Genotype and Subtype Independent Full Genome Sequencing Assay for Hepatitis C Virus

Charlotte Hedskog, Krishna Chodavarapu, Karin S. Ku, Simin Xu, Ross Martin, Michael D. Miller, Hongmei Mo, Evguenia Svarovskaia

Gilead Sciences Inc,
Foster City, CA, United States.

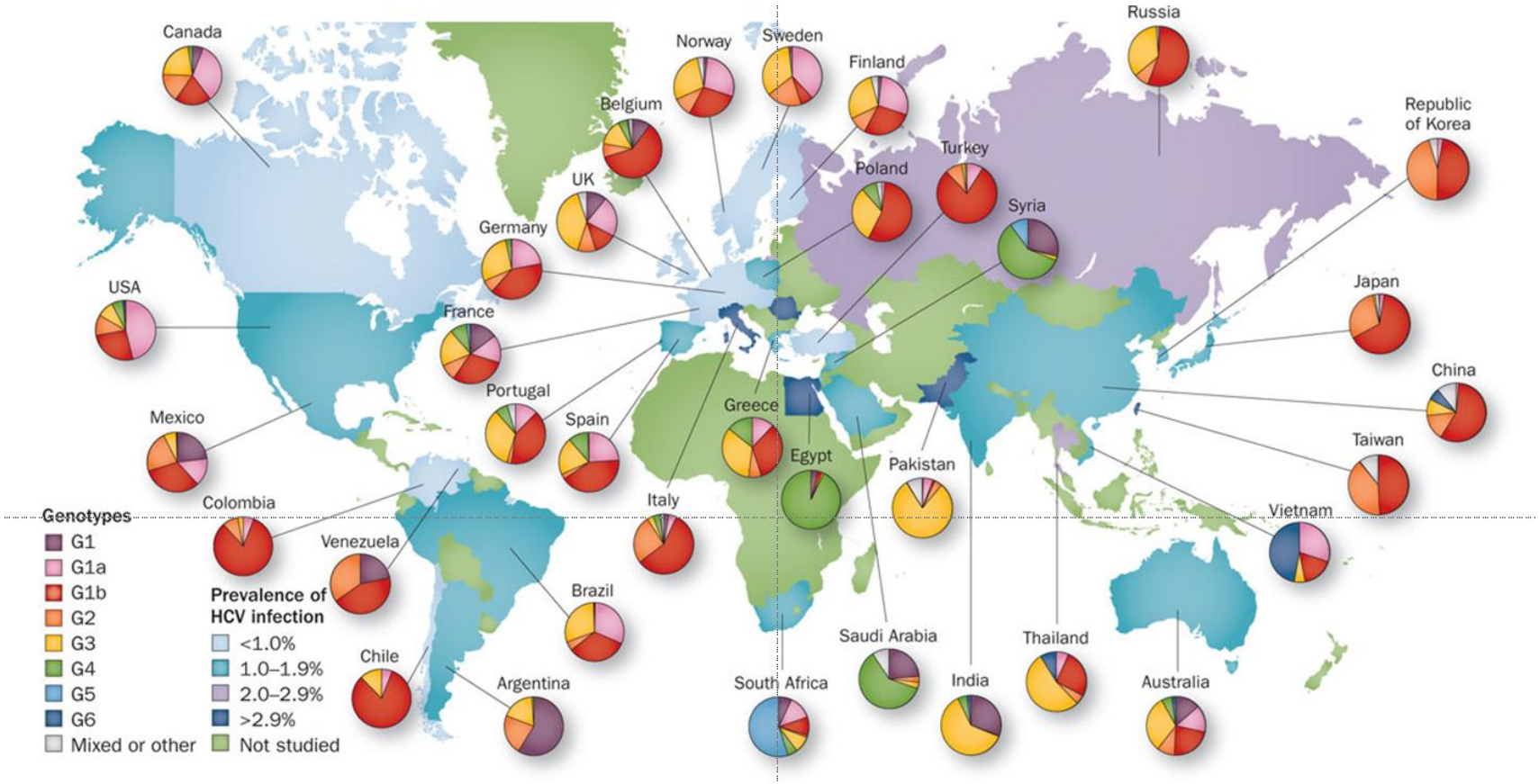
Outline

- Introduction to Hepatitis C virus (HCV)
- Genetic variability of HCV
- Sequencing strategies for HCV drug resistance monitoring in clinical trials
- Subtype independent full genome sequencing assay of HCV

The Global Burden of Disease Due to HCV

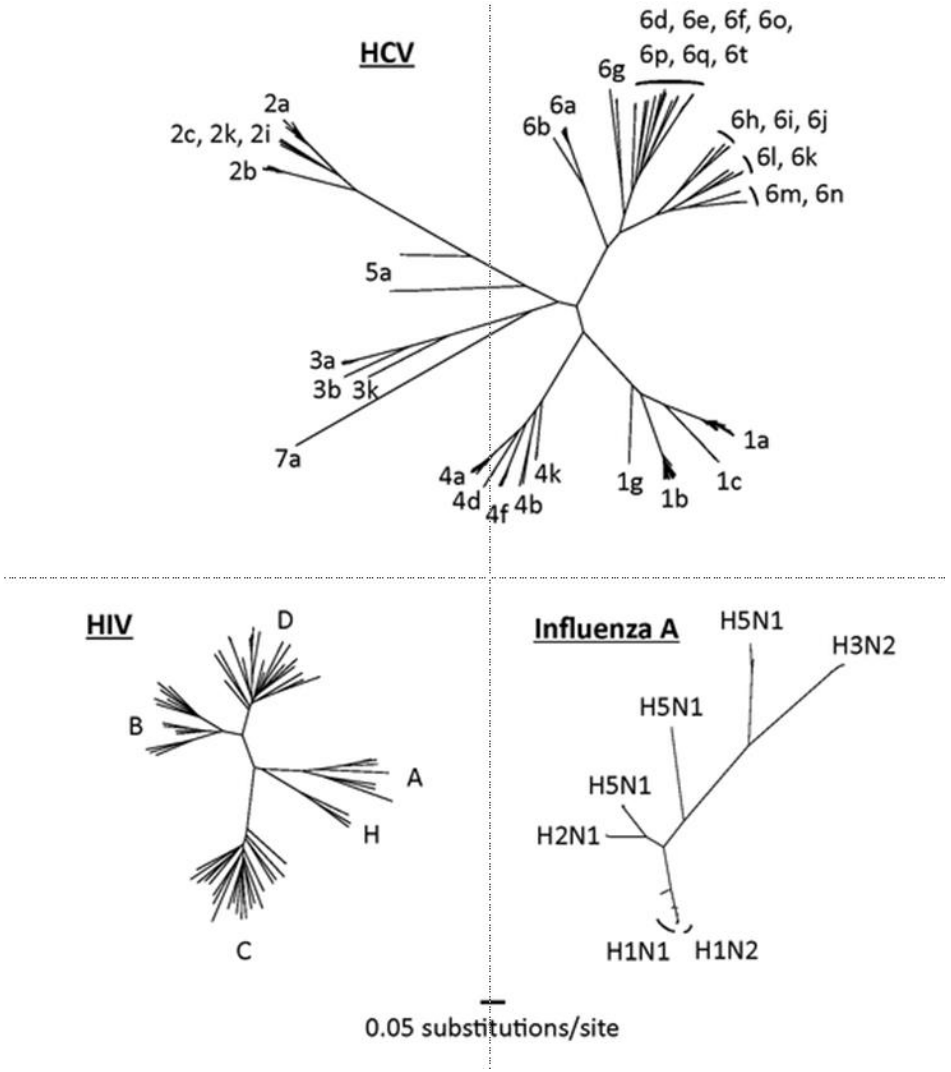
- 170 million people infected (2% world's population)
- 3.2 million infected in the US
- 50% of infected patients have been diagnosed
- Disease complications due to HCV infection
  - hepatic fibrosis
  - cirrhosis
  - hepatocellular carcinoma

The Estimated Prevalence of HCV Infection and the Distribution of HCV Genotypes across the World



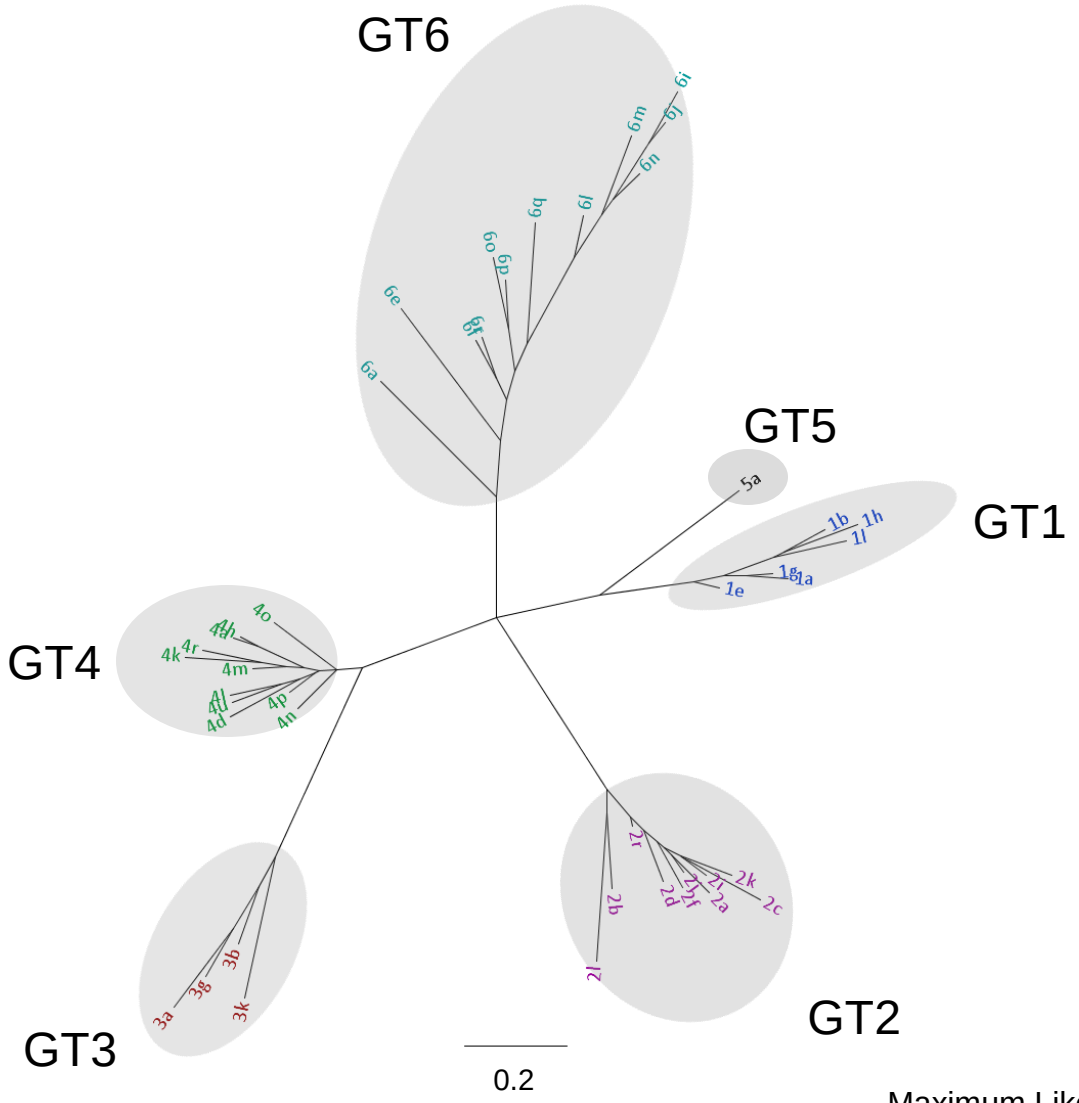
Negro, F. & Alberti, A. Liver Transpl. ***31*** *(Suppl. 2), 1–3 (2011)*

