



UNIVERSITY OF COPENHAGEN



Investigation of viruses in human cancers by application of multiple pre-sequencing enrichment methods

Sarah Mollerup

Centre for GeoGenetics

Natural History Museum of Denmark

University of Copenhagen

3rd International Conference on Clinical Metagenomics

18 October 2018

GENOME
denmark



Innovation Fund Denmark
RESEARCH, TECHNOLOGY & GROWTH

Centre for **Geo**
Genetics

Carcinogenic viruses

Epstein-Barr virus (Human herpesvirus type 4)

Kaposi sarcoma herpesvirus (Human herpesvirus type 8)

Hepatitis B virus

Hepatitis C virus

Human papillomaviruses (certain types, alphapapillomaviruses)

Human T-cell lymphotropic virus type I

Human immunodeficiency virus type 1

Probably/possibly carcinogenic

Merkel cell polyomavirus

BK and JC polyomavirus

Human papillomaviruses (certain types, mainly alpha)

NGS – viruses in cancer

Merkel cell polyomavirus discovered in 2008 using NGS

Feng et al 2008, Science

Several large-scale studies of viral expression in TCGA repository data

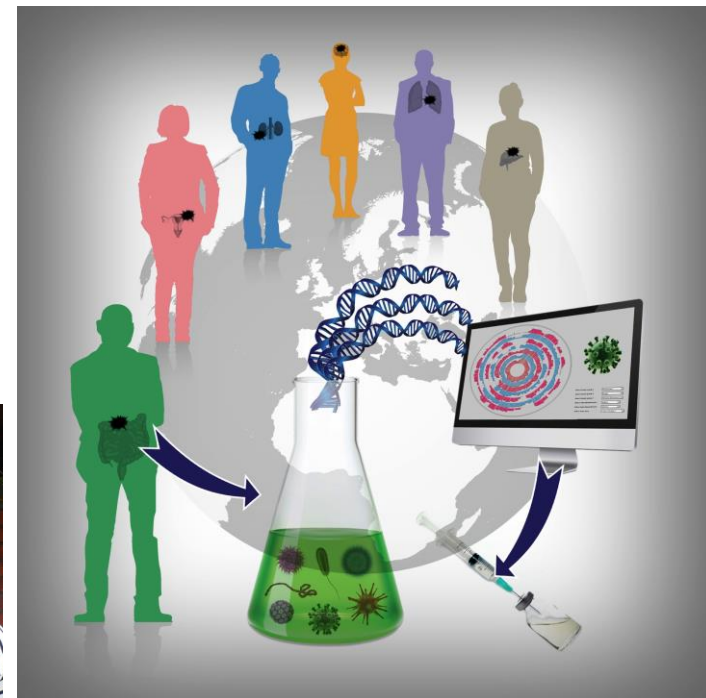
Tang et al 2012, Nat Commun; Khoury et al 2013, J Virology; Strong et al 2016, Acta Neuropath Commun; Cantalupo et al 2018, Virology

Confirmed established associations, but not revealed novel associations

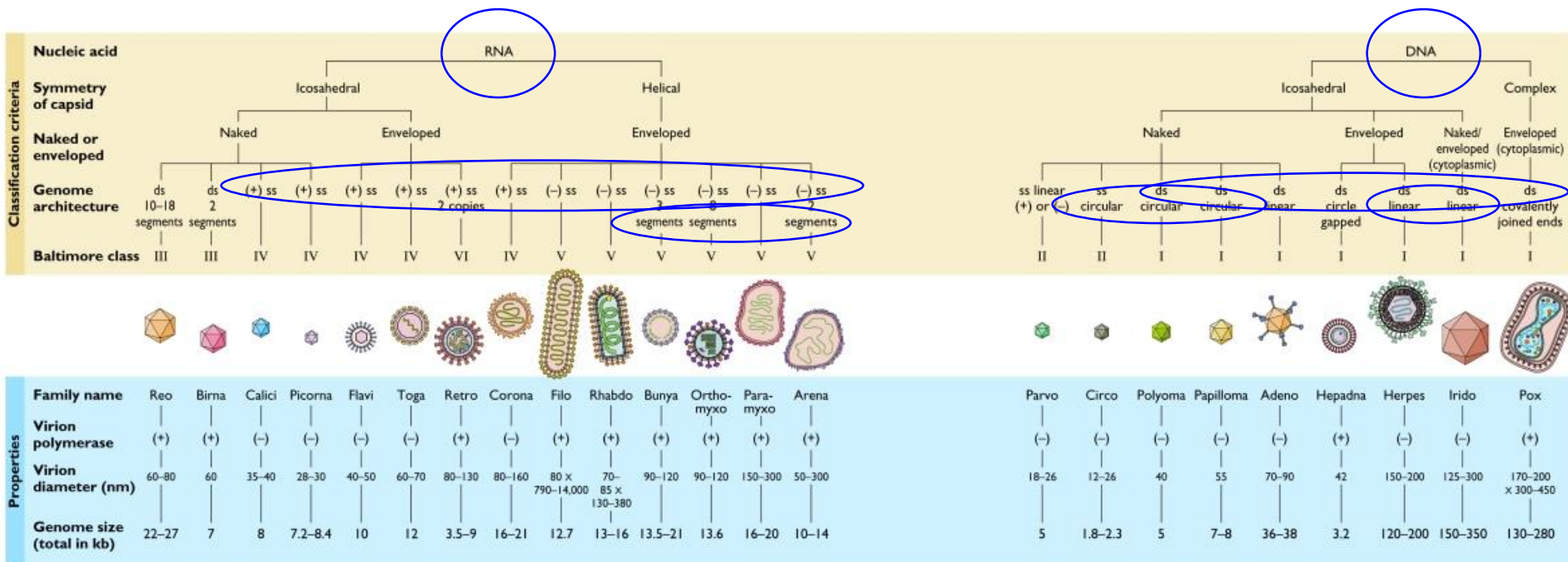
GenomeDenmark Cancer&Pathogens

Comprehensive high-throughput sequencing based virome characterization of cancers

