

Computational Approaches for Microbiome Characterization

Keith A. Crandall, PhD, FAAS, FLS

Director, Computational Biology Institute

Director, GW Genomics Core

Co-Director, Informatics, Clinical and Translational Science Institute CN

Co-Director, Institute for Biomedical Sciences Genomics and Bioinformatics Program

Professor, Department of Biostatistics and Bioinformatics, GWSPH

Professor, Department of Biological Sciences, CCAS

Research Associate, Department of Invertebrate Zoology, US National Museum of Natural History,
Smithsonian Institution

Why Microbiome?

- Trillions of microbes (more microbes in you than human cells!) composed of bacteria, archaea, fungi, and viruses
- Strong connections between microbiomes, health and disease (obesity, type-2 diabetes, cirrhosis, rheumatoid arthritis, inflammatory bowel disease, etc.)
- Strong connections to environmental health, part of eDNA, etc.



How do we survey the microbiome?

Microbiome in Cancer

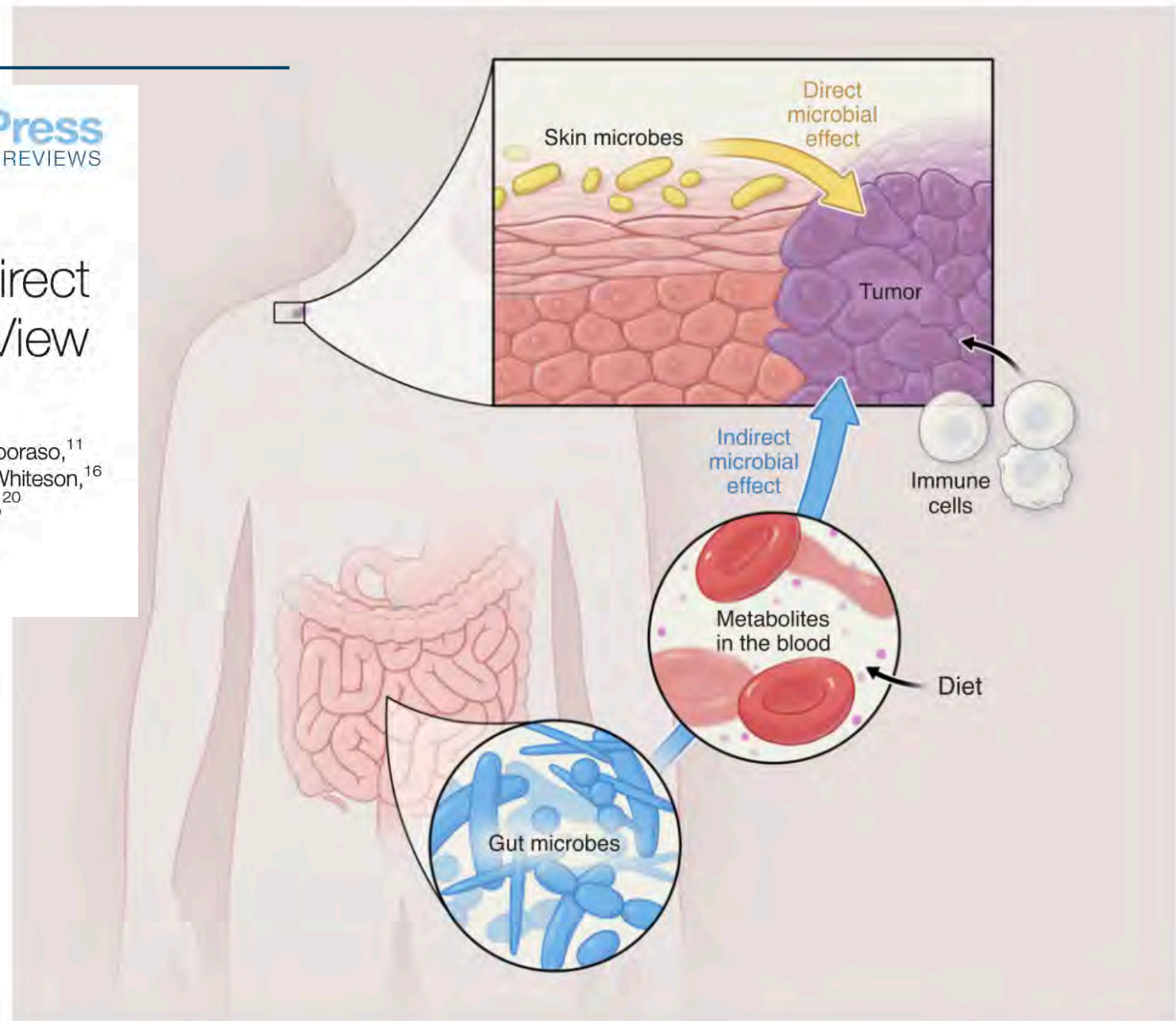
Trends in Cancer

CellPress
REVIEWS

Opinion

The Cancer Microbiome: Distinguishing Direct and Indirect Effects Requires a Systemic View

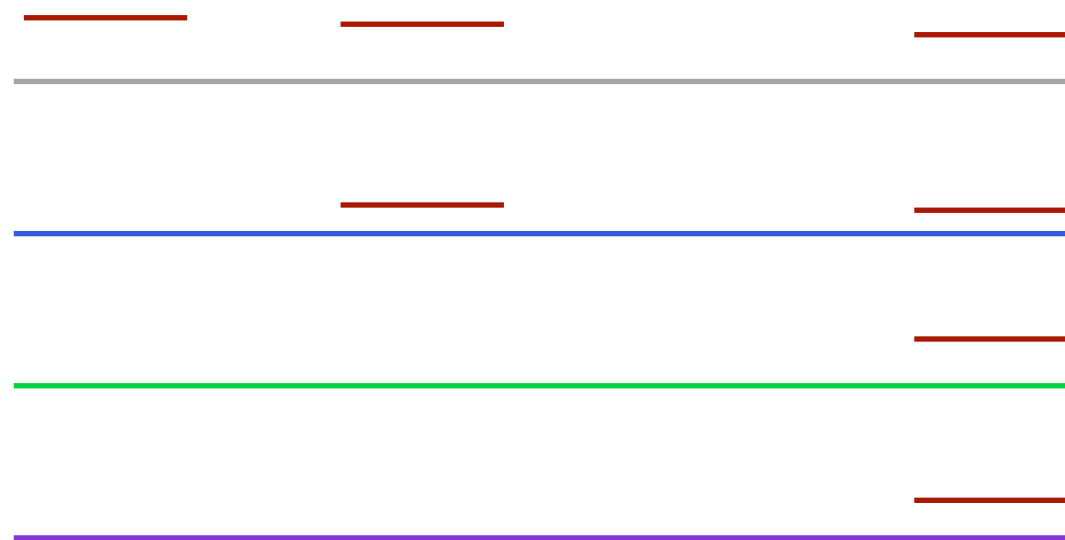
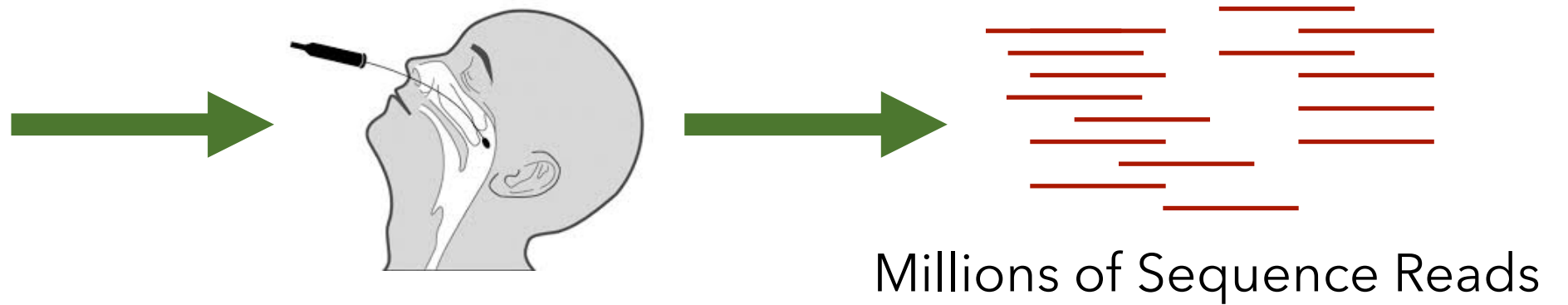
Joao B. Xavier,^{1,*} Vincent B. Young,² Joseph Skufca,³ Fiona Ginty,⁴ Traci Testerman,⁵ Alexander T. Pearson,⁶ Paul Macklin,⁷ Amir Mitchell,⁸ Ilya Shmulevich,⁹ Lei Xie,¹⁰ J. Gregory Caporaso,¹¹ Keith A. Crandall,¹² Nicole L. Simone,¹³ Filipa Godoy-Vitorino,¹⁴ Timothy J. Griffin,¹⁵ Katrine L. Whiteson,¹⁶ Heather H. Gustafson,¹⁷ Daniel J. Slade,¹⁸ Thomas M. Schmidt,¹⁹ Marina R.S. Walther-Antonio,²⁰ Tal Korem,²¹ Bobbie-Jo M. Webb-Robertson,²² Mark P. Styczynski,²³ W. Evan Johnson,²⁴ Christian Jobin,²⁵ Jason M. Ridlon,²⁶ Andrew Y. Koh,²⁷ Michael Yu,²⁸ Libusha Kelly,²⁹ and Jennifer A. Wargo³⁰



Trends in Cancer

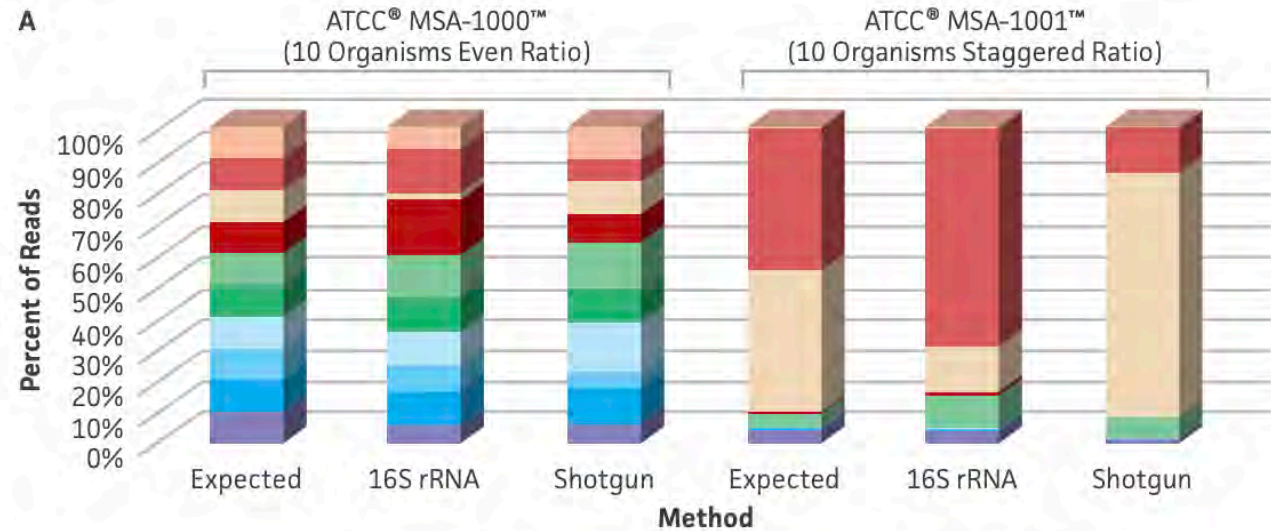
Figure 1. The Human Microbiome May Interact with a Cancer through Intricate Feedback Loops, and Investigating the Interaction Networks May Require a Systems Biology Approach. The effect of a microbiome on a cancer may be direct or indirect, and this is an important distinction. Direct interaction could take place between microbiomes residing in tissues where the cancer emerges, such as the hypothetical roles illustrated here of the skin microbiome interacting directly with melanoma; indirect interactions could occur between a microbiome and a cancer residing in a different tissue. For example, the gut microbiome alters circulating metabolites which in turn affect general host physiology and could have an indirect effect on the progression of skin cancer or its response to therapy. Diet will also play a role because it affects the circulating levels of metabolites and the microbiome composition.

Motivation

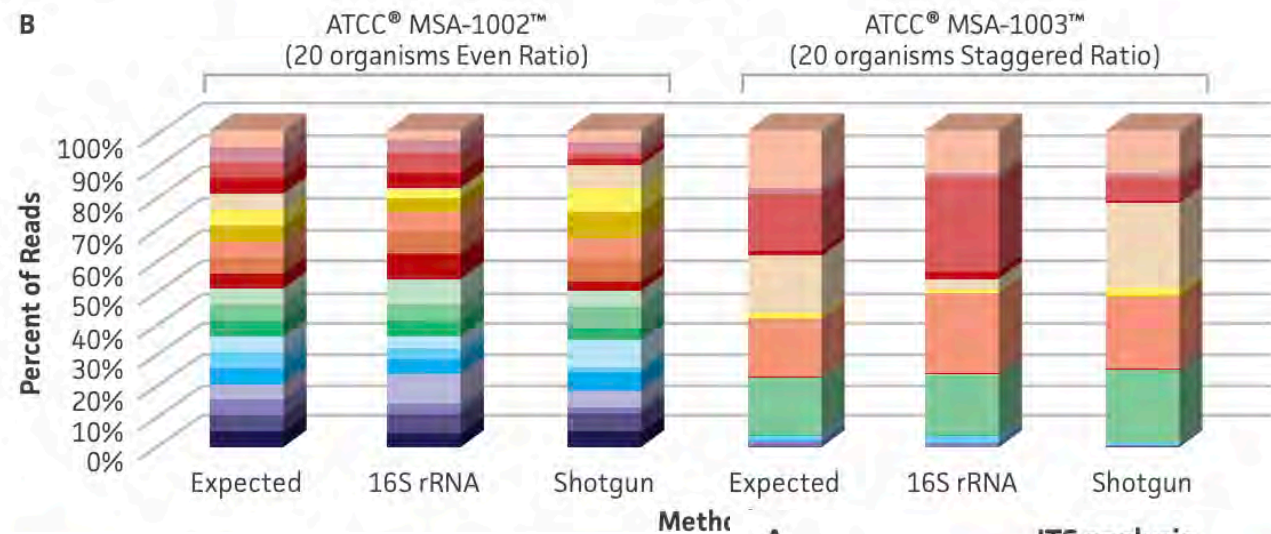


Map Reads to Reference Genomes

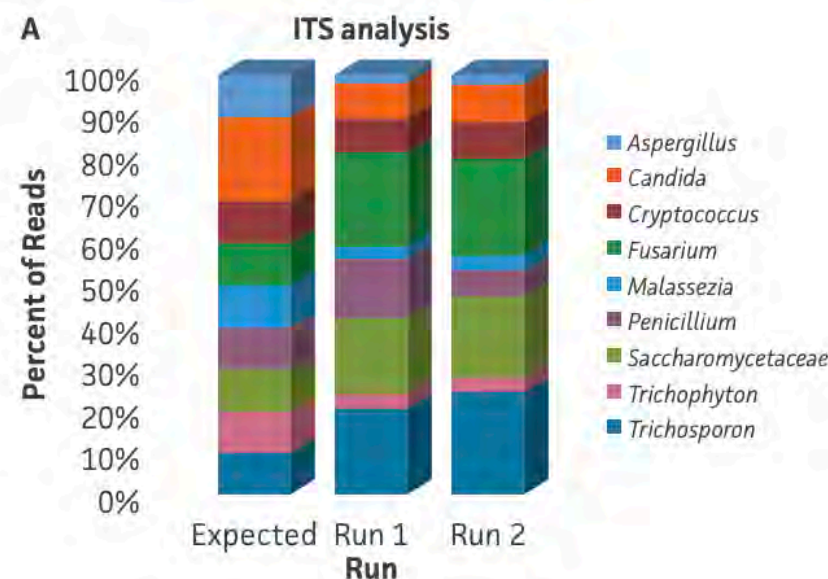
Shotgun vs Targeted Amplicon (16S/ITS)