Genetic Diversity of HCV Compared to HIV-1 and Influenza A



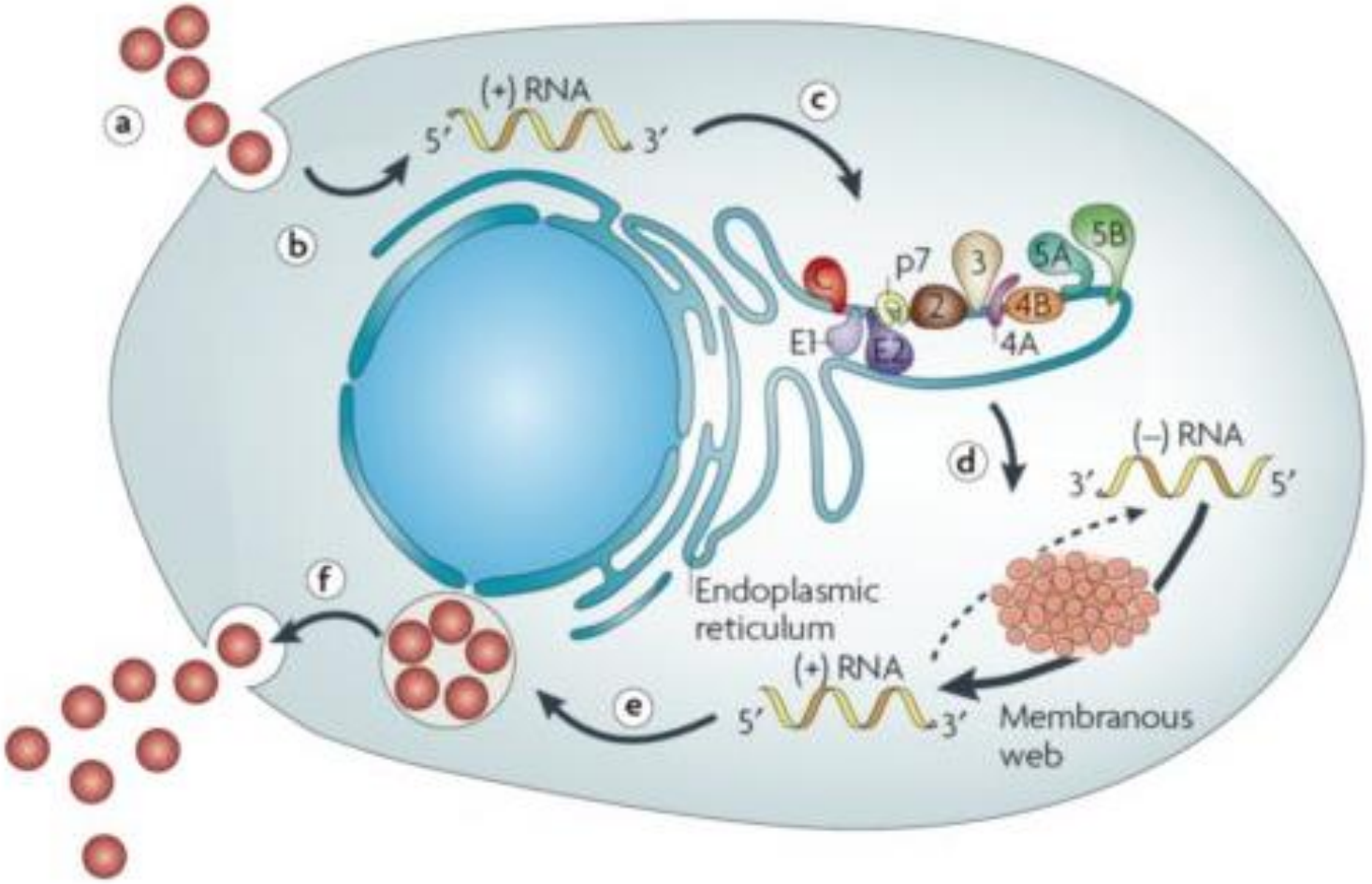
Klenerman P et al. (2009) PLoS Med 6(6): e1000096

44 HCV Subtypes in Gilead Database



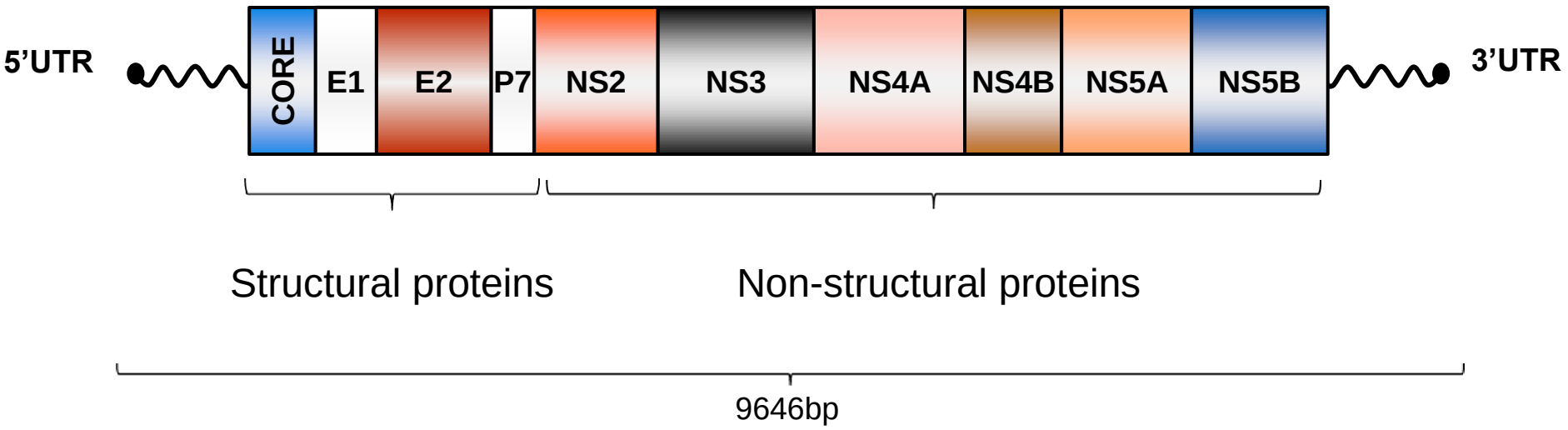
Maximum Likelihood tree

HCV Life Cycle



Moradpour D. et al. Nat. Rev. Microbiol. 2007;5:453-463

Structure of the HCV Genome



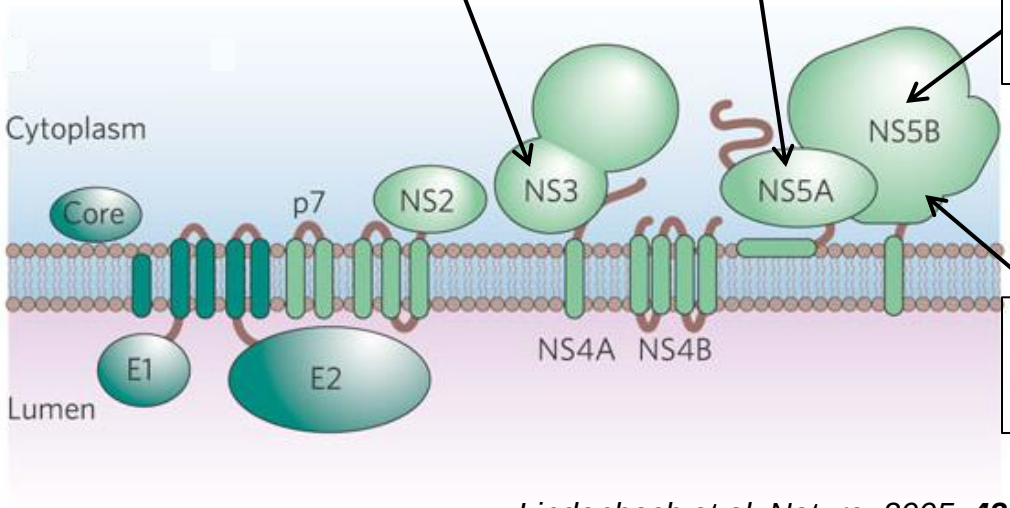
- NS2/3 – Autoprotease
- NS3/NS4A – Serine protease / helicase
- NS4B – Membranous web induction
- NS5A – Exact function is unclear: replication complexes, assembly and secretion
- NS5B – RNA dependent RNA polymerase

Direct Acting Antiviral drugs (DAAs) for HCV treatment

NS3 Protease inhibitors
Telaprevir (GT1)
Boceprevir (GT1)
Simeprevir (GT1)

NS5A inhibitors
Ledipasvir (GT1)
Daclatasvir (GT1)

NS5B Nucleoside analogs (NI)
Sofosbuvir (GT 1-6)
Mericitabine (GT 1-6)



NS5B Non-nucleoside analogs (NNI)

