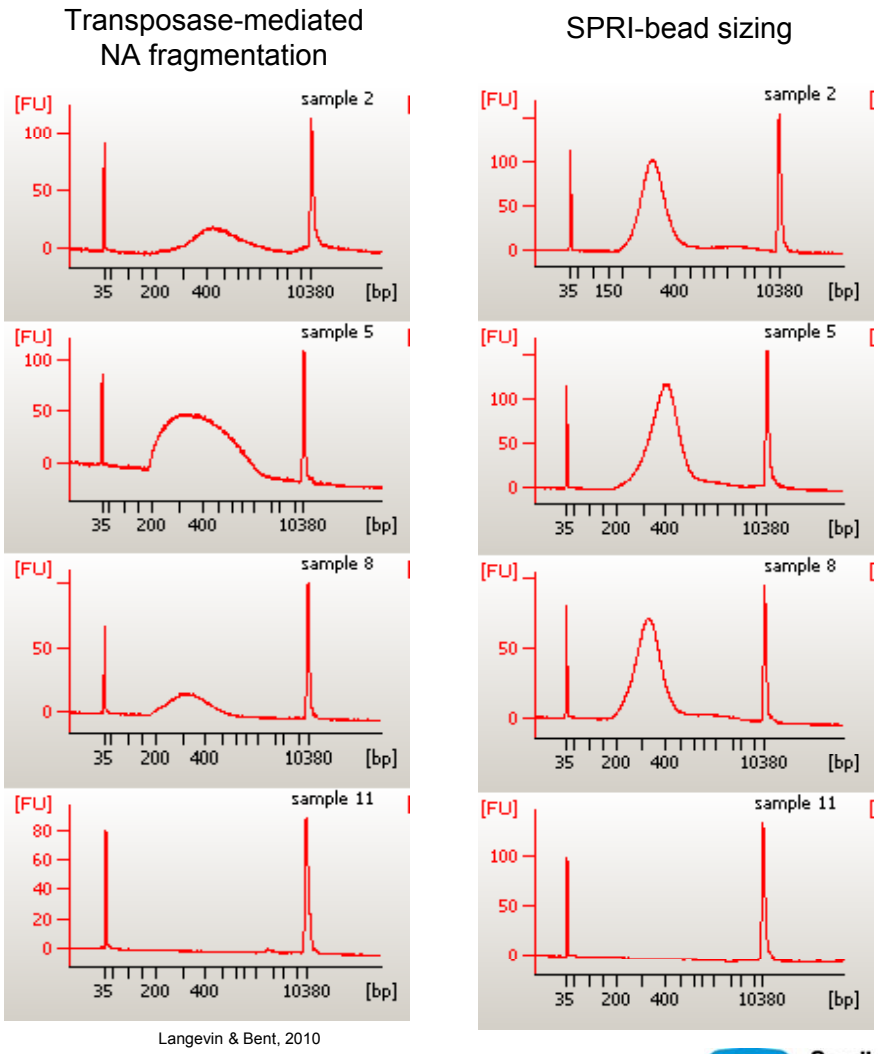
Illumina Protocol (µg)	min	Nextera Transposase-mediated(~50 ng)	min	AMB-adapted Transposase (~pg)	min
Fragmentation	30	Add Nextera™ Enzyme Mix to DNA	5	React transposase + DNA on DMF device with thermal capillary reactors	5
Collection	15				
Concentration	15				
Size Selection	60				
End-Repair	60			**Quantitation of DNA before and after PCR using DMF interfaced chip electrophoresis**	**15**
Clean-Up	15				
A-Tailing	30				
Adaptor Ligation	60				
Clean-Up	15	Clean-Up and size selection	60	Bead-based clean-up and size separation	~20
Benchtop PCR (Enrichment)	~60	Benchtop PCR	60	DMF interfaced PCR	~20 min

~ 6 hrs **~ 2 hrs** **~1 hour**

National Laboratories

Transposase-mediated prep outperforms standard library prep for small samples amounts

- Standard Illumina protocol is inefficient and poorly suited to small sample amounts (> 2-3 µg of DNA)
- 50 ng input DNA Transposase-mediated fragmentation yields after 12 PCR cycles ~250 ng at the benchtop
- Transposase fragmentation produces a “tighter” fragmentation pattern than standard Illumina prep
  - ~200-500 bp range
  - PBMC w/ *Francisella Tularensis* 1:100 and 1:10,000.
- SPRI magnetic beads further decrease the size distribution
  - (Avg:358 bp c_v= 30)



