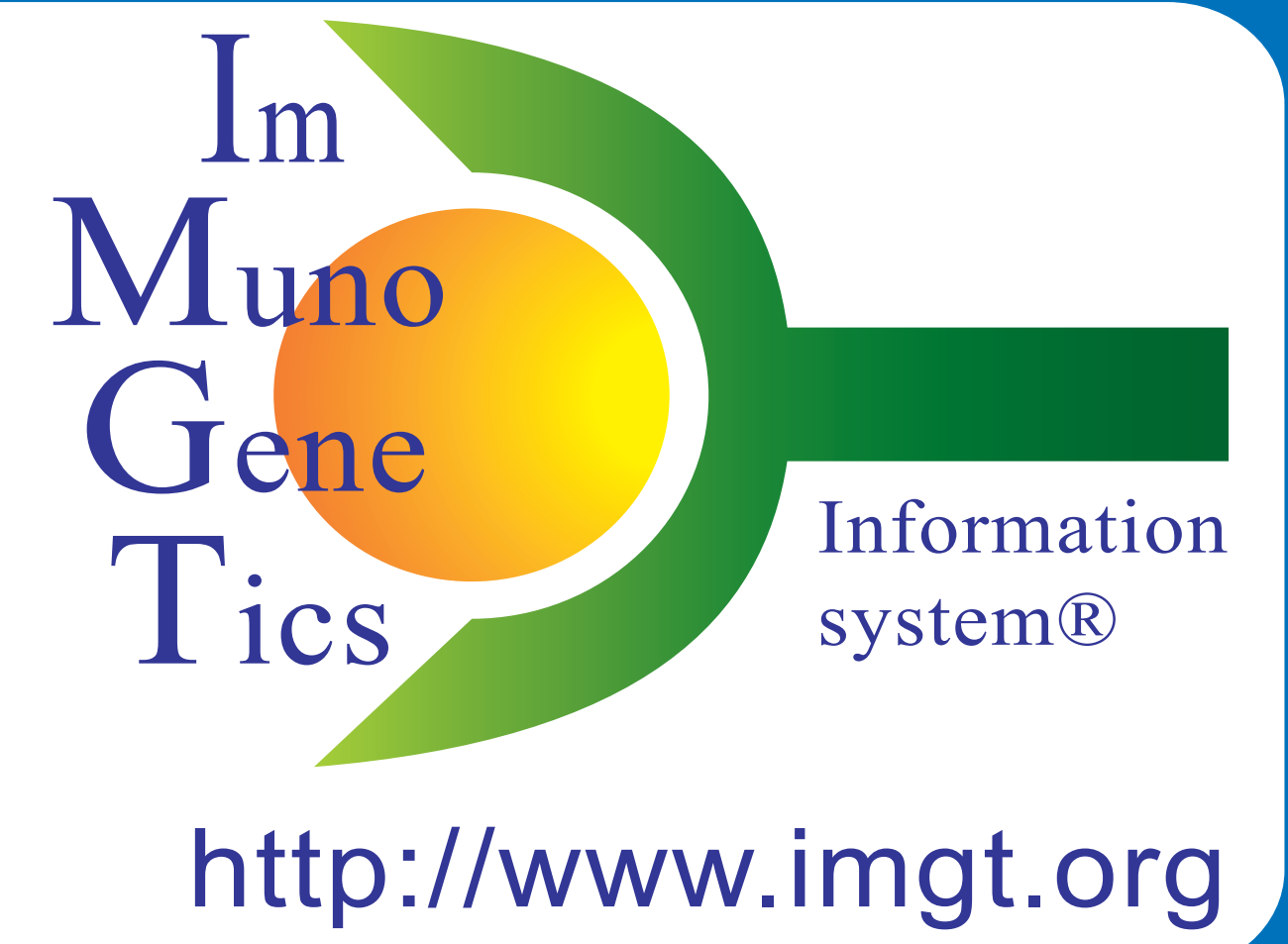# IMGT/HighV-QUEST

Véronique Giudicelli, Patrice Duroux, Arthur Lavoie, Safa Aouinti, Sofia Kossida, Marie-Paule Lefranc
IMGT®, the international ImMunoGeneTics information system®,
Université de Montpellier and CNRS, Laboratoire d'ImmunoGénétique Moléculaire (LIGM),
Institut de Génétique Humaine (IGH), UPR CNRS 1142, Montpellier (France)

