Berkeley School of Public Health

FORUM for Collaborative RESEARCH

HBV

Next Generation Sequencing, data analysis and reporting

DDL DIAGNOSTIC LABORATORY

Partner in advanced diagnostic testing

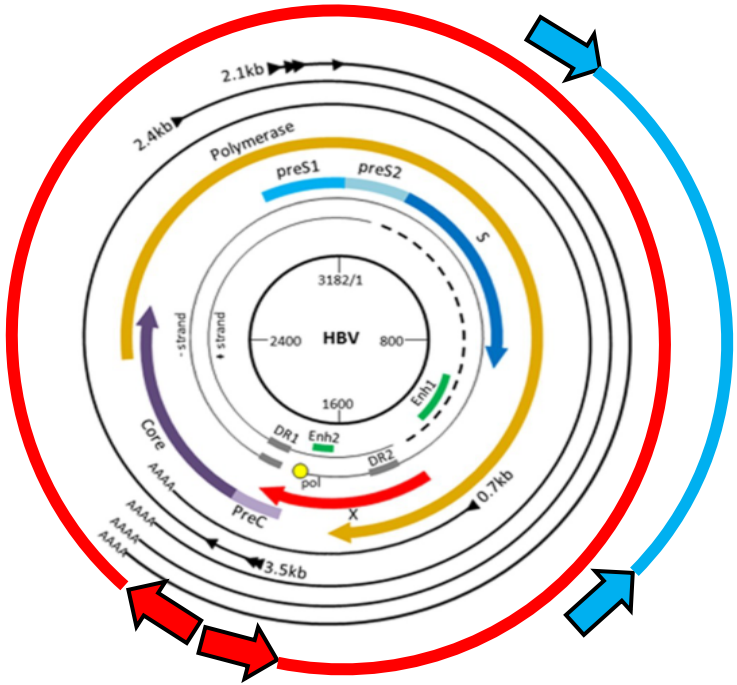
Presenter
Leen-Jan van Doorn

HBV Forum 3
October 24th, 2017
Marriott Marquis, Washington DC

www.forumresearch.org

HBV Biomarkers

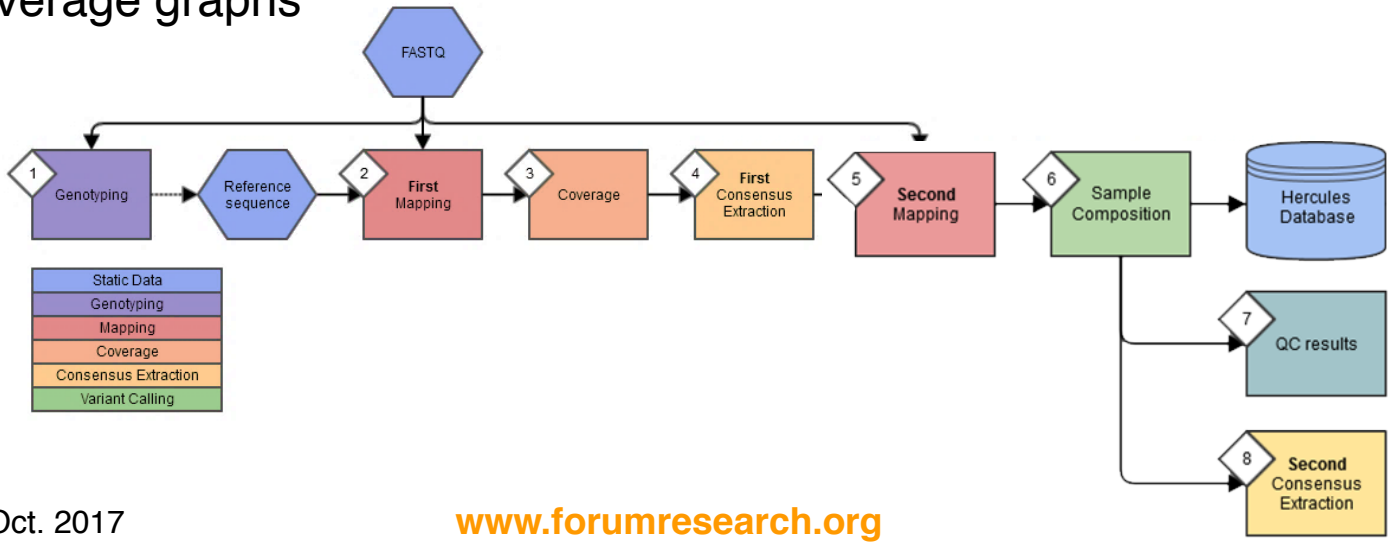
- HBV biomarkers:
  - HBV DNA viral load
  - HBsAg, HBeAg, Anti-HBs, etc
  - HBcrAg
  - HBV RNA quantitative
  - HBV DNA/RNA sequencing
- Aim of HBV DNA/RNA sequencing:
  - Determine HBV geno/subtype
  - Assess RAVs
- HBV-DNA or HBV-RNA amplification:
  - Full length or partial



