SAND2011-1794C

# Automated nucleic acid library preparation for sequence-based unknown pathogen detection

**Hanyoup Kim, 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

- **The emergence of unknown pathogens**
- **Approach to detecting unknown pathogens from clinical samples using next-gen sequencing**
- **Automated Molecular Biology (AMB) Platform**
- **Conclusions**



Sandia National Laboratories

# Novel pathogens increasingly threaten national security & public health

- **Factors promoting pathogen emergence:**
  - Increase  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

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 strains 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 bioengineer 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

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

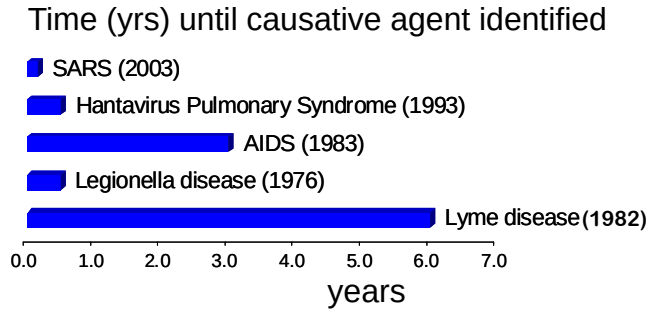
New York Times Feb 10, 2010

Sandia National Laboratories

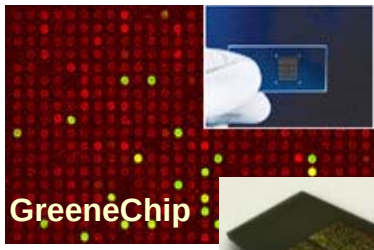
# We need new tools for rapid identification & characterization of novel pathogens

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

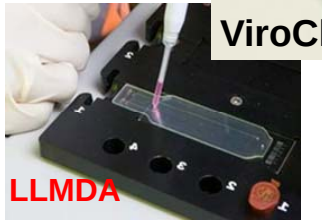
- Identification of a novel causative agent by conventional methods can take ***months to years***.




- Modern probe-based methods are fast, but are often confounded by novel pathogens.
  - pathogens can escape detection
  - unanticipated features giving false negatives
  - Unusual profiles can be difficult to interpret