Im Muno Gene Tics Information system®

http://www.imgt.org

IMGT/HighV-QUEST [1-3] (http://www.imgt.org), created in October 2010, is the high-throughput version of IMGT/V-QUEST, available on the Web for the direct analysis of complete IG and TR domain sequences from NGS. IMGT/HighV-QUEST analyses up to 500,000 sequences per run in, with the same degree of resolution and high-quality results as IMGT/V-QUEST. The functionalities of IMGT/HighV-QUEST include: the introduction of IMGT gaps, the identification of indels and their correction (by default), the identification of the closest V, D and J genes and alleles, the IMGT/JunctionAnalysis results, the description of mutations and amino acid changes, the annotation by IMGT/Automat, the statistical analysis and the characterization of the IMGT clonotypes (AA). An 'IMGT clonotype (AA)' is defined by a unique V-(D)-J rearrangement (with IMGT V and J gene and allele names determined by IMGT/HighV-QUEST at the nt level) and a unique CDR3-IMGT AA (in-frame) junction sequence. Each 'IMGT clonotype (AA)', in a given dataset, has a unique set identifier and, importantly, has a unique representative sequence selected by IMGT/HighV-QUEST, based on the highest percent of identity of the V-REGION ('V %') compared to that of the closest germline, and/or on the sequence length (thus the most complete V-REGION). The sequences of each '1 copy' assigned to a given 'IMGT clonotype (AA)' are available in 'Sequences file'. Clonotype diversity is visualized in the online results with histograms which represent the number of IMGT clonotypes (AA) per V, D (for IGH) and J genes (in pink). Clonotype expression is visualized with histograms which represent the number of sequences assigned to IMGT clonotypes (AA) per V (in green), D (in red) and J (in yellow) genes. Values are normalized, respectively, for 10,000 IMGT clonotypes (AA) to represent IG diversity immunoprofiles per V, D (for IGH) and J genes, and for 10,000 sequences assigned to IMGT clonotypes (AA) to represent IG expression immunoprofiles per V, D (for IGH) and J genes. These normalized values allow comparative analysis studies performed with the same IMGT/HighV-QUEST standards.

[1] Alamyar E et al. Immunome Res 8(1):26 (2012). [2] Alamyar E et al. Mol Biol 882:569-604 (2012). [3] Li S et al. Nat. Commun. 4:2333 (2013).

Biological Context

The adaptive immune response is characterized by an extreme diversity of the specific antigen receptors that comprise the immunoglobulins (IG) or antibodies and the T cell receptors (TR) (10^{12} different IG and 10^{12} different TR per individual, in humans). The complex molecular mechanisms (DNA rearrangements, N-diversity, and for IG, somatic hypermutations) that occur in B cells and T cells are at the origin of that huge diversity.