Cover full diversity of viruses



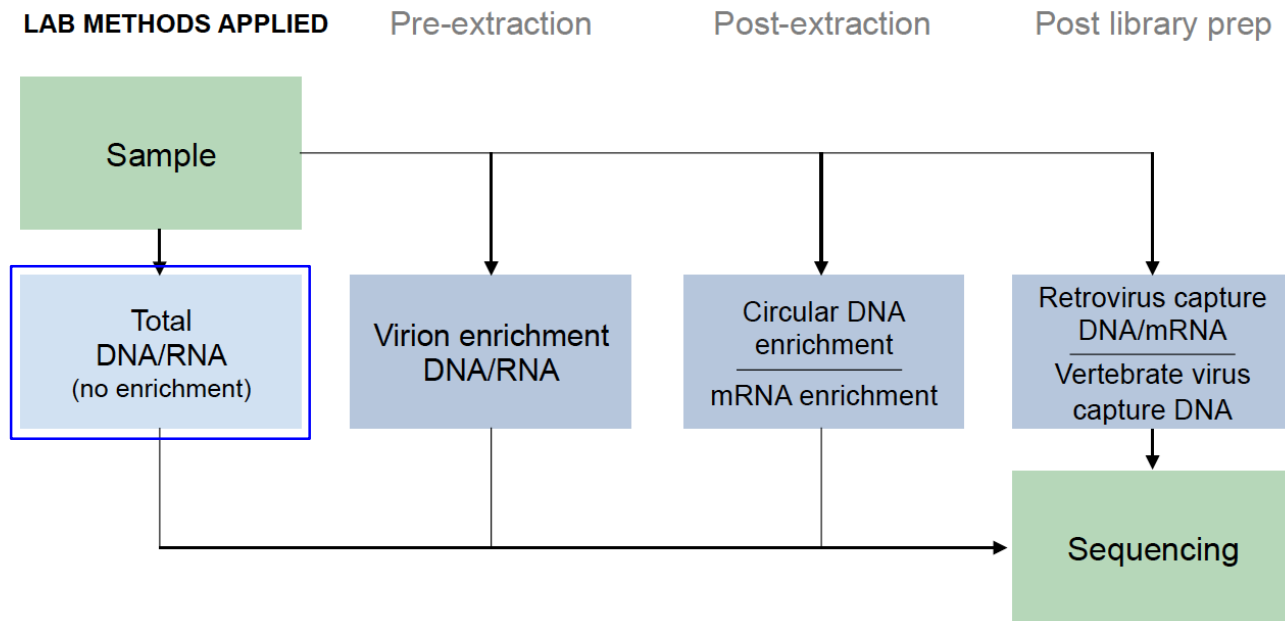
Virus classification



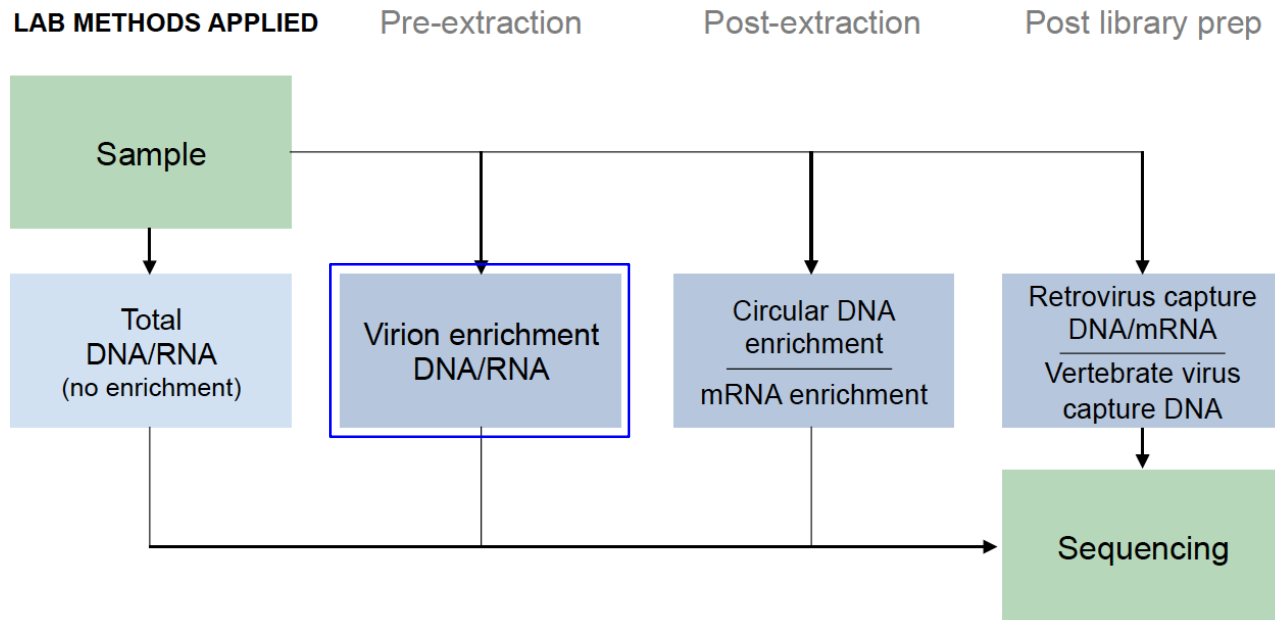
Adapted from M. H. V. van Regenmortel et al. (ed.), *Virus Taxonomy: Classification and Nomenclature of Viruses. Seventh Report of the International Committee on Taxonomy of Viruses* (Academic Press, Inc., San Diego, Calif., 2000).

