

# RNA-Seq based studies of host-pathogen interactions

Zachary W. Bent  
Systems Biology  
Sandia National Labs

100 microns

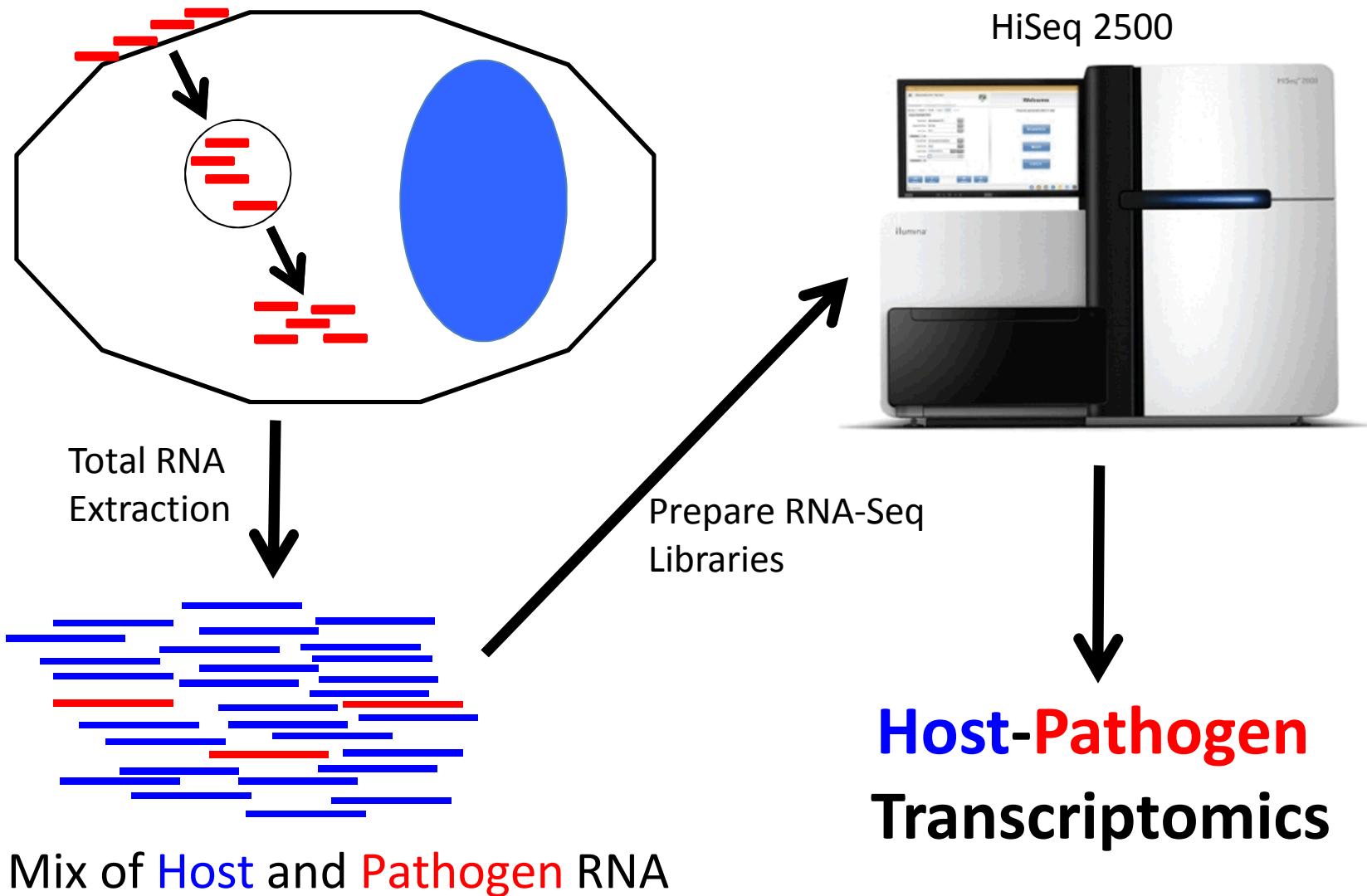


20 microns

# Outline

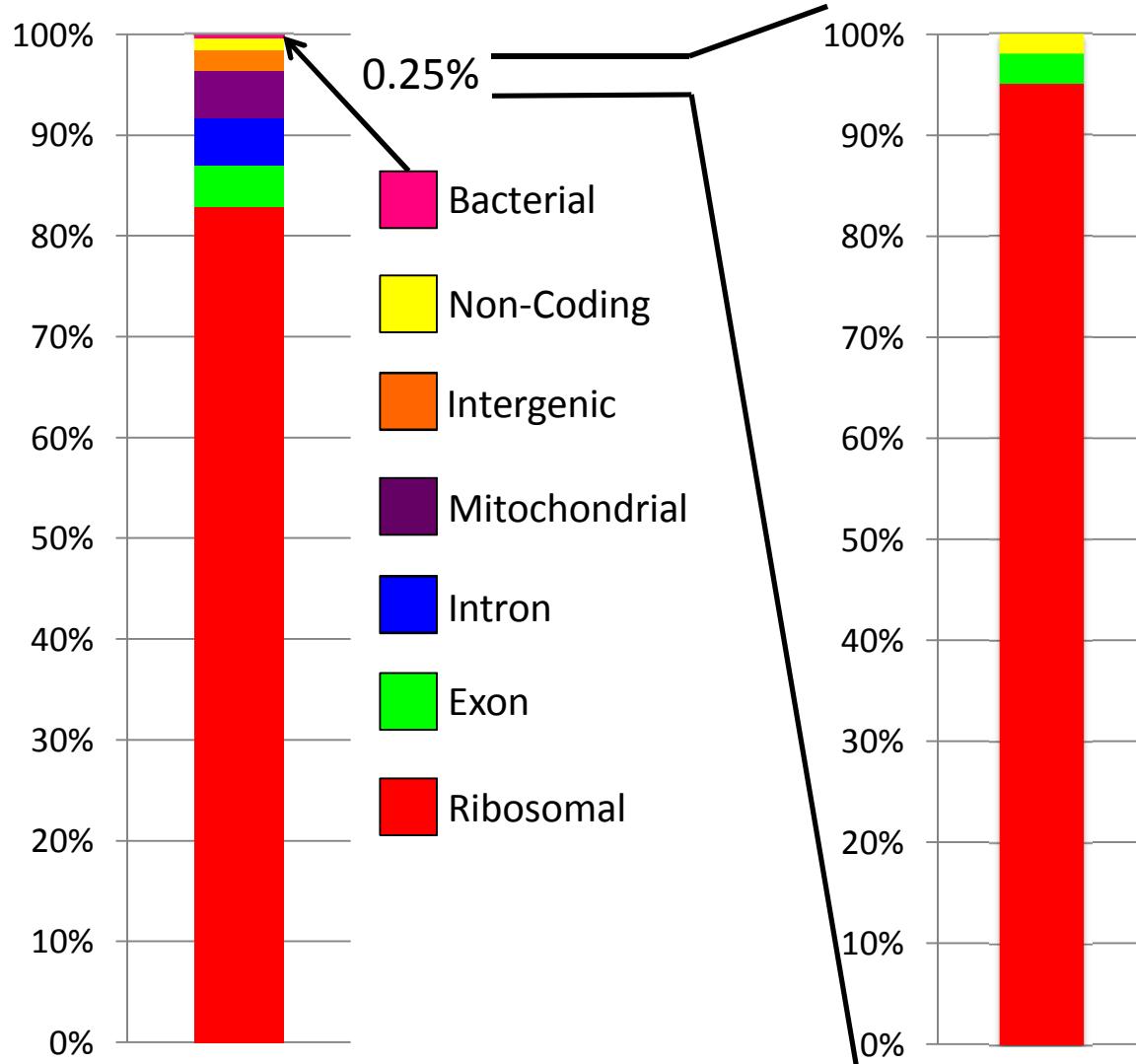
- Capture and host-pathogen transcriptomics
- Proof of concept: *Francisella tularensis*
- Intracellular *Yersinia enterocolitica*
- *In vivo* *Salmonella* Typhimurium transcriptomics and cell specific host response
- Future work and promising applications

# Host-Pathogen Interactions using RNA-Seq



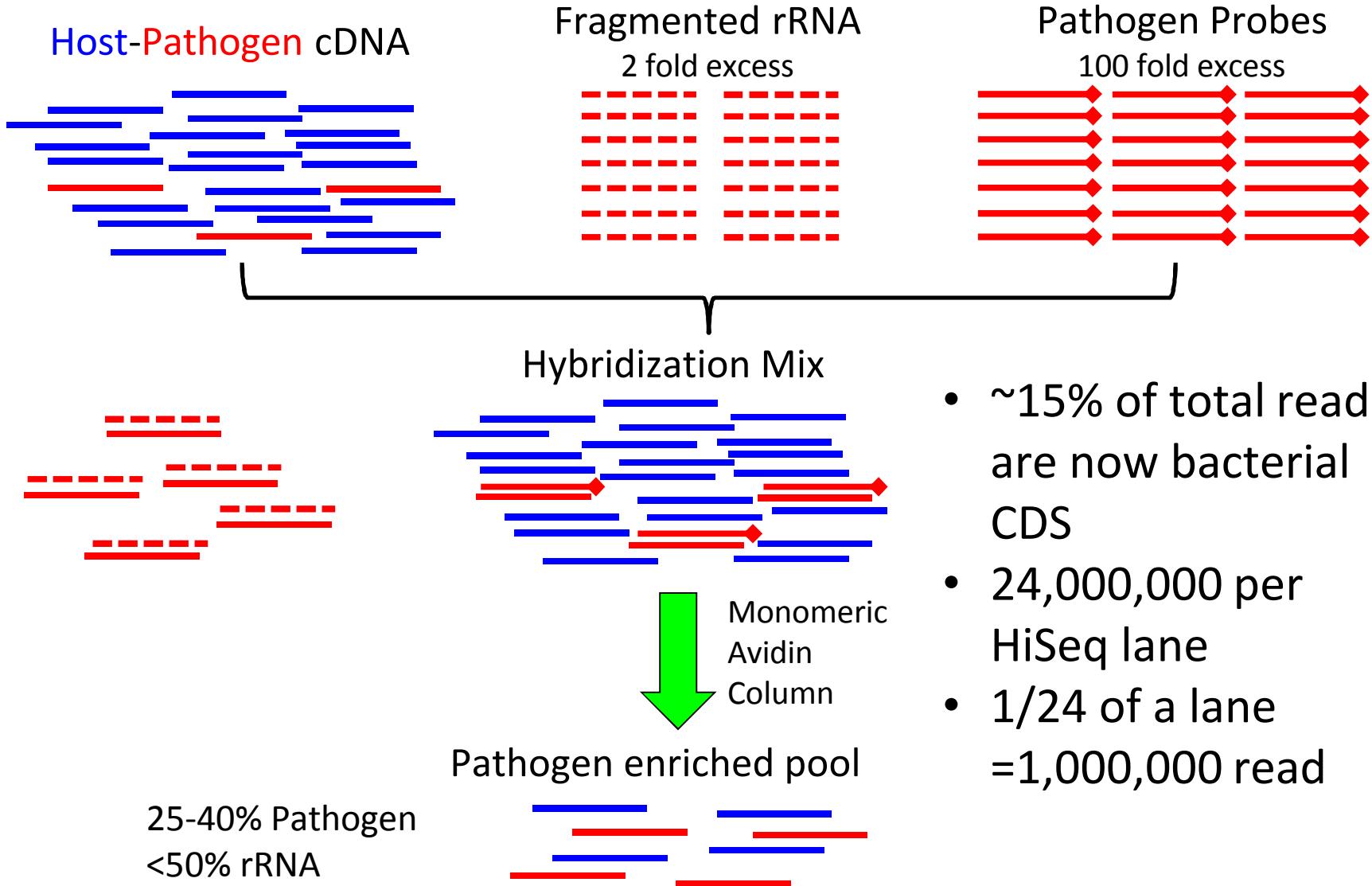
# Brute Force Sequencing is NOT Feasible for Bacterial Transcriptomes

99.75% Host



- 0.0075% of total reads are bacterial CDS
- 12000 reads per HiSeq lane
- ~83 HiSeq lanes = 1,000,000 reads

# Capture-based Bacterial Transcript Enrichment and rRNA Depletion



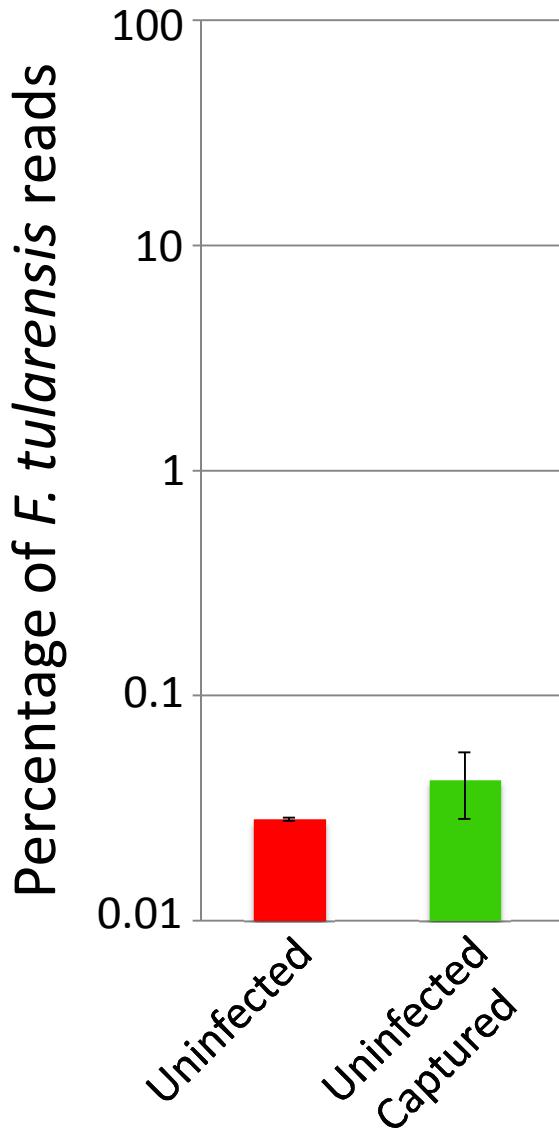
# *Francisella tularensis*

- Gram-negative, zoonotic, facultative intracellular bacteria
- Infects almost anything
  - Amoeba, insects, birds, reptiles, mammals
- 2 forms
  - Cutaneous ulceroglandular
  - Pneumonic
- Aerosol ID<sub>50</sub> = 10
- Developed as a biological weapon



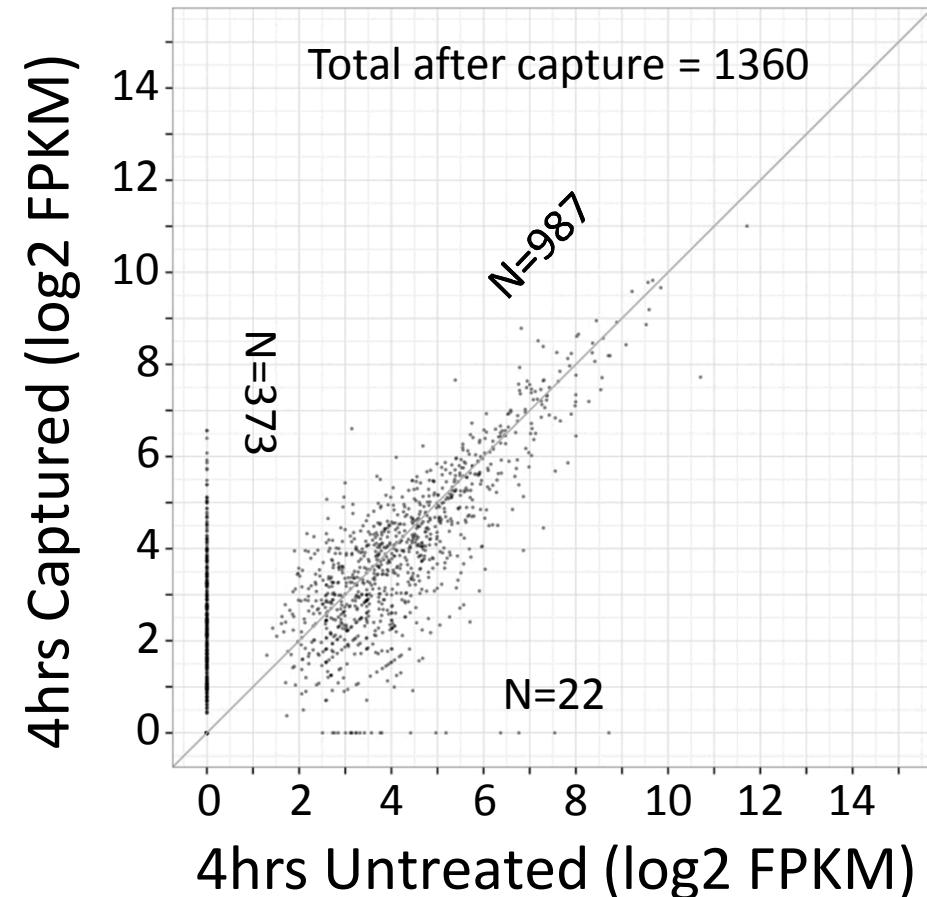
FIGURE 12.—Tularemia lesion following tickbite observed in soldier at Army-Navy General Hospital, Hot Springs, Ark.

