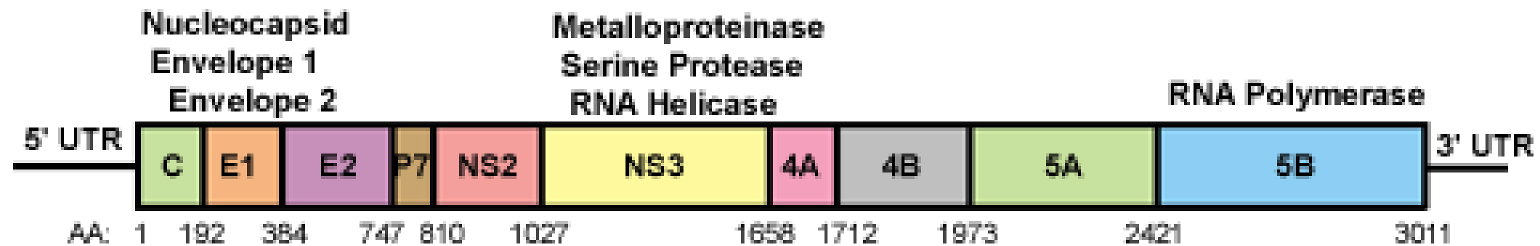
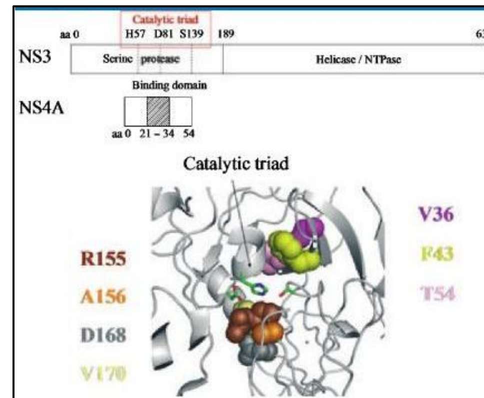
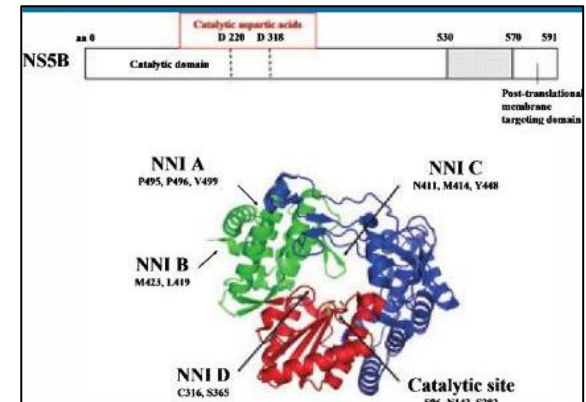
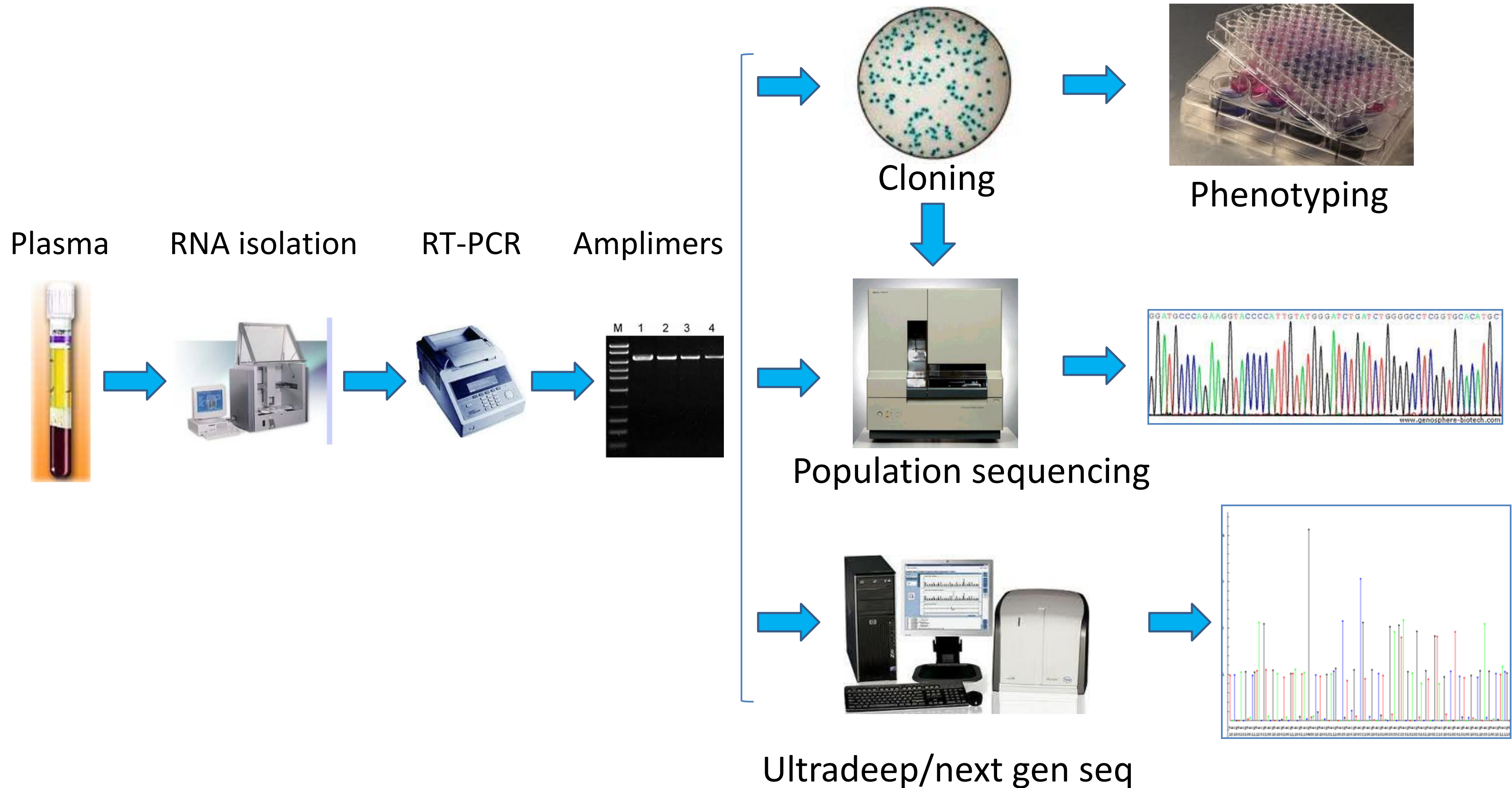
HCV ultradeep sequencing (UDS): technical issues



