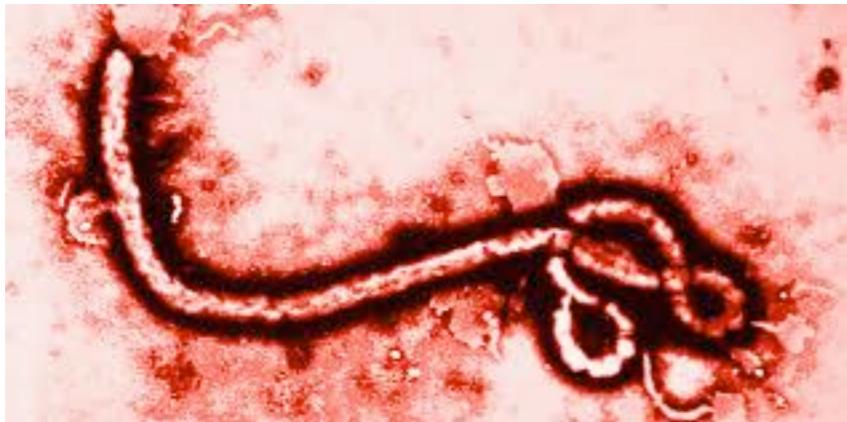
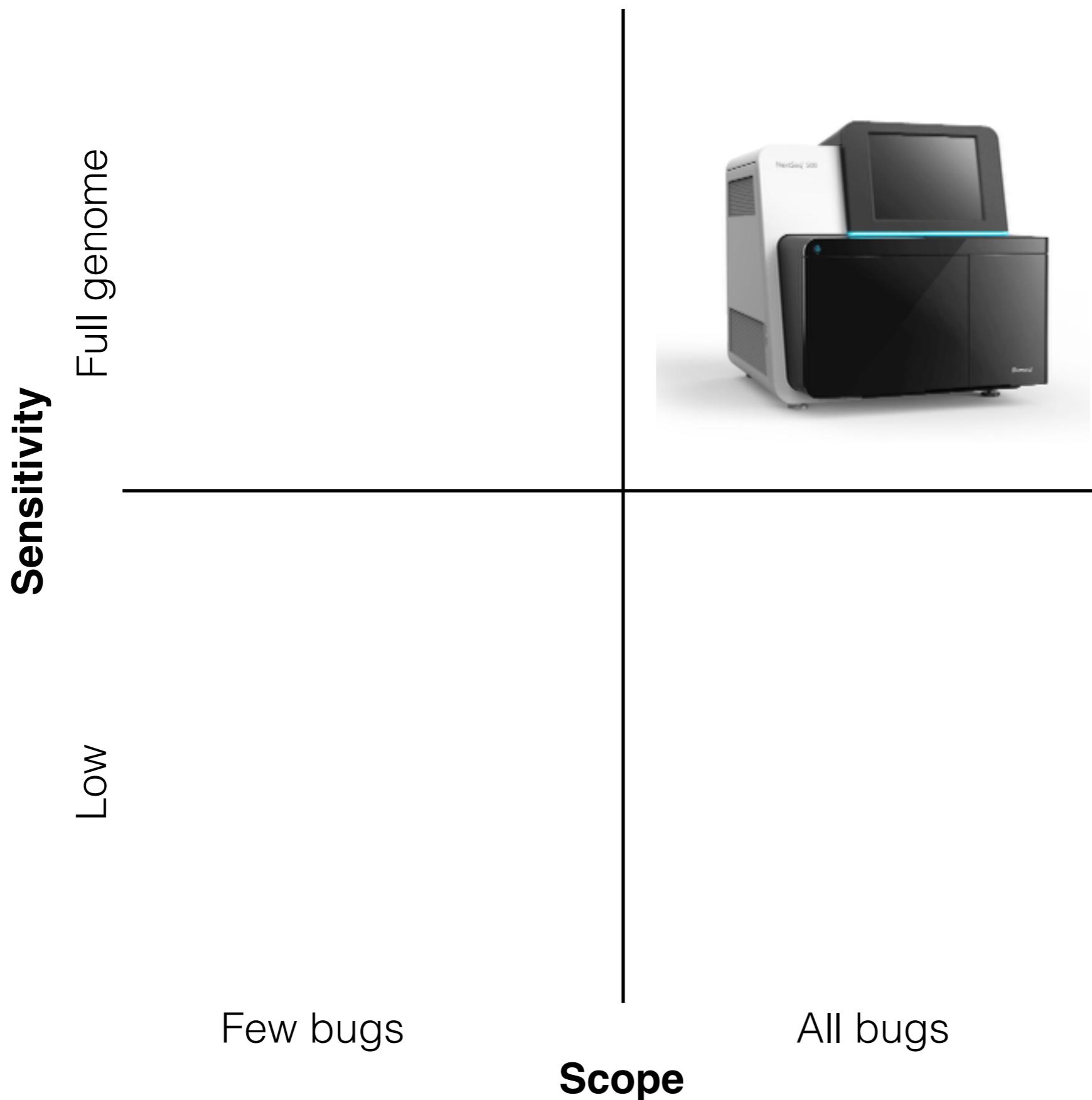


# Monitoring the human infectome

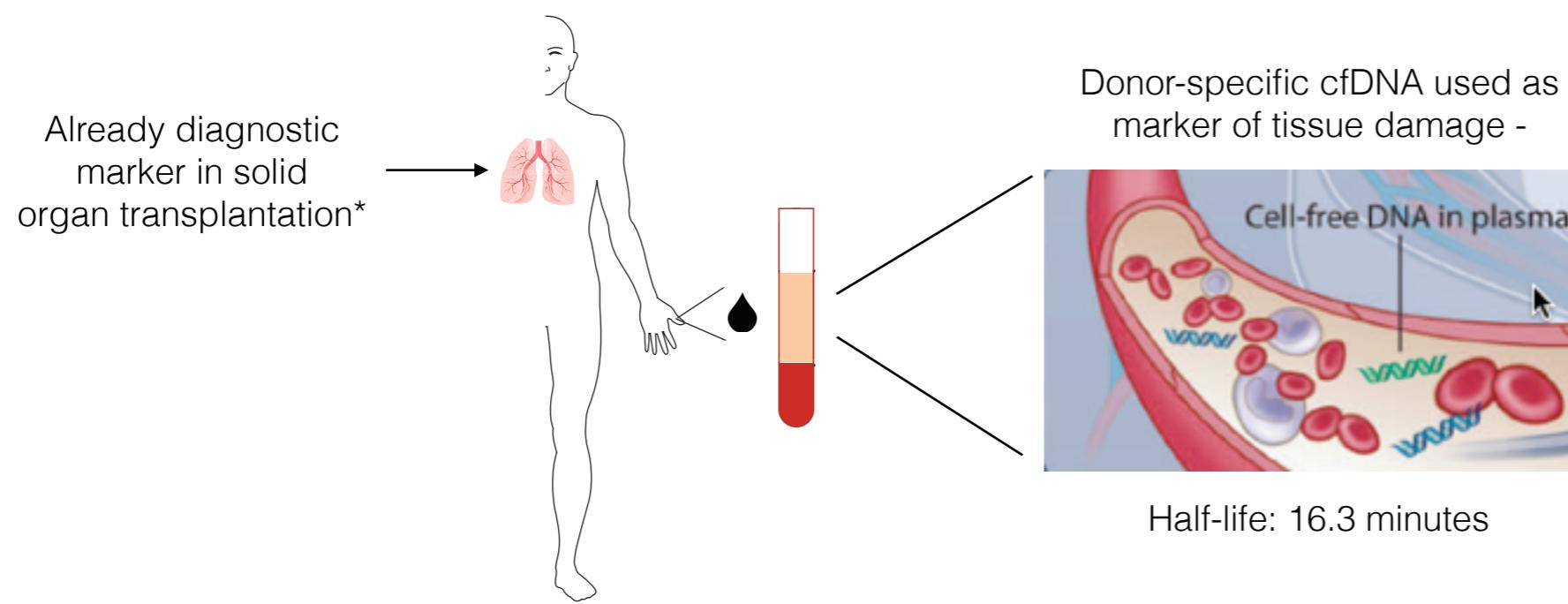
Lance Martin



# Non-invasive detection of any pathogen?



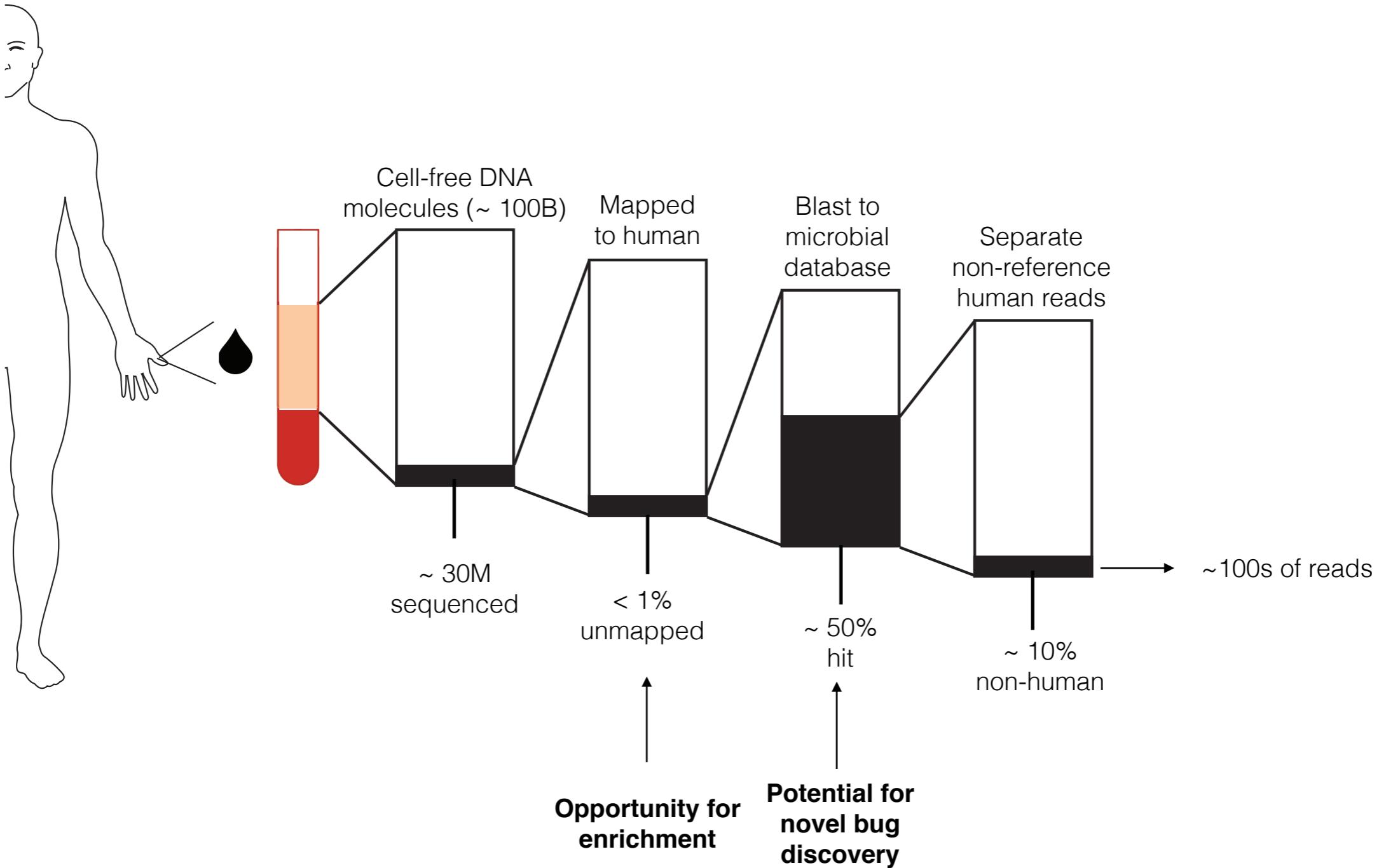
# Possibility of cell-free DNA for infection monitoring.



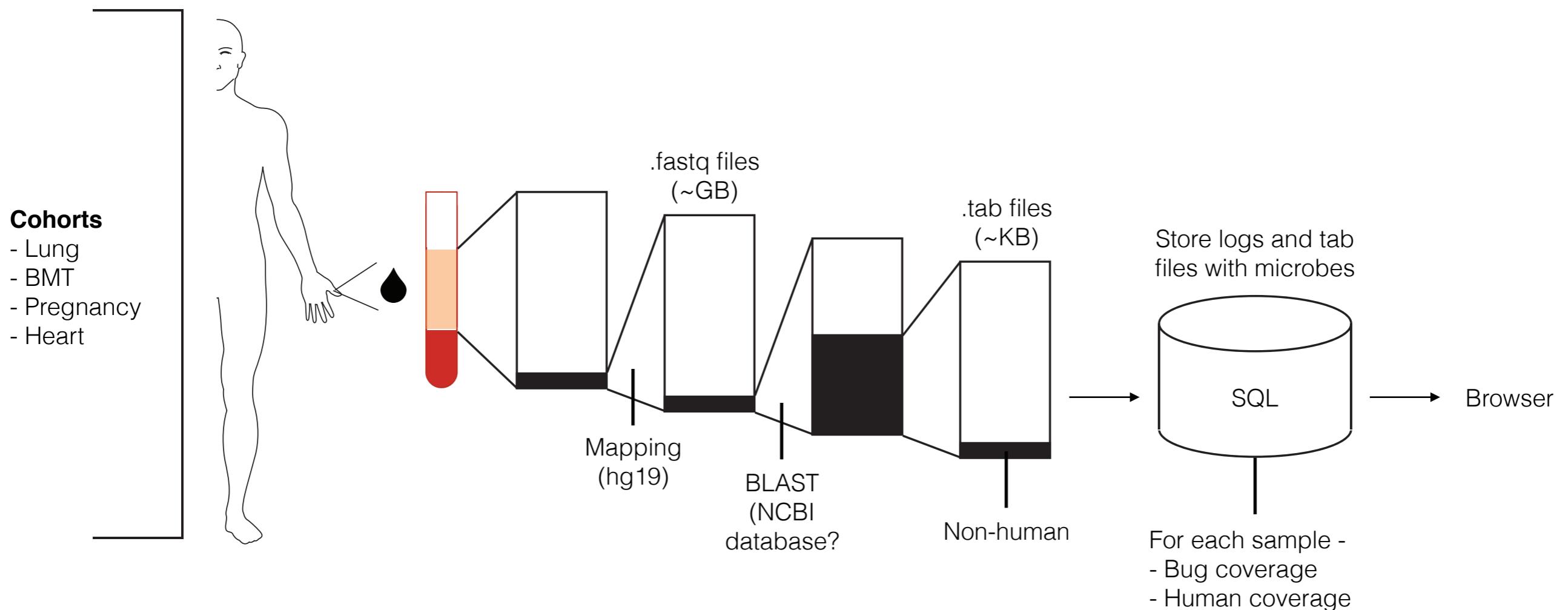
\* As well as rapidly growing use in pregnancy, and emerging effort in cancer monitoring.

De Vlaminck et al Sci Trans Med (2014), Lo et al Lancet (1998)

# Human microbiome in cell-free DNA.

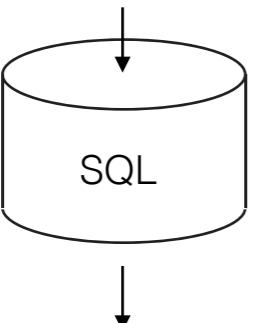


# Pipeline for isolating microbial reads in cell-free data.



# Browser for visualization and navigation.

1000s of samples



Infectome Explorer Cohorts - Taxonomic level -

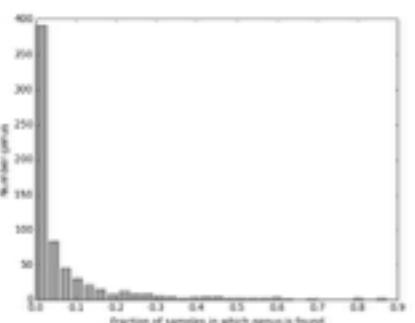
## Welcome to the Infectome Explorer.

Choose a cohort:

All cohorts SMT Lung Pregnancy Ebola Stanford clinic Biopsy Pregnancy RNA CPS

### Cohort data for All.

Use dropdown menu above to switch between cohorts.



Cohort parameter Value

Number of patients: 94

Number of samples: 807

Show: 10 Search: 86

Patient ID: 86 Number\_of\_samples: 16

Showing 2 of 2 records (filtered from 807 total records)

Pages: Previous Next

Cohort parameter Value

Number genera detected: 867

Show: 10 Search:

Name Prevalence

Bacillus 0.67

Propionibacterium 0.66

Acidovorax 0.61

Pseudomonas 0.79

Corynebacterium 0.68

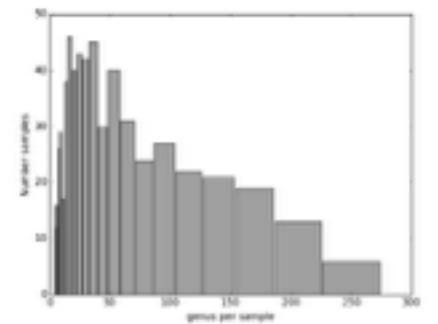
Verinimonas 0.53

Sphaerotilus 0.60

Methylophilus 0.68

Sphaerotilus 0.59

Methylophilus 0.67

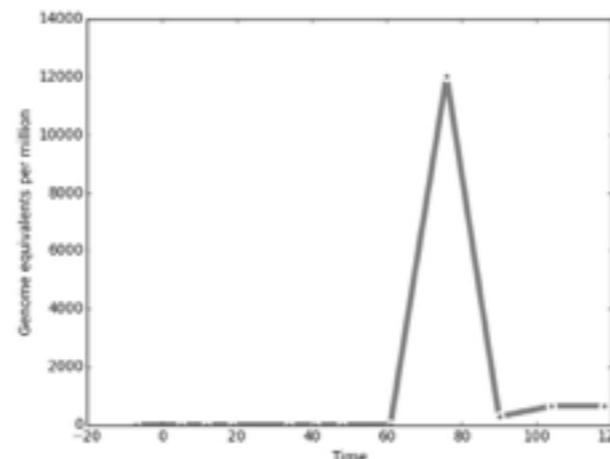


### Infection timeseries for Polyomavirus in I6.

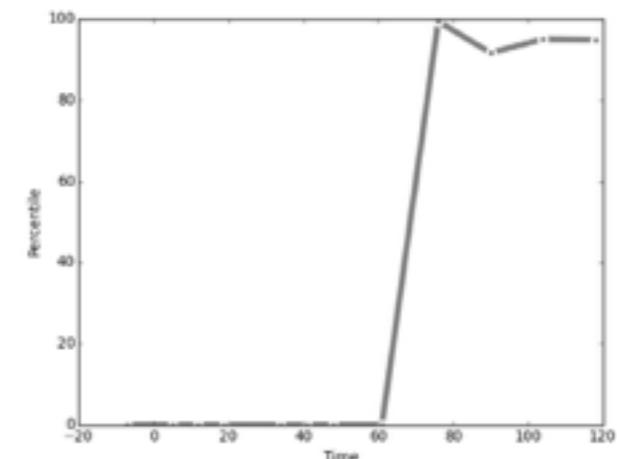
Use back to return to cohort or dropdown menu above to switch between cohorts.