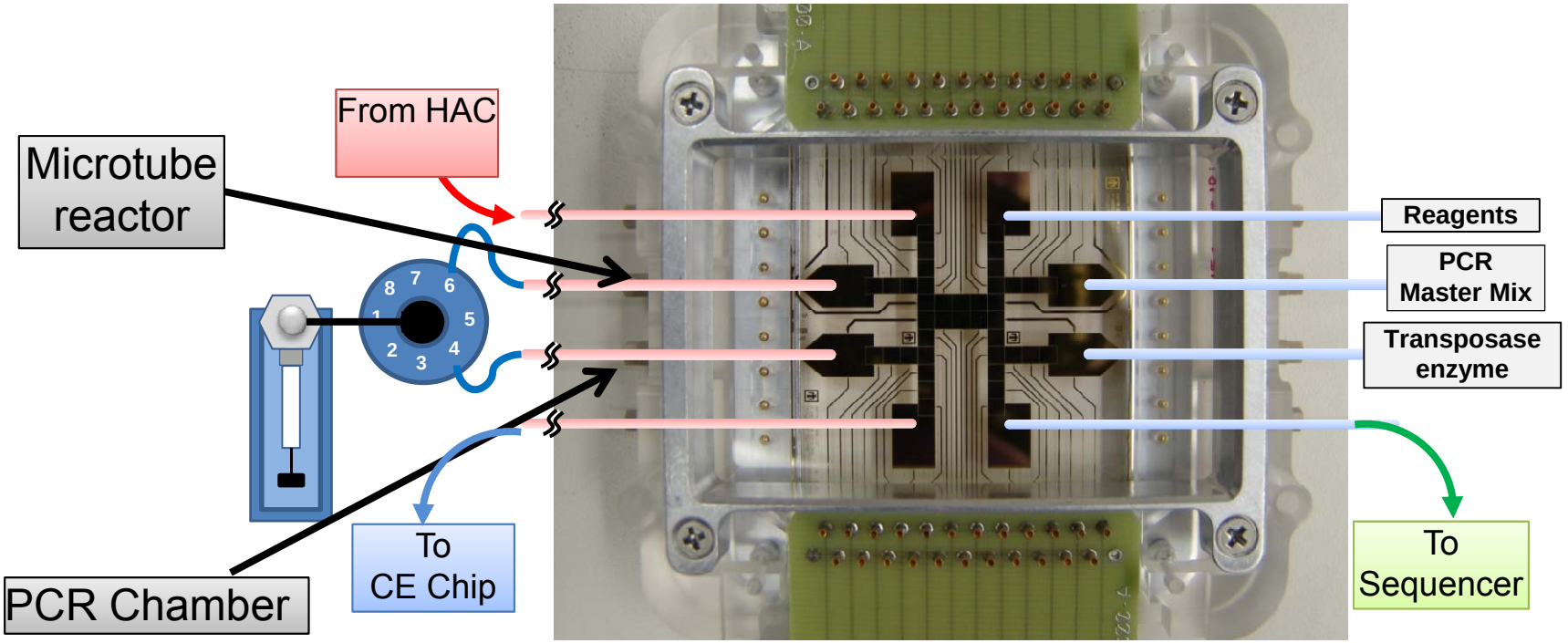
Langevin & Bent, 2010

Bioanalyzer runs

Sandia National Laboratories

Transposase-mediated reactions at pg-scale can be adapted to DMF device through interfaced modules

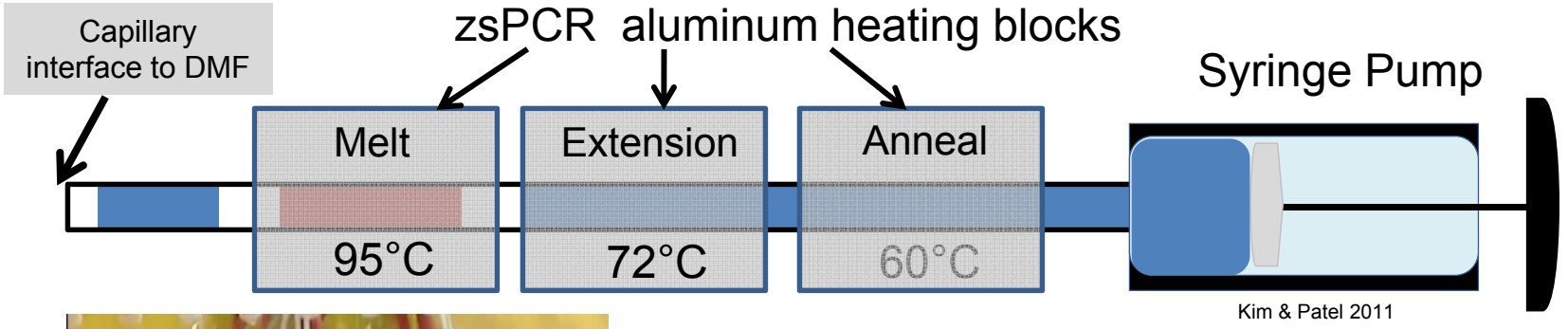
- Enzyme reagents and buffers
- Clean-up/sizing reagents
- Transposase reaction at 55° and PCR amplify
- Method to quantitate DNA through electrophoretic separations



