THE Genome CENTER AT WASHINGTON UNIVERSITY

The characterization of the human virome in children and adults

Kristine M. Wylie
Kathie Mihindukulasuriya
Erica Sodergren
George M. Weinstock
Gregory Storch

Background

- Viruses make up an important part of the human microbiome
  - Healthy individuals
  - Disease
- Many new viruses have been discovered in recent years using high-throughput sequencing
- Sequencing offers advantages to studying the human virome
  - PCR assays focus on the most common viruses (pathogens)
  - Potential for discovery of new and unexpected viruses
  - Can sometimes determine additional information about the virus
    - virus subtype
    - sequence variation from reference genomes

Kristine M. Wylie kwylie@genome.wustl.edu

Purpose

- To develop an analysis pipeline to detect known and novel viruses in metagenomic samples
- To apply this analysis pipeline to the study of the human virome
  - Children with fever and afebrile control children
    - Nasopharyngeal swabs
    - Plasma
  - Human Microbiome Projects whole genome shotgun data from normal subjects
    - Gastrointestinal
    - Oral
    - Vaginal
    - Skin
    - Nasal

Kristine M. Wylie kwylie@genome.wustl.edu

454 vs. Illumina platforms

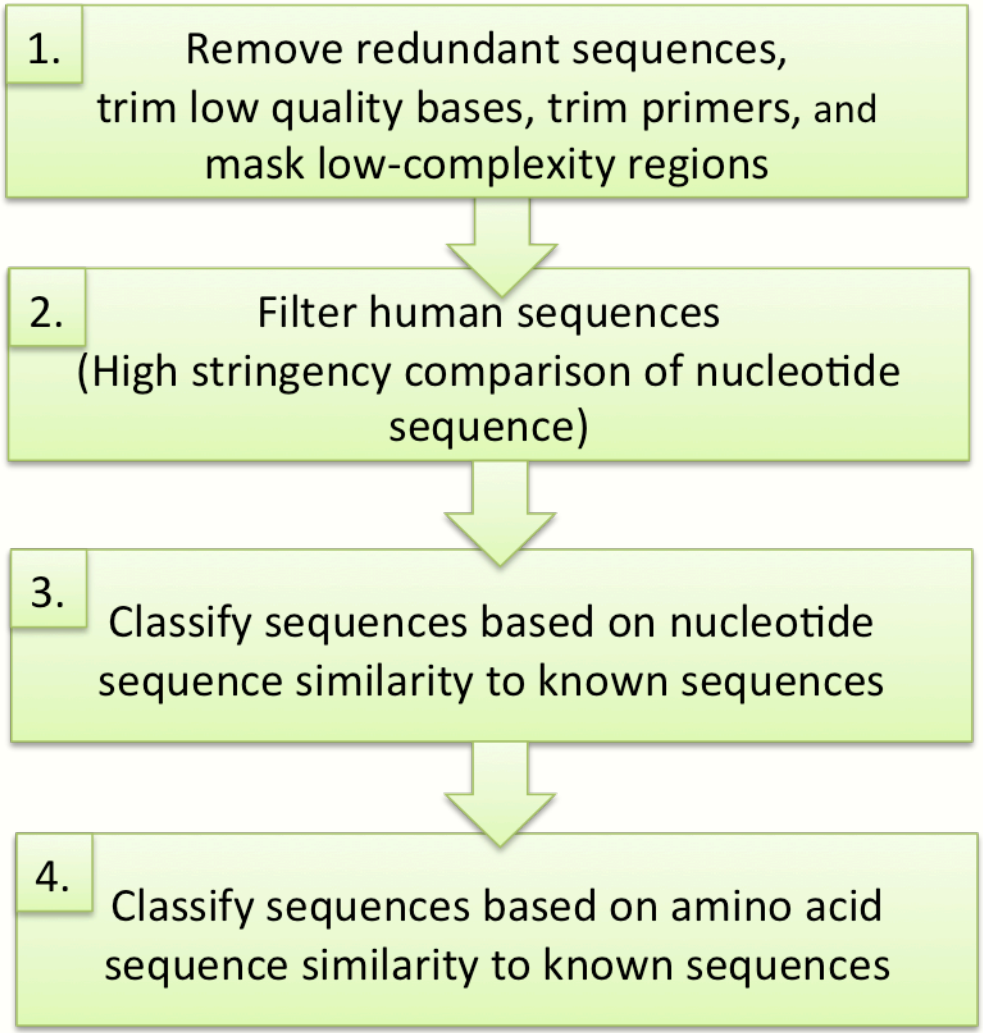
- Extracted total nucleic acids
- cDNA synthesis and amplification
- Library construction and sequencing
- Alignment of sequences to reference genomes and manual review

	454		100 base Illumina		
	Total sequences	Virus sequences	Total sequences	Virus sequences	Virus genome completeness
Enterovirus	52,275	7	4,792,380	342	42%
Human herpesvirus 6	56,614	1	3,931,804	8	NA

- Chose to generate ~5 million reads per sample on the Illumina platform

Kristine M. Wylie kwylie@genome.wustl.edu

Analysis



Important elements of analysis:

- High-speed aligners that allow for more mismatch
  - Initially used cross_match instead of blastn
  - Now integrating RTG map from Real-Time Genomics
- High-speed amino acid alignment
  - used tblastx for most analysis
  - accelerated blastx from Multi Core Ware used for some analysis
  - accelerated tblastx being developed
- Manual review of results and improved understanding of reference databases