**W Polymaviridae:** Polyomaviruses are DNA-based (double-stranded DNA, ~5000 base pairs, circular genome) viruses. They are small (40–50 nanometers in diameter), and icosahedral in shape, and do not have a lipoprotein envelope. Moreover, the genome possess early and late genes, contributing to its complex transcription program. They are potentially oncogenic (tumor-causing); they often persist as latent infections in a host without causing disease, but may produce tumors in a host of a different species, or a host with an ineffective immune system. The name polyoma refers to the viruses' ability to produce multiple (poly-) tumors (-oma). The family Polyomaviridae used to be one of two genera within the now obsolete family Papovaviridae (the other family being Papillomaviridae). The name Papovaviridae derived from three abbreviations: Pa for Papillomavirus, Po for Polyomavirus, and Va for "vacuolating". Clinically, Polyomaviridae are relevant as they contribute to pathologies such as Progressive multifocal leukoencephalopathy (PML virus), nephropathy (BK virus), and Merkel cell cancer (Merkel cell virus). Until recently, the family of Polyomaviridae contained only one genus (Polyomavirus). The recent expansion of known Polyomaviruses called for reclassification of the family into 3 genera: Orthopolyomavirus, Wukipolyomavirus, and Avipolyomavirus. Murine polyomavirus was the first polyomavirus discovered by Ludwik Gross in 1953. Subsequently, many polyomaviruses have been found to infect birds and mammals. For nearly 40 years, only two polyomaviruses were known to infect humans. Genome sequencing technologies have recently discovered seven additional human polyomaviruses, including one causing most cases of Merkel cell carcinoma and another associated with transplant-associated dysplasia (TSV), that are natural infections of humans. Discovery of these polyomaviruses in humans and animals, leading to fundamental insights into carcinogenesis, DNA replication and protein processing. The tumor suppressor molecule p53 was discovered, for example, as a cellular protein bound by the major oncoprotein (cancer-causing protein) T antigen made by Simian vacuolating virus 40 (SV40). The avian polyomavirus sometimes referred to as the Budgerigar fledgling disease virus is a frequent cause of death among caged birds.

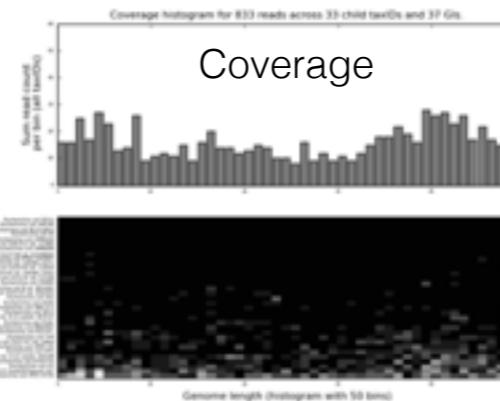
### Infection load for patient



### Percentile (relative to all samples)

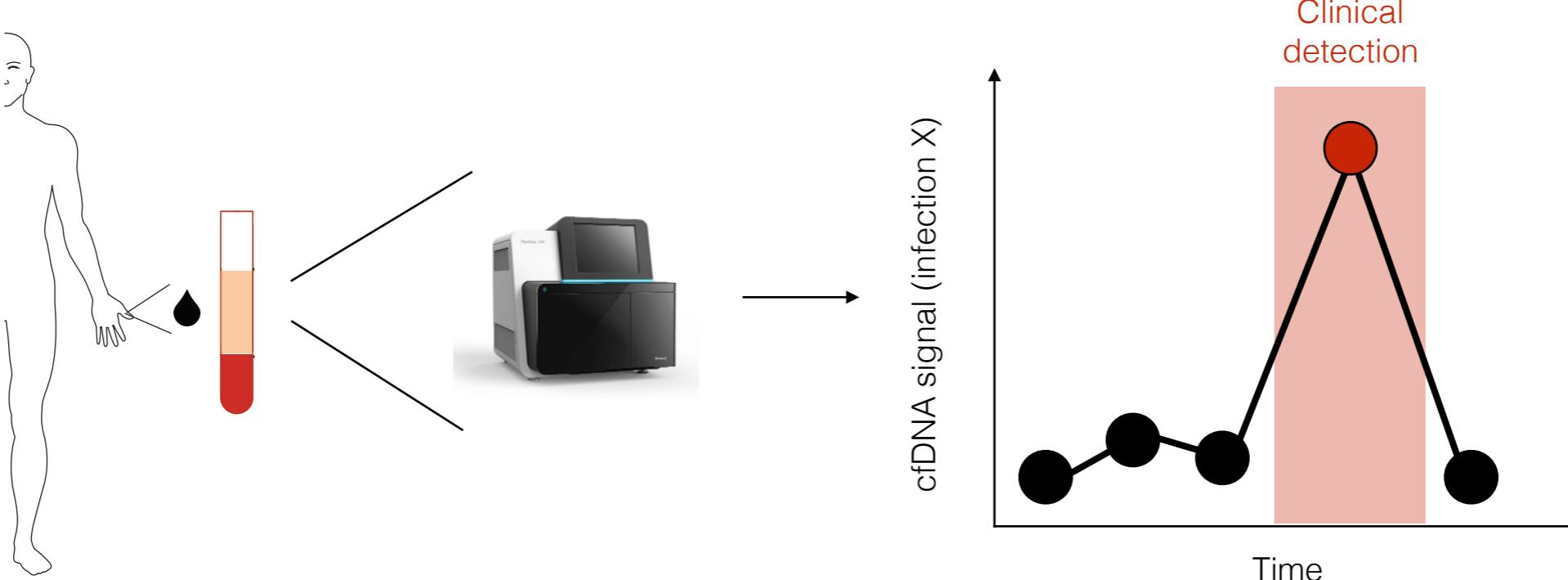


### Coverage



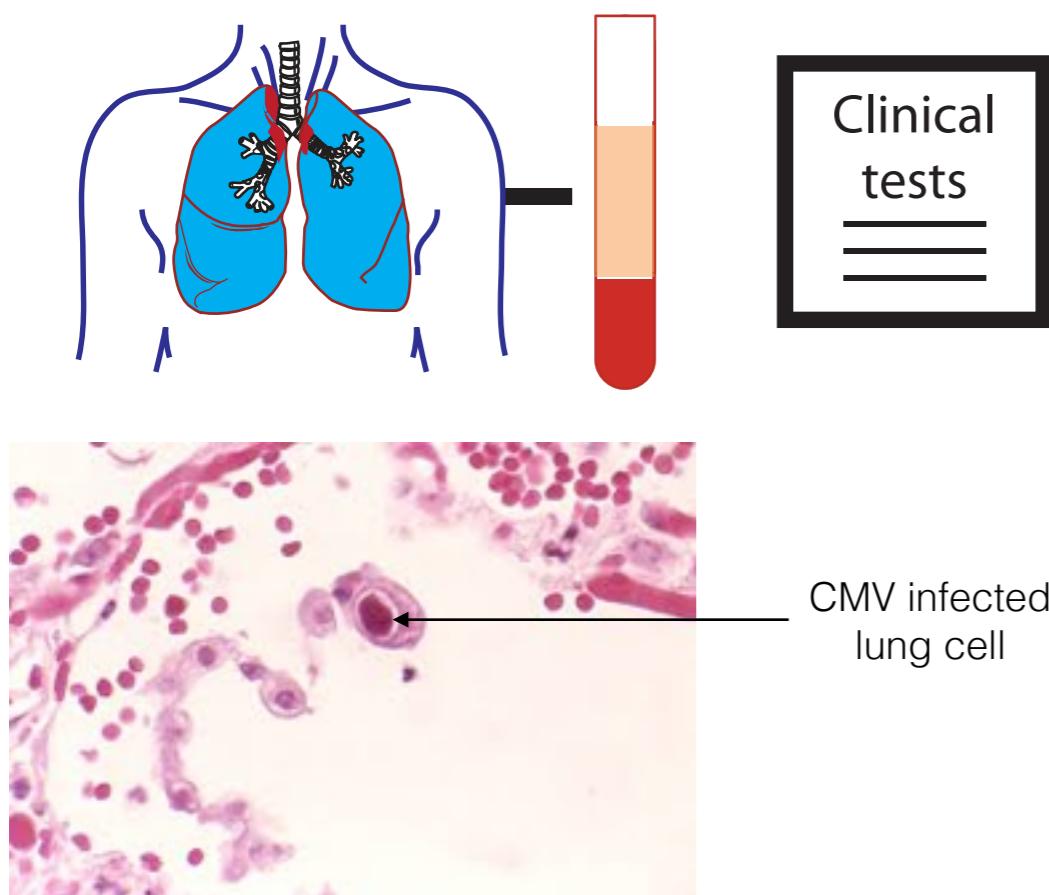
ID	Days	Date	Total Blast	C21 Coverage	Blast for inf
I6_BC	-8	2013-10-23 00:00:00	235	1.590626	27.628748
I6_D-1	-7	2013-10-24 00:00:00	379	1.001016	57.134878
I6_W1	5	2013-11-05 00:00:00	60	1.068258	22.686783
I6_W2	12	2013-11-12 00:00:00	48	0.987716	23.581704
I6_W3	19	2013-11-19 00:00:00	45	0.817058	28.274749
I6_W4	34	2013-12-04 00:00:00	65	0.893145	35.540858
I6_W5	41	2013-12-11 00:00:00	85	1.062171	45.917179
I6_W6	48	2013-12-18 00:00:00	299	1.190450	45.601895
I6_W8	61	2013-12-31 00:00:00	678	1.061075	49.335448
I6_W10	76	2014-01-15 00:00:00	807	0.878492	45.095884

# But how to know if it's useful?

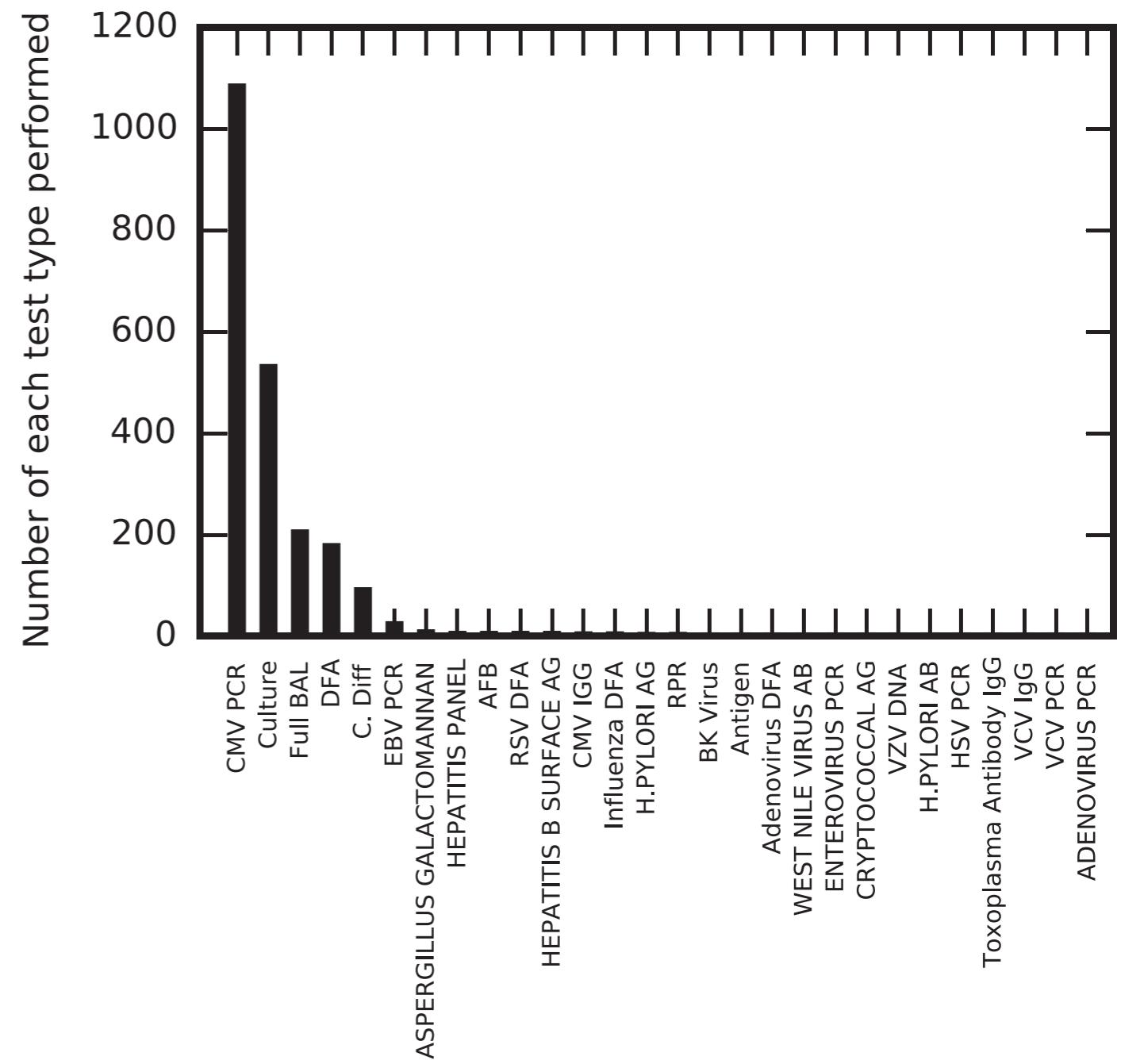


# Look back at the clinical history of our cohorts.

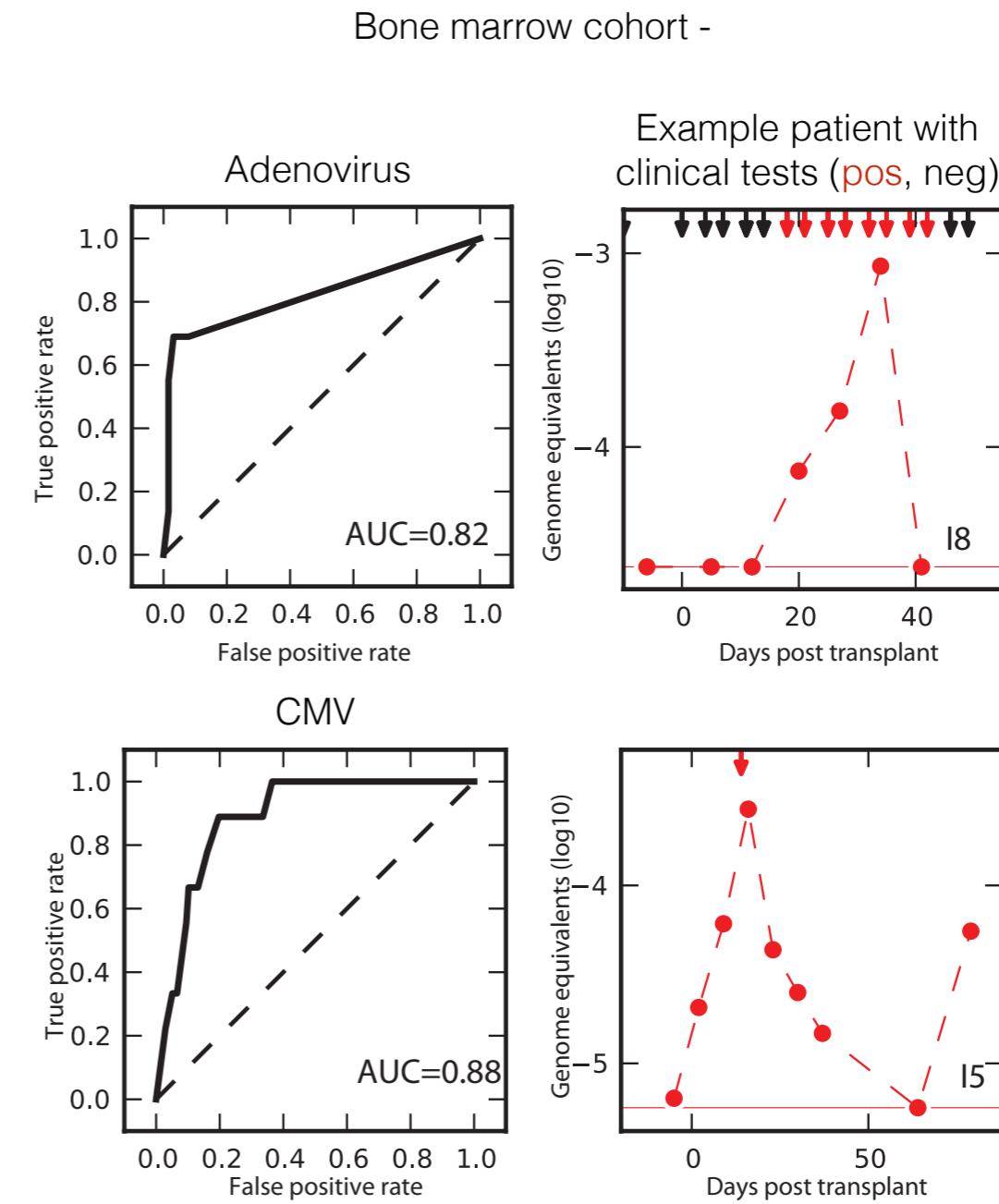
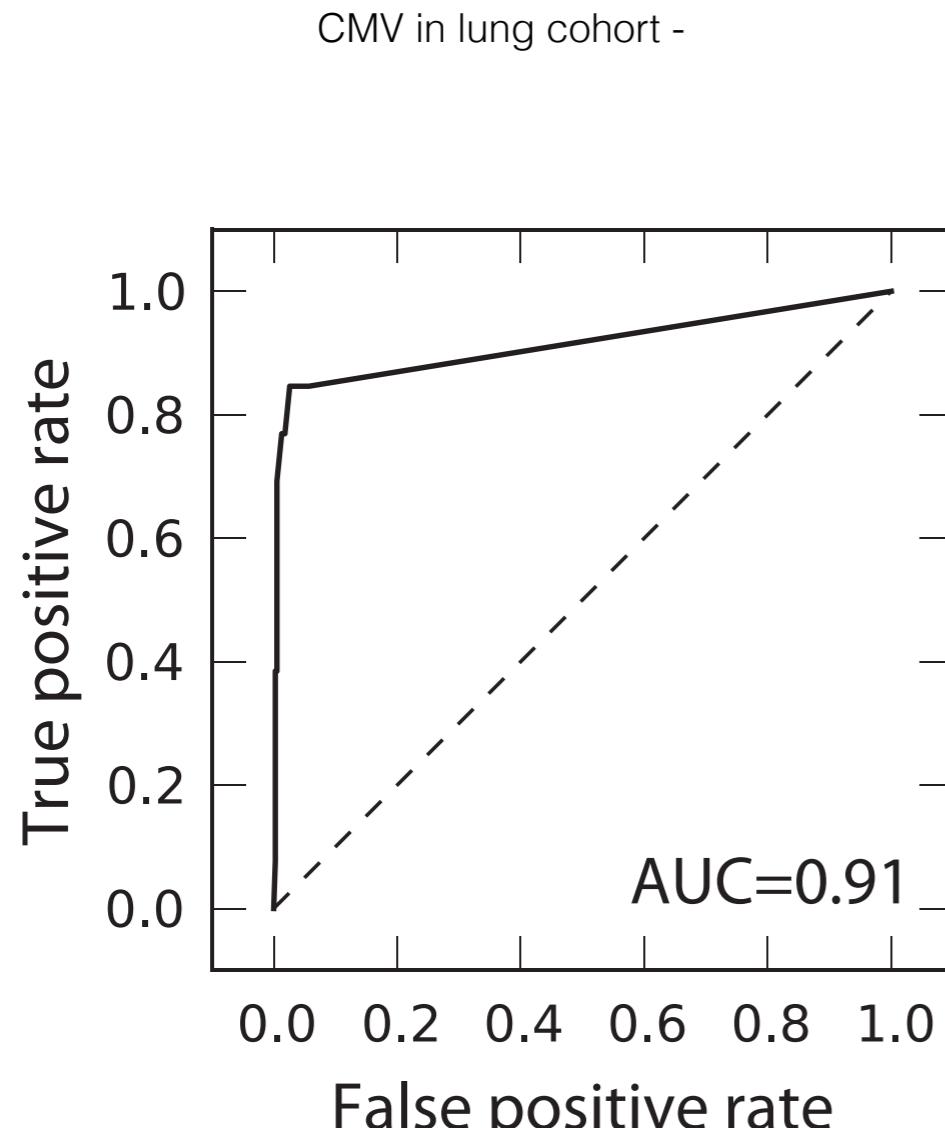
Lung Transplant (431 samples collected and processed) -



