IMGT-ONTOLOGY and IMGT databases, tools and Web resources for immunoinformatics

Marie-Paule Lefranc
Université Montpellier, CNRS
Institut Universitaire de France

Singapore Immunoinformatics Symposium
March 1, 2004, Institute for Infocomm Research, Singapore
What is an ontology? and Why?

An ontology is the definition of the concepts and of their relations, necessary to share, to reuse and to represent the knowledge, in a domain.
What is an ontology?

An ontology is the definition of the concepts and of their relations, necessary to share, to reuse and to represent the knowledge, in a domain.
What is an ontology?

An ontology is the definition of the concepts...

Example: a gene

- **gene type**: concept of IDENTIFICATION
- **gene name**: concept of CLASSIFICATION
- **gene labels**: concept of DESCRIPTION

Controlled vocabulary
What is an ontology?

An **ontology** is the definition of the concepts and of their relations...

- In many ontologies, no distinction between « concepts » and « instances »
- In GO (GeneOntology), only 2 types of relations: « is a », « is part of »

Gene labels: **V-GENE IGLV2-11**

Chain type: **Ig-Light-Lambda**

Gene type: **V-GENE**

Gene name: **IGLV2-11**
Why an ontology?

An **ontology** is the definition of the concepts and of their relations, necessary to share, to reuse and to represent the knowledge, in a domain.

- Human beings
- Information systems
Why an ontology?

An ontology is the definition of the concepts and of their relations, necessary to share, to reuse and to represent the knowledge, in a domain.
The international ImMunoGeneTics information system®
Coordinator: M.-P. Lefranc, Montpellier, France    http://imgt.cines.fr

IMGT domain of research: the adaptive immune system

Vertebrates

\[
\text{T cell Receptor} \quad \text{MHC} \quad \text{peptide} \quad \text{Immunoglobulin}
\]

Trimolecular complex
Immunoglobulin (IG) and T cell receptor (TR) synthesis

150
FUNCTIONAL IG GENES

HEAVY CHAIN

<table>
<thead>
<tr>
<th>V</th>
<th>D</th>
<th>J</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'</td>
<td>39-46</td>
<td>x 23</td>
<td>x 6</td>
</tr>
</tbody>
</table>

6300 POTENTIAL RECOMBINATIONS

N-DIVERSITY SOMATIC MUTATIONS

ABOUT 6.3 x 10^6 POSSIBILITIES

LIGHT CHAIN

<table>
<thead>
<tr>
<th>V</th>
<th>J</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'</td>
<td>34-37</td>
<td>x 5</td>
</tr>
<tr>
<td>30-33</td>
<td>x 4-5</td>
<td>Lambda</td>
</tr>
</tbody>
</table>

185 +165 POTENTIAL RECOMBINATIONS

ABOUT 3.5 x 10^5 POSSIBILITIES

DIFFERENT ANTIBODIES

2 x 10^{12}

DIFFERENT ANTIBODIES
IMGT-ONTOLOGY five main concepts

to share, reuse and represent knowledge in immunogenetics
"CLASSIFICATION" concept

- **group**
  - **is a member of**
  - an instance of
    - **subgroup**
      - **is a member of**
      - an instance of
        - **gene**
          - **is a variant of**
          - an instance of
            - **allele**

- **locus**
  - **is ordered in**
  - an instance of
    - **IGLV**
      - **is a member of**
      - an instance of
        - **human IGL**
          - (22q11.2)

- **IGLV2**
  - **is ordered in**
  - an instance of
    - **IGLV2-11**
      - **is a member of**
      - an instance of
        - **IGLV2-11*02**

- **IGLV**
  - **is a member of**
  - an instance of
    - **human IGL**
      - (22q11.2)

« Concepts »

« Instances »
WELCOME!

to IMGT/Gene-DB

THE INTERNATIONAL IMMUNOGENETICS INFORMATION SYSTEM®

http://imgt.cines.fr
IGLV2-11 index:

Homo sapiens Official Gene Symbol and Name (HGNC)

IGLV2-11: immunoglobulin lambda variable 2-11

LocusID: 28815

- Locus Type: gene, segment
- Alternate Symbols: V1-3. IGLV211

Map Information:
- Chromosome: 22
- Cytogenetic: 22q11.2

NCBI Reference Sequences [RefSeq]
- Category: REVIEWED
  - Genomic: NG_000002

Category: NCBI Genome Annotation
- Genomic Contig: NT_011520

Related Sequences:
- Nucleotide Type Protein
  - D66993
  - Z73365

Additional Links:
- IMGT Repertoire for individual human immunoglobulin and T cell receptor genes
"DESCRIPTION" concept

Label 1 | Label 2 | Relations entre Labels
---|---|---
V-GENE | V-EXON |
FR3-IMGT | CDR3-IMGT |
L-PART1 | DONOR-SPLICE |
V-REGION | FR1-IMGT |
V-REGION | CDR3-IMGT |
IMGT/LIGM-DB ON LINE, HERE YOU ARE!