# Capture Increases *F. tularensis* Reads

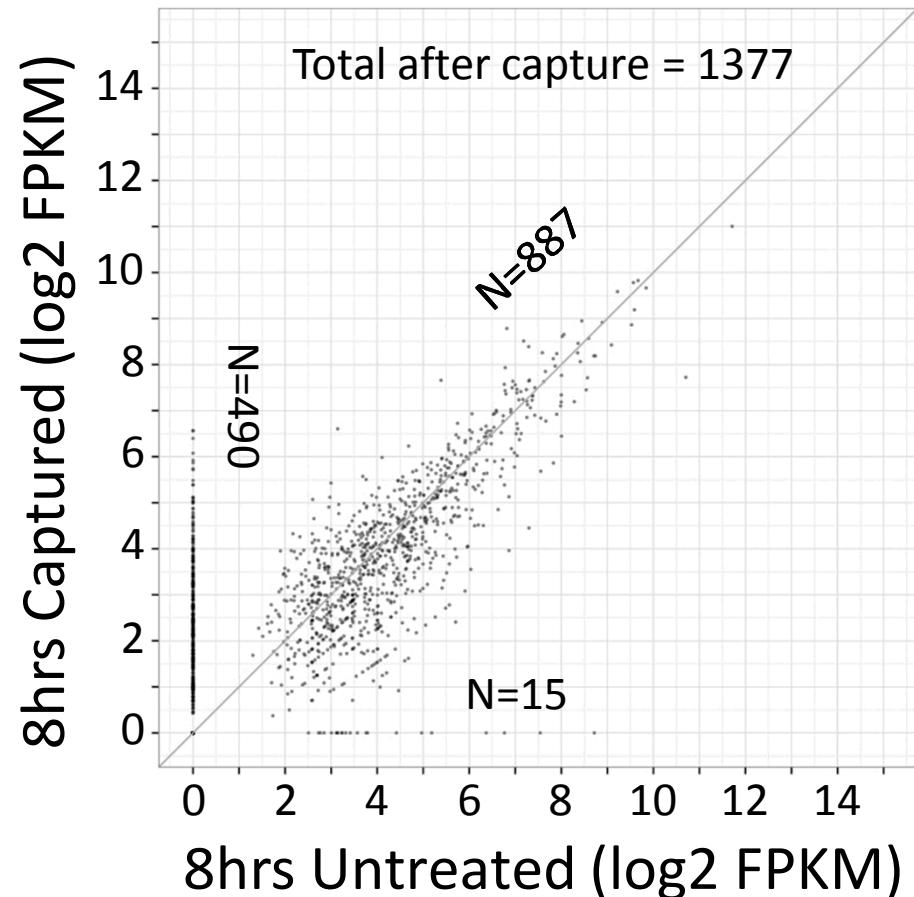


# Capture Leads to Unbiased Enrichment of Bacterial Transcripts

4hrs Post-Infection

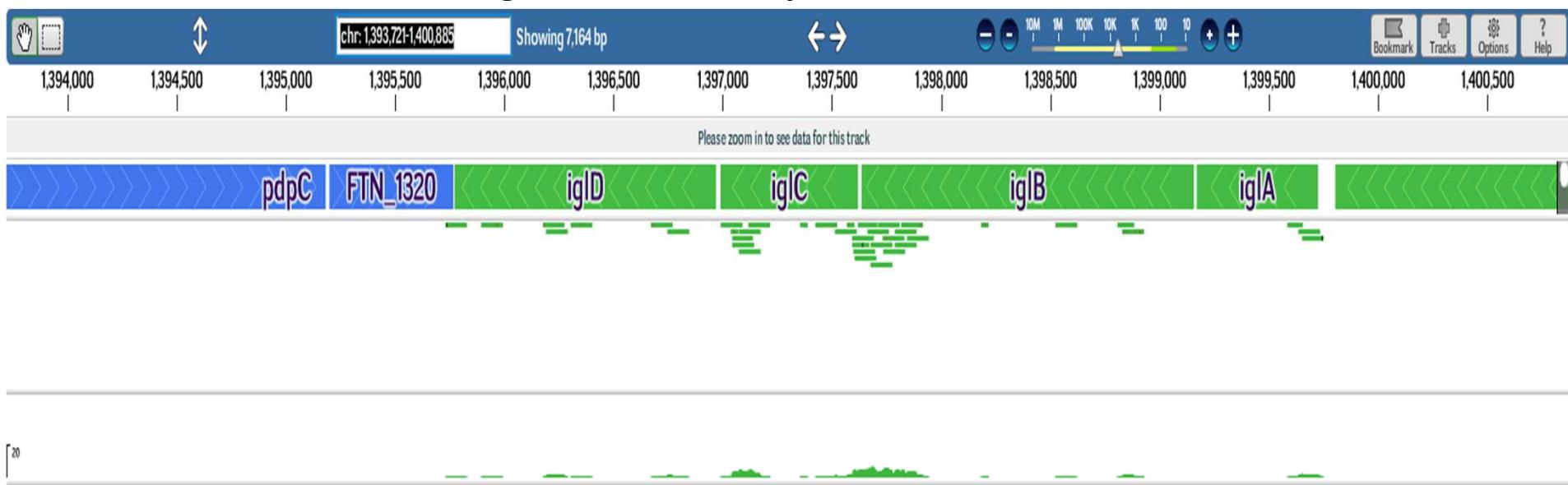


8hrs Post-Infection

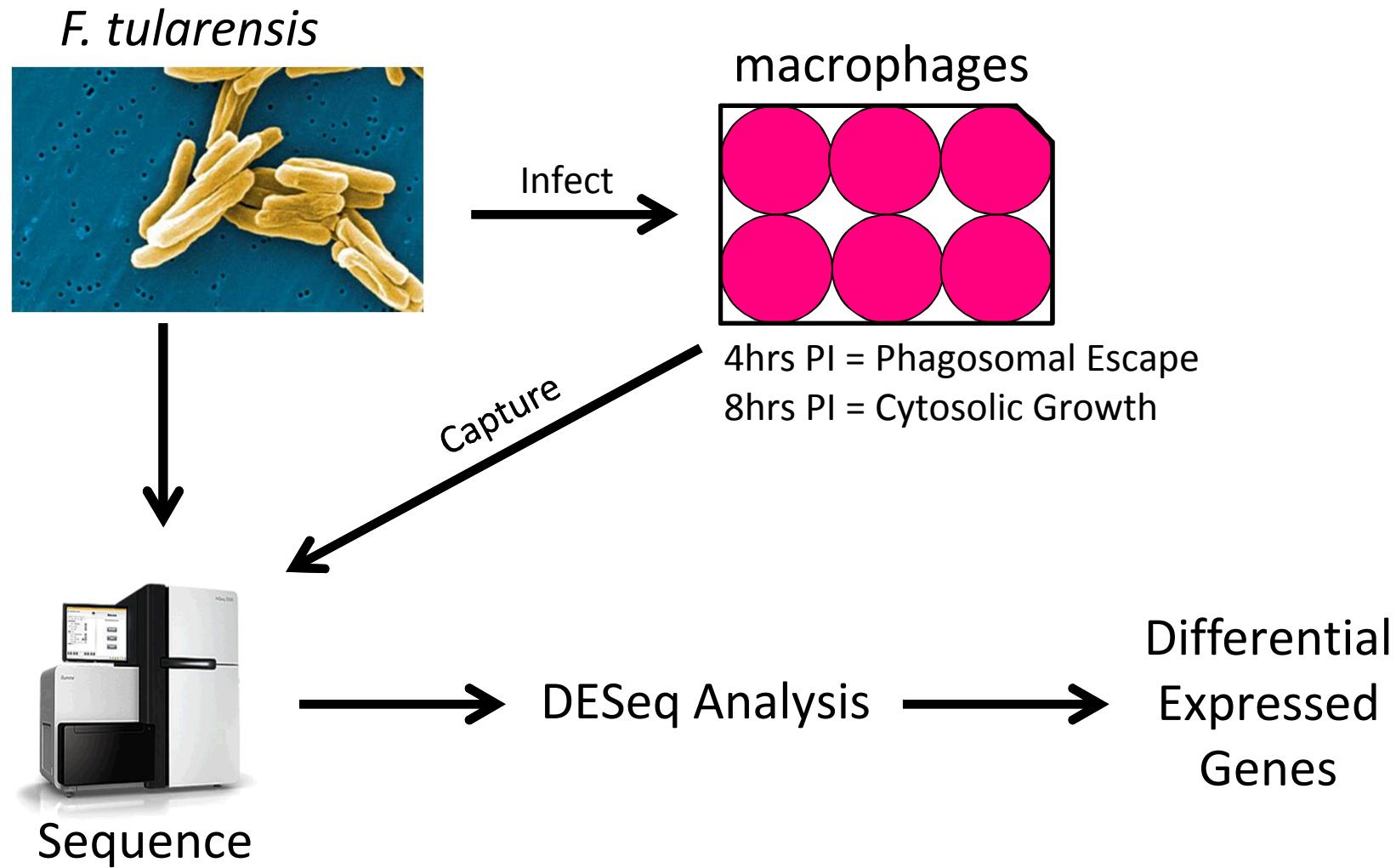


# Capture Improves *F. tularensis* Mapping

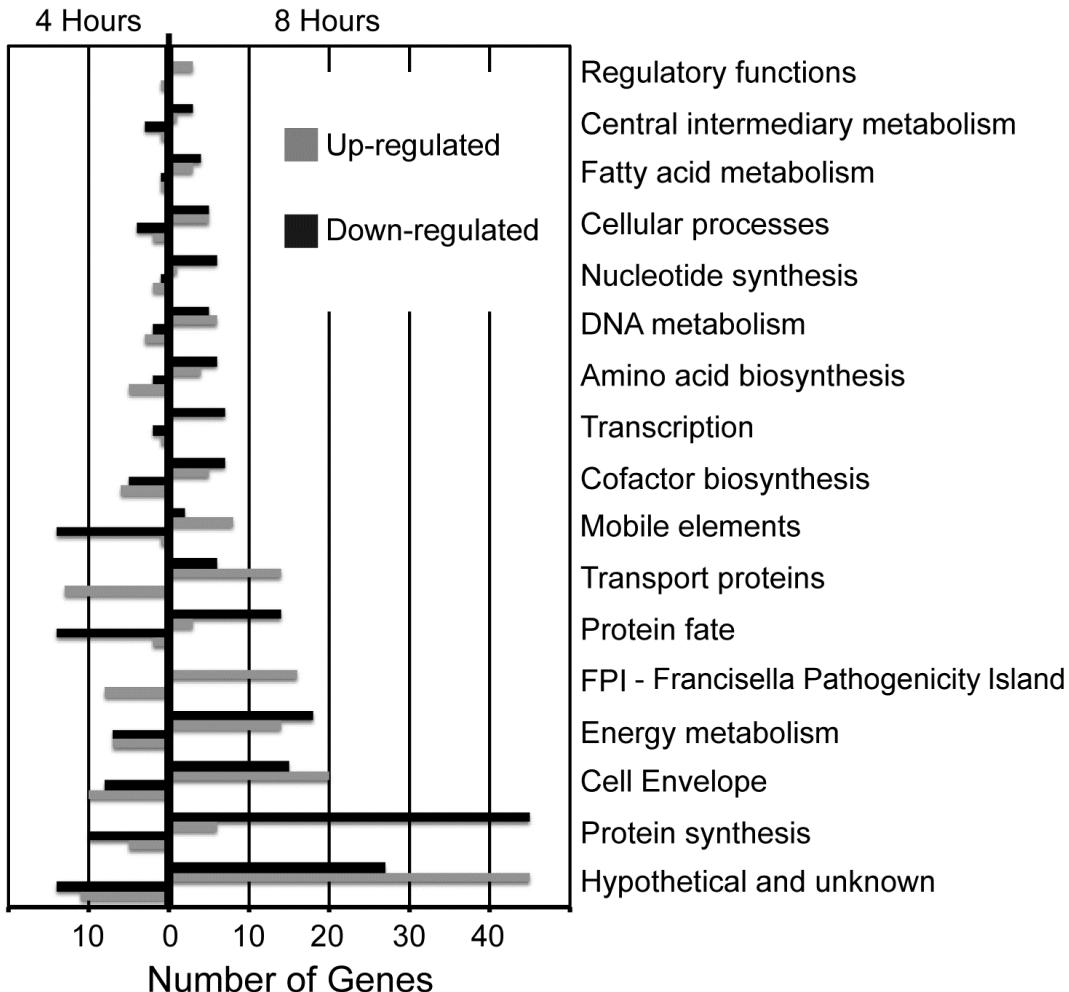
## *igl* locus before capture – 2 million reads



# Differential *F. tularensis* Gene Expression during Phagosomal Escape and Cytosolic Growth



# Differentially Expressed Genes at 4 and 8hrs Post Infection



# Most Highly Up-Regulated Genes Have Unknown Function

4 hours Post Infection			8 hours Post Infection		
Gene ID	Name/Function	4hrs fold change	Gene ID	Name/Function	8hrs fold change
FTL0721	DedA family protein	9.69	FTL0815	PRC-barrel protein	28.51
FTL1213	Unknown	8.39	FTL1402	ISFtu1 transposase	13.91
FTL1216	Unknown	8.16	FTL0953	methyltransferase	12.72
FTL1876	Outer membrane protein	8.14	FTL0814	PRC-barrel protein	12.03
FTL1402	ISFtu1 transposase	7.86	FTL0924	Oligopeptide transporter	10.90
FTL1509	Carboxypeptidase	7.82	FTL1219	aminotransferase	8.76
FTL0765	<i>vacJ</i> - lipoprotein	7.70	FTL0881	Unknown	8.63
FTL0731	YhhQ family protein	6.69	FTL1957	Heat Shock	8.43
FTL0814	Unknown	6.15	FTL0816	Unknown	8.18
FTL0700	<i>comL</i> - competence	5.80	FTL0123	Short chain dehydrogenase	8.13

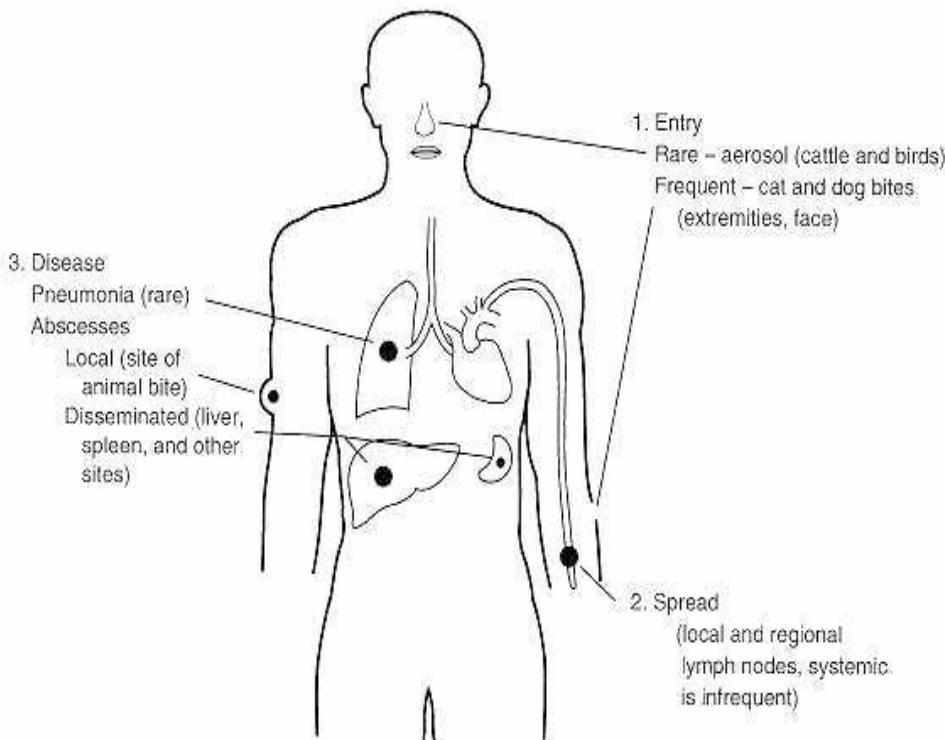
# *F. tularensis* Conclusions

- Phagosomal escape and cytosolic growth have significantly different gene expression profiles
- Many genes up-regulated during infection currently have unknown functions
- These proteins, especially the OM proteins, make promising targets for vaccines or therapeutic intervention

# The Diseases of *Yersinia*

## Plague

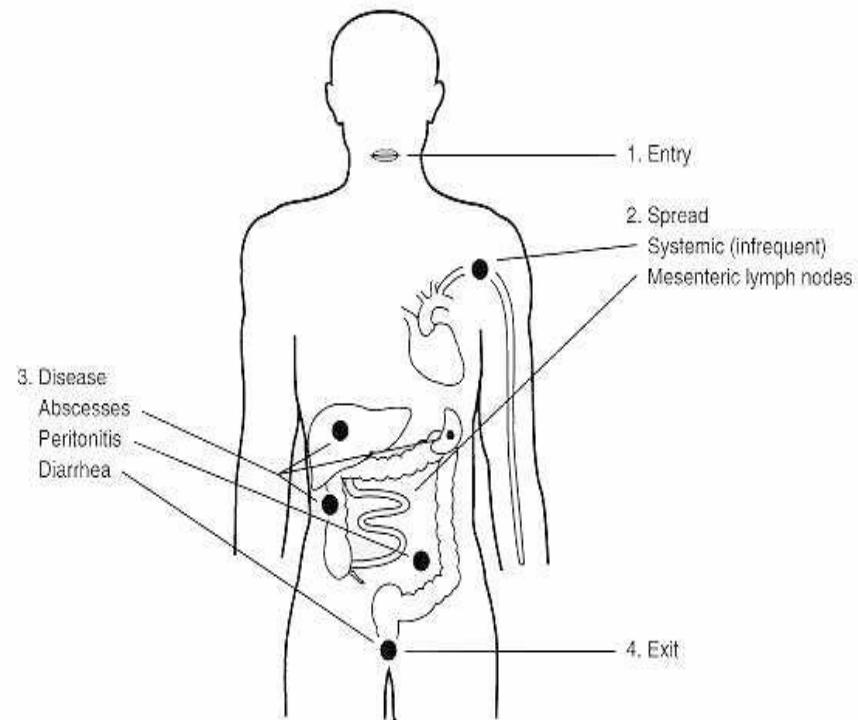
*Y. pestis*



## Gastrointestinal syndromes

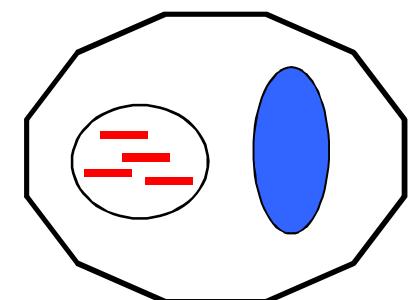
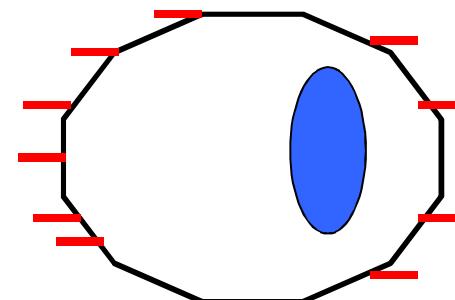
*Y. pseudotuberculosis*

*Y. enterocolitica*



All *Yersinia* are typically considered extracellular pathogens

# *In Vitro Y. enterocolitica* Infections

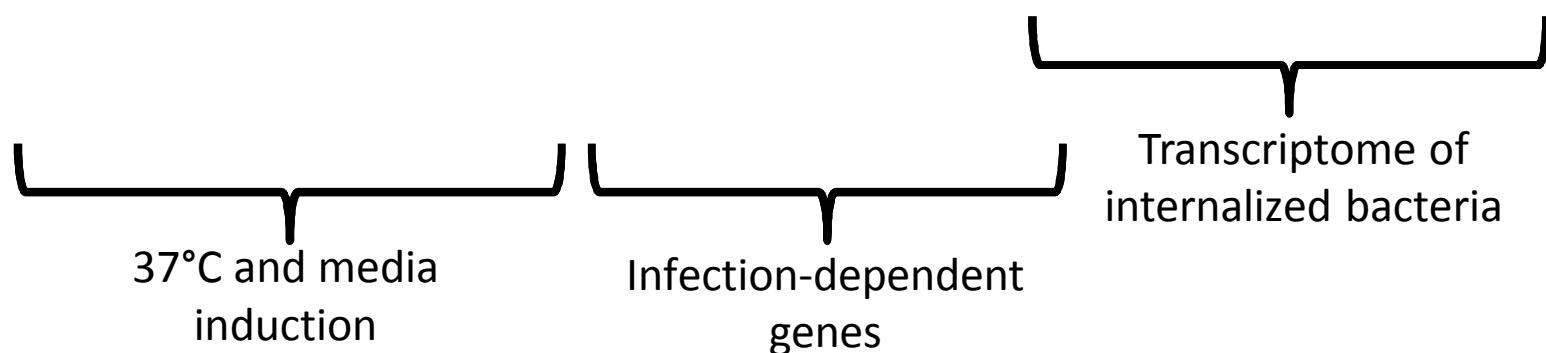


Growth Medium  
(LB w/o NaCl)  
26°C

Conditioned RPMI  
Filter sterilized  
5% CO<sub>2</sub>  
37°C

Infection of MΦ  
MOI = 10  
37°C

Gentamicin Treated  
MOI = 10  
1hr of infection  
1hr of gentamicin



# *Y. enterocolitica* Remains Viable within MΦ

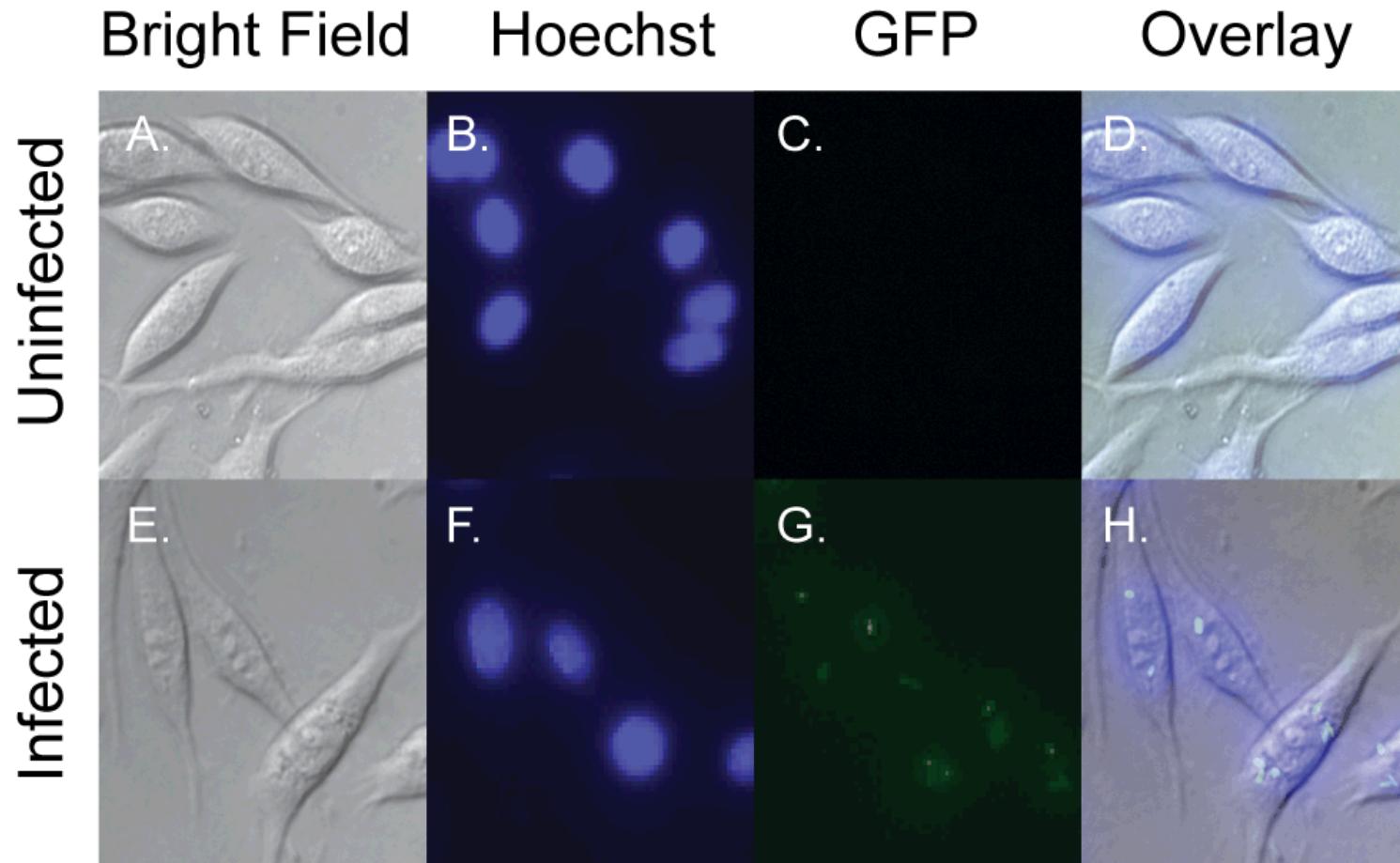


Plate counts show 19% of inoculum is viable after 3 hours within macrophage

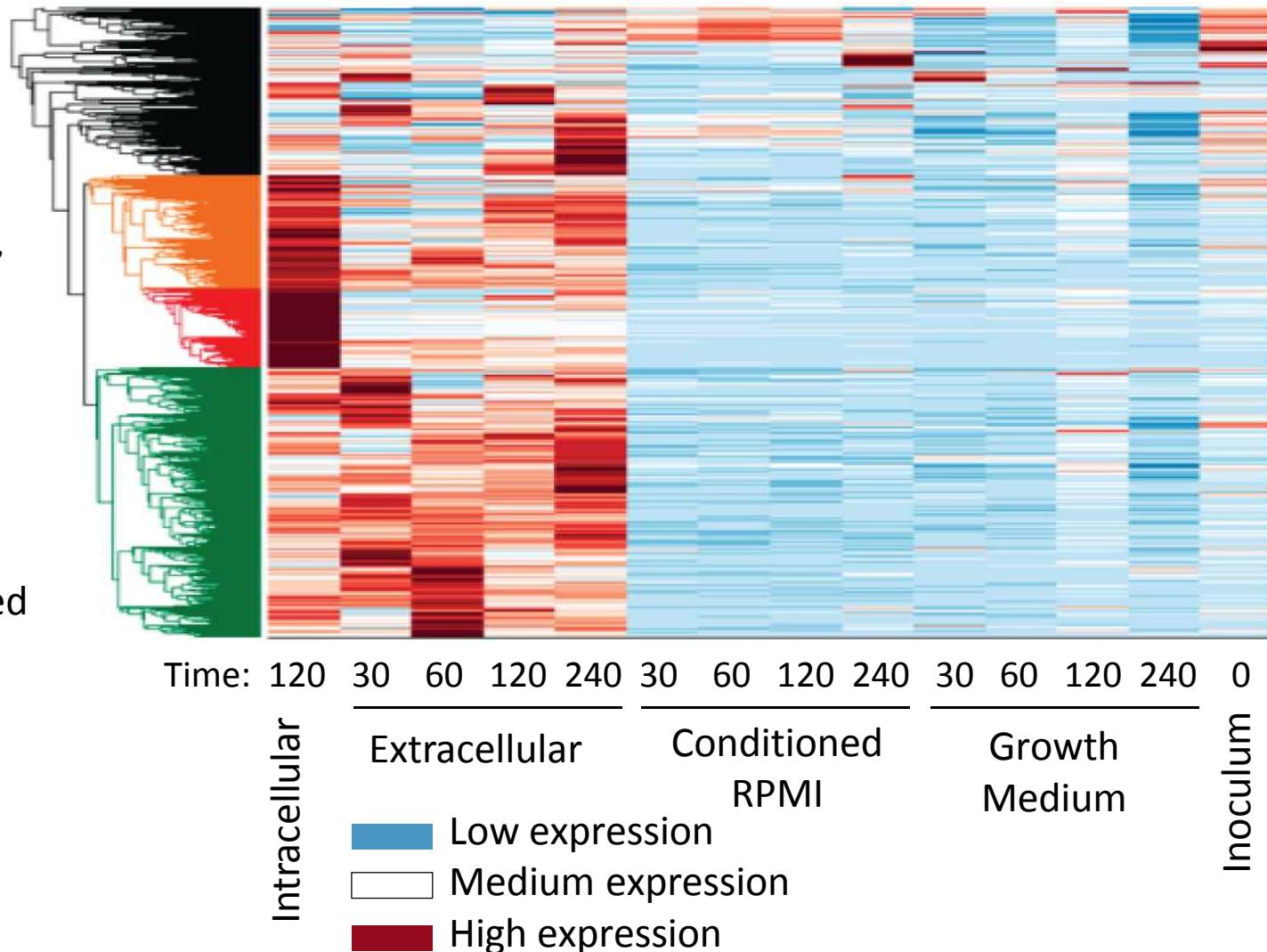
# Gene Expression during Infection is Temporally Dynamic