Thousands of tests recorded (~35k bug measurements) -



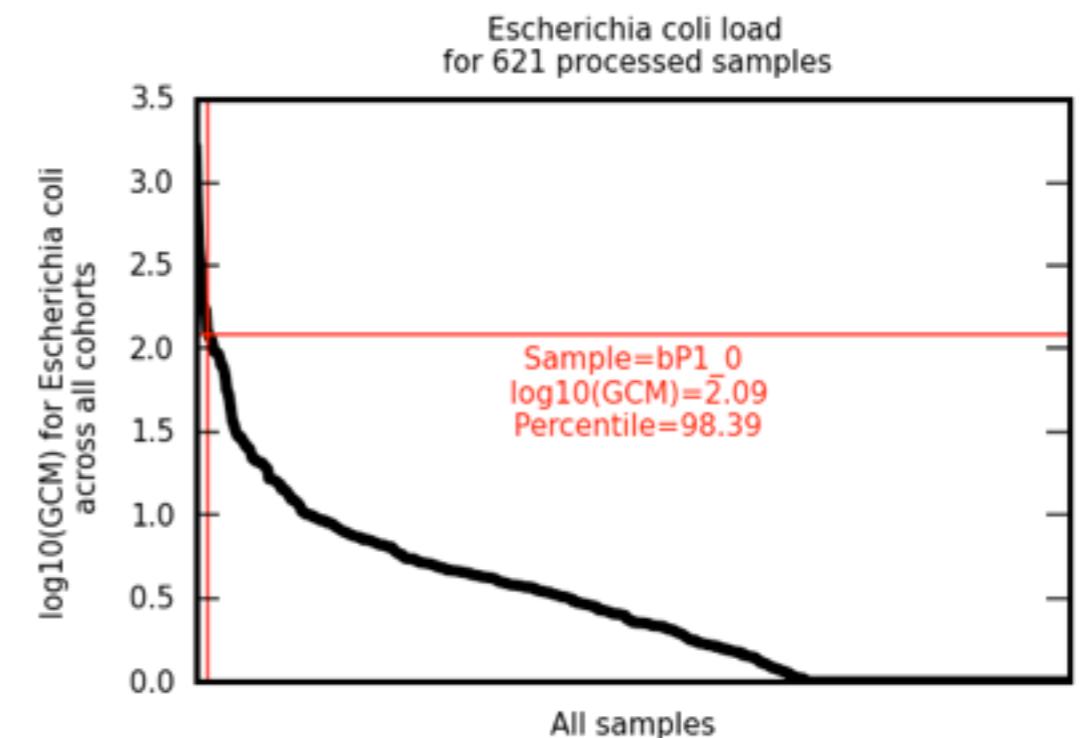
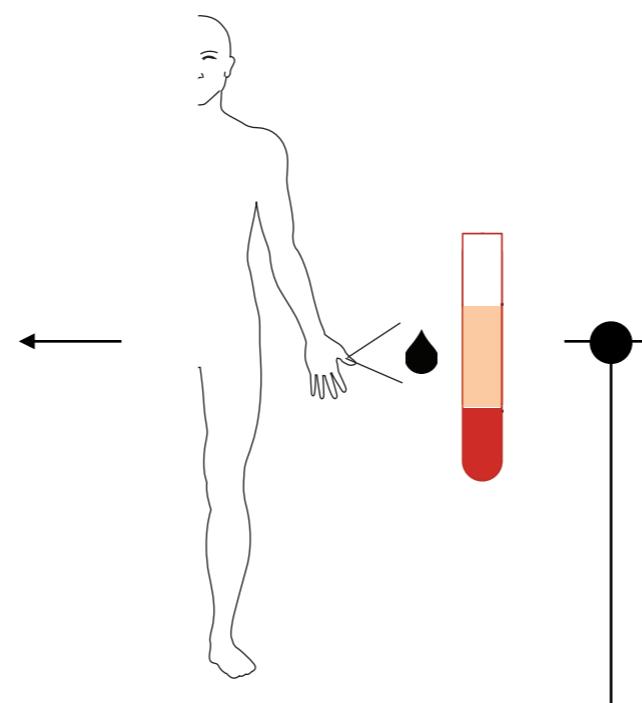
# Performance on viruses.



# Performance on microbes (detected in deep tissues).

Stanford Pathology -  
Biopsy and MALDI-tof

E. Coli



Ranking of all bugs detected in the sample -

## Sorted infection data for bP1\_0.

Use back to return to cohort or dropdown menu above to switch between cohorts.

Show: 10

Search:

Name

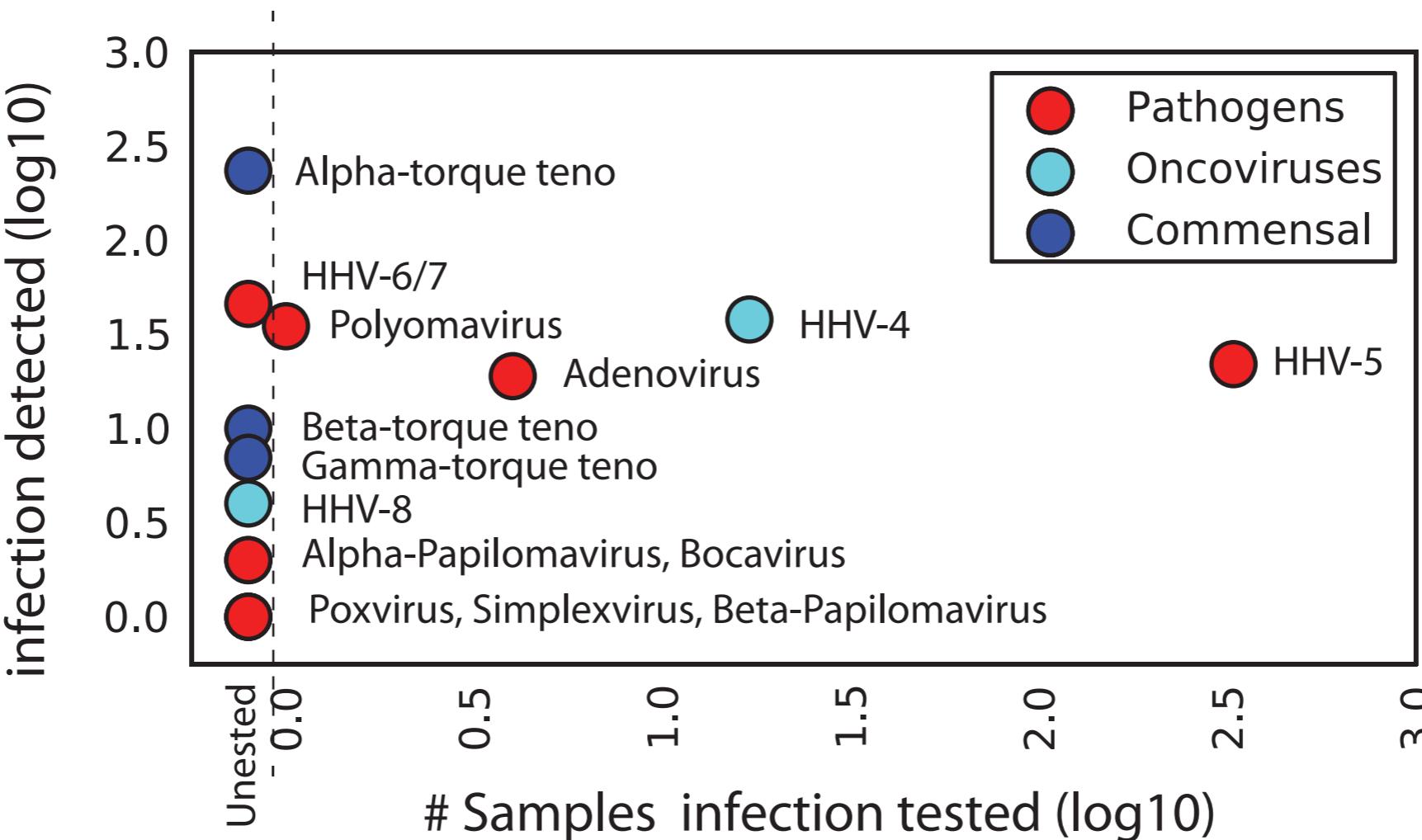
Gen\_Equ

Percentile

Candidatus Midichloria	0.28	99.516908
P2likeviruses	34.09	98.711755
Gammatorquevirus	189.43	98.711755
Escherichia	123.17	98.380694
Alkaliphilus	0.11	98.228663
Shigella	8.45	96.940419
Chelatovorans	0.11	95.169082
Enterocytozoon	0.15	94.363929
Citrobacter	0.31	93.236715
Psychrobacter	0.39	92.914654

# Un-tested / diagnosed infections.

Many potential pathogen that we detected are infrequently clinically tested for.



# Undiagnosed cases of infection.

I6, Cause of death:  
**Respiratory failure.**



## Sorted infection data for I6.

Use back to return to cohort or dropdown menu above to switch between cohorts.

Show: 10

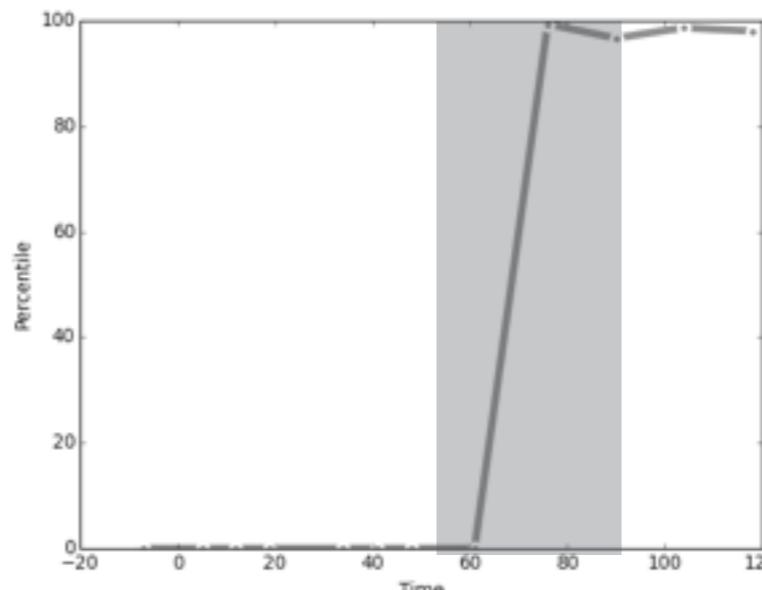
Search:

Name	I6_BC	I6_D-1	I6_W1	I6_W2	I6_W3	I6_W4	I6_W5	I6_W6	I6_W8	I6_W10	I6_W12	I6_W14	I6_W16
WU Polyomavirus	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	99.337748	96.688742	98.675497	98.013245
Human herpesvirus 5	0.000000	75.496689	88.079470	78.145699	0.000000	90.096225	80.132450	89.403974	91.390728	0.000000	82.781457	0.000000	0.000000
Enterocytozoon bieneusi	0.000000	0.000000	90.728477	95.364238	94.039735	0.000000	96.026490	98.013245	98.675497	99.337748	94.701987	92.715232	0.000000

WU Polyomavirus  
(Rare virus that causes severe respiratory infection)

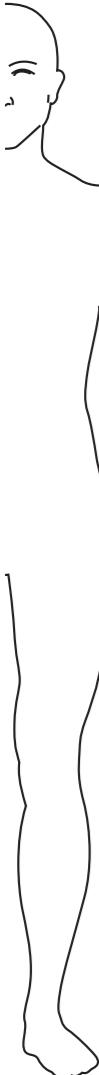


Very high load of WU Polyomavirus during time period of negative clinical test results.

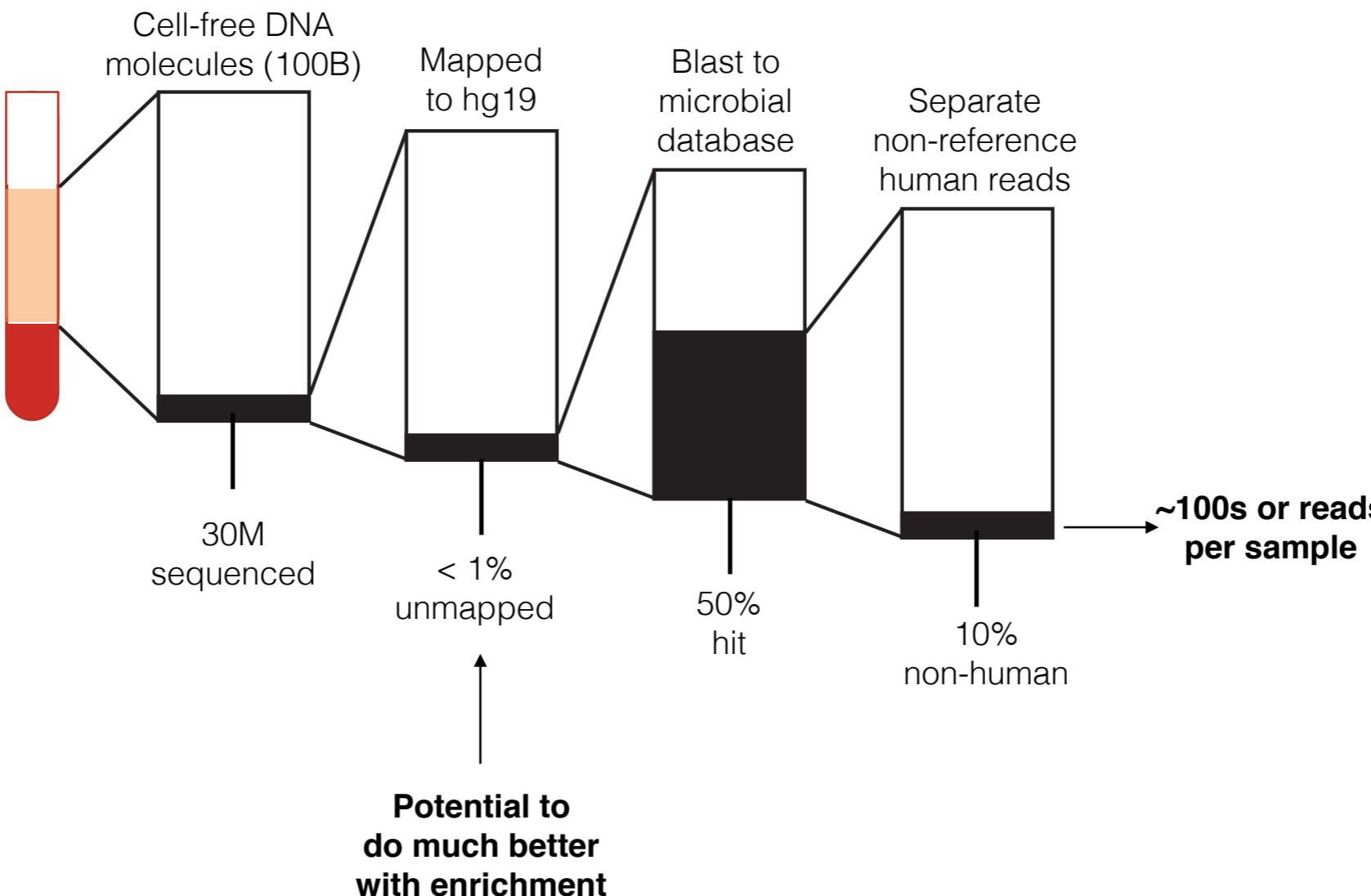


Multiple (-) tests for  
BK polyomavirus

# Also, headroom for improvement.



Favorable results shown with sparse sampling of microbial reads.



**Clinical studies -**

- Lung
- BMT

**Deep tissues -**

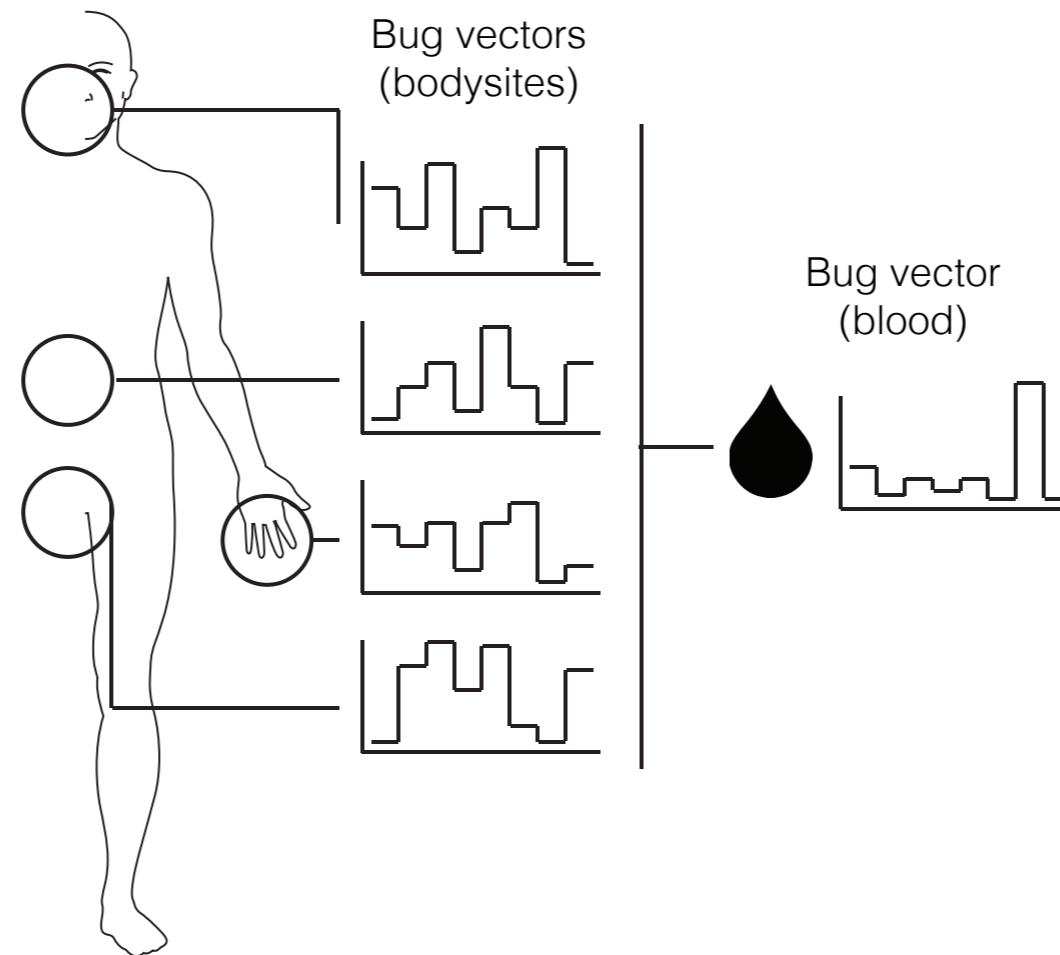
- Biopsy replacement

**Undiagnosed**

- Viruses (I6)
- Fungi (L78)

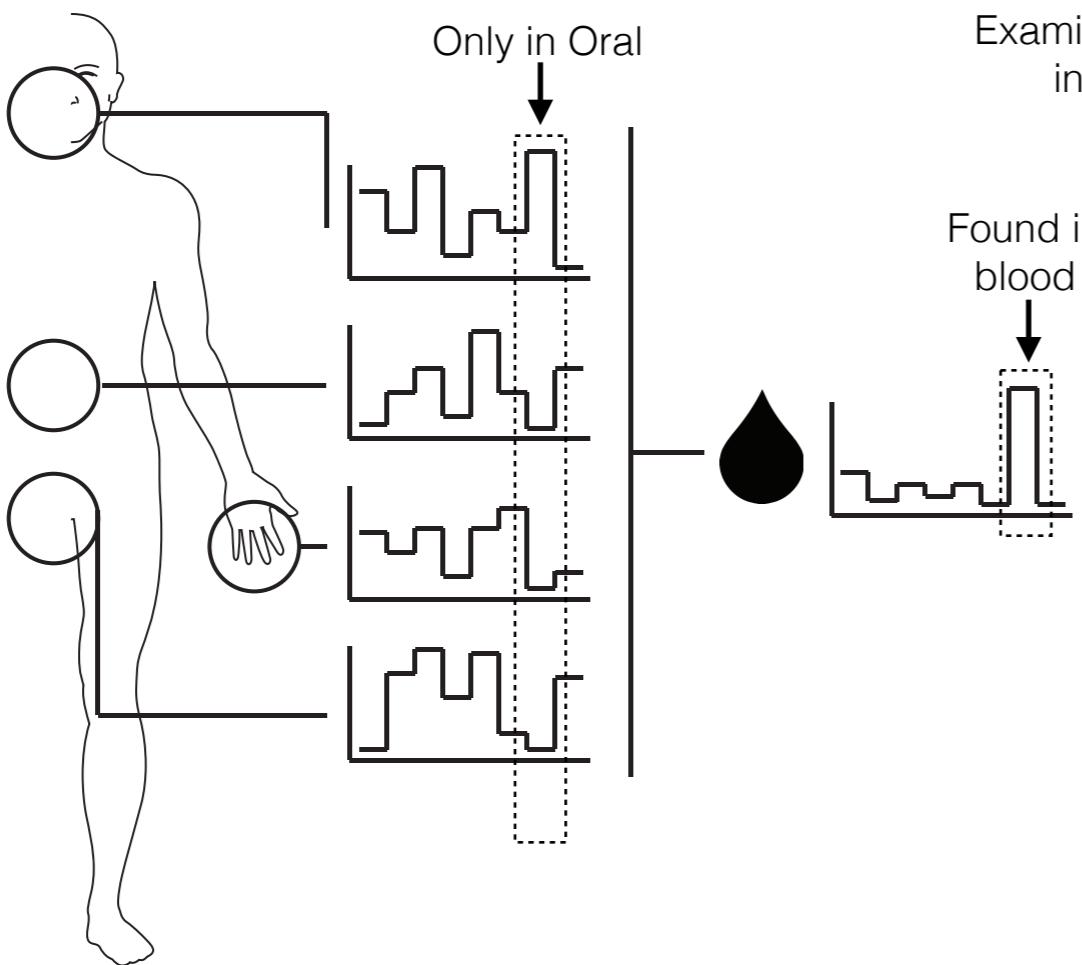
# But, where does microbial cell-free DNA come from?