- High sequence diversity
- No conserved regions
- Different cellular localization

Total DNA/RNA sequencing



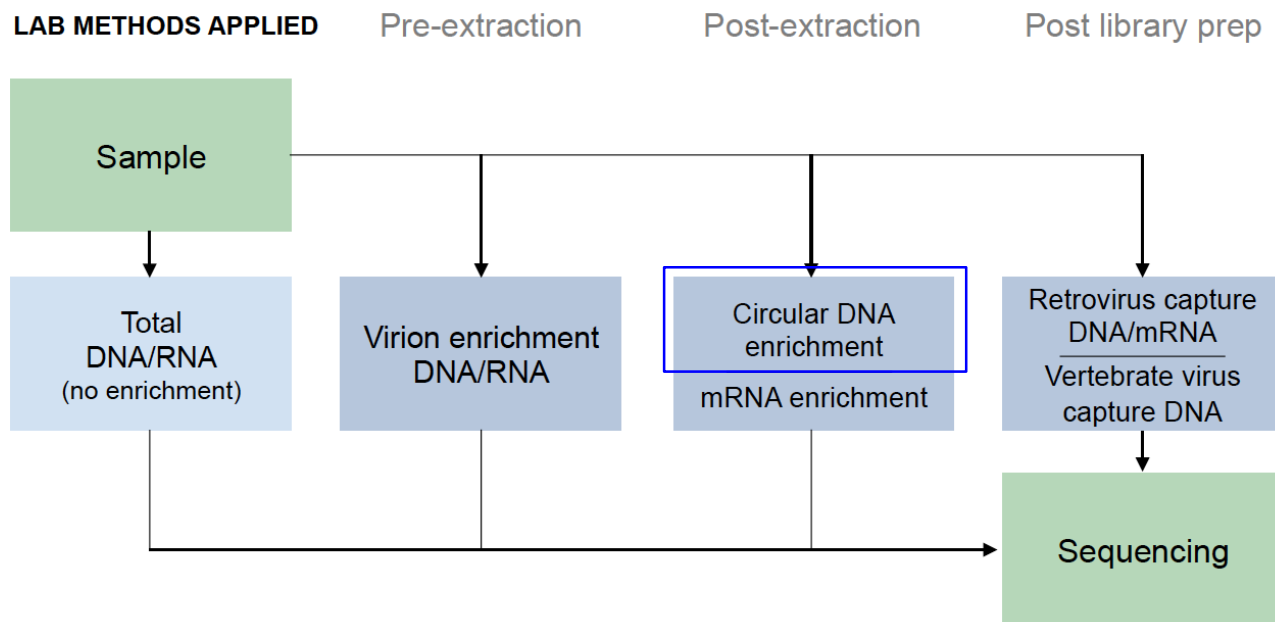
Viral particle purification



Target:

