Immunoinformatics comes of age

Can Keşmir
Theoretical Biology/Bioinformatics
Utrecht University, NL &
CBS, Technical University of Denmark

c.kesmir@bio.uu.nl
Definition: The research field that applies informatics techniques to generate a systems level view of the immune system.

Related field: Theoretical Immunology

Methods and data are different but the aim is the same.
How does our immune system detect infections?
Where is Immunoinformatics now?

Start/Heart of the immune response!
Antigen Processing and Presentation: CBS tools

1. NetChop-20S
2. TAP
3. NetMHC
4. NetChop-Cterm
Are epitope databases complete?
Ingrid Schellens (UMCU)

Let's take HIV and intensively studied MHC molecules, can we still find novel epitopes?
Immunoinformatics comes of age

Known epitopes

Novel epitopes

Peptide

#SFU/million PBMC
More than 400 microbial agents are associated with disease.

However, vaccines for only 34 pathogens are developed!

While deaths due to infectious and parasitic diseases make up 19% of total deaths in the world.
Given reliable predictions, we can now study the immune system at the system-level!

A highly dynamic system because of interactions with pathogens!
Co-evolution of host and pathogen: HIV as a model

- Does HIV adapt to its new human host?
- How did the natural host evolve to deal with SIV?
Specificity of the molecules in Class I pathway

- Proteasome: 30% monomorphistic
- TAP: 70% monomorphistic
- MHC: 1-5% polymorphistic
- TCR: 0.001% highly diverse
Escapes from MHC and T cell

Escapes from MHC and T cells are not beneficial for the virus after transmission thanks to polymorphism!
Are monomorphic molecules (TAP + Proteasome) easy targets of escape during HIV infection?
Adaptation on single host level

- Two full HIV genome sequences (1986 and 1997) of a long term survivor
- Total of 66 non-silent mutations
- 62 mutations can be associated with CTL responses given the patient MHCs

Escape from strong epitopes: 13
MHC escapes: 10
Processing escapes: 5
Adaptation on population level
Why HIV is not evading processing?
Shadowing due to MHC polymorphism
Conclusions: Is HIV evading Processing?

• At the population level processing escape mutants do not get fixed because of the shadowing by MHC polymorphism

• The degeneracy of proteasome and TAP might make it difficult for HIV to evade antigen processing (we need to analyse more data to be able to say this).
Natural host response to SIV S. mangabeys

- Prevalence up to 60% or more, increasing with age
- Normal CD4 T cell counts, no signs of hyper-activation
- High viral loads, however hardly any AIDS
- SIVsm is not ignored by the host, but a mild T cell response is generated

Asian macaques infected with modified SIVsm develop AIDS
Chimpanzee and SIV

- Up to 60% infected with SIVcpz ---> no AIDS
- 200 chimpanzees so far infected experimentally with HIV-1 --> only 1 case of AIDS.
- Low viral loads --> Efficient T cell response???
- MHC diversity is much lower compared to human. Especially protective MHC-B has lower diversity.
  ➢ Has been through a severe selection (bottleneck)
- Common MHC molecules in the chimpanzee population target the same epitopes as human long-term non-progressors

Can we find more signs of adaptation/selection?
Comparative genomics

AIDS
SIVmac

AIDS
SIVcpz

AIDS
HIV-1/2
Comparative analysis

- Lineage specific genes (~500) → difficult to predict the function
- Look for orthologs in macaques and human that are not existing chimpanzee
- PanTro2.1 (March 2006) & Ensemble automatic orthology detection (manual check afterwards)
- There are 312 ortholog groups that are deleted in chimpanzee
- 153 of these have GO annotations
- 12 genes are related to immune response and thus our candidates for possible host adaptation
Almost for sure missing in chimpanzee

- ICEBERG: inhibitor of caspase-1.
- IL1F7 & ILF18: members of IL1 super family
Some genes are very diverged!