Examine a pregnancy cohort of with microbiome sequencing at 4 body sites with matched blood samples -

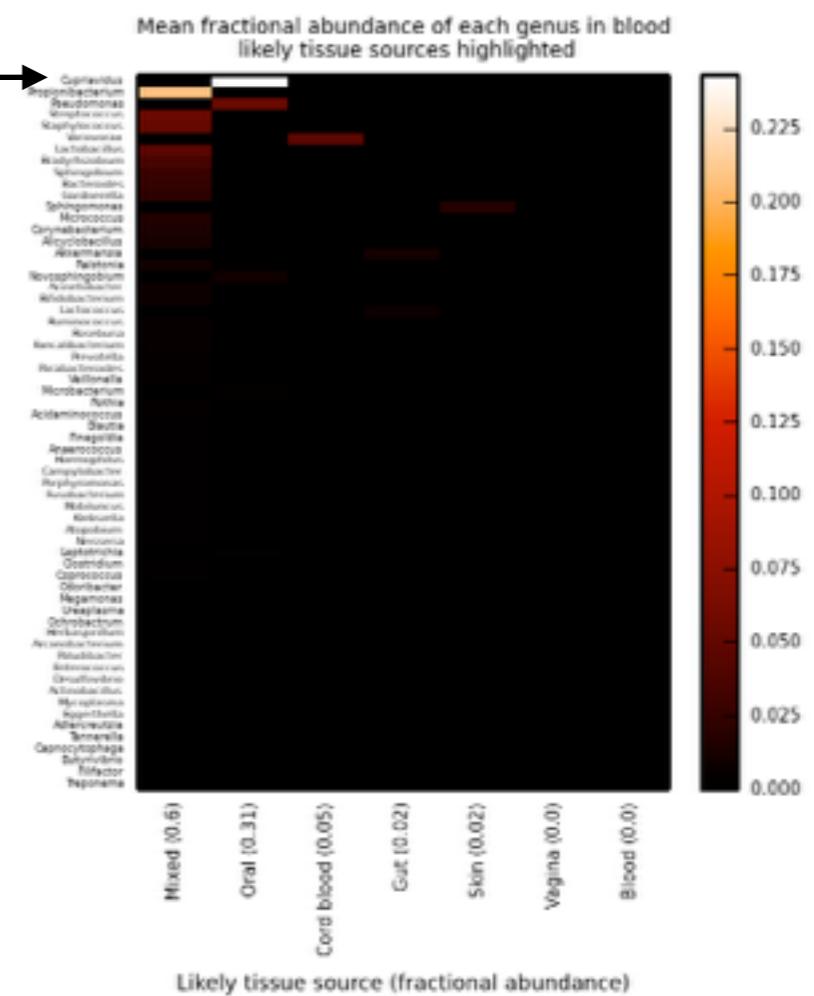


# Tissue specific bugs don't explain blood composition.

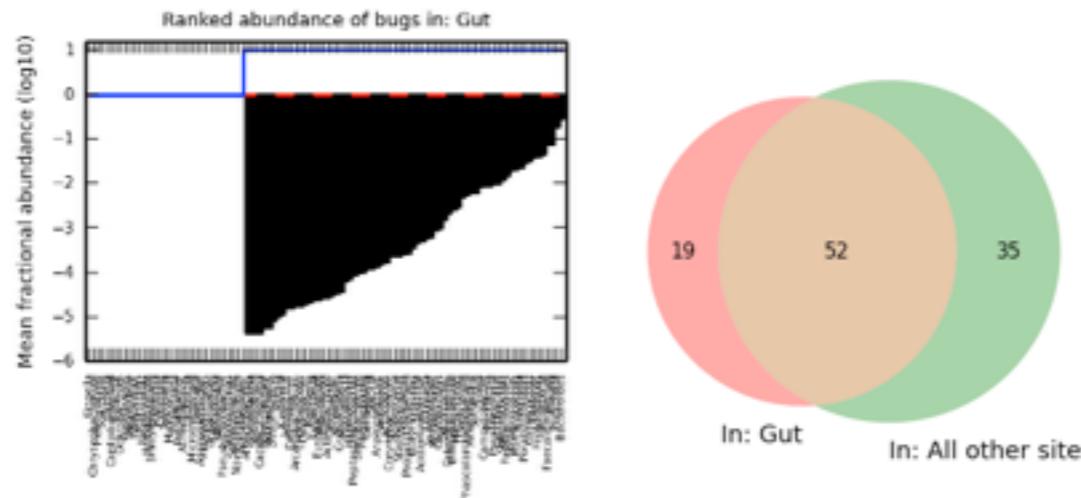
Define tissue specific bugs -



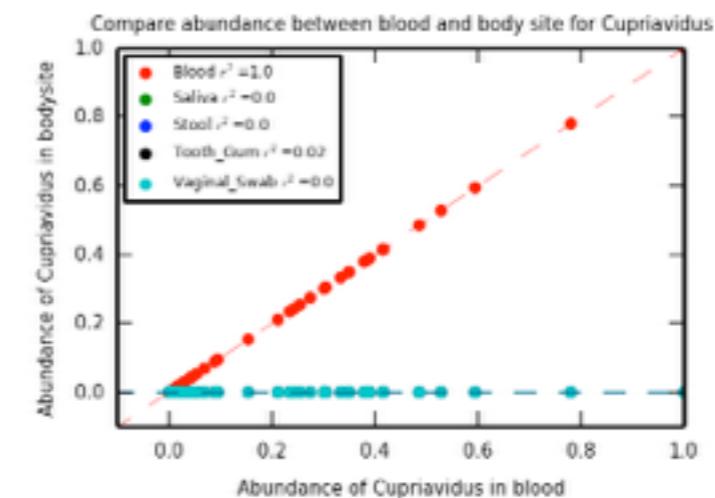
Assign likely source to each bug found in blood -



Bugs at each site are re-cast as bit vectors -

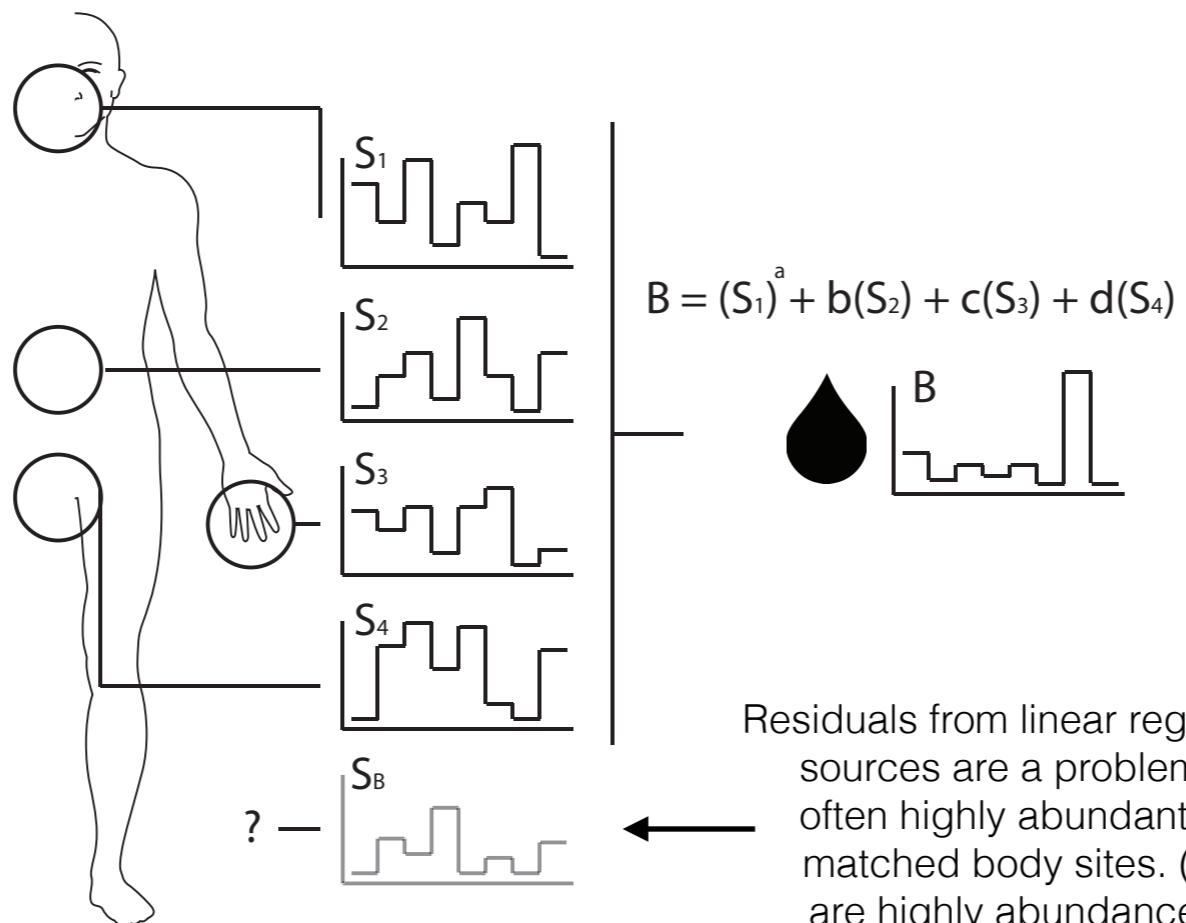


Cupriavidus has trace abundance in other tissues, but high in blood -



# ... Neither does linear model of commonly sampled sites.

The four sampled body sites are insufficient to model blood, as it likely draws from additional tissue sources -



Residuals from linear regression [see appendix] show that blind sources are a problem: (1) Bugs such as *Curprividius* are often highly abundant in blood sample, but not present at matched body sites. (2) Bugs such as *Propionibacterium* are highly abundance in skin, but skin is rarely sampled.

# Try approach that is not constrained to sampled sites.

The cocktail party problem -



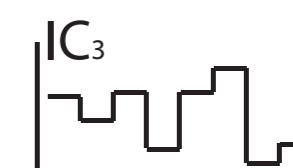
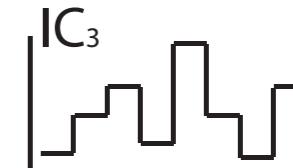
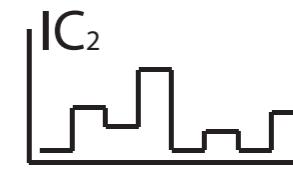
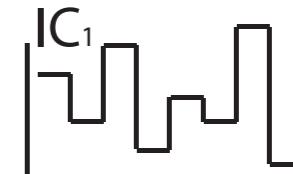
Toy problem for source de-convolution from mixed signals -

- Independent sources (voices)
- From recorded mixed conversations, it is possible to isolate the sources.
- ICA algorithm used to do this (see derivation in appendix).

# Blood microbiome as a cocktail party problem.



Source vectors (bugs) -

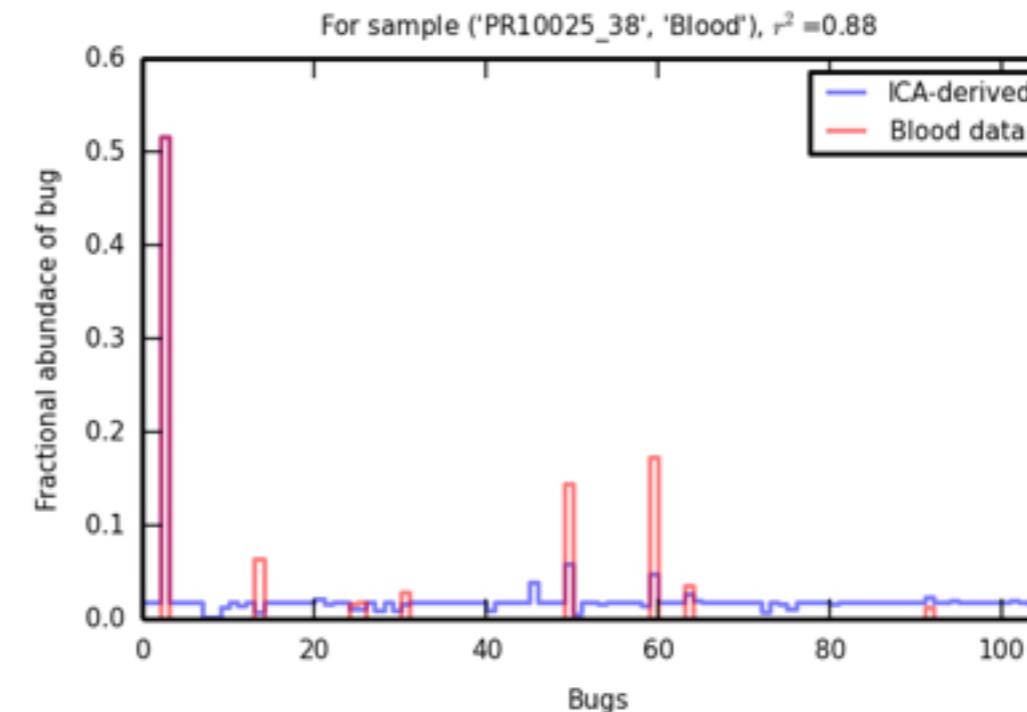
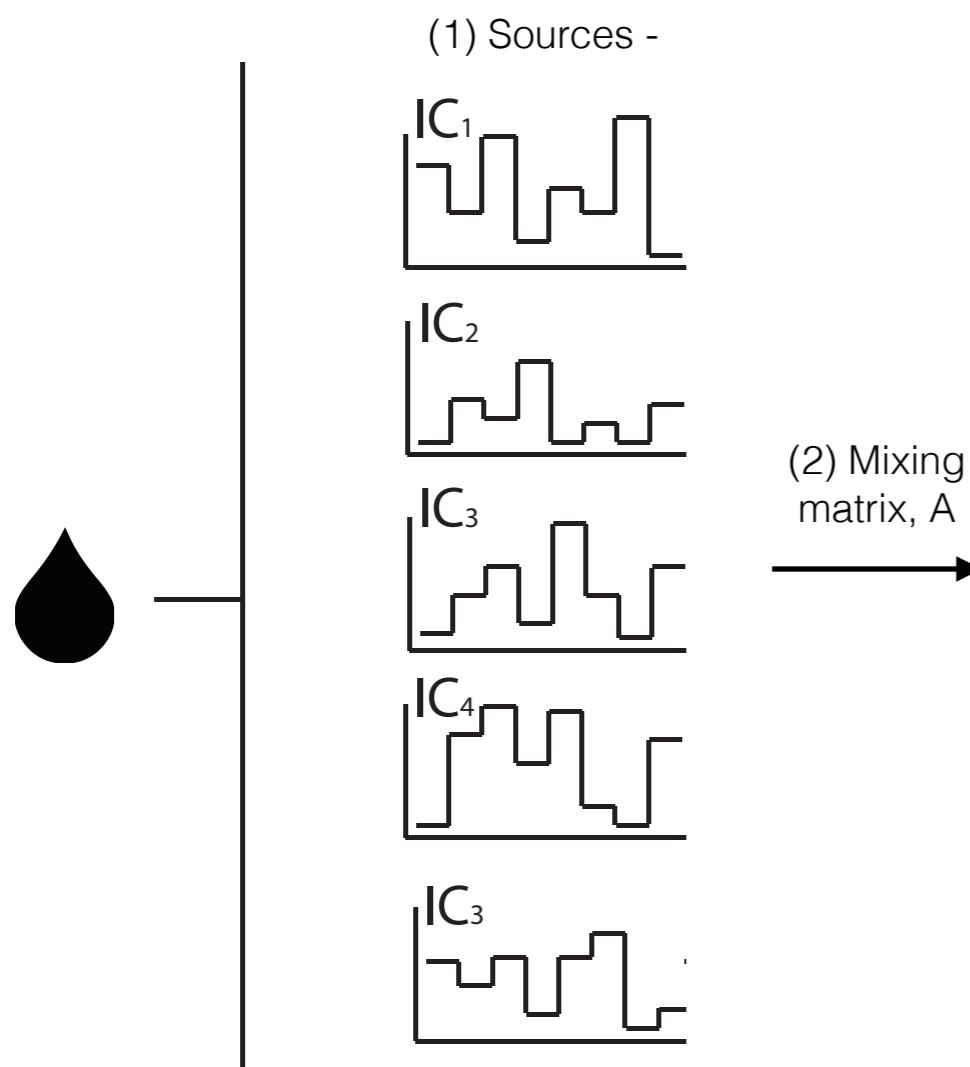


- Independent sources (body sites)
- Measured mixtures (blood samples)
- Use ICA to isolate the body site sources

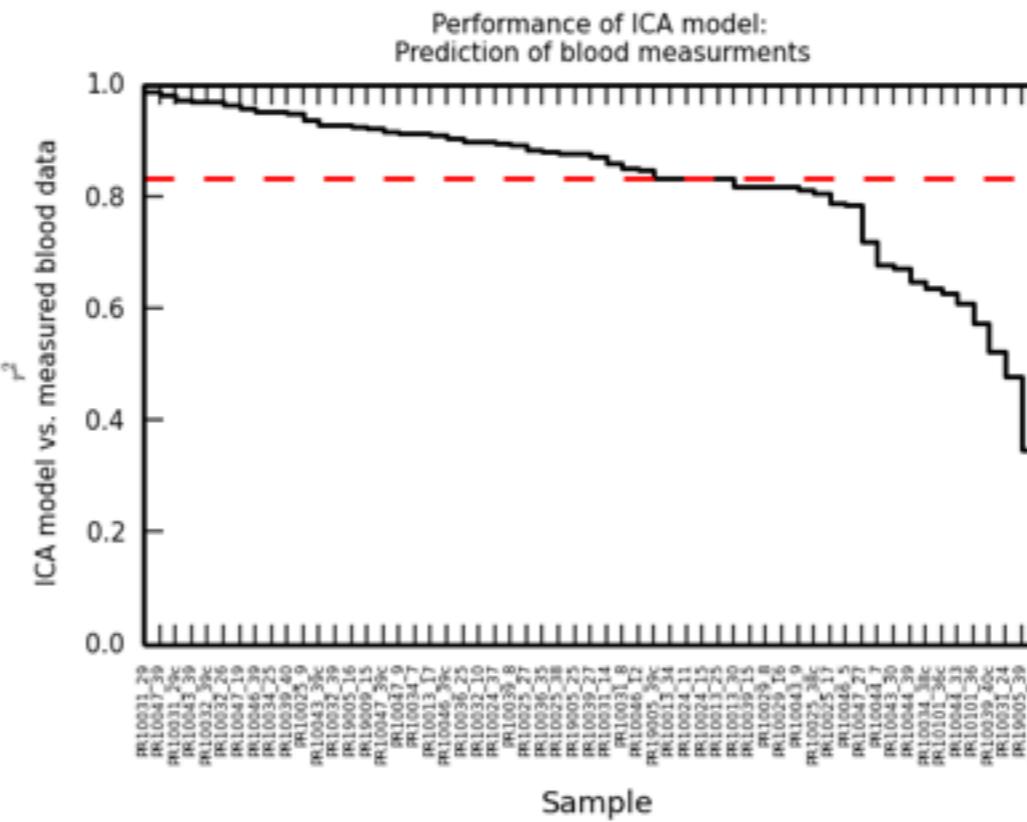
# Unsupervised source detection in pregnancy cohort.