**Black:** Low expression all conditions

**Orange:** Highly expressed intracellular, increased extracellular expression over time

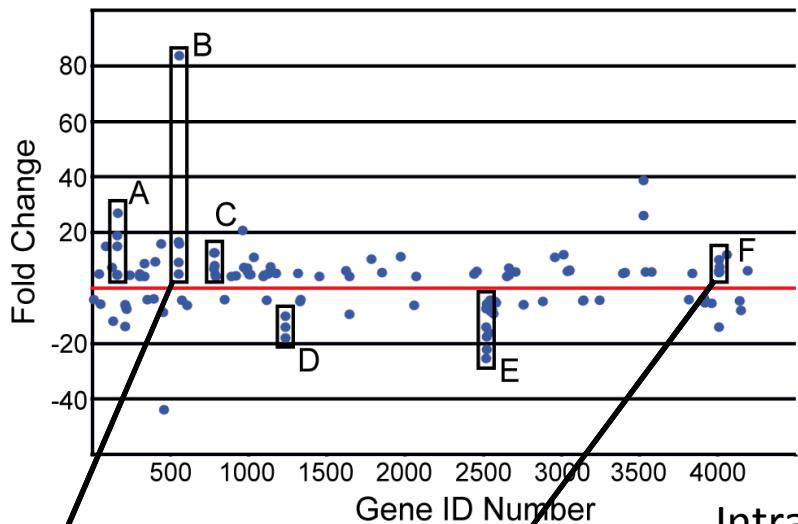
**Red:** Highly expressed intracellular

**Green:** Highly expressed extracellular and intracellular



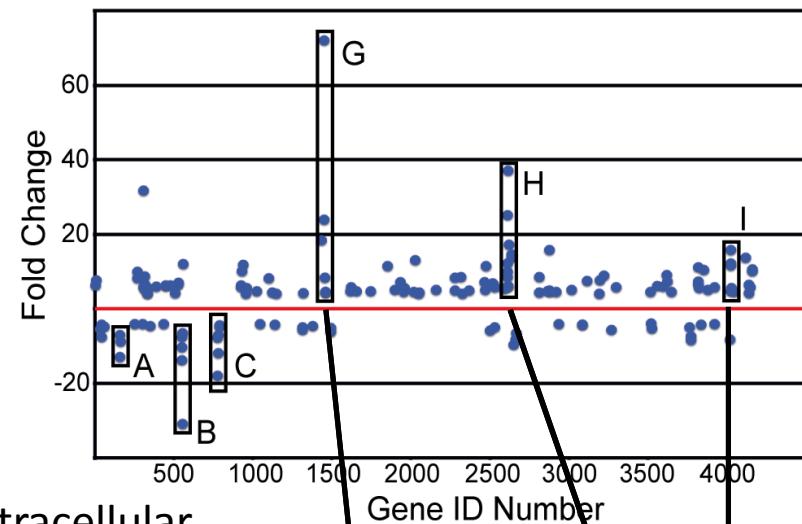
# Differential Expression Analysis Reveals Clusters of Related Genes

Conditioned RPMI vs Growth Medium



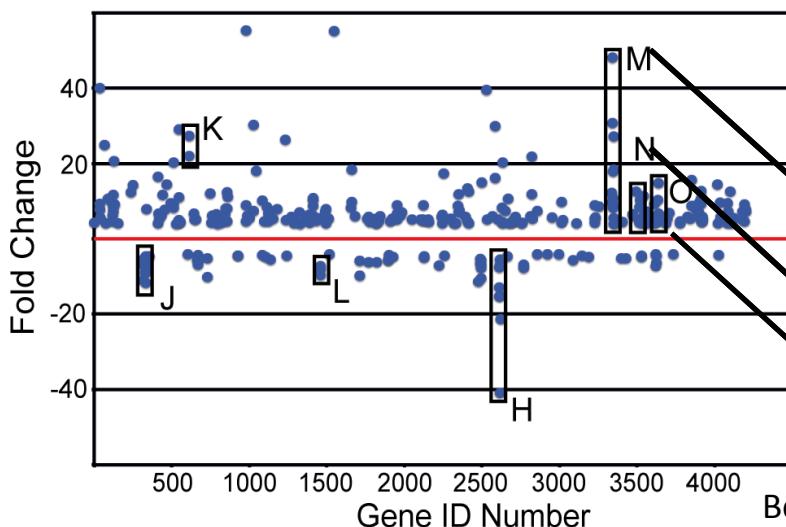
B) Sucrose uptake  
F) Glycogen metabolism

Extracellular vs Conditioned RPMI



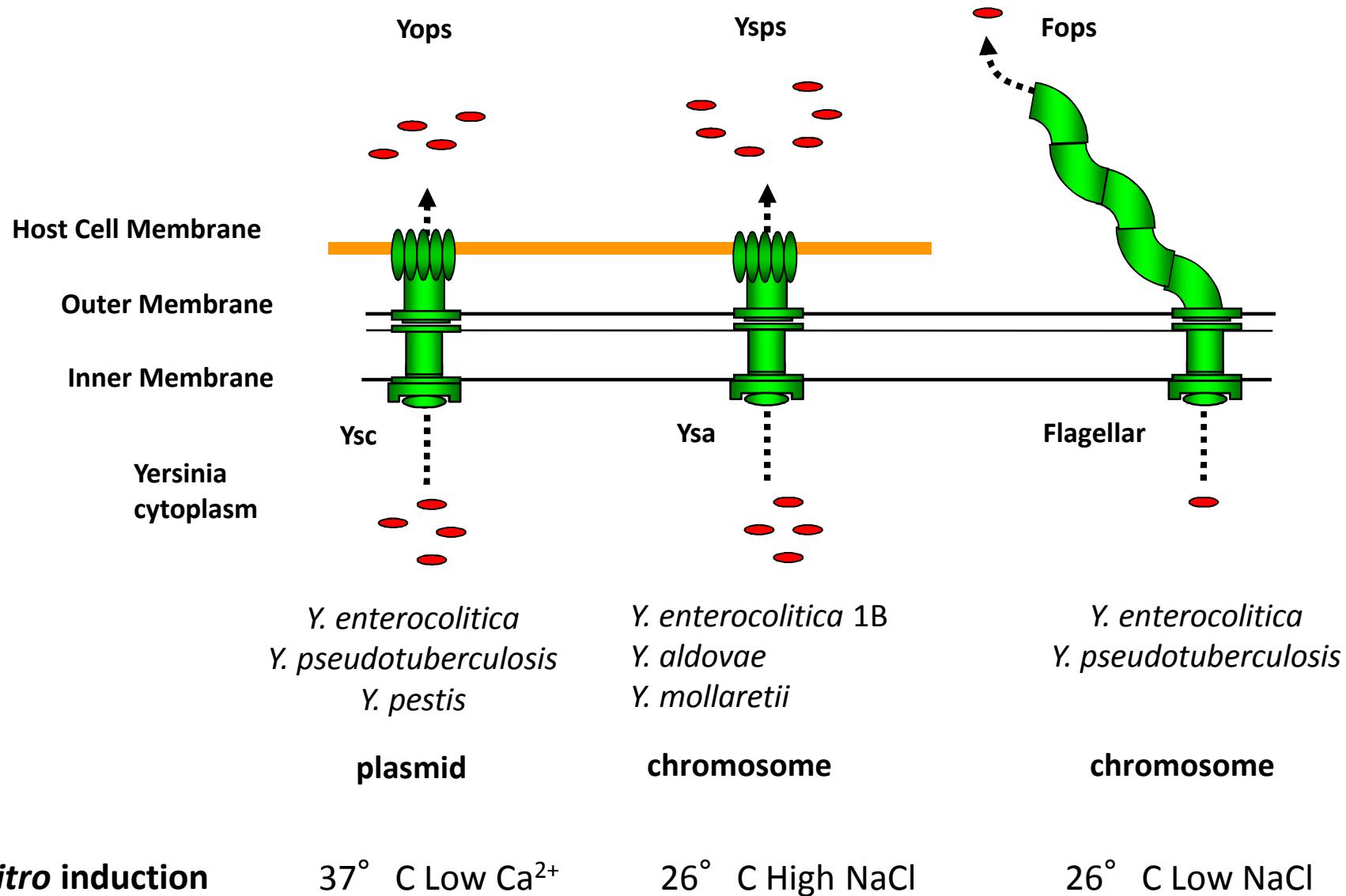
G) Myf pilus  
H) Yersiniabactin  
I) Inositol metabolism

Intracellular vs Extracellular

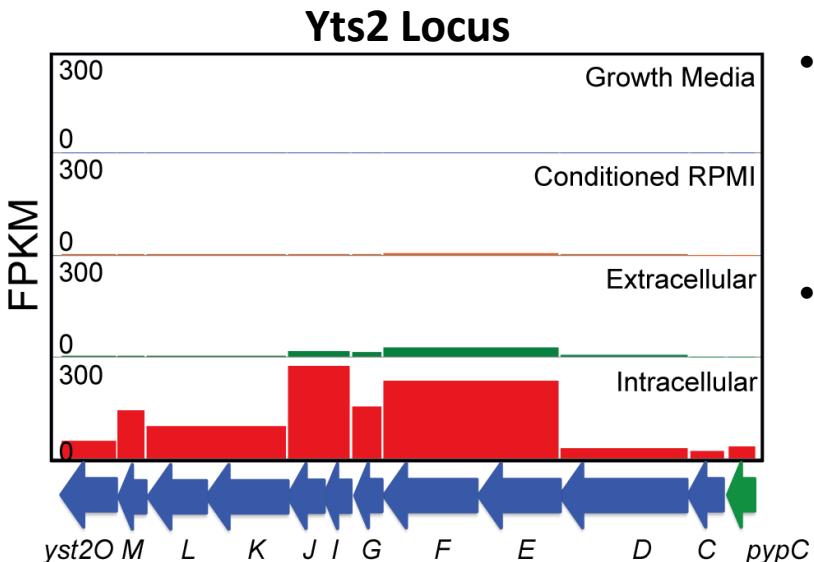
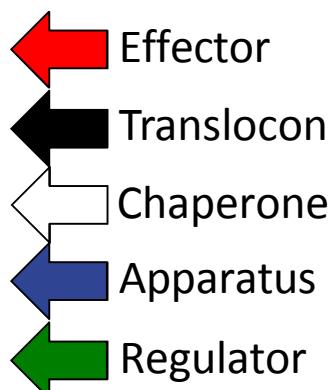
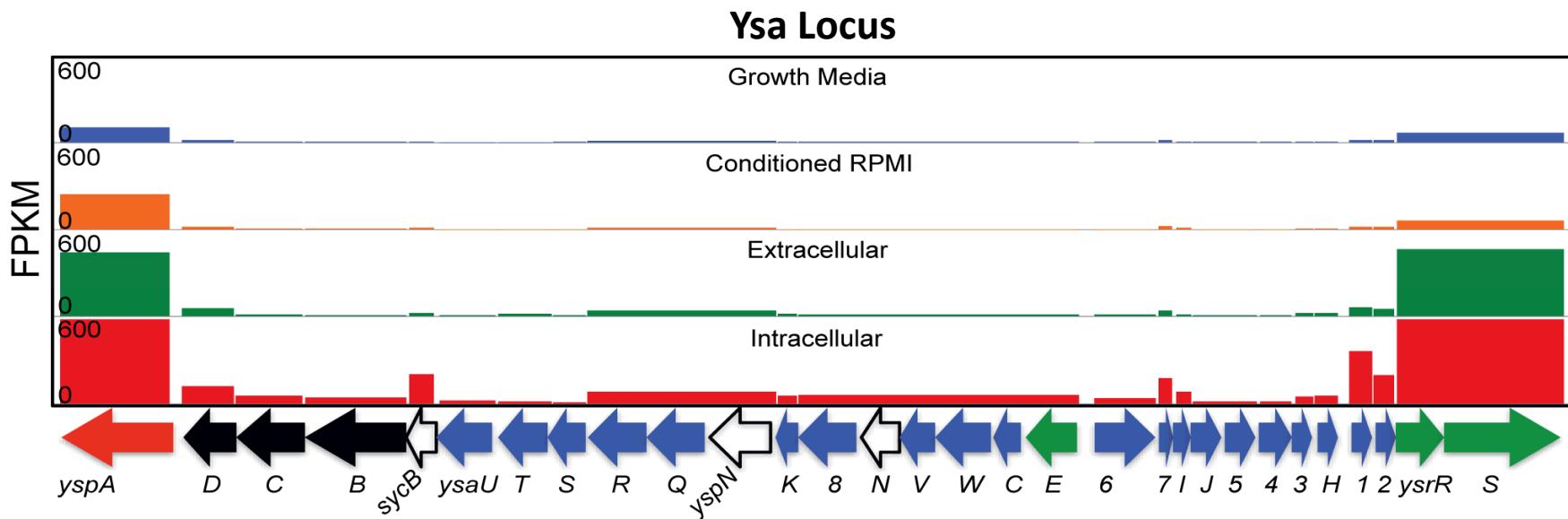


M) Yts2 T2SS  
N) Ysa T3SS  
O) Tad Pilus

# *Y. enterocolitica* Type III Secretion

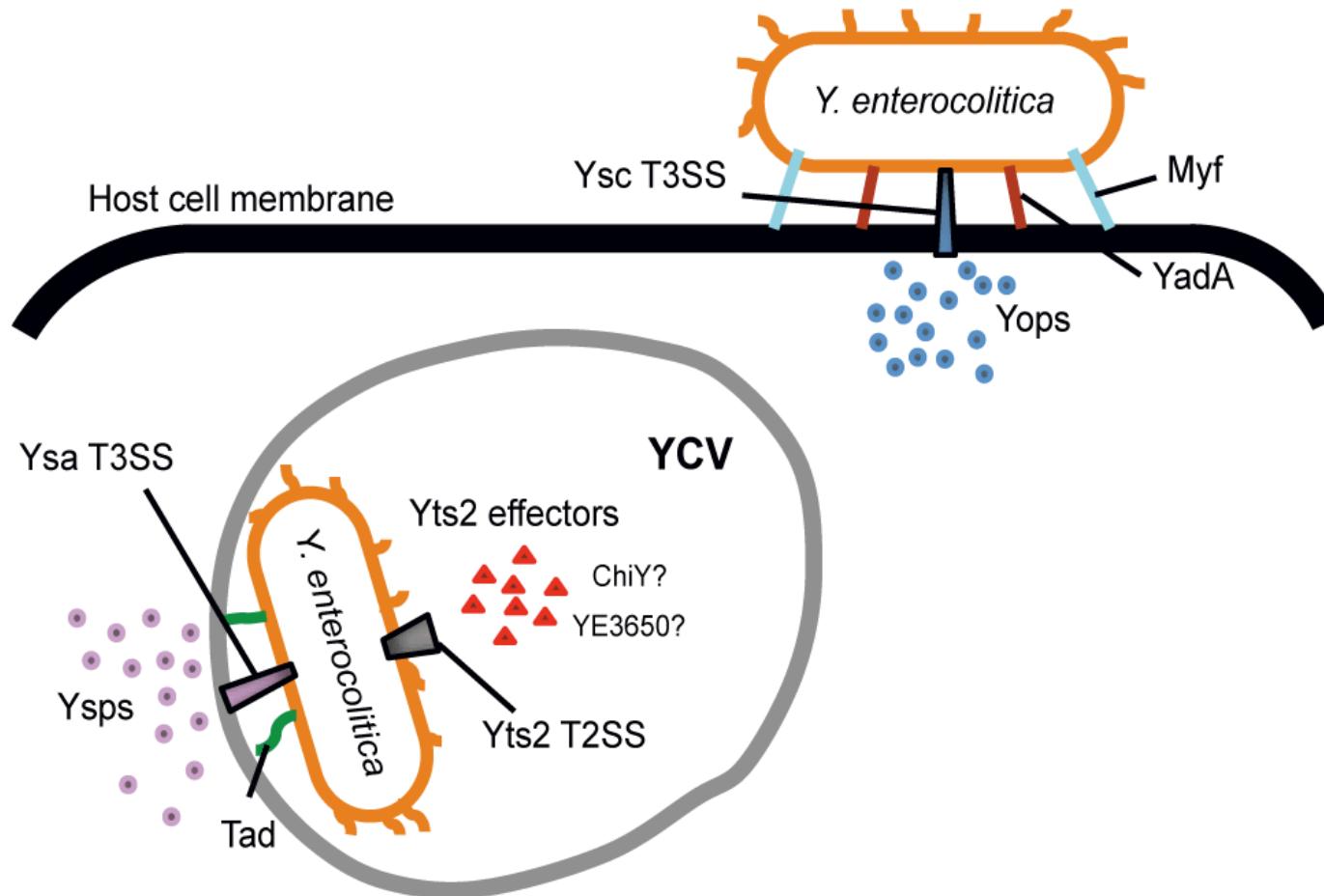


# The Ysa T3SS and Yts2 T2SS are Expressed Intracellularly



- First evidence Ysa T3SS is expressed intracellularly in mammalian cells
- First example of native expression of the Yts2 T2SS

# A New Model of *Y. enterocolitica* Intracellular Infection



Rank	30min	60min	120min	240min
1	YE0418	YE3012	YE0418	YE0418
2	YE3012	YE0418	YE1765	YE3682
3	YE3682	YE3682	YE3012	YPE0066
4	YE1636	YE0889	YE3682	YE1765
5	YE2165	YE1913	YE2165	YE1452
6	YE1913	YE1765	YPE0066	YE1913
7	YE1765	YE1636	YE1913	YE2165
8	YE0889	YE2165	YE1636	YE3821
9	YE0419	YE0419	YE0419	YE1636
10	YE1563	YPE0066	YE3926	YE0419
11	YE1914	YE1914	YE0064	YE0064
12	YE0064	YE3926	YE0889	YE1820
13	YE3915	YE0064	YE0402	YE3926
14	YE0437	YE1623	YE1914	YE3012
15	YE0300	YE1563	YE2590	YE1914
16	YE3926	YPE0058	YPE0040	YE3822
17	YE0613	YE1820	YE1820	YPE0095
18	YE0402	YPE0057	YE0065	YPE0040
19	YE1630	YPE0095	YE1631	YPE0057
20	YE1631	YPE0040	YE0613	YE0613
21	YPE0057	YE3996	YE3821	YE0437
22	YE1820	YE0300	YPE0095	YE2590
23	YPE0058	YE0402	YE0437	YE1631
24	YE1915	YE0613	YE1452	YE0101
25	YE0101	YE0437	YPE0057	YE0889

- **16 of top 25 most highly expressed genes are shared at each time point**
- 7 ribosomal proteins
- 2 cold shock proteins
- 1 ferridoxin
- 1 sigma 54 modulator
- 1 translation initiation factor
- 1 acyl carrier protein – fatty acid biosynthesis
- 1 plasmid maintenance toxin
- **2 outer membrane proteins**
  - Ail
  - Lpp

Despite its minor role in *Y. enterocolitica* virulence, *ail* is highly expressed during infection

Lpp has not been studied in *Y. enterocolitica*, but *Y. pestis* mutants show attenuated virulence

Given high levels of expression and localization to the outer membrane these may be promising therapeutic targets

# *Y. enterocolitica* Conclusions

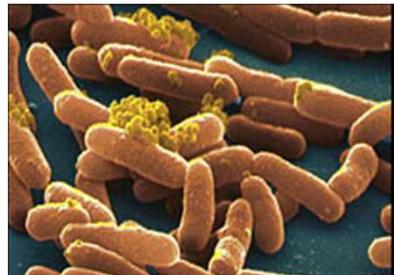
- Despite extracellular classification, *Y. enterocolitica* survives within macrophages
- Infection is a dynamic process, even *in vitro*
- Differential gene expression analysis reveals systems that are up or down regulated during different stages of infection
- Ysa T3SS, Yts2 T2SS, and Tad pilus are highly expressed after internalization

# *Salmonella* Typhimurium



- Gram-negative, facultative intracellular, gastrointestinal pathogen
- Most common cause of bacterial food-borne illness in U.S. < 1 million cases → 19,000 hospitalizations
- Frequent outbreaks in the U.S.
  - Most commonly associated with poultry
  - Other sources = peanut butter, melons, beef, exotic pets (frogs, turtles, hedgehogs)