All viruses present as virions

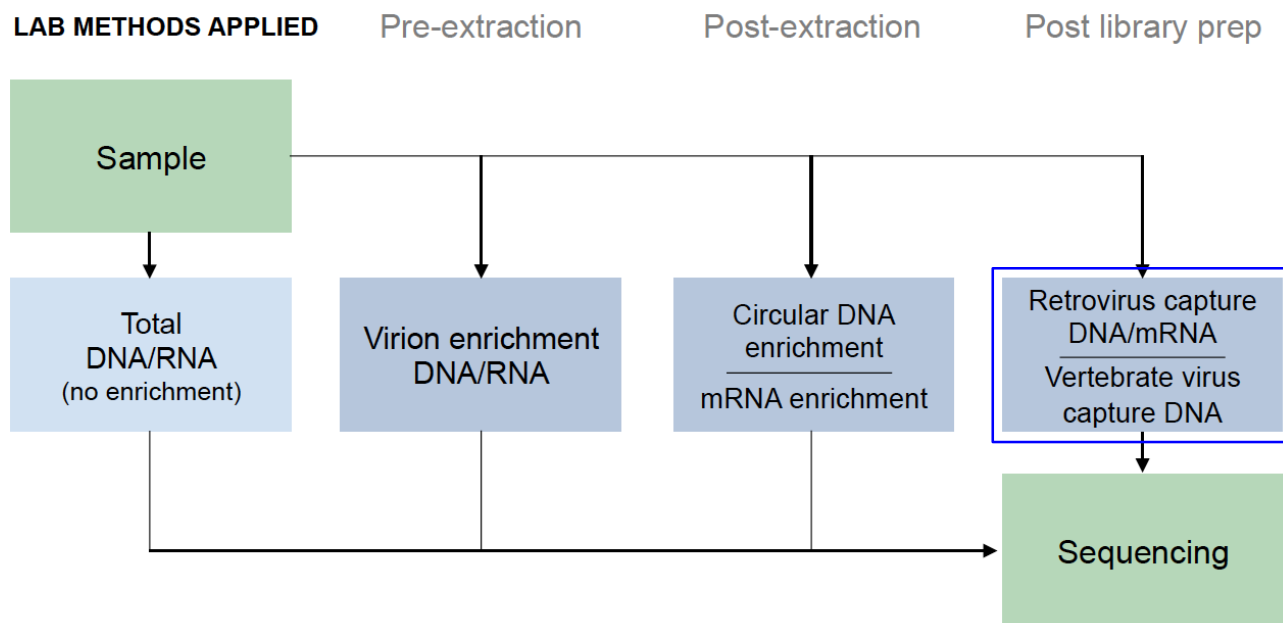
Circular DNA enrichment



Target:

Viruses with circular genomes

Viral sequence capture



Target:

All viruses including episomes
and integrated genomes

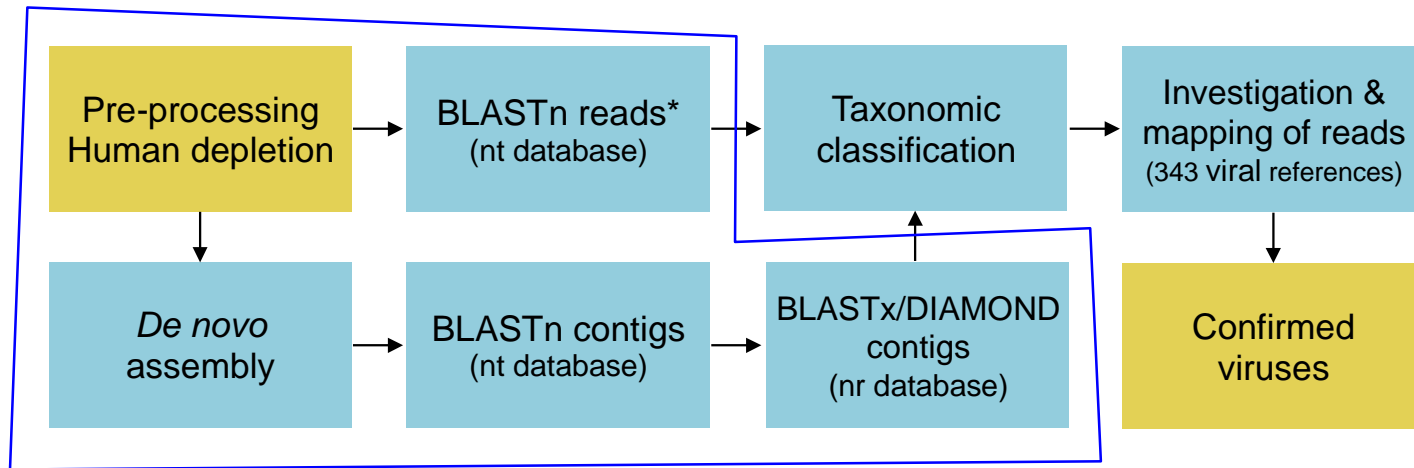
Samples and datasets

Sample type	Samples (n)	Total DNA	Total RNA	Virion enrichment		Circular DNA enrich.	Capture			mRNA	Datasets (n)
				DNA	RNA		Retro-virus DNA	Retro-virus mRNA	Vert. virus		
Basal cell carcinoma (cutaneous)	11	11		11	11	4	6		11		54
Mycosis fungoides (cutaneous)	11	11		11	11	10	10		11		64
Melanoma (cutaneous)	10	10		10	10	8			10		48
Oral cancer	10	9		10	10	10			10		49
Oral healthy	1					1			1		2
Vulvar cancer	3			3	3	3			3		12
Bladder cancer	7			7	7	5			7		26
Bladder cancer urine	10		2			10			4		16
Colon cancer	16	12	11	3	3		6			6	41
Colon healthy	2									2	2
Breast cancer	20	20	19	17	20	15					91
Testicular cancer	20	5		20	20						45
AML	9		6	9	9	7					31
B-CLL	9		8	9	9	8	9		8		51
BCP-ALL	8			8	8	8					24
CML	10		10	10	10	10			10		50
T-ALL	11		9	11	11	9					40
DLBCL*	5	5						3		3	11
Lymphoblastic lymphoma*	1	1						1		1	3
Multiple myeloma*	6	6						2		2	10
Colon cancer blood	8	8									8
Colon cancer ascites	1	1					1				2
Breast cancer ascites	1	1	1	1	1	1					5
Ovarian cancer ascites	5	5	4	3	3	5					20
Pancreatic cancer ascites	2	2	2				1				5
NTC				19	18	5	1		7		50
Total (without NTC)	197	107	72	143	146	114	33	6	75	14	710

* Cell lines. NTC: non-template control.

