IMGT/V-QUEST and IMGT/JunctionAnalysis: The standardized approach for IG and TR rearrangement analysis

Xavier Brochet

Laboratoire d’ImmunoGénétique Moléculaire
Université Montpellier, UPR CNRS 1142, IGH

First Immunomics Summer School
24 August - 3 September 2007
IMGT was created in June 1989, during the Human Genome Meeting HGM 10

IMGT® was created to answer the need to manage the immunoglobulin (IG) and T cell receptor (TR) related data.
IMGT/V-QUEST and IMGT/JunctionAnalysis, part of IMGT®, the international ImMunoGeneTics information system®, analyse the IG and TR rearranged V-D-J and V-J sequences.

1) identify the V, (D) and J genes and alleles involved in the rearrangements.
2) provide a detailed and accurate description of the sequences.

Results are according to the IMGT Scientific chart rules, based on the IMGT-ONTOLOGY axioms.
IMGT/V-QUEST and IMGT/JunctionAnalysis

**IMGT/V-QUEST**
- compares the user’s sequences with the IMGT reference sequences (IMGT reference directory: CLASSIFICATION axiom)
- displays the nucleotide and protein alignments according to the IMGT Scientific Chart rules (IMGT labels: DESCRIPTION axiom)
- provides an extensive analysis of the mutations (IMGT unique numbering: NUMEROTATION axiom)

**IMGT/JunctionAnalysis**
- analyses accurately the junctions of IG and TR rearranged sequences: identifies IGHD, TRBD and TRDD genes and alleles, N- and P REGION, "gc“ content, amino acid physicochemical properties, PI, etc.

Specificities of IMGT/V-QUEST algorithm

IMGT unique numbering rules

- maintain conserved AA, and therefore codons, at the same positions
- standardize the delimitations of FR-IMGT and CDR-IMGT

Gaps are inserted in IMGT reference sequences, taking into account the variable lengths of the CDR-IMGT

IMGT proposes for the first time a standardized description of the V-DOMAIN, whatever the antigen receptor, the chain type and the species.
IMGT/V-QUEST algorithm

In order to respect the IMGT unique numbering rules:

➡ The IMGT/V-QUEST algorithm is based on a global alignment which does not accept any insertion and/or deletion

➡ Gaps are integrated into the user sequence, according to the IMGT unique numbering, by comparison with a model sequence of the IMGT reference directory.
WELCOME!
to the IMGT/V-QUEST Search page

THE INTERNATIONAL IMMUNOGENETICS INFORMATION SYSTEM®


You are in the new IMGT/V-QUEST, upgraded for multiple sequences and with new functionalities.

Analyze your Immunoglobulin nucleotide sequences

- Human
- Mouse
- Rat (test for IGHV2 and IGHV5 subgroups)
- Chondrichthyes
- Teleostei
  - Atlantic cod
  - Channel catfish
  - Rainbow trout
- Sheep

Analyze your T cell Receptor nucleotide sequences

- Human
- Mouse
- Non-human primates
Analysis of several sequences (in FASTA format) in a single run

Selection of output parameters

Detailed view for each sequence

Synthetic view per allele
1. "Detailed view" results: Summary

Sequence number 1: AF184762

Sequence compared with the human IG set from the IMGT reference directory

```
>AF184762
atggagtttggctagctgggttttcctgtgcttatattttttataggtgtccaccttgaag
agtgcagctgttgagtcgaggctgttagctcacgcttgggaagcatcttccctgctcttgcaagtctttgtccctcgctacctctcatgtgctcgaggctgctcttgacagtctcttgcagcgacacagcaggtttttaaaccgcagagtgggagggaacccttgctaccgcctttctctcaagcatccgaccagcctagctggcctctgcagccagccagat
```

Automatic interpretation

<table>
<thead>
<tr>
<th>Result summary:</th>
<th>Productive IGH rearranged sequence (no stop codon and in frame junction)</th>
</tr>
</thead>
<tbody>
<tr>
<td>V-GENE and allele</td>
<td>IGHV3-73*01</td>
</tr>
<tr>
<td>J-GENE and allele</td>
<td>IGHJ1*01 (a)</td>
</tr>
<tr>
<td>D-GENE and allele by IMGT/JunctionAnalysis</td>
<td>IGHD3-10*01</td>
</tr>
<tr>
<td>CDR-IMGT lengths and AA JUNCTION</td>
<td>[8,10,10]</td>
</tr>
</tbody>
</table>