Kristine M. Wylie kwylie@genome.wustl.edu

Febrile children study - background

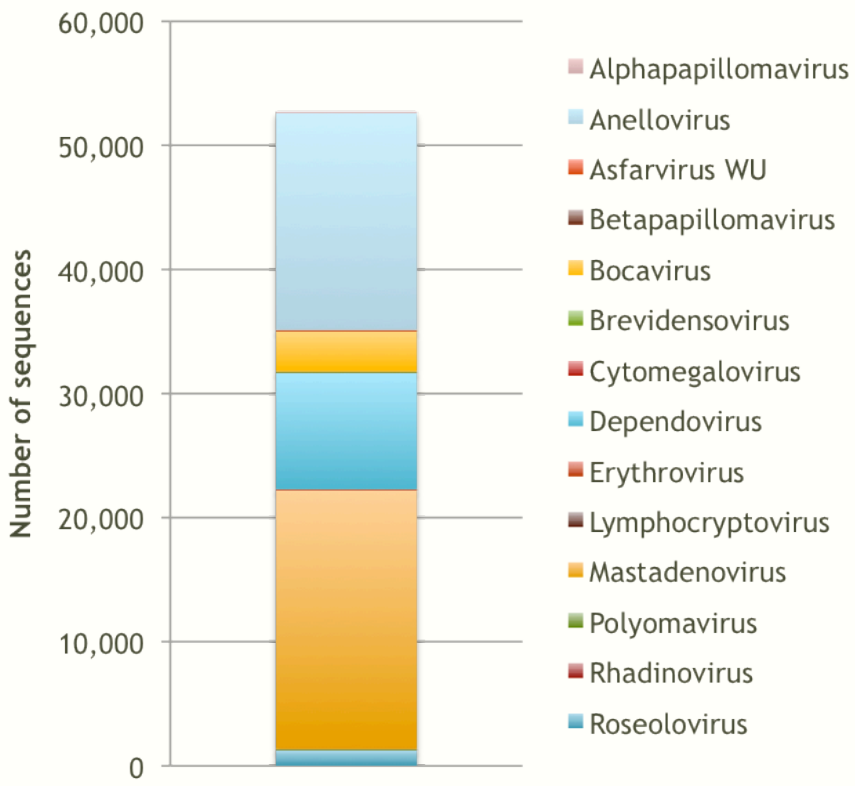
- Fever without a focus
  - Common reason for emergency room visits in children under 3 years old
  - 85-90% of these fevers are unexplained
  - Cause thought to be viral, but no comprehensive analysis has been done
- Samples
  - Nasopharyngeal swabs and plasma
  - Febrile children and afebrile controls

	Afebrile	Febrile	Total
Nasopharyngeal	82	67	149
Plasma	22	28	50
Total	104	95	199

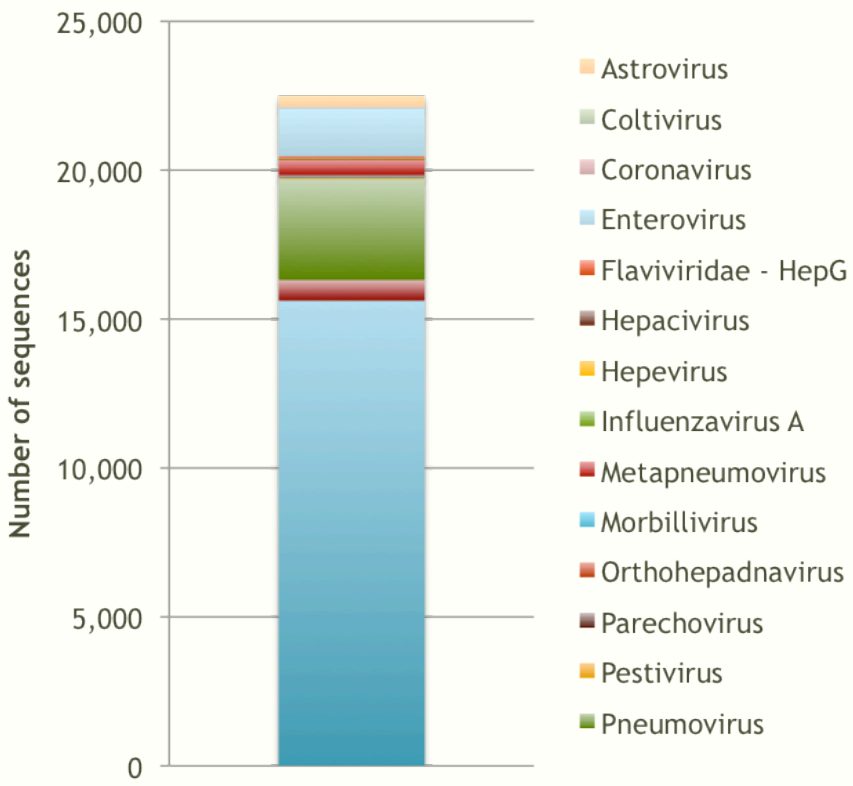
Kristine M. Wylie kwylie@genome.wustl.edu