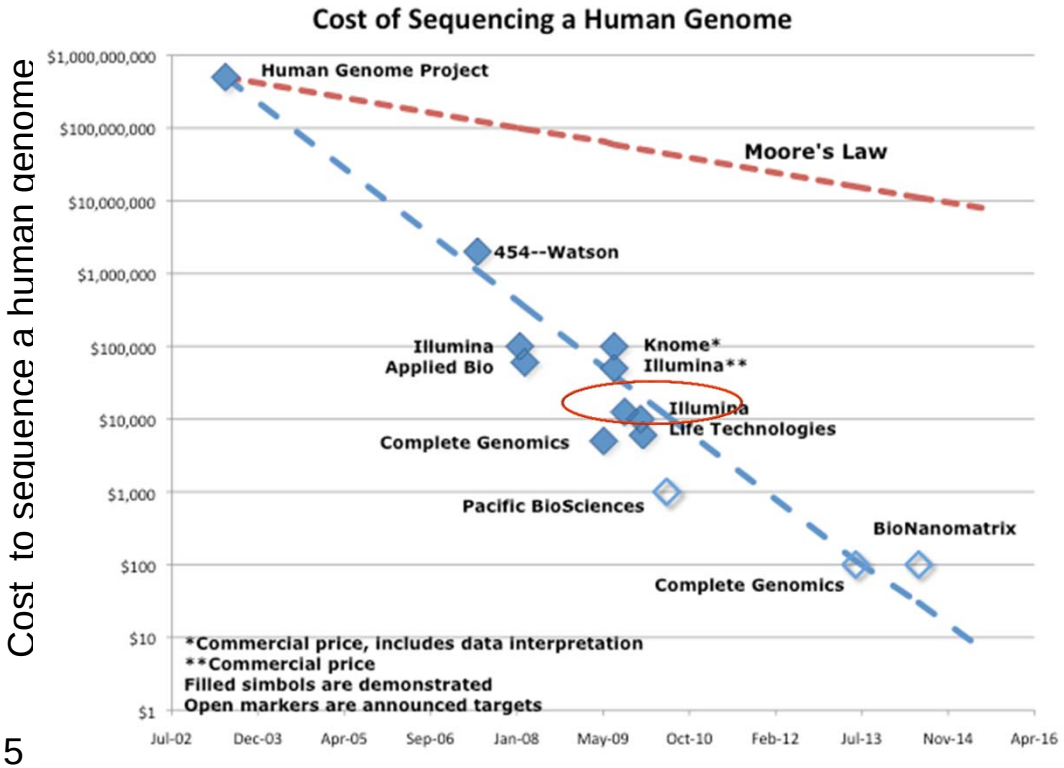

Sandia National Laboratories

# Next Generation Sequencing is a transformational technology for pathogen characterization

- 150-200 Gb (2 x 100 bp run)
- Two human genomes (30 x coverage) in a single run under $10,000 per sample

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

| 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 |

*Deplete non-informative NA to improve efficiency of NGS analysis*



Sandia National Laboratories

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

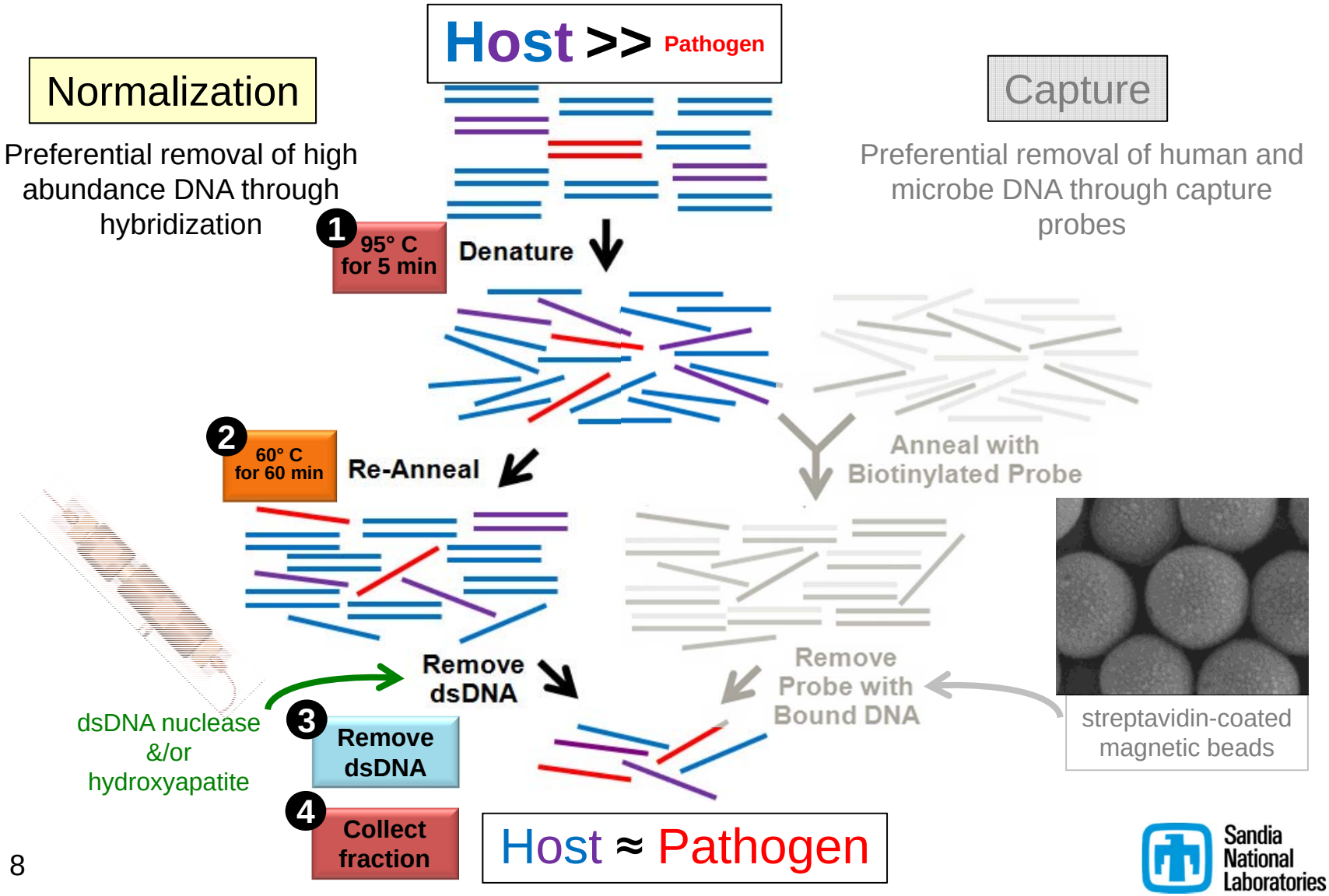
**Rapid Threat Organism Recognition (RapTOR) system**