Five types of search are available: select one by clicking on the button

Catalogue

accession number, mnemonic, definition, creation date, length, annotation level

IMGT/IGMD-DB Consultation module v3 - Netscape
http://imgt.cines.fr

V-GENE
V-REGION
FR1-IMG
1st-CYS
CDR1-IMG
FR2-IMG
CONSERVED-TRP
CDR2-IMG
FR3-IMG
2nd-CYS
CDR3-IMG
XX

Sequence 297 BP; 60 A; 93 C; 71 G; 73 T; 0 other;
cagctgacc tcagcagc tcggggtgc tcgagacgtc agtacacagt 60
tcggagcgc cacccacac tagatgttct gctataact atctctcctc gttcaaacag 120
Collier de Perles: HUMAN IGHV V-DOMAIN from B12 (PDB: 1hzh_H)

[8.8.20]
V-DOMAIN 3D representation (TR A6, 1ao7)

WELCOME!

to IMGT/3Dstructure-DB

THE INTERNATIONAL IMMUNOGENETICS INFORMATION SYSTEM®

http://imgt.cines.fr
IMGT-ML schema

IMGT Ontology
- Identification
- Classification
- Obtention
- Description
- Numerotation

IMGT Data
- Knowledge
- SeqData
- External schema

http://imgt.cines.fr
IMGT-ML architecture

IMGT-ONTOLOGY

XML schema

Information system modelling

Data distribution format

Controlled vocabulary

Documentation

Biological expertise
Informatic answers to the biological problems

- Use IMGT-ONTOLOGY (and IMGT-ML)
- Allow IMGT components to dynamically interact

- The Web Services

Diagram:
- IMGT/LIGM-DB
- IMGT/V-QUEST
- Web Service IMGT/LIGM-DB
- Web Service IMGT/V-QUEST
- Sequences
- Gene names + ref. seq.

http://imgt.cines.fr
Example of IMGT/V-QUEST results

### Alignment for V-GENE

<table>
<thead>
<tr>
<th>Accession</th>
<th>Gene</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF402940</td>
<td>IGHV1-3*01</td>
<td>GTGCAGCTGCTCGACAGCTCTGGGGCT</td>
</tr>
<tr>
<td>X62109</td>
<td>IGHV1-3*02</td>
<td>CA.GTC.A...T.T........................AG.G............G..................</td>
</tr>
<tr>
<td>X62107</td>
<td>IGHV1-5*02</td>
<td>CA.GTT.A....G.T........................AG.G............G..................</td>
</tr>
<tr>
<td>M99637</td>
<td>IGHV1-8*01</td>
<td>CA.GT.A....G.T........................AG.G............G...........C......</td>
</tr>
<tr>
<td>L06612</td>
<td>IGHV1-46*03</td>
<td>CA.GT.A....G.T........................AG.G............G..................</td>
</tr>
<tr>
<td>X92343</td>
<td>IGHV1-46*01</td>
<td>CA.GT.A....G.T........................AG.G............G..................</td>
</tr>
</tbody>
</table>

### Alignment for J-GENE

<table>
<thead>
<tr>
<th>Accession</th>
<th>Gene</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF402940</td>
<td>IGHSJ3*01</td>
<td>CTTACGCGGCGGGCTTTTGAGCTCTGGGGGCAAGGACCACGCTCACGCTCTCTCA</td>
</tr>
<tr>
<td>J00256</td>
<td>IGHSJ6*01</td>
<td>T.A.TACTACTACT...G.A........................A.T...............T...G</td>
</tr>
<tr>
<td>X86355</td>
<td>IGHSJ6*02</td>
<td>T.A.TACTACTACT...G.A........................A.T...............T...G</td>
</tr>
<tr>
<td>X86355</td>
<td>IGHSJ3*02</td>
<td>T.A.TACTACTACT...G.A........................A.T...............T...G</td>
</tr>
</tbody>
</table>