Viruses detected in samples from febrile and afebrile children

A



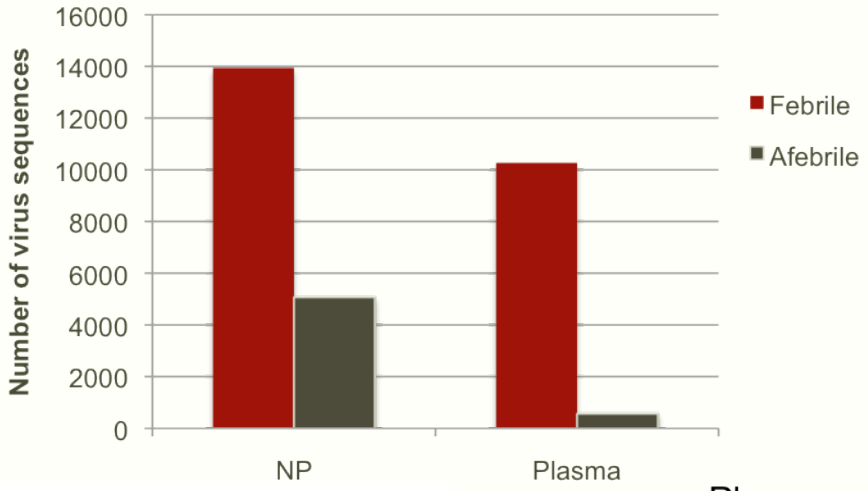
B

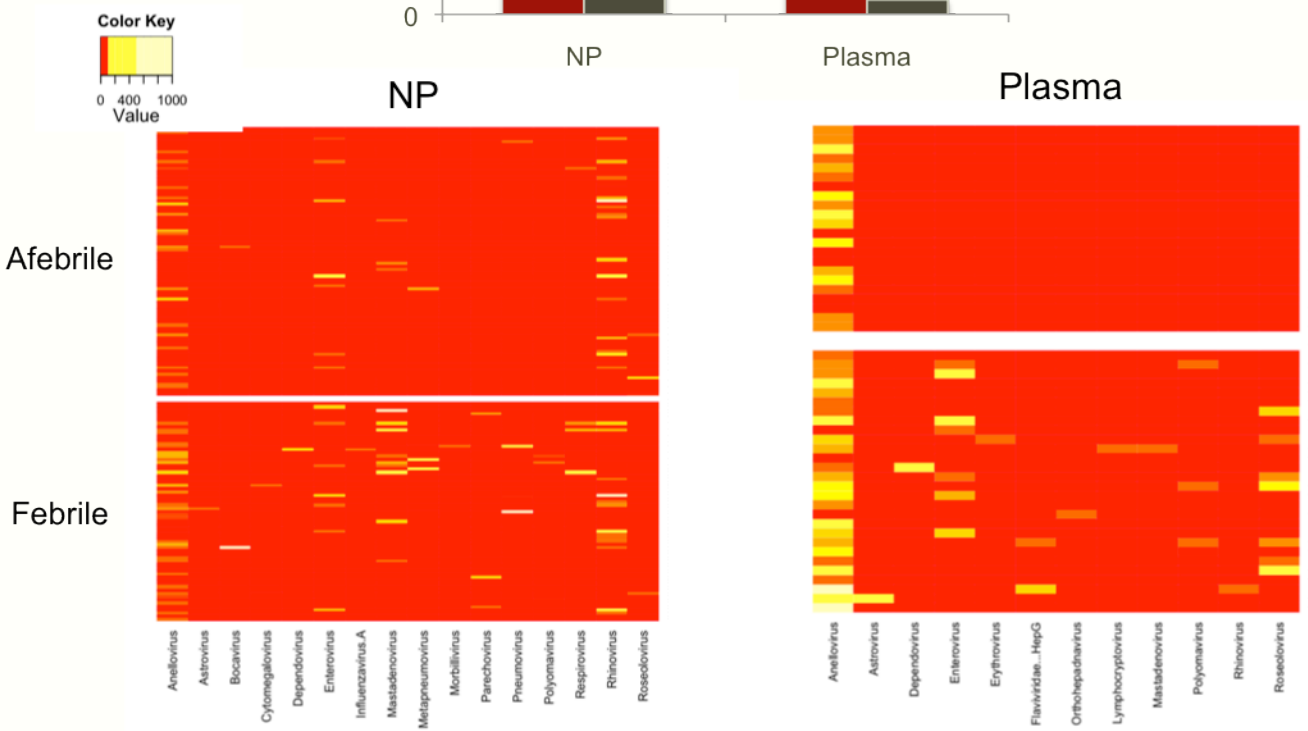


Kristine M. Wylie kwylie@genome.wustl.edu

THE Genome CENTER AT WASHINGTON UNIVERSITY

Virus sequences are more abundant in febrile children