# *In Vivo S. Typhimurium Infections*



*S. Typhimurium*

Infect →



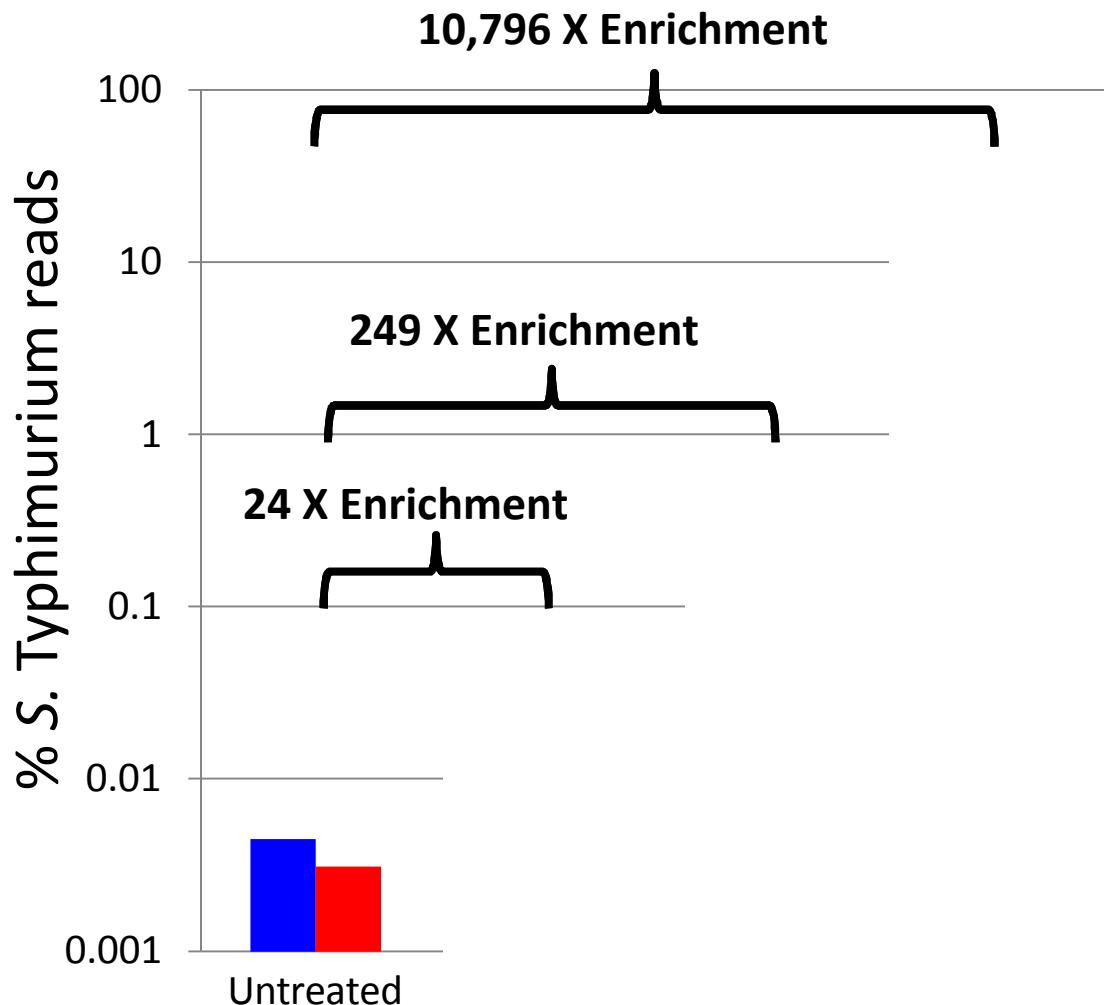
→ Liver (4X)  
→ Spleen (5X)

Livers  $2.5 \times 10^2 - 5.5 \times 10^3$  CFU/sample

Spleens  $1.3 \times 10^4 - 5.3 \times 10^5$  CFU/sample

~0.005% of reads map to *S. Typhimurium*

# *In Vivo* Enrichment



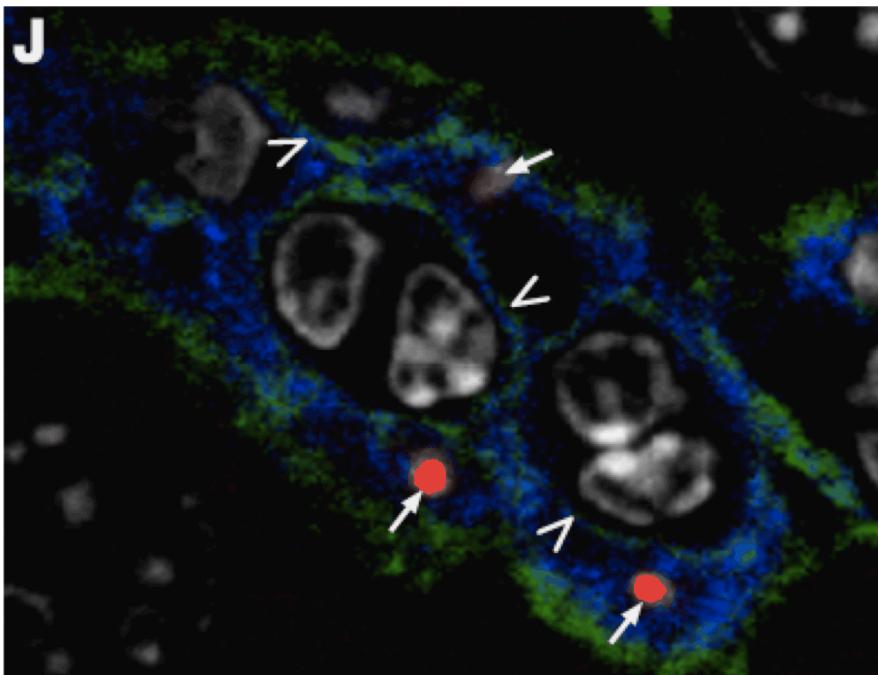
- Synergistic effect of combining Ribo-Zero with capture ~2X
- Despite use of Ribo-Zero, significant amount of bacterial rRNA is present

# Cell Type Specific Host-Pathogen Interactions

- Many bacterial species preferentially infect specific tissues or host cell types
- Understanding these preferences is critical to understanding pathogenesis
- *S. Typhimurium* chronic infection is associated with the presence of hemophagocytic macrophages
- This cell type may be important in controlling infection or may be a reservoir of bacteria
- Can we use RNA-Seq to understand this interaction?

# Hemophagocytic Macrophages

*S. Typhimurium* infects hemophagocytic macrophages



Confocal image

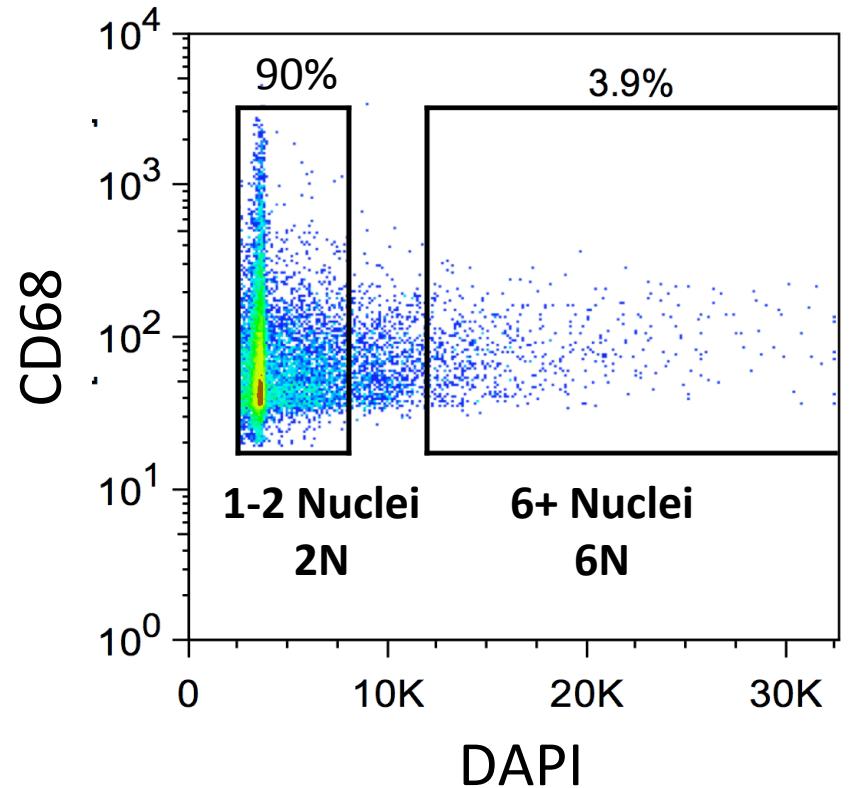
Macrophage

Actin

DNA

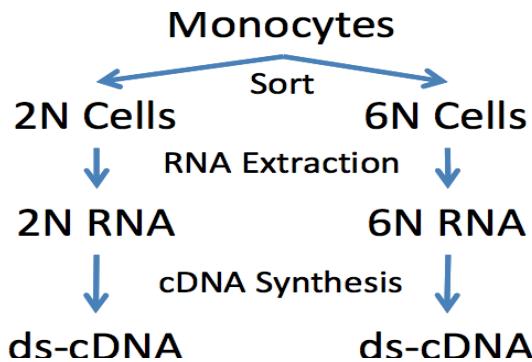
Salmonella

~4% of macrophages become hemophagocytes during chronic infection



- What is the role of the hemophagocytic macrophages during infection?
- What are the bacteria doing within this cell type?

# Host-Pathogen Interactions Using a Dual Transcriptomics Approach



Analysis is underway!

# *In Vivo* Conclusions

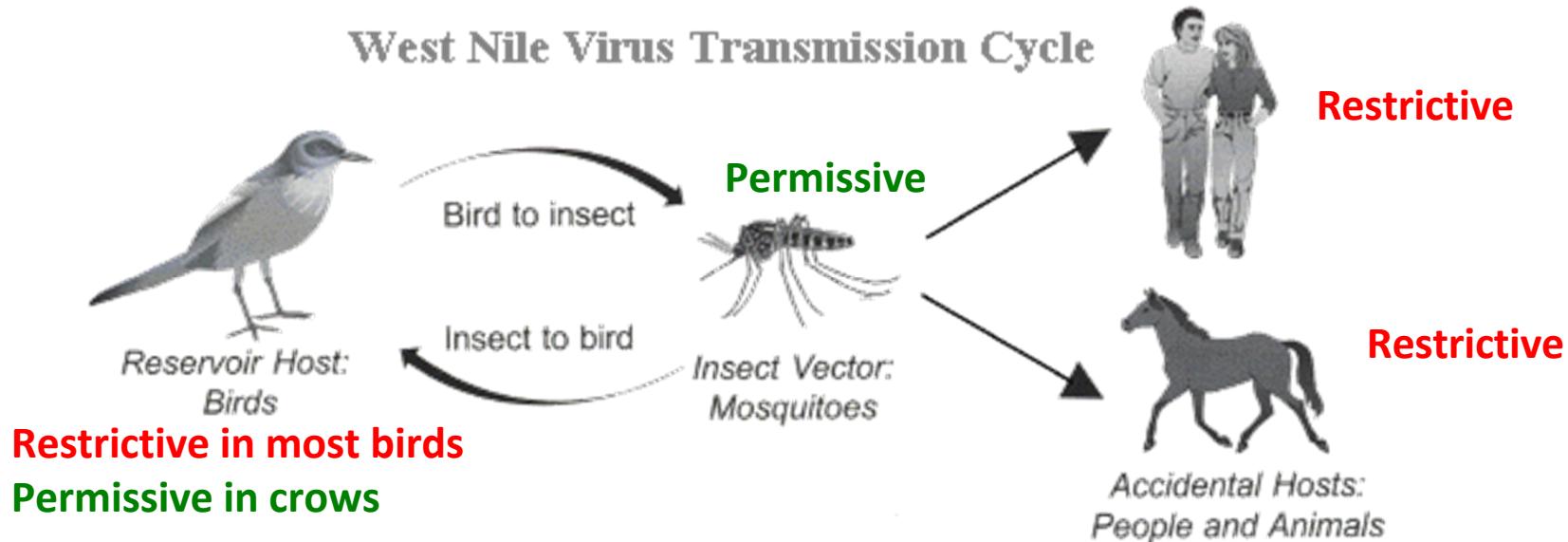
- Capture is effective at enriching for bacterial transcripts from infected tissues
  - More work necessary to deplete bacterial rRNA
- Host-pathogen transcriptomics in *S. Typhimurium* infected hemophagocytic macrophages will lead to improved understanding of chronic infection

# Future Directions

- Complete transcriptomic analysis of pathogen and host through the course of an infection
  - Tissue by tissue analysis from infection to death
    - Discover novel bacterial virulence mechanisms
    - Improved understanding of host response
    - Design better vaccines and therapeutics
- Comparative bacterial transcriptomics
  - What are the non-genetic factors that make strains more pathogenic?
  - *F. tularensis* LVS and SCHU S4

# Applications in Virology

- Viral quasi-species and bottlenecking effects
  - Dominant viral type changes from vector to host



# Summary

- Capture makes it possible to analyze bacterial transcriptomes during *in vitro* and *in vivo* infections
- *F. tularensis* has dynamically changing gene expression profiles at 4 and 8 hours post infection
- *Y. enterocolitica* uses the Ysa T3SS, Yts2 T2SS, and Tad pilus to promote intracellular survival
- Dual transcriptomic analysis of *S. Typhimurium* infected hemophagocytic macrophages will improve our understanding of chronic infection
- Capture and RNA-Seq are still in their infancy, there are many diverse and promising applications of the technology

# Acknowledgments

## Sandia-CA

Robert Meagher  
Steve Branda  
Anupama Sinha  
Kunal Poorey  
Deanna Curtis  
Mary Tran-Gyamfi  
Kelly Williams  
Victoria VanderNoot  
Todd Lane

## UC Davis

Glenn Young  
Karen LeGrand  
Rachelle Hamblin  
Walter Boyce  
Nam Tran

## University of Washington

Stan Langevin  
Michael Katze

## University of Colorado

Corrie Detweiler  
Eugenia Silva-Herzog  
Angelika Krivenko  
Dave Brazel  
Erin MacDonald

## UC Berkeley

Minyong Chung  
Vincent J. Coates Genomic Sequencing Lab

## Funding



National Institute  
of Allergy and  
Infectious Diseases



**Sandia**  
**National**  
**Laboratories**

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-



Questions?

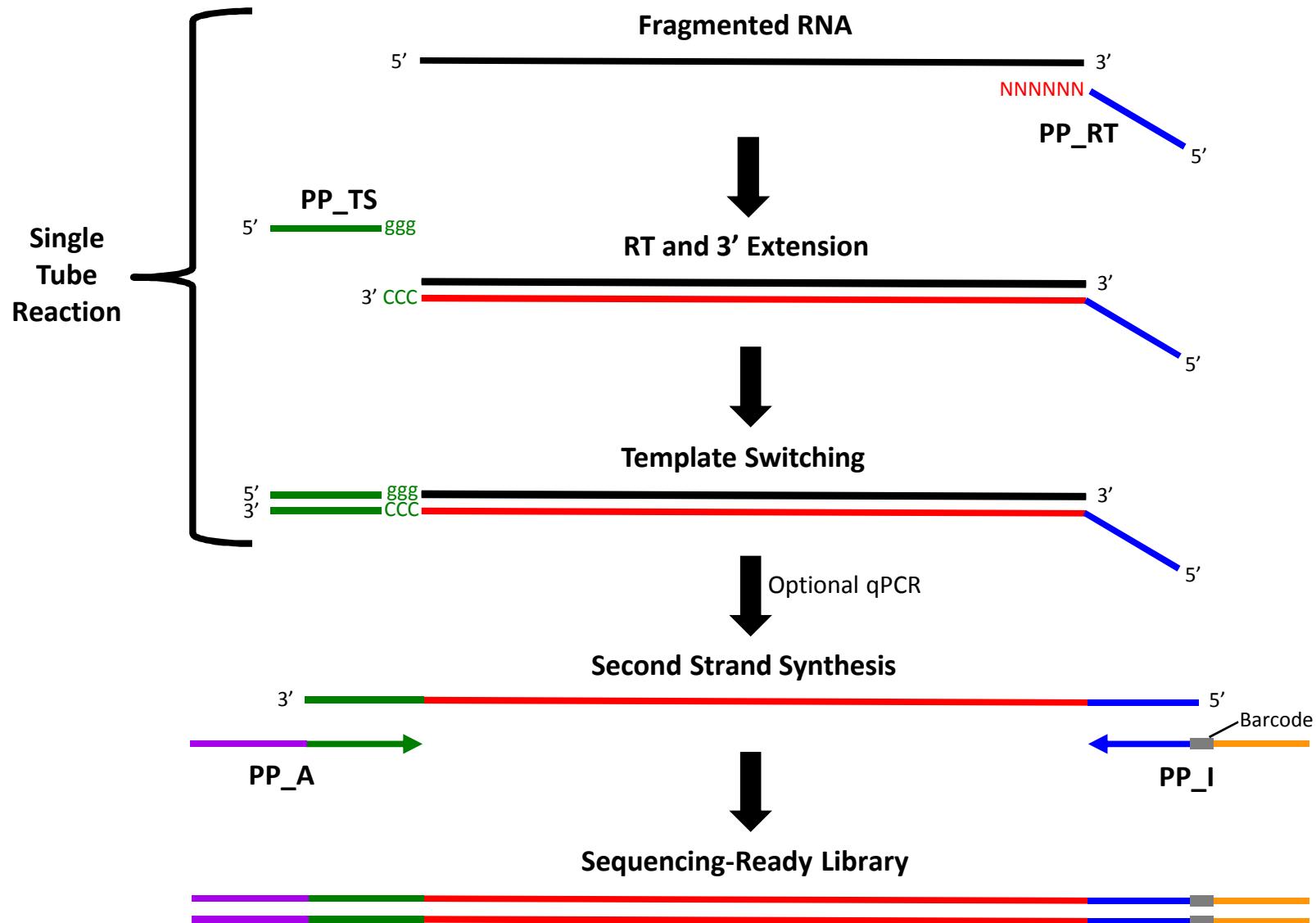
# Bonus Slides

# RNA-Seq

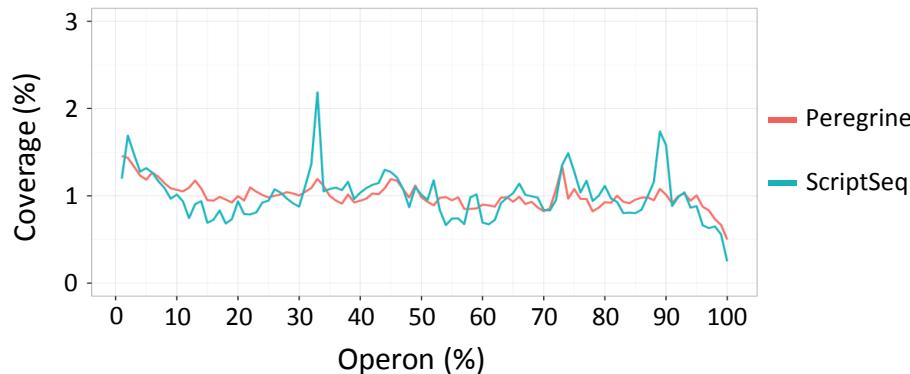
- Sequencing all RNA transcripts in a sample
  - mRNA, miRNA, lncRNA, snoRNA, etc.
- Compared to microarrays
  - Better sensitivity, better dynamic range, discovery of non-coding RNA and splice variants, cheaper
- Initial drawback of sample preparation
  - All RNA must first be converted to cDNA
  - First generation kits were biased, time consuming, and expensive

# Creating a New RNA-Seq Library Prep

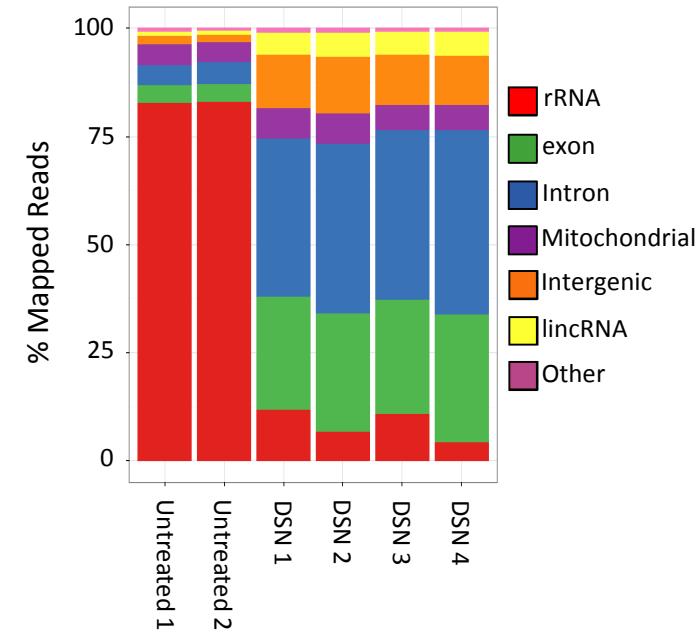
- Unbiased
  - Results accurately reflects presence and abundance
- Flexible
  - Create libraries from any sample and RNA type
- Strand-Specific
  - Vital to discovery of anti-sense RNAs
- Compatible with enrichment techniques
  - rRNA removal (DSN/HAC) and capture
- Rapid
  - Older protocols can take days
- Cost-Effective
  - Commercial kits can cost >\$75/sample



## Unbiased Coverage



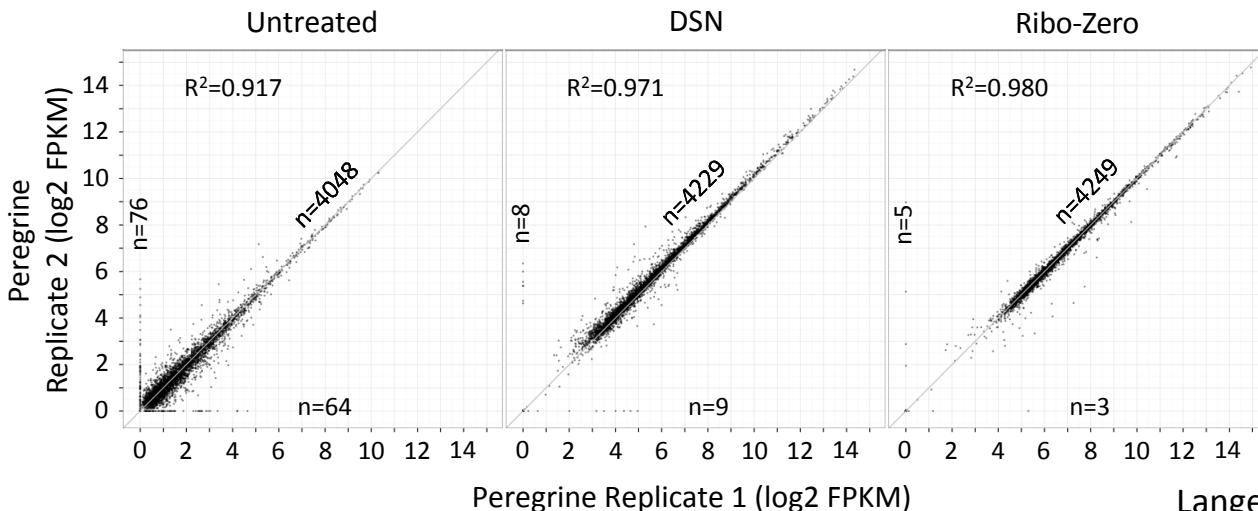
## Compatibility with Enrichment



## Strand specificity

Prep & treatment	rRNA	CDS	CDS w/known as RNA
Peregrine untreated	99.77 ± 0.01	94.77 ± 0.04	86.97 ± 0.35
Peregrine DSN	99.86 ± 0.01	94.20 ± 0.23	87.17 ± 0.19
Peregrine Ribo-Zero	2.21 ± 0.09	87.15 ± 1.28	81.72 ± 1.51
ScriptSeq untreated	99.53 ± 0.10	97.85 ± 0.19	94.04 ± 1.24
ScriptSeq Ribo-Zero	22.33 ± 4.27	97.87 ± 0.30	93.11 ± 0.92

## Reproducibility



## Rapid and Cost-Effective

Total Time = < 5.5 hrs  
 Hands on Time = ~1.5 hrs  
 Cost = < \$5/sample

# Commercial Adoption

April 2013

TECHNICAL PAPER

RNA Biology 10:4, 502-515; April 2013; © 2013 Landes Bioscience

## Peregrine

A rapid and unbiased method to produce strand-specific RNA-Seq libraries from small quantities of starting material

Stanley A. Langevin,<sup>1,†</sup> Zachary W. Bent,<sup>1,†</sup> Owen D. Solberg,<sup>1</sup> Deanna J. Curtis,<sup>1</sup> Pamela D. Lane,<sup>1</sup> Kelly P. Williams,<sup>1</sup> Joseph S. Schoeniger,<sup>1</sup> Anupama Sinha,<sup>1</sup> Todd W. Lane<sup>1</sup> and Steven S. Branda<sup>2,\*</sup>

<sup>1</sup>Systems Biology; Sandia National Laboratories; Livermore, CA USA; <sup>2</sup>Biotechnology and Bioengineering; Sandia National Laboratories; Livermore, CA USA

October 2013

### SMARTer Stranded RNA-Seq Kit User Manual

The SMARTer Stranded RNA-Seq Kit starts with less than nanogram amounts of RNA. A modified N6 primer (the SMART Stranded N6 Primer) primes the first-strand synthesis reaction (Figure 2). For added simplicity, the RNA is chemically fragmented during denaturation.