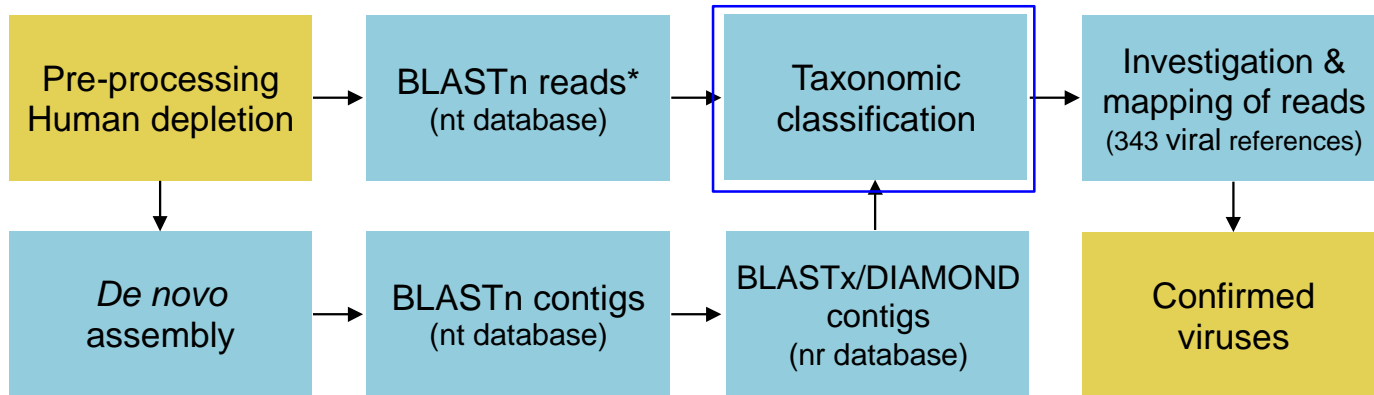
Sequenced on Illumina HiSeq 2000
Total of >57 billion read pairs

Data analysis



- Unbiased approach with BLAST of contigs and reads

Data analysis



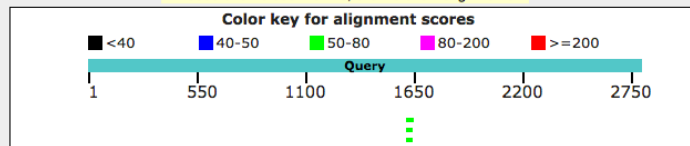
- “Real” viral sequences
- False positives/artefacts
- Contamination
- Bleedover/incorrect index assignment

Data analysis

Graphic Summary

Distribution of the top 34 Blast Hits on 3 subject sequences

Mouse over to see the title, click to show alignments



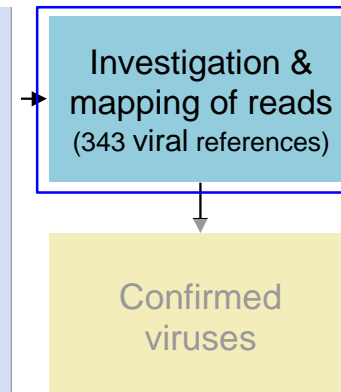
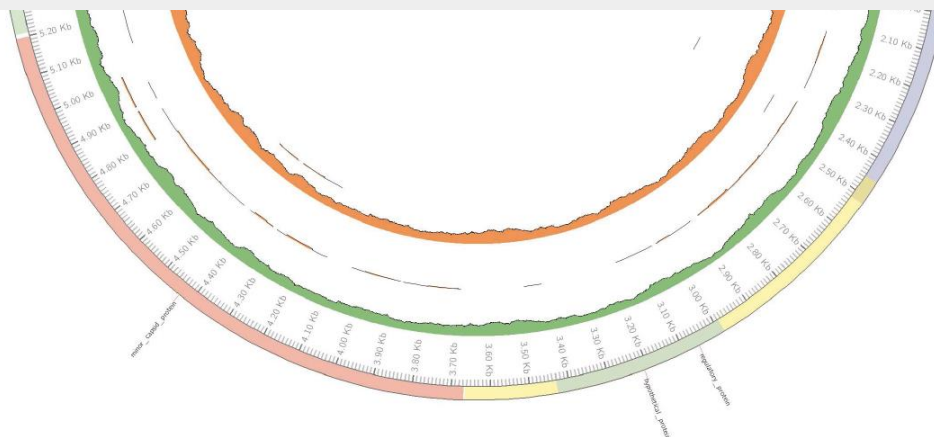
Descriptions

Sequences producing significant alignments:

