

A new bioinformatic pipeline for the analysis of hepatitis B virus transcriptome by Nanopore sequencing coupled to 5'RACE

Xavier Grand^{1,2}, Doohyun Kim², Delphine Bousquet², Guillaume Giraud², Cyril Bourgeois¹,

Fabien Zoulim² and Barbara Testoni²

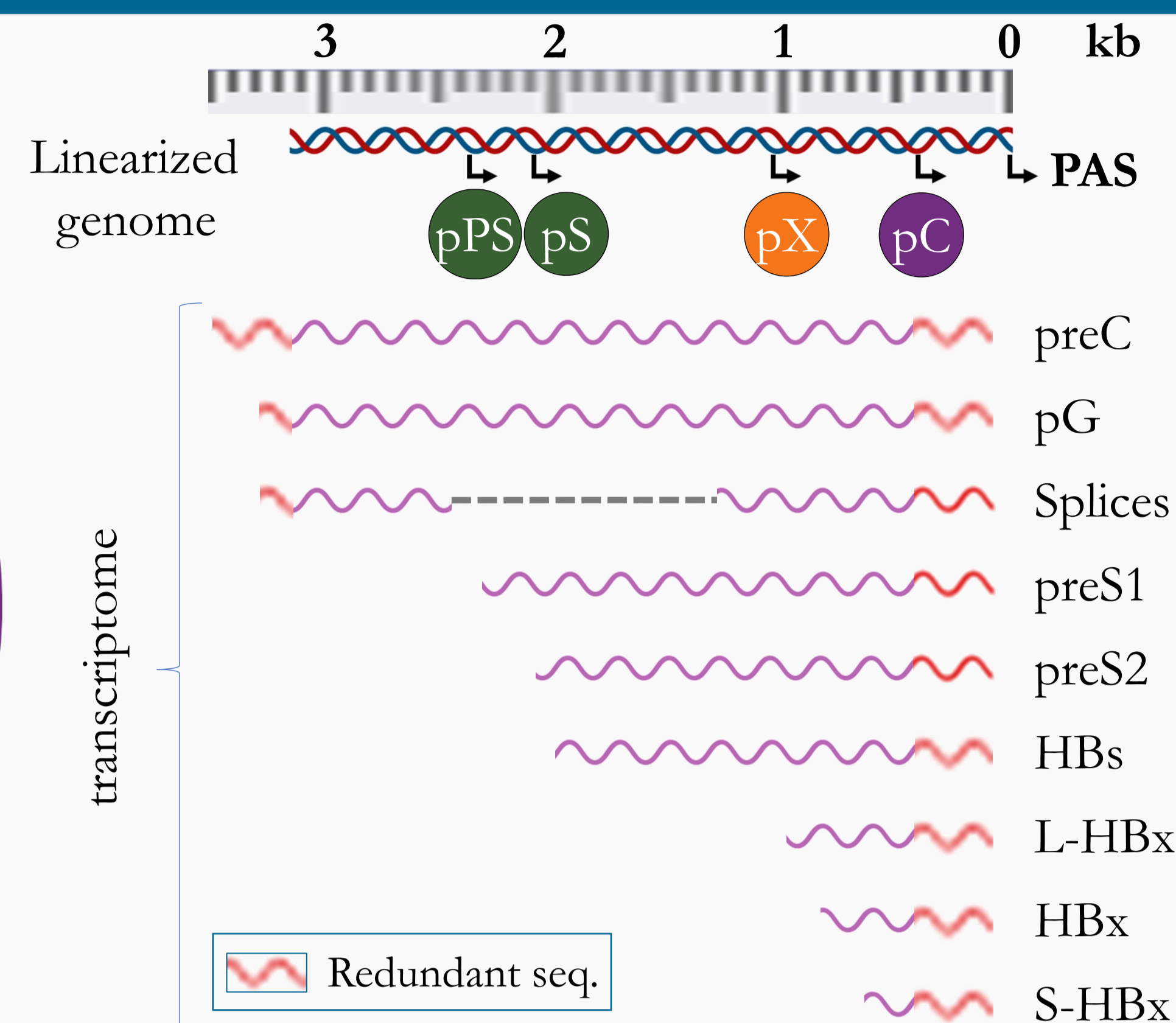
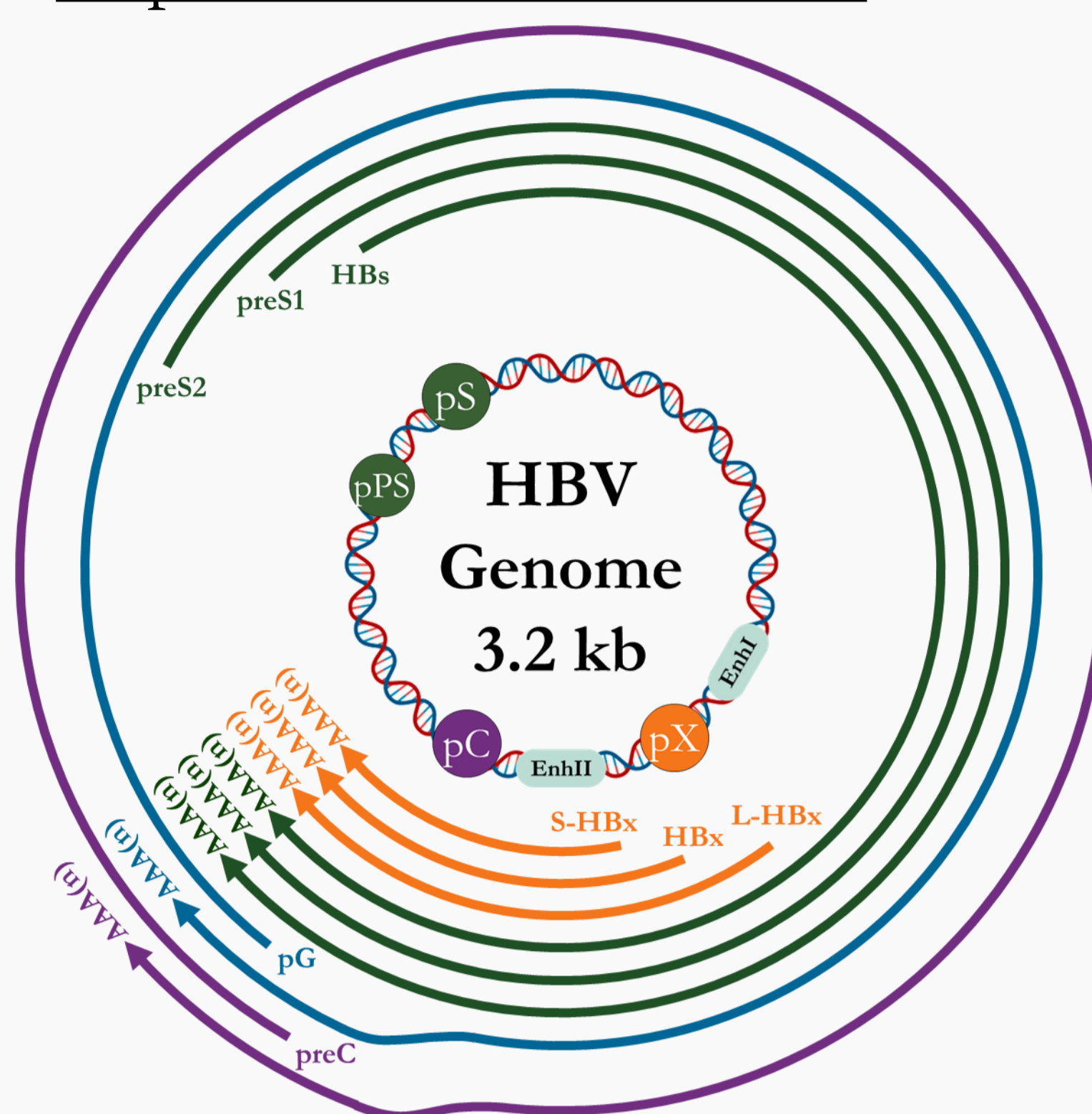


Problem & Biological model

How to evaluate the expression of HBV RNAs ?