10^{12} IG (antibodies) per individual
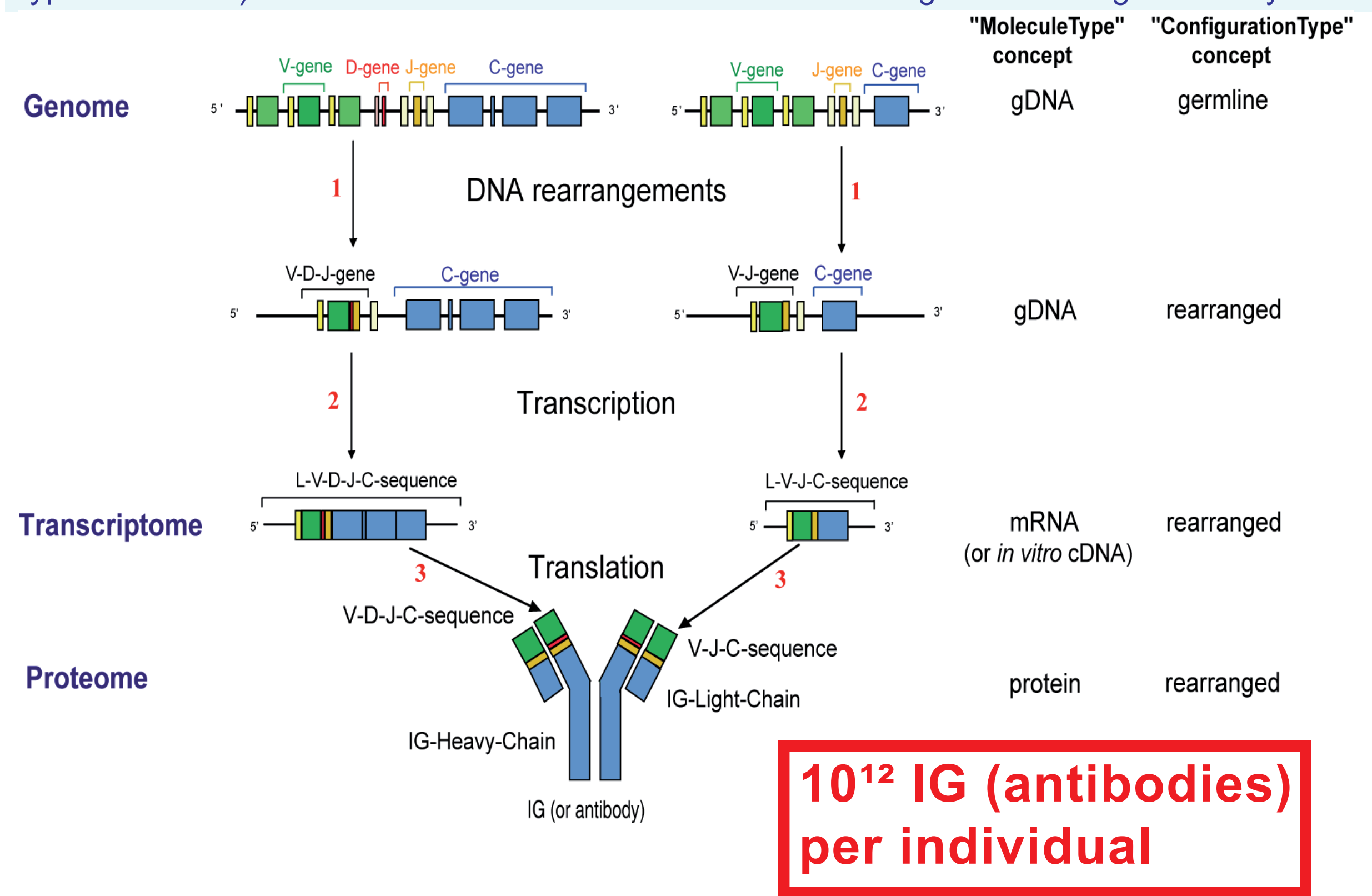
IMGT-ONTOLOGY concepts

Prototypes of IG and TR V, D, J genes