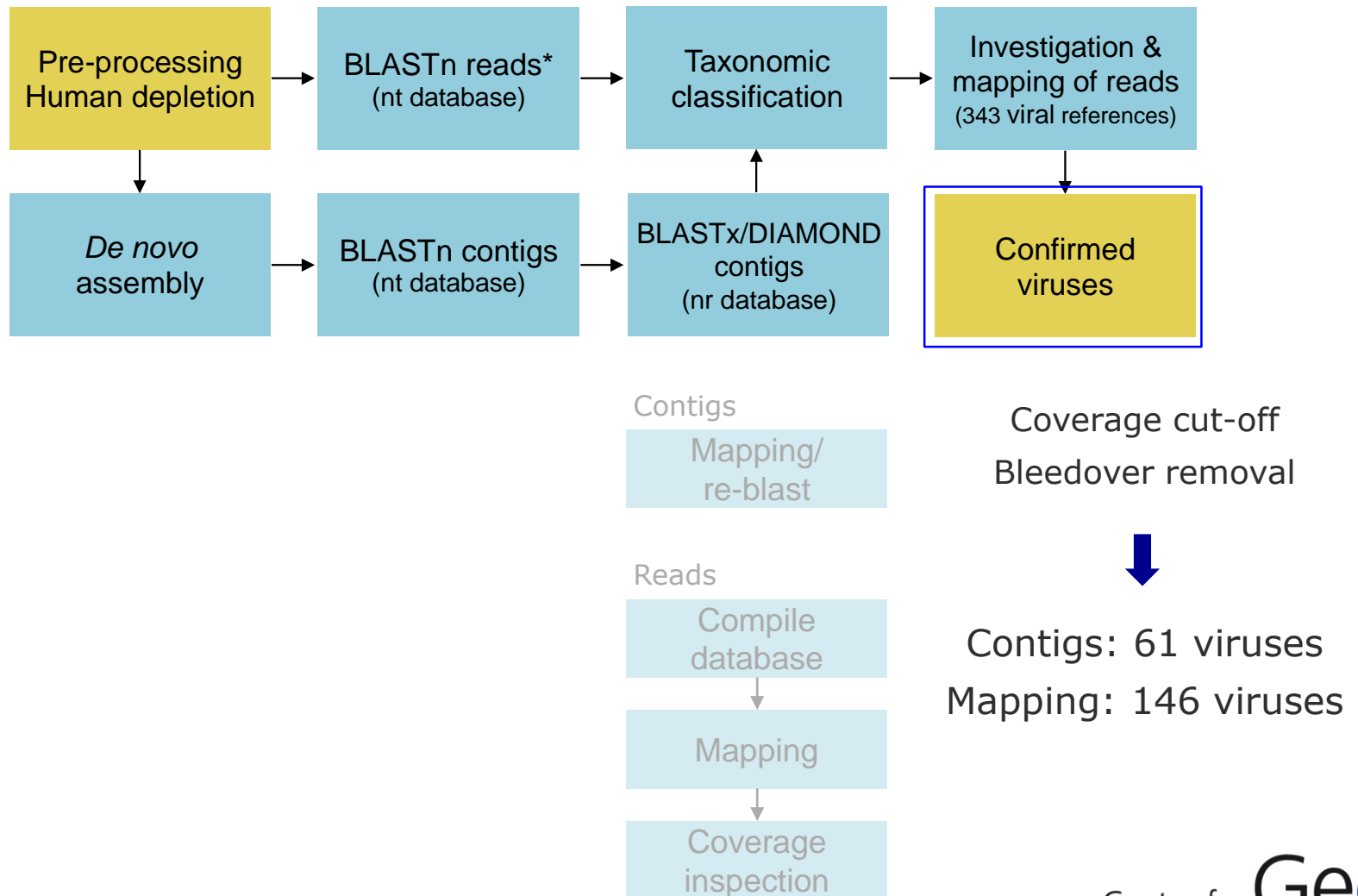
Select: [All](#) [None](#) Selected:0

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> Human herpesvirus 7 strain RK, complete genome	65.8	2056	1%	9e-06	100%	AF037218.1
<input type="checkbox"/> Human betaherpesvirus 6B strain HGDP00092, partial genome	63.9	63.9	1%	3e-05	100%	KY316037.1
<input type="checkbox"/> Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold109	62.1	62.1	1%	1e-04	100%	LK064703.1