- Overlapping ORFs: 8 major viral RNAs on the minus strand and 22 splice variants. Massively **redundant RNA sequences**.
- Longest RNAs are longer than the genome itself. Redundant, **strictly identical sequences** at the beginning and the end of RNAs, Induces issues to select a genomic reference sequence.
- Highly similar splice variant sequences: Only distinguishable by **splice-junctions**, Non-canonical splice-junction donor and acceptor motifs.

Hepatitis B Virus Genome:



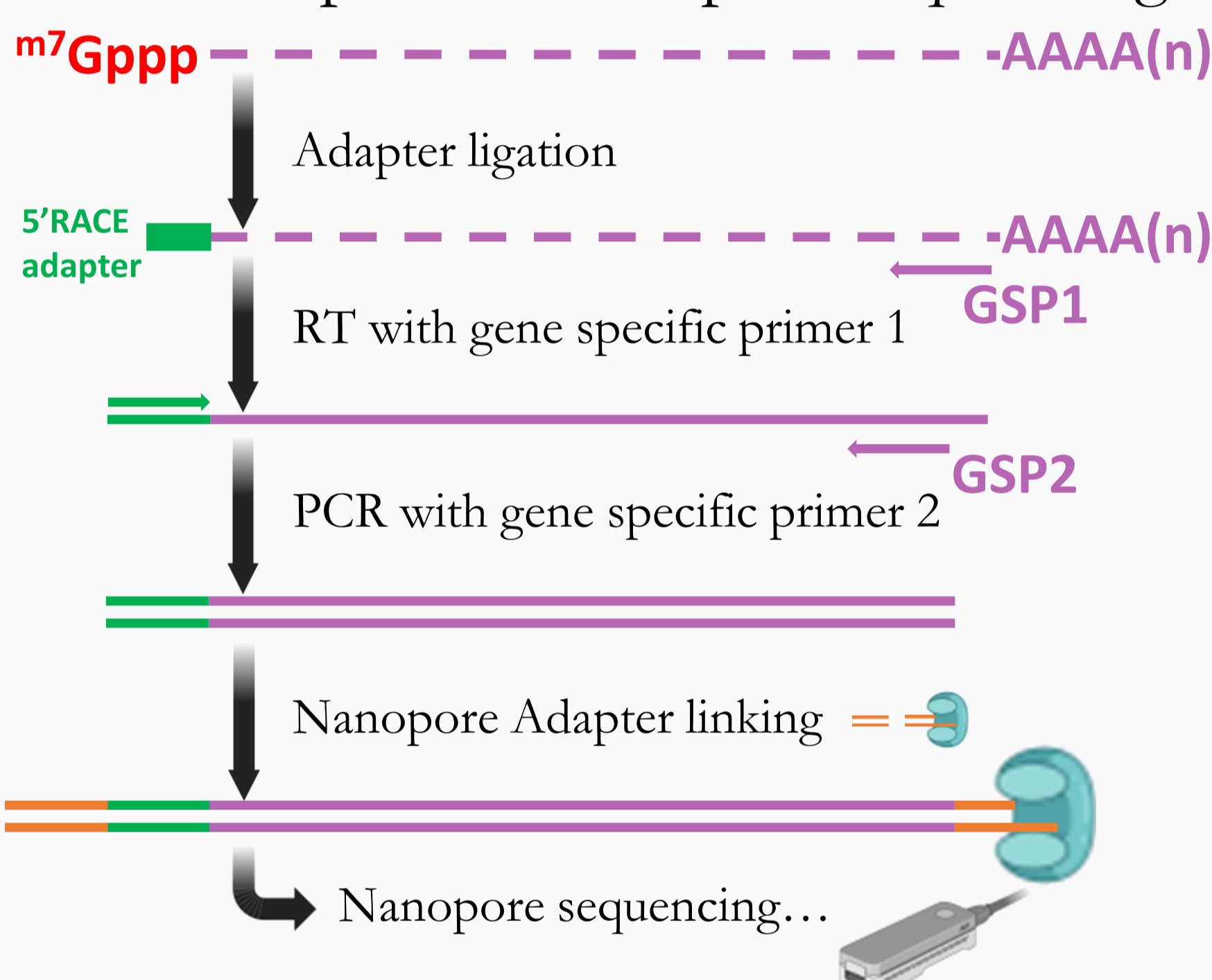
5'RACE coupled to Nanopore sequencing

5'RACE to isolate full-length HBV RNAs:

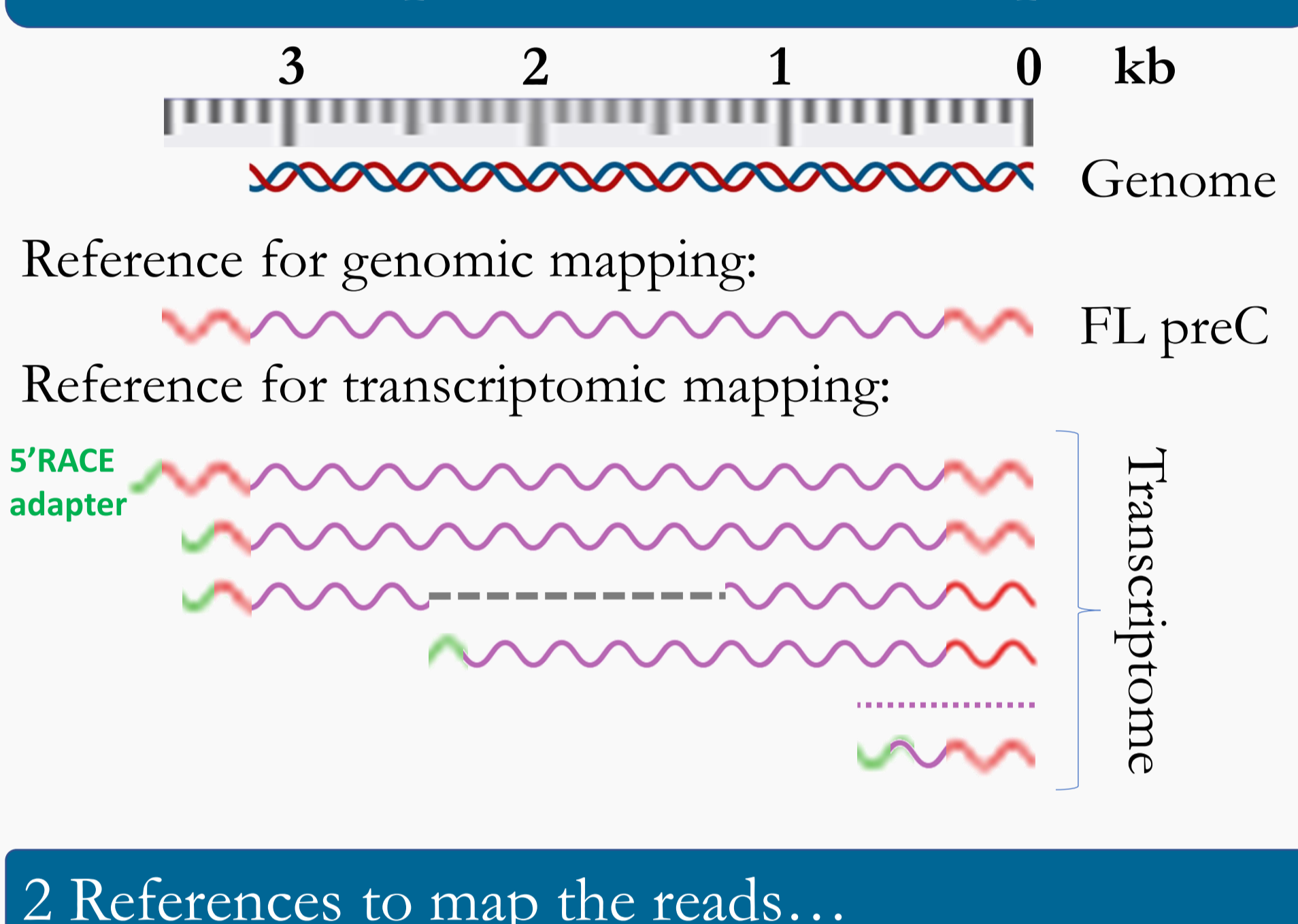
- Total RNAs** are extracted from infected hepatocytes or patient sera.
- Capped mature RNAs** are linked with a 5'RACE Adapter.
- HBV RNAs** are retro-transcribed and amplified with 2 **gene specific primers** (GSP 1&2).
- Finally, **Nanopore Adapters** are linked to process the sequencing.