*Lindenbach et al. Nature. 2005. **436**, 933-938*

Benefits of DAA based treatment compared to peg-IFN/RBV treatment

- Shorter duration of treatment
- Higher cure rates
- Less side effects

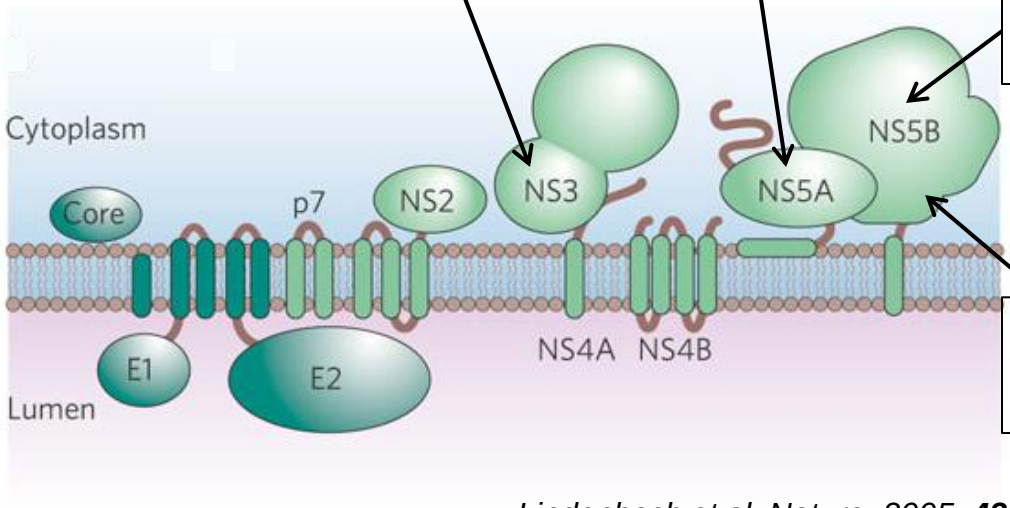
Drug Resistance Development to DAAs

NS3 Protease inhibitors
Telaprevir (GT1)
Boceprevir (GT1)
Simeprevir (GT1)

NS5A inhibitors
Ledipasvir (GT1)
Daclatasvir (GT1)

High resistance barrier

NS5B Nucleoside analogs (NI)
Sofosbuvir (GT 1-6)
Mericitabine (GT 1-6)



NS5B Non-nucleoside analogs (NNI)

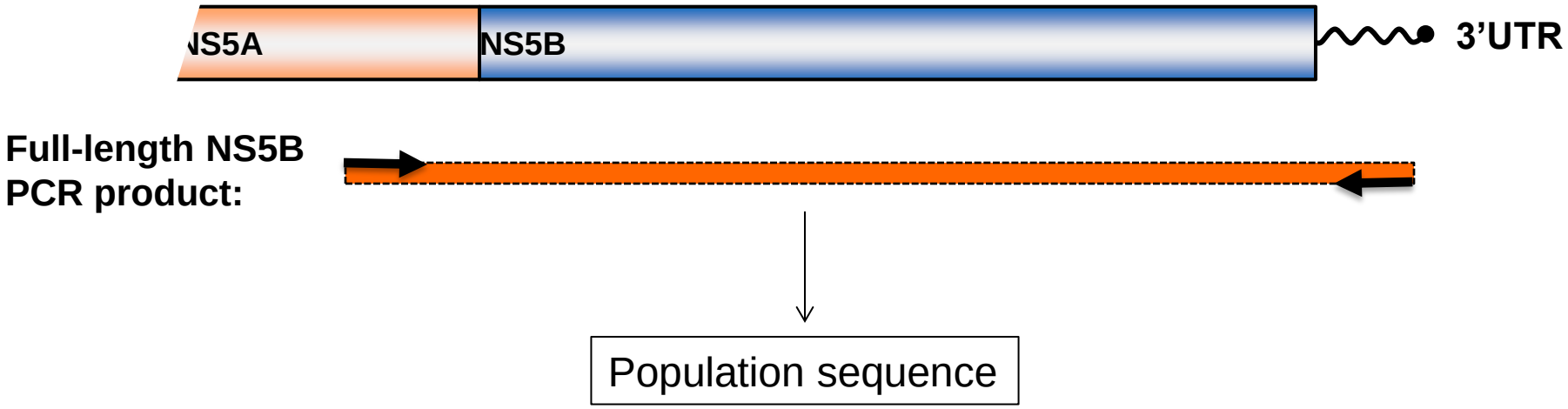
*Lindenbach et al. Nature. 2005. **436**, 933-938*

- NS3 and NS5A is relativity non-conserved
- NS5B nucleoside analogs resistance mutations associated with high fitness cost (S282T)

HCV Sequencing Approaches for Drug Resistance Testing

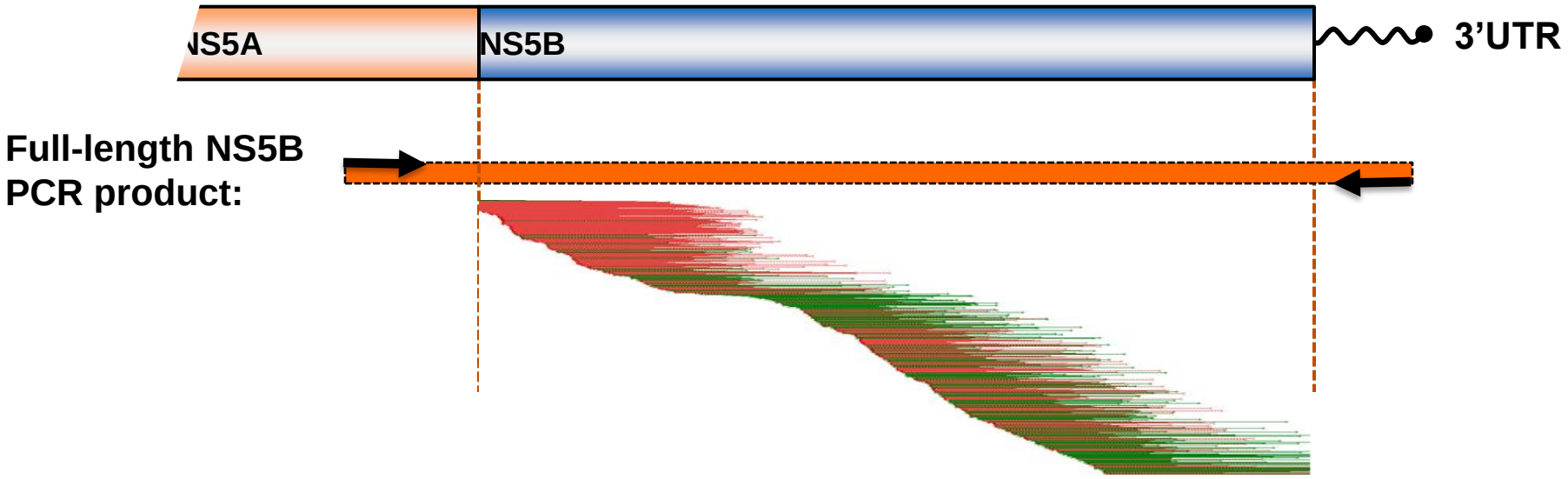
- Population Sanger sequencing of target gene
- Deep sequencing approaches of target gene
- Full HCV genome sequencing

Population Sanger Sequencing of HCV



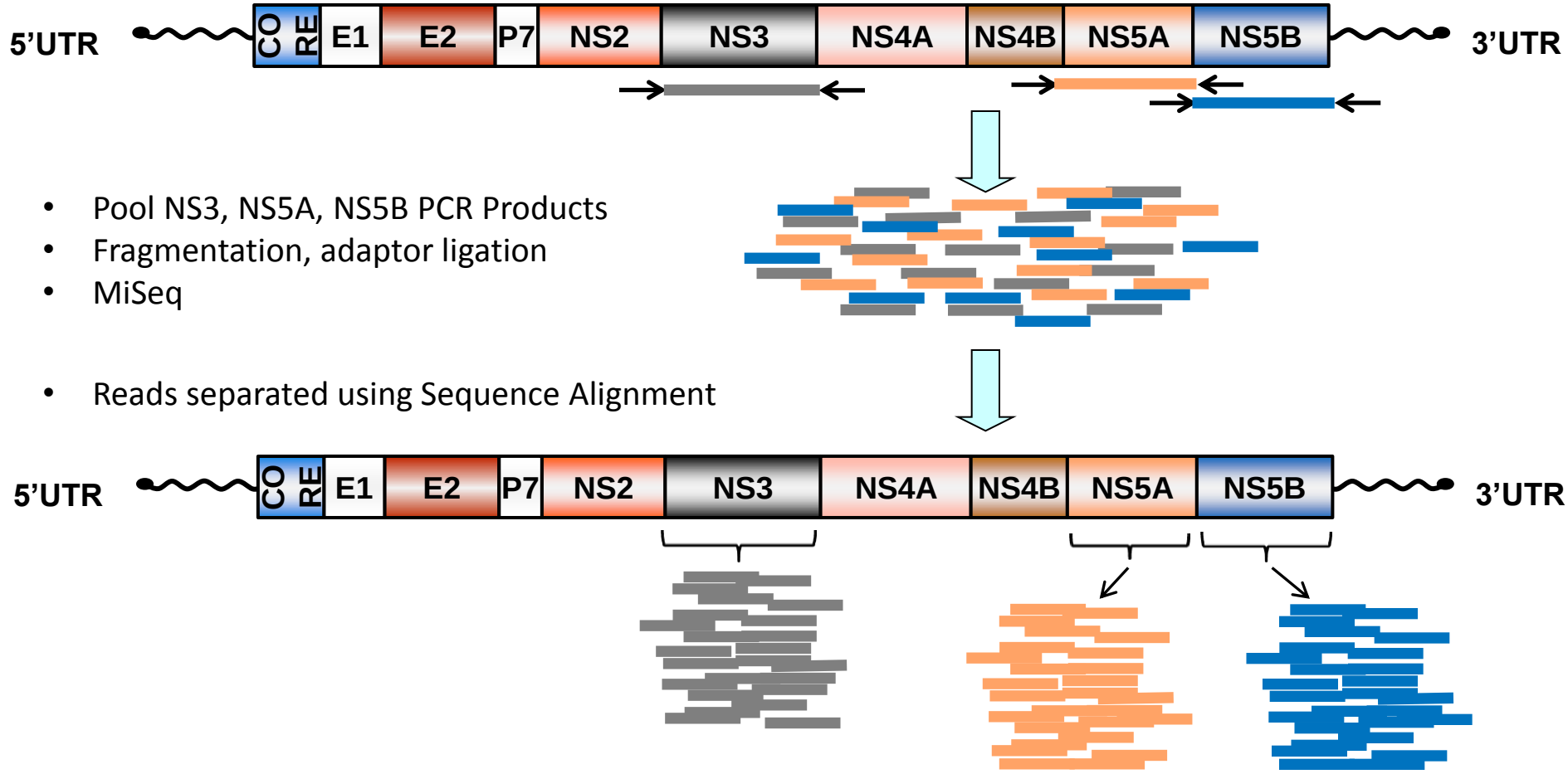
- Well established
- Sensitivity 15-20% of variant/mutant detection
- Primers for amplification established for common GTs
- Primers for sequencing established for common GTs
- Assay not available for all GTs