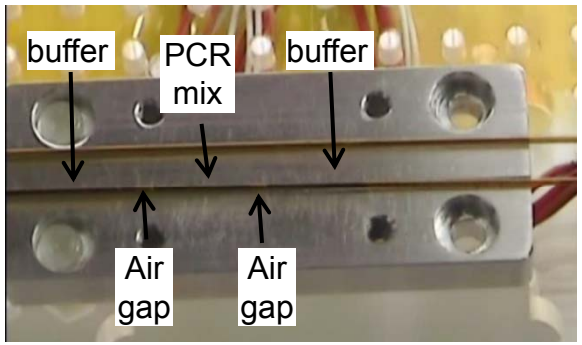
Sandia National Laboratories

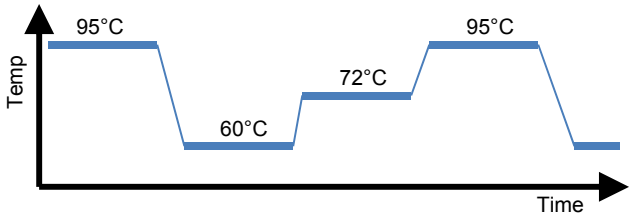
Zone shuttling PCR for rapid amplification is well-suited for the DMF platform

- Sample bolus with air separators is shuttled between three temperature zones for amplification
  - Scalable from nL to µL (change tubing diameter)
  - Rapid temp cycling (seconds)
  - Interface with DMF and droplet reagents
  - Disposable (inserted tube)
  - Multiplex using multiple capillaries



Kim & Patel 2011

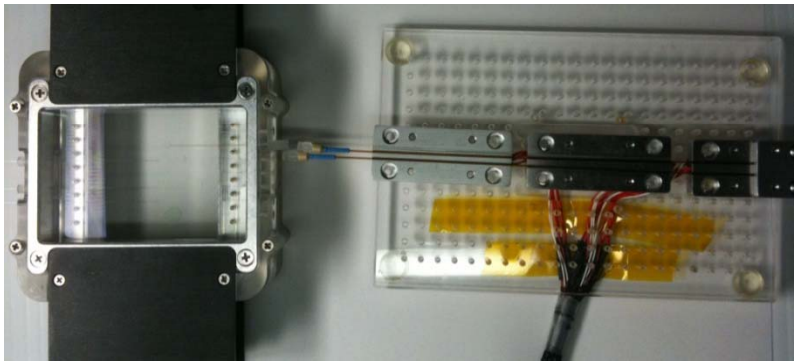




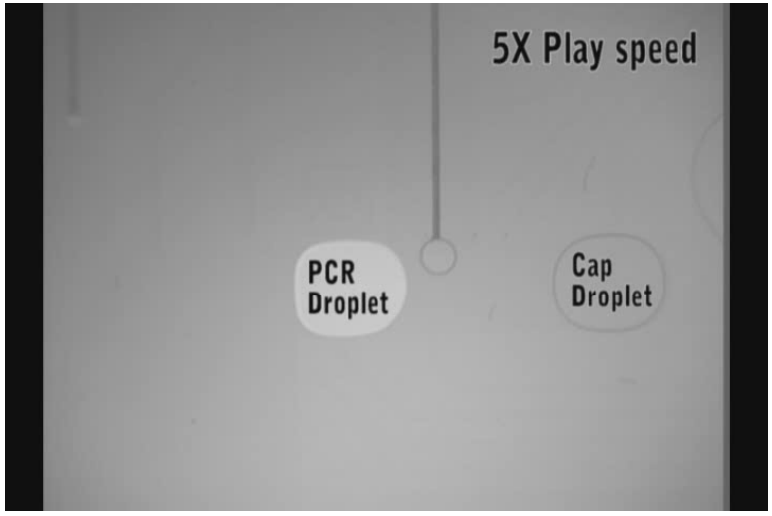
Sandia National Laboratories

Hardware and fluidic control for zone shuttling PCR has been demonstrated

- 2-6 µL zsPCR reactor
  - (three temp zones 95, 72, 60° C)
- Repurposed for enzyme-based thermal reactors
    - Nextera @ 55 deg for 5 min
- Demonstrating proof-of-principle