1. Mapping and reporting: reference sequences

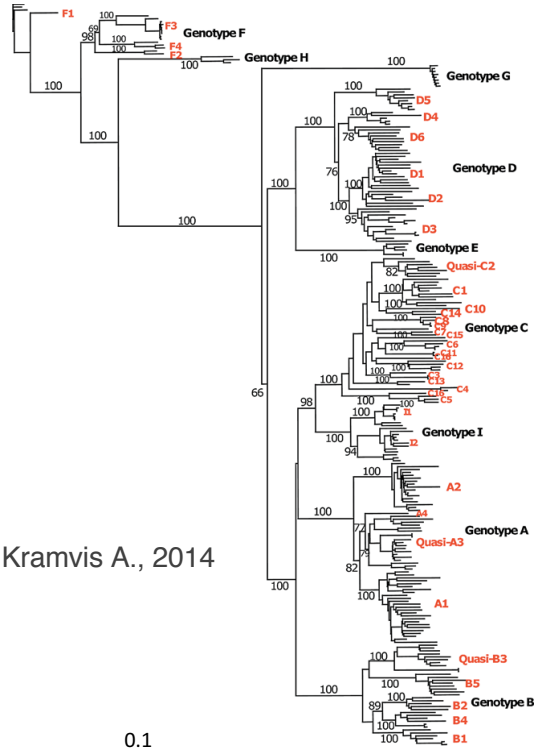
HBV NGS data analysis + reporting

- DDL's *Athena* pipeline for Quality Based Variant Detection
- Input: FASTQ file containing sequencing reads
- Output: reference-based variant table, consensus sequence and coverage graphs



Mapping of NGS sequence reads

HBV genotypes/subtypes



Kramvis A., 2014

- Which reference sequences to use?
- HBV genotype divergence >7.5% across the complete genome. Nine genotypes described A-I
- Genotypes A, B, C, D, F, H, and I are classified into at least 35 subtypes. HBV subtypes differ 4-8%
- Different HBV genome length between genotypes
  - **A**: 3221bp, **B, C, F, H, I**: 3215bp, **D**: 3182bp, **E**: 3212bp, **G**: 3248bp.
- Deletion variants
- HBV recombinants

Reference-based variant calling

```
Refer 1:    AGTTCCATGCCAATG      Refer 2:    ACATGCATGCTAATC
Sample:     AGATCCATCCCAATC      Sample:     AGATCCATCCCAATC
Variants:     A     C     C      Variants:    G   C    C C
```

The same sequence data yield different variant lists if compared to different references

Mapping vs. reporting

Accurate mapping of reads to (sub)type reference → Variant calling + reporting to 'universal' reference

HCV	Genotypes: 1 - 8 Subtypes: > 80	→ 1a = H77, 'universal' reference
HBV	Genotypes: A - I Subtypes: ?	→ ?? = 'universal' reference

HBV References

Reference source	Accession ID	Genotype	Subtype**	Geographic Origin**
ICTV	X02763	A	A2	USA
	D00330	B	B2	Japan
	AY123041	C	?	?
	V01460/NC_003977*	D	D3	France
	X75657	E	E	France
	X69798	F	F2	Brazil
	AF160501	G	G	USA
	AY090454	H	H	Nicaragua
HBV Database	X02763	A	A2	USA
	EU054331	A	?	?
	AB219428	B	B3	Philippines
	D00331	B	B3	Indonesia
	GQ924620	C	?	?
	GQ358158	C	?	?
	AF121240	D	?	?
	FJ904433	D	D6	Tunisia
	AB106564	E	?	?
	AY738145	E	?	?
	AY090458	F	F1	Costa Rica
	X756558	F	F2	France
	AF160501	G	G	USA
	AB064313	G	G	USA
	AY090454	H	H	Nicaragua
	FJ356716	H	?	?