Deep Sequencing of HCV



- NGS (evaluated 454, Iontorrent, PacBio and Illumina MiSeq)
- MiSeq sensitivity down to 1% of variant/mutant detection
- Primers for amplification established for common GTs
- Assay not available for all GTs

Combining Multiple HCV Targets



- Pool NS3, NS5A, NS5B PCR Products
- Fragmentation, adaptor ligation
- MiSeq

- Reads separated using Sequence Alignment

- Primers for amplification established for common GTs
- Assay not available for all GTs

Benefits and Disadvantages of HCV Deep Sequencing Approach

+ Sensitivity 1% of variant/mutation detection
+ Primers established for common GTs
+ Increased efficiency by combining multiple targets

- Assay not available for all GTs: For pan-genotypic HCV drug development an subtype independent assay is needed
- Investigation of potential drug resistance associated variants outside of drug target gene – approaches to sequence the whole HCV genome is needed

Approaches for Sequencing Full HCV Genome

- Amplification of overlapping regions using gene-specific primers coupled with either Sanger sequencing or NGS *(Newman et al. 2013, Okamoto et al. 1992, Hmaied et al. 2007, Lauck et al. 2012)*
  - + Efficient for known GT
  - - Less useful for rare/unknown GTs
  - - Primer skewing

- RNA-Seq: Random amplification of total RNA in the sample *(Niomiya et al. 2012)*
  - + Subtype independent
  - - High human background (>99% non-HCV)
  - - Assembly difficulty through highly variable regions such as E1/E2
  - - Not complete HCV genome generated

- NuGEN random amplification coupled with NGS (*Malbeouf et al. 2013*)
  - + Complete coding regions of HIV, RSV and West Nile Virus generated
  - + Subtype independent

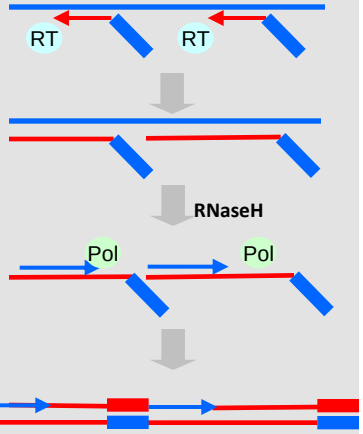
Full HCV Genome Sequencing Assay

RNA extraction

HCV genome

cDNA synthesis (NuGEN kit)

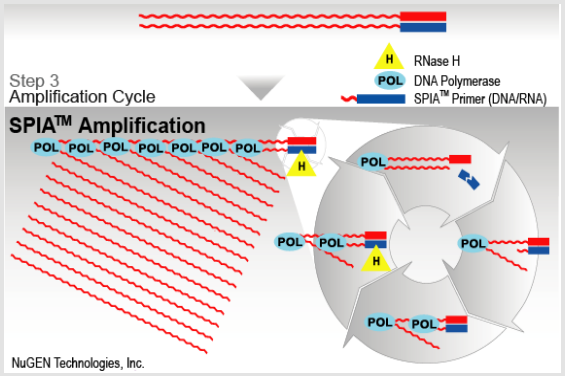
cDNA synthesis
random primers



Second strand
DNA synthesis

dsDNA

DNA amplification (NuGEN kit)



Quality control

Bioanalyzer and
Nanodrop

Deep sequencing

Library prep and
MiSeq at CRO

Data analysis

- Approx. 2 million sequences are generated per sample
- *De novo* assembly by Vicuna – creating contigs
- Contigs are aligned to HCV reference – a draft full genome assembly
- Iterative refinement of the full genome assembly by alignment of trimmed and filtered reads to the assembly (x3)
- Full genome consensus sequence is generated

Assay validation

In vitro transcribed RNA experiments

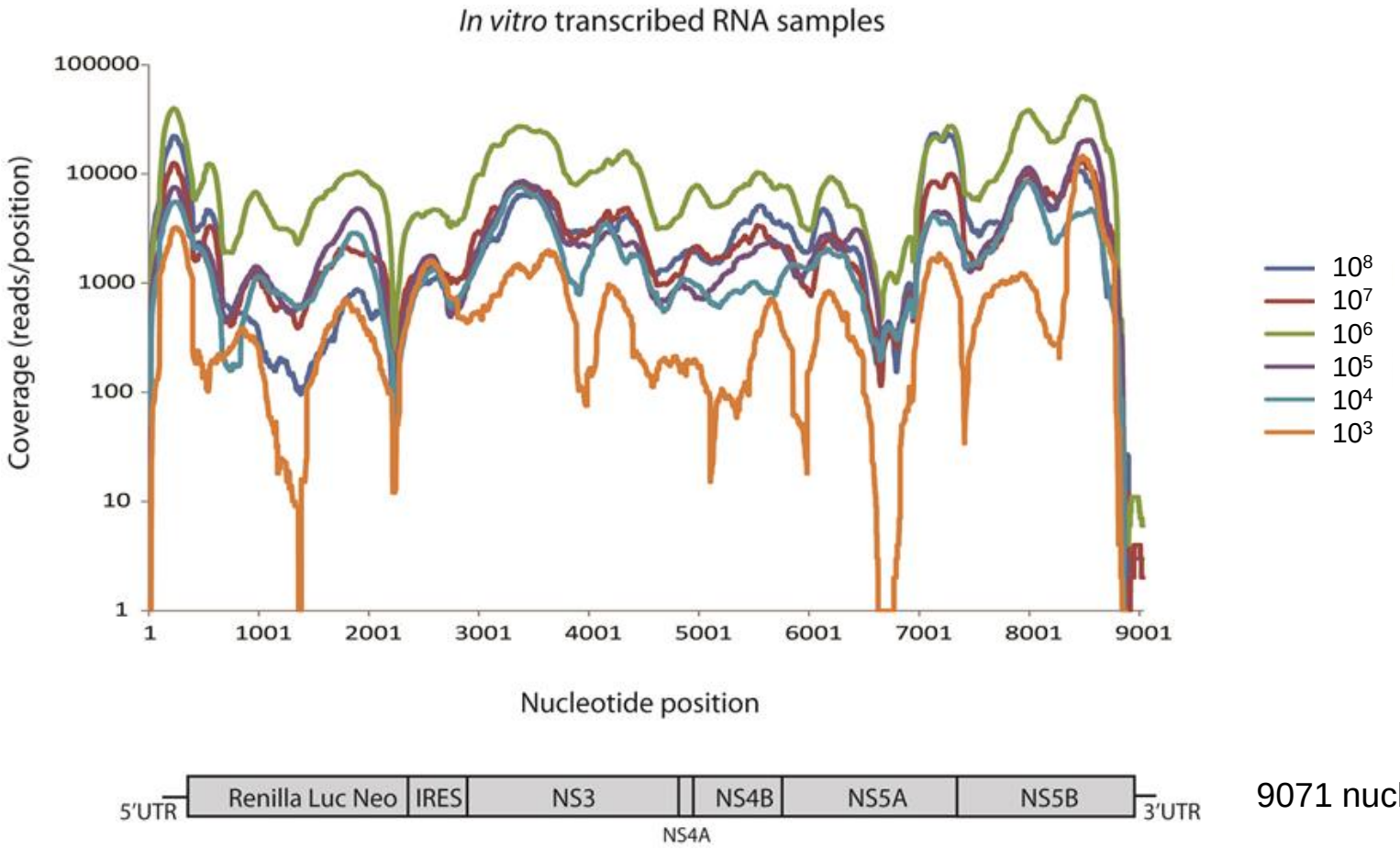
Patients' samples
17 patients with GT 1 to 6

Validation of Assay by *in vitro* Transcribed HCV RNA

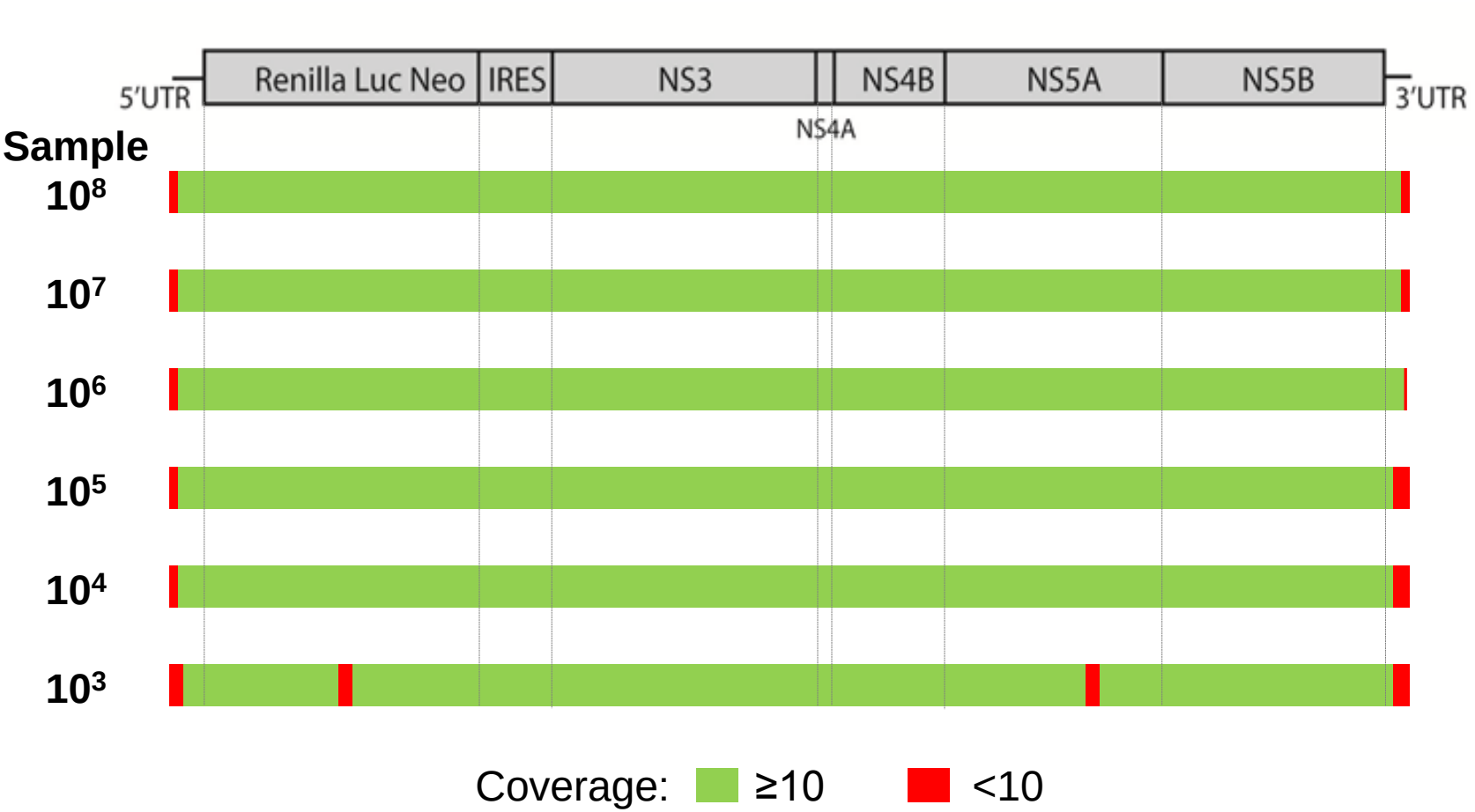
- RNA transcripts were generated from HCV genotype 2a replicon (2a-RlucNeo)
- The RNA was spiked with replicon containing 10% NS5B S282T
- RNA was 10-fold diluted into six RNA input copies per reaction (10^8 – 10^3 molecules)
- NuGEN amplification and MiSeq was performed