Blood data for any sample re-capitulated using ICA sources and mixing matrix -

$$B_k = A_{k1}(IC_1) + A_{k2}(IC_2) + A_{k3}(IC_3) + A_{k4}(IC_4) + A_{k5}(IC_5)$$

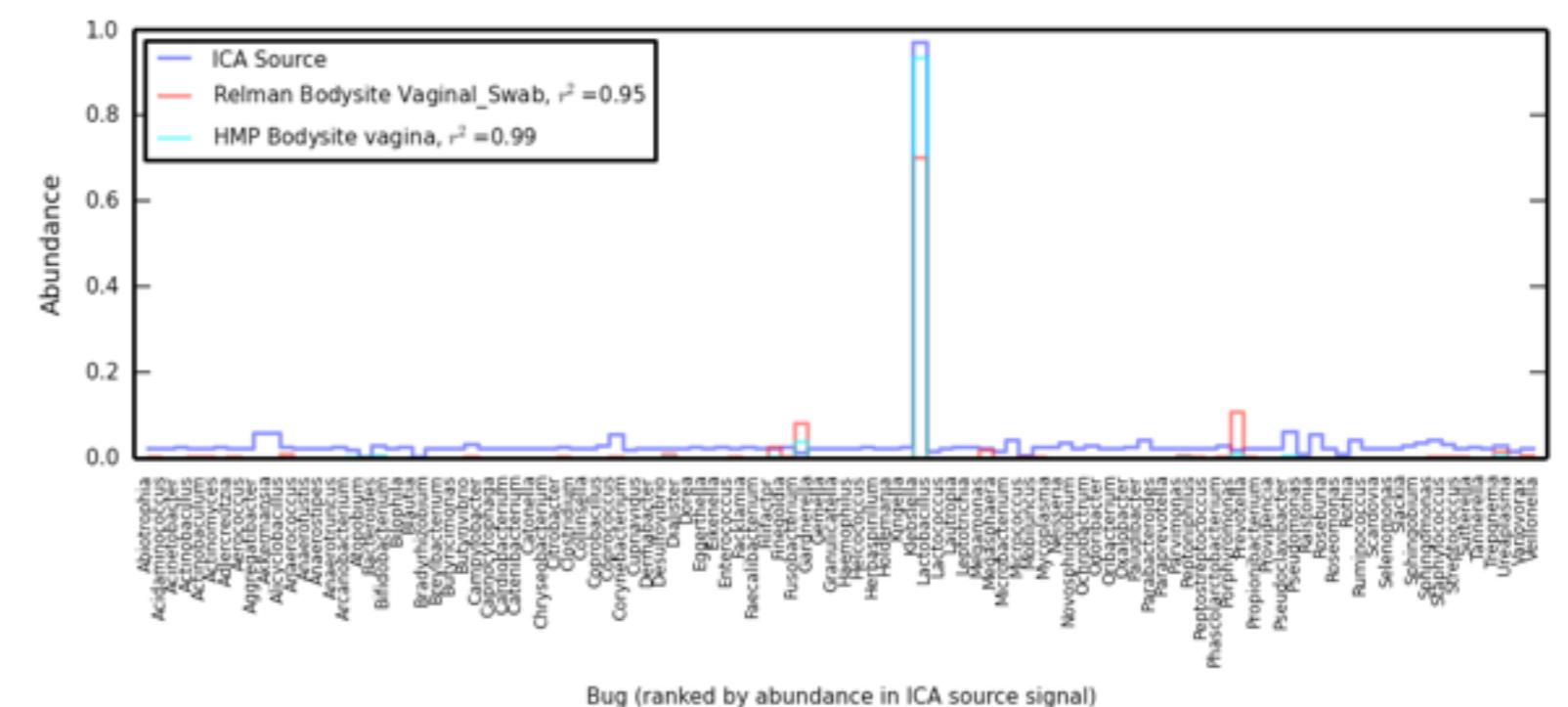
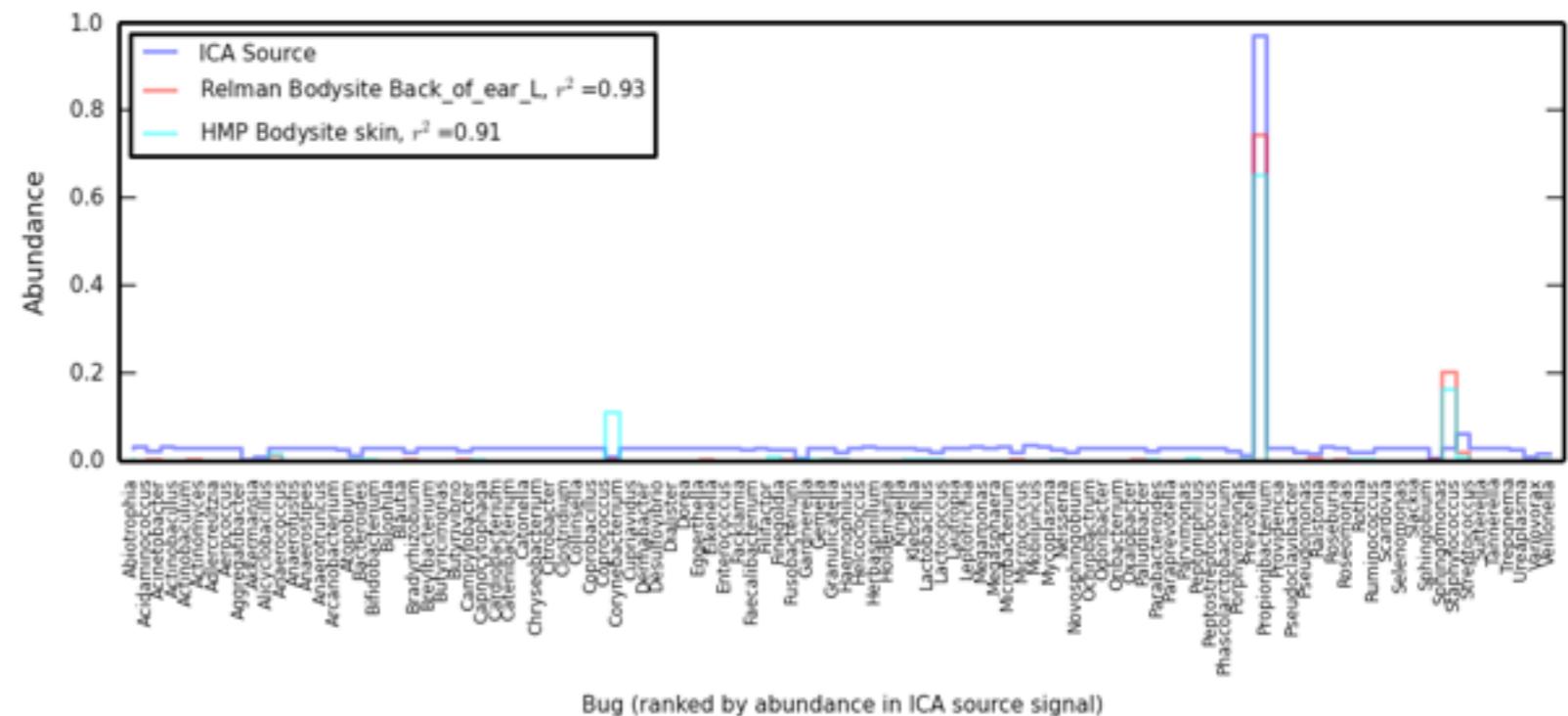
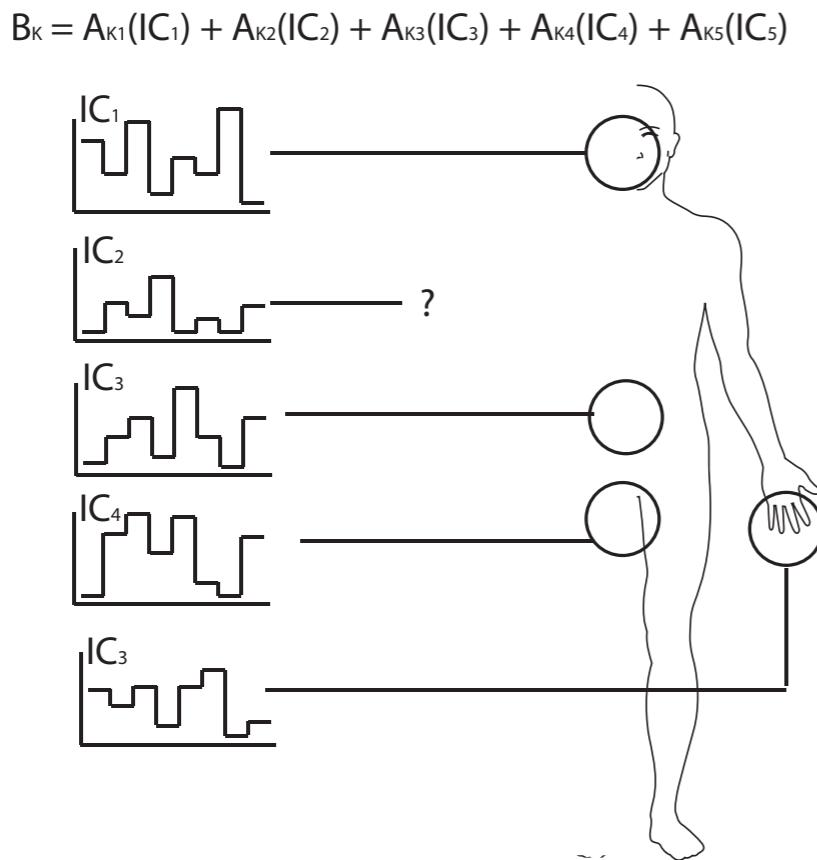


All samples (with mean correlation highlighted) -



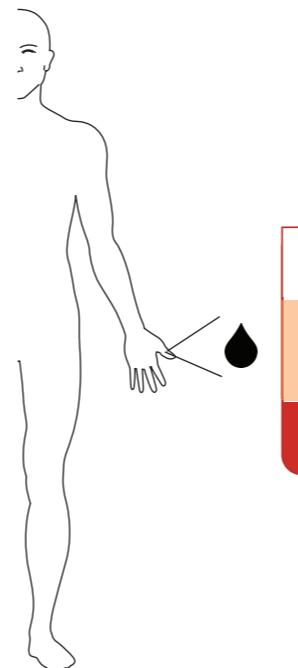
# Correlate ICA sources with composition of body sites.

We can then try to correlate “discovered” ICA sources with known tissues using HMP and Relman data -

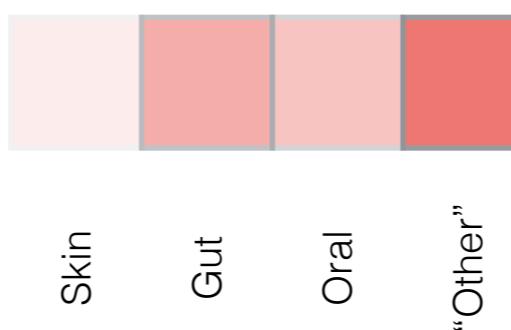


# Each blood sample is a combination of ICA sources.

The mixing matrix tell us the weight of each source per sample.



Relative weight  
of ICA sources:

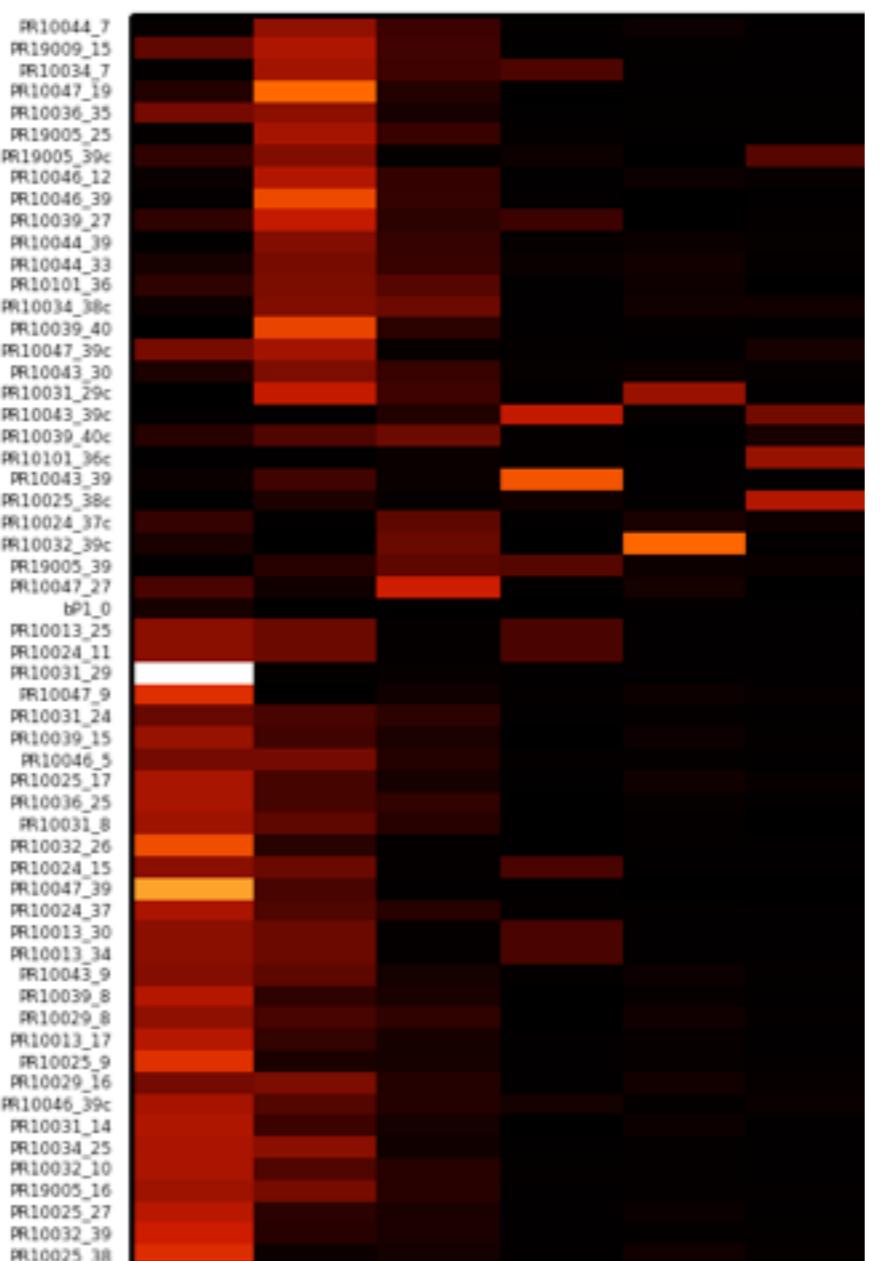


# Mixing matrix for pregnancy cohort shows blood sources.

Cluster (1): Enriched in skin-like signal (high in *Propionibacterium*), which may come from skin or a different (unsampled) deep tissue.



Cluster (2): Enriched in source with high load of *Cupriavidius*, which has no body site association (literature says it may grow in blood, so possible blood commensal).



Dominant sources found in pregnancy cohort.