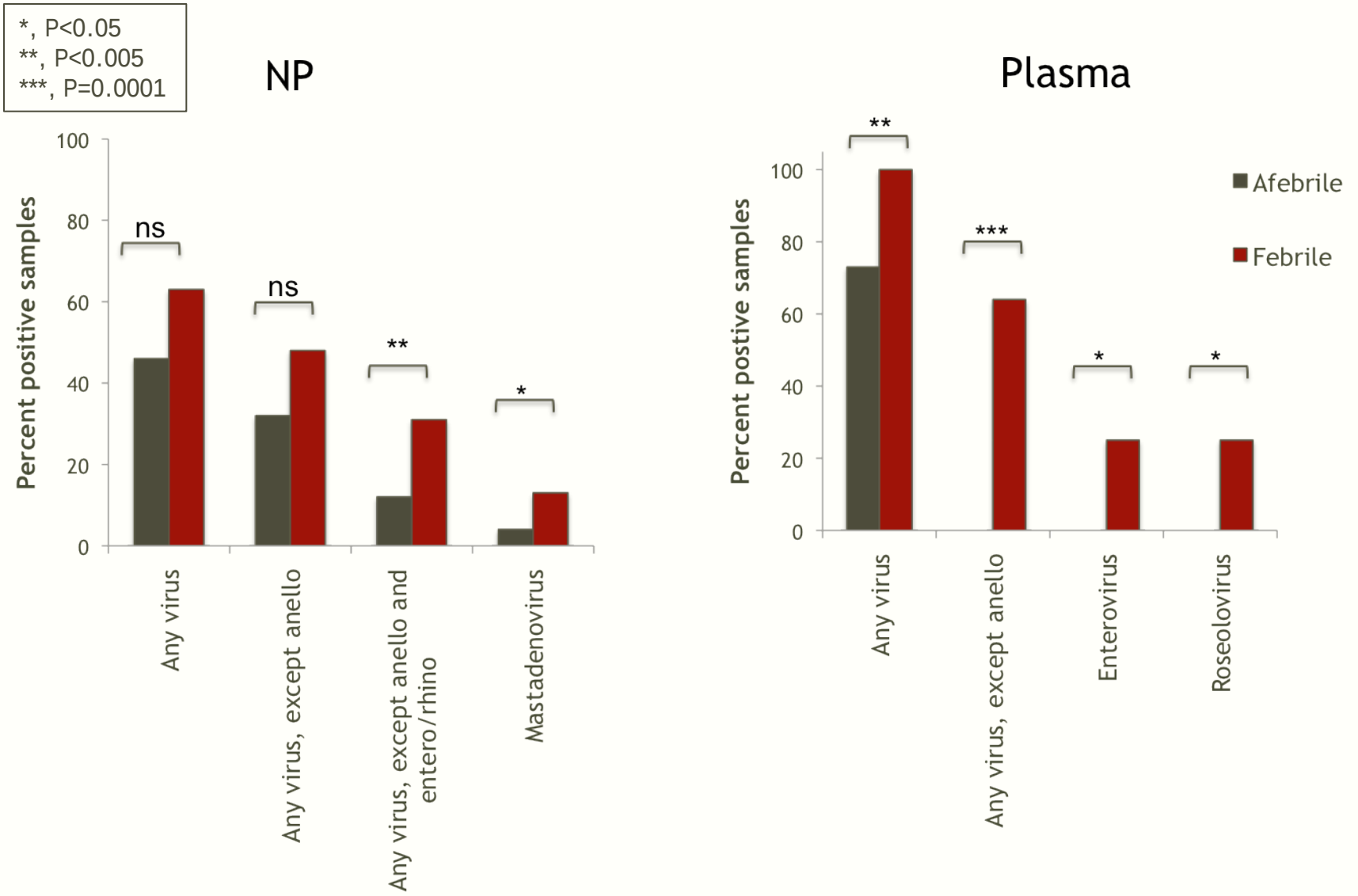




Kristine M. Wylie kwylie@genome.wustl.edu

THE Genome CENTER AT WASHINGTON UNIVERSITY

Viruses more commonly associated with febrile children

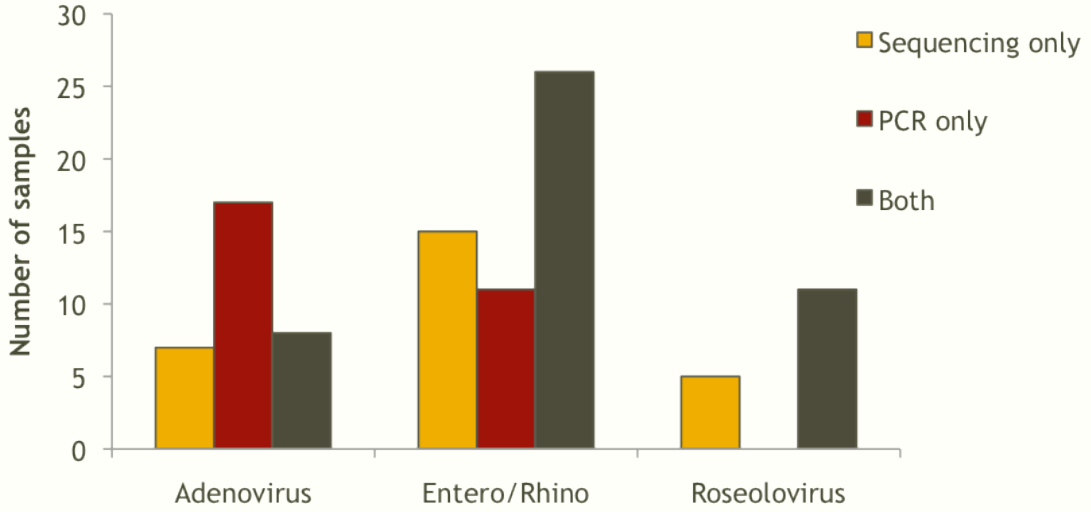


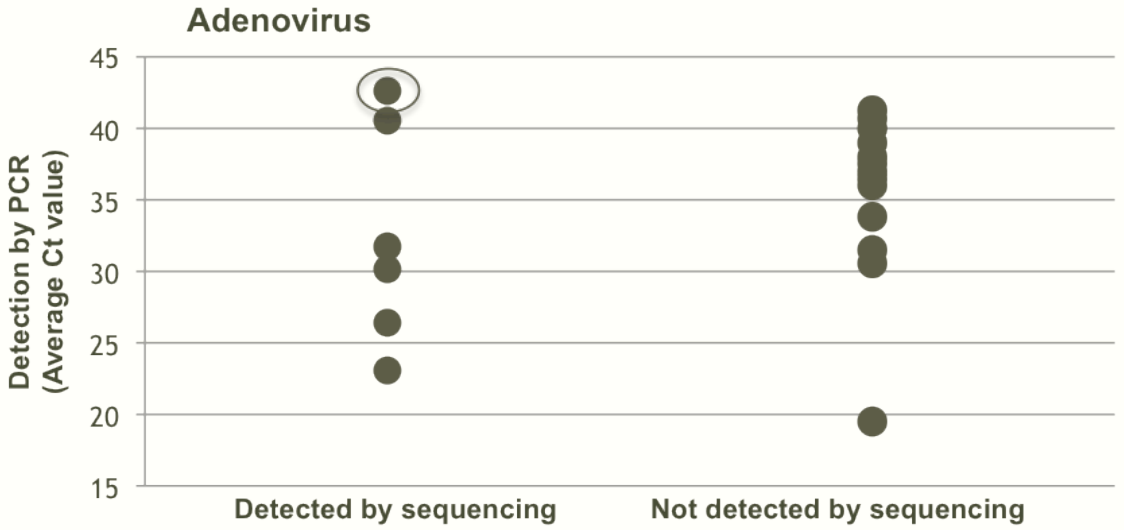
Kristine M. Wylie kwylie@genome.wustl.edu