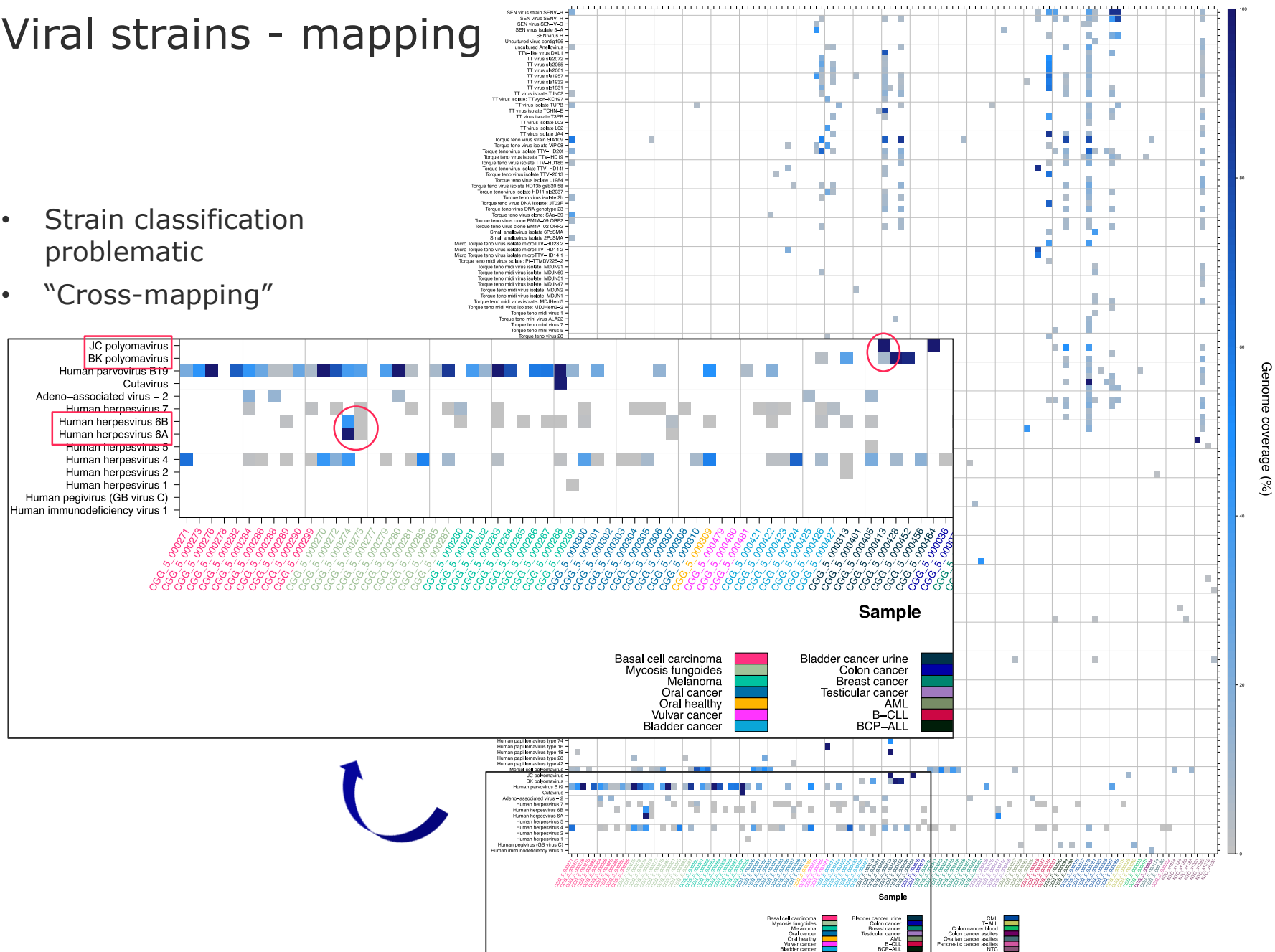


Data analysis

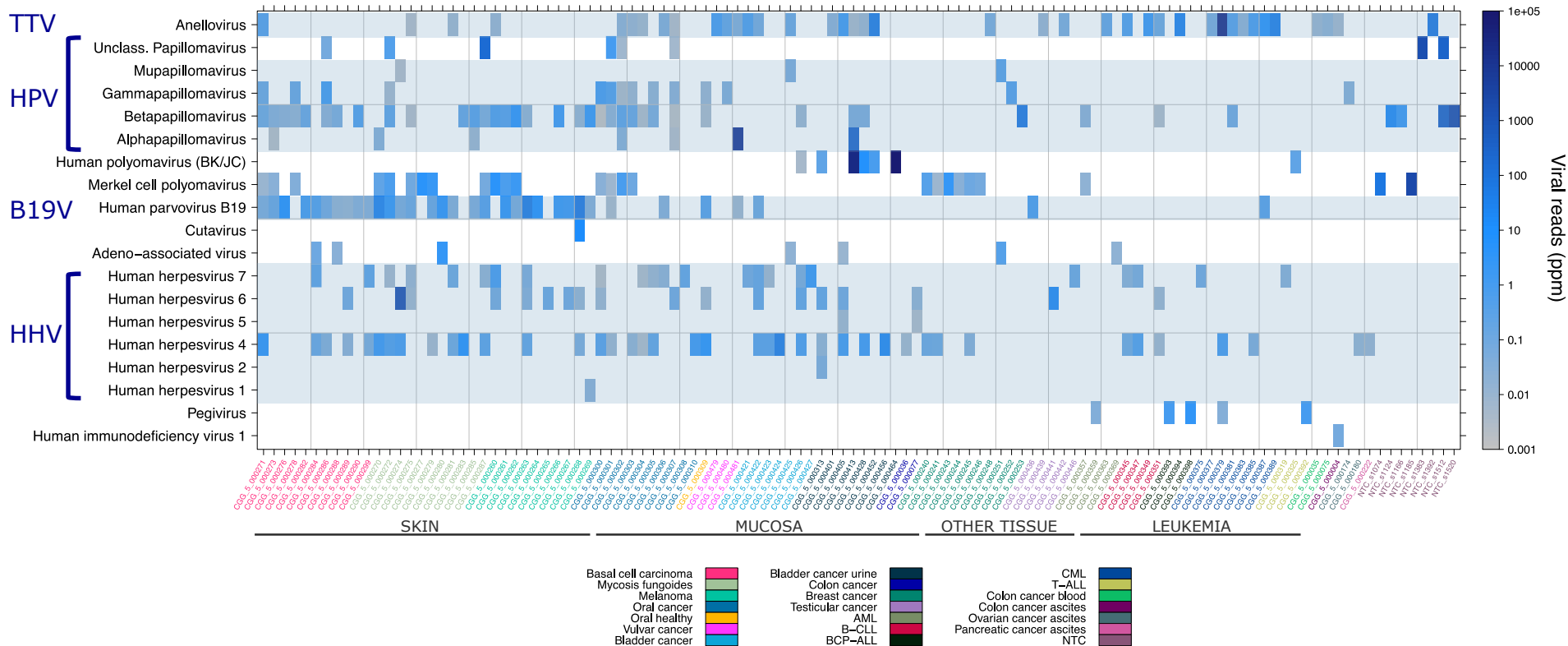


Viral strains - mapping

- Strain classification problematic
- “Cross-mapping”