- **Goal:** Efficient analysis of pathogen nucleic acids (NA) in clinical samples *via* targeted Next Generation Sequencing (NGS)
- **Key advance:** Automated microfluidic platform to enable molecular suppression and NA 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 prior knowledge of a pathogen or culturing of organism; (deep genomic sequencing)
- Automate the nucleic acid processing for operation at federal and state-wide laboratories

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

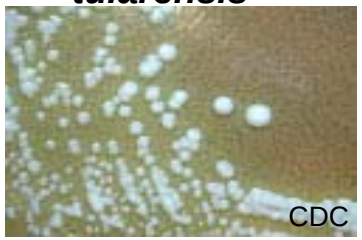




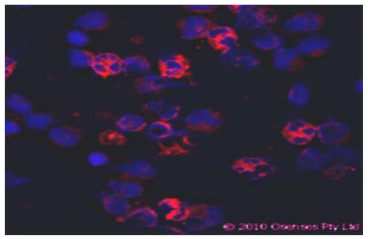

Sandia National Laboratories

# 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**

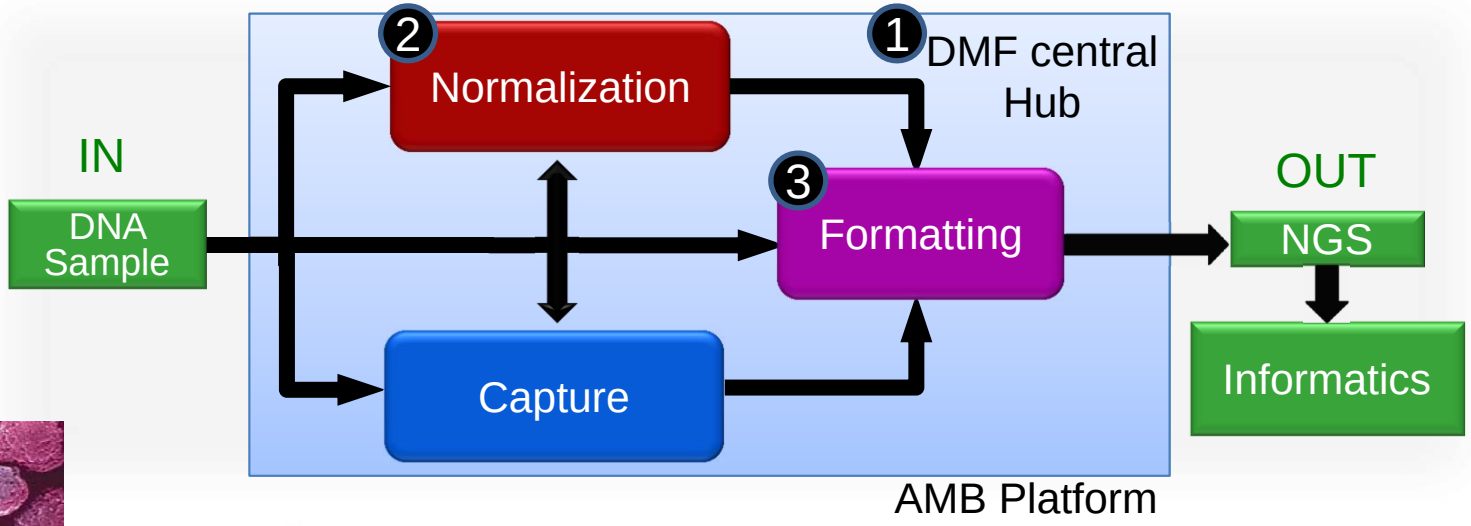
**Benchtop Library Prep for Illumina NGS**