Zone shutting PCR interfaced to DMF



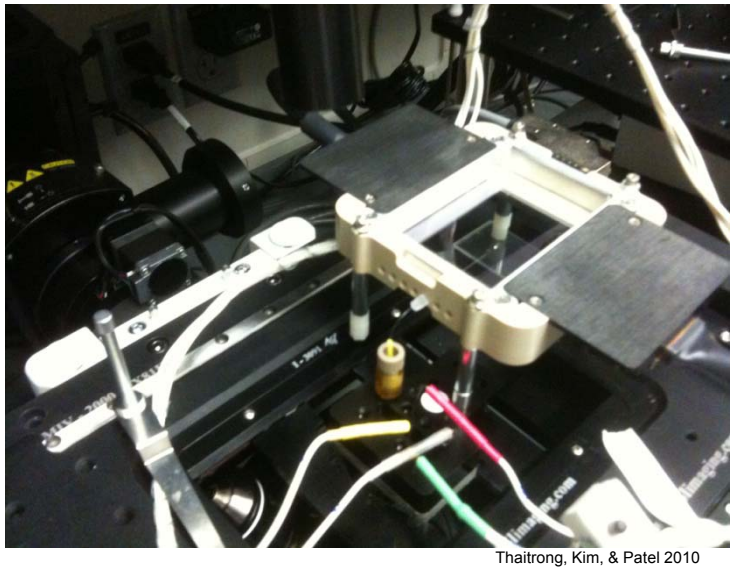
Load PCR reactor

Collect PCR products Patel ,2010

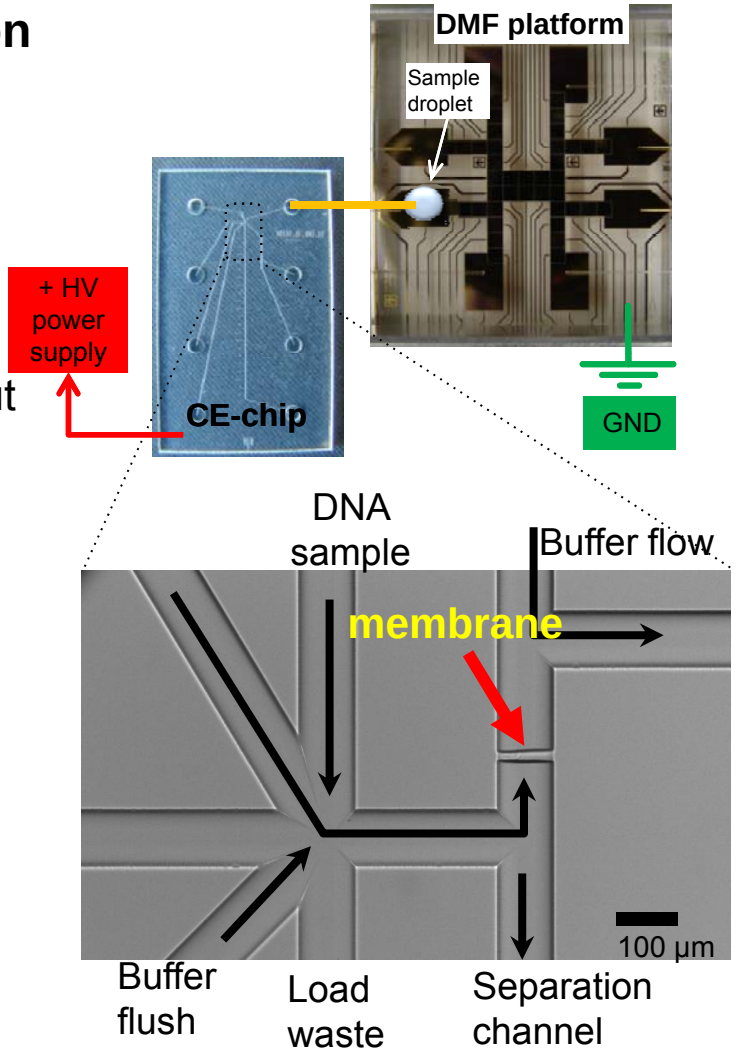
Sandia National Laboratories

Quantitative analysis of fragmentation and PCR products from the DMF droplet with CE