Consequently, quasi **full-length HBV RNAs** isolated and amplified can be sequenced.

5'RACE to isolate and amplified HBV RNAs coupled to Nanopore sequencing:

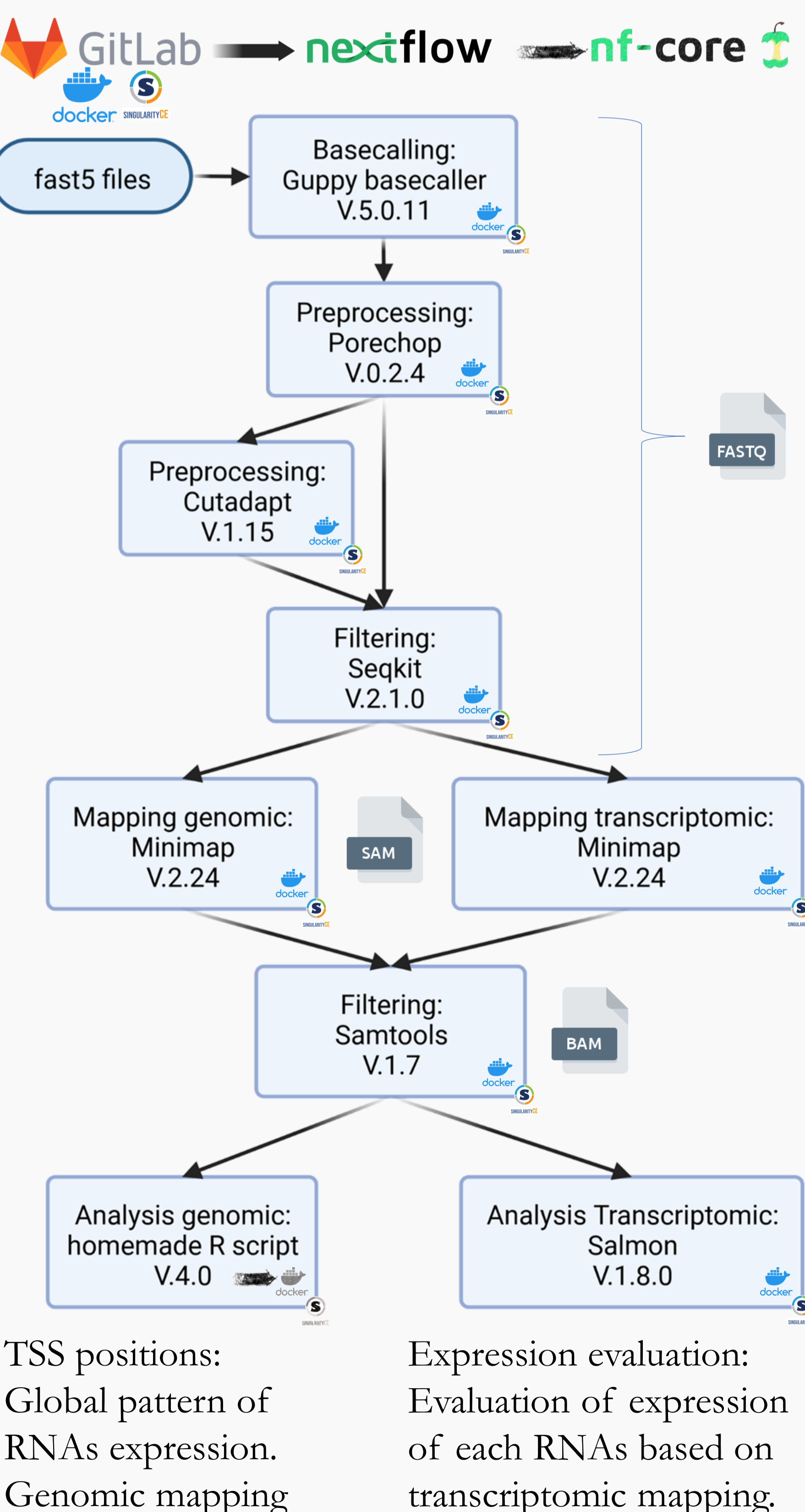


Full-length preC sequence as genomic & 5'RACE adapters linked transcriptome.



References to map the reads

Analysis pipeline:



TSS positions: Global pattern of RNAs expression. Genomic mapping

Expression evaluation: Evaluation of expression of each RNAs based on transcriptomic mapping.