↑  
Blind: Cupriavidius

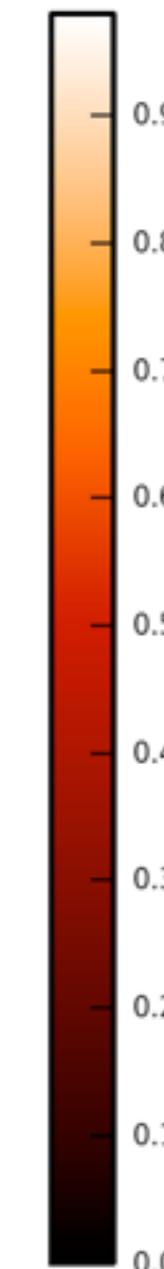
↑  
Skin

Oral

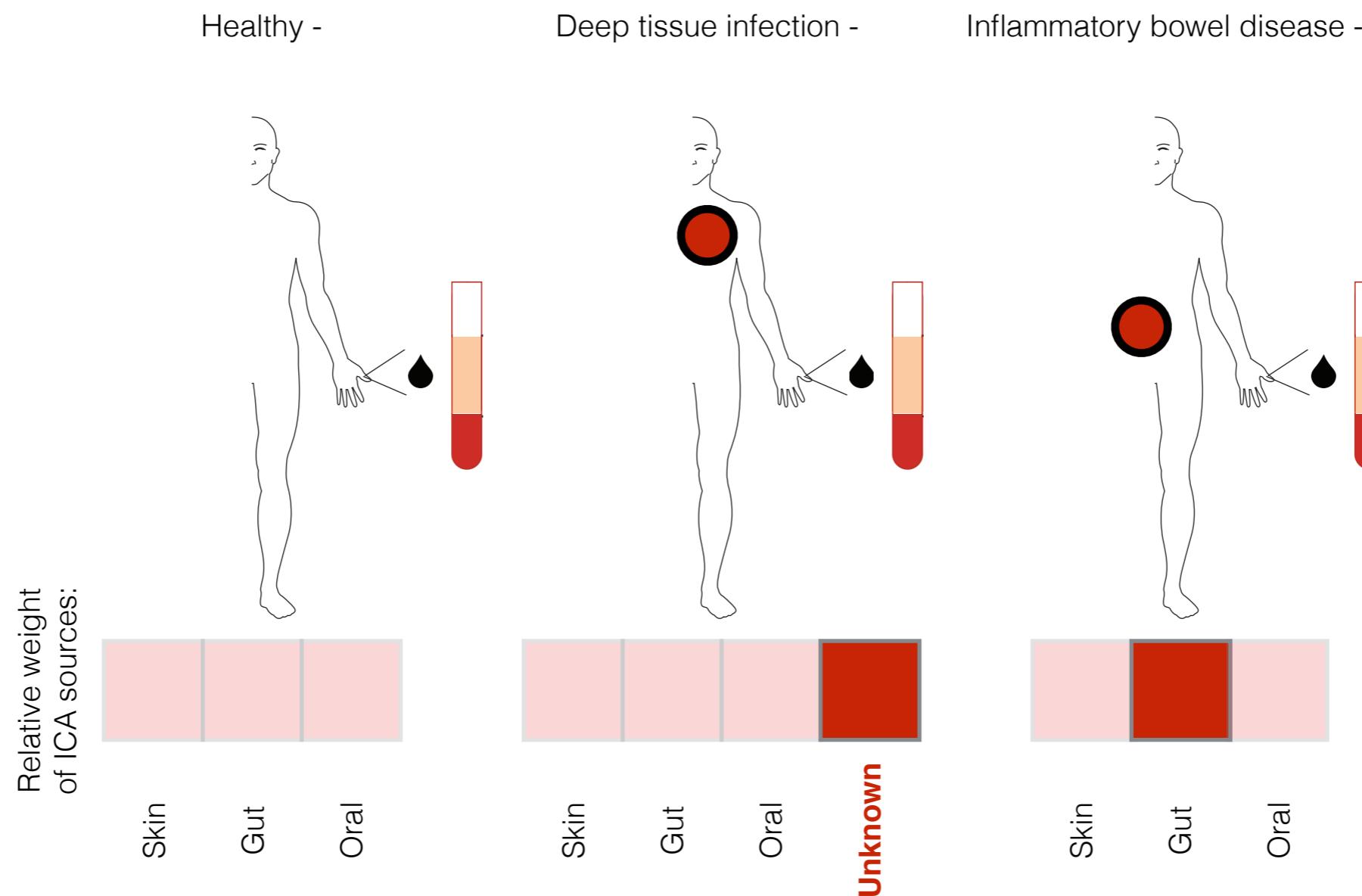
Vagina

Blind: Gardnerella

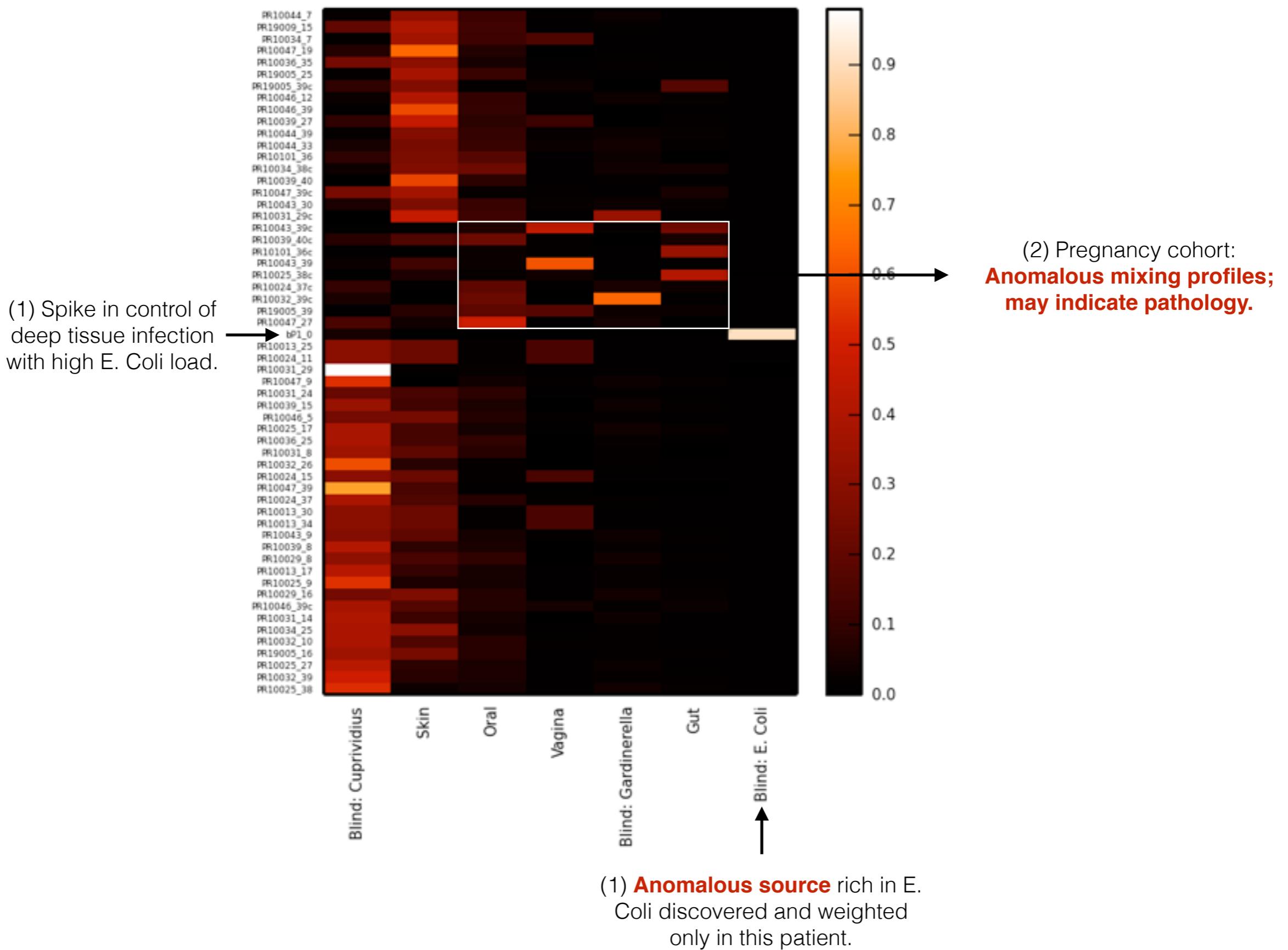
Gut



# Works for anomaly detection in cfDNA microbiome data?



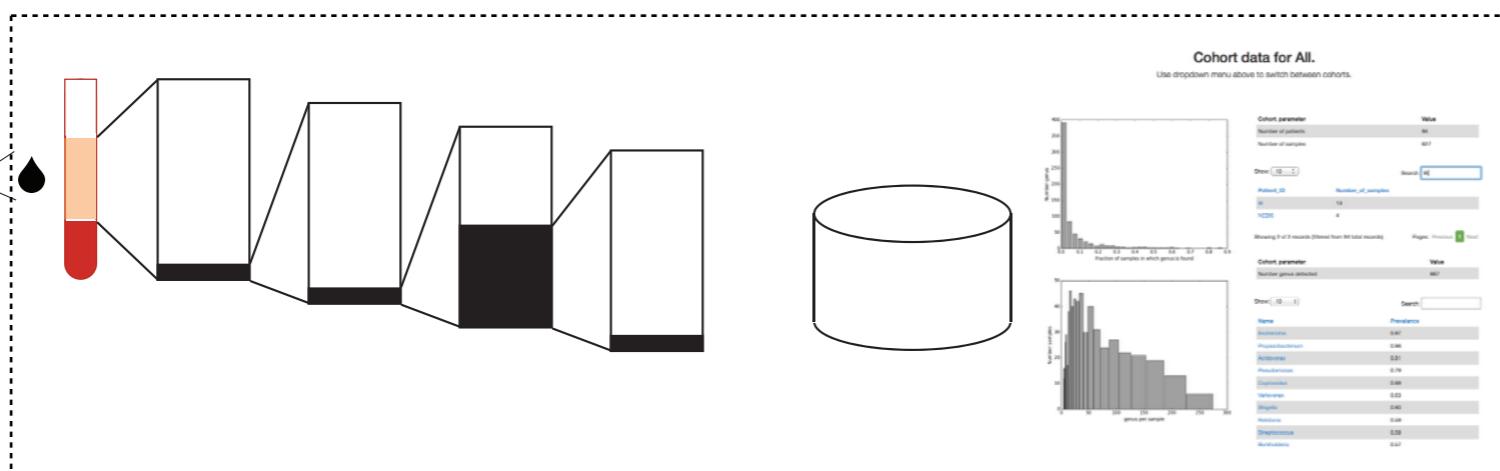
# Mixing matrix can pinpoint anomalies.



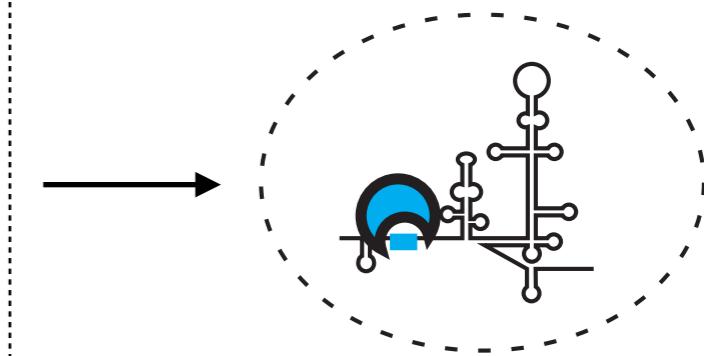
# Sequencing also as a tool to study infection mechanism?



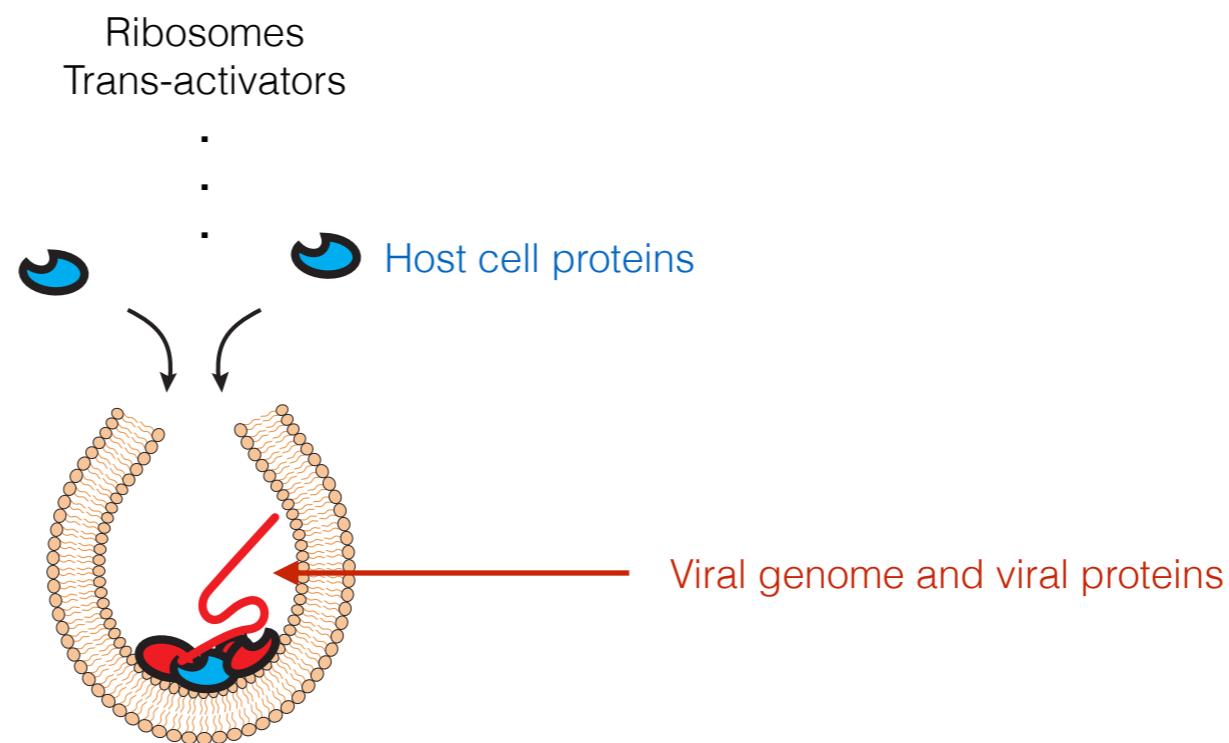
Sequencing as diagnostics -



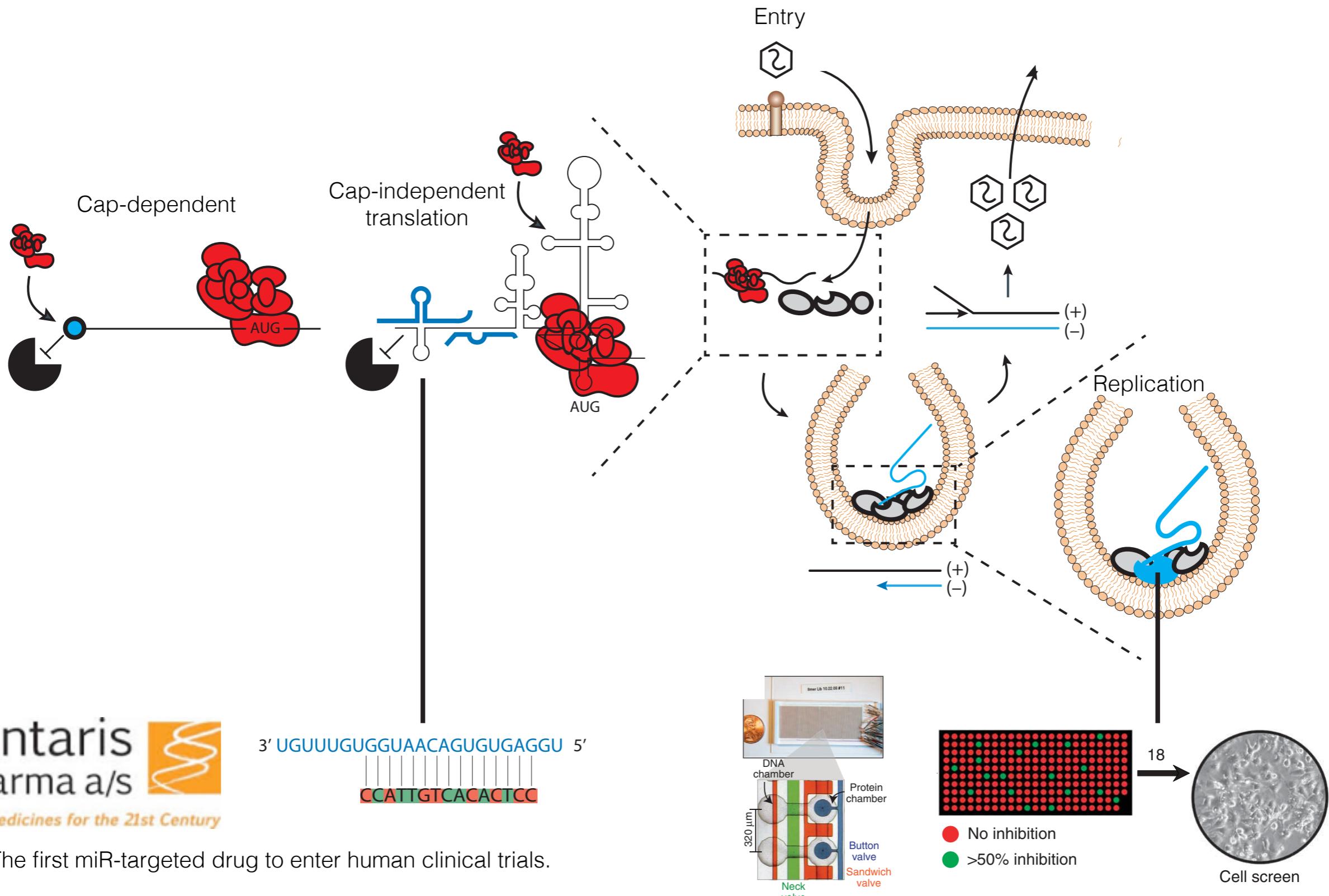
Mechanism relies on interactions -



# Viral recruitment of cell host factors is critical.



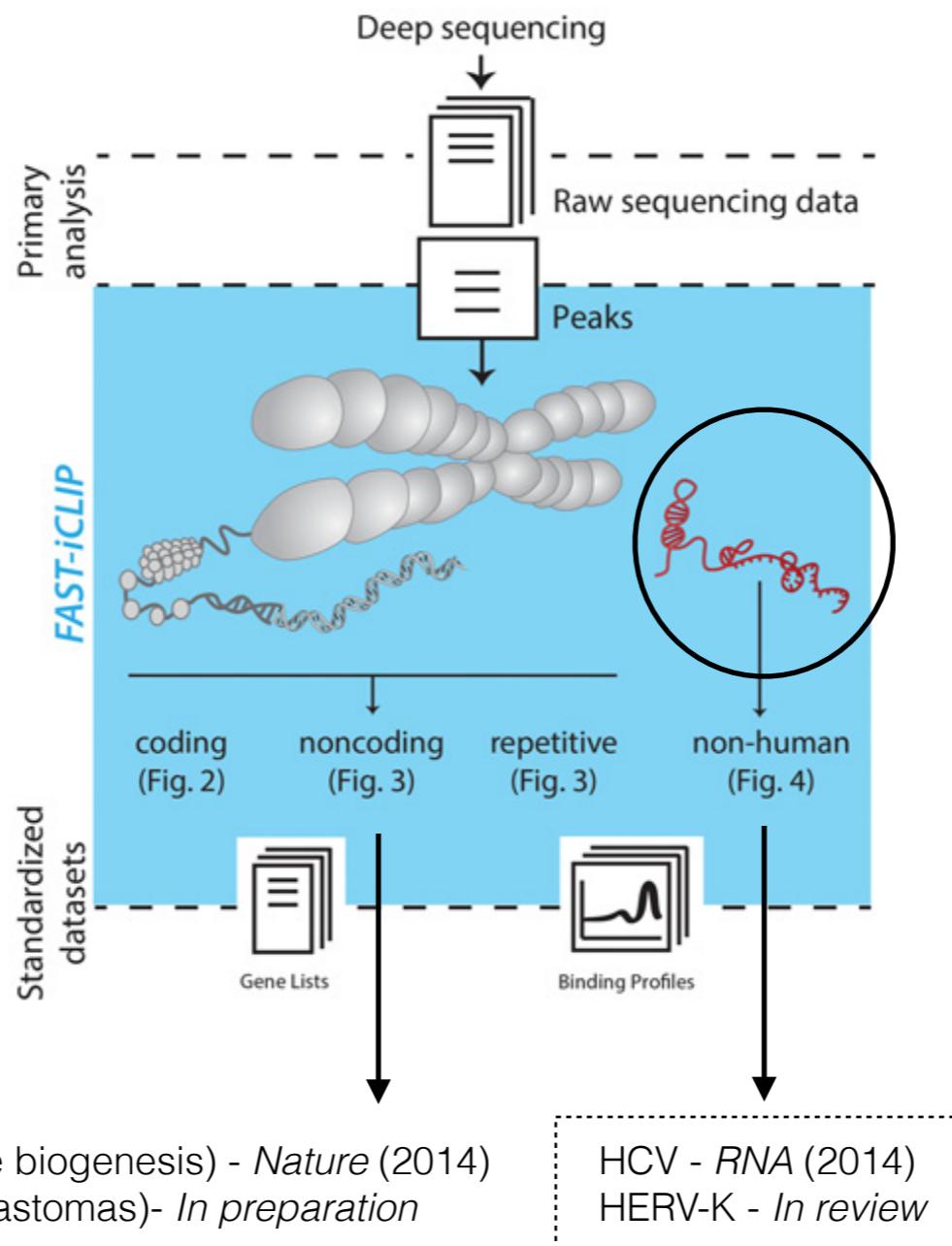
# Human host proteins required Hepatitis C virus lifecycle.