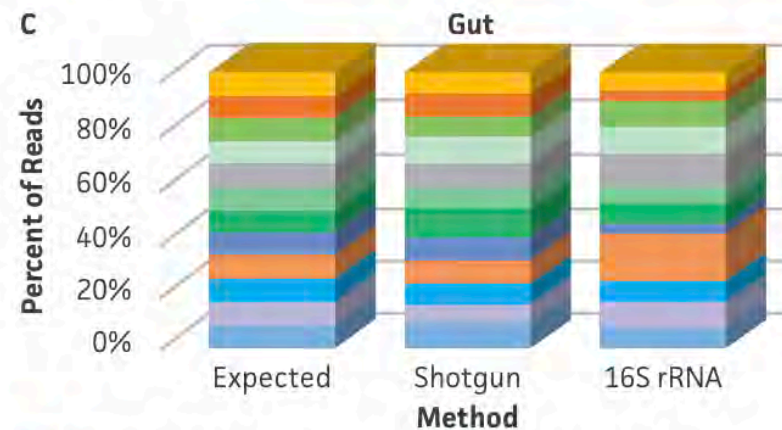
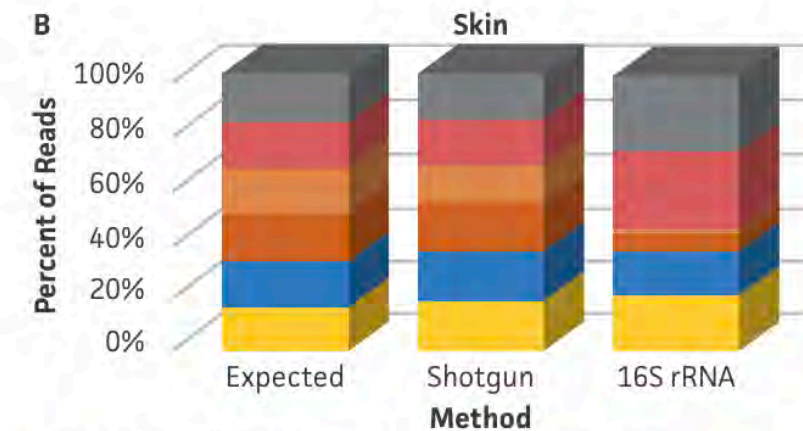
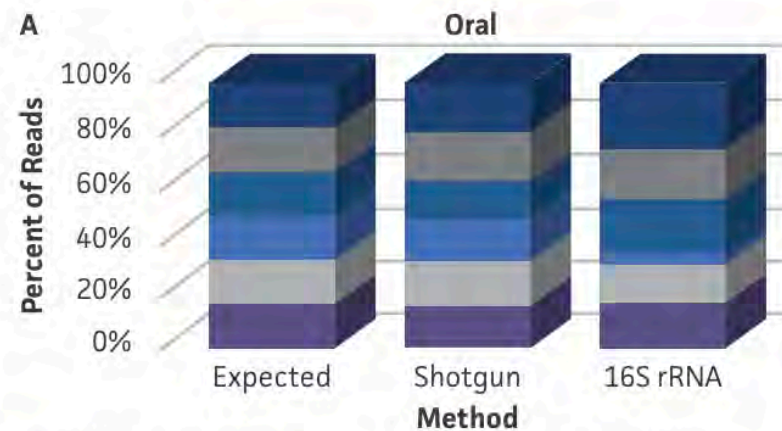
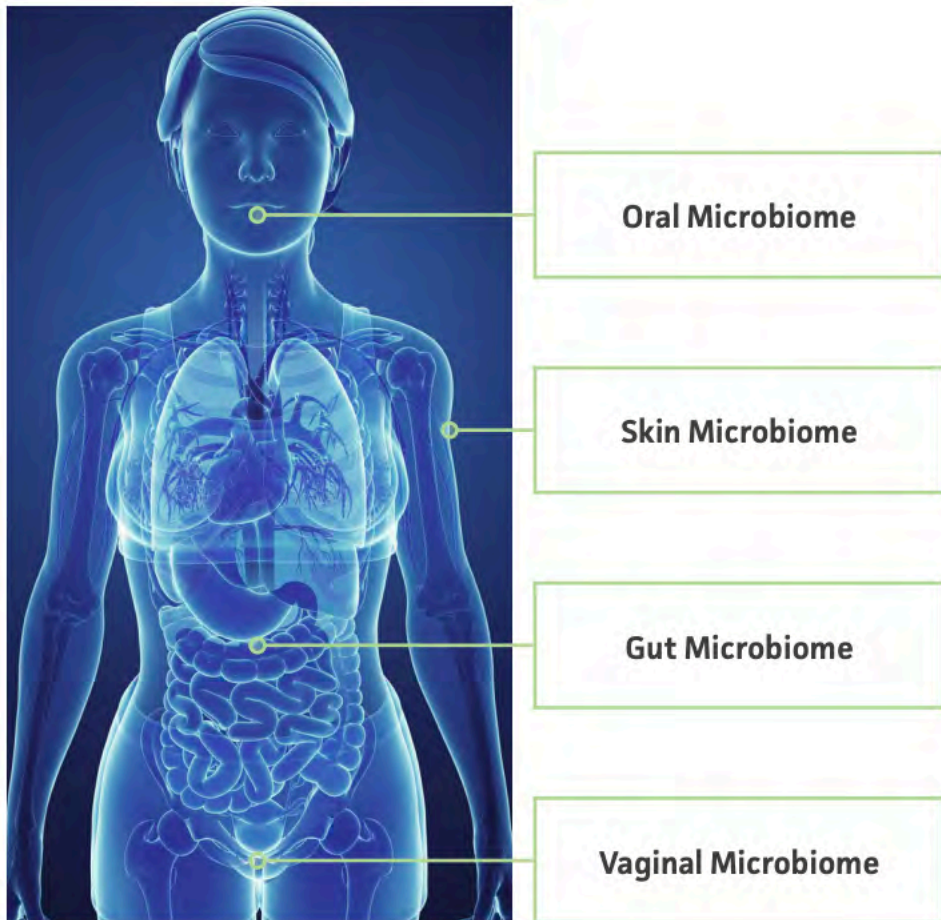
16S v Shotgun



ITS v Shotgun



Shotgun vs Targeted Amplicon (16S)



Method Development and Software Implementation



Pathoscope: Species identification and strain attribution with unassembled sequencing data

Owen E. Francis, Matthew Bendall, Solaiappan Manimaran, et al.

Hong et al. *Microbiome* 2014, **2**:33
<http://www.microbiomejournal.com/content/2/1/33>



METHODOLOGY

Open Access

PathoScope 2.0: a complete computational framework for strain identification in environmental or clinical sequencing samples

Changjin Hong^{1*†}, Solaiappan Manimaran^{1†}, Ying Shen¹, Joseph F Perez-Rogers^{1,2}, Allyson L Byrd², Eduardo Castro-Nallar³, Keith A Crandall³ and William Evan Johnson^{1,2*}

Bayesian Mixture Model

$$L(\pi, \theta | x_i, q_i, y) \propto \prod_{i=1}^R \prod_{j=1}^G [\pi_j \theta_j^{(1-y_i)} q_{ij}]^{x_{ij}}$$

π = proportion of reads from the j th genome

θ = reassignment parameter

x = unique reads/missing data for non-unique reads

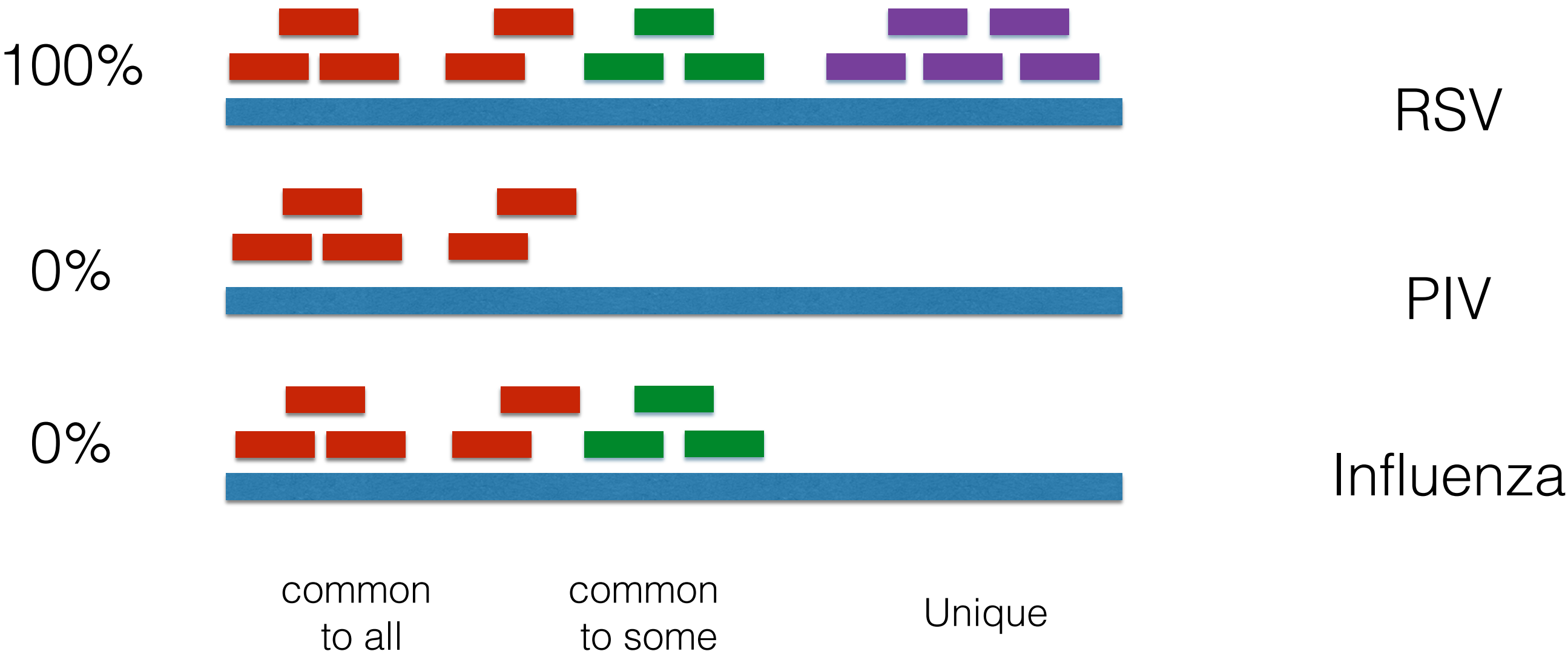
q = mapping score

y = uniqueness indicator

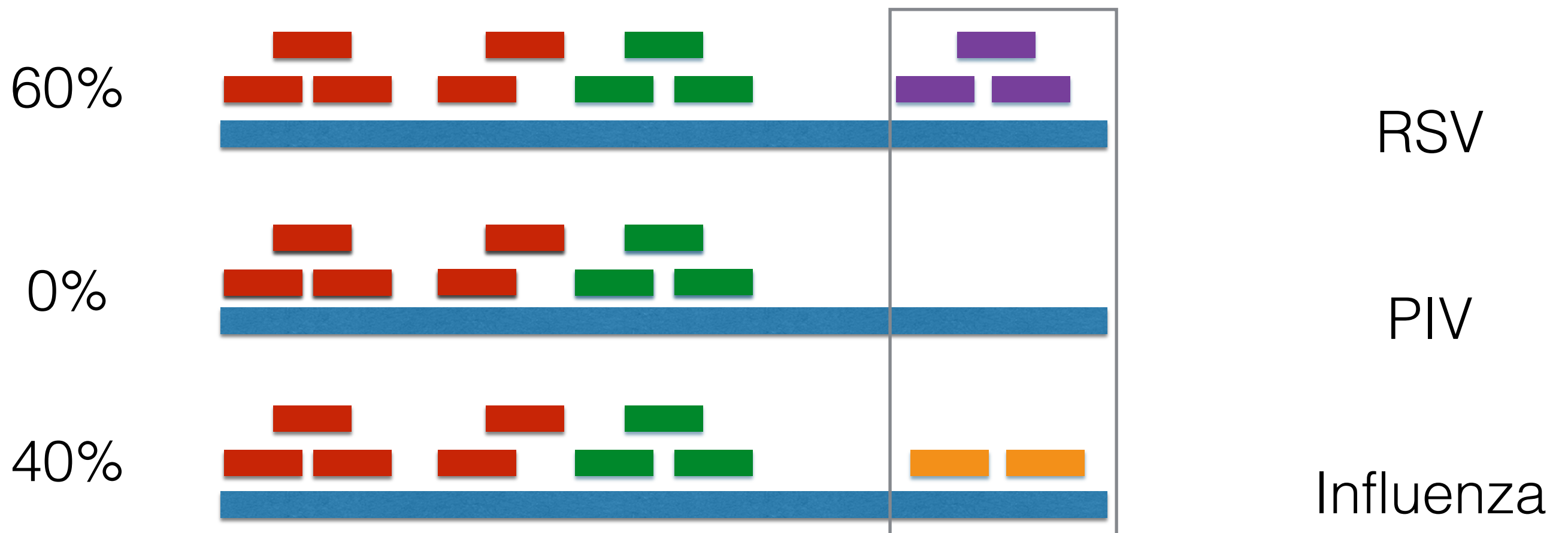
let $i = 1, \dots, R$ reads

let $j = 1, \dots, G$ genomes

PathoScope Concept



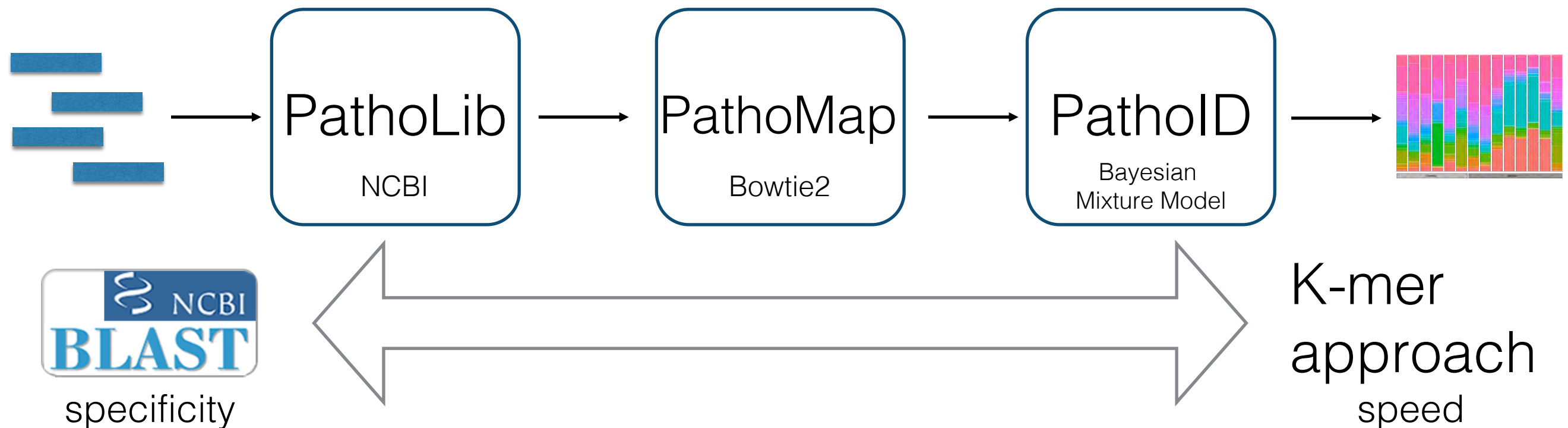
PathoScope Concept



PathoScope

Unassembled sequence data
Assigns reads to genomes of origin
Bayesian mixture model
Dirichlet distribution

