**Appendices**

**Appendix 1: ELISpot analyses; detailed methodology**

## PBMCs: Blood samples were obtained in EDTA Vacutainer tubes and PBMCs separated by Ficoll density gradient centrifugation at 800g for 20min at 20˚C. Cells were washed and centrifuged twice in PBS and 1ml of ACK lysing buffer then resuspended in FCS and freezing mix (20% DMSO 80% RPMI) in a 1:1 ratio. Cells were stored at -150˚C in liquid nitrogen. Cells were thawed at 37oC in a water bath using 5x dilution with R10 medium supplemented with DNAse and centrifuged at 20oC 250g for 10 minutes. Cells were resuspended with 4 ml/thawed vial of R10 medium and incubated overnight at 37°C in a 5% CO2 incubator.

**ELIspot assay.** Polyvinylidene plates (96-well; Millipore) were coated with 100µl 1:200 dilution of mouse anti-human gamma interferon (IFNγ) antibody (MAbTech) in sterile phosphate-buffered saline at 4°C overnight. Thawed peripheral blood mononuclear cells (PBMC) were added at 100-200,000 cells/well in 100µl of R10 medium (RPMI 1640 [Sigma-Aldrich], 10% foetal calf serum [Sigma-Aldrich], and 10 mM HEPES buffer [Sigma-Aldrich] with 2mM glutamine and antibiotics [50U of penicillin-streptomycin/ml]). Peptides were added directly to the wells at a final concentration of 6µg/ml. The plates were incubated for 18 h at 37°C in 5%CO2. Plates were then washed, labelled with a 1:2000 dilution of biotinylated mouse anti-human IFN-gamma monoclonal antibody (MAbTech), and developed by incubation with alkaline phosphatase-conjugated anti-biotin monoclonal antibody (Vector Laboratories), followed by incubation with 5-bromo- 4-chloro-indolylphosphate (BCIP)–nitroblue tetrazolium (NBT) (Pierce) in Tris buffer (pH 9.5). The reaction was stopped by washing with tap water, and the plates were dried overnight, prior to counting on an Elispot reader. If the negative control (DMSO) background was greater than 55 spot-forming cells (SFC)/106 PBMC, the sample was retested. Responses were considered positive if the number of spots per well minus the background was at least 25 SFC/106 PBMC; concavalin A served as a positive control.

**Appendix 2: Reducing PCR and cloning error rates**

*PCR contamination error*: Universal precautions to avoid PCR contamination were followed at all times. PCR mixes were prepared in an allocated room separate from clinical samples. In order to confirm the possibility of multiple superinfection events and to rule out PCR contamination error, PCR amplification of the HCV NS5B region (in addition to E2 HVR-1 amplification) was carried out. Phylogenetic analysis of cloned amplicon sequences resulted in the production of similar trees with high bootstrap values.

**Supplementary Figure 1: Phylogenetic tree using HVR-1 and NS5B sequences derived from the same time points in a) a patient with evidence of 3 circulating strains at baseline and b) a patient with similar strains at 2 time points**

1. **HVR-1 NS5B**
2. **HVR-1 NS5B**

**Supplementary Figure 1:** Phylogenetic trees were constructed using sequences derived from 6 patients at 2 time points following amplification of HVR-1 and NS5B. The genetic distance between NS5B sequences was 7.9% and HVR-1 sequences was 25.3% and thus branches were shorter on the tree constructed using NS5B sequences. However, trees derived from sequences obtained at the same time points showed similar results on both trees. Bootstrap scores >70% are indicated for the relevant branches. SA4a shows branches derived from a patient with 2 genotypes at baseline (genotype 1a in red; genotype 4d in blue). SA4b shows samples taken at baseline (filled circles) and 3 months later (open circles).

*PCR error rates*: Error rates due to incorporation of nucleotide changes by Taq polymerase were calculated by subcloning (x100) of a single colony following end-point dilution PCR.