- **Electrophoresis of DNA from a droplet on DMF device to a nanoporous membrane**
  - Use DMF ground plane to complete circuit
  - Sample small fraction of droplet for separation analysis
- **Nanoporous membrane**
  - Traps DNA to concentrate at the surface but allow ions to pass
  - Photo-polymerized *in situ* with a UV laser



Thaitrong, Kim, & Patel 2010

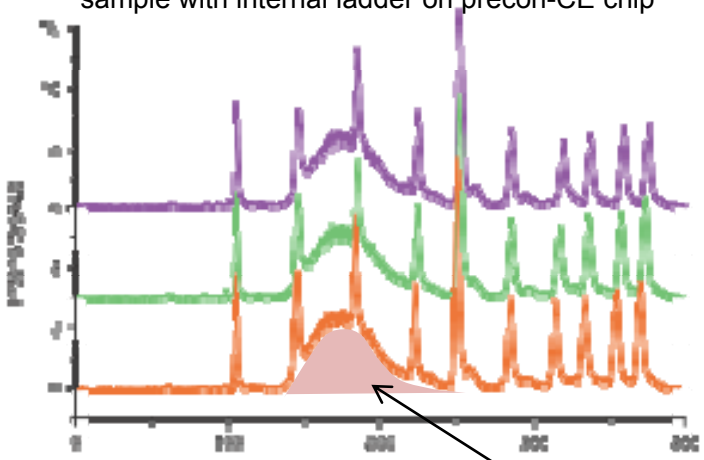


Sandia National Laboratories

Preconcentration integrated with CE-chip enables sensitive LIF detection of NA reaction products

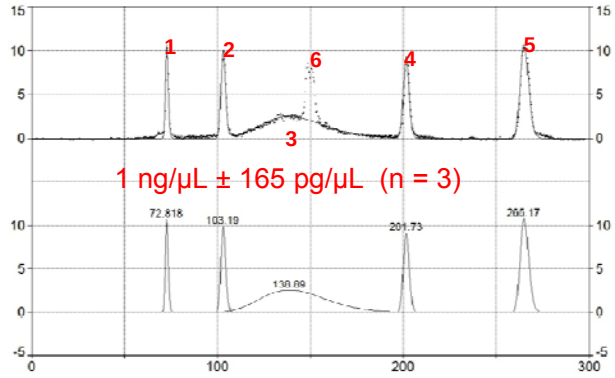
- Preconcentration minimizes effect of diffusion and dilution of DNA sample
- DNA stacks at the membrane interface a well-defined injection plug at the separation column
- Preconcentration increase detection limits (10-100-fold)
- Add an internal ladder to quantitate our reaction products
  - Calculated using peak area of co-injected ladder as an internal standard

Separation of transposase-mediated fragmentation sample with internal ladder on precon-CE chip