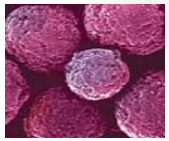
**Transcript Analysis with Databases**

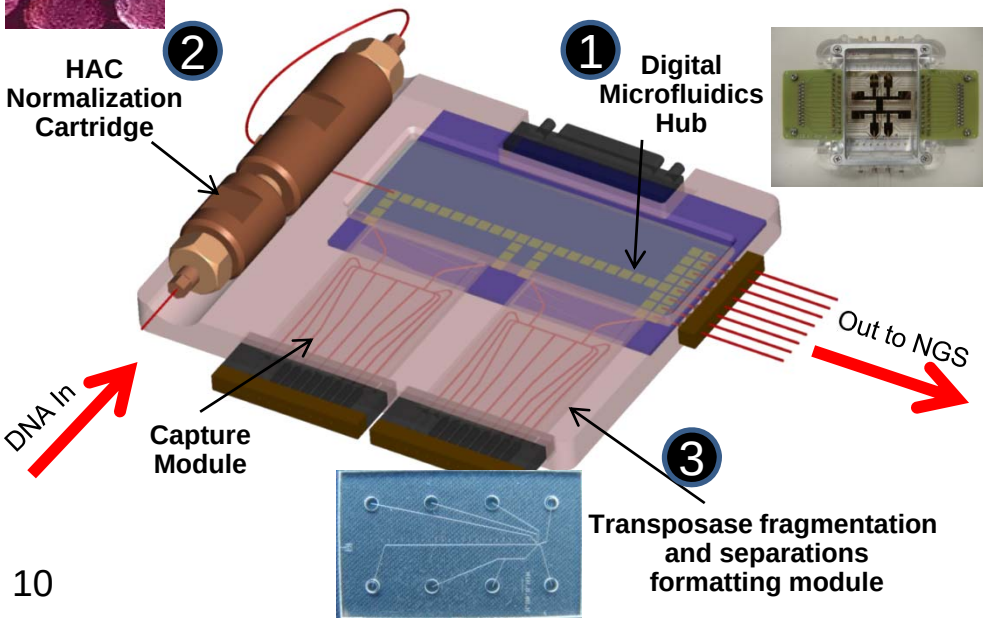


Sandia National Laboratories

# AMB platform integrates suppression & library prep methods into a flexible architecture





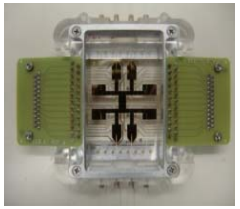

**Gen 1 requirements:**

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

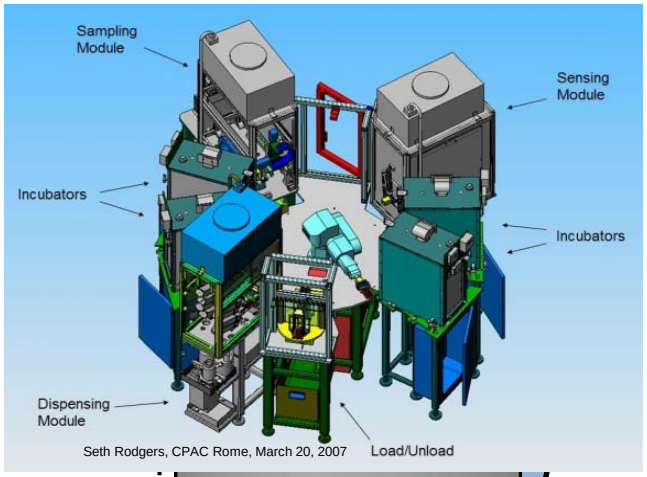


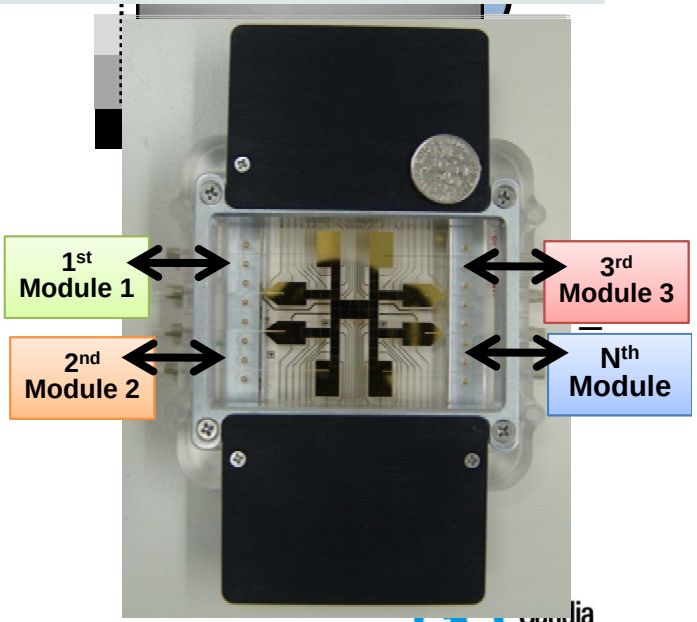
Sandia National Laboratories

# Core architecture of the AMB platform for NA processing is the Digital Microfluidic (DMF) Hub



- Use droplets as sample 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
  - Advantage
    - 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