Successful Sequencing of *in vitro* Transcribed RNA

- On average 653,935 reads were generated per sample where 64% were aligning to HCV



Full Genome Consensus Sequences generated from *in vitro* Transcribed RNA



Sequence identity to ref
100%
100%
100%
100%
100%
99.98%

- Generated consensus sequences spanned 96-100% of the HCV coding region (8474 nucleotides) with high accuracy

Sensitivity of the Full Genome Sequencing Assay

Input RNA copies per reaction	S282T 10%
10^8	10.6% (688/6464)
10^7	9.7% (874/9036)
10^6	9.7% (1024/10552)
10^5	11% (1144/10364)
10^4	14.6% (786/5392)
10^3	5.9% (16/272)

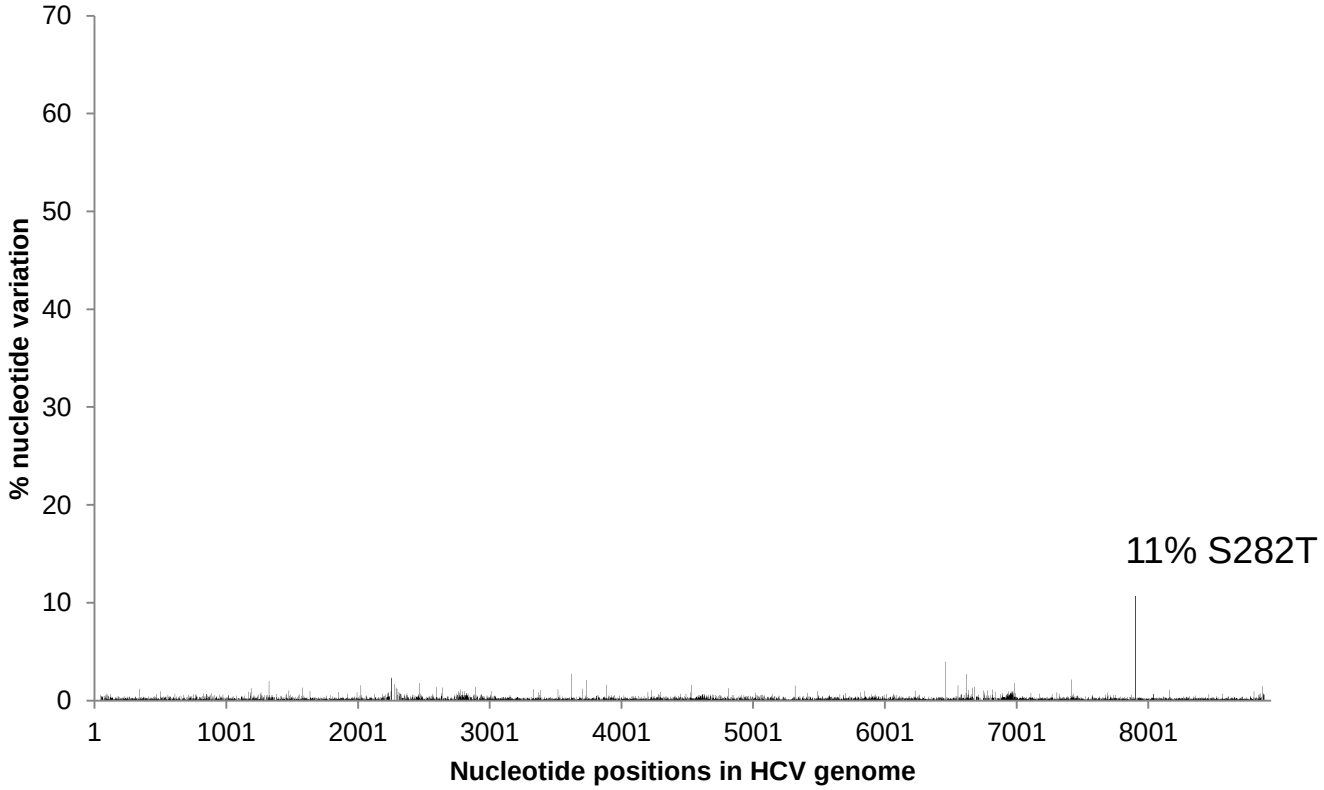
- The frequency of S282T was consistent with the 10% S282T addition

Background Noise across the HCV Genome

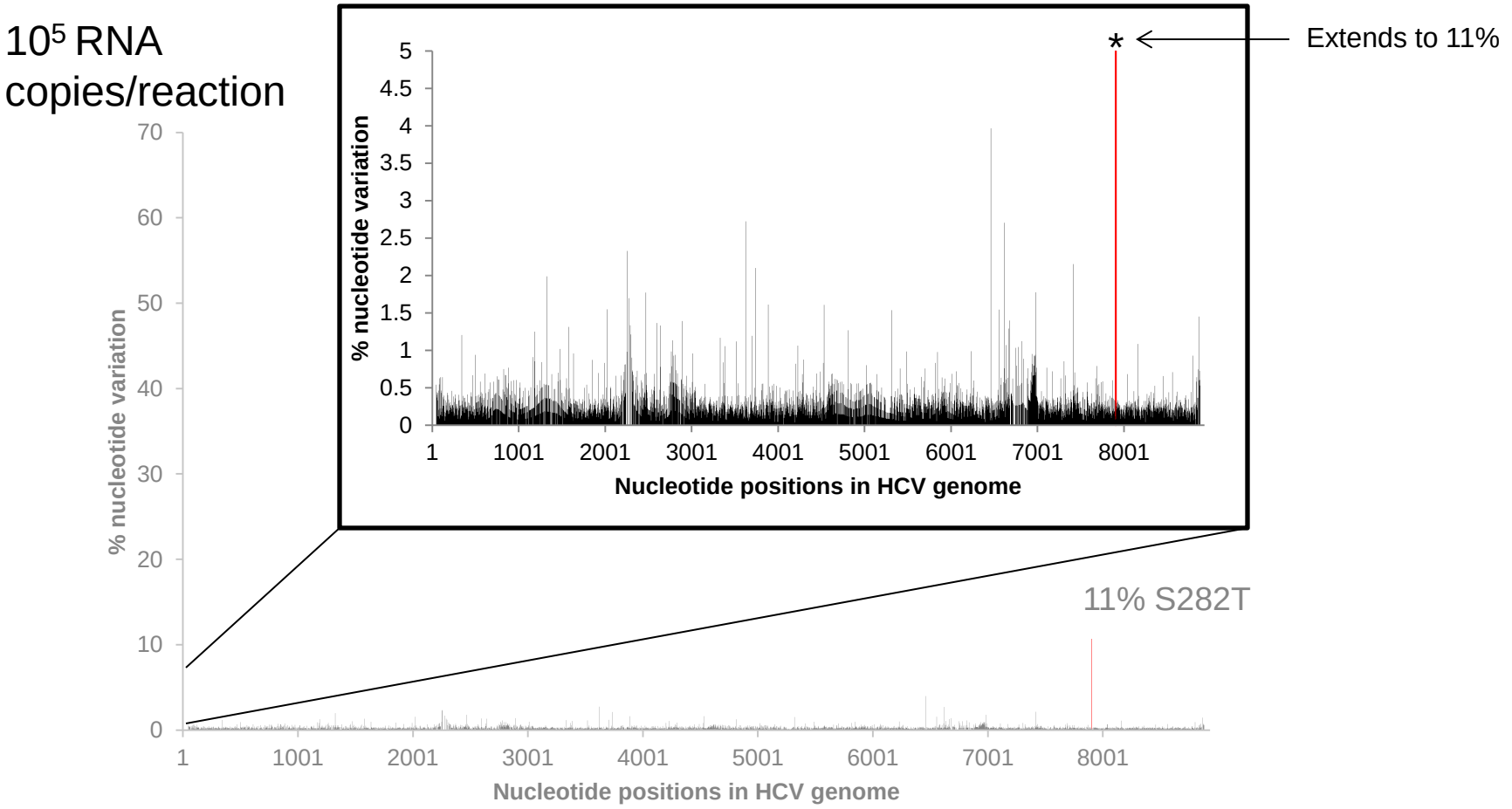
- Background noise across the HCV genome was evaluate at nucleotide level
- Due to the clonal origin of the *in vitro* transcribed RNA all genetic variations in the generated sequence reads was likely to be due to amplification or sequencing errors (except the 10% S282T)
- Variation was defined as the percent composition of all but the most prevalent nucleotide at each position using the trimmed and filtered reads
- The average background noise and a 95% confidence interval were calculated for each sample

Background Noise in the Generated Full Genome Sequences from the *in vitro* Transcribed RNA

10^5 RNA copies/reaction



Background Noise in the Generated Full Genome Sequences from the *in vitro* Transcribed RNA



- Average background noise 0.16%, 95% CI [0.156, 0.164]

Background Noise in *in vitro* Transcribed RNA Samples

Input RNA copies per reaction	Average % background noise ± 95% CI	Maximum background noise
10^8	0.28% ±0.0049	4.5%
10^7	0.34% ± 0.0078	5.3%
10^6	0.28% ± 0.0039	2.8%
10^5	0.16% ± 0.0037	4.0%
10^4	0.22% ± 0.0067	8.0%
10^3	0.19% ± 0.0107	21.3%

- The average nucleotide variation was similar in all samples
- Maximum nucleotide variation was higher in the 10^3 sample which had the lowest amount of input RNA molecules

Full Genome Sequencing of HCV

- Method established and validated on *in vitro* transcribed RNA
- Consensus sequences >99% match to reference
- Average background noise is ~0.2%
- Sensitivity of at least 10% of variant/mutant detection