**NOTE:** If your sample is degraded or of low quality, see Appendix A for a fragmentation-free protocol.

When SMARTScribe™ Reverse Transcriptase reaches the 5' end of the RNA fragment, the enzyme's terminal transferase activity adds a few additional nucleotides to the 3' end of the cDNA. The carefully-designed SMARTer Stranded Oligo base-pairs with the non-template nucleotide stretch, creating an extended template to enable SMARTScribe RT continue replicating to the end of the oligonucleotide (Chenchik *et al.*, 1998). The resulting full-length, single-stranded (ss) cDNA contains the complete 5' end of the mRNA, as well as sequences that are complementary to the SMARTer Stranded Oligo.

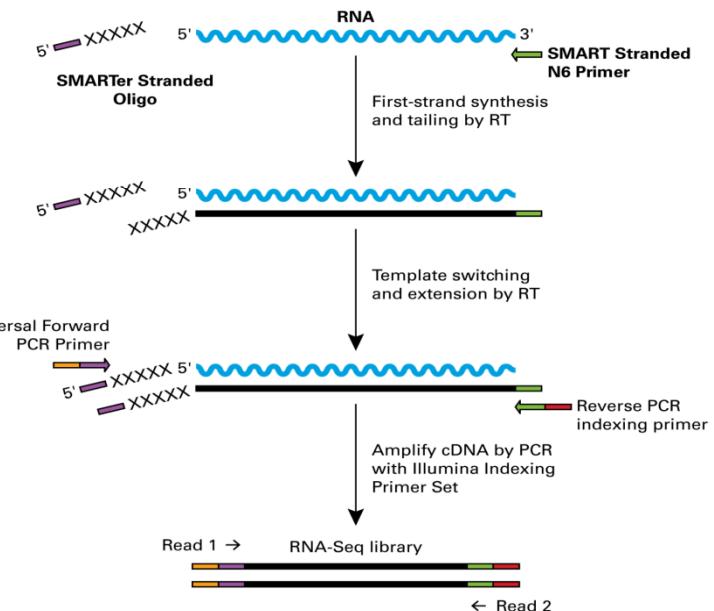
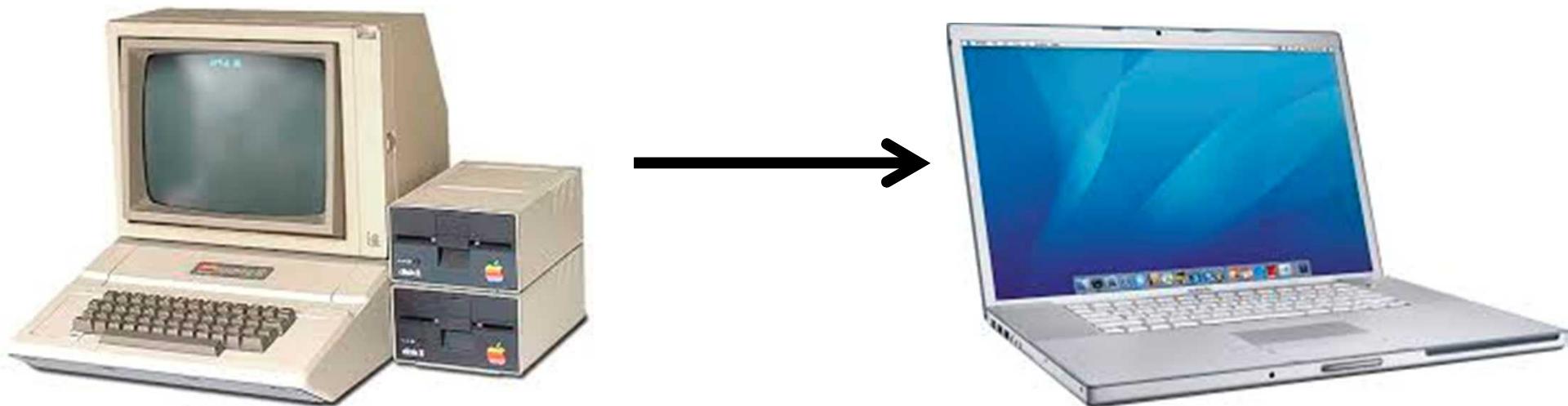
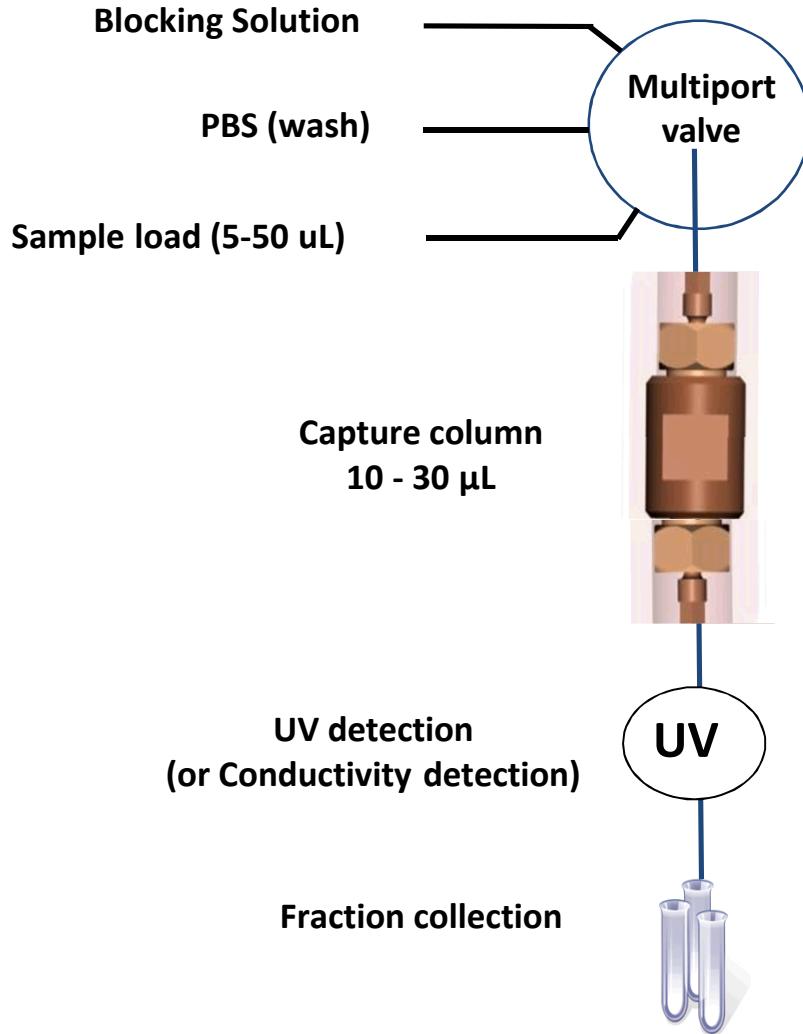


Figure 2. Flowchart of SMARTer Stranded RNA-Seq library generation.

# The Evolution of “The Device”



# Meso-fluidic Capture



- True chromatography approach
- Less manipulation
- Less hands on time

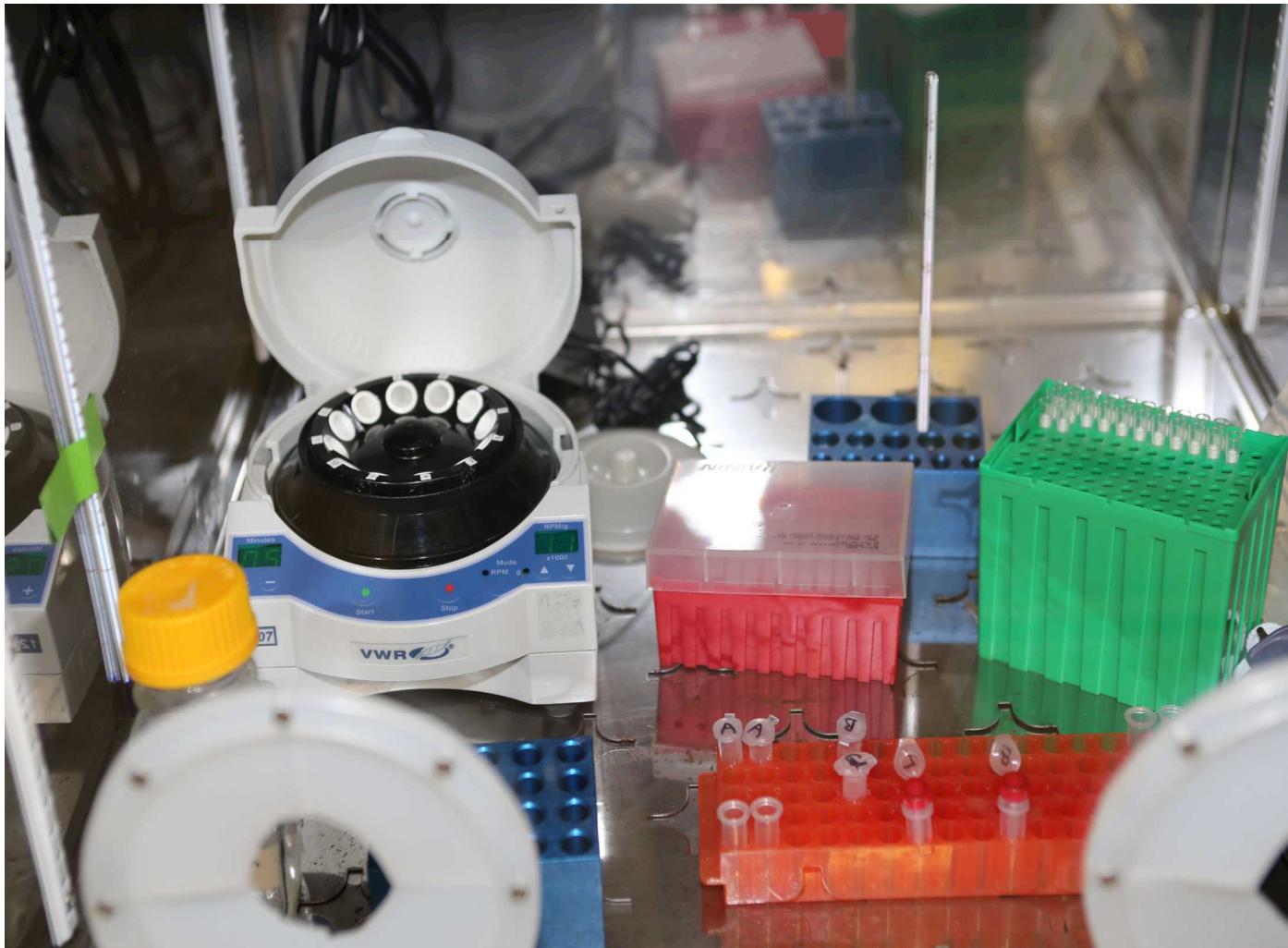
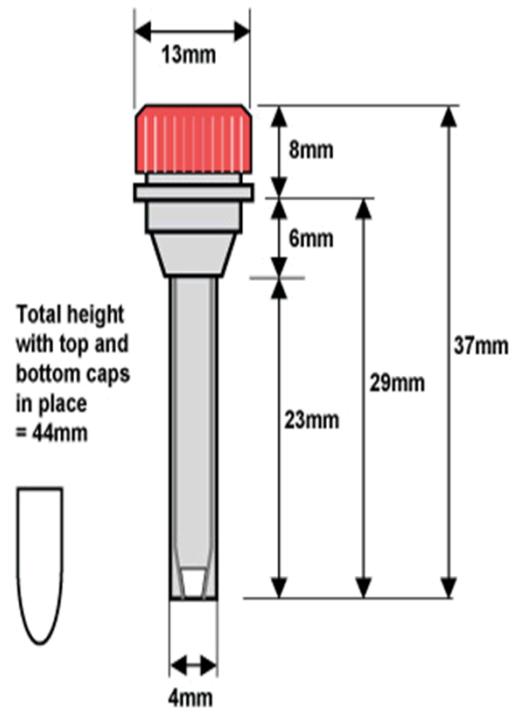
# Version 1 – Syringe Pumps



# Version 2 – Spin Columns

Thermo Scientific Pierce  
Micro-Spin Columns  
(Part No. 89879)

Total column capacity = 0.4mL  
(resin bed = 0.1mL; reservoir = 0.3mL)

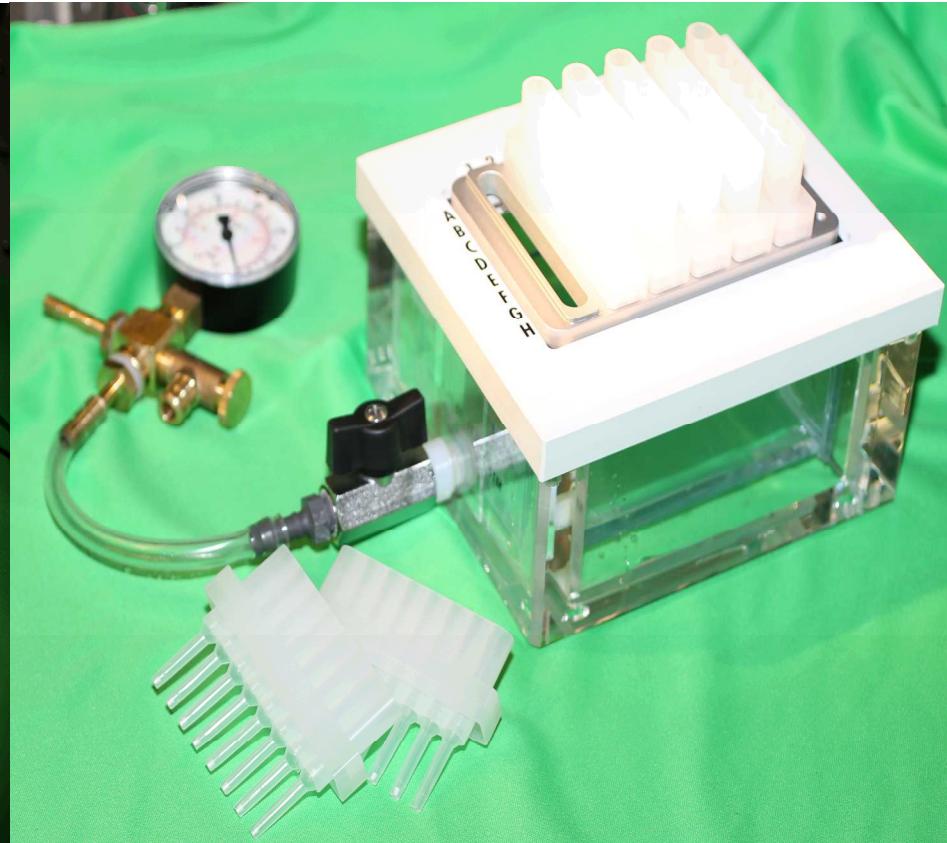


# Version 3 – Vacuum Driven 48-plex Capture

## 8-Well strips



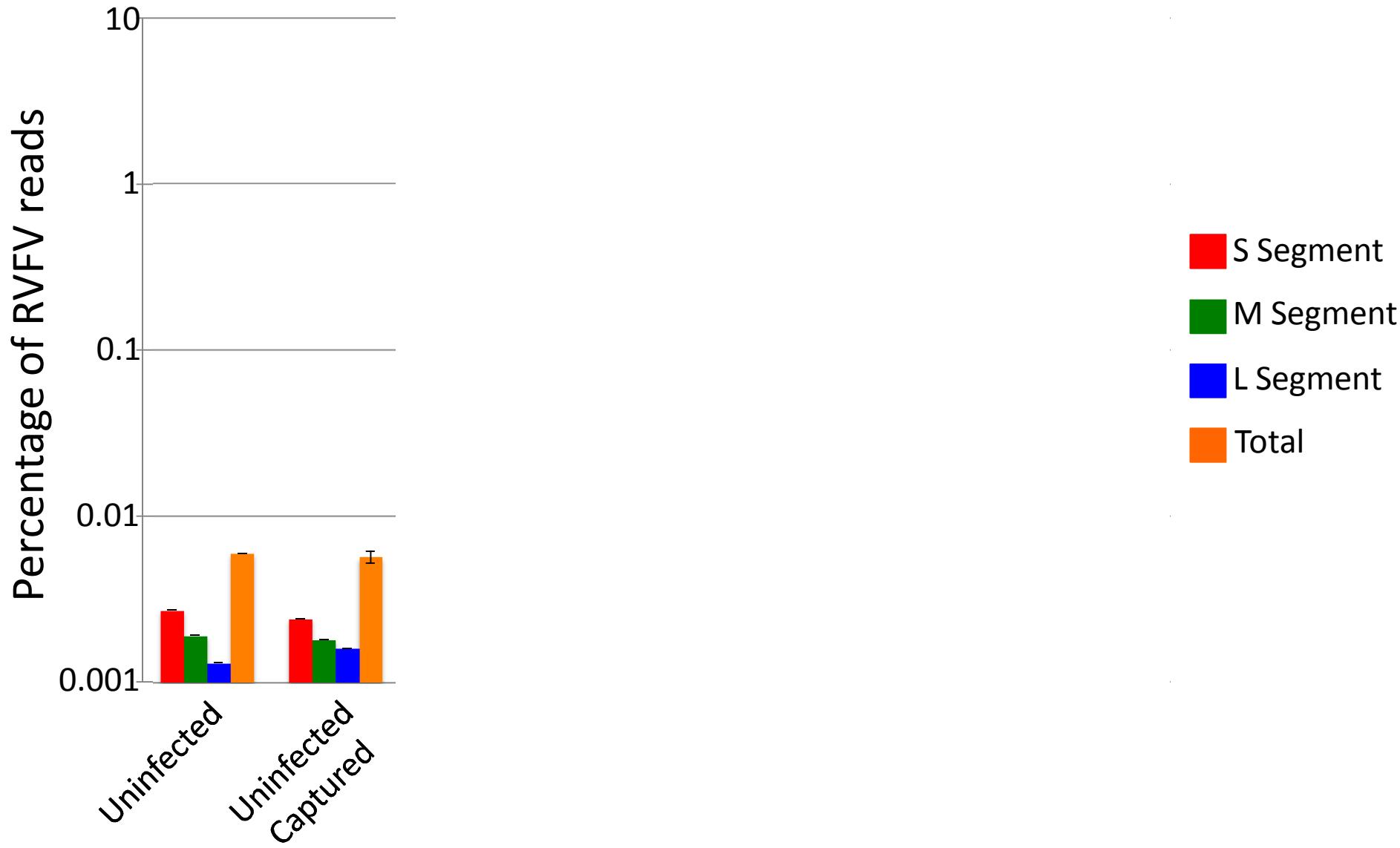
Vacuum manifold holds 6 strips



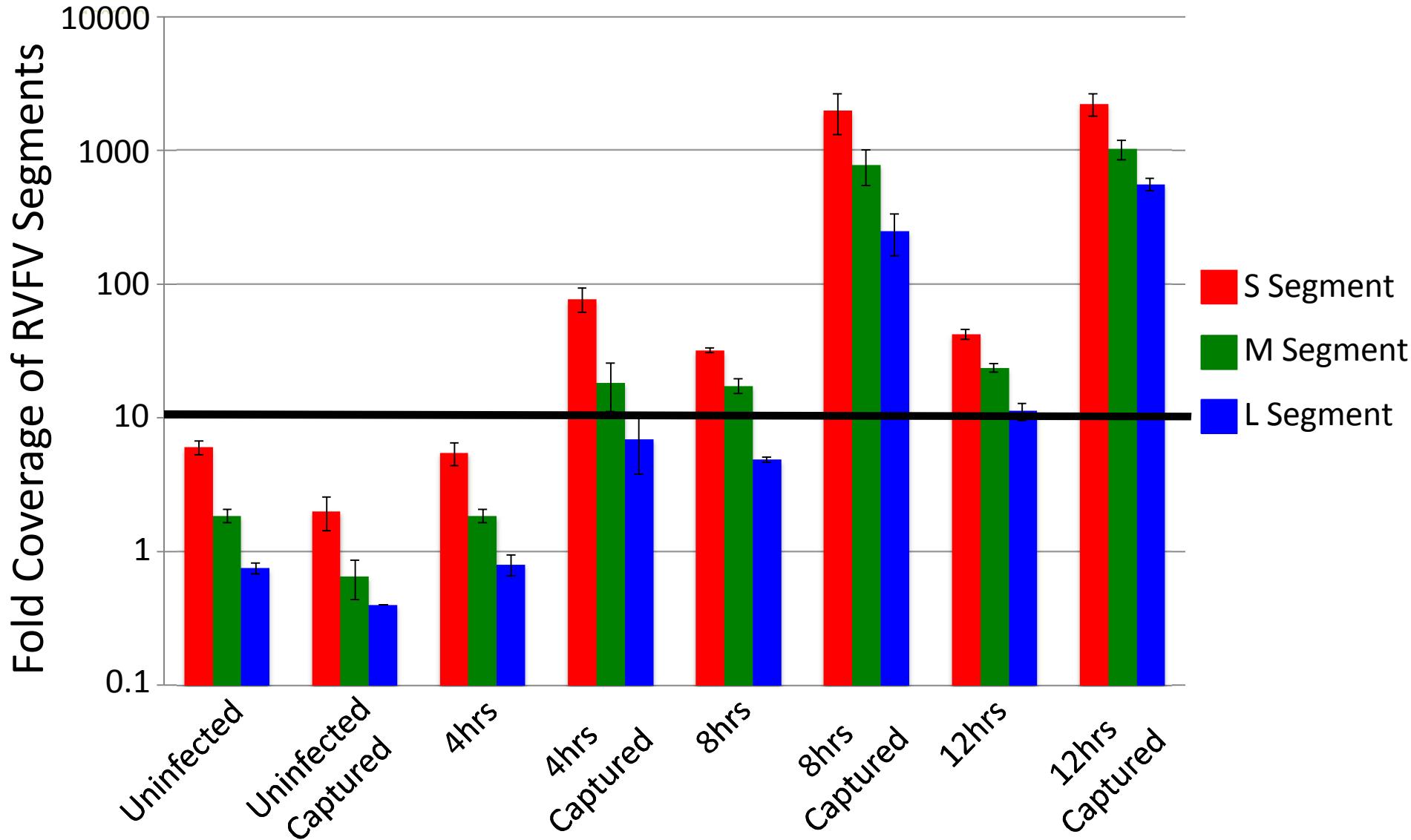
# Rift Valley Fever Virus (RVFV)