(a) Other possibilities: IGHJ4*02 and IGHJ5*02 (highest number of consecutive identical nucleotides)
2. "Detailed view" results: V,D and J alignments

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

<table>
<thead>
<tr>
<th>V-GENE</th>
<th>Score</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z22768</td>
<td>1240</td>
<td>91.55%</td>
</tr>
<tr>
<td>AB011937</td>
<td>1231</td>
<td>91.18%</td>
</tr>
<tr>
<td>X92206</td>
<td>1024</td>
<td>63.33%</td>
</tr>
<tr>
<td>X92216</td>
<td>979</td>
<td>81.63%</td>
</tr>
<tr>
<td>M96406</td>
<td>970</td>
<td>81.63%</td>
</tr>
</tbody>
</table>

Alignment with FR-IMGT and CDR-IMGT delimitations

AF184752
Z22768
AB011937
X92206
X92216
M96406

Score and nucleotide identity

2. Alignment for D-GENE and allele identification

Closest D-REGIONS

<table>
<thead>
<tr>
<th>D-GENE</th>
<th>Score</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>X13072</td>
<td>28</td>
<td>72.83%</td>
</tr>
<tr>
<td>X93615</td>
<td>19</td>
<td>63.64%</td>
</tr>
<tr>
<td>J00226</td>
<td>14</td>
<td>60.00%</td>
</tr>
<tr>
<td>X97051</td>
<td>13</td>
<td>62.50%</td>
</tr>
<tr>
<td>X93614</td>
<td>10</td>
<td>54.55%</td>
</tr>
</tbody>
</table>

Alignment

AF184752
gtgcgtcgggagtctcgtacgtgaagccgacccgctggcagggcctgtggctacacgtgcttcgcttccc
X13072
X93615
J00226
X97051
X93614

3. Alignment for J-GENE and allele identification

Closest J-REGIONS

<table>
<thead>
<tr>
<th>J-GENE</th>
<th>Score</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>J00226</td>
<td>179</td>
<td>92.69%</td>
</tr>
<tr>
<td>X86355</td>
<td>150</td>
<td>75.17%</td>
</tr>
<tr>
<td>J00226</td>
<td>141</td>
<td>77.01%</td>
</tr>
<tr>
<td>X86355</td>
<td>138</td>
<td>74.51%</td>
</tr>
<tr>
<td>M25625</td>
<td>175</td>
<td>75.00%</td>
</tr>
</tbody>
</table>

Note that the highest number of consecutive identical nucleotides has been found in the alignment with IG

Alignment

AF184752
gtgcgtcgggagtctcgtacgtgaagccgacccgctggcagggcctgtggctacacgtgcttcgcttccc
J00226
X86355
J00226
X86355
M25625

The international ImMunoGeneTics information system®
Founder and director: M.-P. Lefranc, Montpellier, France
http://imgt.cines.fr
3. “Detailed view” results: IMGT/JunctionAnalysis

4. Results of IMGT/JunctionAnalysis

Maximum number of accepted mutations in: 3′V-REGION = 2, D-REGION = 4, 5′-REGION = 2

Analysis of the JUNCTION

D-REGION is in reading frame 3.

Click on mutated (underlined) nucleotide to see the original one:

<table>
<thead>
<tr>
<th>Input</th>
<th>V name</th>
<th>3′V-REGION</th>
<th>H1</th>
<th>D-REGION</th>
<th>H2</th>
<th>5′-REGION</th>
<th>J name</th>
<th>D name</th>
<th>Vmut</th>
<th>Dmut</th>
<th>Jmut</th>
<th>Hgc</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF184762</td>
<td>IGHV3-73*01</td>
<td>tgt.....</td>
<td>g</td>
<td>...........tgatccgaggagatgtt....</td>
<td>tcaacgga</td>
<td>.............cagttg</td>
<td>IGHD1*01</td>
<td>IGHD3-10*01</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>5/10</td>
</tr>
</tbody>
</table>

Translation of the JUNCTION

Click on mutated (underlined) amino acid to see the original one:

<table>
<thead>
<tr>
<th>Frame</th>
<th>CDR3-IMGT length</th>
<th>Molecular mass</th>
<th>pI</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF184762</td>
<td>tgt gtt atc cgg gga gat gtt tac aac cga cag tgg</td>
<td>+</td>
<td>10</td>
</tr>
</tbody>
</table>

Be aware that some allele reference sequences may be incomplete or from cDNAs. In those cases, IMGT/JunctionAnalysis uses automatically the allele *01 for the analysis of the JUNCTION.