Fast
Taxonomic profile with proportions
Accurate
Analyze all the reads collectively



Relative Performance - Metagenomics

Methods

- Read reassignment - PathoScope2 and metaMix
- Marker genes - MetaPhlan2 and ConStrains
- Kmer - Kraken and Centrifuge
- Genomic distance - Taxator-tk
- Mapping with ML model - Sigma

Approach



MetaSim

Daniel H. Huson and Felix Ott

with contributions from:
R. Schmid, A.F. Auch and D.C. Richter

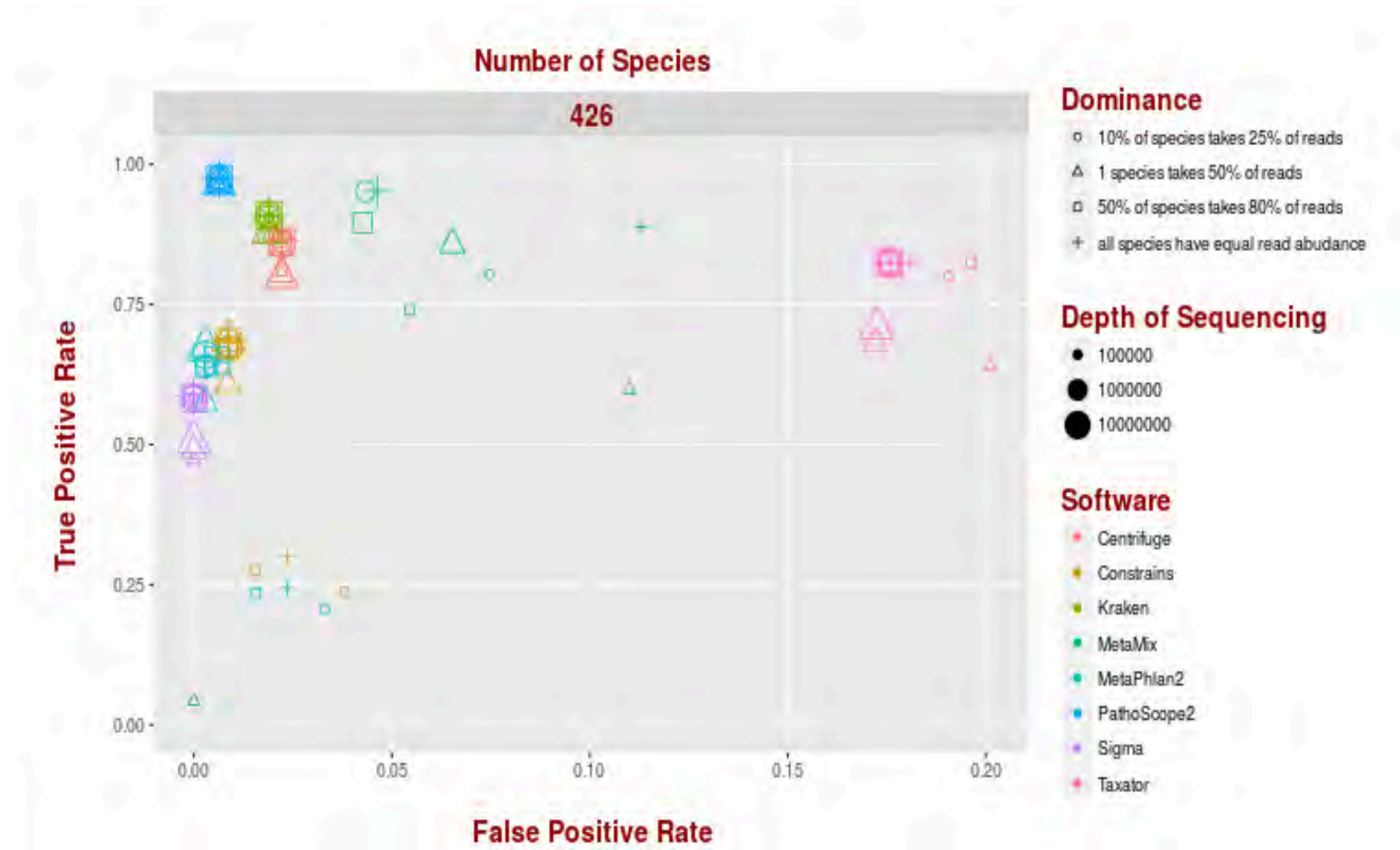
www-ab.informatik.uni-tuebingen.de/software/metasim

Parameters

Condition	Parameter values	Number of values
Read length	75; 150; 300; 1000 bp	4
Sequencing depth	100k; 1M; 10M	3
Dominance	1->50%; 10-> 25%; 50->80%; all equal	4
Number of species	10; 100; 426	3
Methods	MetaPhlan2; ConStrains; Sigma; metamix; Kraken; PathoScope2	6
Total number of analysis		864

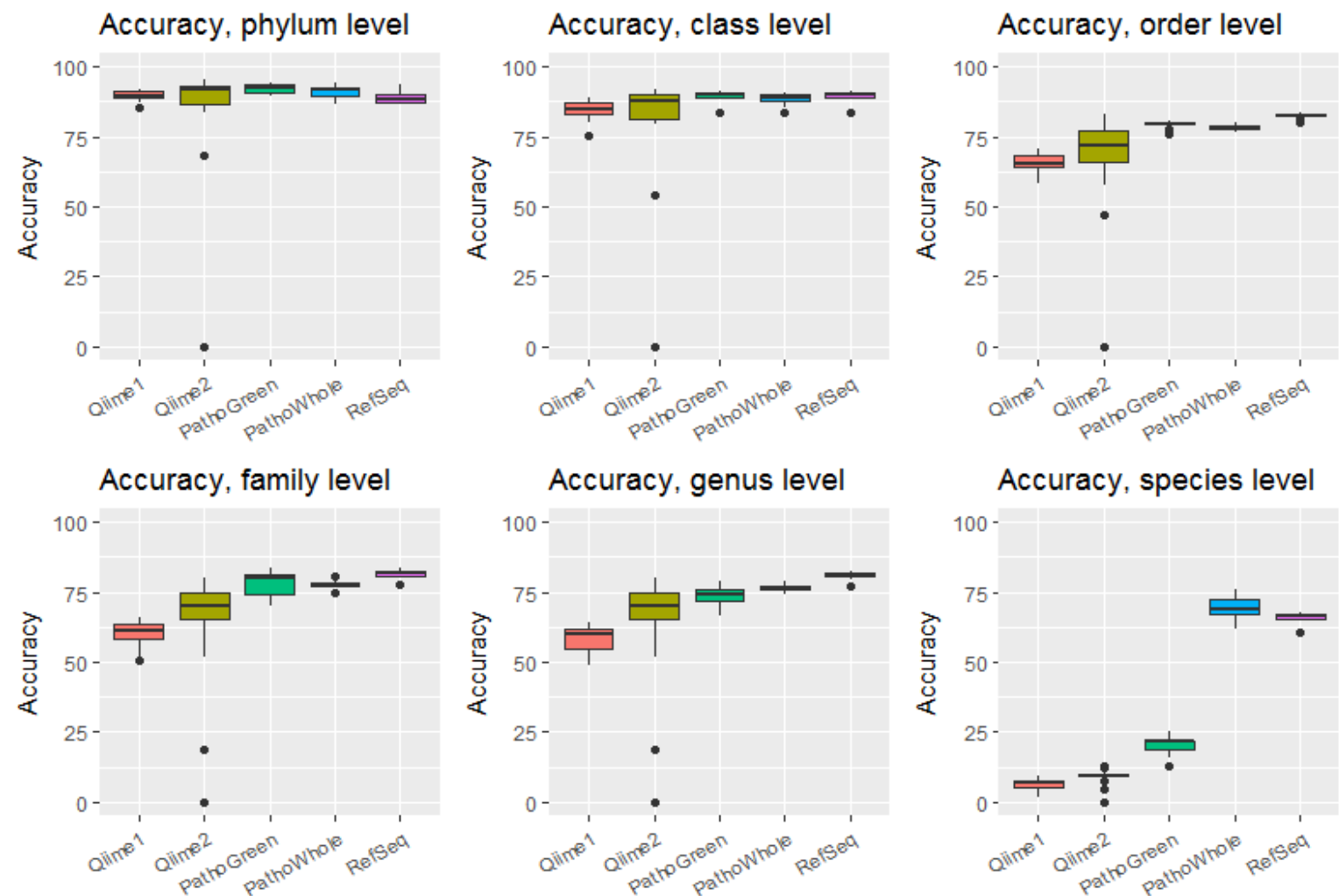
Relative Performance

- Centrifuge, Kraken, PathoScope perform well across the board compared to other approaches
- Read length does not impact performance, but current software will not work well with longer read lengths
- Sequencing depth does not negatively impact these methods – you can get away with low coverage!

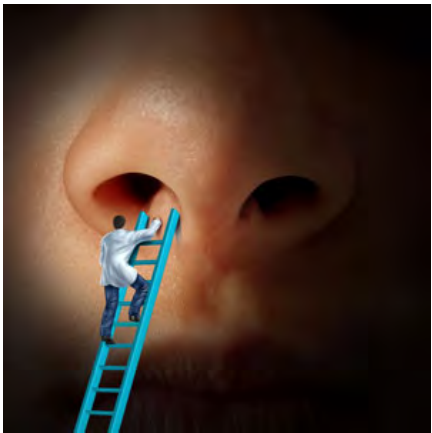


Relative Performance - 16S

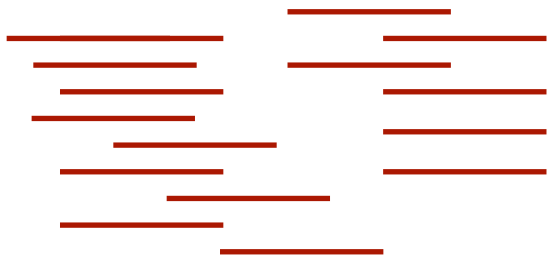
- Direct comparison with QIIME and QIIME 2 on 16S
- PathoScope with
 - Whole genome
 - RefSeq
 - GreenGenes
- Similar performance at higher taxonomic levels
- PathoScope outperforms at family, genus, and really kills it at species level!



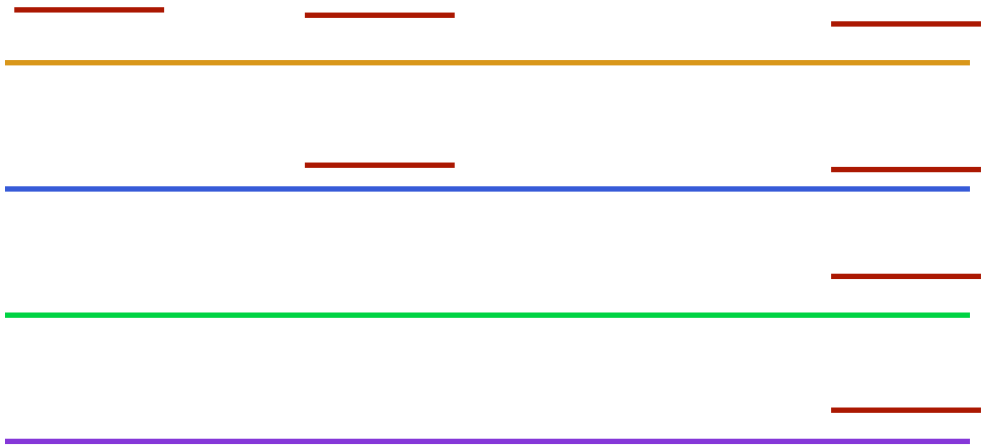
Pediatric Asthma



(Metagenomics)



Sequence Reads

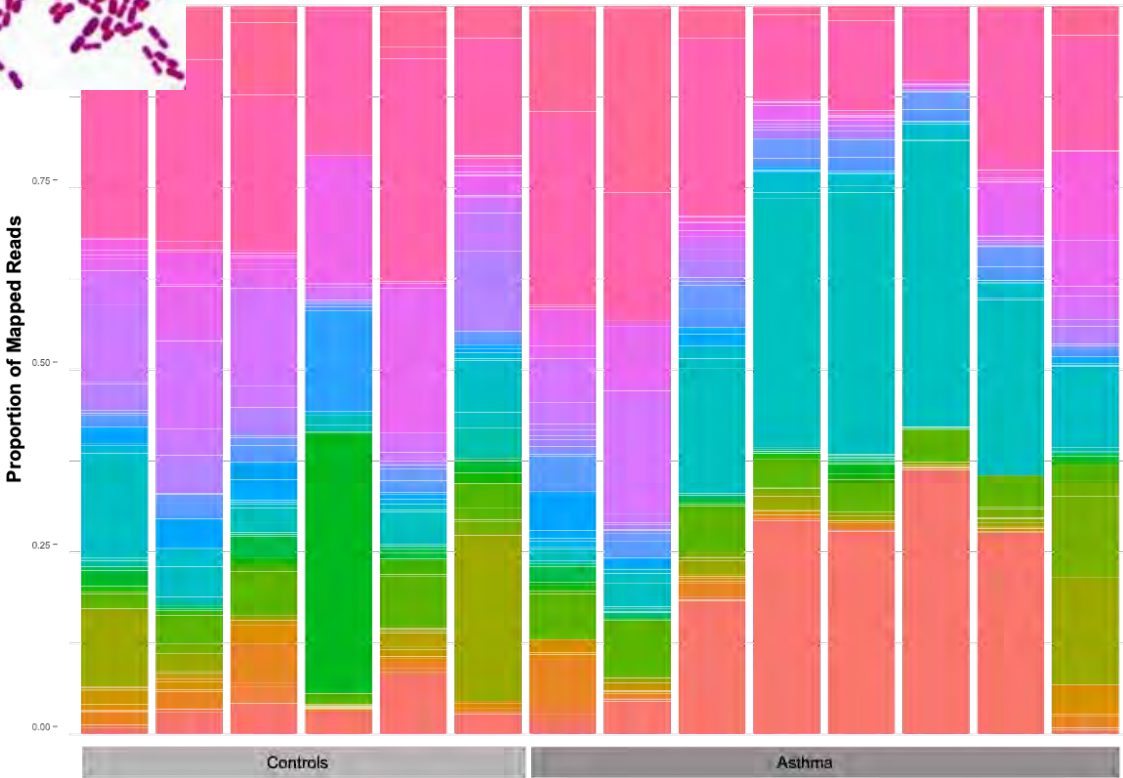


Map Reads to Reference
Genomes

(Informatics)