- Zoonotic, mosquito-borne, 3 segment, -ssRNA virus
  - Typically infects livestock
  - Bats may act as viral reservoirs
- 98% of people have mild symptoms
  - Fever, headache, general malaise
- 2% of people develop hemorrhagic fever
- Outbreak in Egypt in 1977-78 infected ~200,000 people with ~600 deaths
- Potential biological weapon

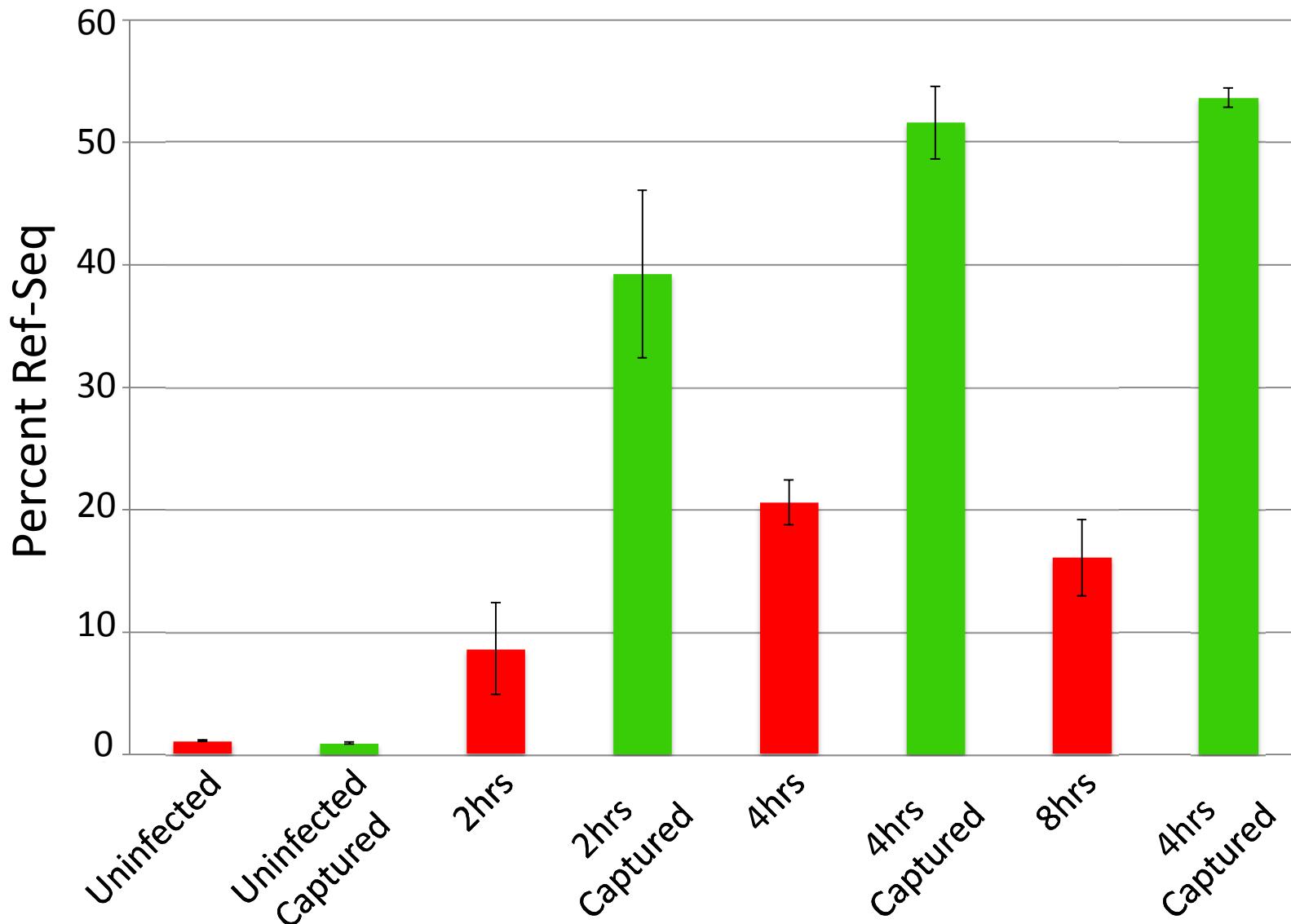
# RVFV Enrichment



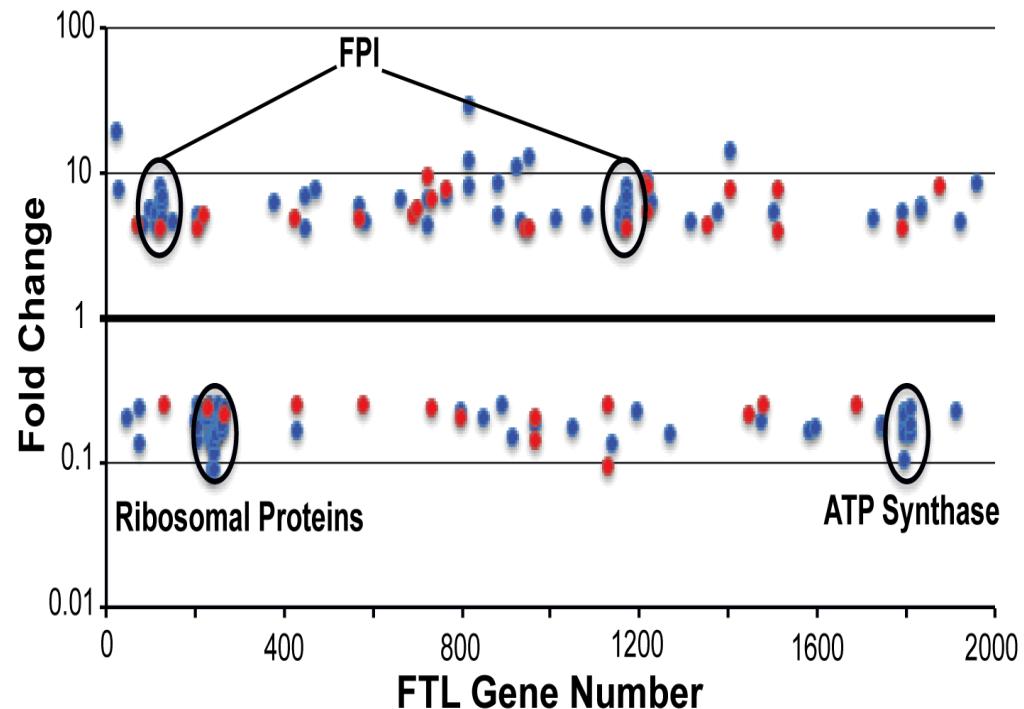
# RVFV Coverage



# Increase in Genes Covered

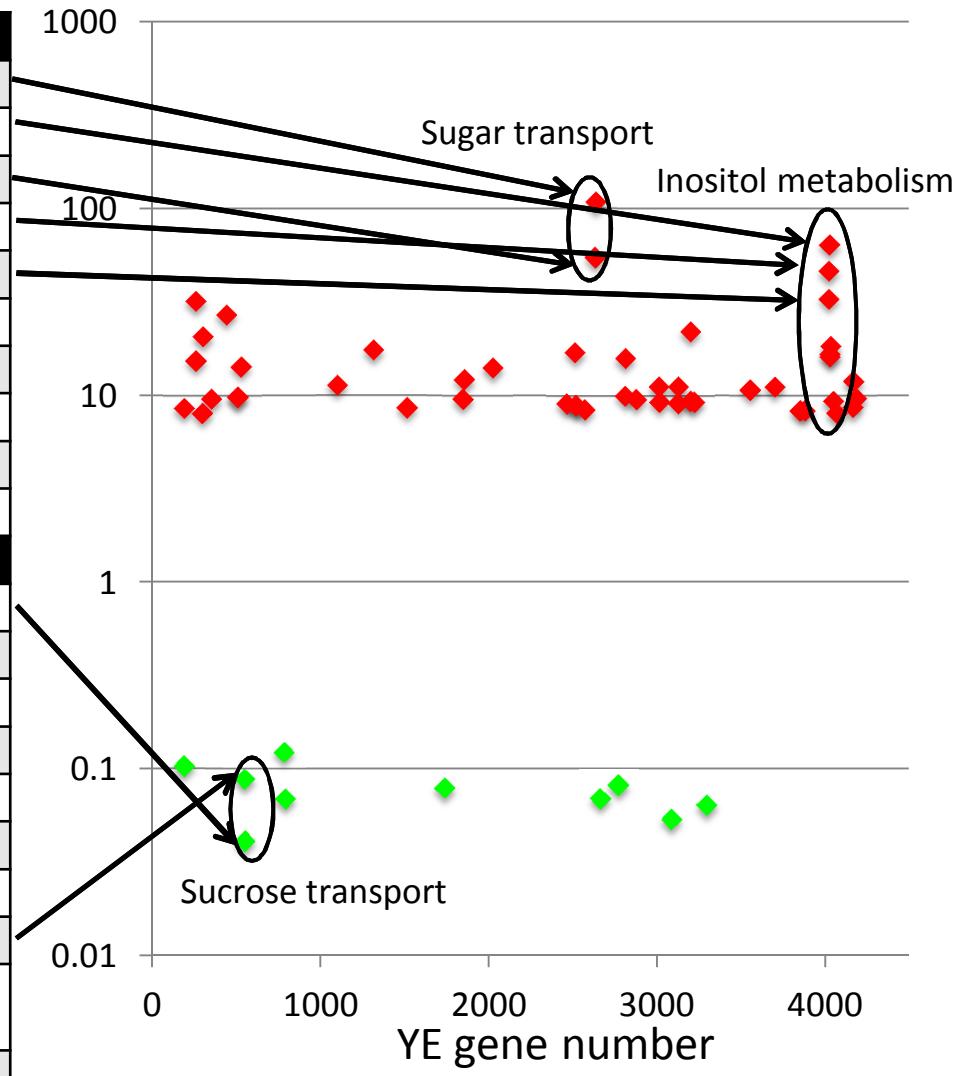


# Clustering by Gene Expression



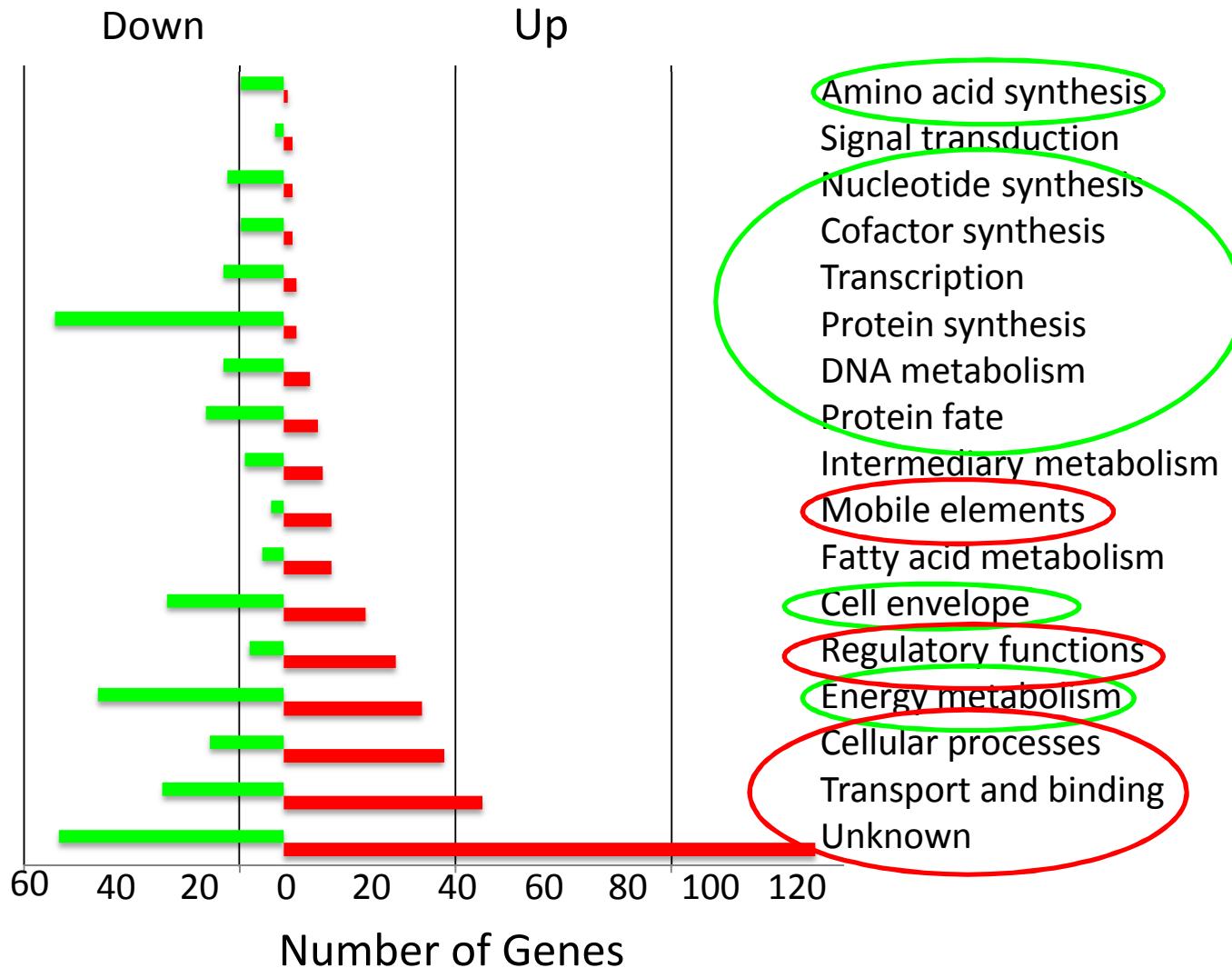
# Infection-dependent Differential Gene Expression

Gene ID	Name	Function	Fold Change
YE2639	-	Sugar Transporter	107.65
YE4027	-	Epi-inositol hydrolase	63.44
YE2638	-	Sugar Transport	54.79
YE4025	-	Inosose dehydratase	46.21
YE4026	<i>idh</i>	Myo-inositol 2-dehydrogenase	32.73
YE0267	<i>fadA</i>	Fatty acid oxidation	31.52
YE0447	-	Unknown membrane transporter	26.63
YE3198	<i>aglB</i>	6-phospho- $\alpha$ -glucosidase	21.75
YE0309	<i>acs</i>	acetyl-CoA synthetase	20.38
YE4031	-	aldehyde dehydrogenase	18.21
Gene ID	Name	Function	Fold Change
YE0553	<i>scrY</i>	sucrose porin	-24.15
YE3092	-	Unknown	-18.47
YE3297	<i>xni</i>	exonuclease	-15.63
YE0789	-	thiol:disulfide interchange	-14.42
YE2667	-	Unknown	-14.32
YE1740	-	Unknown	-12.59
YE2772	<i>hisI</i>	ATP pyrophosphatase	-12.19
YE0552	<i>scrA</i>	Sucrose transporter	-11.30
YE0195	<i>yigB</i>	Flavin mononucleotide phosphatase	-9.69
YE0781	<i>mrfC</i>	Fimbrial protein	-8.12

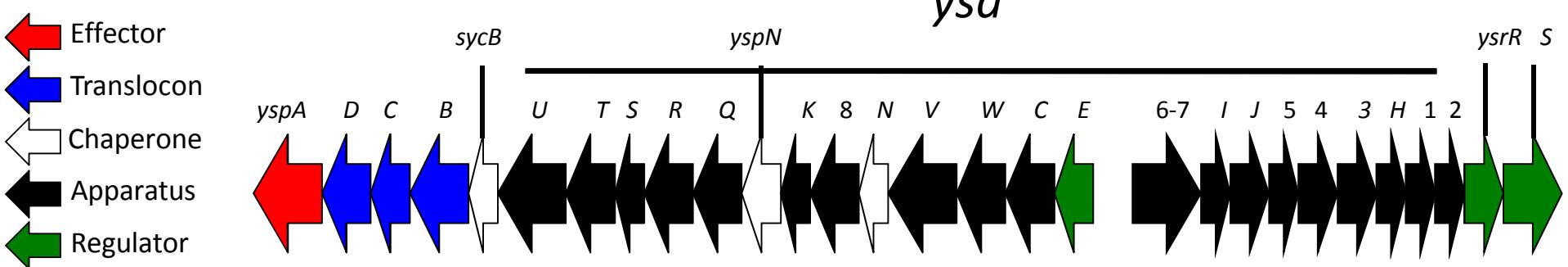


# Infection-dependent Differential Expression by Functional Category

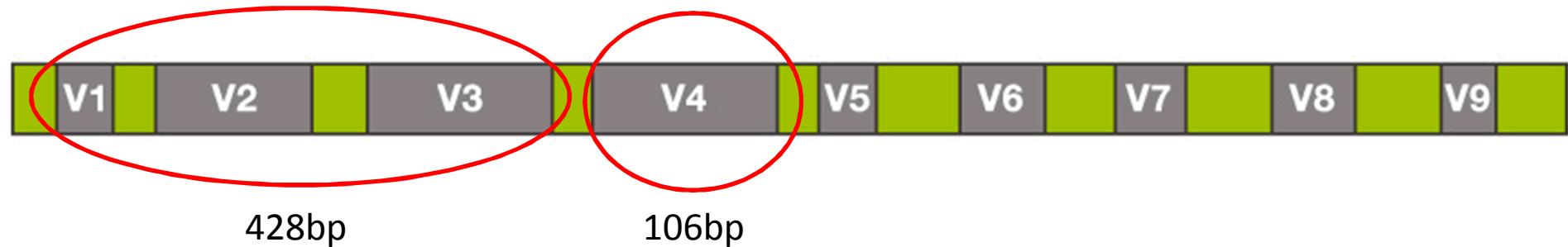
(Genes differentially expressed between RPMI and infection)