Next

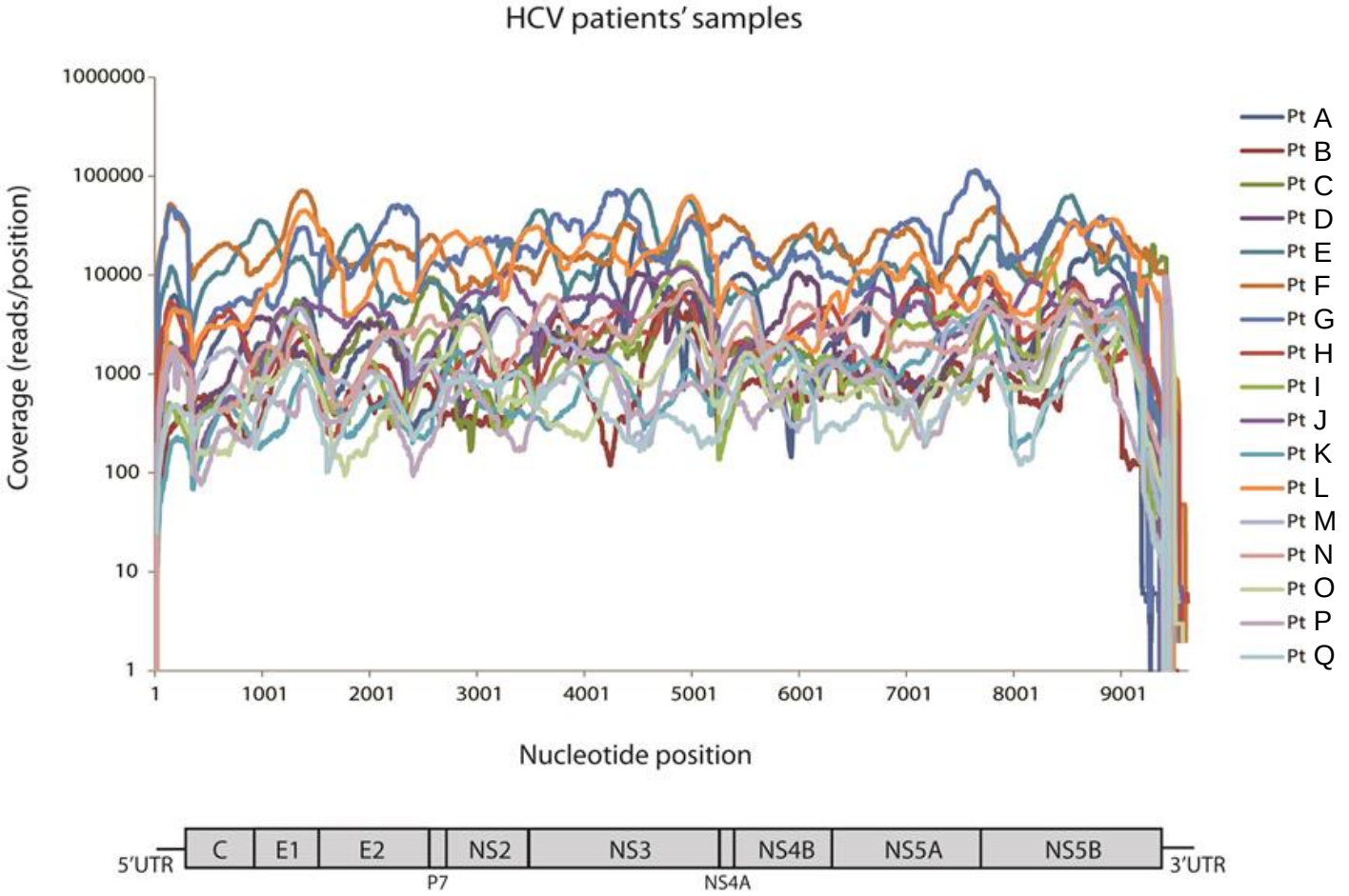
Full genome sequencing of patient's samples with genotype 1 to 6

Full Genome Sequencing of Patients' Samples

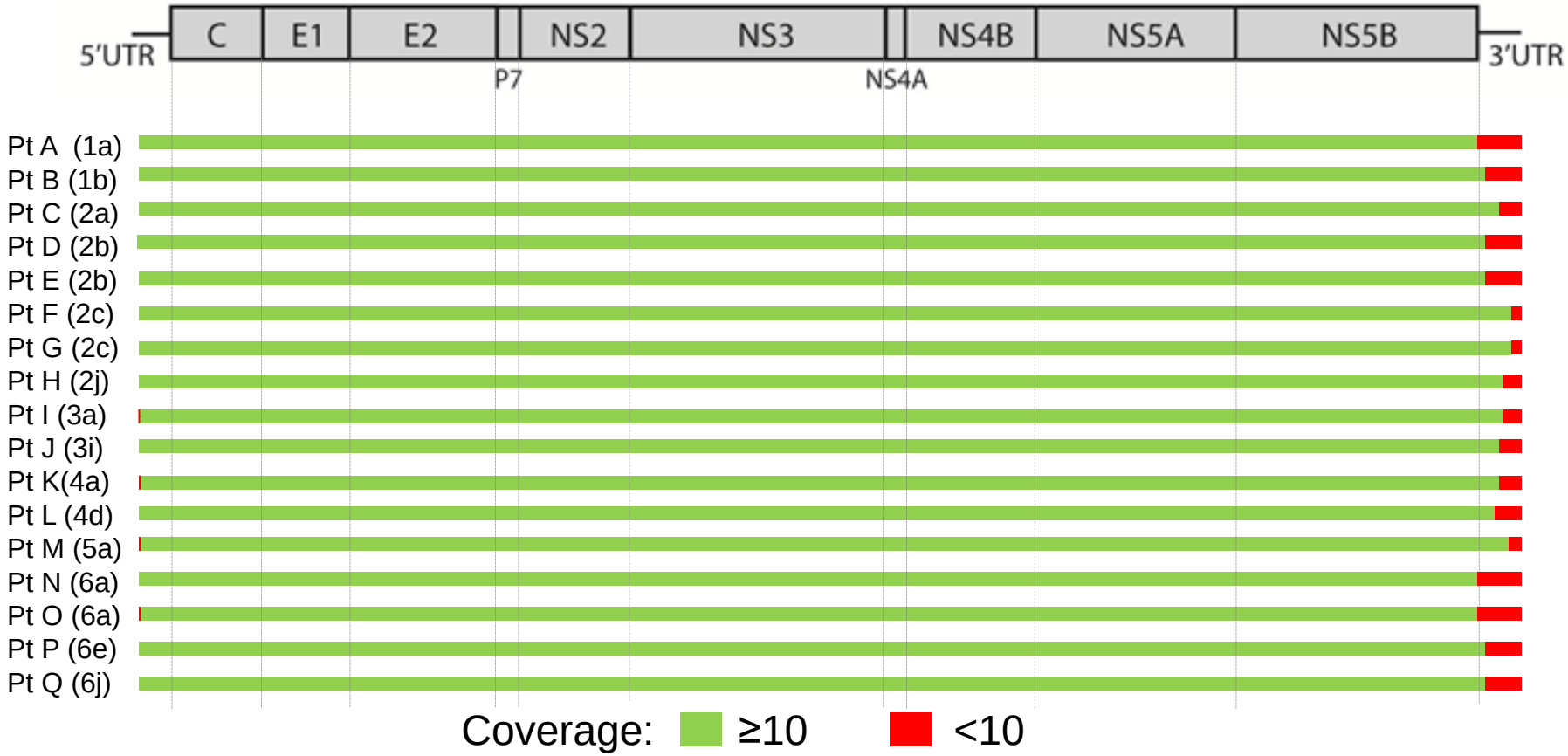
Patient ID	HCV viral load IU/mL	HCV RNA copies per reaction	GT
Patient A	5,660,000	~10^5	1a
Patient B	23,700,000	~10^6	1b
Patient C	9,100,000	~10^5	2a
Patient D	9,980,000	~10^5	2b
Patient E	1,952,000	~10^5	2b
Patient F	18,700,000	~10^6	2c
Patient G	22,600,000	~10^6	2c
Patient H	5,620,000	~10^5	2j
Patient I	2,300,000	~10^5	3a
Patient J	1,820,000	~10^5	3i
Patient K	8,400,000	~10^5	4a
Patient L	5,820,000	~10^5	4d
Patient M	5,820,000	~10^5	5a
Patient N	32,300,000	~10^6	6a
Patient O	13,300,000	~10^6	6a
Patient P	5,500,000	~10^5	6e
Patient Q	21,300,000	~10^6	6j

Full HCV Genome Sequencing of HCV Patients' Samples

- On average 3.7 million reads were generated per sample where 15% were aligning to HCV