E_{sep} = 125 V/cm
T_{conc} = 100 s
SYBR green intercalating dye

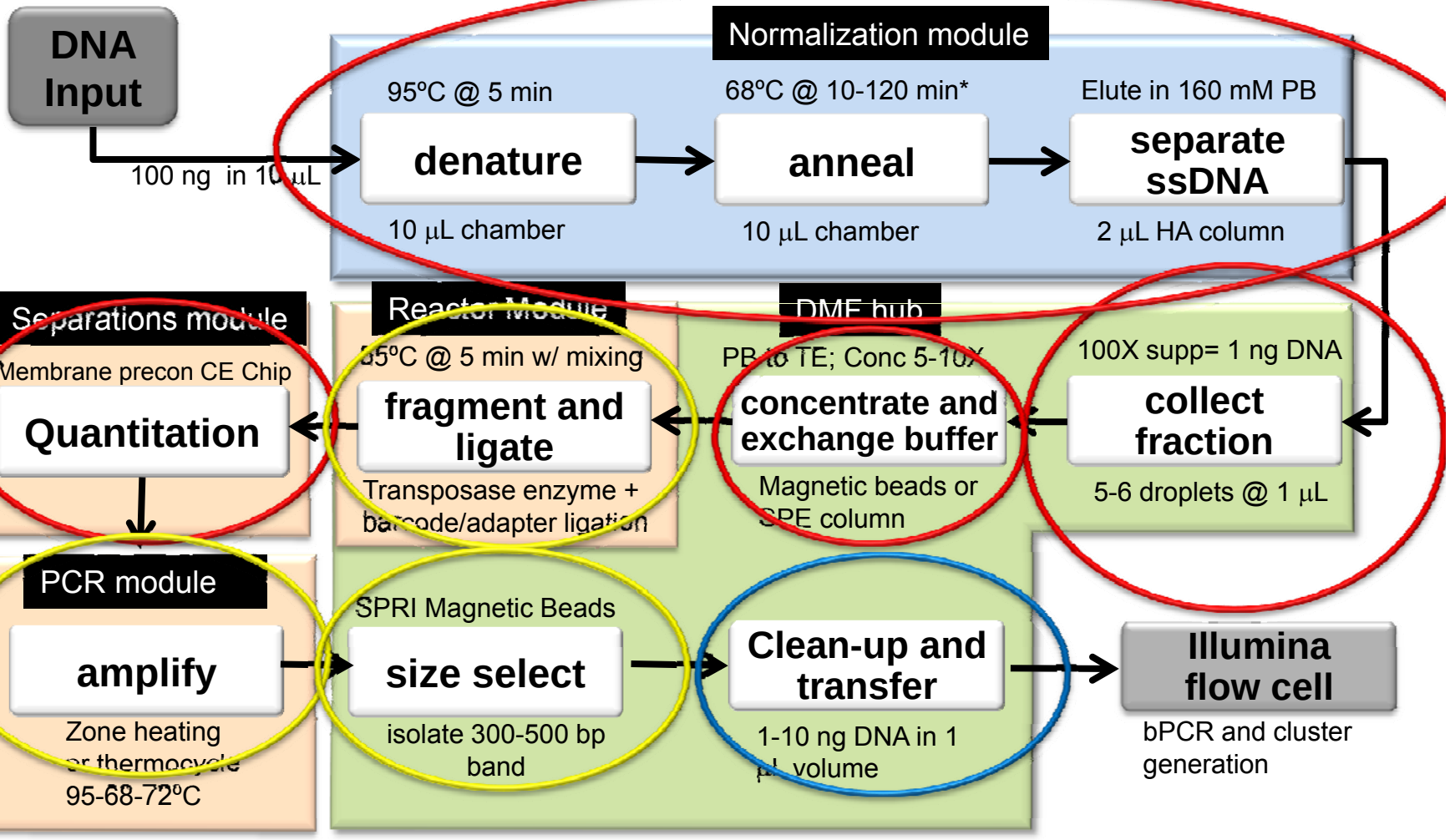
cDNA PBMC w/ F. Tularensis 1:10,000



Thaitrong, 2010

Sandia National Laboratories

NA prototype process flow diagram for normalizing + NA sequencing preparation



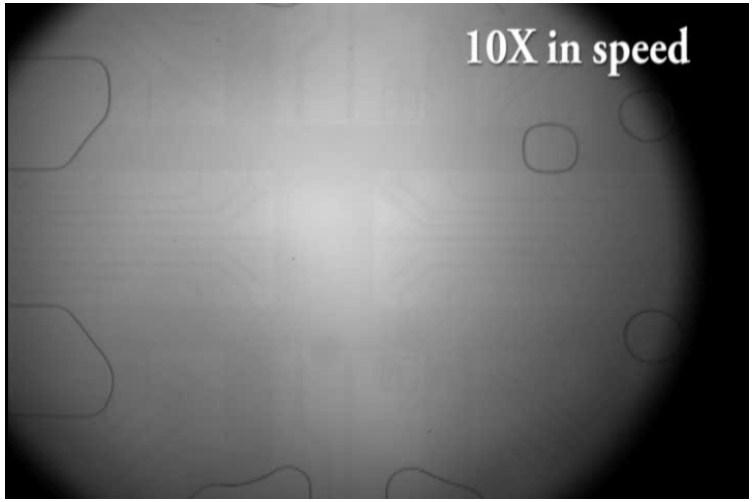
Sandia National Laboratories

Acknowledgments

- **Automated Molecular Biology Team**
  - Team lead: Kamlesh Patel
  - **Hanyoup Kim, Numrin Thaitrong**, Robert Meagher, **Victoria VanderNoot**, Conrad James, Carlton Brooks,
  - Engineering Team **Michael Bartsch**, **Ron Renzi**, Jim He, Jim Van De Vreugde, Ron Renzi, Mark Claudnic
- **Host Pathogen Molecular Biology Team**
  - Team Lead: Steve Branda
  - **Stan Langevin, Zach Bent**, Sadie LaBauve, Bryan Carson, Julie Kaiser, Pam Lane, Bryce Ricken, Deanna Curtis
- **Data knowledge and Discovery Team**
  - Team Lead: Joe Schoeniger
  - Milind Misra, Kelly Williams, Amy Powell, Chi-Chi May
- **Project Management and PI:**
  - Duane Lindner, Malin Young, and Todd Lane