# Expression of the Ysa T3SS

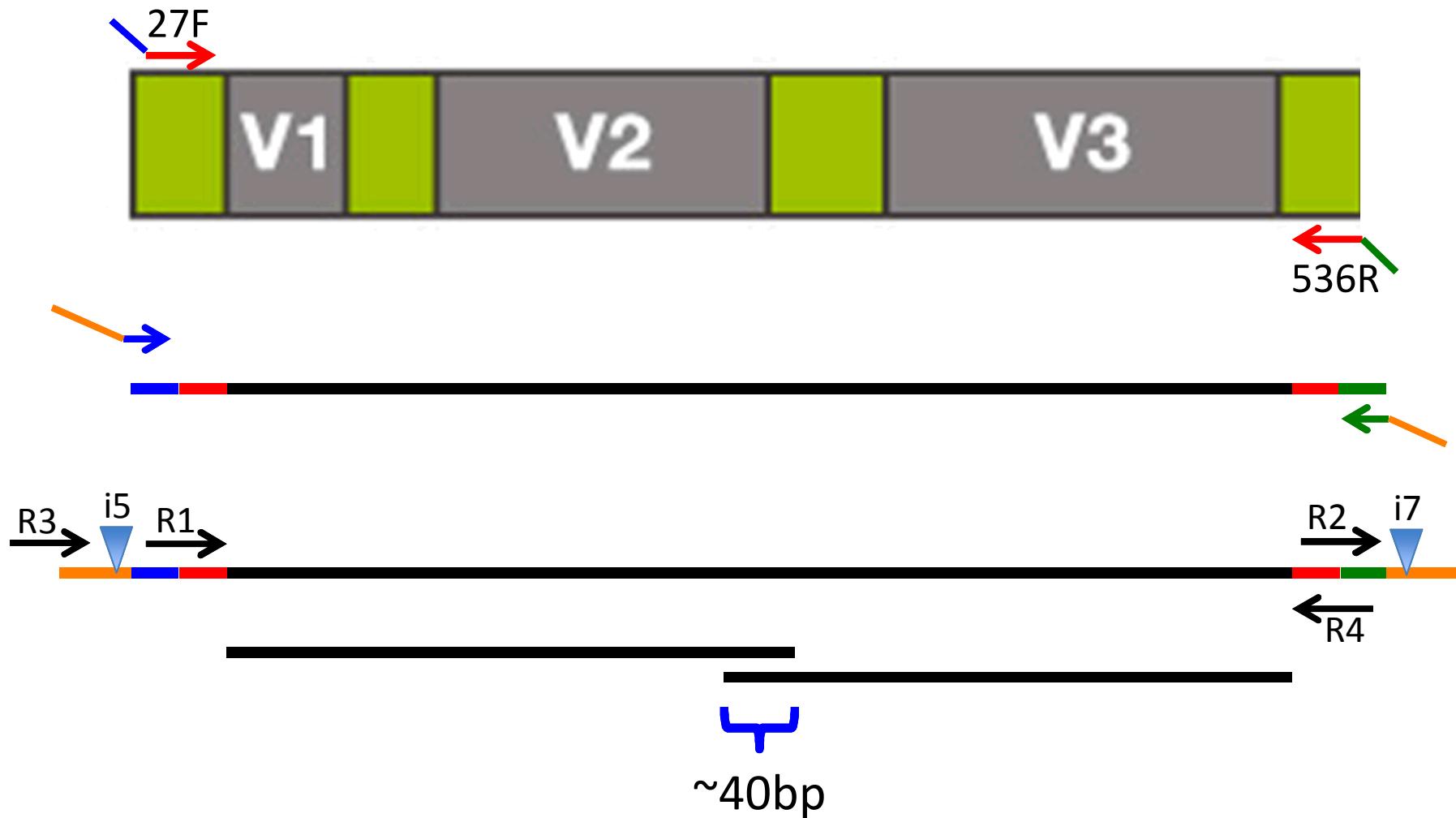


# 16S Ribosomal RNA



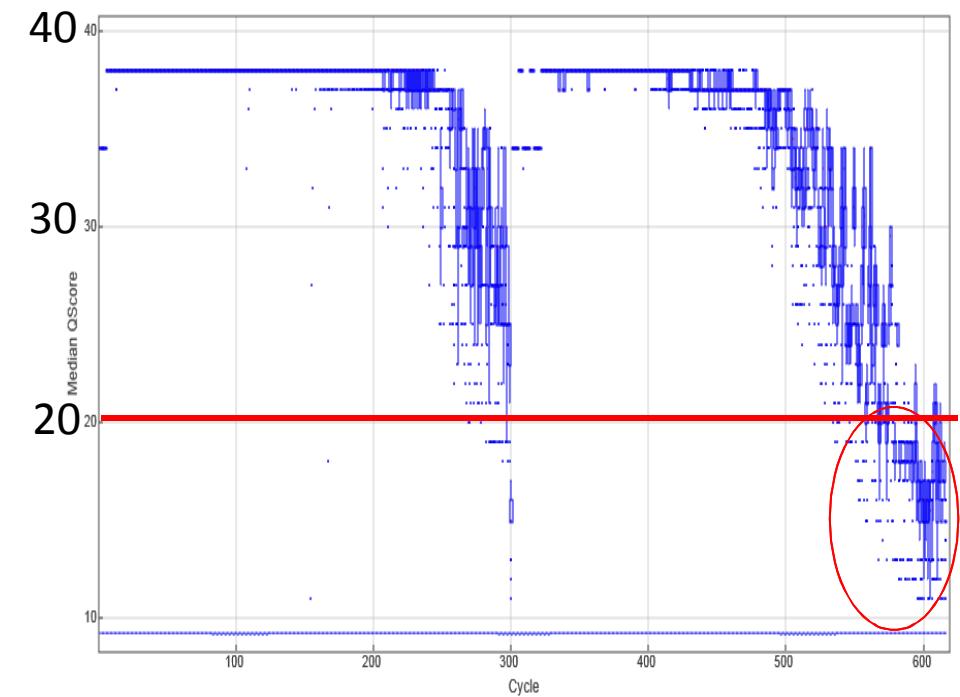
Variable Region	Bases Covered	Length	Identification Level
V4	576 - 682	106 bases	Family
V1-V3	69 - 497	428 bases	Genus/Species

# Creating the V1-V3 Amplicon

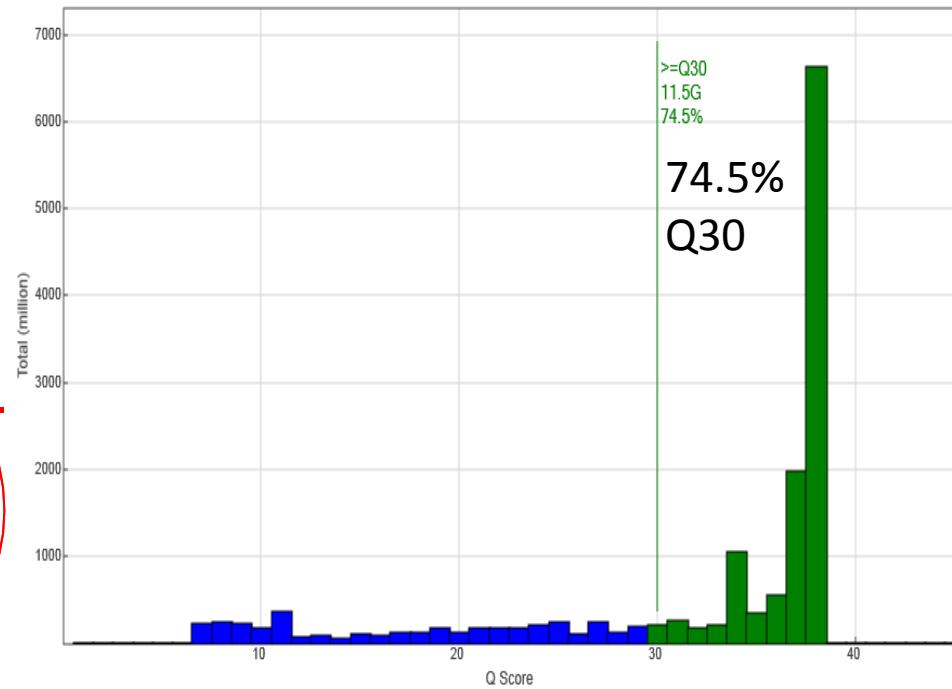


# Sequencing V1-V3 Amplicon

Median Q Score



Q Scores



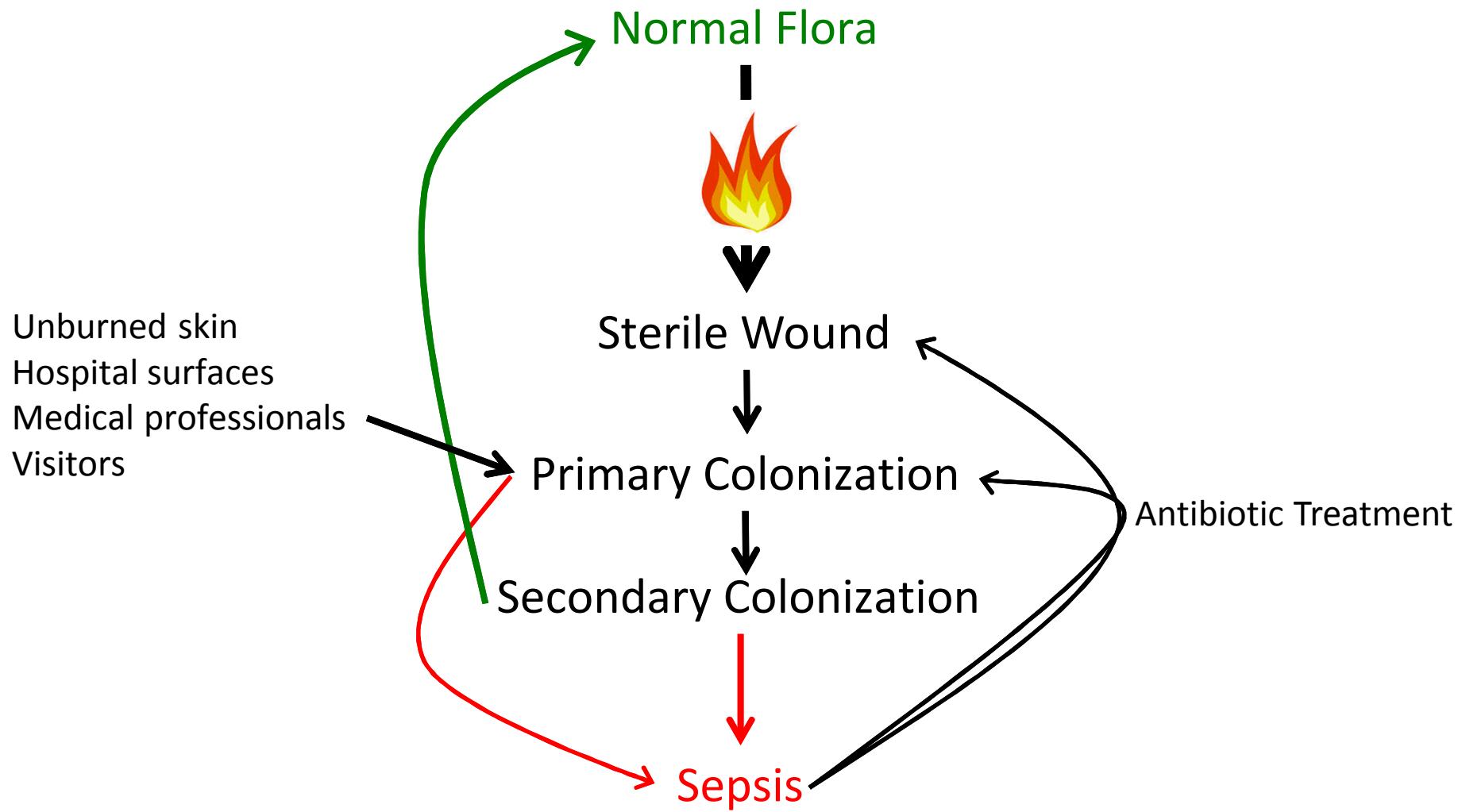
1053k/mm<sup>2</sup>  
25 million reads PF  
94% Indexed

# Typical Burn Wound Progression

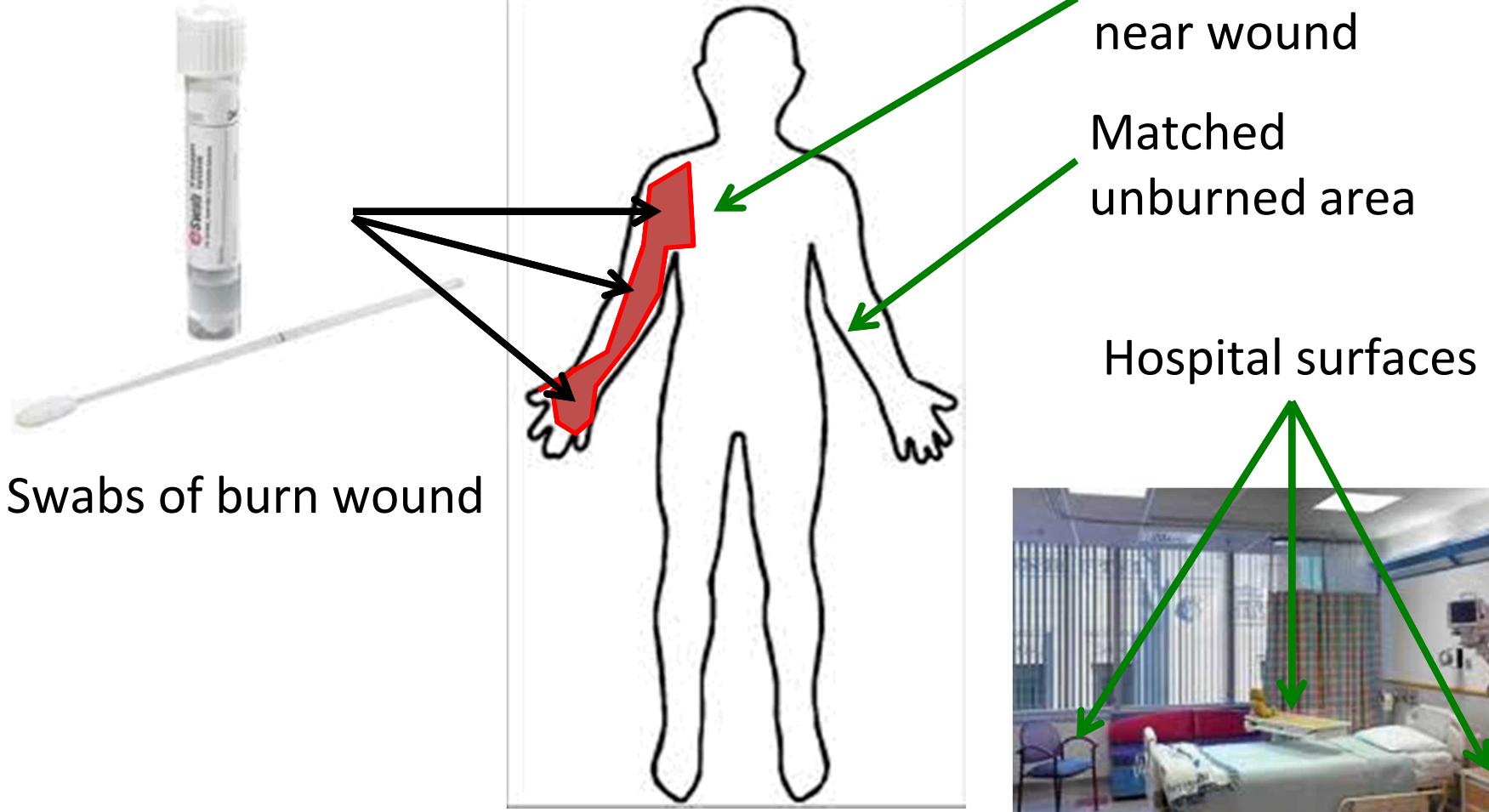
- Most patients with >15% surface area burn will develop sepsis at least once
- Week 1 = Gram-positive
- Week 2 = Gram-negative
- Later time points = fungal
- All patients are on some type of antimicrobial
  - IV/oral/topical
- Are there populations of bacteria that lead to better wound healing?

# Burn Wound Metagenomics

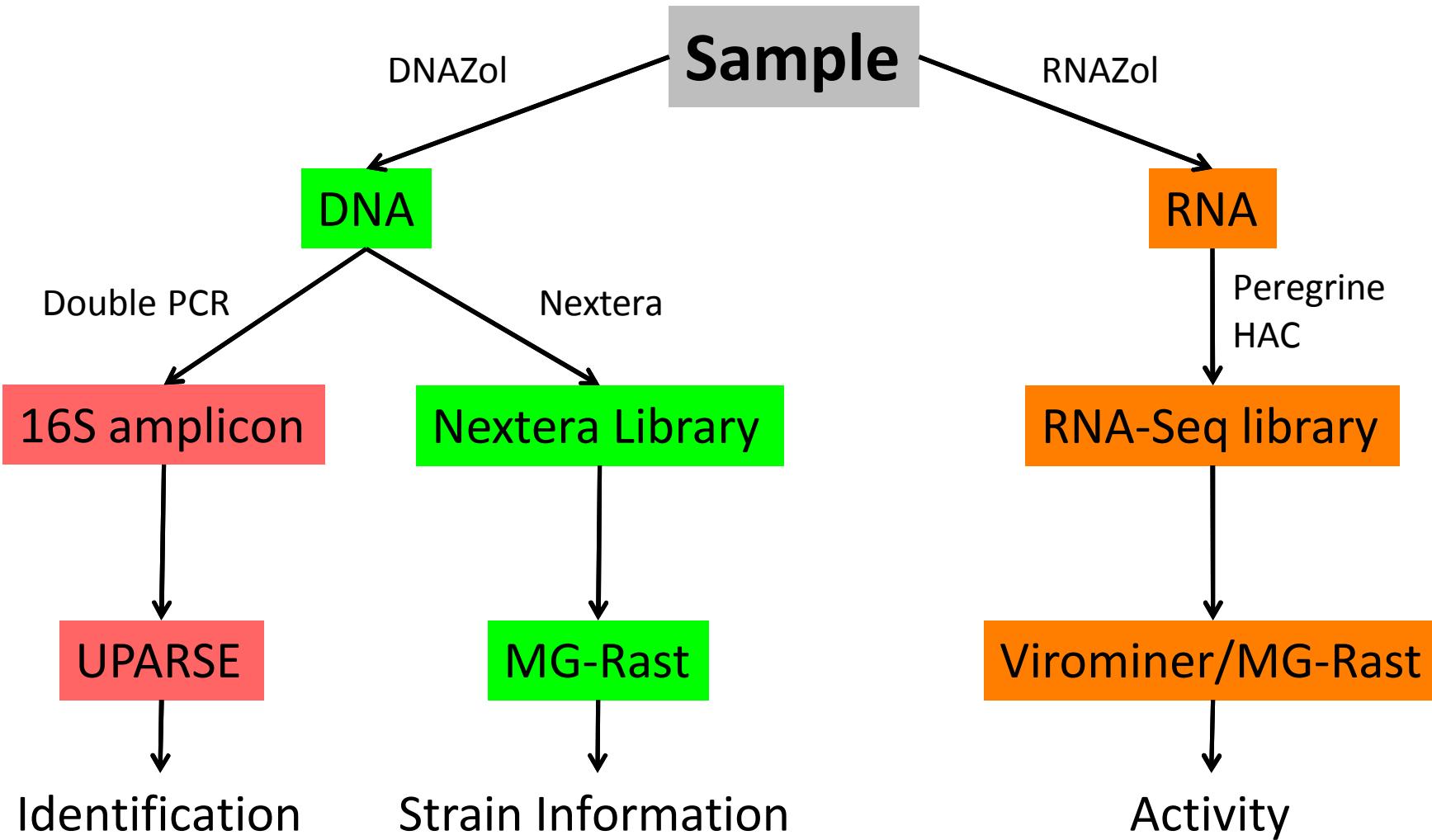
## An Ecological Approach



# Human Burn Wound Study Design



# Disease Diagnostics using Metagenomics and Metatranscriptomics



# Mountain Lions and Elephant Seals



- Mountain lion found visibly ill
- Tranquilized + antibiotic injection
- Blood sample collected
- Swabs collected
  - Oral
  - Nasal
  - Rectal
- Mother and pup found ill
- Tranquilized + antibiotic injection
- Blood sample collected
- Swabs collected
  - Oral
  - Nasal
  - Rectal

# Diagnosis from 16S V1-V3 Amplicon



Sample	Bacteria	Hits	% ID
Oral	<i>Fusobacterium</i>	212	98.5
Nasal	<i>Streptobacillus moniliformis</i>	1511	94.7
Rectal	<i>Fusobacterium</i>	1005	99.3
Blood	<i>Fusobacterium</i>	1761	99.8

Diagnosis: Unknown *Fusobacterium*

Other possibilities:

*Salmonella Typhimurium*

*Streptobacillus moniliformis*



©Bernard Zee

Sample	Bacteria	Hits	% ID
Oral	<i>Coenonia anatina</i>	6166	95.2
Nasal	<i>Bisgaardia Jack1</i>	935	98.8
Rectal	<i>Mycoplasma mirounga</i>	1571	99.4
Blood	<i>Mycoplasma mirounga</i>	1515	99.4

Diagnosis: *Mycoplasma mirounga*

Other possibilities:

*Actinomyces marimammalium*

*Coenonia anatina*

# Microbial Forensics

## Forensic identification using skin bacterial communities

Noah Fierer<sup>a,b,1</sup>, Christian L. Lauber<sup>b</sup>, Nick Zhou<sup>b</sup>, Daniel McDonald<sup>c</sup>, Elizabeth K. Costello<sup>c</sup>, and Rob Knight<sup>c,d</sup>

<sup>a</sup>Department of Ecology and Evolutionary Biology, <sup>b</sup>Cooperative Institute for Research in Environmental Sciences, and <sup>c</sup>Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309; and <sup>d</sup>Howard Hughes Medical Institute

Edited by Jeffrey I. Gordon, Washington University School of Medicine, St. Louis, MO, and approved February 13, 2010 (received for review January 05, 2010)

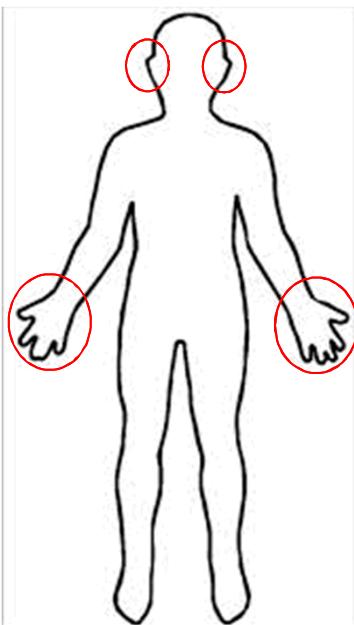
### Conclusions

- Structure of bacterial communities can be used to differentiate objects handled by different people
- Objects can be matched to individual people based on their hand microbiome
- Communities on objects remain stable if untouched for up to 2 weeks

### Questions

- Are microbial populations stable over longer periods of time with normal use?
  - On people?
  - On objects?
- Is differentiation possible if larger numbers of people are tested?
- Is it possible to tell if multiple people have used an object?
  - Which people?