THE Genome CENTER AT WASHINGTON UNIVERSITY

Correlation of metagenomic sequencing data with PCR data

Common viruses

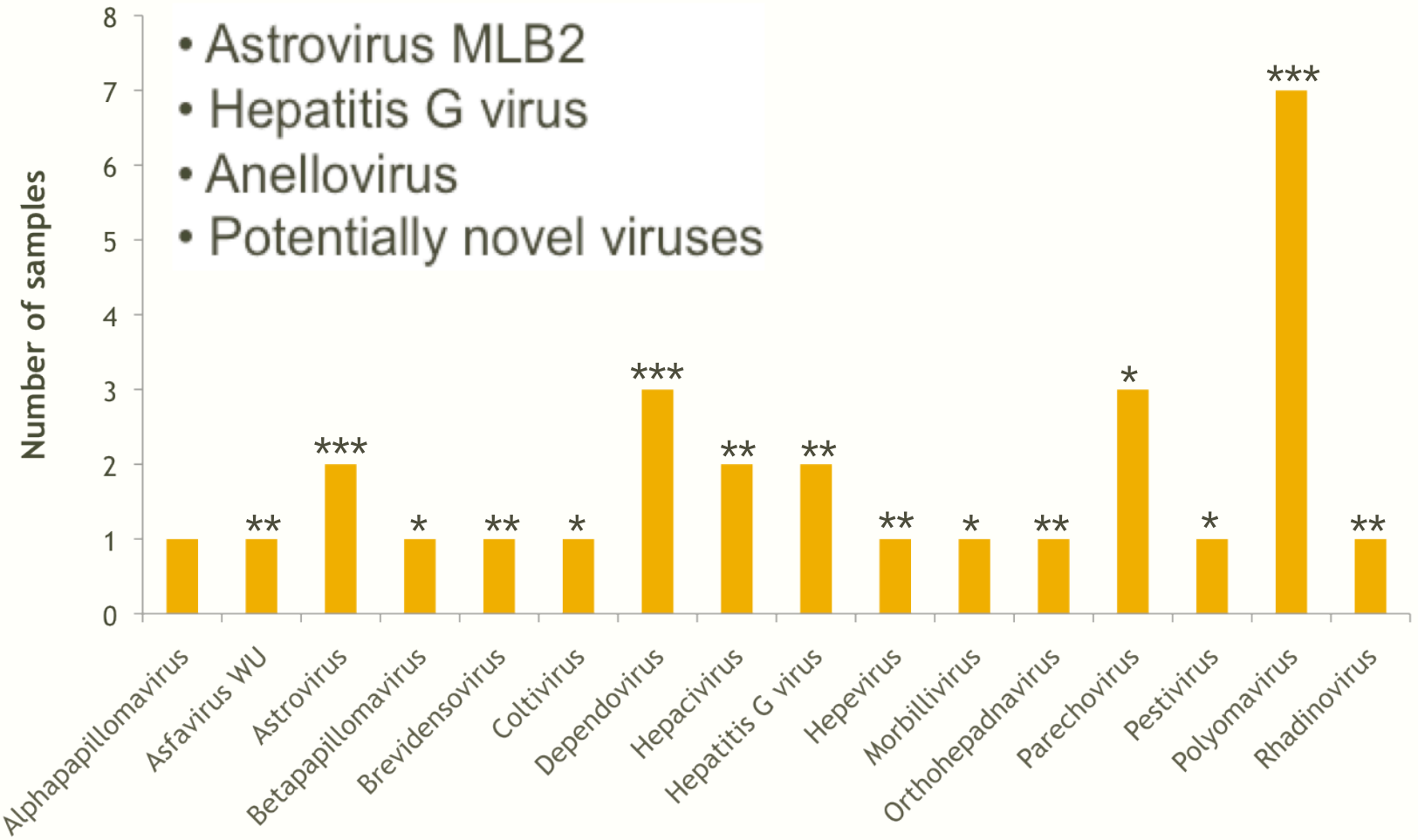




Kristine M. Wylie kwylie@genome.wustl.edu

THE Genome CENTER AT WASHINGTON UNIVERSITY

Metagenomic sequencing allows detection of unexpected viruses



No star NP afebrile
* NP febrile
** Plasma febrile
*** NP and Plasma febrile
Hepacivirus also NP afebrile (translated alignment)

Kristine M. Wylie kwylie@genome.wustl.edu

THE Genome CENTER AT WASHINGTON UNIVERSITY

Additional information gained by sequencing

- Genome coverage

Genome	Contigs	Smallest Contig	Largest contig	Genome size	Completeness	Depth of coverage
Human bocavirus	5	100 nt	1886 nt	5299 nt	92.6%	63X
Respiratory Syncytial virus	24	102 nt	1882 nt	15,191 nt	58.4%	19.5X
Human rhinovirus QPM	2	798 nt	5962 nt	6948 nt	94.5%	112X
Human parainfluenza virus	10	105 nt	803 nt	15,462 nt	14.2%	2X

- Genome subtype
  - o HHV-6
  - o Adenovirus
  - o Enterovirus/Rhinovirus

Kristine M. Wylie kwylie@genome.wustl.edu

Febrile children study – summary and conclusions

- Viruses are more diverse and abundant in samples from febrile children compared with afebrile children.
- Enterovirus/rhinoviruses and roseoloviruses are statistically more common in plasma samples from febrile children compared with samples from afebrile children (consistent with PCR data).
- Adenoviruses are more commonly found in NP samples from febrile children compared with samples from afebrile children.
- Short read Illumina sequencing improves the sensitivity of sequencing (consistent with PCR data).
- PCR and sequencing can be used together to generate a more complete description of the microbiome.
- Advantages of sequencing include:
  - ability to detect unexpected viruses
  - additional genome sequence/virus subtype information