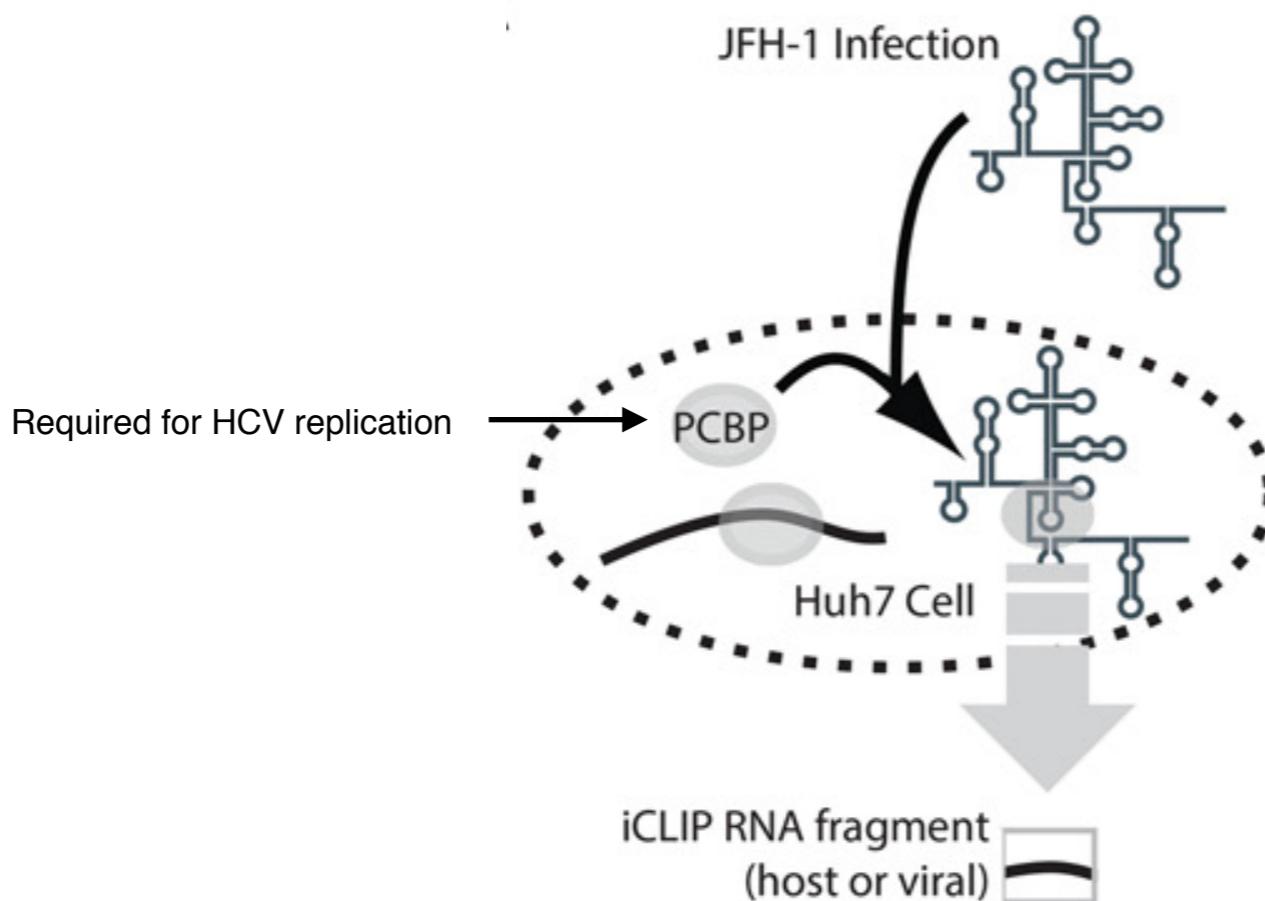
santaris  
pharma a/s   
*RNA Medicines for the 21st Century*

The first miR-targeted drug to enter human clinical trials.

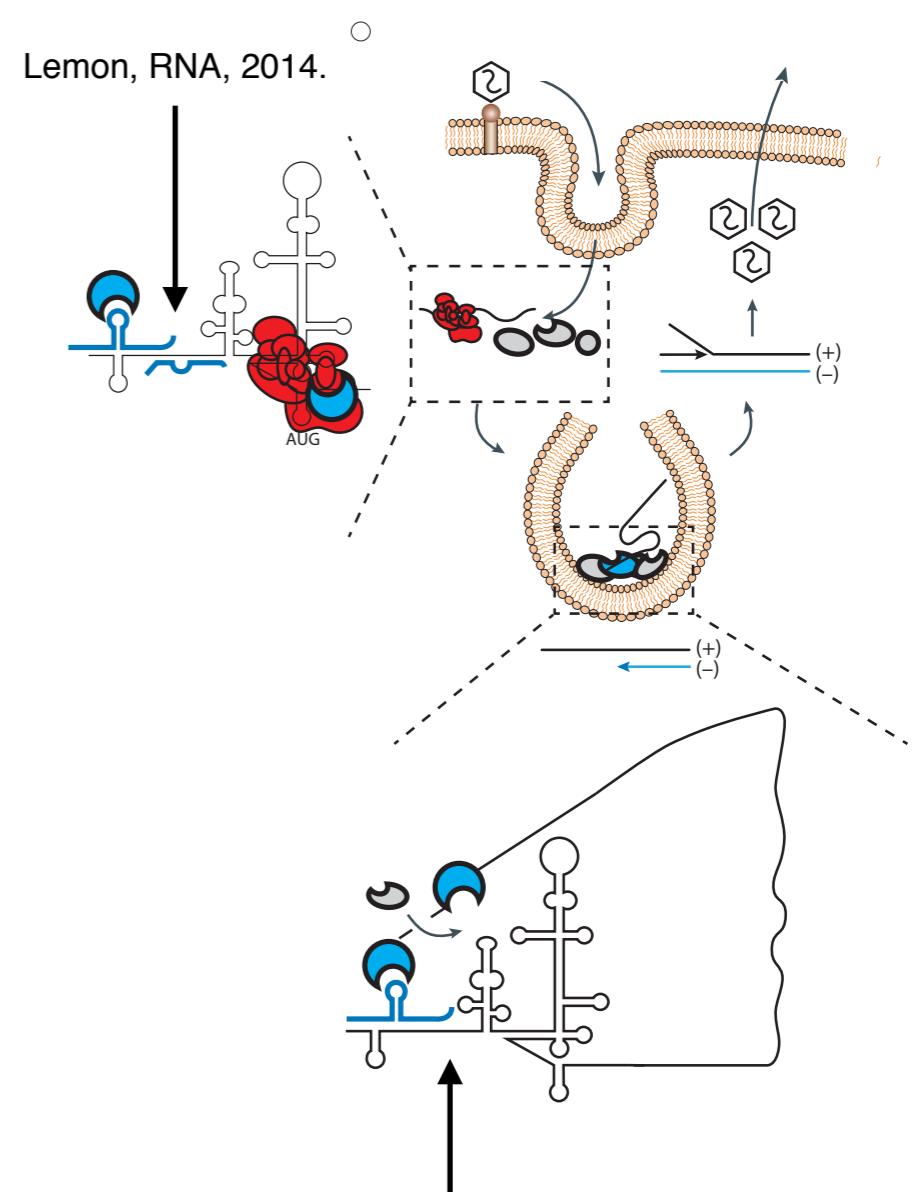
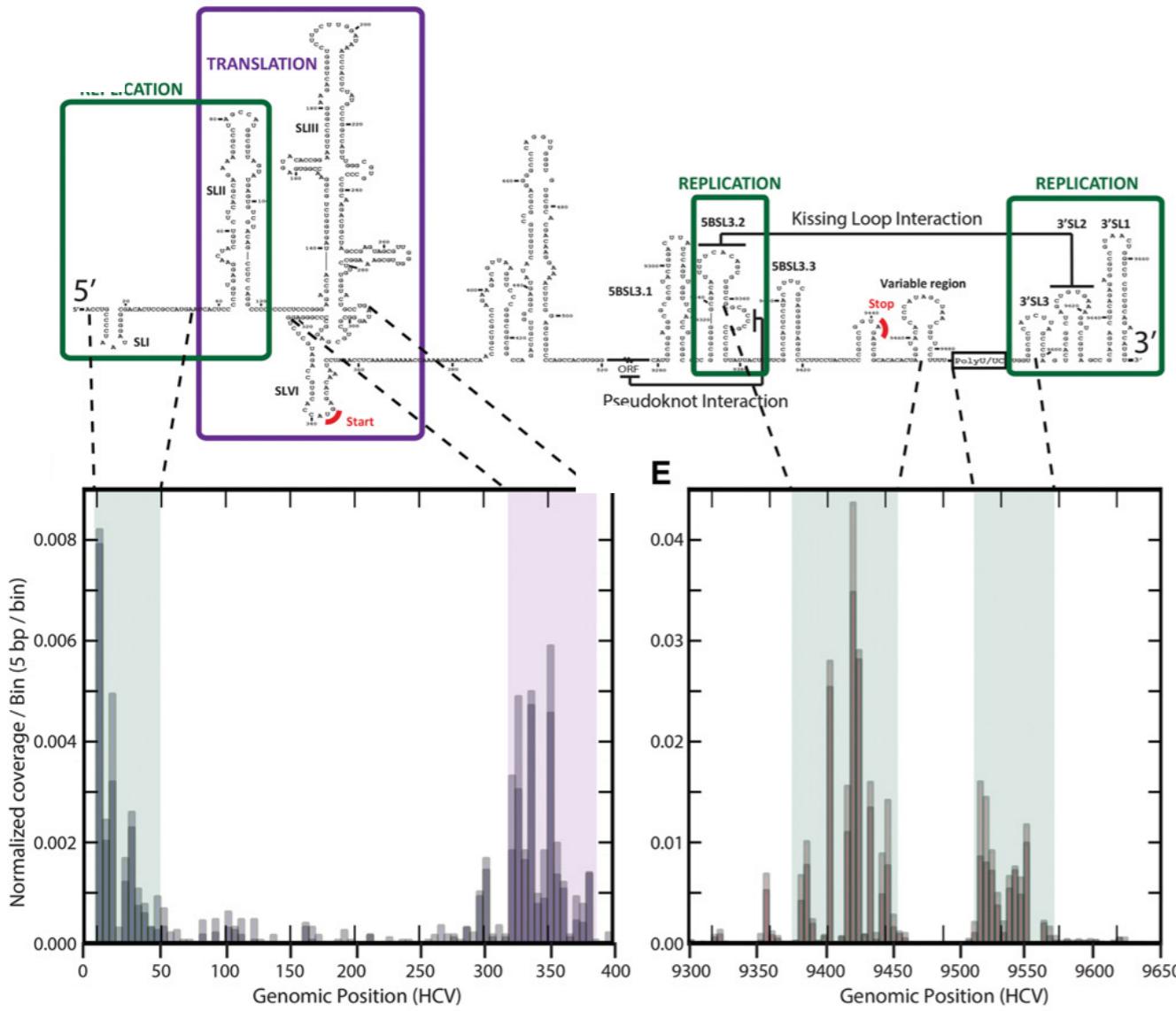
# Pipeline for RNA-protein interactome of viruses.



# Apply to HCV viral-host factor interactome.



# New insights about HCV lifecycle.

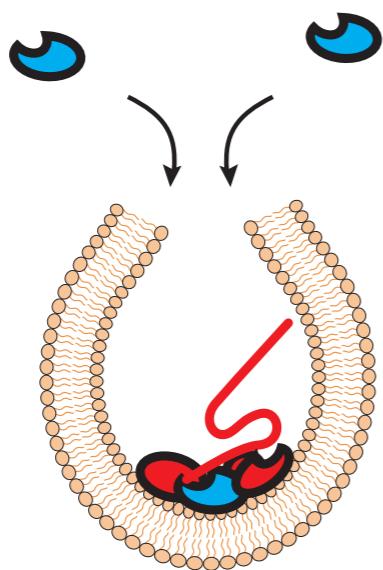


Wang et al, J. Virology, 2011.

# But, what about viral products interacting with host?

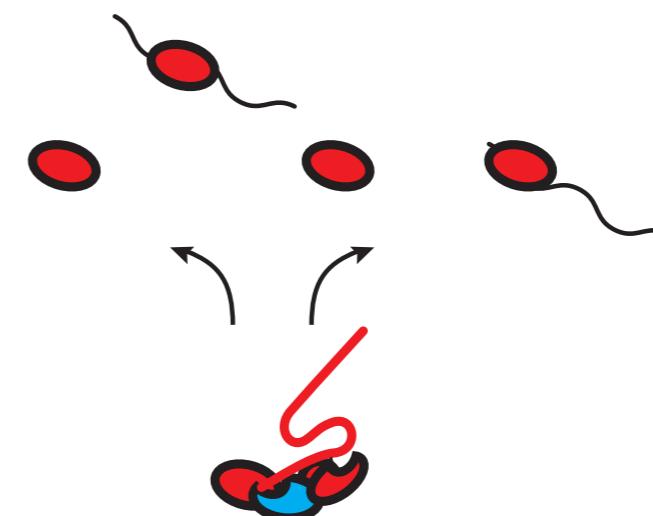
Host factor co-option -

- HIV
- Ebola
- HCV
- etc.

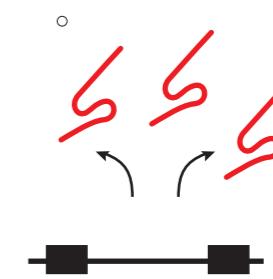
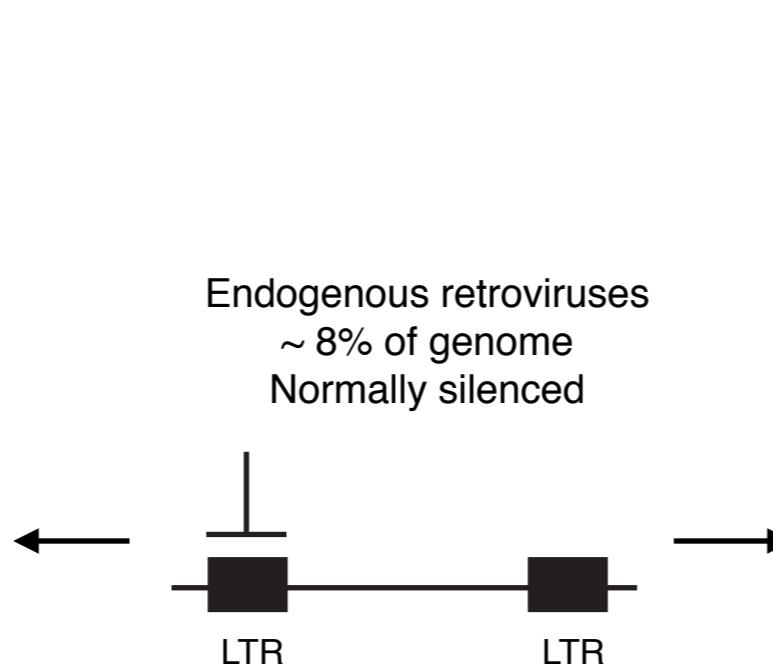
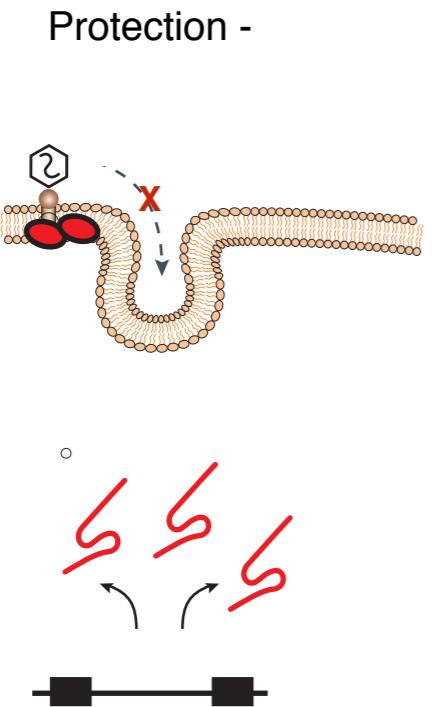


Viral protein interactome -

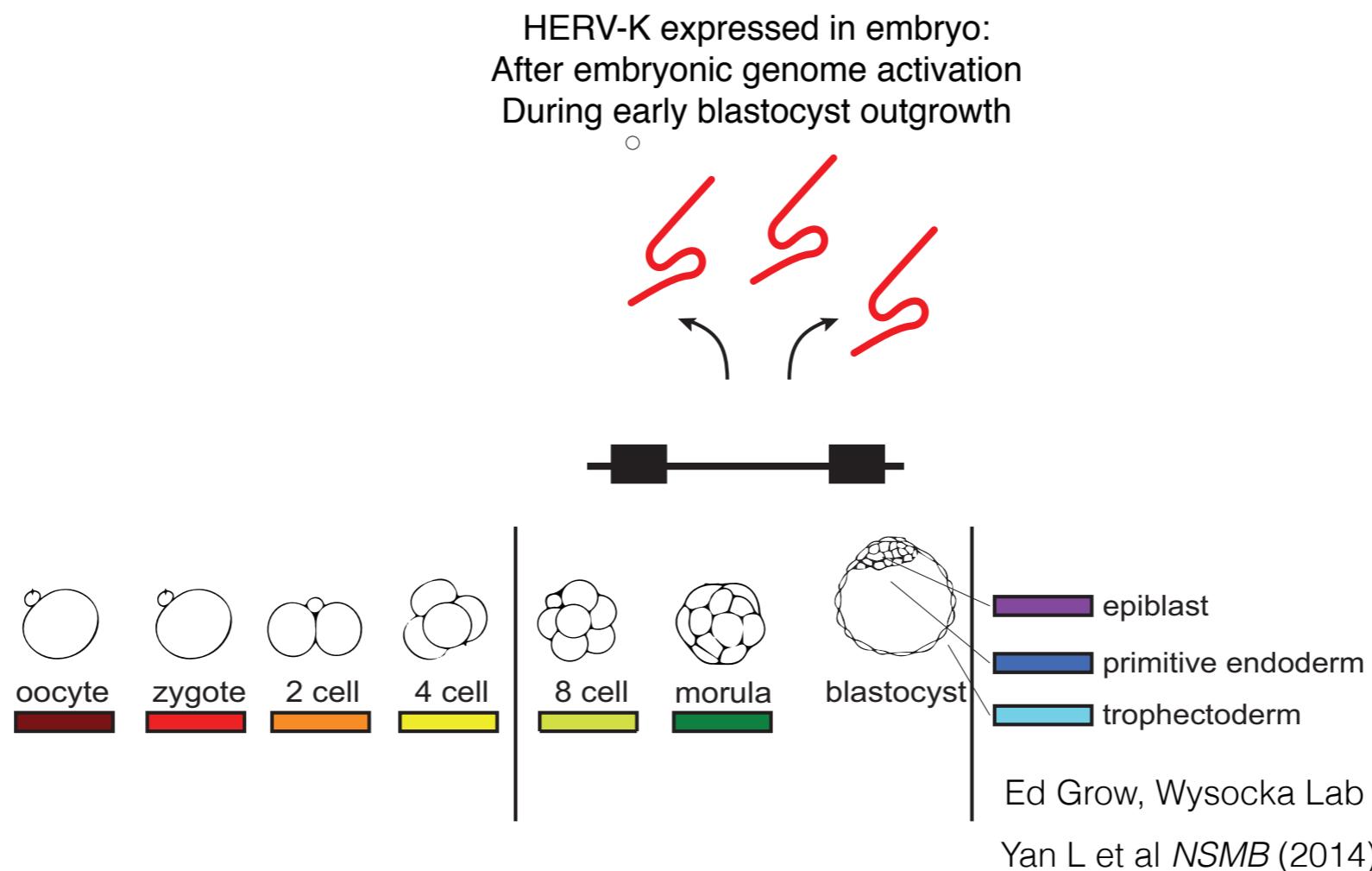
- Interaction with host?