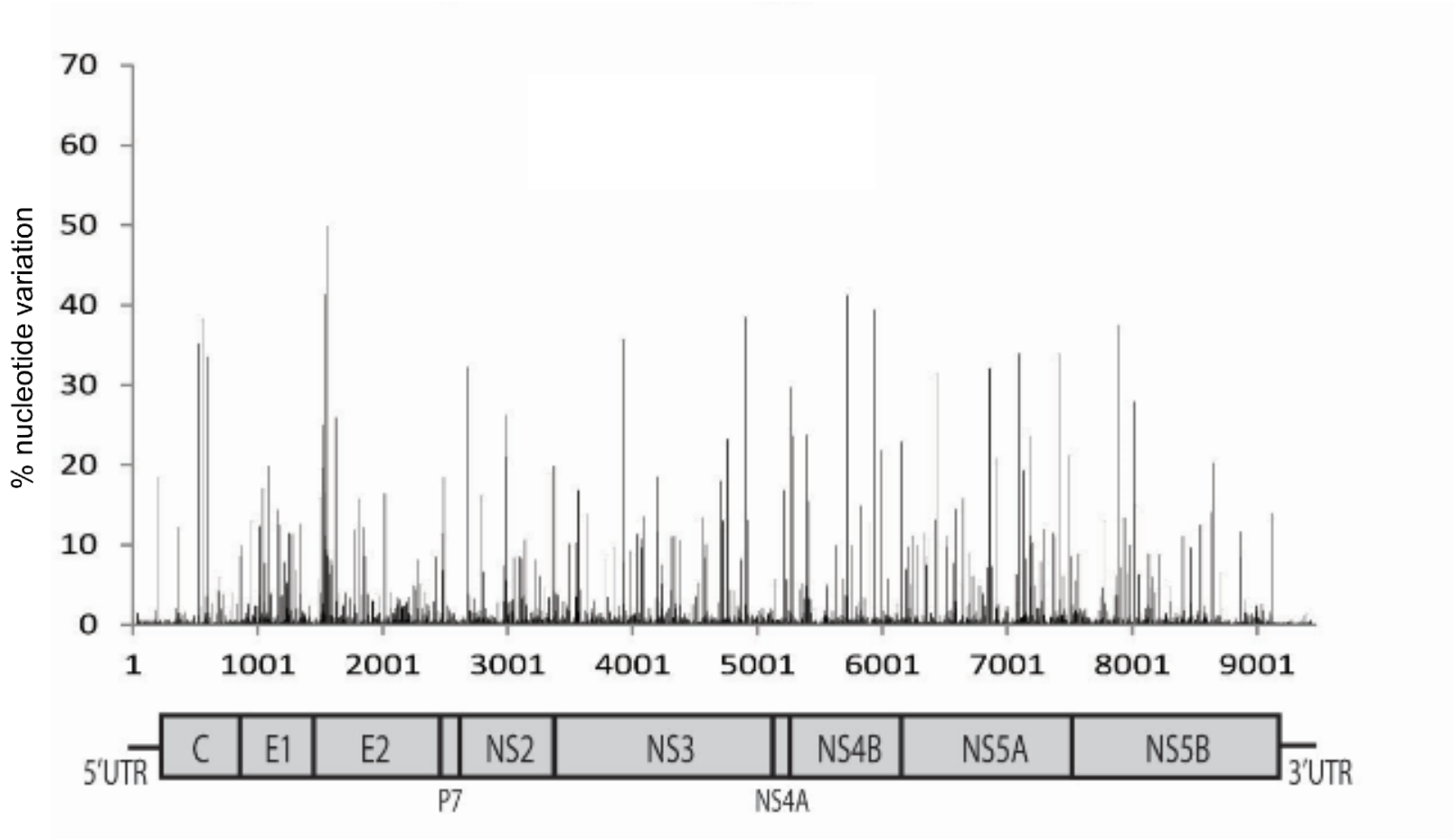


Full HCV Genome Consensus Sequences Generated from Patients' Samples



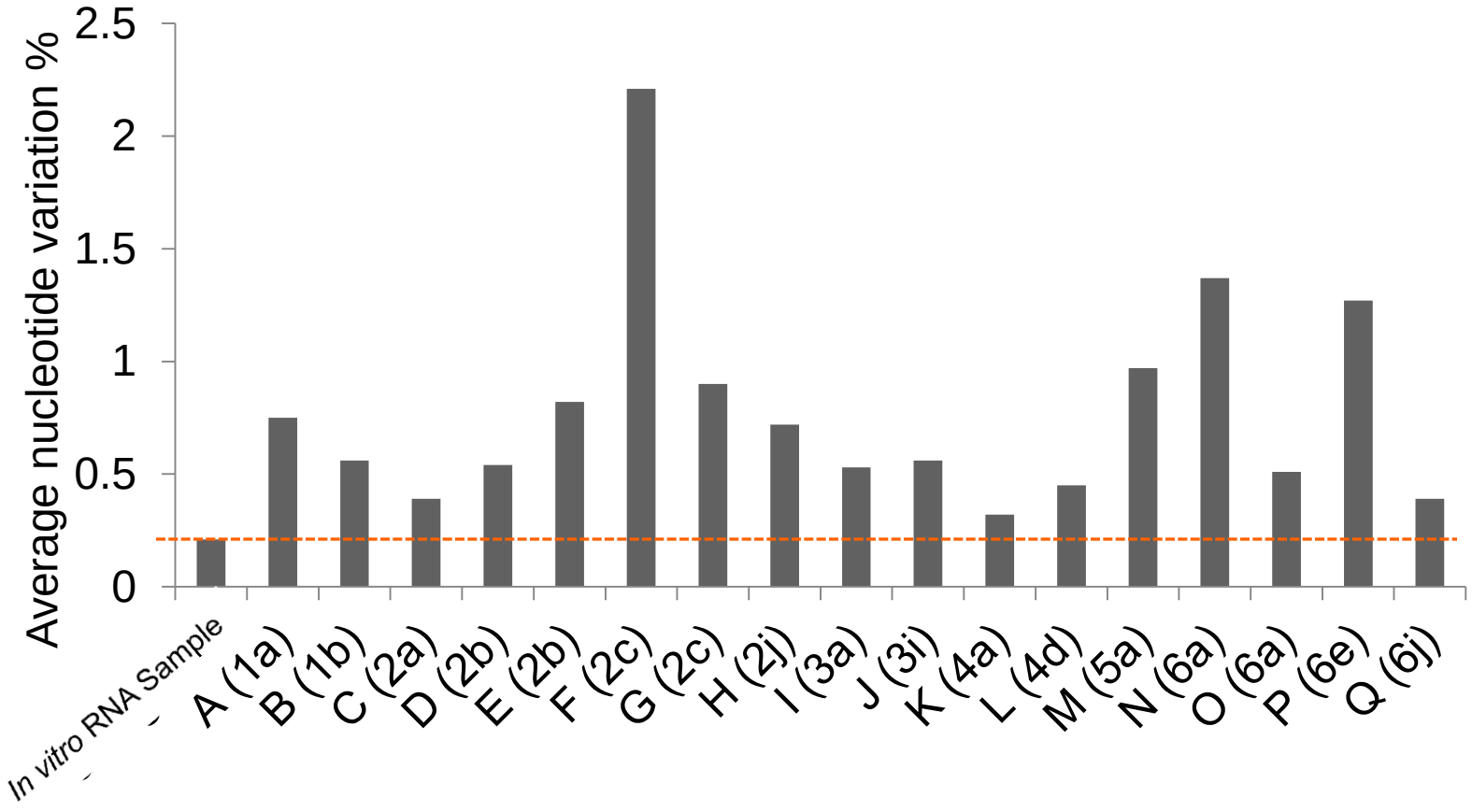
- Generated consensus sequences spanned 99.3-100% of the HCV coding region (9,036 nucleotides) for GT 1-6

Nucleotide Variability Within the HCV Quasispecies in Patient D (2b)



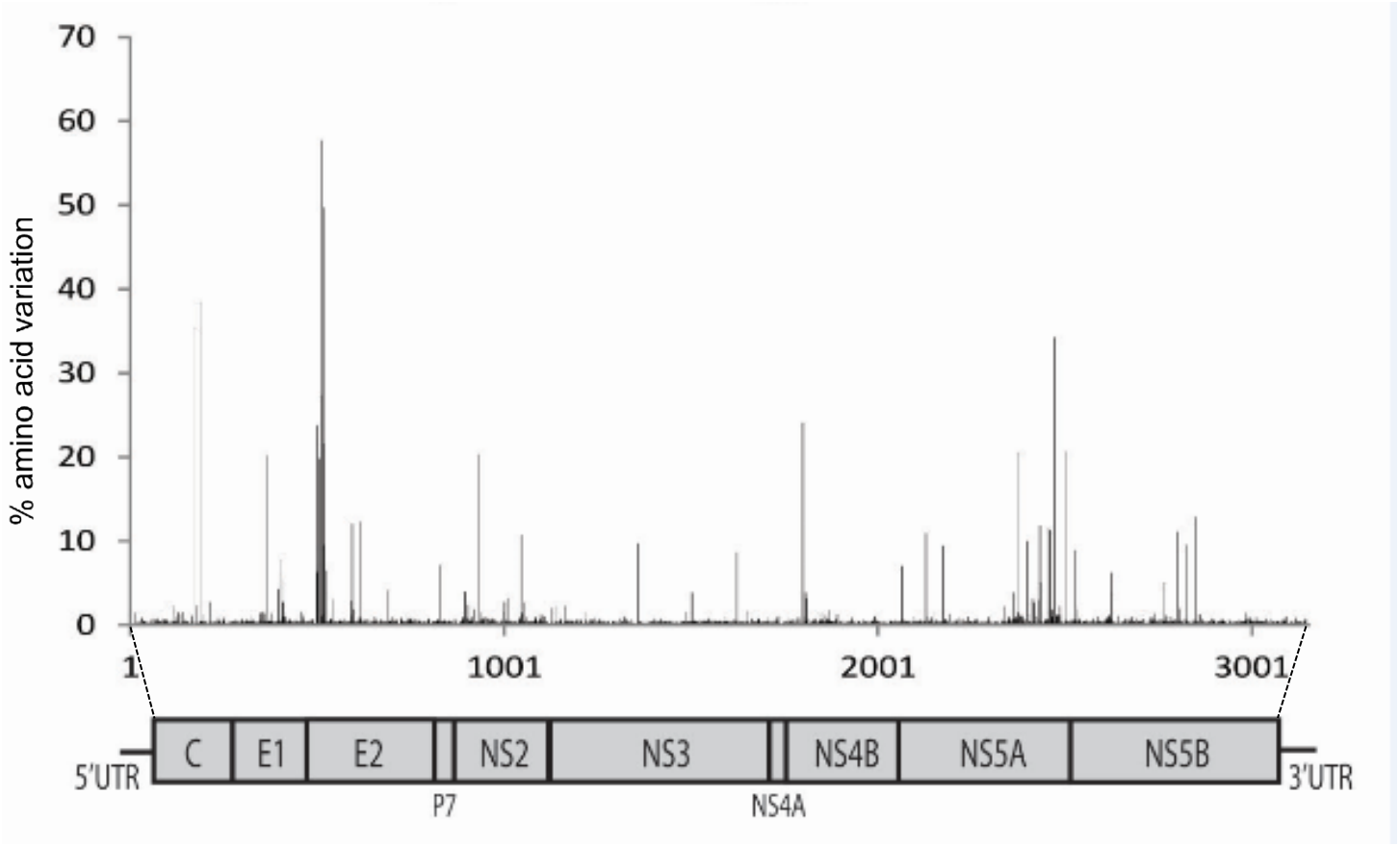
- Nucleotide variation was spread out throughout the HCV genome
- No specific hotspots for nucleotide variation

Average Genetic Nucleotide Variation within the HCV Quasispecies



- Nucleotide variation was substantially different between the patients
- Variation was not significantly correlated with VL or coverage

Amino Acid variability within the HCV Quasispecies in Patient D (2b)

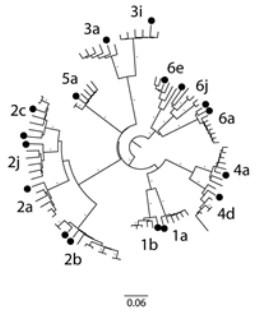


- Amino acid variations were predominantly found in E1/E2 and end of NS5A for all patients
- E2 and NS5A targeted by immune system *(Simmonds et al. 2004)*

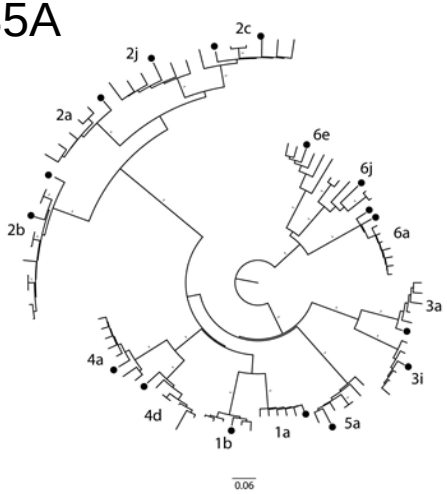
Genotype Diversity at Different HCV Genomic Regions