• *Moraxella catarrhalis*



Subjects

Sequencing Reads



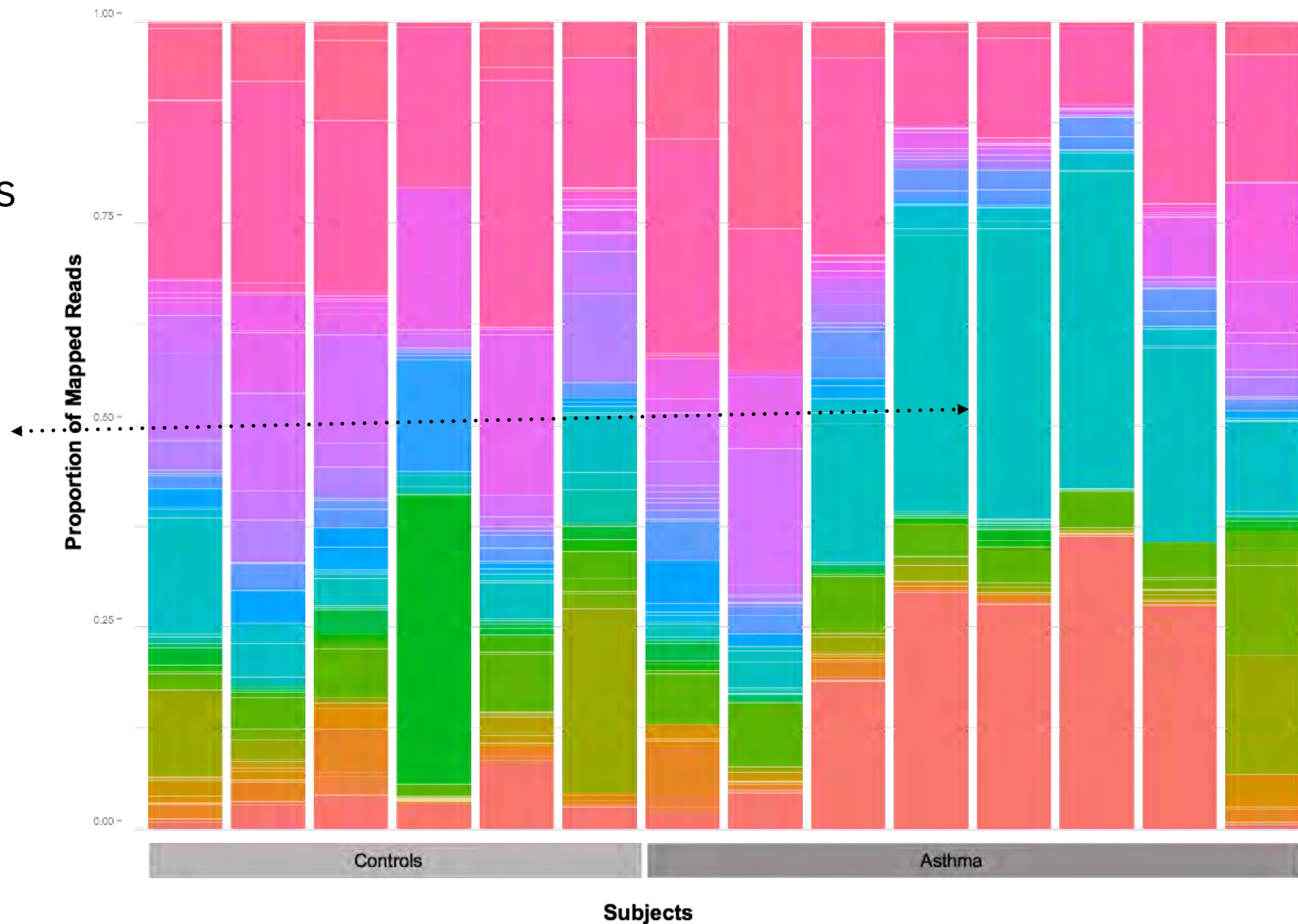
Host ~90%



Pathogens ~10%
including fungi, virus,
bacteria

Asthma Pathogen Detection

- RNAseq from 6 cases and 8 controls
- *Moraxella catarrhalis* more abundant in asthma patients



Sequencing Reads

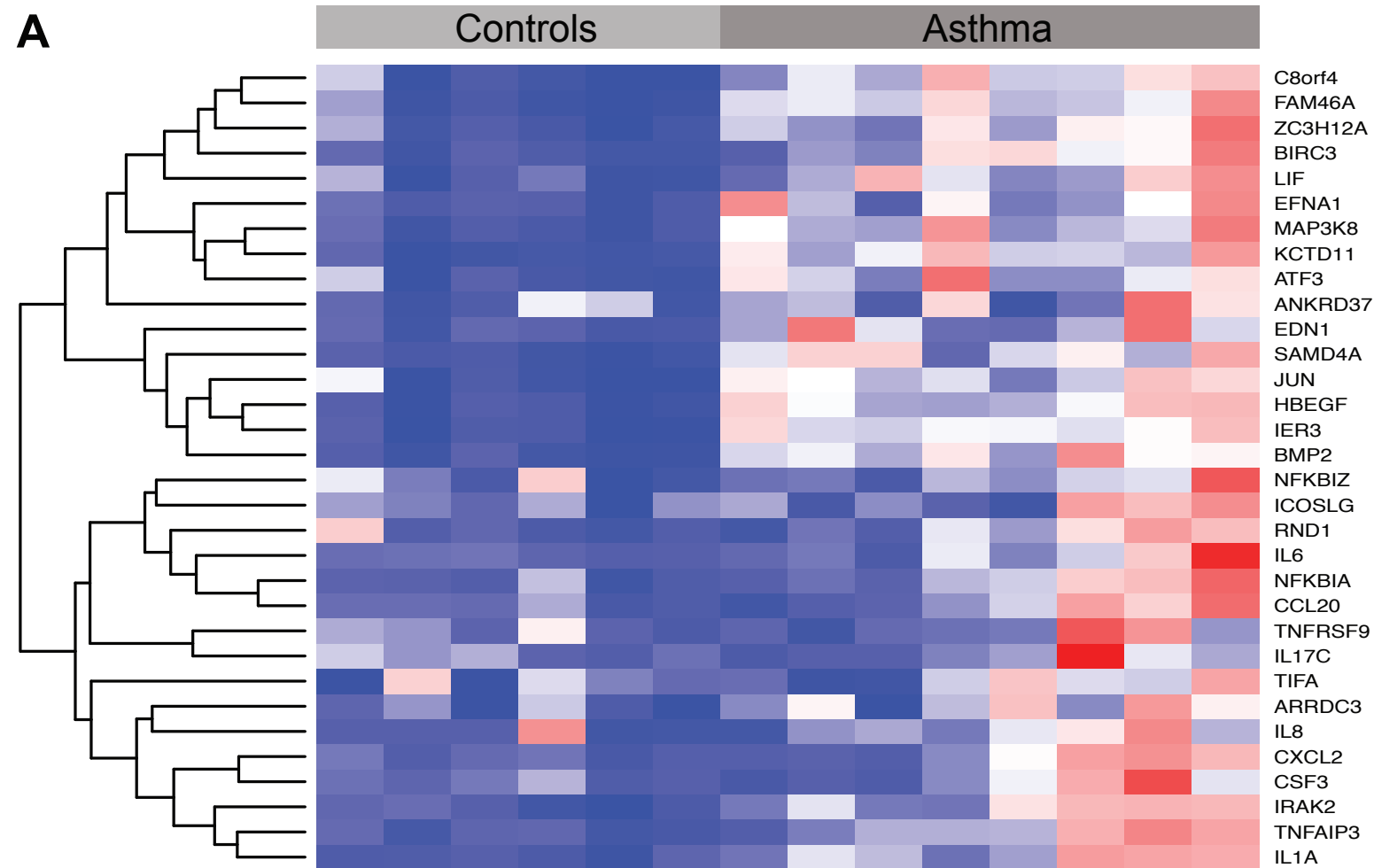


Host ~90%

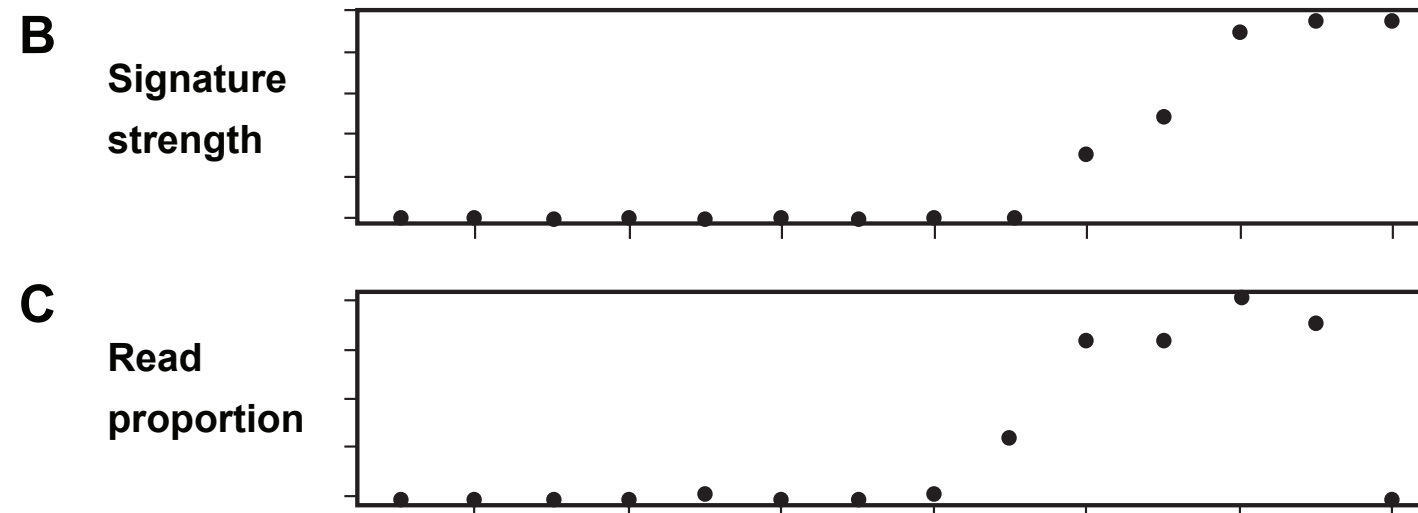


Pathogens ~10%
including fungi, virus,
bacteria

Host Impact

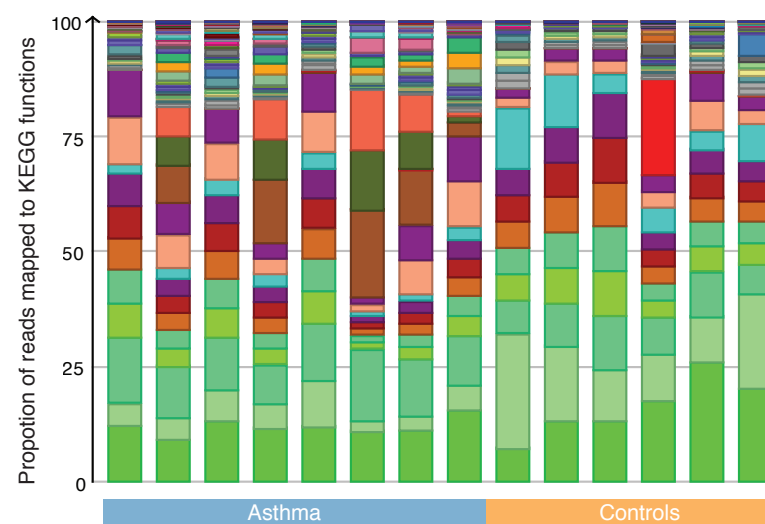
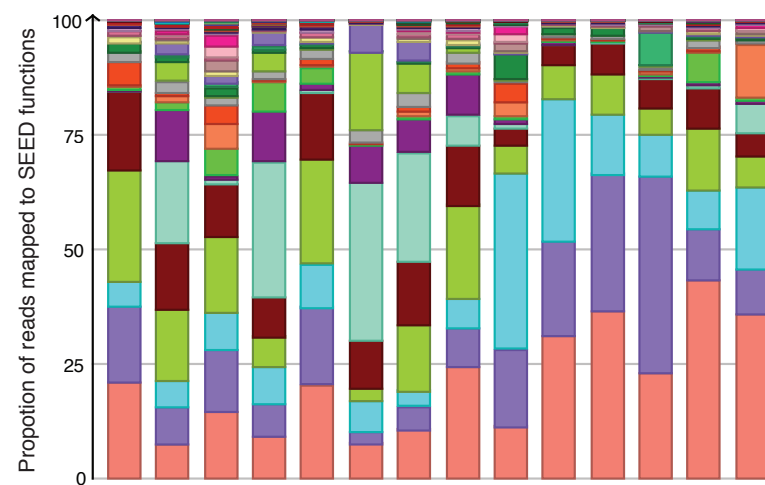
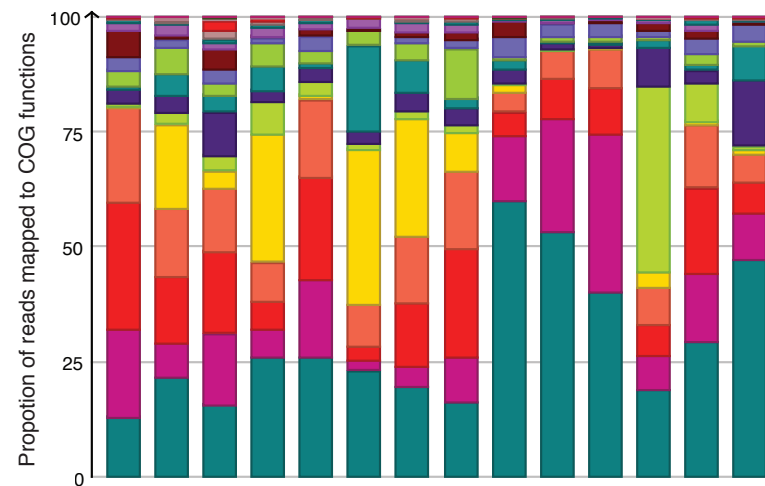


Candidate gene
expression for
immune response
to *Moraxella*



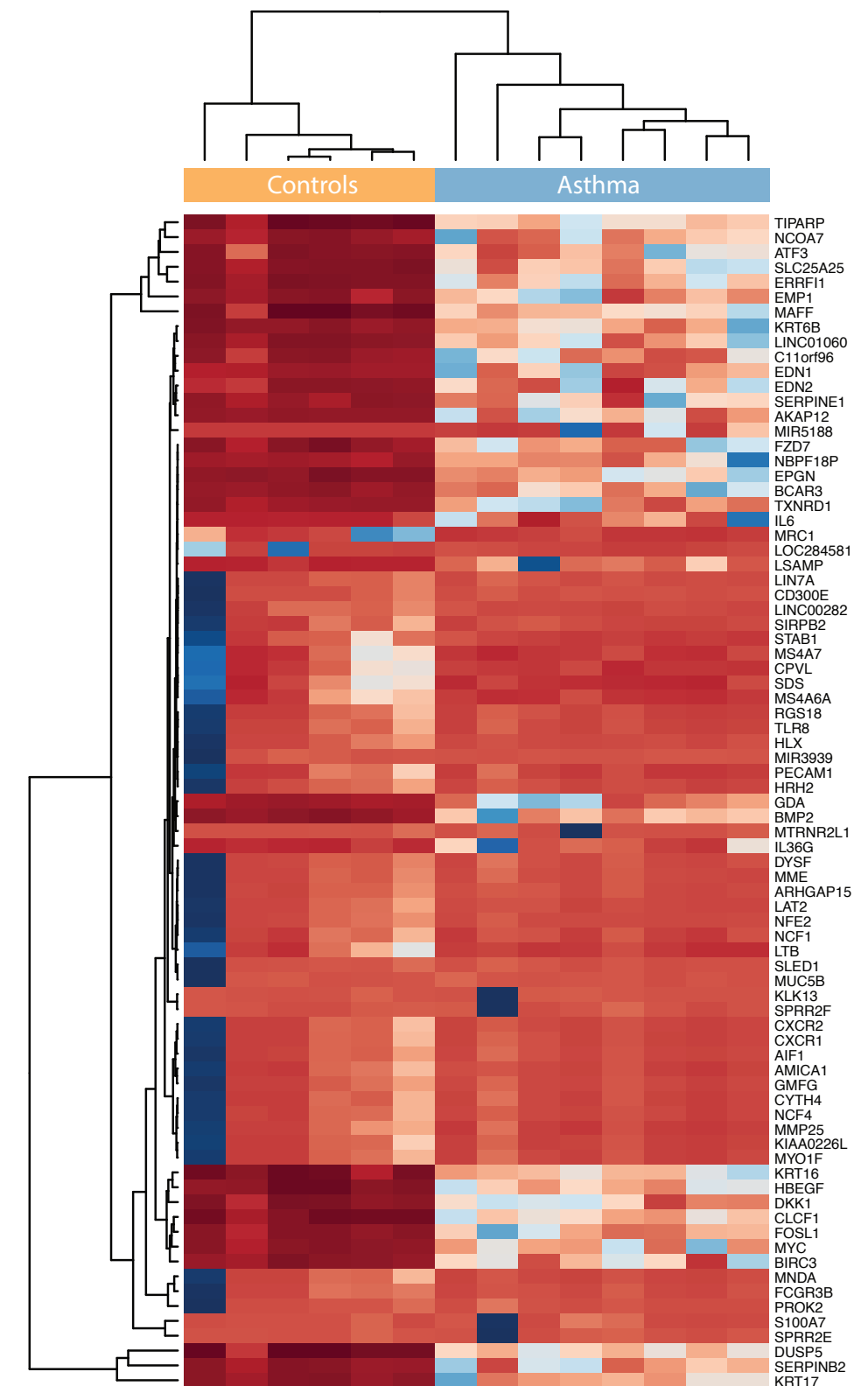
- 28 billion nucleotides total
- 30 TB data post analysis