# Resurrection of endogenous retroviruses.

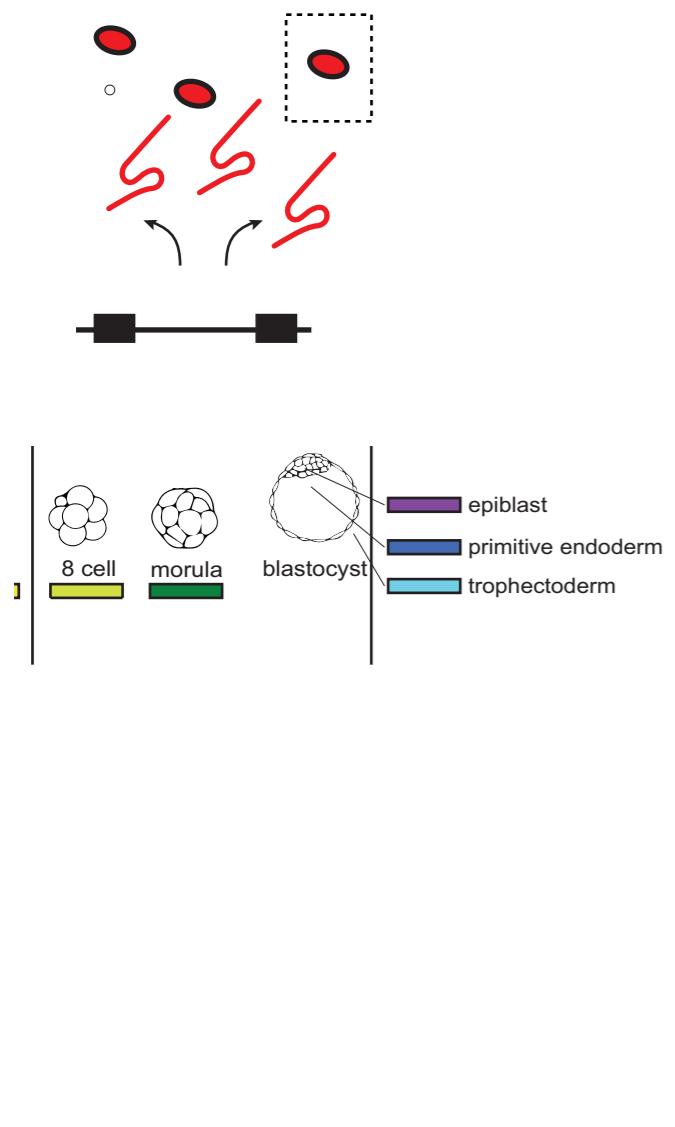


# HERV-K retrovirus detected in signal cell embryo RNA-seq.

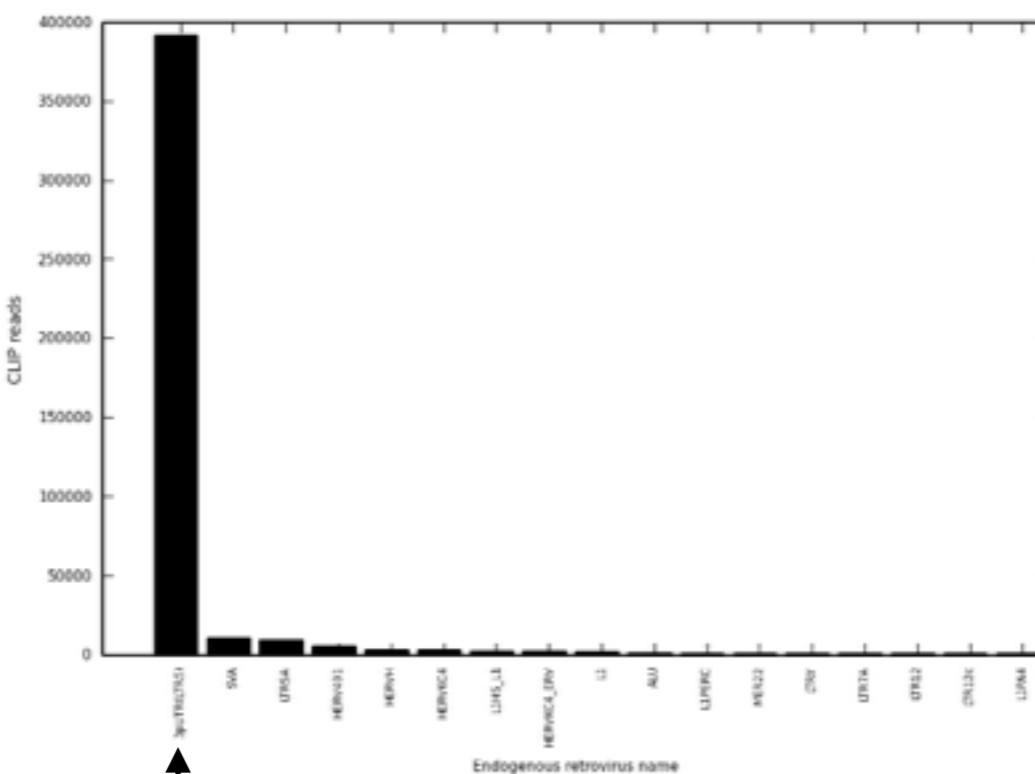


# CLIP shows retroviral proteins active in embryo.

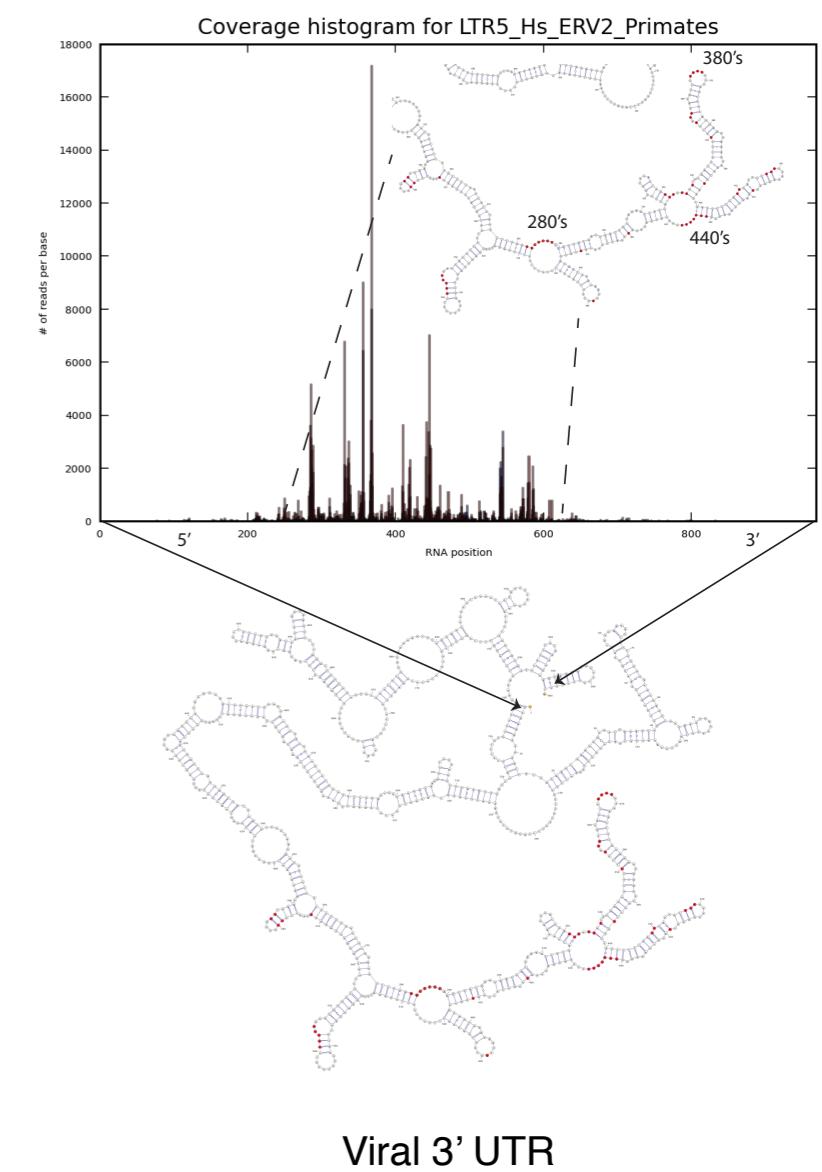
CLIP Rec, a viral protein  
required for nuclear export in  
embryonic carcinoma cells (hECCs)



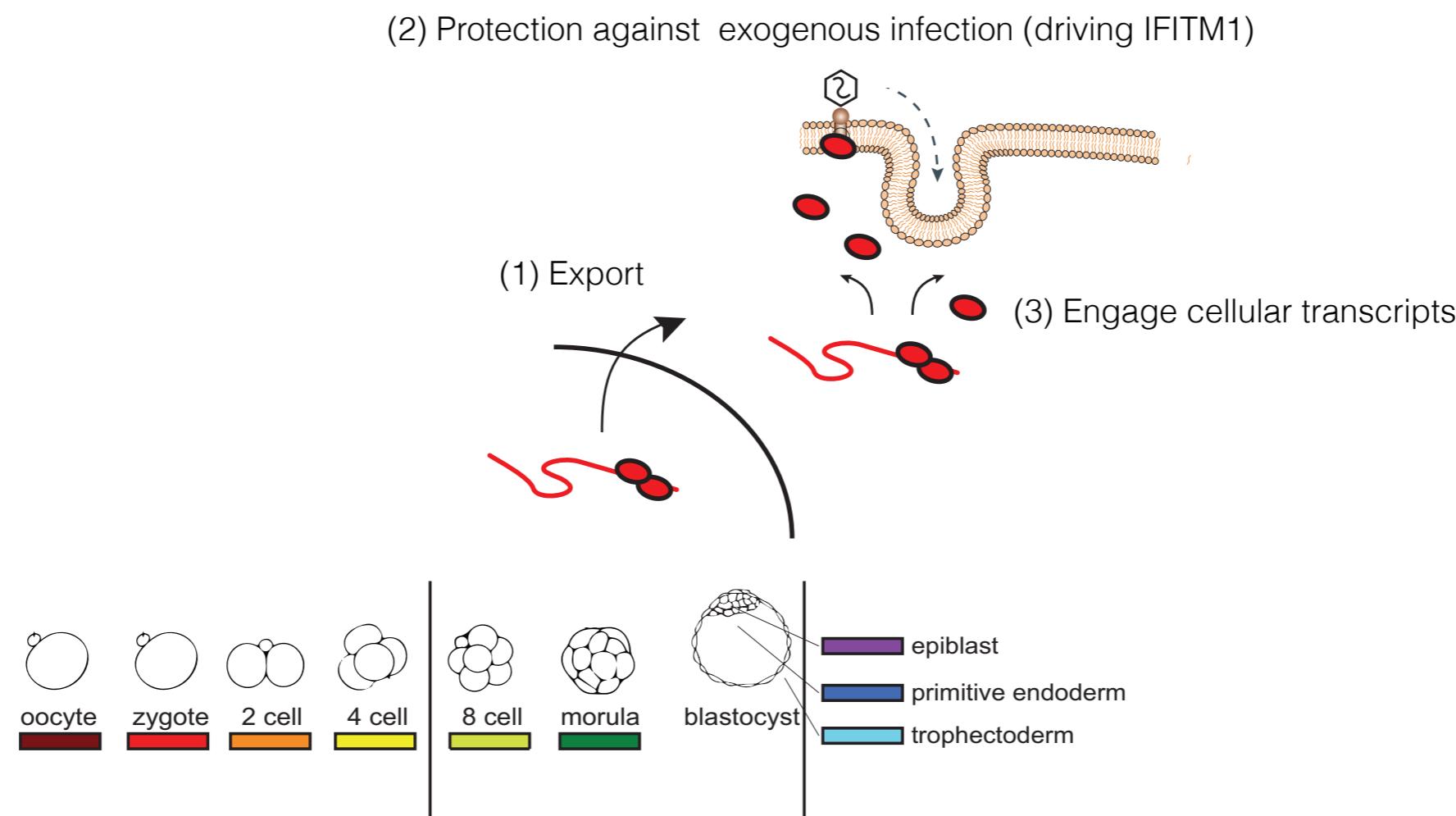
Read counts to regions in retroviral index -



Hits in Rec-responsive element region -  
Lower et. al. PNAS (1993, 1996)

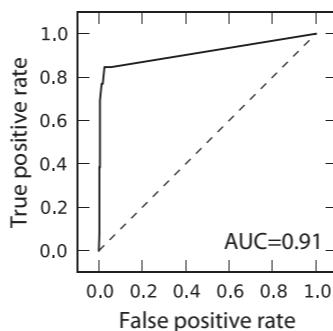


# Retroviral proteins present in early human development.

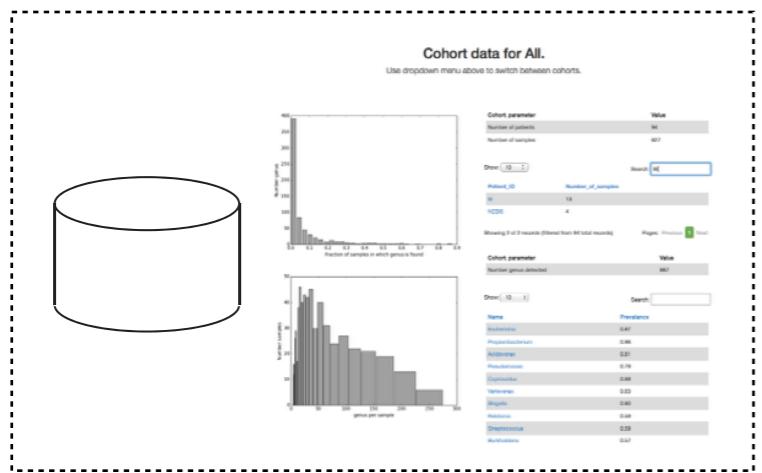


# Diagnostics

(3) Clinical studies: Lung, BMT  
*In review, 2014*



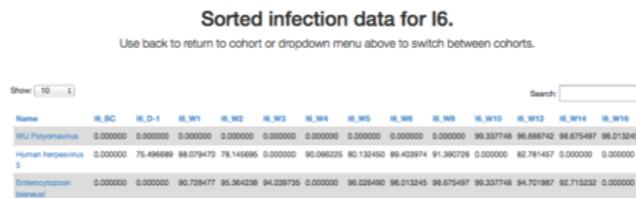
Infectome (1) Pipeline and (2) Browser



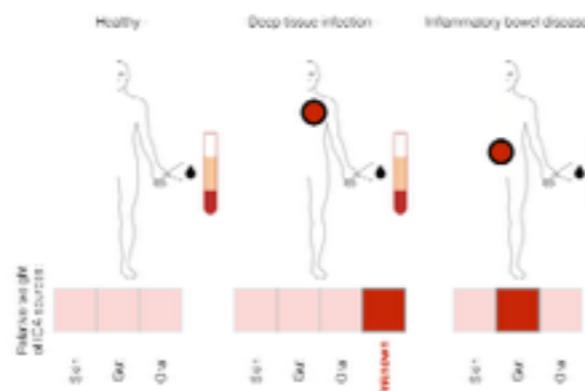
(4) Biopsy replacement  
*Stanford pathology*



(5) Undiagnosed infections  
*Sandhya Kharbanda, Stanford BMT*



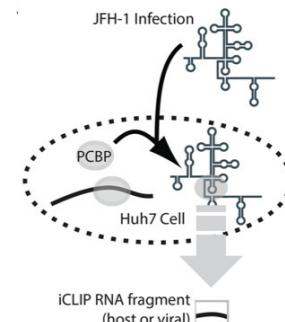
(6) Body site composition and anomalies  
*Relman lab, March of Dimes*



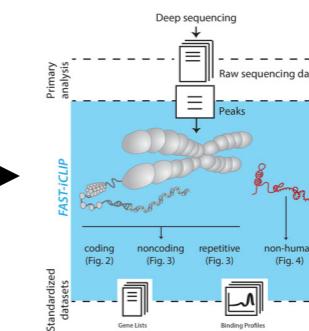
# Mechanism

(8) Human  
- *Nature, 2014*  
- *In preparation*  
(Cho, Doudna Labs)

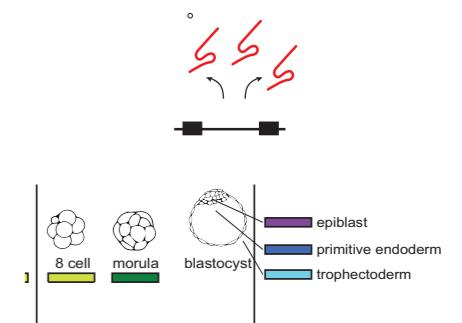
(9) Viral genomes  
*RNA, 2014*



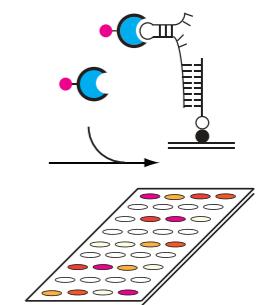
(7) CLIP pipeline  
*RNA, 2014*



(10) Retroviral genomes  
*In review, 2014*

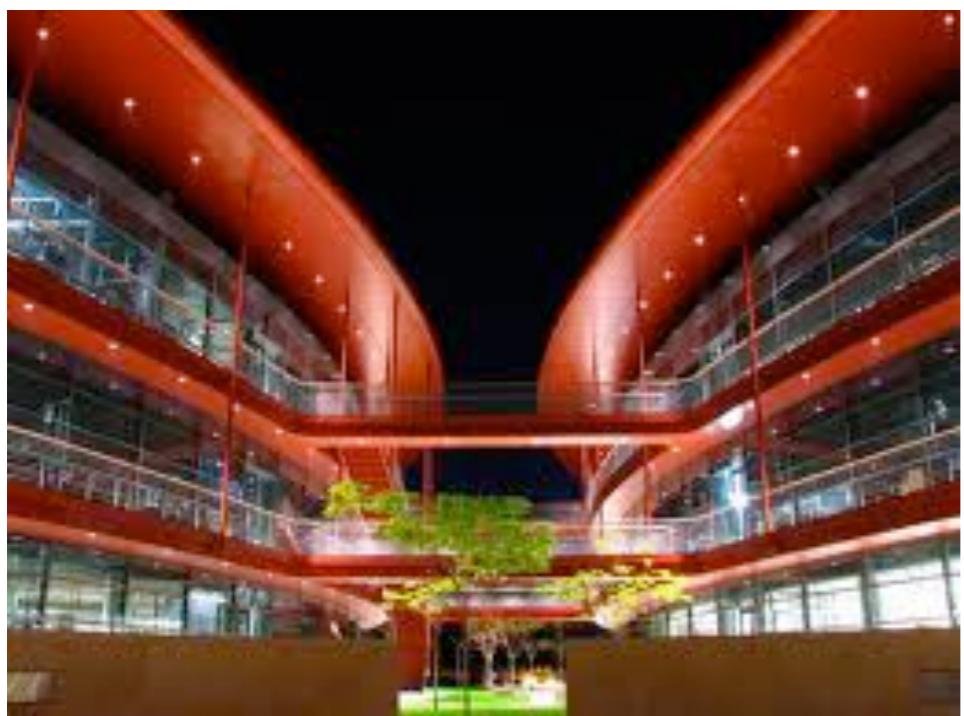


(11) Validation tools  
*Nature Methods, 2012*



# Thanks

Funding



People



+ Quake and Chang labs, Sequencing Core