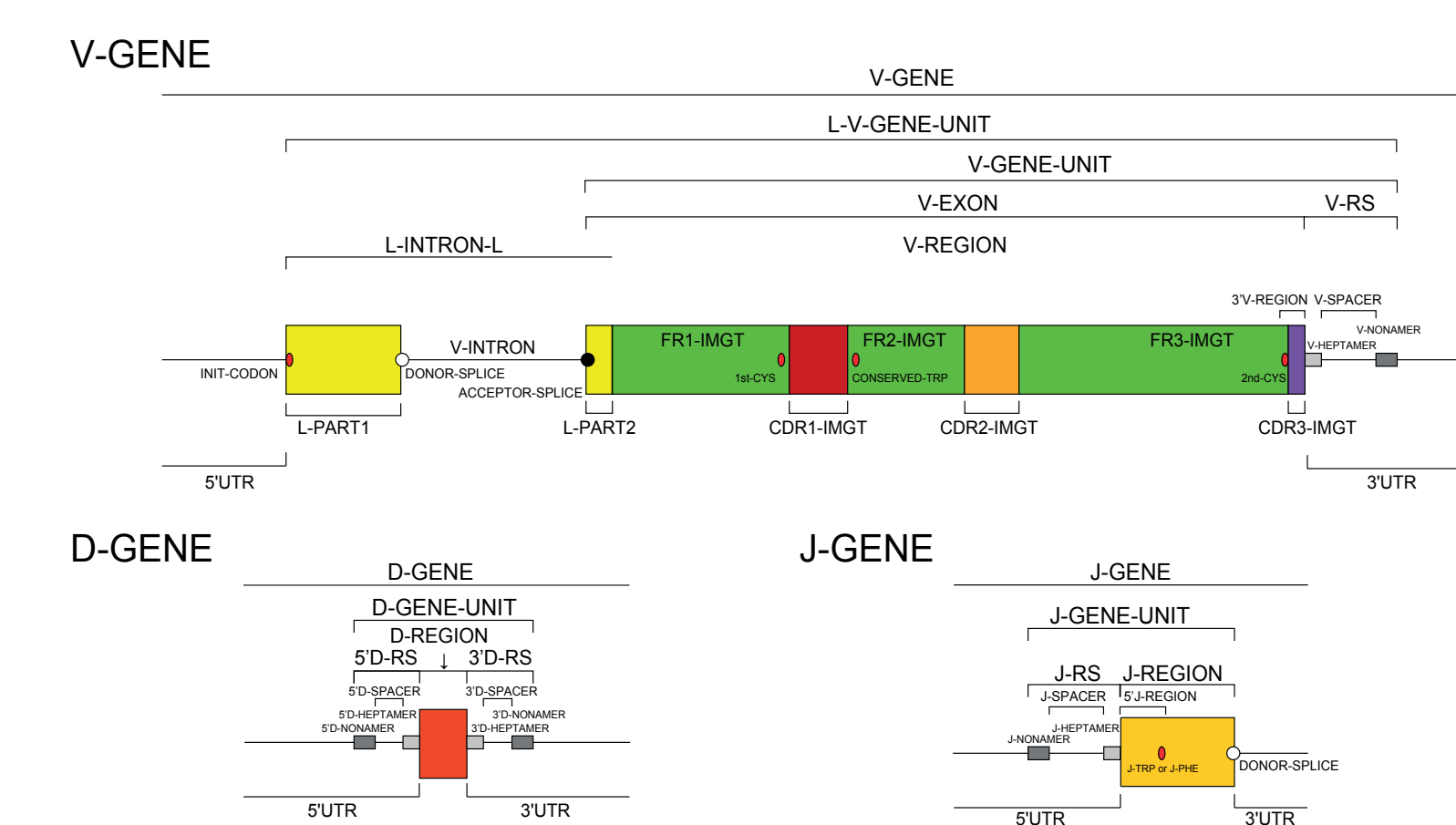
Prototypes are graphical representation based on the concepts of description