Simulation:

NanoSim

Datasets simulation & Salmon quantification:

- Datasets containing **only 1 RNA species target**.
- Datasets combining all transcript species (not shown).
- Evaluation of assignment for each transcript species (% TPM).

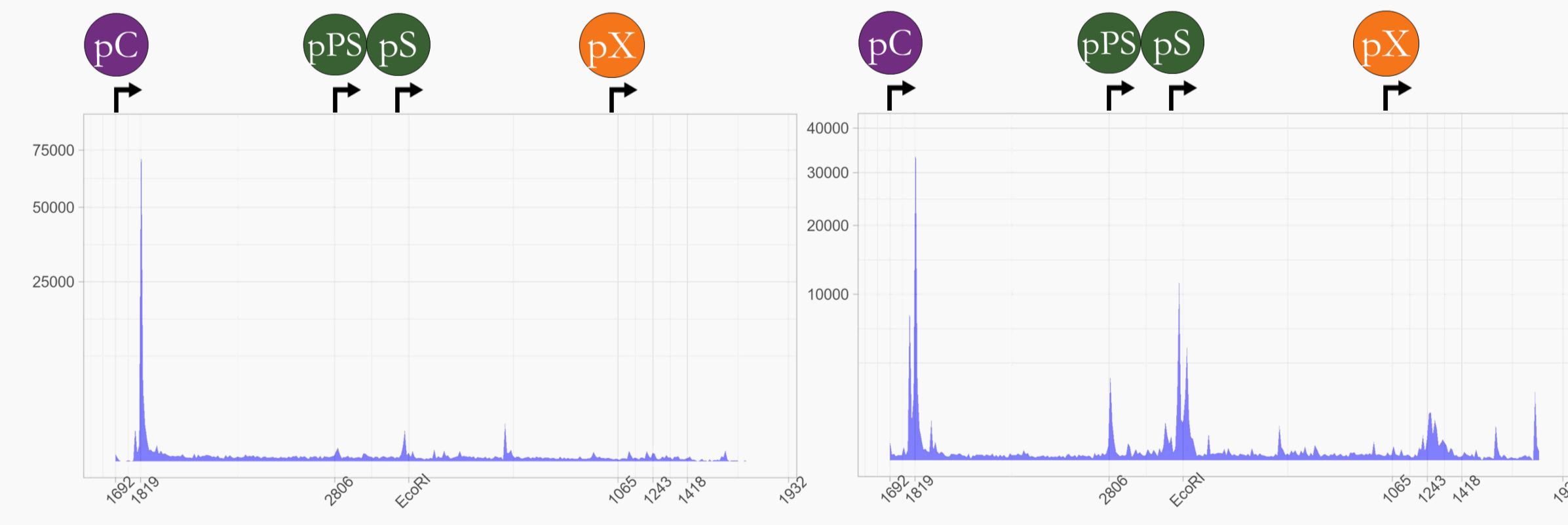
More than 90% of assignment for "canonical" transcript species. Poor assignment of Splice Variant species, e.g. SP09 or SP19.