E1/E2

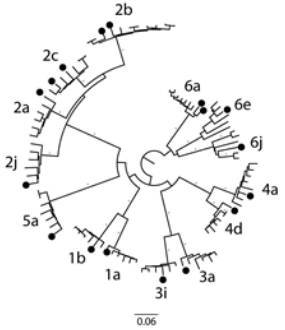
NS3/NS4A



NS5A



NS5B



0.06

- HCV subtype classification was confirmed for each patient
- Genetic distance between the sequences was higher in E1/E2 and NS5A compared to NS3/4A and NS5B regions

Variants of Potential NS3 and NS5A Resistance Associated Amino Acid

		NS3										NS5A								
Patient sample	**GT**	V 36 any	F 43 any	T 54 any	V 55 any	Q 80 any	S 122 R	R 155 any	A 156 any	D 168 any	I 170 A/T/L	L 175 L	K 24 G/N/R	M 28 A/G/T	Q 30 any	L 31 any	P 32 L	S 38 F	H 58 D	A 92 K/L	Y 93 any
A	**1a**											L									
B	**1b**														R	L/M					
C	**2a**	L				G						L			K	M					
D	**2b**	L				G	R					L			K						
E	**2b**	L				G	R					L			K	M					
F	**2c**	L				G	R					L			K	F					
G	**2c**	L				G	R					L			K	M					
H	**2j**	L				G	R					L			K	M					
I	**3a**	L								Q		L			A						
J	**3i**	L								Q		L			K						
K	**4a**	L										L			S	M					
L	**4d**	L										L			R	M					
M	**5a**	L				K						L									T
N	**6a**					K									R						T
O	**6a**					K									R						T
P	**6e**														S						S
Q	**6j**														A						T

Variants of Potential NS5B Resistance Associated Amino Acid

		NI					NNI																			RBV	
Patient sample	GT	L 159 any	S 282 any	C 289 any	L 320 any	V 321 any	S 96 any	N 142 T	C 316 any	M 414 I/T/V	L 419 any	R 422 K	M 423 any	C 445 F	Y 448 any	Y 452 any	I 482 any	A 486 any	V 494 A	P 495 any	P 496 S	A 499 A	G 554 S	S 556 G	D 559 G	T 390 I	F 415 Y
A	1a																					A					
B	1b																										Y
C	2a			M							I			F			L		A			A		G			Y
D	2b			M							I			F			L		A			A		G			Y
E	2b			M							I			F			L		A			A		G			Y
F	2c			M							I			F			L		A			A	S	G			Y
G	2c			M							I			F			L		A			A	S	G			Y
H	2j			M							V			F			L		A			A		G			Y
I	3a			F							I			F			L					A		G			Y
J	3i			F							I			F			L					A		G			Y
K	4a			F						V	I			F			L					A		G			Y
L	4d			F						I	I			F			L					A		G			Y
M	5a			M									I	F								A		G			Y
N	6a			M							I			F			L	G	A			A					
O	6a			M							I			F			L	G	A			A					
P	6e			L							I			F			L		A			A					Y
Q	6j			M							I			F			L		A			A					Y

- NS5B S282T mutation associated with resistance to sofosbuvir and mercitabine was not detected in any of the subtypes

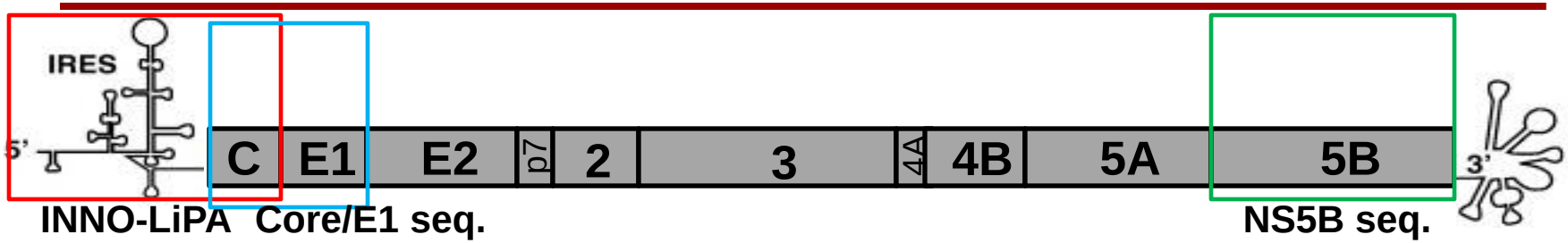
Full Genome Sequencing of HCV Patients' Samples

- Amplification and sequencing was successful for 17 patients' samples with genotype 1 to 6, including subtypes with limited sequence information
- Consensus sequences spanned 99.3-100% of the HCV coding region, including the highly variable E2 region
- Full genome sequences enables investigation of genetic variability of viral quasispecies and presence of potential resistance associated variants

Genotype Discordance Between Inno-LiPA and NS5B Sequencing Genotyping Methods

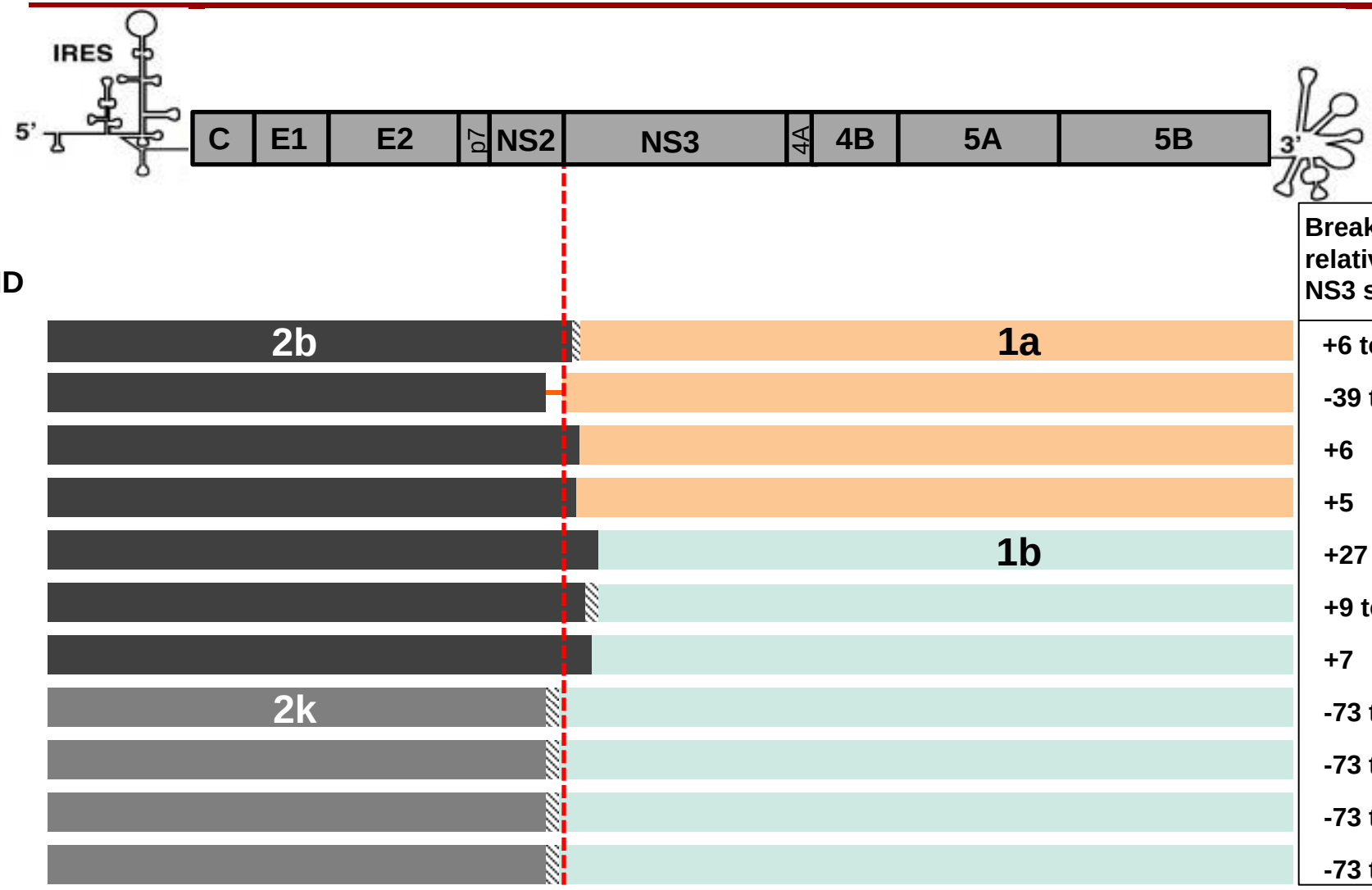
	Patients, N	Genotype	
		Concordant	Discordant
GT 1	840	840 (100%)	0
GT 2	487	475 (97.5%)	12 (2.5%)
GT 3	986	986 (100%)	0
GT 4	39	39 (100%)	0
GT 5	1	1 (100%)	0
GT 6	10	10 (100%)	0
Total	2363	2351 (99.5%)	12 (0.5%)

Discordant Genotyping Results



		Population sequencing	
Subj ID	INNO-LiPA 5'UTR	Core/E1	NS5B
A	2b	2b	1a
B	2b	2b	1a
C	2b	2b	1a
D	2b	2b	1a
E	2b	2b	1b
F	2a/2c	2k	1b
G	2b	2b	1b
H	2b	2b	1b
I	2	-	1a
J	2	-	1b
K	2a/2c	-	1b
L	2	-	1b

Full Genome Sequencing of HCV Inter-Genotypic Recombinant Viruses



Hedskog et al. Hepatology 2014

- **Recombination breakpoint was located within ~80 aa of the NS2-NS3 junction**

Summary

- Hepatitis C infection is a global health problem and pan-genotypic drugs are in development
- Efficient sequencing strategies for drug resistance analysis are essential to support pan-genotypic testing
- Subtype independent full HCV genome sequencing assay has been established
  - Successful sequencing of genotype 1 to 6
  - Low background noise in control experiments
  - Sensitivity down to at least 10%
- Twelve inter-genotypic HCV recombinant viruses characterized using the Full HCV genome sequencing assay

Acknowledgements

Gilead Sciences
Clinical Virology

Jenny Svarovskaia
Hongmei Mo
Michael Miller
Krishna Chodavarapu
Viktoria Gontcharova
Ross Martin
Brian Doehle
Karin Ku
Simin Xu
Hadas Dvory-Sobol
Bin Han
Joe McCarville
Wen Ouyang
Angela Worth

CRO
Centrillion Biosciences

NuGEN

Broad Institute
Vicuna program

Patients