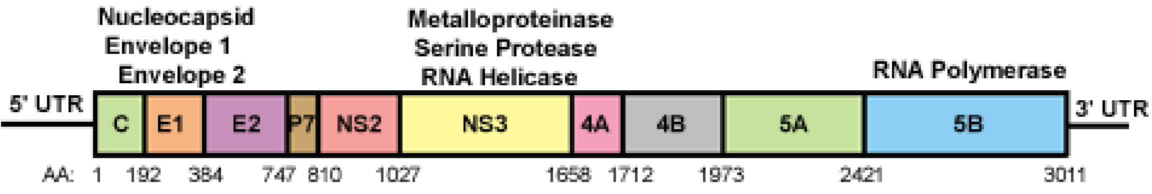


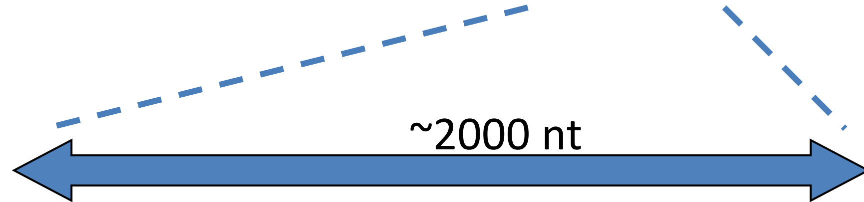


Workflow



HCV NS5B





Population sequencing

1 consensus sequence

Clonal sequencing

40-90 clonal sequences

Ultradeep sequencing

A. DNA shearing

B. Subfragment amplification

Statistical distribution

Potential sources for errors in ultradeep sequencing of amplimers (UDSA)