Host-Pathogen Function

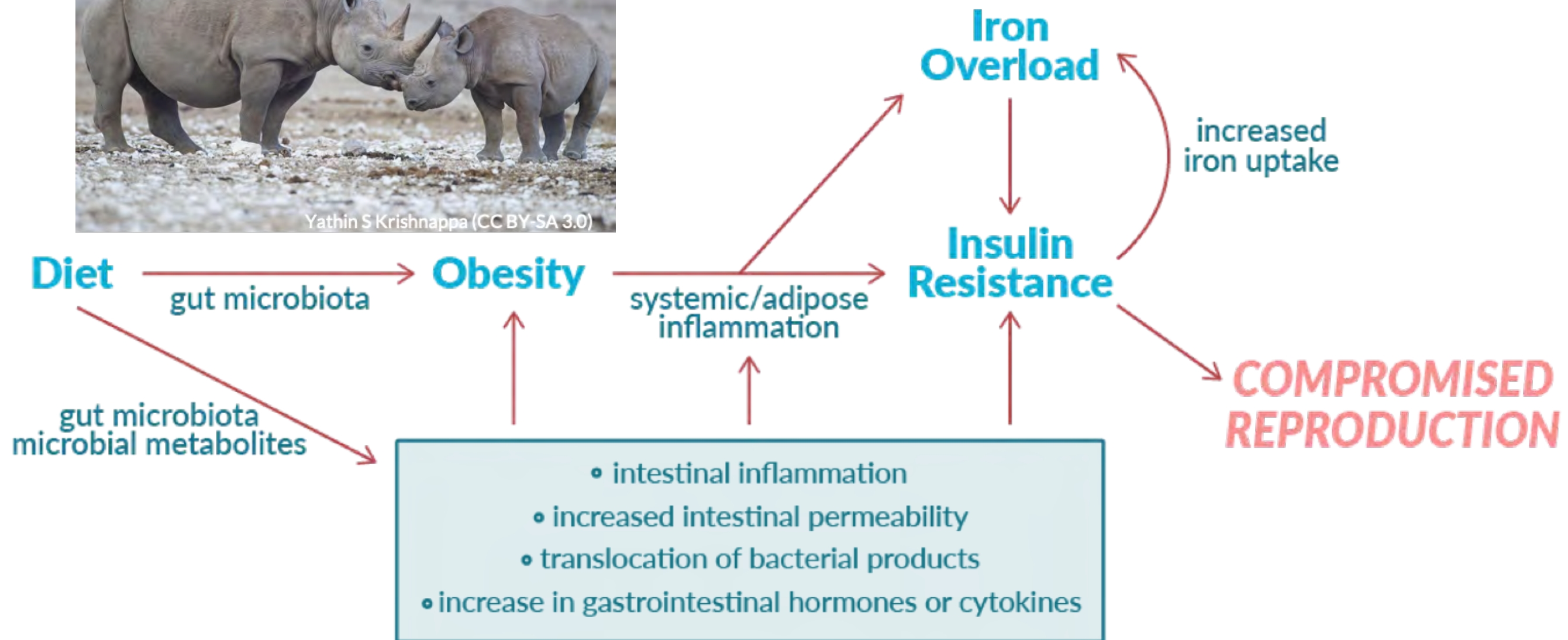


Microbe Function

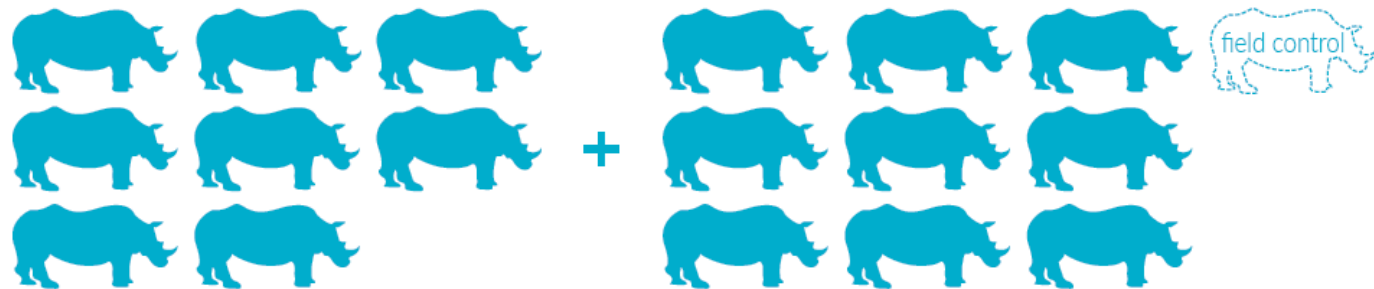
Host Function



Conservation – Wild versus Captive Rhinos



Experimental Design



17 wild rhino fecal samples


8 captive samples



SCIENTIFIC REPORTS

Article | [Open Access](#) | Published: 28 May 2019

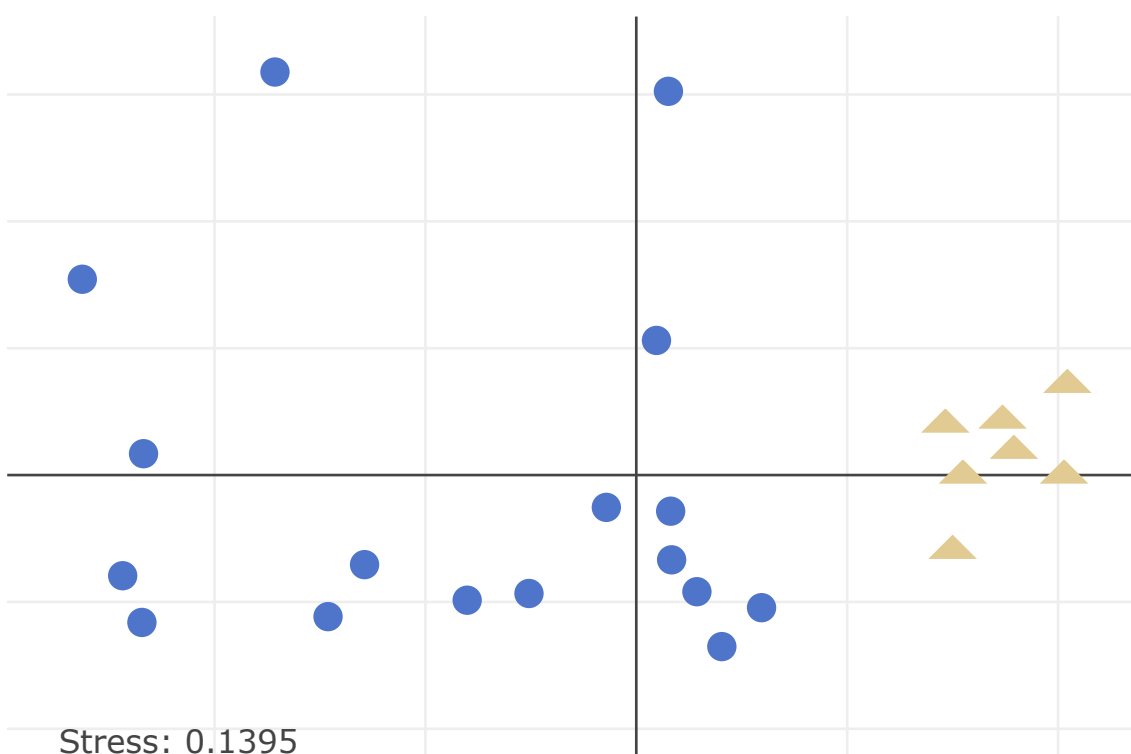
Gut microbiome differences between wild and captive black rhinoceros – implications for rhino health

Keylie M. Gibson, Bryan N. Nguyen, Laura M. Neumann, Michele Miller, Peter Buss, Savel Daniels, Michelle J. Ahn, Keith A. Crandall & Budhan Pukazhenthil 

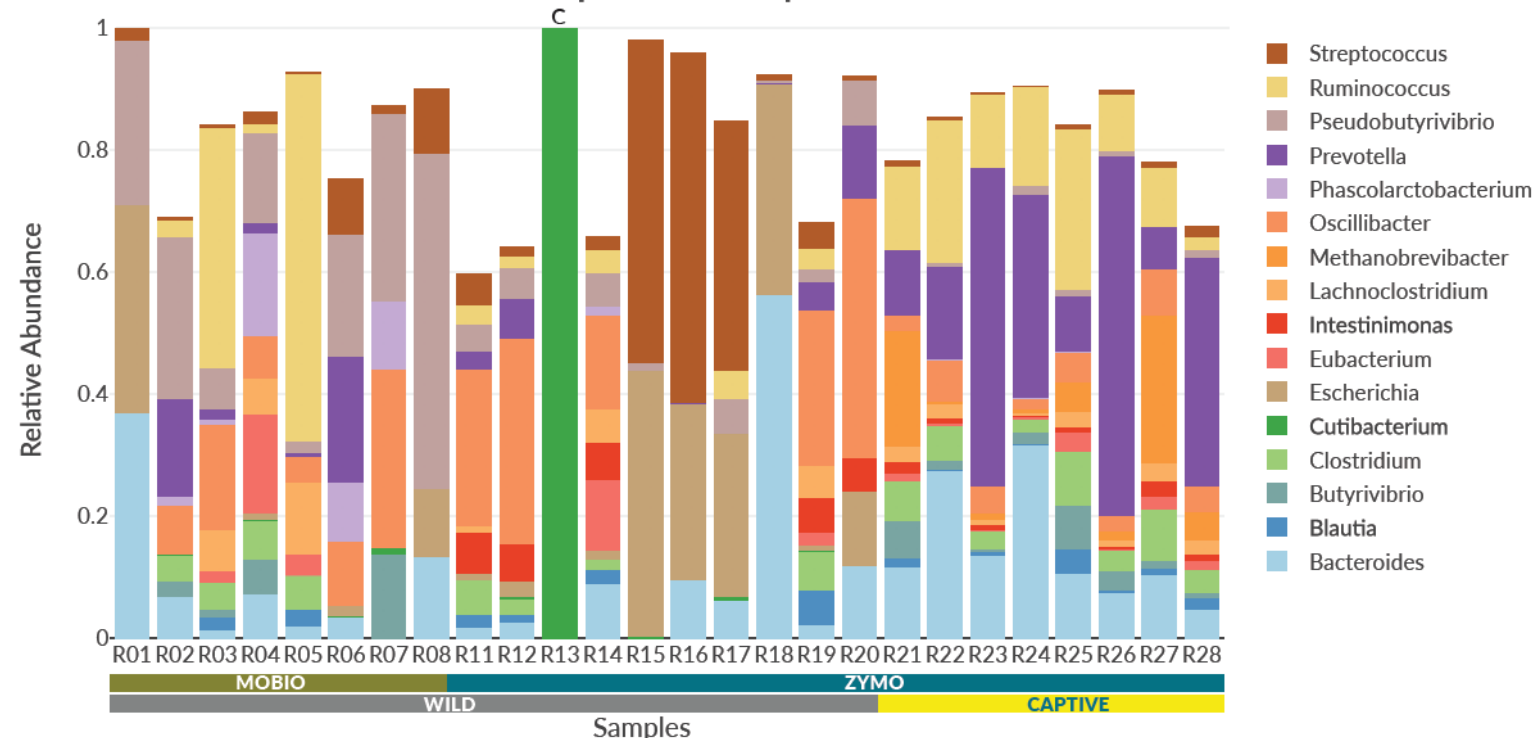
Significant Differences in Beta Diversity

- Significant differences in beta-diversity between captive and wild with **wild being much more taxonomically diverse** ($P=0.012$), PERMANOVA
- PathoScope reports **bacterial taxa traditionally associated with ruminant guts are higher in captive**
- PathoScope results suggest the **unknown gut symbionts of wild rhinos are being replaced** by those found in conventional human-domesticated livestock

B) PathoScope Bray-Curtis NMDS



Pathoscope Genus Composition



Fecal Microbiome Transplant (FMT)

- Test if FMT is an effective treatment of *Clostridioides difficile* infection (CDI)
- 9 children with recurrent CDI underwent FMT
- 'Donation' provided by relative or commercial
- Conducted metagenomic surveys longitudinally to test for
 - Diversity changes over time
 - CDI
 - Potential pathogen passage
 - Antimicrobial resistance genes



TREATMENTS

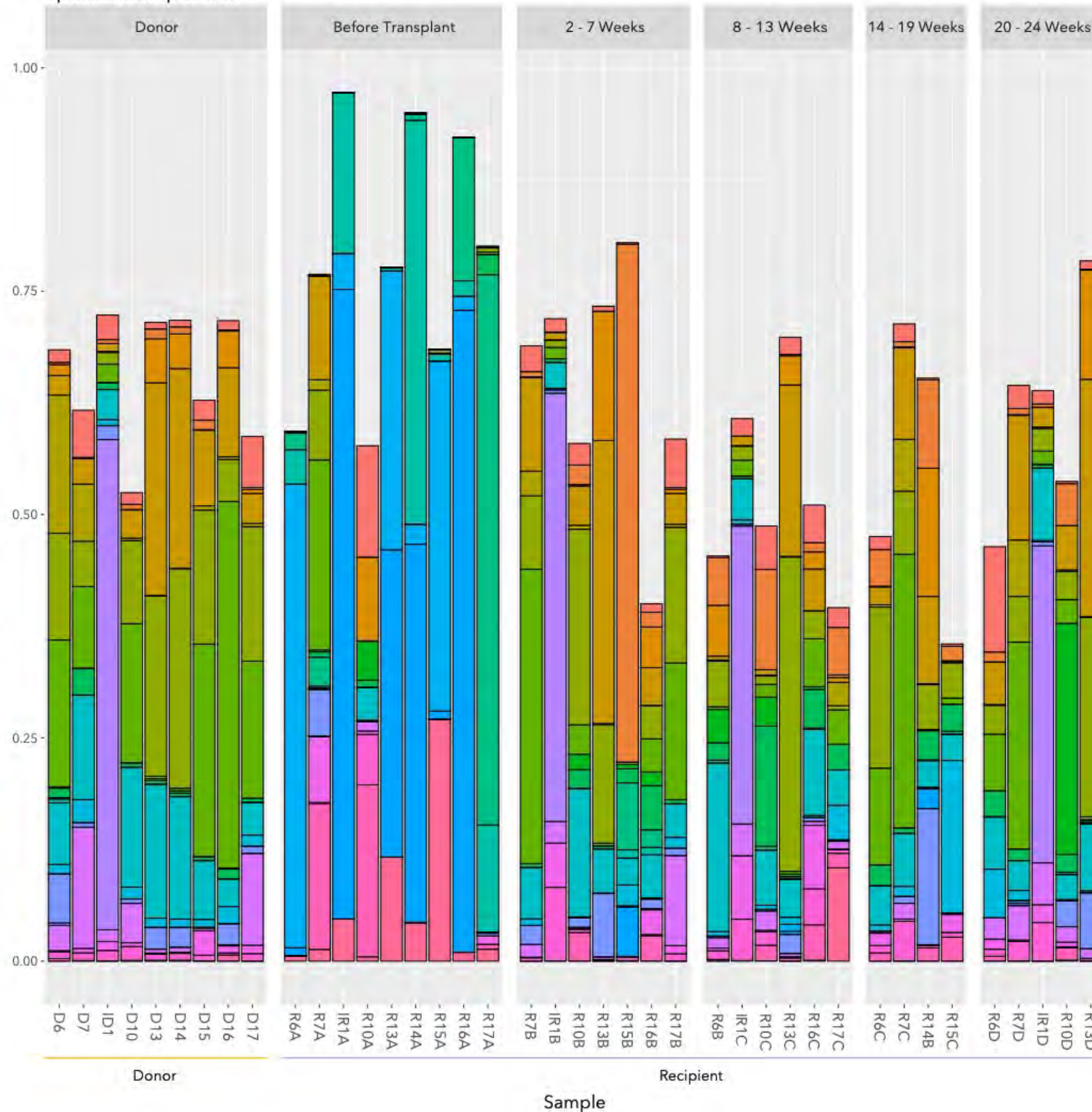
Be Careful Of Fecal Transplants, Warns FDA, After Patient Death