IMGT-ONTOLOGY Concepts

DESCRIPTION
The concepts of description correspond to IMGT® standardized labels. They are more than 560 standardized labels (available in the IMGT Scientific chart), 277 for the nucleotide sequences and 285 for the 3D structures.

CLASSIFICATION
The concepts of classification allowed to classify and name the human IG and TR genes and alleles which were approved by HGNC and endorsed by WHO-IUIS. They provide the frame for the standardized IG and TR nomenclature of jawed vertebrates.
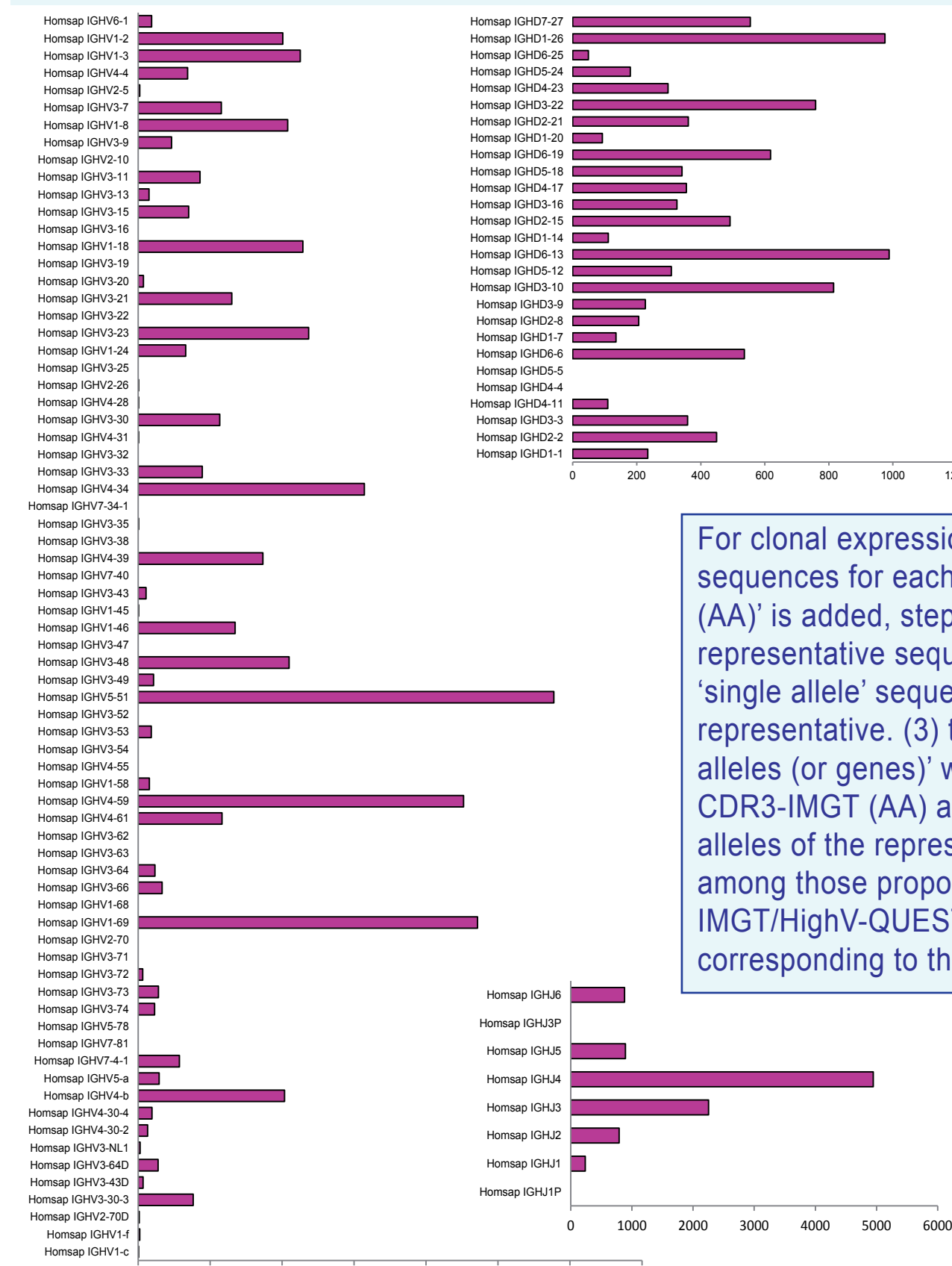
NUMEROTATION
The concepts of numerotation comprise the 'IMGT unique numbering' and 'IMGT Collier de Perles'.
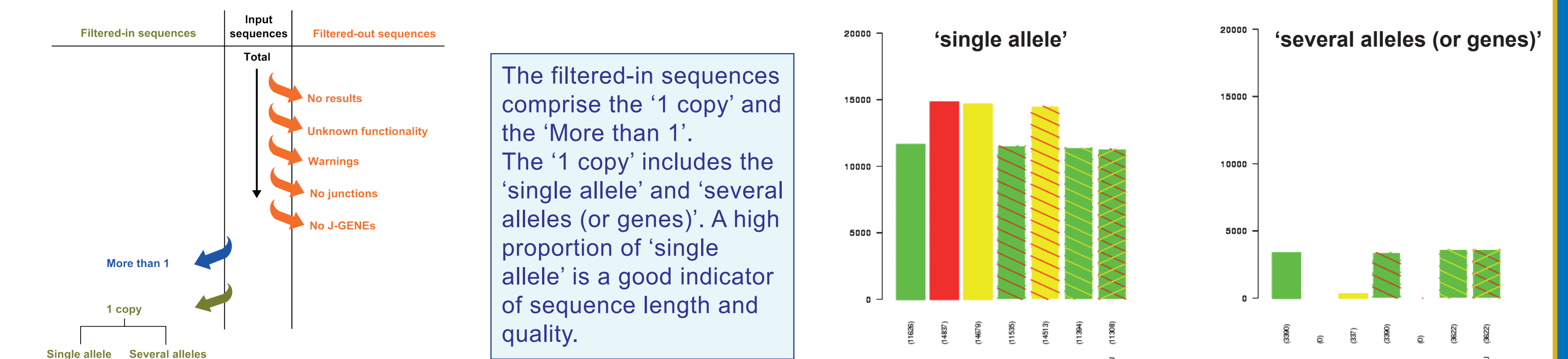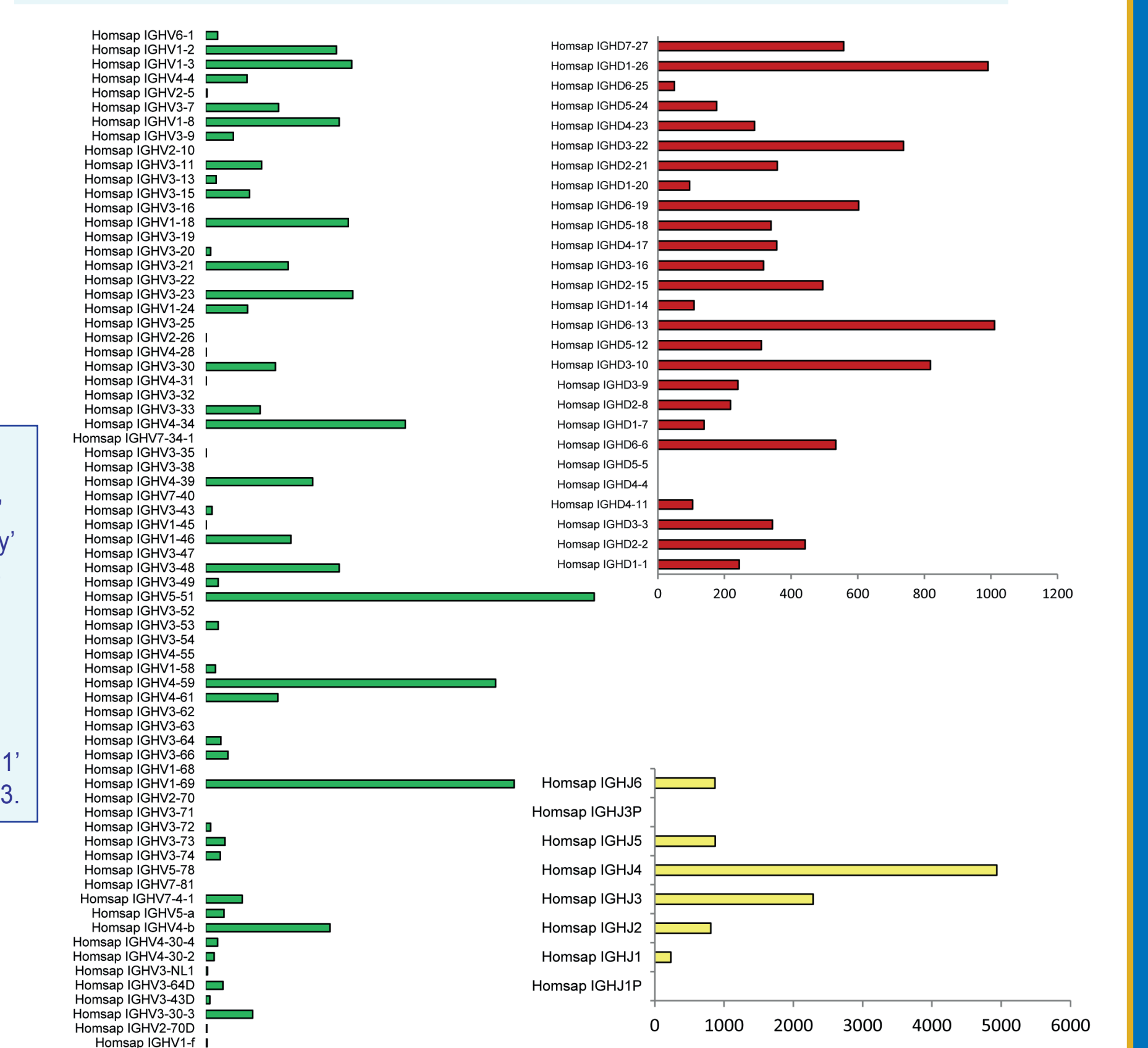
IMGT Clonotypes (AA)

