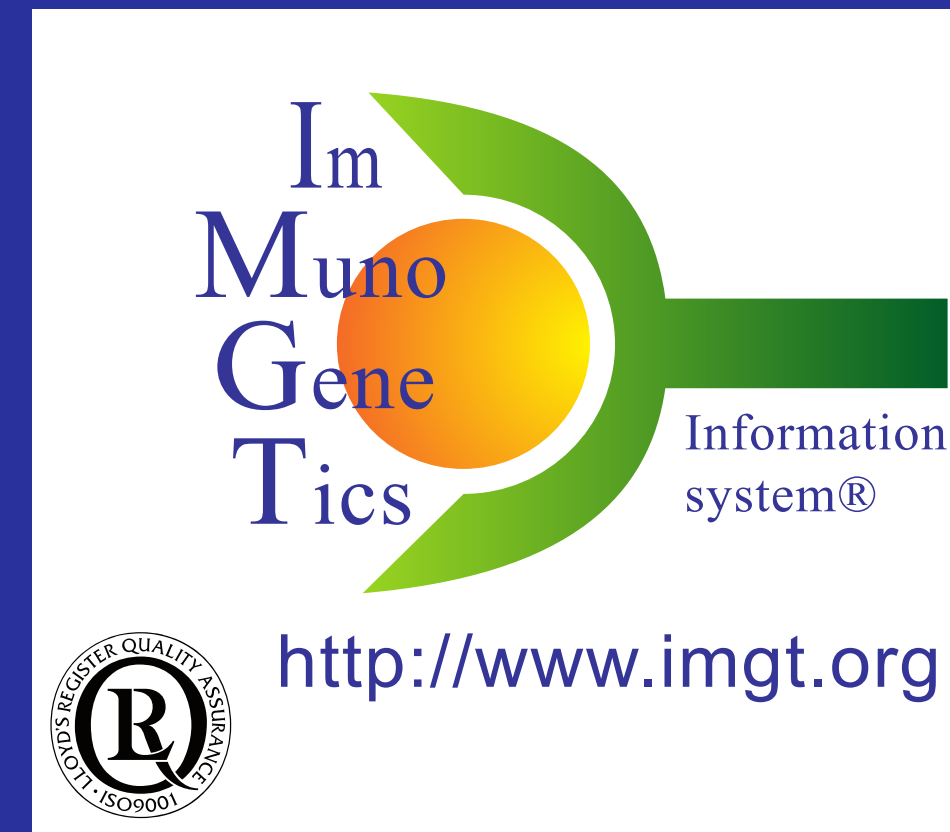


# 'IMGTStatClonotype': An R package with integrated web tool for pairwise evaluation and visualization of IMGT clonotype diversity and expression from IMGT/HighV-QUEST output

Safa Aouinti<sup>1,2</sup>, Patrice Duroux<sup>1</sup>, Véronique Giudicelli<sup>1</sup>, Dhafer Malouche<sup>2</sup>, Sofia Kossida<sup>1</sup>, Marie-Paule Lefranc<sup>1</sup>

<sup>1</sup>IMGT®, the international ImMunoGeneTics information system®, Laboratoire d'ImmunoGénétique Moléculaire (LIGM), Institut de Génétique Humaine (IGH), UPR CNRS 1142, Montpellier University and CNRS, Montpellier (France)