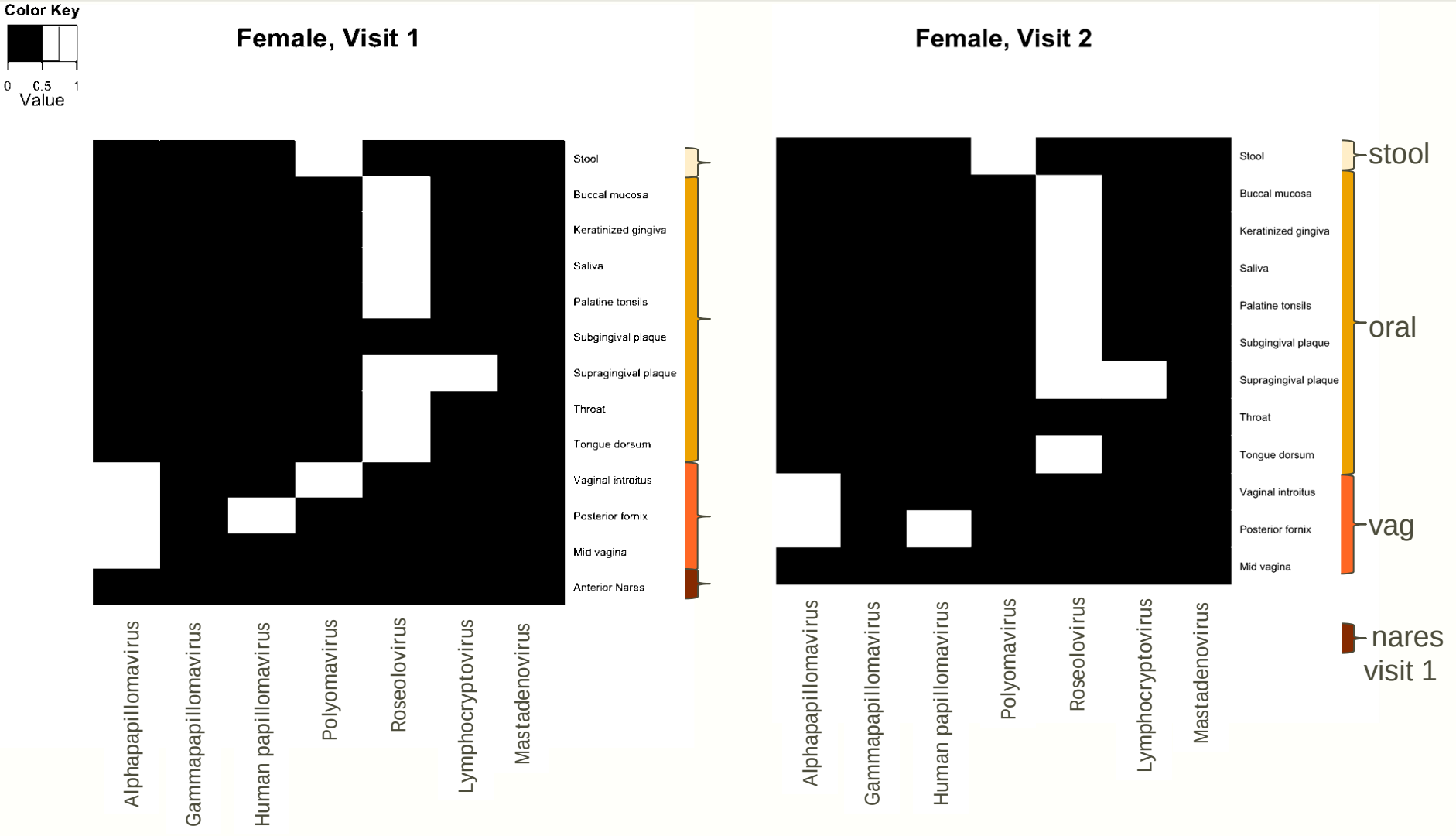
Kristine M. Wylie kwylie@genome.wustl.edu

Human Microbiome project normal subjects - background

- 743 samples
  - 100 subjects
  - 5 major body sites
- DNA sequenced
  - 2 lanes of Illumina sequencing per sample
- Sequences processed by HMP consortium
  - Quality trimming
  - Mask low complexity sequences
- Screened for viruses
  - Analysis ongoing