# Microbial Forensics Study Design



## People Swabs

- Both hands
- Phone ear
- Phone cheek
- Inner elbow

## Smart Phone Swabs

- Front and back
- Every Monday - 10 weeks
- Once a month after



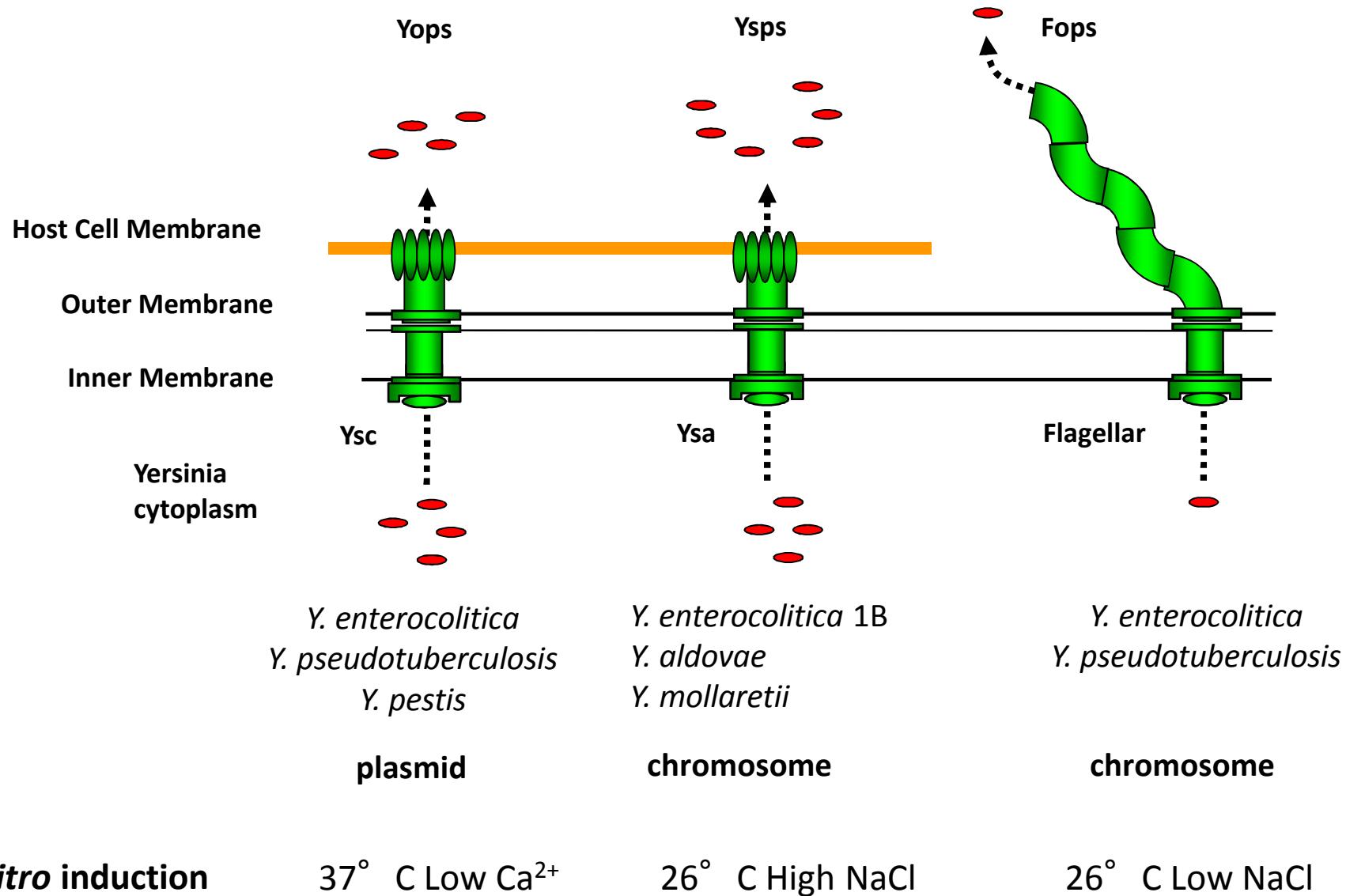
## Office Swabs

- Various objects
  - Personal
  - Communal

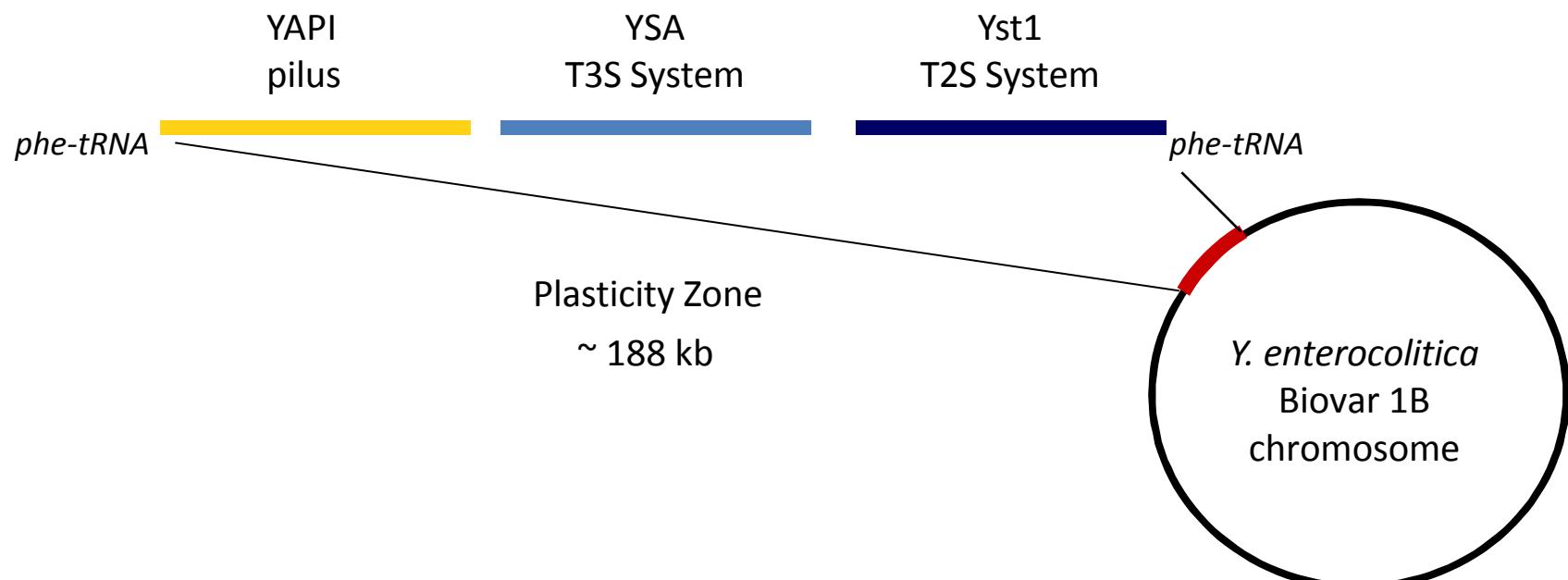
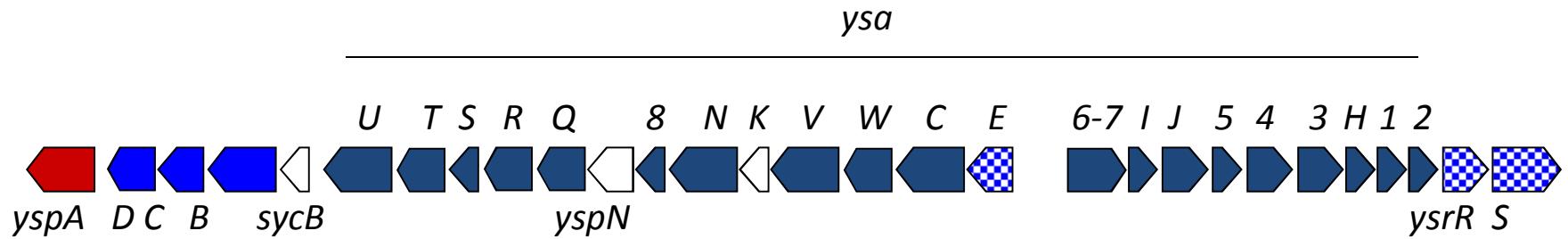
# Metagenomics Conclusions

- V1-V3 sequencing approach leads to more informative MiSeq runs
- Microbial succession of burn wounds may lead to better understanding of healing and non-healing communities
- Combined metagenomic and metatranscriptomic approach can lead to diagnosis of unknown diseases
- It may be possible to identify the users of objects using microbial forensics

# *Y. enterocolitica* Type III Secretion

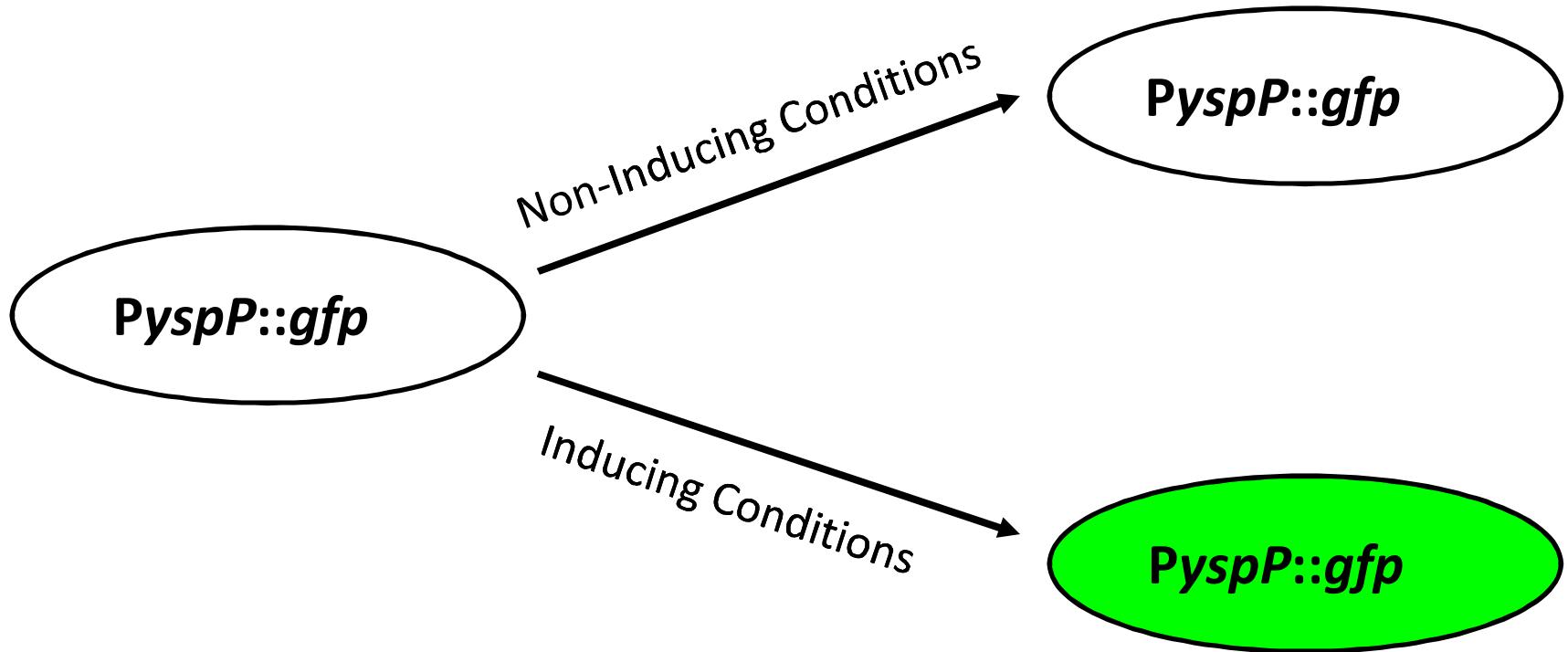


# *Y. enterocolitica* biovar 1B



# Experimental Conditions

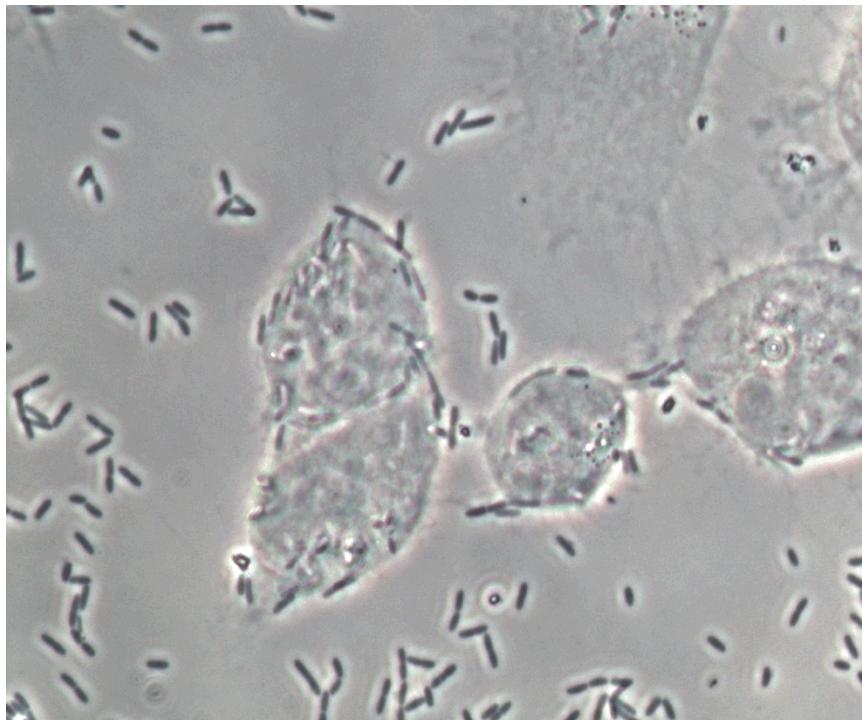
Transcriptional fusion between *PyspP* and *gfp*



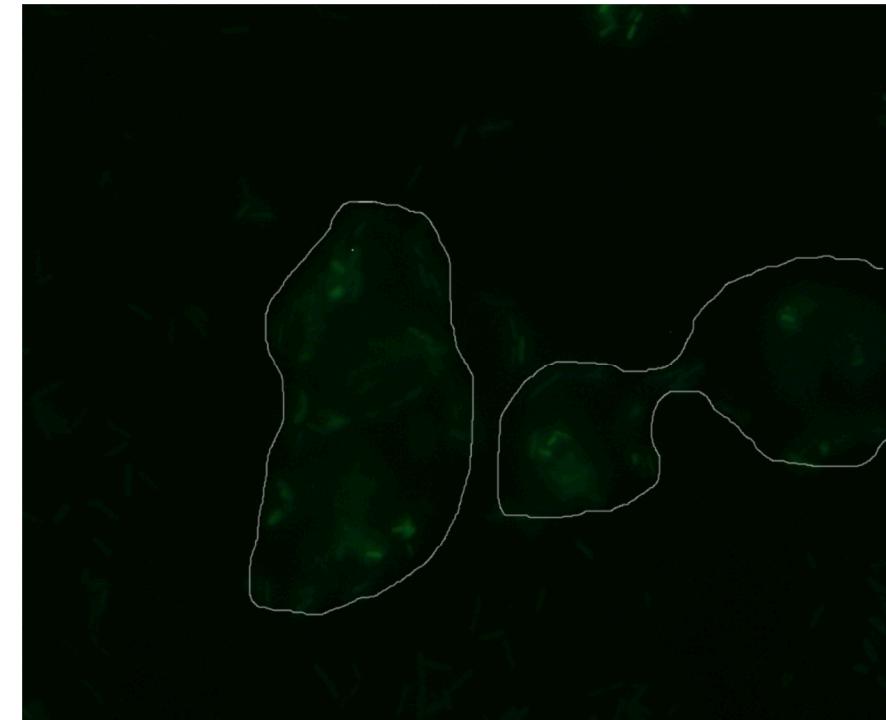
Fluorescence can be quantified by microscopy or by flow cytometry

# HeLa Cell Infections

*PyspP::gfp*



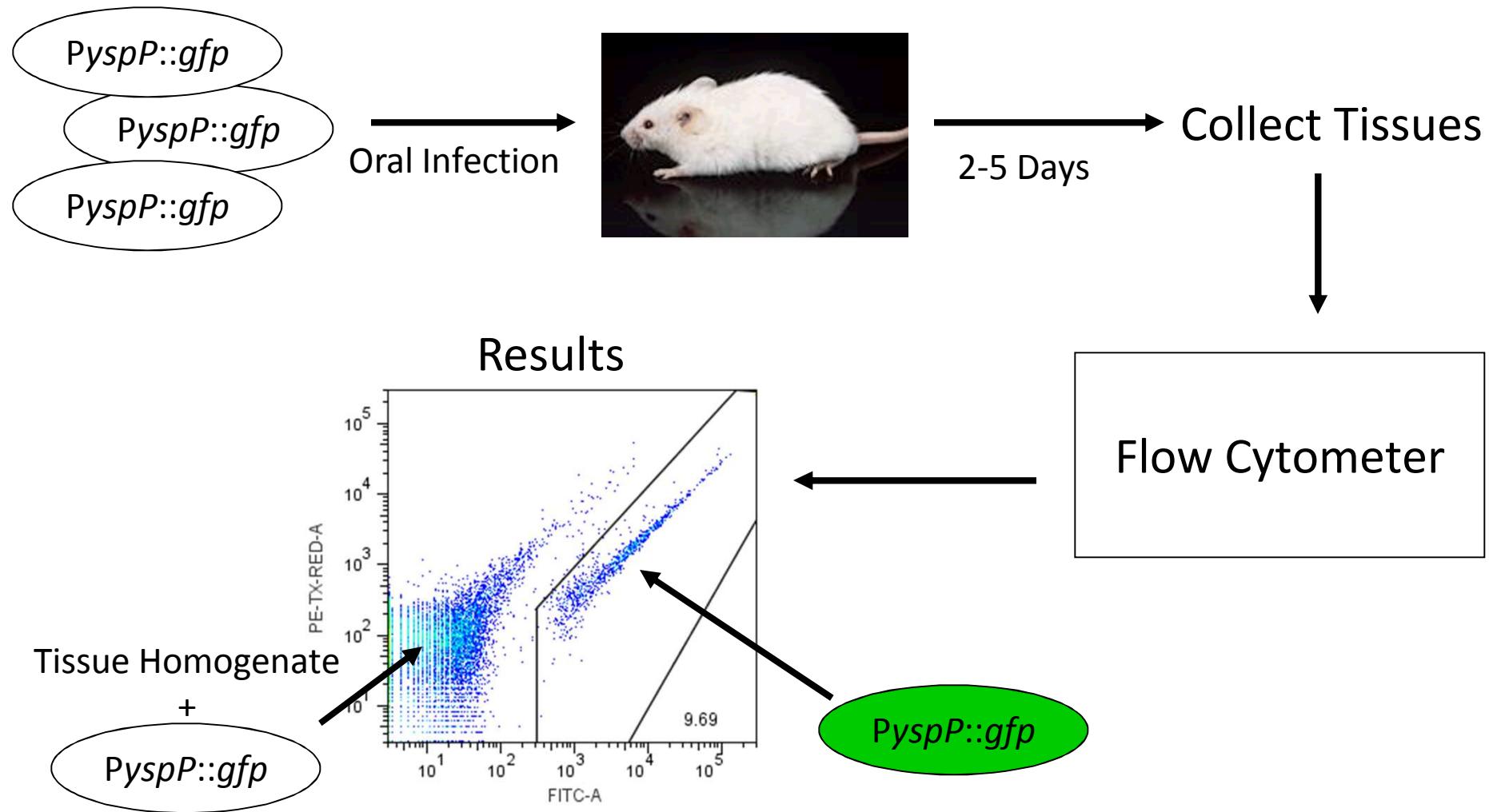
Phase Contrast



Fluorescence

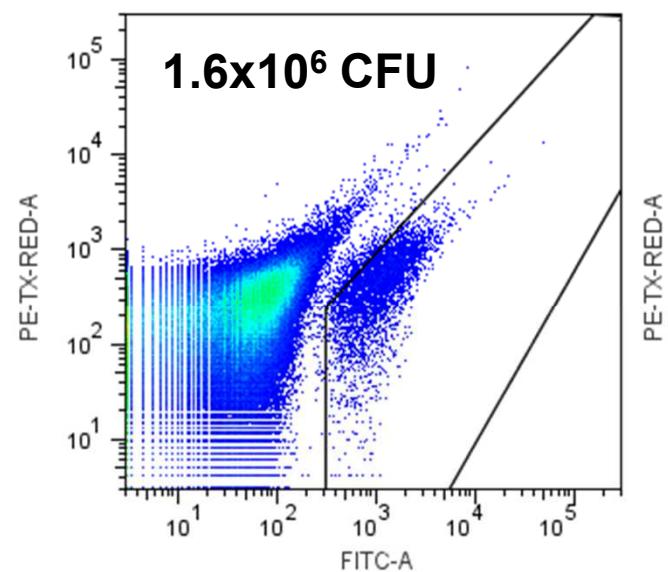
Observations confirmed by qRT-PCR

# *In Vivo* Protocol

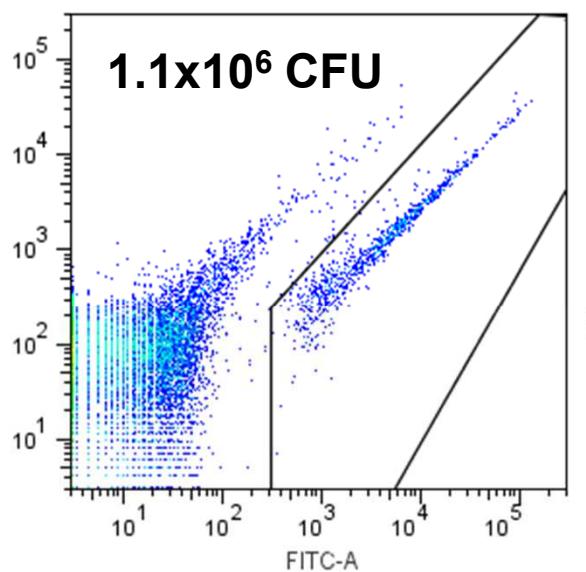


# Results by Tissue

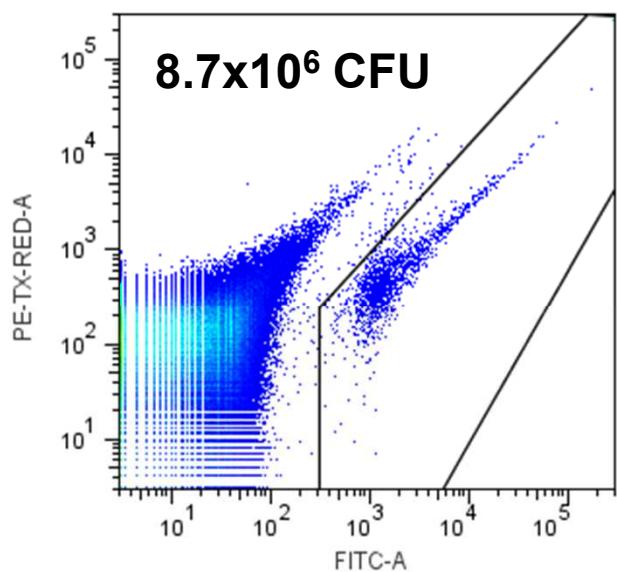
Terminal Ileum



Peyer's Patches



Mesenteric Lymph Nodes



# *Y. enterocolitica* Conclusions I

- Ysa T3SS is expressed *in vitro* in a contact dependent manner
  - Physiologically relevant conditions
- Ysa T3SS is expressed through the course of a murine infection in each tissue examined
- Problem: conclusions based on analysis of only two genes