June 14, 2019 · 5:22 PM ET

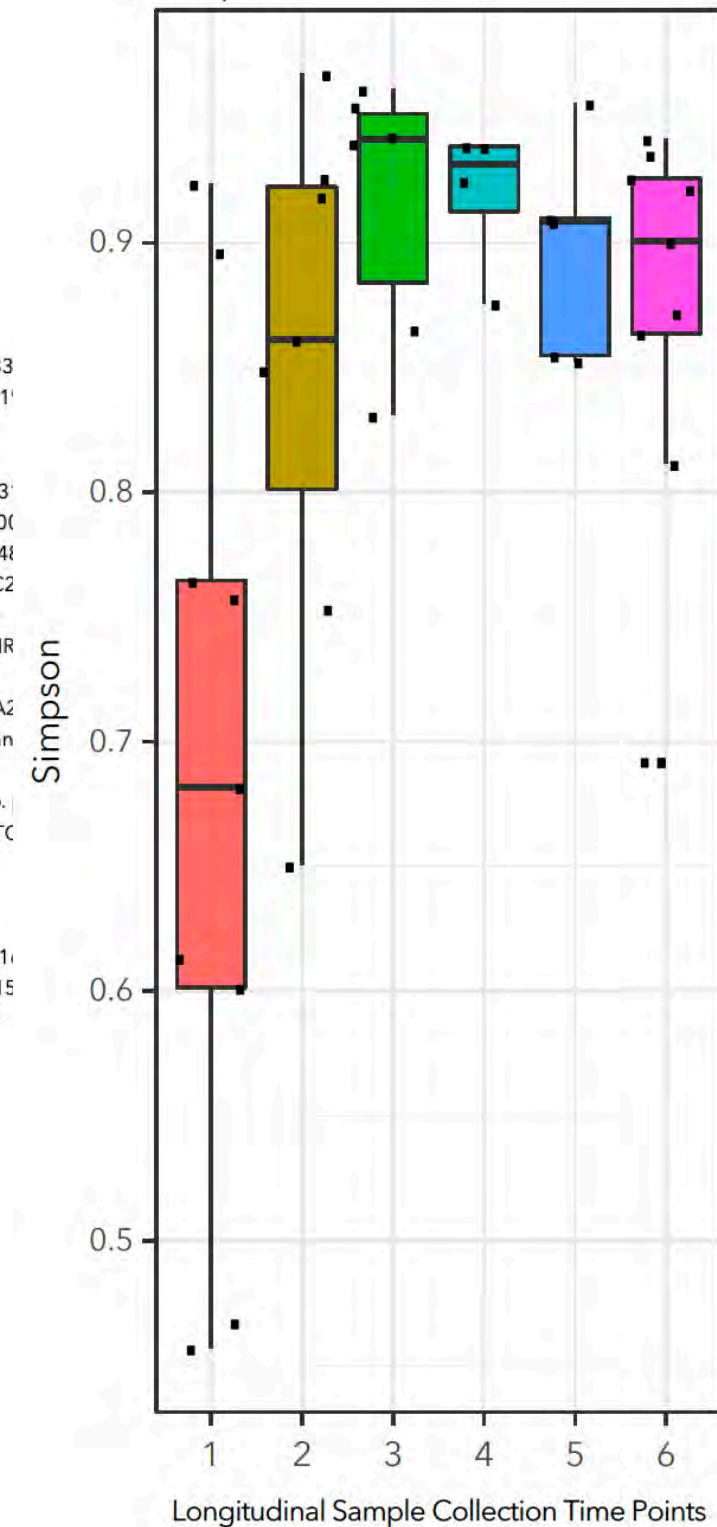
JONATHAN LAMBERT

FMT is effective in children

Species Composition

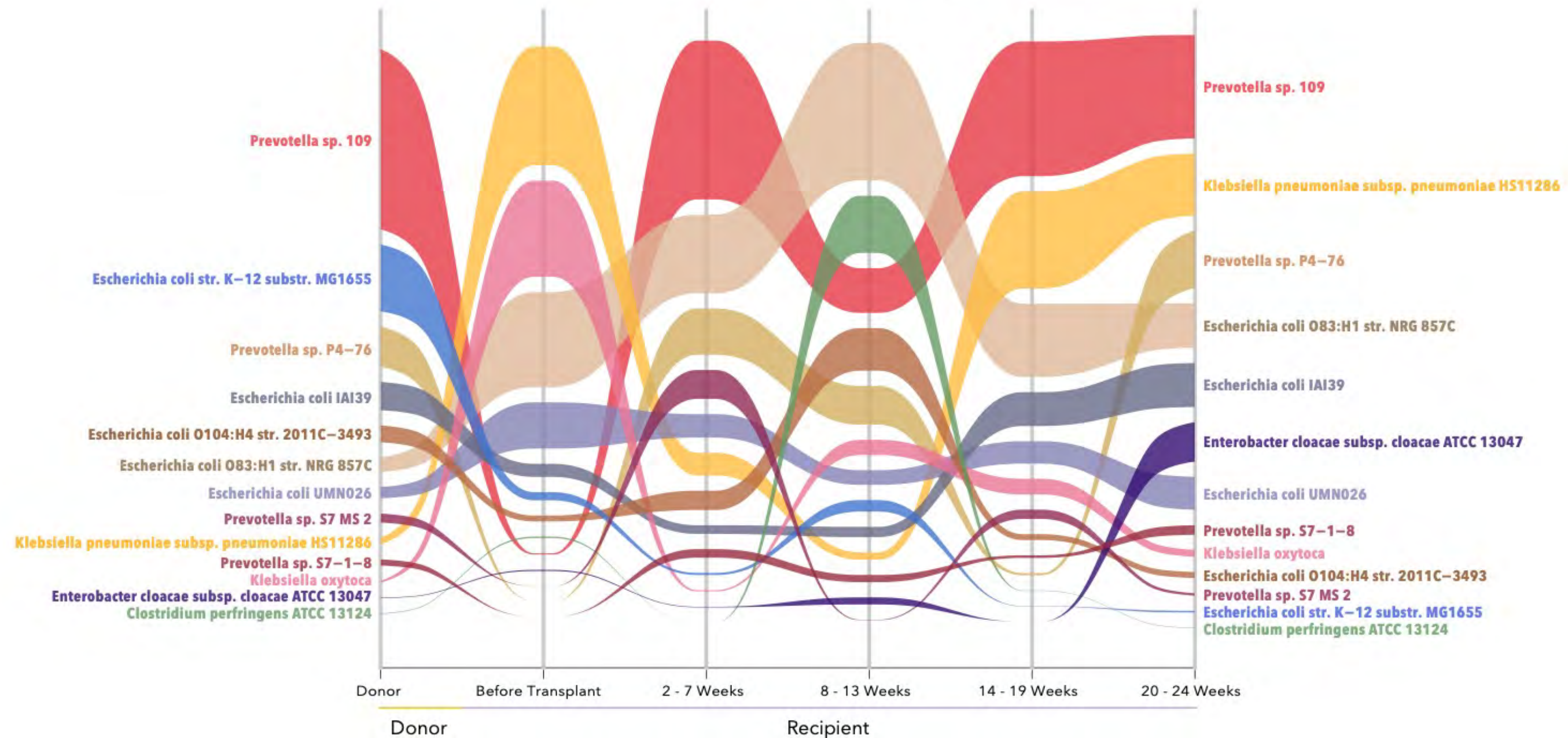


Simpson

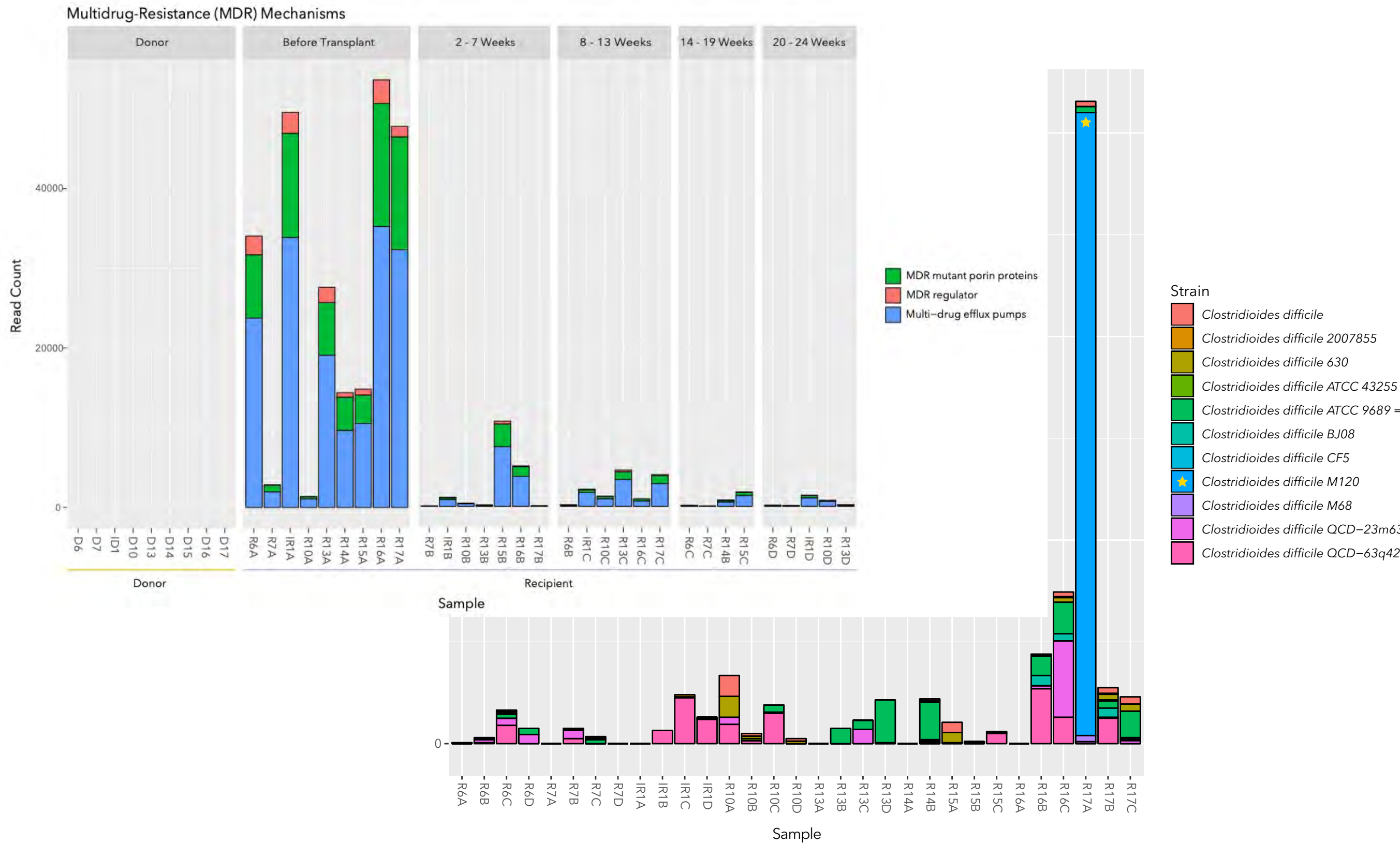


Potential pathogens identified

Potential Pathogens Identified



AMR and *C. diff.*



Linking Research and Education: Project-based Learning in Genomics

- Testing the 'Grandma Hypothesis'
- Do Grandma Hotspots have different microbiomes?