Kristine M. Wylie kwylie@genome.wustl.edu

Virome of an individual over two visits

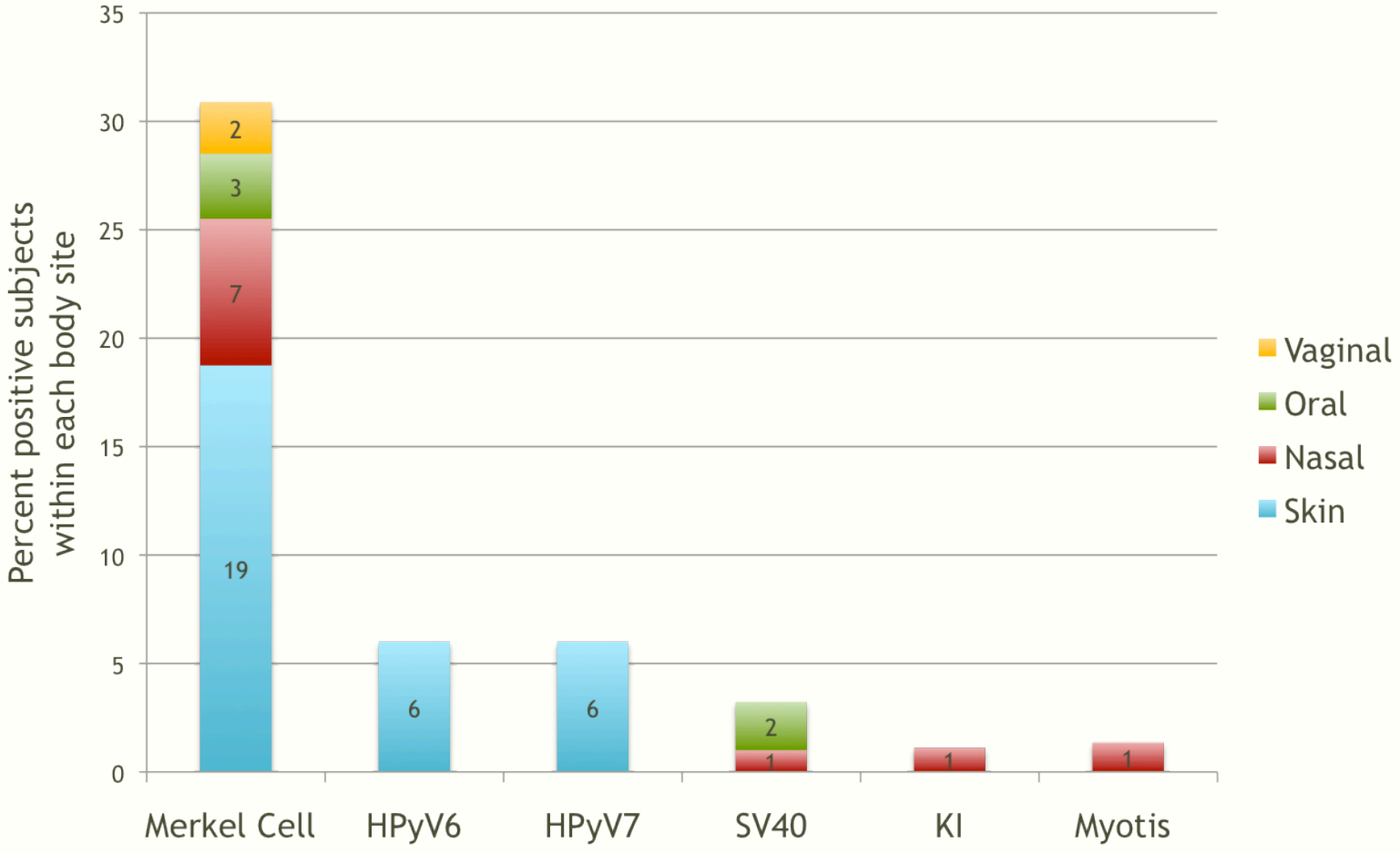


Presence/absence
White=virus present

Kristine M. Wylie kwylie@genome.wustl.edu

THE Genome CENTER AT WASHINGTON UNIVERSITY

Distribution of polyomaviruses among body sites

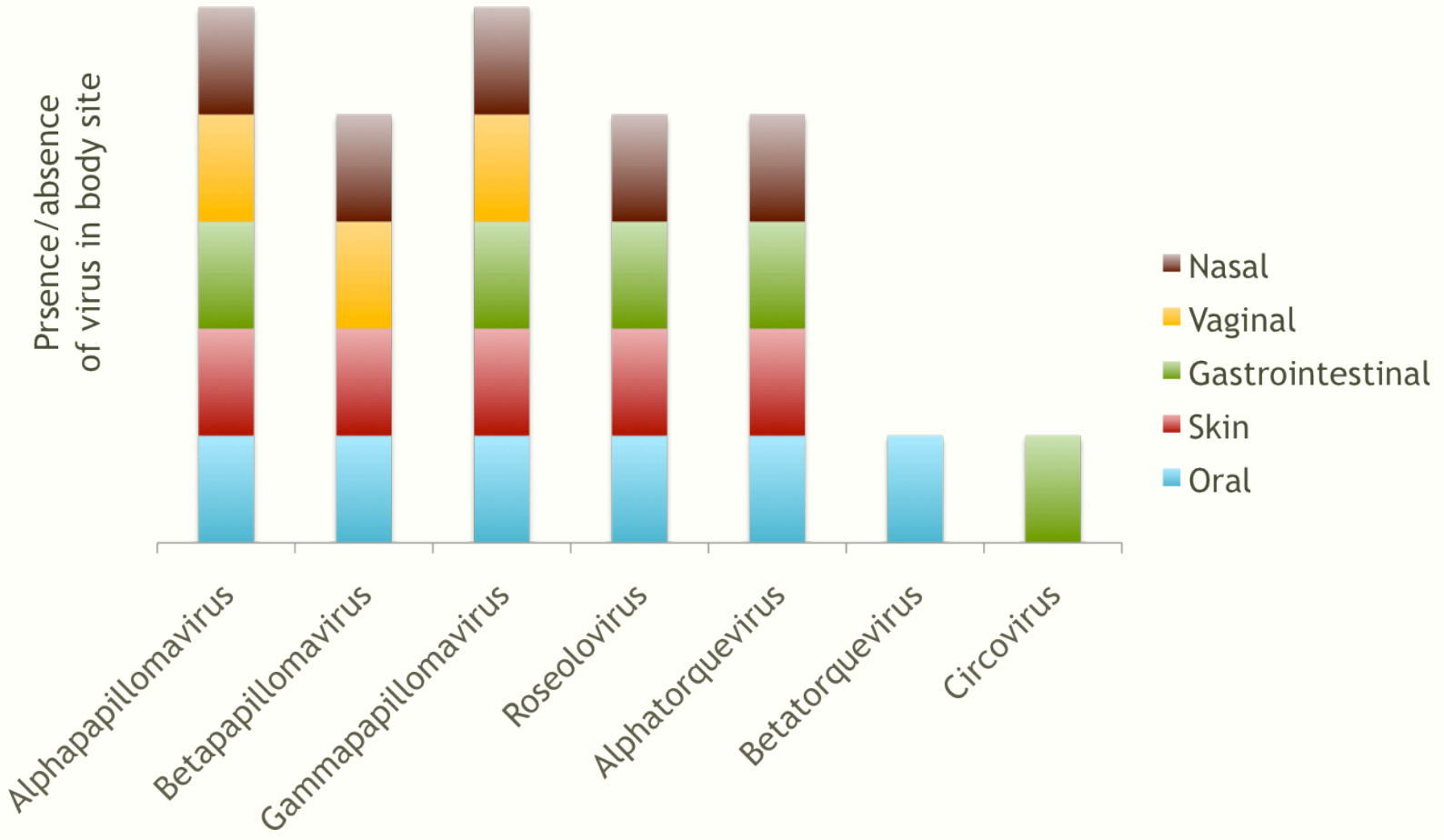


Additional analysis ongoing.