*Cloning error*: In order to investigate whether new viral strains observed in individual patients were a result of superinfection or recrudescence of minority variant strains, PCR primers were designed to specifically identify the introduced strains and were used in a third semi-nested PCR step to amplify the cDNA obtained from the first positive (baseline) sample. 11 patients had sequences sufficiently divergent to design primers to detect the new strain exclusively (Supplementary Figure 2).

**Supplementary Figure 2: Strain-specific PCR a) Strain-specific primer sequences b) Alignment of old and new strains with strain-specific primer sequence for subject 13 c) PCR gel**

1. **Strain-specific primer sequences**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Primer name** | **Patient number** | **Primers** |  |  |
| ET26643 | 2 | ATGCTTAGTCCGGGCGCC |
| ET4485 | 3 | GGCCGCGTTTGGTCTC |
| ET7609 | 6 | CGGGCCGCGTTTGCTC |
| ET22503 | 8 | CTGCGAATAGCCTTGCC |
| ET1117 | 12 | CTTCACCTCACTGTTCAA |
| ET5322\_4 | 13 | GACCACGTCTGCTGTCG |
| ET5322\_1 | 13 | TGGTGCCAATGCTATAACTG |
| ET55108 | 18 | CGCCCGCACTGTGAATAG |
| ET25196 | 21 | AAGTGCTGCTAAGACCACGTC |
| ET55427 | 27 | CTGAGACCATGTCTGGTCTCG |
| ET8438 | 28 | GCCGGGCTTGTCGATCTC |
| ET21754 | 29 | GGCGCGTTGGGAATTAG |
| ET5645 | 31 | GGTGCCAATGCTATAACTGG |
| ET21860 | 37 | GCTATTGCCAAATCCGCATTC |
| ET27758 | 39 | AAGTGCTGCTCGGACCATG |
| ET2387 | 44 | CCGCTGGATTTGCCAGCTTCAG |

1. **Alignment of old and new strains with strain-specific primer sequence (patient 13)**

Primer specific to sample 1 strains

Sample 1 sequences

Sample 2 sequences



1. **Amplified strain-specific PCR products run on a 2% agarose gel (patients 13 and 37)**



Negative 6/1/4 14/8/3 6/3/6 13/11/6 control Patient 13 Patient 37

100 bp product

200 bp ladder

**Supplementary Figure 2:** Strain-specific primers were designed to amplify specific strains and were used to amplify PCR product obtained by HVR-1 PCR in a third semi-nested PCR cycle. a) Primer sequences used for the PCR b) Alignment of strain-specific primer for patient 13 with baseline and subsequent sample c) 2% agarose gels showing amplified product from subjects 13 and 37 at 2 time points. Strain-specific product is present as expected at the second time point in subject 13, but not at the first, while strain-specific product is detected at both time points in subject 37

Of the 11 patients tested, sequences characteristic of the new majority strain were identified in baseline samples in 4 patients (36%) by strain-specific PCR. Thus 64% of patients demonstrating evidence of new strains were likely to have been superinfected. The experiment was repeated to exclude contamination and the products were gel extracted and sequenced in order to confirm that non-specific binding had not occurred.

**Appendix 3: Multiple strains of HCV are circulating in the HIV-positive MSM population**

**Appendix 3**: A neighbour-joined phylogenetic tree was constructed (using MEGA 4.0 and Figtree; Kimura 2-parameter method) with patient majority quasispecies sequences (n=97; coloured lines) and reference sequences (n=212; black lines) obtained from the Los Alamos database. This revealed several monophyletic clusters of infection suggesting several introductions into the MSM population. The majority of infections were genotype 1a but genotypes 1b, 3a and 4d were also identified.

**Appendix 4: Genetic diversity (uncorrected) over time**



**Appendix 4:** Cumulative genetic distance was calculated for each group of patients at multiple time points (range 0-450 days) showing marked non-linear variation over time.

**Appendix 5: T cell responses and VL in individual patients**

**Appendix 5:** T cells responses (IFNγ release; red lines) were assessed by ELISpot in representative individual patients and plotted against log HCV viral load (blue lines).