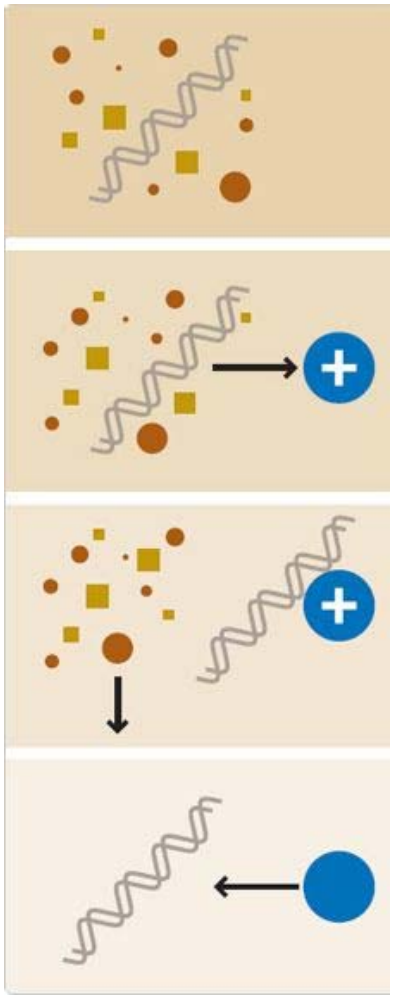





Sandia National Laboratories

# Magnetic beads assay performed on DMF captures DNA effectively



**DNA + Contaminants**

DNA with contaminants
Ex) Excessive salts, enzymes

**DNA binding**

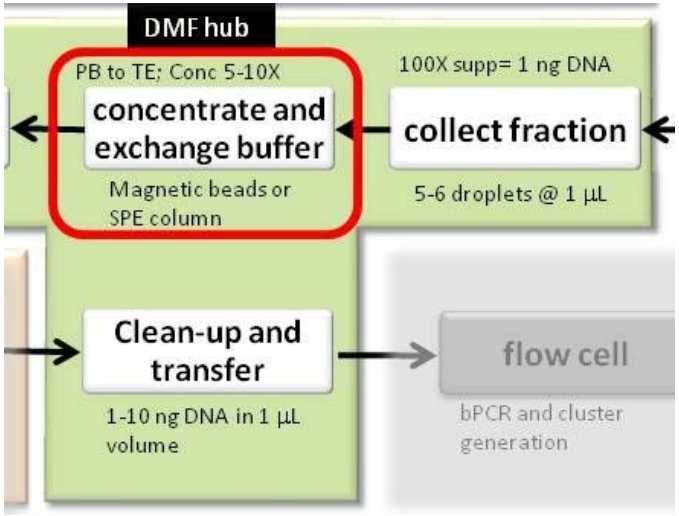
Purification buffer lowers pH
Beads added take positive charges & bind DNA