*GT D reference from ICTV has been curated in NCBI Refseq. **Subtype and origin Information from Norder et al., 2004 and Kramvis, 2014

HBV References

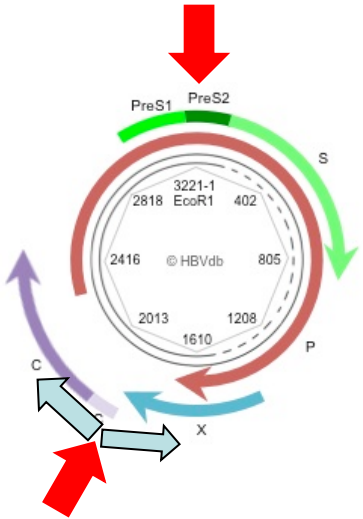
Genotype	ICTV	Cai et al., 2016*	HBVdb	Selection
A	X02763	X02763	X02763 EU054331	**X02763**
B	D00330	AB073858	AB219428 D00331	?
C	AY123041	AB048704	GQ924620 GQ358158	?
D	V01460 NC_003977	X02496	AF121240 FJ904433	?
E	X75657	X75657	AB106564 AY738145	?
F	X69798	X69798	AY090458 X756558	?
G	AF160501	AF160501	AF160501 AB064313	?
H	AY090454	AY090454	AY090454 FJ356716	?

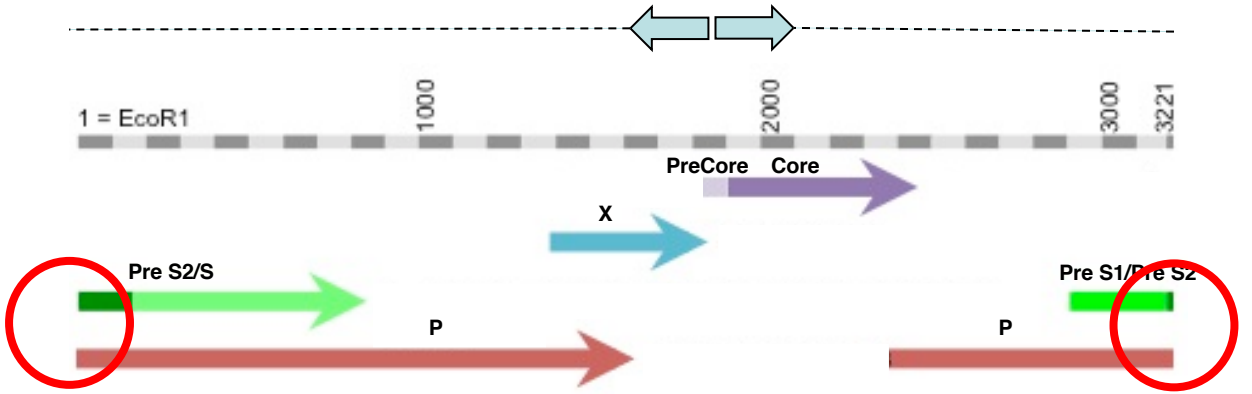
* Cai at al., Virus Genes 2016

2. Annotation

HBV Reference Annotation

- Which starting point to use for absolute numbering of genes (circular genome)?
- *Eco*RI (TTC) mostly used as starting point; located in the PreS2/Polymerase



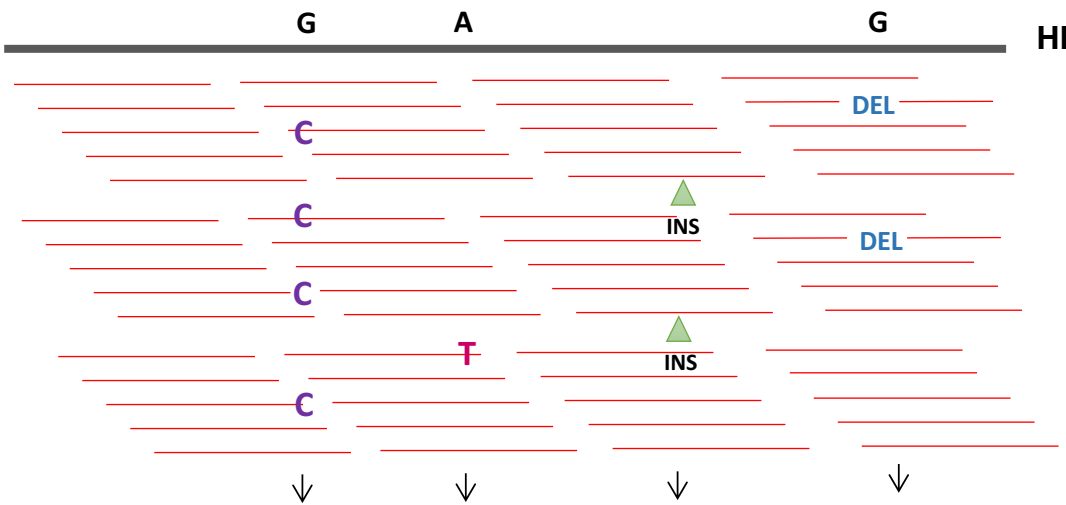


Gene	Pre S1	Pre S2	S	P	X	Precore	Core
Absolute numbering NT	2854-3210	3211-3221 **1-154**	155-835	2307-3221 **1-1623**	1374-1835	1814-1900	1901-2455
Relative numbering AA	1-119	1-55	1-227	1-846	1-155	1-29	1-186