Kristine M. Wylie kwylie@genome.wustl.edu

THE Genome CENTER AT WASHINGTON UNIVERSITY

Body site distribution of common viruses

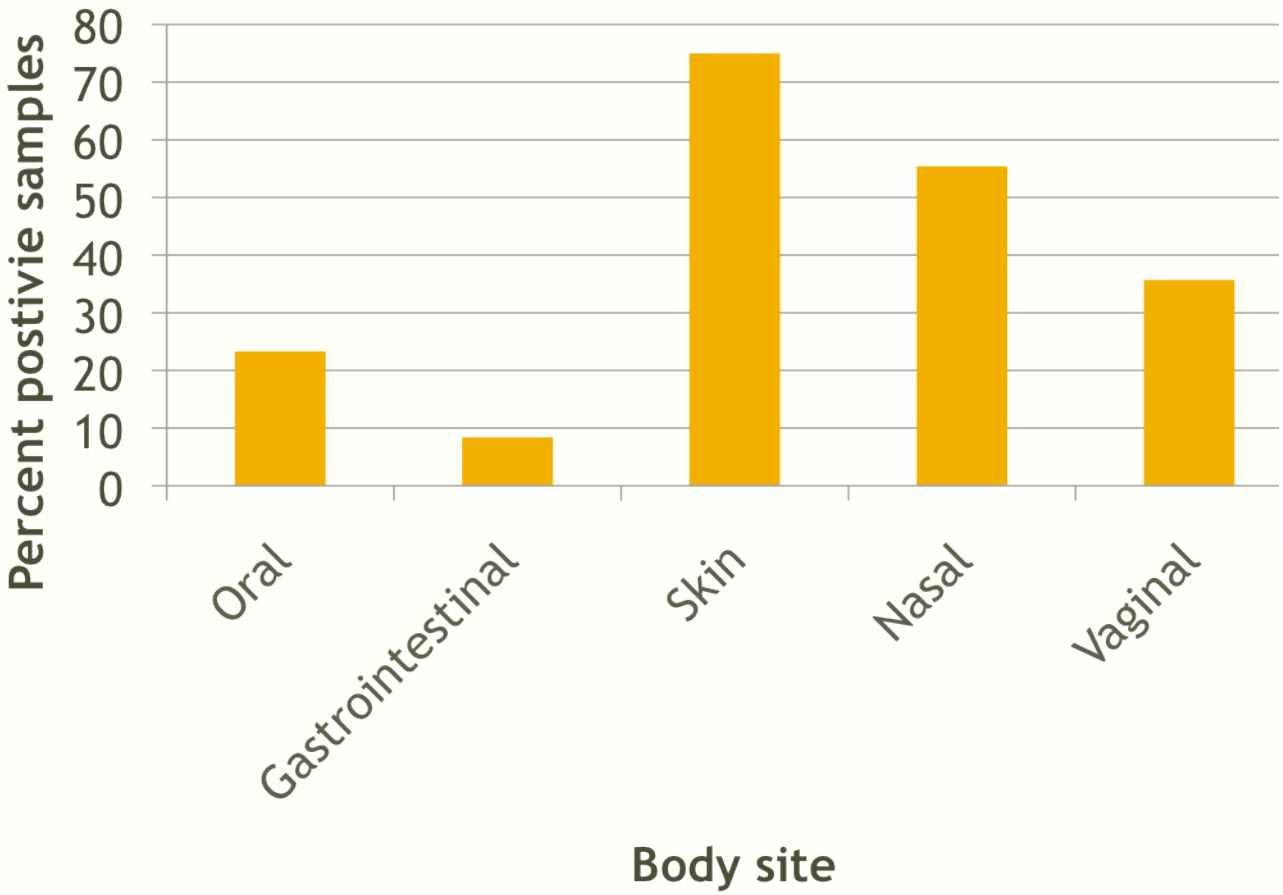


Additional analysis ongoing; data may be added.

Kristine M. Wylie kwylie@genome.wustl.edu

THE Genome CENTER AT WASHINGTON UNIVERSITY

Papillomaviridae distribution and variation



Alphapapillomavirus
Betapapillomavirus
Deltapapillomavirus
Dyodeltapapillomavirus
Dyoetapapillomavirus
Epsilonpapillomavirus
Gammapapillomavirus
Lambdapapillomavirus
Phipapillomavirus
Pipapillomavirus
Rhopapillomavirus
Taupapillomavirus
Upsilonpapillomavirus
Xipapillomavirus

Additional analysis ongoing; data may be added.

Kristine M. Wylie kwylie@genome.wustl.edu

THE Genome CENTER AT WASHINGTON UNIVERSITY

HMP adult subjects - summary and conclusions

- The virome analysis pipeline developed for the febrile children study can be applied to multiple sample types.
- The extensive sample collection of the HMP allows us to begin to define the human virome and its stability in a range of body sites from the same individual.
  - How common and abundant are these viruses?
  - How much sequence variation do we detect in viruses from the same individual at different body sites?
  - How do these data relate to what has previously been described in the literature?
  - How do chronic infections affect human health?
- Recently discovered viruses are detected in these samples, and potentially novel virus sequences have been identified that will be explored in future work.

Kristine M. Wylie kwylie@genome.wustl.edu

THE Genome CENTER AT WASHINGTON UNIVERSITY

Acknowledgements

- The Human Microbiome Project Consortium
- The Genome Center
  - Brandi Herter
  - Makedonka Mitreva
  - John Martin
  - Sahar Abubucker
  - Karthik Kota
  - Todd Wylie
  - Eric Becker
  - Matt Callaway
  - Otis Hall
  - Hongyu Gao
  - Yanjiao Zhou
- Collaborators
  - David Wang
- Emergency Medicine
  - Joshua Colvin
  - Jared Muenzer
  - David Jaffe
- SLCH Virology Laboratory
  - Richard Buller
  - Monique Gaudreault
- Virome Laboratory
  - Max Arens
  - Maria Cannella
- Study coordinator
  - Avraham Smason

Kristine M. Wylie kwylie@genome.wustl.edu

THE Genome CENTER AT WASHINGTON UNIVERSITY