Carry over experiments

Well / Set	Avg C(t)	Remarks
New Set 1	N/A	Elution buffer
New Set 2	N/A	Elution buffer
New Set 3	N/A	Baseline of DMF device
	DNA droplet passage	
New Set 4	32.27	1st buffer droplet
New Set 5	35.92	2nd buffer droplet
	EB Flush with 3X droplet volume	
New Set 6	39.49	1st buffer droplet
New Set 7	N/A	2nd buffer droplet
New Set 13	13.48	19ng/ul cDNA from DMF device
New Set 14	13.6	19ng/ul cDNA before DMF

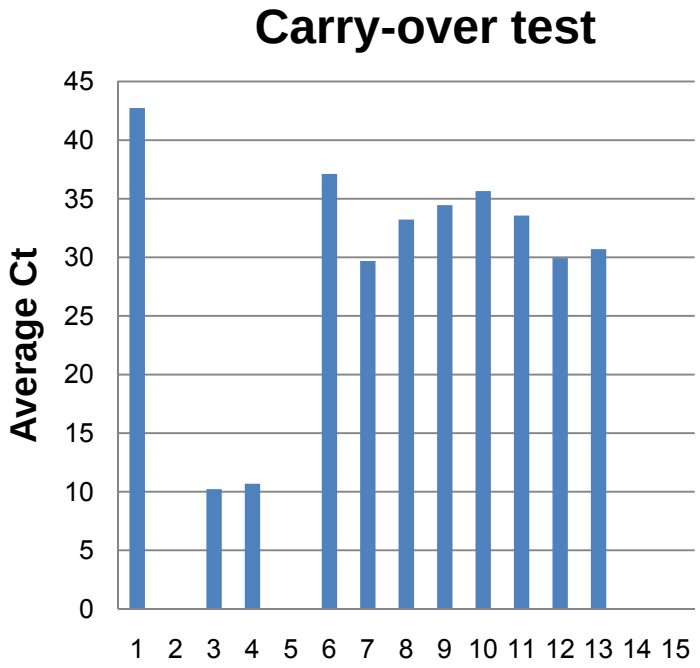


Other cleaning agents: Bleaching solution, UV treatment, DNA zap

Sandia National Laboratories

Ignorable carry-over on DMF

Well / Set	Efficiency	Avg C(t)	Remarks
New Set 1	73.06%	42.74	Elution buffer
New Set 3	85.19%	10.22	cDNA@19ng/ul
New Set 4	116.47%	10.69	cDNA@19ng/ul
New Set 5	N/A	N/A	Baseline of DMF device
New Set 6	56.57%	37.11	Baseline of DMF device
	DNA droplet passage		
New Set 7	81.37%	29.69	1st buffer droplet
New Set 8	81.39%	33.22	2nd buffer droplet
New Set 9	79.94%	34.45	3rd buffer droplet
New Set 10	69.47%	35.65	4th buffer droplet
	Ethanol droplet passage		
New Set 11	76.90%	33.57	buffer droplet
	DNA droplet passage		
New Set 12	85.18%	29.93	1st buffer droplet
New Set 13	93.05%	30.7	2nd buffer droplet
New Set 14	N/A	N/A	DI water
New Set 15	N/A	N/A	No DNA control



Sandia National Laboratories

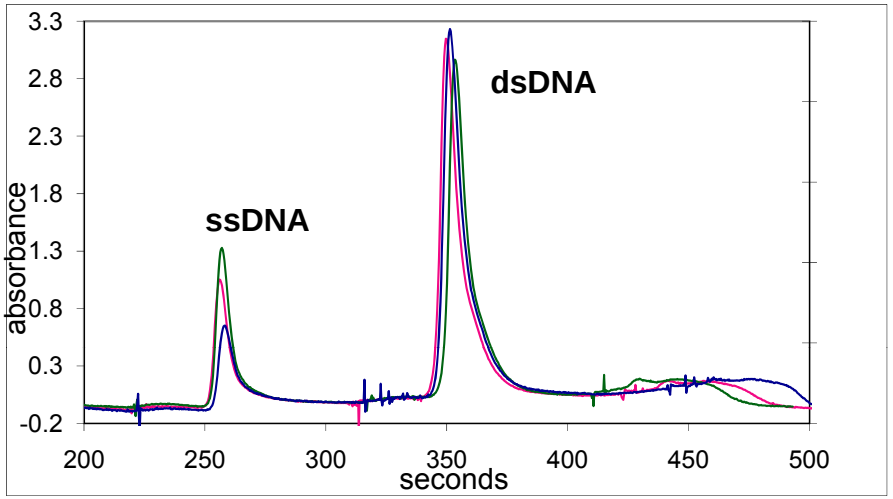
DMF hub operation optimized and tested compatible with DNA processing buffers