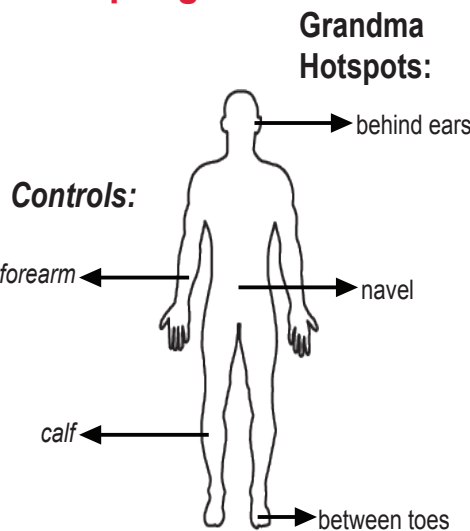



Project-Based Learning

STEP 1

Technology:
 Swabs

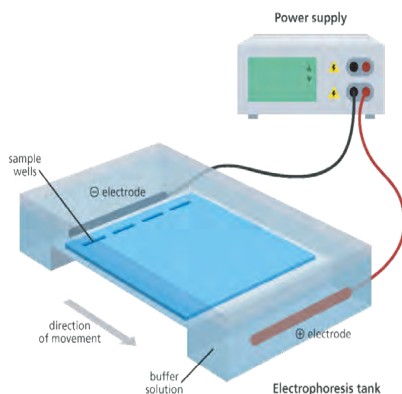
Activity:
Sampling



Formative Assessment
(FA):
 Lab notebook

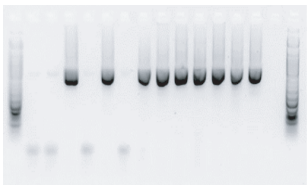
STEP 2

Technology:
Gel Electrophoresis



Activity:
DNA extraction &
PCR of 16S

FA:
Agarose gel



STEP 3

Technology:
DNA Sequencing



Activity:
Targeted
Amplicon
Sequencing

FA:
Sequence data
GGT**T**GATAG
GG**C**TGATAG
GGT**T**GATAT

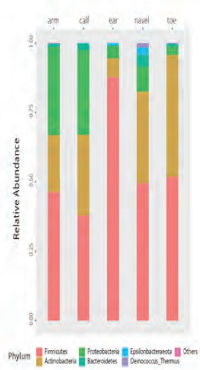
STEP 4

Technology:
Bioinformatics



Activity:
Bioinformatic
Microbiome
Characterization

FA:
Relative abundance



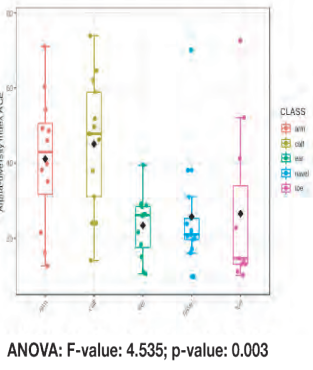
STEP 5

Technology:
Bioinformatics



Activity:
Statistical
Analysis

FA:
Statistical test

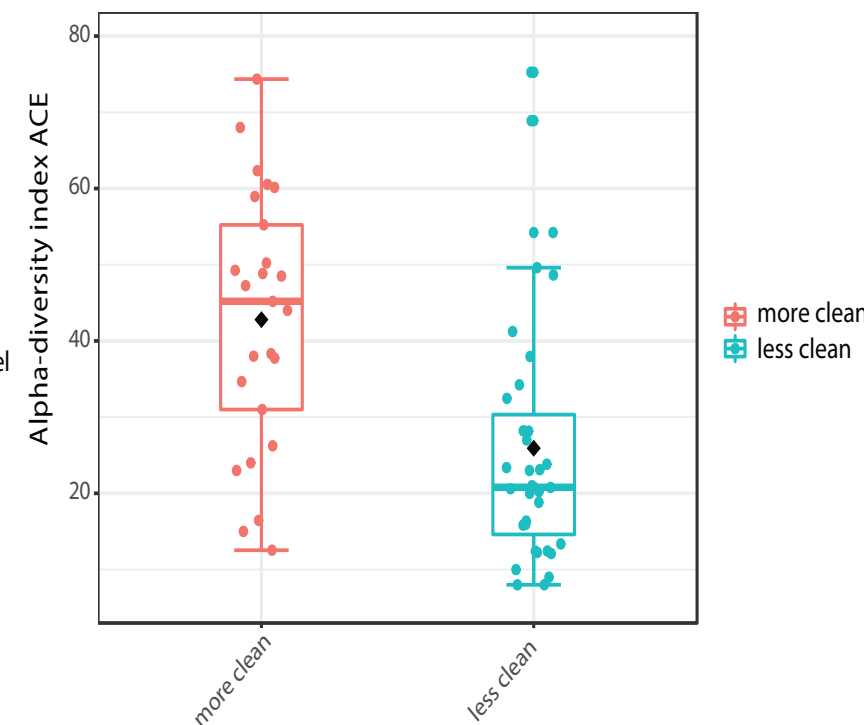
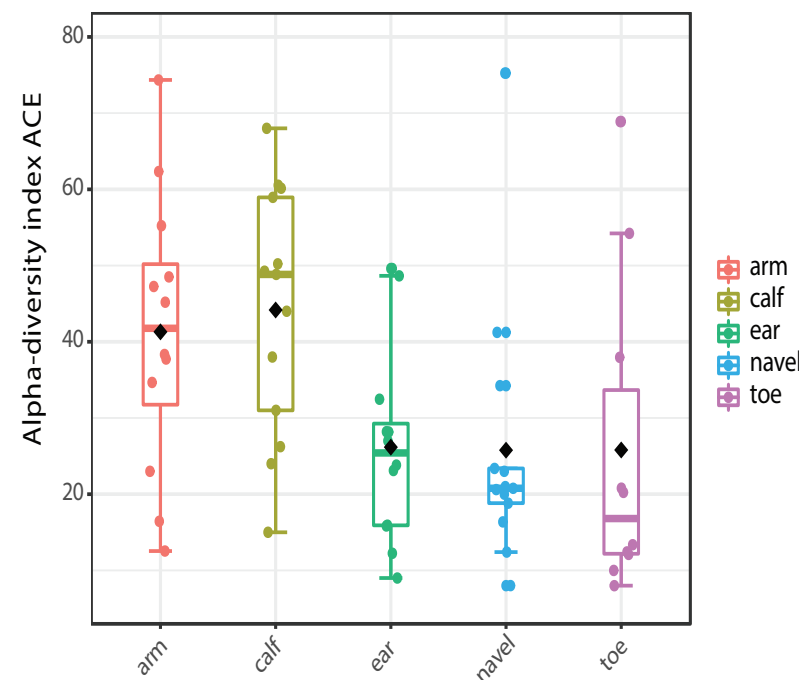
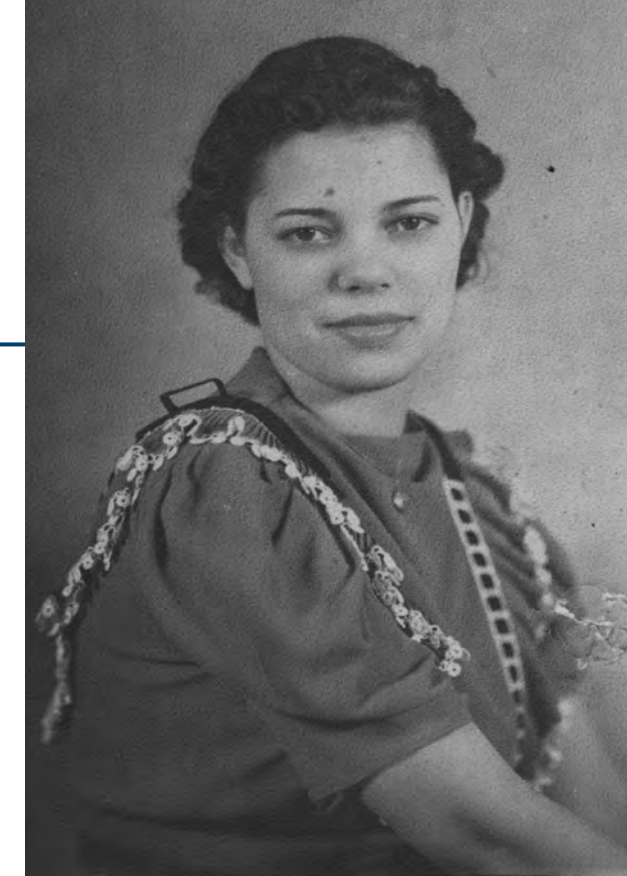


STEP 6

Summative
Assessment:
Scientific
report

Linking Research and Education: Project-based Learning in Genomics

- Significant differences between sections with PBL and without ($P < 0.05$ on two different assignments [$P = 0.040, 0.027$])
- Students had a more positive experience as reported on in-class surveys and post-course evaluations
- Significant difference in richness and evenness between Grandma hotspots and controls



Curriculum

DOI: <https://doi.org/10.1128/jmbe.v2i1.2019>

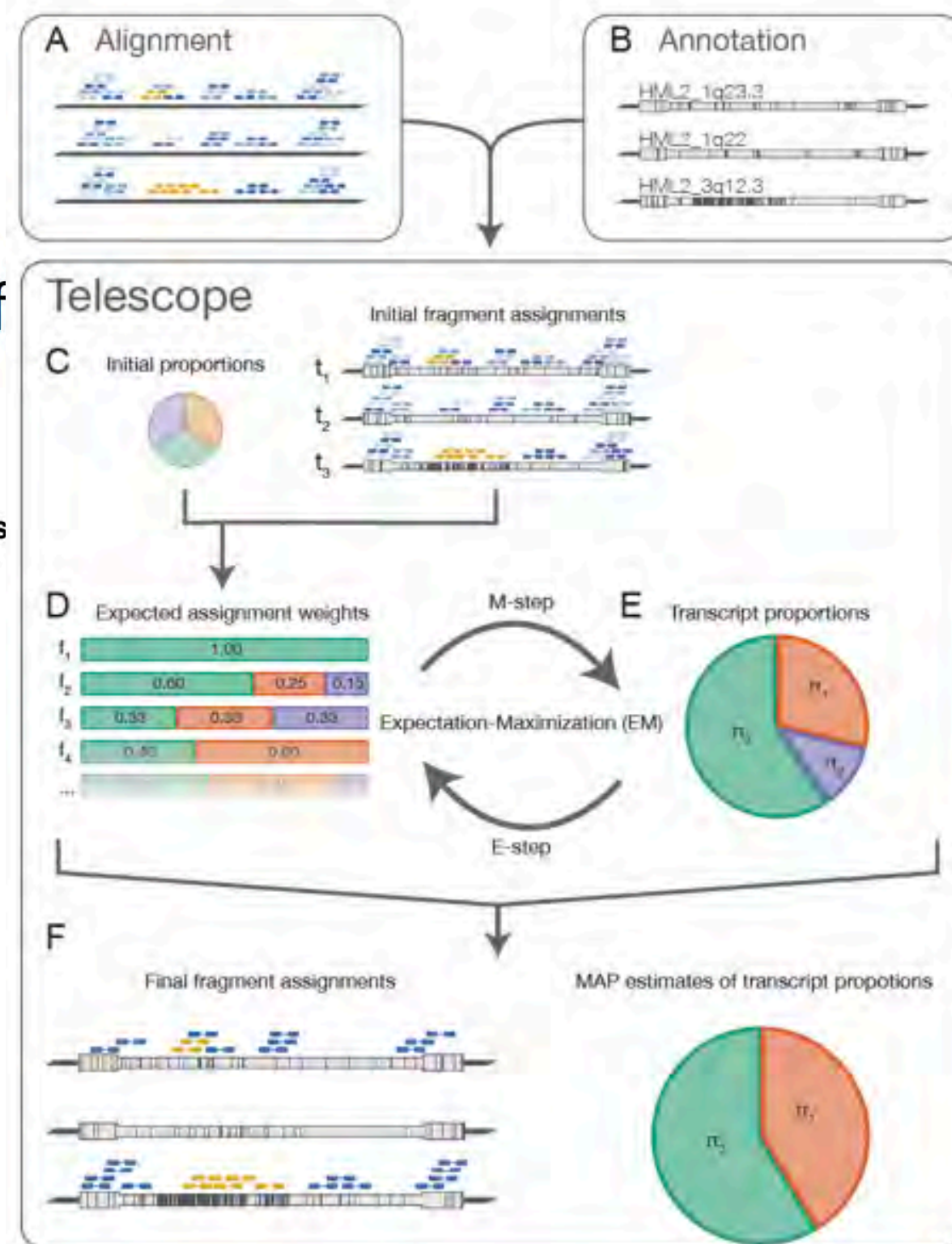
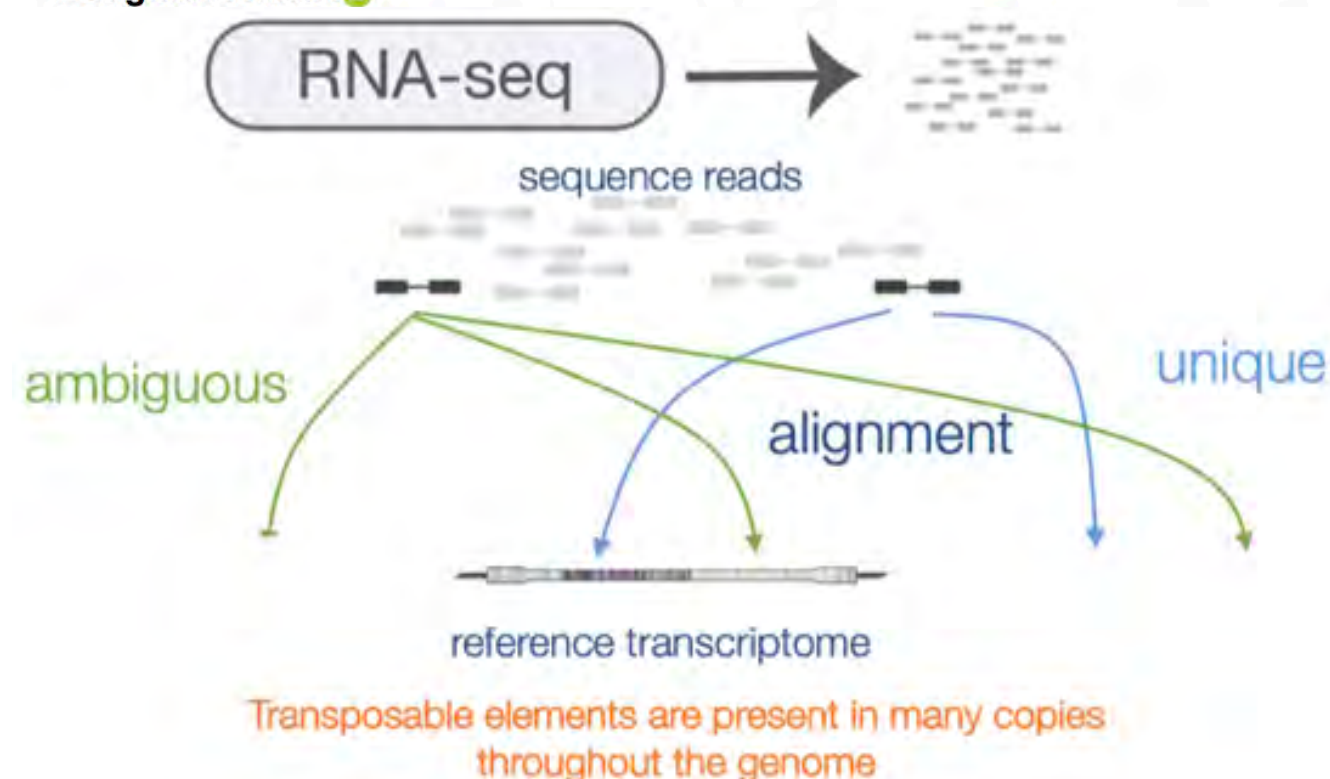


Testing the “Grandma Hypothesis”: Characterizing Skin Microbiome Diversity as a Project-Based Learning Approach to Genomics[†]

Telescope – method and implementation

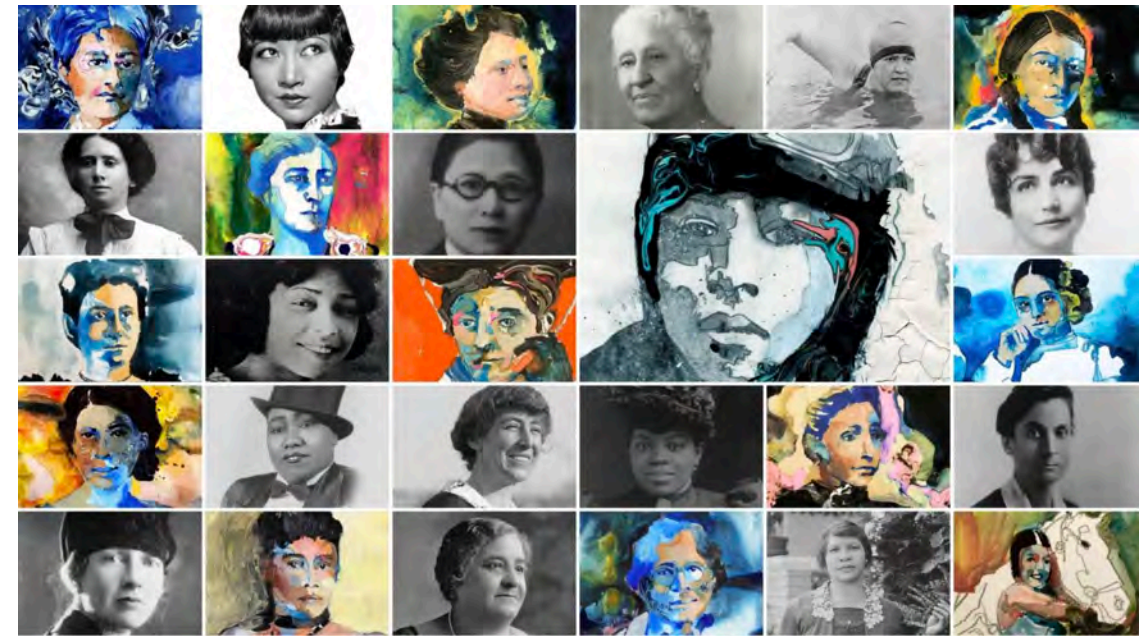
Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression

Matthew L. Bendall^{1,2*}, Miguel de Mulder², Luis Pedro Iñiguez^{2,3}, Aarón Lecanda-Sánchez³, Marcos Pérez-Losada^{1,4,5}, Mario A. Ostrowski^{6,7}, R. Brad Jones², Lubbertus C. F. Mulder^{8,9}, Gustavo Reyes-Terán³, Keith A. Crandall^{1,4}, Christopher E. Ormsby³, Douglas F. Nixon²



Transposable Elements in Health

- Transposable Elements (TE) make up 45% of the human genome
- Discovered by Barbara McClintock doing maize experiments
- Initially characterized as 'junk DNA'
- Expressed TEs cause disease (disrupt normal function)



Women's History Month



Barbara McClintock – 1983 Nobel Prize in Medicine (still the *only* woman to win unshared in this category!)

