Pathogen and population variability
How to hit them all in one blow
HCV

http://farm1.static.flickr.com/226/502166262_7d04df3987_o.jpg
The disease

- More than 4 million Americans (1.3% of the U.S. population) and 170 million individuals in the world (3% worldwide) are infected with hepatitis C virus.

- The prevalence of hepatitis C virus in Scandinavia is less than 0.5% of the population, whereas the prevalence in Egypt is over 20%.

- 85% of individuals initially infected with this virus will become chronically infected.

- Chronic hepatitis C virus infection causes an inflammation of the liver called chronic hepatitis.

- This condition can progress to scarring of the liver, called fibrosis, or more advanced scarring, called cirrhosis.

- Some patients with cirrhosis will go on to develop liver failure or the complications of cirrhosis, including liver cancer.
The genome

Hepatitis C virus (HCV): model structure and genome organisation

Expert Reviews in Molecular Medicine ©2003 Cambridge University Press
HCV is highly variable

Fig. 3. Evolutionary tree of all known subtypes and genotypes of HCV, including those found in high-diversity areas of genotypes 1, 2 and 4 (sub-Saharan Africa) and 3 and 6 (South-East Asia). HCV variants still fall into six distinct clades, but with far greater numbers of genetic variants corresponding to subtypes in industrialized countries. The tree was constructed by using the neighbour-joining method as implemented in the MEGA package, using Jukes–Cantor-corrected distances between partial NS5B sequences (320 bp).
**HCV genotypes**

Fig. 2. Evolutionary tree of the principal genotypes of HCV that are found in industrialized countries and their main epidemiological associations with specific risk groups. These genotypes of HCV are believed to have become prevalent over the course of the 20th century. The tree was constructed by using the neighbour-joining method as implemented in the MEGA package, using Jukes–Cantor-corrected distances between complete genome sequences.
HCV Genotype distribution

1a & 1b: 38% (1b)
1a: 36%
1 & 2: 2%
2a: 6%
2b: 9%
3: 6%
4: 1%

©Hepatitis Central (TM)
http://hepatitis-central.com
Genotype Variation
Genotype variation

HIV-1 CRF02_AG (a), HCV genotype 4 (b) and HCV genotype 1 (c)

de Oliveira et al., Nature 444, 836-837 (14 December 2006)
Epitope selection for maximal strain coverage

- Use high performing Epitope predictions to predict potential epitopes for all supertypes
- Select top scoring peptides
- Use EpiSelect tool for maximal strain coverage
EpiSelect - GenoCover

**Top Scoring Peptides**
- Genotype 1
- Genotype 2
- Genotype 3
- Genotype 4
- Genotype 5
- Genotype 6

Select peptide with maximal coverage

Select peptide with maximal coverage preferring uncovered strains

Select peptide with maximal coverage preferring lowest covered strains

*Repeat until the desired number of peptides is selected*
HCV Results - B7

Of the 5 selected B7 peptides, 2 have been verified CTL epitopes and are included in the Los Alamos database.
Population Diversity

http://static.howstuffworks.com/gif/population-six-billion-1.jpg
MHCCover

Select peptide with maximal MHC coverage

Select peptide with maximal MHC coverage preferring uncovered MHCs

Select peptide with maximal MHC coverage preferring lowest covered HLAs

Repeat until the desired number of peptides is selected

Top Scoring Peptides

- HLA-A*0101
- HLA-A*0201
- HLA-A*0301
- HLA-B*0702
- HLA-B*2705
- HLA-B*4402
Population diversity

Allele Frequency

\[ X = \frac{5}{12} = 0.42 \]
\[ Y = \frac{4}{12} = 0.33 \]
\[ Z = \frac{3}{12} = 0.25 \]
Allele coverage

\[ \begin{align*}
X &= \frac{4}{6} = 0.66 \\
Y &= \frac{3}{6} = 0.5 \\
Z &= \frac{3}{6} = 0.5
\end{align*} \]
Frequency to coverage

- Assumption:
  - Alleles are independently distributed

\[ X_i = q_i + (1-q_i) \times q_i \]

\[ X_i = 2q_i - q_i^2 \]
Coverage of peptide binding to multiple alleles

- Assumption:
  - Independent distribution

\[ C_{jk} = X_j + (1-X_j) \times X_k \]

\[ C_{jk} = X_j + X_k - X_jX_k \]
• HLA-A*0101 has a frequency of xx in the European population

• HLA-A*0301 has a frequency of yy in the European population

• What are the population coverage of each of the two alleles?

• What are the total coverage of the two alleles?
MHC-PopCover
Top Scoring Peptides

Select peptide with maximal population coverage

HLA-A*0101
HLA-A*0201
HLA-A*0301
HLA-B*0702
HLA-B*2705
HLA-B*4402

Select peptide with maximal coverage preferring uncovered MHCs with highest population coverage

Select peptide with maximal coverage preferring lowest covered HLAs with highest population coverage

Repeat until the desired number of peptides is selected
EpiSelect

Select peptide with maximal population coverage and maximal genotype coverage

Select peptide with maximal coverage preferring uncovered MHCs with highest population coverage and maximal genotype coverage

Select peptide with maximal coverage preferring lowest covered HLAs with highest population coverage and maximal genotype coverage