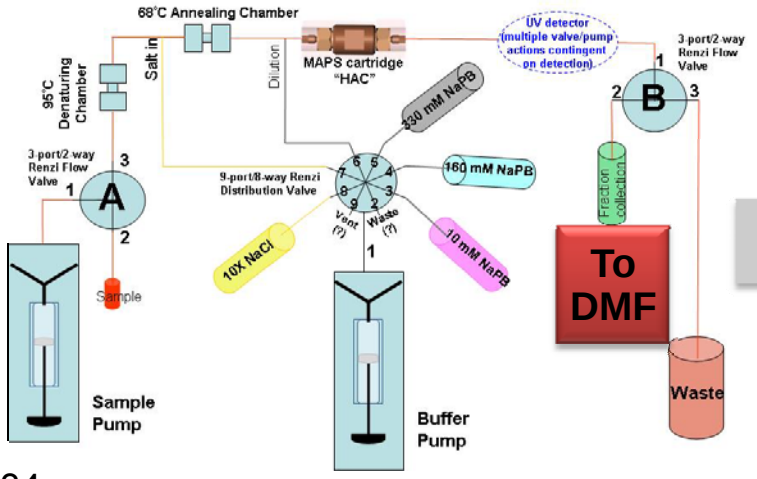
- Frequency: 15kHz (0 to 20kHz Tested)
- Dielectric: Parylene C (Su-8, PMMA, Glass tested)
- Actuation voltage

		Actuation voltage	Surface fouling
DI water		>100 Vrms	No
TE buffer		77 Vrms	No
Elution buffer (10 mM Tris-HCl, pH 8.5)		~70 Vrms	No
Buffers from normalization module	NaPB (10mM)	56 Vrms	No
	NaPB (160mM)	66 Vrms	No
	NaPB (330mM)	<70 Vrms	No
Ethanol		< 70 Vrms	No
DNA zap (DNA cleaning agent)		<70 Vrms	No
5% bleaching solution		<70 Vrms	No
Buffers for magnetic beads DNA capturing kit	Purification buffer	<70 Vrms	Controllable
	Beads solution	<70 Vrms	No
	Wash buffer	<70 Vrms	No
PBS		<70 Vrms	No

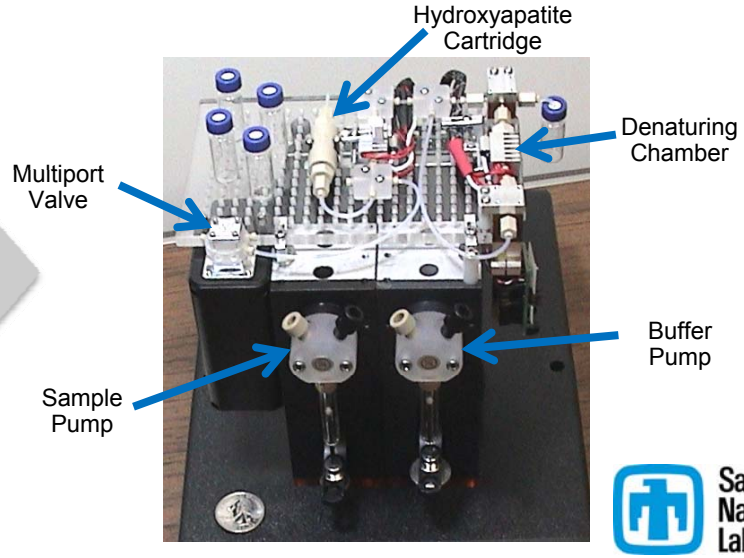
Hydroxyapatite chromatography separates ssDNA from dsDNA for low abundance collection

- **Normalization**
  - Preferential removal of high abundance DNA through selective hybridization
  - Denature, anneal, separate, collect
- ssDNA fraction = low abundance
  - 160 mM Na_3PO_4 buffer
- dsDNA fraction = high abundance
  - 330 mM Na_3PO_4 buffer
  - Fluidically easy to integrate
  - Increased reliability and automation
  - Decrease sample processing time





Automated mesoscale normalization module



Sandia National Laboratories