Primer design
- Primer selection → primer bias
- Primer dimers
- Non-specific binding

Amplification
- PCR bias
- Secondary structures
- Accuracy/Fidelity of the polymerase

Ultradeep sequencing of amplimers (UDSA):
- Chemistry
- Read-length
- Homopolymeric tracts
- Linkage

Data analysis (system)
e.g., Roche 454 software (AVA, Reference Mapper, De Novo Assembler)

Data analysis (external)
- Commercial software
- Freeware
- Home-brew tools

Provided by Ina Vandenbroucke

Sample preparation

- RNA isolation:
  - no specific technical requirements (high sensitivity/quality; same as for consensus sequencing)
  - Preferably: larger volume, elution in small volume
- Adequate sample ‘representation’:
  - IU/mL in sample → copies per PCR → sensitivity
  - Sampling bias: Representation of minority variants?

Dependence of Sensitivity of Detection of Minor Variants and Input Viral Load

- Assume that 200 µl plasma used for RNA extraction, 25% used for RT-PCR; RT successful for 20% of RNA molecules; minority variant present at 10% of total.

Viral load (copies/ml)	RNA copies in RT rxn	Amplifiable genomes in PCR	Copy no. (minor variant)
100,000	5,000	1,000	100
10,000	500	100	10
1,000	50	10	1

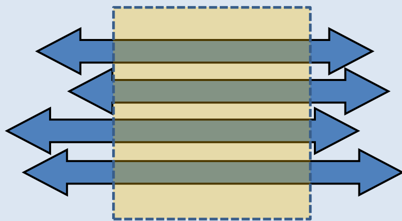
Provided by Neil Parkin

- Low viral load → high sampling bias, high PCR bias
- Threshold ~10,000 IU/ml to perform UDS?

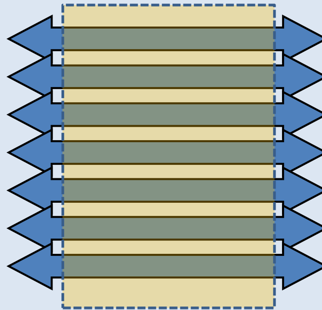
Amplification (Rev. transcription + PCR):

- High-fidelity enzymes RT and PCR step
  - (< 0.1% substitution error rate)
- Amplification bias
  - Robustness of PCR primer sets (genotypes/subtypes)
  - Optimized protocols (Mg^{2+}, temperature,…)

<u>Primer selection</u>:
Overlapping fragments, generated by different PCR primers



<u>PCR bias</u>:
Replicate RT-PCRs



<u>Sampling bias</u>:
- Increase sample input volume
- Replicate entire procedure (RNA isolation, RT, PCR)

Sequencing and data analysis

- Homopolymeric tracts
- Technical sensitivity (# of reads; forward and reverse; coverage per position)
- Analytical sensitivity (sample & viral load)
- Data analysis software (platform specific (454); Bowtie, CLCbio, home-brew).

Reporting issues

- frequency table listing all amino acid positions (also at nt level?), including technical sensitivity
- Assess impact on the protein level → key positions related to resistance
- % wildtype vs. mutant
- Include background sequencing error rate

Reporting to FDA

- No specific reporting format requirements.

Issues to cover (FDA-Division of Antiviral Products; DAVP):

- % of 'resistant' variant relative to wild-type
- absolute concentration of 'resistant' variant (i.e., extrapolating from viral load)
- specifically for long-term persistence analyses:
  - trends over time and modeling approaches to predict return of 'resistant' variant to pre-treatment background level
  - relationship, if any, between 'sensitive' sequence analysis results and population-based results (e.g., is there a relationship between duration while undetectable by population sequencing with time to undetectable by ultradeep sequencing?)
- some descriptive information on technical performance and limitations of the assay (e.g., sensitivity, sampling reproducibility, effect of HCV genotype/subtype, etc.)

Provided by Patrick Harrington