Repeat until the desired number of peptides is selected

Genotype 1
- HLA-A*0101

Genotype 2
- HLA-A*0201

Genotype 3
- HLA-A*0301

Genotype 4
- HLA-B*0702

Genotype 5
- HLA-B*2705

Genotype 6
- HLA-B*4402
Reaching optimal coverage

### Table

<table>
<thead>
<tr>
<th>Peptide</th>
<th>1b.fsa.pep.preds</th>
<th>1a.fsa.pep.preds</th>
<th>2.fsa.pep.preds</th>
<th>3.fsa.pep.preds</th>
<th>4.fsa.pep.preds</th>
<th>5.fsa.pep.preds</th>
<th>6.fsa.pep.preds</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLHGFTPLL</td>
<td>47.00</td>
<td>47.00</td>
<td>47.00</td>
<td>47.00</td>
<td>47.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>ETTYRLRAY</td>
<td>47.00</td>
<td>49.00</td>
<td>50.00</td>
<td>49.00</td>
<td>47.00</td>
<td>52.00</td>
<td>5.00</td>
</tr>
<tr>
<td>IMTCSADL</td>
<td>47.00</td>
<td>49.00</td>
<td>50.00</td>
<td>49.00</td>
<td>47.00</td>
<td>5.00</td>
<td>49.00</td>
</tr>
<tr>
<td>LLRHHNLVY</td>
<td>47.00</td>
<td>50.00</td>
<td>50.00</td>
<td>47.00</td>
<td>47.00</td>
<td>34.00</td>
<td>49.00</td>
</tr>
<tr>
<td>WAKHMWNFI</td>
<td>58.00</td>
<td>73.00</td>
<td>26.00</td>
<td>65.00</td>
<td>47.00</td>
<td>34.00</td>
<td>49.00</td>
</tr>
<tr>
<td>GIYLFWAV</td>
<td>58.00</td>
<td>73.00</td>
<td>26.00</td>
<td>65.00</td>
<td>47.00</td>
<td>34.00</td>
<td>49.00</td>
</tr>
<tr>
<td>ATLCASALYV</td>
<td>58.00</td>
<td>73.00</td>
<td>60.00</td>
<td>65.00</td>
<td>47.00</td>
<td>65.00</td>
<td>49.00</td>
</tr>
<tr>
<td>SSIVSWAVK</td>
<td>58.00</td>
<td>73.00</td>
<td>60.00</td>
<td>65.00</td>
<td>47.00</td>
<td>65.00</td>
<td>62.00</td>
</tr>
<tr>
<td>RAWRHRARA</td>
<td>58.00</td>
<td>73.00</td>
<td>60.00</td>
<td>79.00</td>
<td>65.00</td>
<td>65.00</td>
<td>78.00</td>
</tr>
<tr>
<td>GSRMITGPK</td>
<td>58.00</td>
<td>73.00</td>
<td>71.00</td>
<td>79.00</td>
<td>76.00</td>
<td>79.00</td>
<td>78.00</td>
</tr>
<tr>
<td>LLADARVCA</td>
<td>58.00</td>
<td>73.00</td>
<td>71.00</td>
<td>79.00</td>
<td>76.00</td>
<td>79.00</td>
<td>78.00</td>
</tr>
<tr>
<td>LSDFKTWLK</td>
<td>58.00</td>
<td>81.00</td>
<td>71.00</td>
<td>79.00</td>
<td>76.00</td>
<td>79.00</td>
<td>84.00</td>
</tr>
<tr>
<td>NLRAFTEAM</td>
<td>58.00</td>
<td>85.00</td>
<td>79.00</td>
<td>79.00</td>
<td>76.00</td>
<td>79.00</td>
<td>84.00</td>
</tr>
<tr>
<td>LSSWFVAGY</td>
<td>74.00</td>
<td>85.00</td>
<td>79.00</td>
<td>79.00</td>
<td>76.00</td>
<td>79.00</td>
<td>84.00</td>
</tr>
<tr>
<td>LSAFTLHGY</td>
<td>74.00</td>
<td>85.00</td>
<td>79.00</td>
<td>79.00</td>
<td>76.00</td>
<td>79.00</td>
<td>84.00</td>
</tr>
<tr>
<td>RLYCGGPME</td>
<td>74.00</td>
<td>85.00</td>
<td>79.00</td>
<td>79.00</td>
<td>76.00</td>
<td>79.00</td>
<td>85.00</td>
</tr>
<tr>
<td>FTDNSSTPA</td>
<td>74.00</td>
<td>85.00</td>
<td>79.00</td>
<td>79.00</td>
<td>76.00</td>
<td>79.00</td>
<td>85.00</td>
</tr>
<tr>
<td>VTQHGSYHY</td>
<td>74.00</td>
<td>85.00</td>
<td>85.00</td>
<td>79.00</td>
<td>85.00</td>
<td>79.00</td>
<td>85.00</td>
</tr>
<tr>
<td>LSAFTLHYS</td>
<td>74.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
</tr>
<tr>
<td>LRKLGAPPL</td>
<td>74.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>86.00</td>
</tr>
<tr>
<td>YYYDHTLPL</td>
<td>81.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>86.00</td>
</tr>
<tr>
<td>VLLTLSHPY</td>
<td>86.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>86.00</td>
</tr>
<tr>
<td>VLWTVVYGA</td>
<td>86.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>86.00</td>
</tr>
<tr>
<td>SYTWGTALI</td>
<td>87.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>85.00</td>
<td>87.00</td>
<td>88.00</td>
</tr>
<tr>
<td>NEITLTHPI</td>
<td>87.00</td>
<td>85.00</td>
<td>86.00</td>
<td>85.00</td>
<td>85.00</td>
<td>90.00</td>
<td>88.00</td>
</tr>
</tbody>
</table>

Color scale
85.00
75.62
66.25
56.88
47.50
38.12
28.75
19.38
10.00

HCV Genotypes