In IMGT®, the clonotype designated as 'IMGT clonotype (AA)' is defined among the '1 copy' 'single allele' (for V and J) by a unique V-(D)-J rearrangement (IMGT genes and alleles determined at the nt level), conserved anchors (C104, W or F 118), and a unique CDR3-IMGT AA in frame junction [4]. Each 'IMGT clonotype (AA)' is characterized by a selected unique representative sequence.

ID		IMGT clonotype (AA) definition					IMGT clonotype (AA) representative sequence				Nb			IMGT clonotypes (nt)
#	Exp. ID	V gene and allele	D gene and allele	J gene and allele	CDR3-IMGT length (AA)	CDR3-IMGT sequence (AA)	Anchors 104,118	V %	Sequence length	Sequence ID	Total nb of '1 copy'	Total nb of 'More than 1'	Total	Sequences file ('1 copy')
Homsap IGHV1-2*02 F														
1	137-mid5	Homsap IGHV1-2*02 F	Homsap IGHD2-2*01 F	Homsap IGHJ6*03 F	22 AA	ARDLYCSSTSCYGGWYYYYMDV	C,F	95.14	425	GJNZT840ZHBX9K length=425	1	0	1	Sequences file
2	157-mid5	Homsap IGHV1-2*02 F	Homsap IGHD6-6*01 F	Homsap IGHJ6*03 F	22 AA	ARERVGRSIAARRAPDYYYMDV	C,F	97.92	426	GJNZT840ZH4DK W_length=426	1	0	1	Sequences file
3	305-mid5	Homsap IGHV1-2*02 F	Homsap IGHD3-22*01 F	Homsap IGHJ4*02 F	21 AA	ARGPYHRPTYYYDSSGYYGDY	C,F	96.15	374	GJNZT840ZFSHFL length=374	1	0	1	Sequences file
4	331-mid5	Homsap IGHV1-2*02 F	Homsap IGHD3-10*02 F	Homsap IGHJ3*02 F	21 AA	ARNVGHRPGSSDAWEADAFDI	C,F	99.31	422	GJNZT840ZIF2IO length=422	1	0	1	Sequences file
5	374-mid5	Homsap IGHV1-2*02 F	Homsap IGHD3-3*01 F	Homsap IGHJ3*02 F	21 AA	ATTHEPAITIFGVVINDAFDI	C,F	96.18	422	GJNZT840ZH3QEP length=422	1	0	1	Sequences file
6	647-mid5	Homsap IGHV1-2*02 F	Homsap IGHD6-19*01 F	Homsap IGHJ6*03 F	19 AA	AXGAIAVAGTNYYYYYMDV	C,F	97.74	330	GJNZT840ZH4VTXP length=330	1	0	1	Sequences file
7	693-mid5	Homsap IGHV1-2*02 F	Homsap IGHD3-22*01 F	Homsap IGHJ5*02 F	19 AA	ARDGSTYYYDSSGYYWFDP	C,F	99.65	417	GJNZT840ZFI7QZ length=417	1	0	1	Sequences file
8	709-mid5	Homsap IGHV1-2*02 F	Homsap IGHD4-23*01 ORF	Homsap IGHJ2*01 F	19 AA	ARDMGRYGGNURRYWYFDL	C,F	97.57	416	GJNZT840ZG7EN4 length=416	1	0	1	Sequences file

IMGT/HighV-QUEST based on IMGT® standard

The filtered-in sequences comprise the '1 copy' and the 'More than 1'. The '1 copy' includes the 'single allele' and 'several alleles (or genes)'. A high proportion of 'single allele' is a good indicator of sequence length and quality.