---

V-GENE - JUNCTION - J-GENE
Diagram of collaboration: Analyse de repertoires

Web Service
IMGT/LIGM-DB

DYNAMIC INTERACTIONS

Web Service
IMGT/GENE-DB

DYNAMIC INTERACTIONS

Web Service
IMGT/LocusView

Gene localization

Web Service
IMGT/GeneFrequency

Sequences with specificity

Web Service
IMGT/V-QUEST
Example of IMGT/GeneFrequency results

Your Selection:

**Human IGH, IGK and IGL Locus Specificity anti-thyroid peroxidase (TPO)**

For the D and J genes, the number of genes is shown between parentheses when genes names could not be indicated for a click on the zoom for the D and J genes names.
Diagram of collaboration: Analyse des jonctions

Web Service
IMG'T/LIGM-DB

Web Service
IMG'T/V-QUEST

Web Service
IMG'T/GENE-DB

Web Service
IMG'T/JunctionAnalysis

DYNAMIC INTERACTIONS

Gene and allele names

http://imgt.cines.fr
**Example of IMGT/JunctionAnalysis results**

### Analysis of the JUNCTIONs

<table>
<thead>
<tr>
<th>Input</th>
<th>V name</th>
<th>V-REGION</th>
<th>D-REGION</th>
<th>N2</th>
<th>J-REGION</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 AF402940 IGHV1-3*01</td>
<td>tgtgcgagag.</td>
<td>..........</td>
<td>gcttcacgggg.........</td>
<td>cgggac</td>
<td>..........</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input</th>
<th>J name</th>
<th>D name</th>
<th>Vmut</th>
<th>Dmut</th>
<th>Jmut</th>
<th>Ngc</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 AF402940 IGHJ3<em>01 IGHD3-10</em>01</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>5/6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Translation of the JUNCTIONs

<table>
<thead>
<tr>
<th>105</th>
<th>107</th>
<th>109</th>
<th>112</th>
<th>114</th>
<th>116</th>
<th>118</th>
<th>CDR3-IMGT</th>
<th>frame</th>
<th>length</th>
</tr>
</thead>
<tbody>
<tr>
<td>104</td>
<td>106</td>
<td>108</td>
<td>110</td>
<td>113</td>
<td>115</td>
<td>117</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>A</th>
<th>R</th>
<th>G</th>
<th>F</th>
<th>T</th>
<th>G</th>
<th>R</th>
<th>D</th>
<th>A</th>
<th>L</th>
<th>D</th>
<th>V</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>tgtgcgagag.</td>
<td>ggc</td>
<td>ttc</td>
<td>acg</td>
<td>ggg</td>
<td>cgg</td>
<td>gac</td>
<td>gct</td>
<td>ttg</td>
<td>gac</td>
<td>gtc</td>
<td>tgg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+</td>
<td></td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Analysis of the JUNCTIONs

<table>
<thead>
<tr>
<th>Input</th>
<th>V name</th>
<th>V-REGION</th>
<th>N1</th>
<th>D-REGION</th>
<th>N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>M62724</td>
<td>IGHV7-4-1*02</td>
<td>tgtgccgagagaaga</td>
<td>.tagcaatgggtacaa....</td>
<td>aata</td>
</tr>
<tr>
<td>#2</td>
<td>Z47269</td>
<td>IGHV1-69*06</td>
<td>tgtgccgagagggggaotaagg</td>
<td>....togaatatgtgggtggttt.........</td>
<td>tcatagggt</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input</th>
<th>J-REGION</th>
<th>J name</th>
<th>D name</th>
<th>Vmut</th>
<th>Dmut</th>
<th>Jmut</th>
<th>Nc</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>M62724</td>
<td>....tttgccagacaag</td>
<td>IGHJ4*02</td>
<td>IGHD5-24*01</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>#2</td>
<td>Z47269</td>
<td>...actgtcgacaccctgg</td>
<td>IGHJ5*02</td>
<td>IGHD3-3*02</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Translation of the JUNCTIONs

<table>
<thead>
<tr>
<th>105</th>
<th>106</th>
<th>107</th>
<th>108</th>
<th>110</th>
<th>111</th>
<th>111.2</th>
<th>112</th>
<th>112.1</th>
<th>112.2</th>
<th>112</th>
<th>113</th>
<th>114</th>
<th>115</th>
<th>116</th>
<th>117</th>
<th>CDR3-IMGT</th>
<th>frame</th>
<th>length</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAREDSNG</td>
<td>VKI</td>
<td>F</td>
<td>DY</td>
<td>W</td>
<td>tac</td>
<td>aaa</td>
<td>ata</td>
<td>ttt</td>
<td>gac</td>
<td>tac</td>
<td>tgtg</td>
<td>+</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1 M62724 tgtgccgagagaatggc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#2 Z47269 tgtgccgagagggggcgtcttttggtggttcatagggt</td>
<td>tac</td>
<td>tgt</td>
<td>ttc</td>
<td>gag</td>
<td>ccc</td>
<td>tgtg</td>
<td>+</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