5. Sequence of the JUNCTION ('nt' and 'AA')

104 105 106 107 108 109 113 114 115 116 117 118
CVIRGDNRYNRQWntgtgtcatcgggagatgtttacacgcacagtgg

The international ImMunoGeneTics information system®
Founder and director: M.-P. Lefranc, Montpellier, France    http://imgt.cines.fr
### V-REGION mutation table

<table>
<thead>
<tr>
<th>FR1-IMGT</th>
<th>CDR1-IMGT</th>
<th>FR2-IMGT</th>
<th>CDR2-IMGT</th>
<th>FR3-IMGT</th>
<th>CDR3-IMGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>g36A-a</td>
<td>188&gt;C, F30&gt;D (+ -)</td>
<td>a115&gt;D, M30&gt;V (+ -)</td>
<td>t168&gt;C</td>
<td>g211&gt;L, V71&gt;M (+ -)</td>
<td>a313&gt;G, T105&gt;V (+ -)</td>
</tr>
<tr>
<td>g51A-a</td>
<td>193&gt;C</td>
<td>t125&gt;C</td>
<td>g170&gt;D, P75&gt;R (+ ++)</td>
<td>t215&gt;G, K72&gt;R (+ +++)</td>
<td>c170&gt;G, S58&gt;D (+ -)</td>
</tr>
<tr>
<td>g100&gt;A, A346&gt;N (+ -)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### V-REGION mutation statistics

#### Amino acids

<table>
<thead>
<tr>
<th>IMGT labels</th>
<th>V-REGION</th>
<th>FR1-IMGT</th>
<th>CDR1-IMGT</th>
<th>FR2-IMGT</th>
<th>CDR2-IMGT</th>
<th>FR3-IMGT</th>
<th>CDR3-IMGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total amino acids (AA) with gaps</td>
<td>164 (106)</td>
<td>26</td>
<td>12</td>
<td>17</td>
<td>10</td>
<td>39</td>
<td>0 (2)</td>
</tr>
<tr>
<td>Identical AA</td>
<td>89</td>
<td>26</td>
<td>10</td>
<td>16</td>
<td>5</td>
<td>32</td>
<td>0</td>
</tr>
</tbody>
</table>

#### AA changes

- **Conserved IMGT AA classes (hydrophathy, volume, chemical)**
  - Total: 15 (17)
  - (++): 3 (4)
  - (+ +): 2 (0)
  - (+ +): 7 (0)
  - (- -): 1 (2)
  - (++): 0 (0)
  - (+ +): 2 (0)

---

The international ImMunoGeneTics information system®
Founder and director: M.-P. Lefranc, Montpellier, France  
http://imgt.cines.fr
Who are the IMGT/V-QUEST users?

1. IG and TR genome diversity and evolution studies

- genome diversity and genome evolution studies of the adaptive immune responses
- veterinary research (analysis of IG and TR repertoires in farm, wild life and model animal species)

Comparison of IG and TR from newly sequenced genomes is done versus complete known loci. For example: non-human primates vs human, rat vs mouse, etc.

2. Biotechnology related to antibody engineering

- single chain Fragment variable (scFv)
- phage displays
- combinatorial libraries
- chimeric, humanized and human antibodies

Humanization of antibodies based on standardized human FR-IMGT and non-human CDR-IMGT allows to decrease immunogenicity and to maintain specificity.
Who are the IMGT/V-QUEST users?

3- Fundamental and medical research

Repertoire analysis of the IG antibody sites and of the TR recognition sites:
- in normal situations
- in pathological situations (autoimmune diseases, infectious diseases, AIDS, leukemias, lymphomas, myelomas)

Diagnostics
- clonalities in leukemias, lymphomas
- detection and follow up of residual diseases

Prognostic factor in Chronic Lymphocyte Leukemia (CLL)
- “unmutated” IGH V-REGION (identity > 98%): bad prognostic
- “mutated” IGH V-REGION (identity < 98%): good prognostic

ERIC, the European Research Initiative on CLL (www.ericll.org) recommends IMGT/V-QUEST for such studies
IMGT, the international ImMunoGeneTics information system®
http://imgt.cines.fr

Acknowledgements:

- IMGT team
- ImmunoGrid partners
- MARIE Network

Many thanks to the First Immunomics Summer School organizers

The IMGT team at Montpellier, France