**Contaminants Removal**

Collect beads pellet
Wash & collect beads pellet

**DNA Elution**

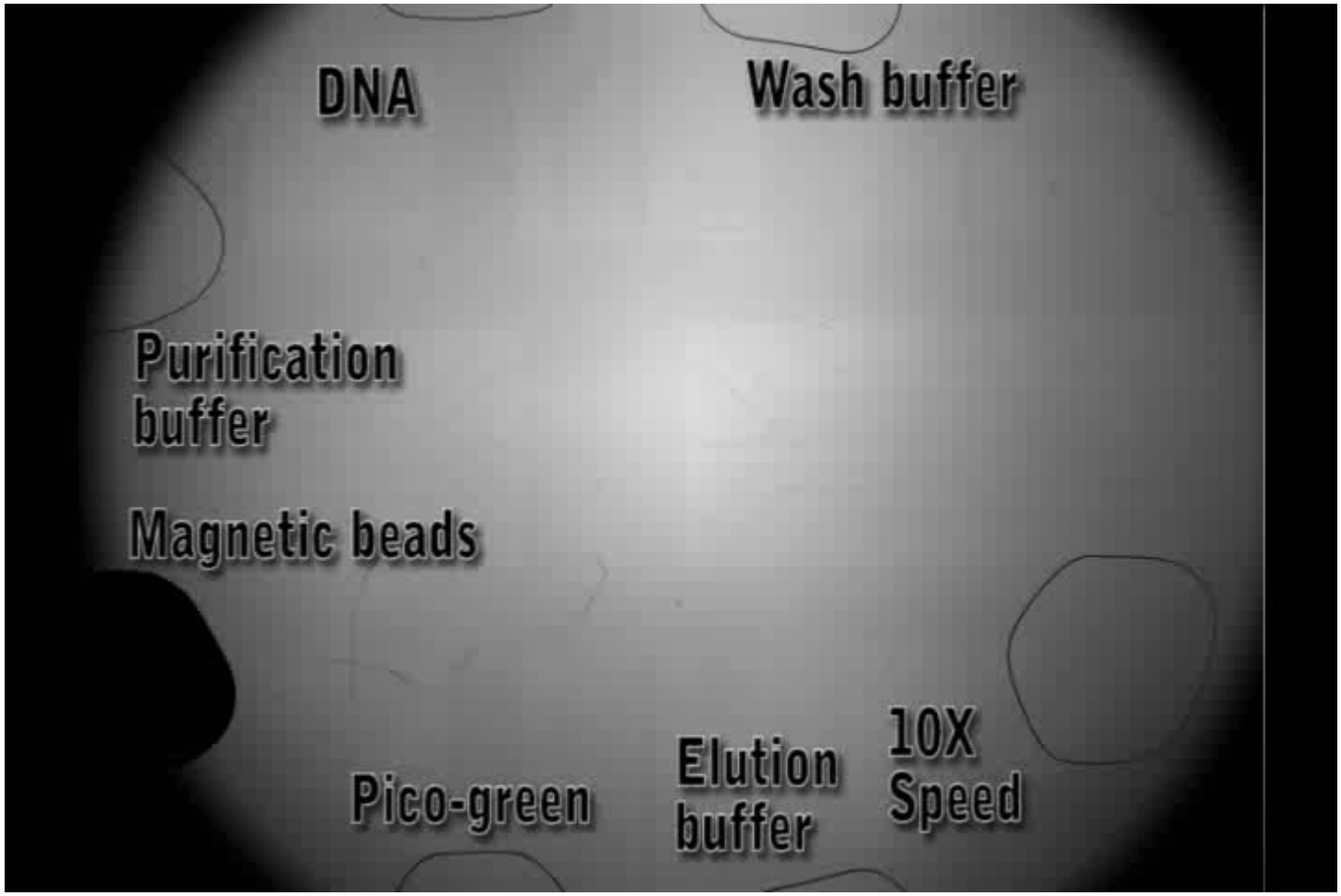
Elution buffer raises pH
Collect elution buffer only






Sandia National Laboratories

# Magnetic beads assay performed on DMF captures DNA effectively



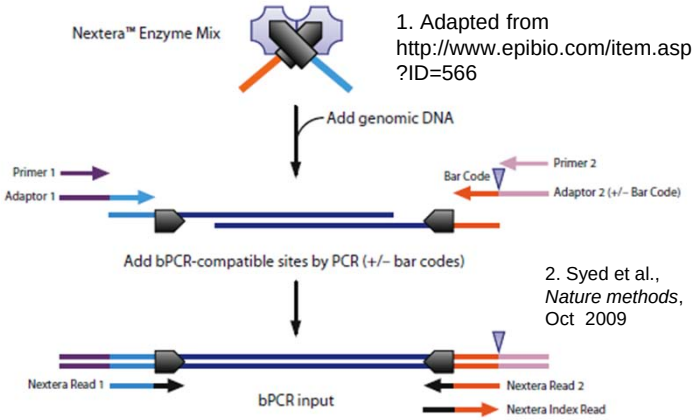

Kim, 2010



Sandia National Laboratories

# Transposase-mediated fragmentation and ligation simplifies library preparation

- **Single tube reaction,**
  - **fragments and ligates Illumina adapters DNA (w/ barcodes)**
  - **Drastically reduces number of steps and preparation time**




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

# 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

# Questions

Thank You



Sandia National Laboratories