transcript	preC	pC	preS1	preS2	HBs	L-HBx	HBx	S-HBx	SP01	SP02	SP03	SP04	SP05	SP06	SP07	SP08	SP09	SP10	SP11	SP12	SP13	SP14	SP15	SP16	SP17	SP18	SP19	SP20	SP21	SP22
preC	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pC	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
preS1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
preS2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HBs	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L-HBx	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HBx	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S-HBx	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP01	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP02	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP03	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP04	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP05	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP06	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SP09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
SP10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
SP11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
SP12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
SP13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
SP14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
SP15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
SP16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
SP17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
SP18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
SP19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
SP20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
SP21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
SP22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Results:

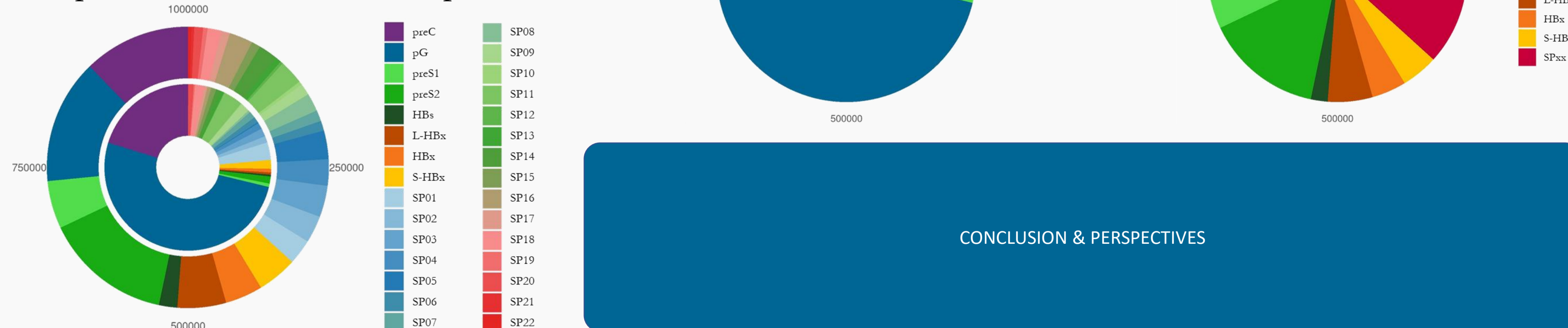
Graphs of **start-position** for reads mapped on preC RNA sequence with genomic coordinates:

- Give the use of the promoters,
- Global pattern of RNAs expression,
- No splice variants expression information.



Graphs of **expression evaluation** of HBV RNAs using Salmon software:

- Considering TPM values as expression levels in each sample.



Simulation and results