Review Article | Published: 12 September 2019

Transposable elements in human genetic disease

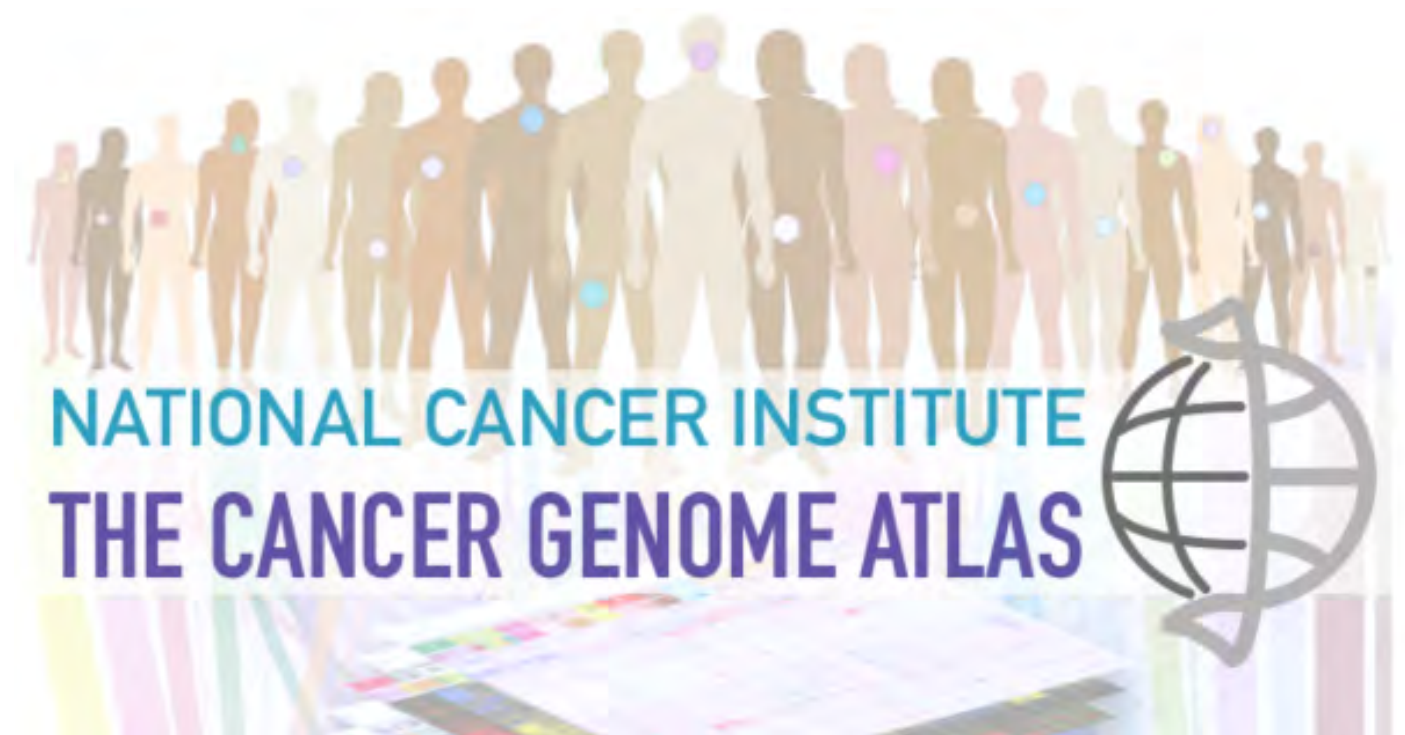
Lindsay M. Payer & Kathleen H. Burns 

Nature Reviews Genetics 20, 760–772(2019) | [Cite this article](#)

10k Accesses | 27 Citations | 67 Altmetric | [Metrics](#)

Telescope application: head and neck cancer

- Head and Neck Cancer impacts ~600,000 people per year world-wide
- Mortality rate ~ 50%
- HERV expression implicated in other cancers
- From TCGA, 43 paired tumor-normal RNA-seq data sets



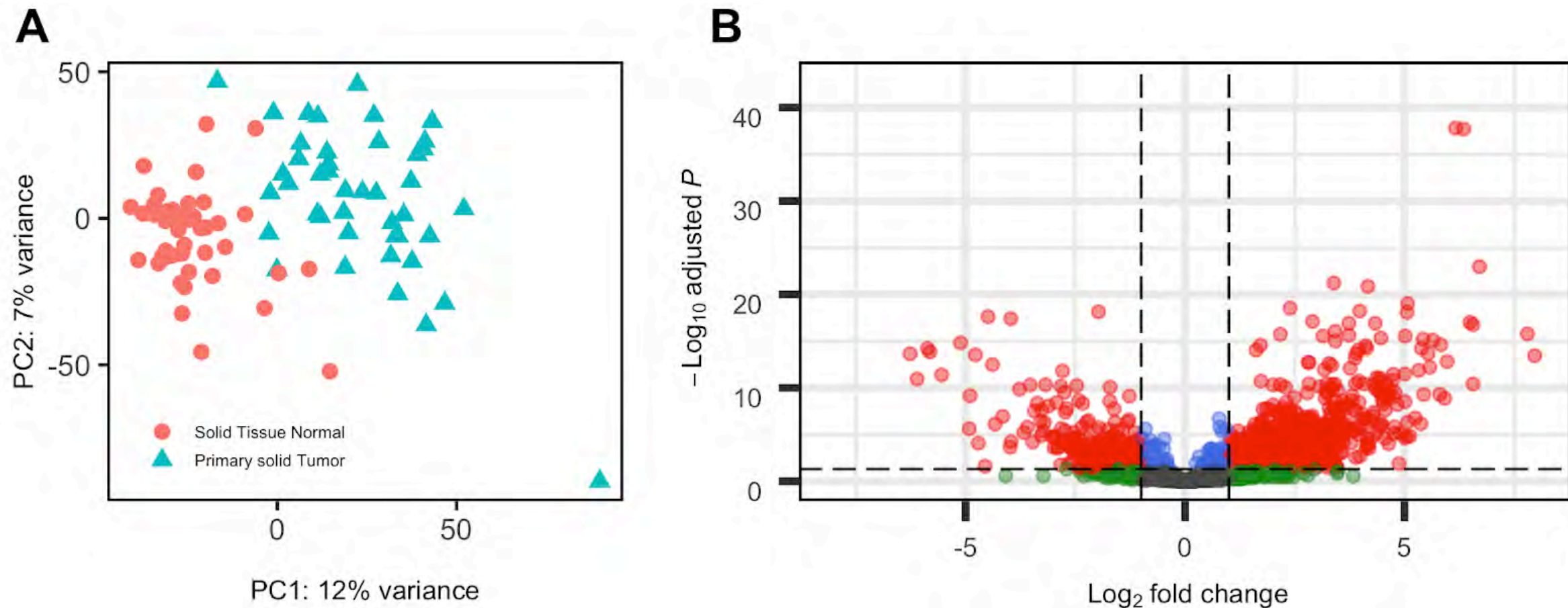
Article

Human endogenous retrovirus expression is associated with head and neck cancer and differential survival

Allison R. Kolbe¹, Matthew L. Bendall², Alexander T. Pearson³, Doru Paul⁴, Douglas F. Nixon², Marcos Perez-Losada^{1,5}, and Keith A. Crandall^{1,5*}

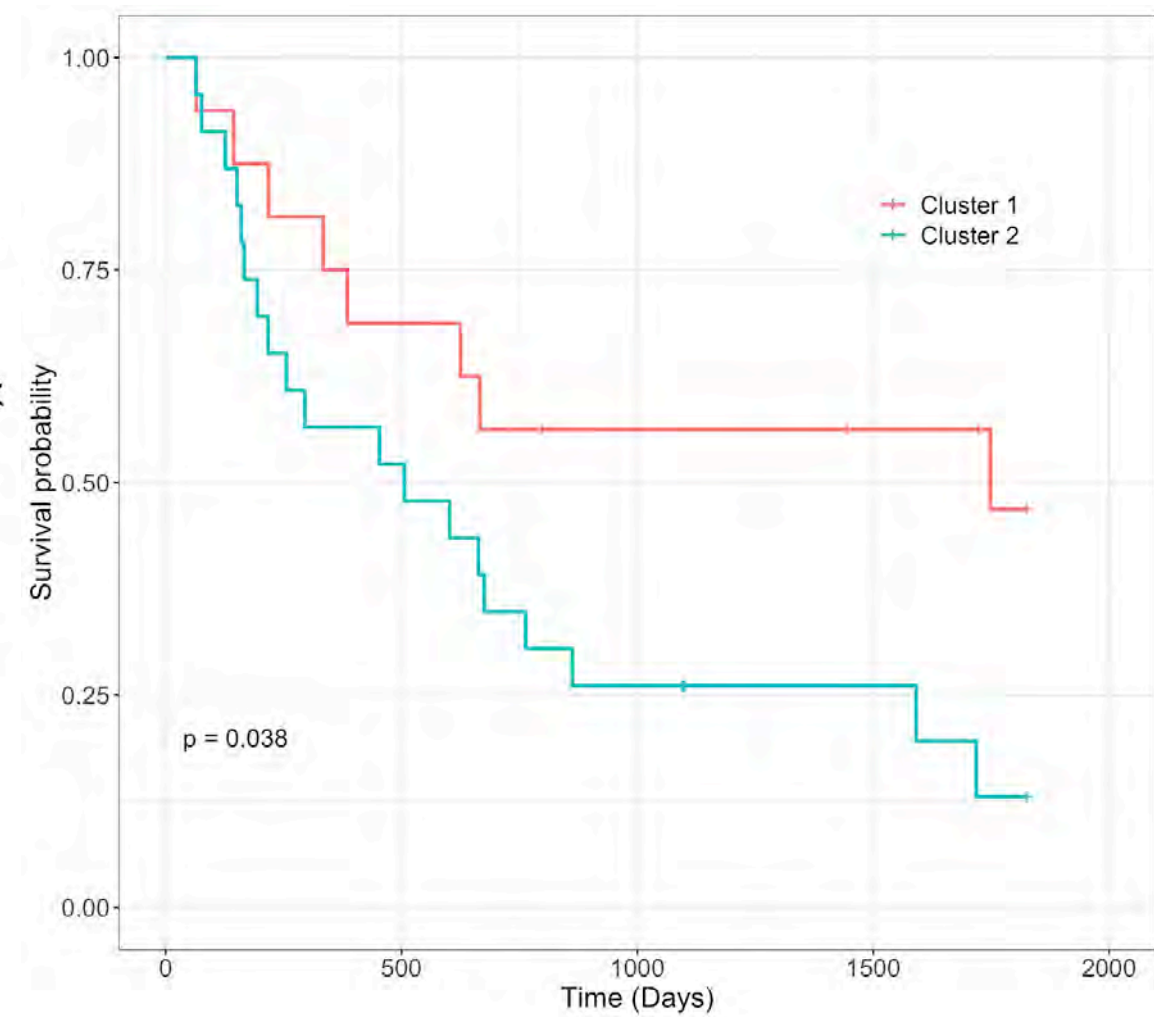
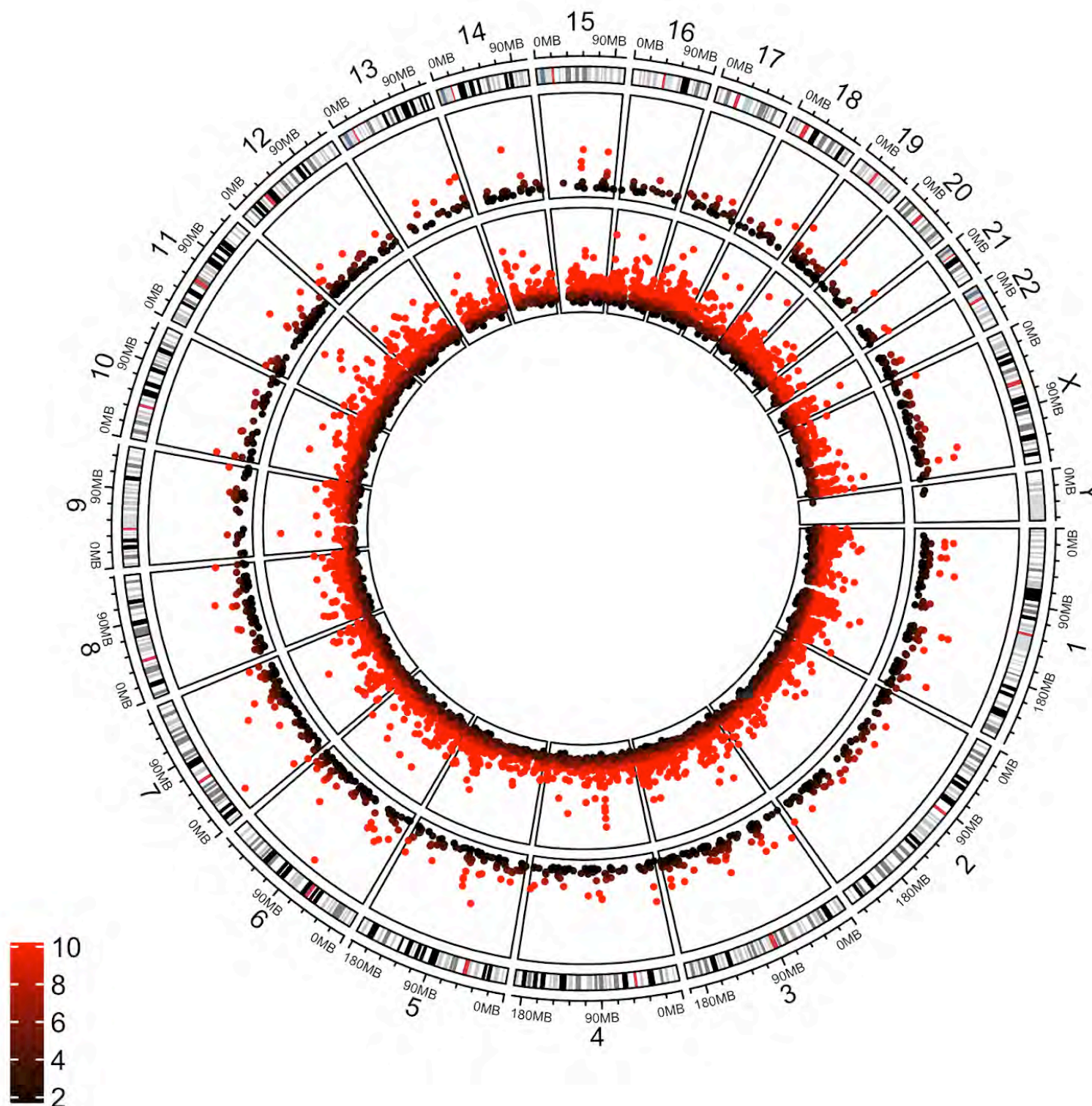


Telescope - head and neck cancer



- Telescope identified 3502 expressed HERVs
- 1078 differentially expressed between tumor and normal tissue
- 802 over expressed in tumor tissue

Telescope - head and neck cancer



Summary



- Metagenomics allow for characterization of microbiome *and* host at a taxonomic *and* functional level
- PathoScope is a useful and functional tool for diverse analysis of metagenomic data – a cog in the informatics pipeline!
- Telescope opens the door to investigations of the impact of the retrotranscriptome
- Informatics is an essential skill area for future biologists!
- <https://github.com/PathoScope>
- <https://github.com/mlbendall/telescope>

Thanks!

- Evan Johnson - PathoScope
 - Eduardo Castro Nallar - Asthma/PathoScope
 - Rob Freishtat, Marcos Pérez-Losada - Asthma
 - Keylie Gibson, Buddha Pukazhenthithi - Rhino
 - Matthew Bendall, Doug Nixon - Telescope
 - Michelle Ahn, Suchi Hourigan, Maria Oliva-Hemker - FMT
-
- NIH, DHS, Inova, The Rhino Foundation



**Weill
Cornell
Medicine**



JOHNS HOPKINS
SCHOOL of MEDICINE



**Universidad
Andrés Bello**



**South African
NATIONAL PARKS**



Smithsonian



**Computational
Biology Institute**

THE GEORGE WASHINGTON UNIVERSITY



Childrens National



CTSI

**Clinical and Translational
Science Institute**
at Children's National

A PARTNERSHIP WITH THE GEORGE WASHINGTON UNIVERSITY



National Institutes of Health
NIH...Turning Discovery into Health