<table>
<thead>
<tr>
<th>Location (human)</th>
<th>Cluster</th>
<th>Median $K_A/K_I^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q21</td>
<td>Epidermal differentiation complex</td>
<td>1.46</td>
</tr>
<tr>
<td>6p22</td>
<td>Olfactory receptors and HLA-A</td>
<td>0.96</td>
</tr>
<tr>
<td>20p11</td>
<td>Cystatins</td>
<td>0.94</td>
</tr>
<tr>
<td>19q13</td>
<td>Pregnancy-specific glycoproteins</td>
<td>0.94</td>
</tr>
<tr>
<td>17q21</td>
<td>Hair keratins and keratin-associated proteins</td>
<td>0.93</td>
</tr>
<tr>
<td>19q13</td>
<td>CD33-related Siglecs</td>
<td>0.90</td>
</tr>
<tr>
<td>20q13</td>
<td>WAP domain protease inhibitors</td>
<td>0.90</td>
</tr>
<tr>
<td>22q11</td>
<td>Immunoglobulin-γ/breakpoint critical region</td>
<td>0.85</td>
</tr>
<tr>
<td>12p13</td>
<td>Taste receptors, type 2</td>
<td>0.81</td>
</tr>
<tr>
<td>17q12</td>
<td>Chemokine (C-C motif) ligands</td>
<td>0.81</td>
</tr>
<tr>
<td>19q13</td>
<td>Leukocyte-associated immunoglobulin-like receptors</td>
<td>0.80</td>
</tr>
<tr>
<td>5q31</td>
<td>Protocadherin-β</td>
<td>0.77</td>
</tr>
<tr>
<td>1q32</td>
<td>Complement component 4-binding proteins</td>
<td>0.76</td>
</tr>
<tr>
<td>21q22</td>
<td>Keratin-associated proteins and uncharacterized ORFs</td>
<td>0.76</td>
</tr>
<tr>
<td>1q23</td>
<td>CD1 antigens</td>
<td>0.72</td>
</tr>
<tr>
<td>4q13</td>
<td>Chemokine (C-X-C motif) ligands</td>
<td>0.70</td>
</tr>
</tbody>
</table>

*Maximum median $K_A/K_I$ if the cluster stretched over more than one window of ten genes.

Chimp Consortium, Nature, 2005
<table>
<thead>
<tr>
<th></th>
<th>Human</th>
<th>Chimp</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR5</td>
<td>MNYQTSTPYYIDYGTEPCQKVNRQIAARLPLLPSLVVFIFGFGVGNVLVLILIDCKK</td>
<td>MDYQVSSPIYIDYTTEPCQKINVQIAARLLPLLPSLVVFIFGFGVGNMLVLILINCKR</td>
</tr>
<tr>
<td></td>
<td>LKSMTDIYLLNLALSDLFLLLTIPFWAHAADQQWTFGKMCQOLLGTGYYIGFTGNNFII</td>
<td>LKSMTDIYLLNLALSDLFLLLTVFPWAHYAADAQQWDGNTCMQOLLGLYFIGFFSGIFFI</td>
</tr>
<tr>
<td></td>
<td>LLTMDRYLAIVHAVSAKRTVGVTGSIGIAWVAVLASFPIIFTRSKKGSRFTCSP</td>
<td>LLTIDRYLAIVHAVFALKARTVGVTGSIVTWWAFAISLPGIIFTRSKGREEIHYTCSS</td>
</tr>
<tr>
<td></td>
<td>HFPPSQHHFWKNNFQALKMSVGLILPLLVMIIYGSAIOKLTLRRCNKRHKRAERLIVF</td>
<td>HFPYSQYQFWKNNFQTLKIVILGLVLPLLVMICYSGilKLTLRRCNKRHKRAVRLIFTI</td>
</tr>
<tr>
<td></td>
<td>MIVYFLFWAPYNVLUPSTTQEFFGNNCNSSRNLQAMQITETLGMTCCINPIIIYAFV</td>
<td>MIVYFLFWAPNYLNNLTQEFFGNNCSRNLQAMQVTETLGMTCCINPIIIYAFV</td>
</tr>
<tr>
<td></td>
<td>GEKFRRYLSLFFRKHIRRFCKCPIFQGELPDRVSSYTRSTGQEISVAL</td>
<td>GEKFRNYLLVFQHICRCKCQFQQEAPERASSYTRSTGQEISVGL</td>
</tr>
</tbody>
</table>
Conclusions

- Processing do not shape HIV evolution, because:
  1. MHC polymorphism **shadow** the immune selection pressure on processing
  2. The processing is **degenerate** and thus difficult to escape from
- Chimpanzee do not develop AIDS, because they lack/adapted some immuno-regulatory genes?
Take home messages

For new comers:

- Technically: several big challenges
- Immune system and pathogenic world are full of mysteries to discover!

For experts:

- Time to focus on the immune system and less on the tools!
- Bioinformatics is not only supervised methods!
- Let's work more on verification data that will reach immunologists
People

TBB, Utrecht, NL
• Boris Schmid
• Jose' Borghans
• Rob J. de Boer

UMCU, Utrecht, NL
• Ingrid Schellens

LANL, NM, USA
• Karina Yusim
• Bette Korber

CBS, DTU, Denmark
• Anne Mølgaard
• Ole Lund
• Morten Nielsen
• Claus Lundegaard
• Søren Brunak
• Sune Frankild
• Ilka Hoof
• Nicolas Rapin