https://hbvdb.ibcp.fr/HBVdb/HBVdbGenome

3. Complete sample composition

Athena: Sample composition



Mapping

Annotation based on reference used

Nucleotides

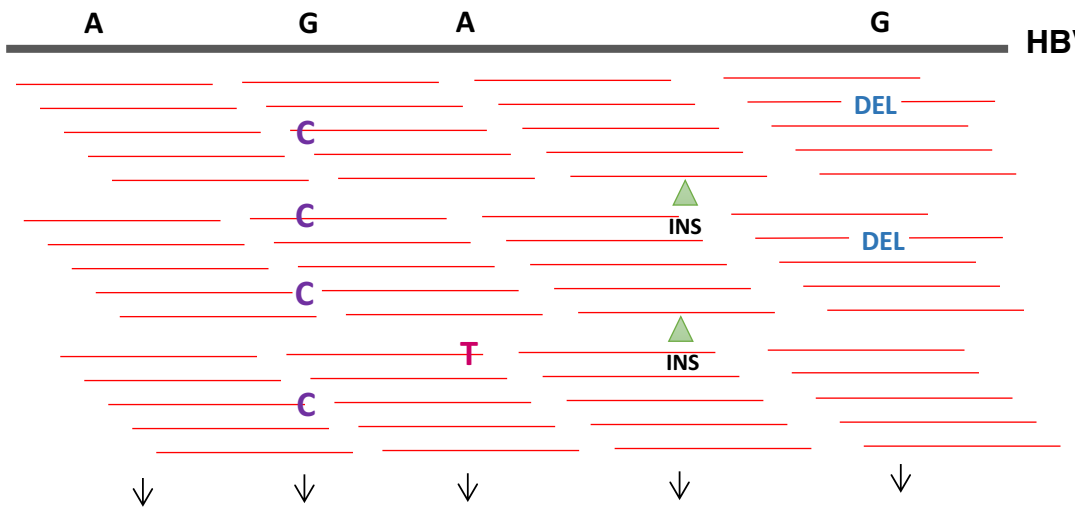
G100C 25% | A238T 0.5% | 438.1Y 20% | G303- 20%

Variants

Amino acids

P 25% | V 0.5% | INS 20% | DEL 20%

Athena: Sample composition



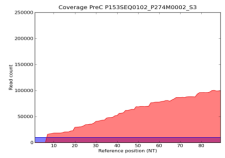
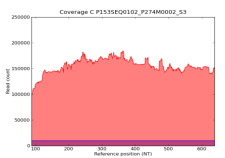
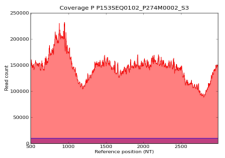
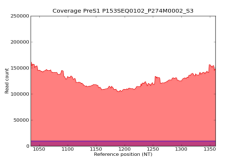
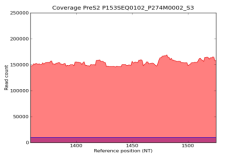
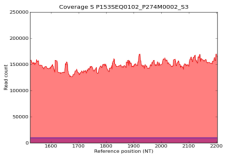
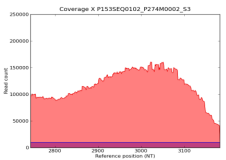
Mapping

Annotation based on reference used

Total sample composition

Nucleotides	10A 100%	100G 75% 100C 25%	238A 99.5% 238T 0.5%	438.1Y 20%	303G 80% 303- 20%
Amino acids	H 100%	R 75% P 25%	D 99.5% V 0.5%	H 80% INS 20%	H 80% DEL 20%

HBV NGS data analysis

HBV PreCore	HBV Core	HBV Polymerase	HBV PreS1	HBV PreS2	HBV S-gene	HBV X-gene
						

- All data is stored in database so multiple options available for data reporting
- Reporting output: per gene, coverage graphs, consensus sequences (different cut-offs), variant tables (nucleotide, codon and amino acid level)
- Comparison against subtype, genotype and universal reference is possible
- This allows comparisons within and between patient groups

Summary

- NGS Sequencing from HBV DNA and RNA:
  - Set of genotype references for mapping and reporting
  - Use Genotype A (X02763) as ‘universal reference’
  - Need for improved subtyping system
  - Use of *Eco*RI (TTC) as annotation start site
  - Complete ‘sample composition’ instead of only listing reference-based variants, to allow multiple queries and flexible reporting

Acknowledgements

- DDL's sequencing and bio-informatics teams