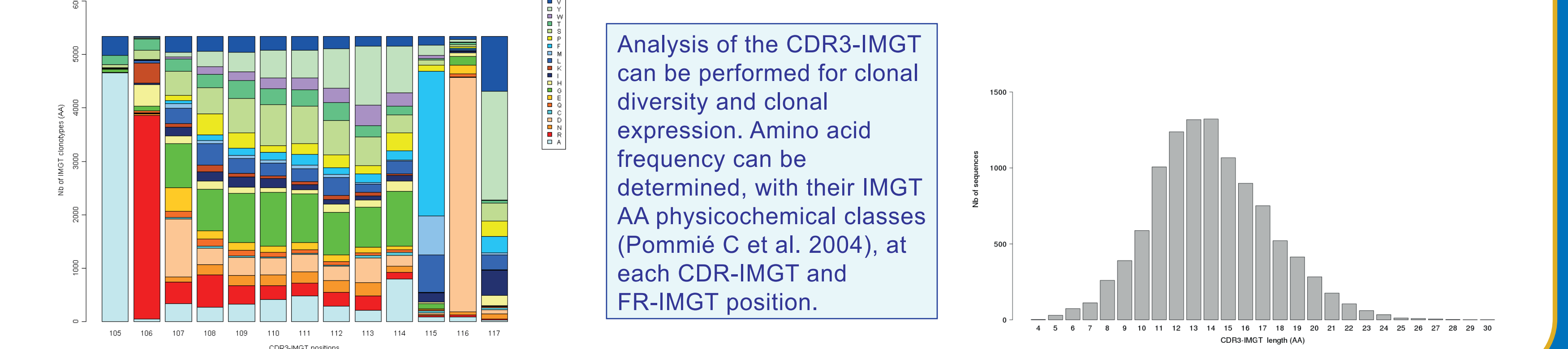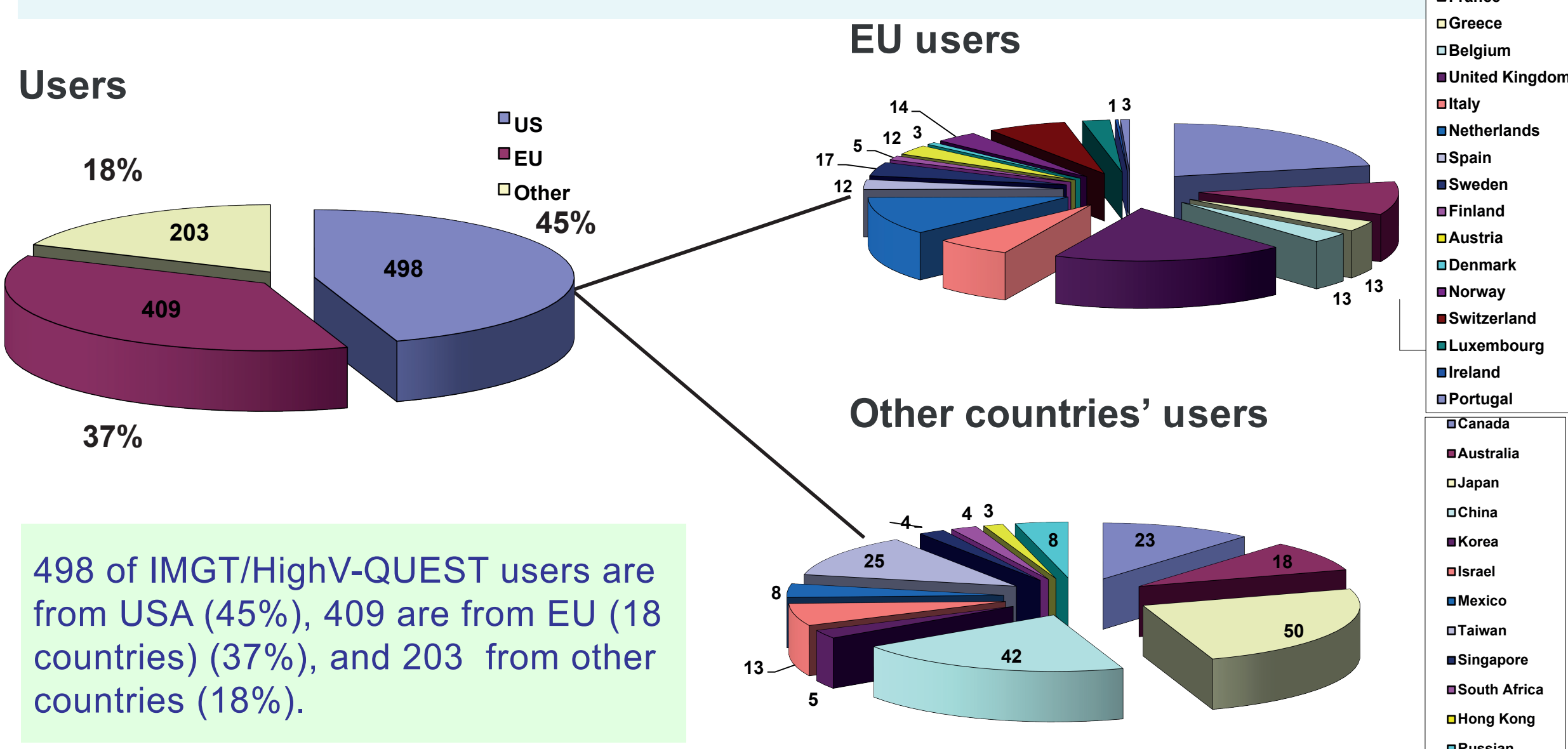**'single allele'** / **'several alleles (or genes)'**

VH clonal diversity
Clonal diversity is the number of 'IMGT clonotypes (AA)' per V, D and J gene.

VH clonal expression
Clonal expression is the number of sequences assigned to 'IMGT clonotypes (AA)' per V, D and J gene.

For clonal expression, the number of sequences for each 'IMGT clonotypes (AA)' is added, stepwise: (1) the '1 copy' representative sequence. (2) the '1 copy' 'single allele' sequences not selected as representative. (3) the '1 copy' 'several alleles (or genes)' with the same CDR3-IMGT (AA) and the same V and J alleles of the representative sequence among those proposed by IMGT/HighV-QUEST. (4) the 'More than 1' corresponding to the '1 copy' of steps 1-3.

Analysis of the CDR3-IMGT can be performed for clonal diversity and clonal expression. Amino acid frequency can be determined, with their IMGT AA physicochemical classes (Pommié C et al. 2004), at each CDR-IMGT and FR-IMGT position.

Users and Analyses

Since the availability of IMGT/HighV-QUEST in October 2010, more than 6.1 billions of sequences (from external users) have been submitted. They required more than 3.2 million hours of computational resources. About 146 terabytes of results were generated.

Users — US 45%, EU 37%, Other 18%

498 of IMGT/HighV-QUEST users are from USA (45%), 409 are from EU (18 countries) (37%), and 203 from other countries (18%).