--> IMGT/JunctionAnalysis Search page
--> IMGT/JunctionAnalysis Documentation
IMGT-Choreography: Expressed IG and TR repertoires

THANK YOU
for using IMGT/JunctionAnalysis

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

http://imgt.cines.fr

Analysis of the JUNCTIONs

<table>
<thead>
<tr>
<th>Input</th>
<th>V name</th>
<th>V-REGION</th>
<th>N</th>
<th>J-REGION</th>
<th>J name</th>
<th>Vmut</th>
<th>Jmut</th>
<th>Ngc</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>AF490920</td>
<td>IGKv1-33*01</td>
<td>tgcgaactatagatgatcttccc...</td>
<td>attcactttc</td>
<td>IGKJ3*01</td>
<td>3</td>
<td>0</td>
<td>0/0</td>
</tr>
<tr>
<td>#2</td>
<td>AF490935</td>
<td>IGKv4-1*01</td>
<td>tgcgaactatataattagatctttc</td>
<td>...tcactttc</td>
<td>IGKJ4*01</td>
<td>0</td>
<td>0</td>
<td>0/0</td>
</tr>
<tr>
<td>#3</td>
<td>AF490937</td>
<td>IGKv4-1*01</td>
<td>tgcgaactatataattatgctctcc</td>
<td>...ctacacttt</td>
<td>IGKJ2*01</td>
<td>2</td>
<td>0</td>
<td>0/0</td>
</tr>
<tr>
<td>#4</td>
<td>AF490932</td>
<td>IGKv3-15*01</td>
<td>tgcgaactataataactgctgcctcc</td>
<td>ccttgacacttt</td>
<td>IGKJ2*01</td>
<td>1</td>
<td>0</td>
<td>2/2</td>
</tr>
</tbody>
</table>

Translation of the JUNCTIONs

<table>
<thead>
<tr>
<th>CDR3-IMGT</th>
<th>105</th>
<th>107</th>
<th>109</th>
<th>113</th>
<th>115</th>
<th>117</th>
<th>frame</th>
<th>length</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>AF490920</td>
<td>tgt caa oac tat gat gat</td>
<td>ttc oca ttc act ttc</td>
<td>+</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#2</td>
<td>AF490935</td>
<td>tgt cag oaa tat tat agt</td>
<td>act cct ctc act ttc</td>
<td>+</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#3</td>
<td>AF490937</td>
<td>tgt cag oaa tat tat agt</td>
<td>ggt cct ceg tac act ttt</td>
<td>+</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#4</td>
<td>AF490932</td>
<td>tgt cag cac tat aat aat tgg cct ccc ctc tac act ttt</td>
<td>+</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IMGT-Choreography: 3D structures/specificities

V-D-J-JUNCTION  V-J-JUNCTION

V-DOMAINS (Mus musculus E5.2 Fv)
"OBTENTION" concept

Origin
- cell, tissue, organ
  - PBL
  - liver

Autoimmune diseases
- autoantibody
- rheumatoid factor

Clonal expansion diseases
- leukemia
- lymphoma
- myeloma

Methodology
- transgenic
  - animal
  - plant
- libraries
  - genomic
  - cDNA
  - combinatorial
- PCR
- hybridoma
  - monoclonal antibody
Immunoinformatics

Data integration specific to Immunology

*interactions host-pathogens
*vaccinology
*immunomodulation...

Gene
Transcript
Protein
Organelle
Cell
Organ
Organism
Population

Microarrays
3D

Collection of clinical data
Gene regulation
Pathways
Networks
Mathematical and computational models

Bioinformatics, databases and tools

http://imgt.cines.fr
Who is using IMGT?

Medical research:
- repertoire in autoimmune diseases, AIDS, leukemias, lymphomas, myelomas, translocations, detection of residual diseases

Therapeutic approaches:
- immunotherapy, grafts, immunomodulation, immunosuppression

Veterinary research:
- IG and TR repertoire of domestic and farm species

Biotechnology related to antibody engineering:
- chimeric, humanized, human antibodies, scFv, combinatorial libraries, intrabodies

Genome diversity:
- comparative and developmental immunology, evolution of the adaptive immune system

http://imgt.cines.fr
IMGT, the international ImMunoGeneTics information system®
http://imgt.cines.fr

The IMGT team at Montpellier, France