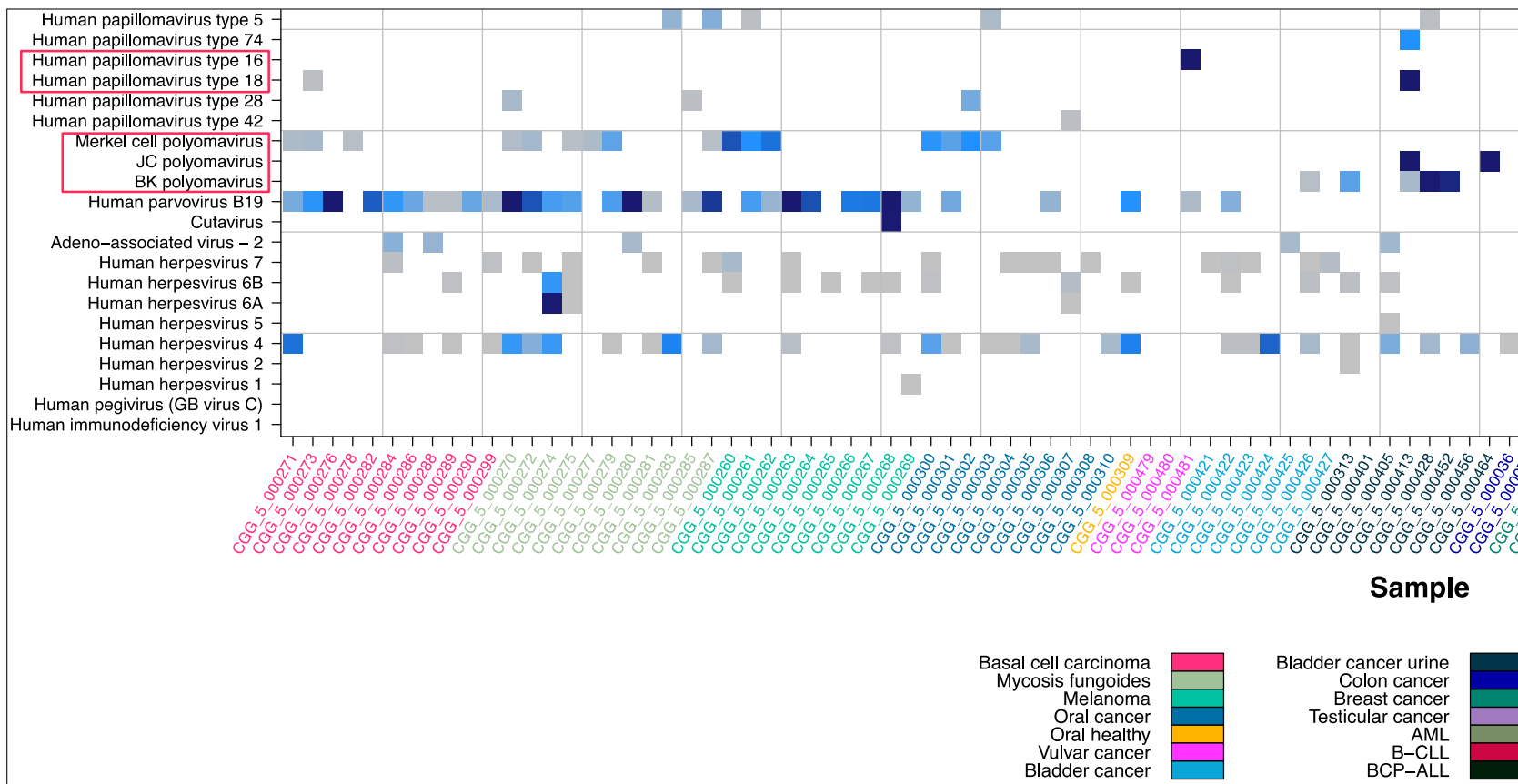


Viral groups - mapping



- DNA viruses commonly found in humans
- No apparent associations
- Half of samples virus-positives
- Skin/mucosal all samples virus-positive

Cancer associated viruses - mapping



- HPV 16 and 18
- Polyomaviruses

Summary and conclusions

- Comprehensive virome characterization covering major viral groups
- False positives, bleedover, and cross-mapping needs to be considered and accounted for
- Half of samples contained viral sequences
- Majority of the viruses commonly found in humans
- No apparent links between specific viruses and cancers
- Low frequency associations cannot be ruled out
- Methods comparison – come by poster no. 34



Thank you for your attention

Acknowledgements

Centre for GeoGenetics

Anders Johannes Hansen
 Kristín Rós Kjartansdóttir
 Maria Asplund
 Lasse Vinner
 Helena Fridholm
 Tobias Mourier
 Thomas Arn Hansen
 Christopher J. Barnes
 Randi Holm Jensen
 Eske Willerslev
 Stine Raith Richter
 Ida Broman Nielsen
 Carlotta Pietroni
 David E. Alquezar Planas
 Alba Rey de la Iglesia
 Pernille S. Olsen

Statens Serum Institut

Lars Peter Nielsen

Technical University of Denmark

Jens Friis-Nielsen
 Jose M. G. Izarzugaza
 José A. R. Herrera
 Ole Lund
 Thomas Sicheritz-Ponten
 Søren Brunak

All sample collectors