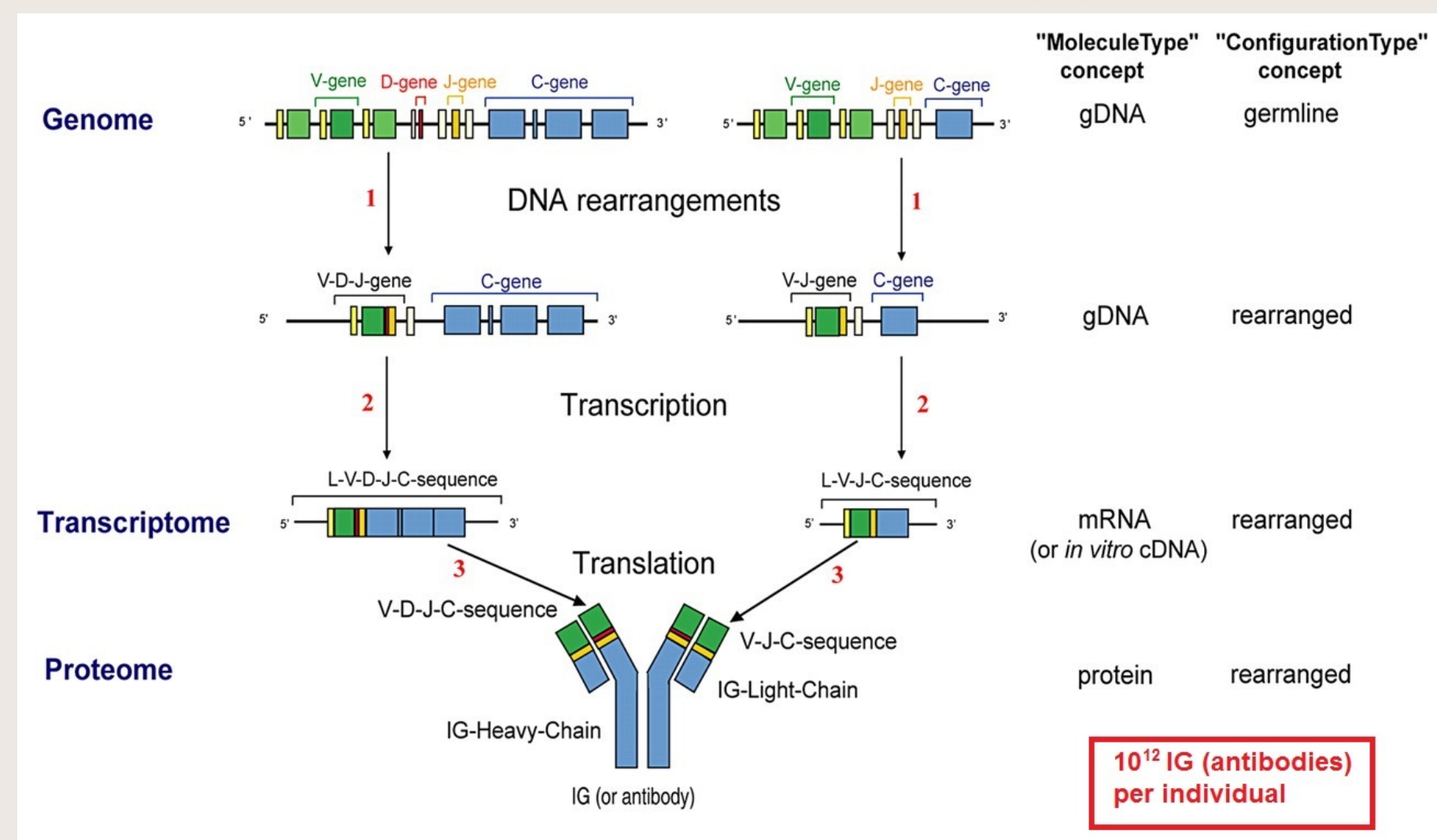
<sup>2</sup>Higher School of Statistics and Information Analysis, Unité Modélisation et Analyse Statistique et Economique, University of Carthage, Tunis (Tunisia)



## IMGT-ONTOLOGY for IG and TR

The adaptive immune response is our ability to produce up to 2.10<sup>12</sup> different immunoglobulins (IG) or antibodies and T cell receptors (TR) per individual to fight pathogens.

IMGT®, the international ImMunoGeneTics information system®, was created in 1989 by Marie-Paule Lefranc (Montpellier University and CNRS) to manage the huge and complex diversity of these antigen receptors [1].



IMGT® is at the origin of immunoinformatics [2], a science at the interface between immunogenetics and bioinformatics. IMGT® is built on the IMGT-ONTOLOGY concepts of identification (keywords), description (labels), classification (gene and allele nomenclature) and numerotation (IMGT unique numbering) [3].

## 'IMGTStatClonotype' R package

'IMGTStatClonotype' is an R package for statistical analysis of sets from IMGT/HighV-QUEST output.

'IMGTStatClonotype' includes a generic and standardized procedure for evaluating the statistical significance of pairwise comparison between differences in proportions of the IMGT clonotypes (AA) diversity and expression per gene of a given IG or TR variable (V), diversity (D) or joining (J) group [8].

### User inputs

- Choose IMGT/HighV-QUEST outputs
- Specify the range of CDR3-IMGT
- Select IMGT clonotype Diversity or expression

### User outputs

#### 'IMGTStatClonotype' analysis modules

- CDR3-IMGT outlier lengths elimination.
  - Normalized bar graph of the proportions displays.
  - Significance of the difference in proportions with 95% confidence interval (CI) for IMGT clonotype (AA) diversity and expression between two sets from IMGT/HighV-QUEST output: Testing difference in proportions of IMGT clonotype (AA) per gene of a given group (z-scores, Fisher-test). Adjustment of the p-values is made through a multiple testing procedure by the two strategies [9]:
    - Family-wise error rate (FWER): Bonferroni, Holm, Sidak (single step and step down) and Hochberg procedures.
    - False discovery rate (FDR): The Benjamini & Hochberg (BH) and Benjamini & Yekutieli (BY) procedures.
  - Multiple testing procedures displays.
  - Differences in proportions graph displays.
  - IMGT/StatClonotype Shiny web application launch with 6 functionalities: 1 to 6.
- Package dependencies:** shiny, shinyjs, plotly, data.table, DT, reshape2, multtest, ggplot2, gridExtra, d3heatmap.

'IMGTStatClonotype' incorporates a user-friendly Shiny web interface, allowing use of the IMGT/StatClonotype tool, in users' own browser.

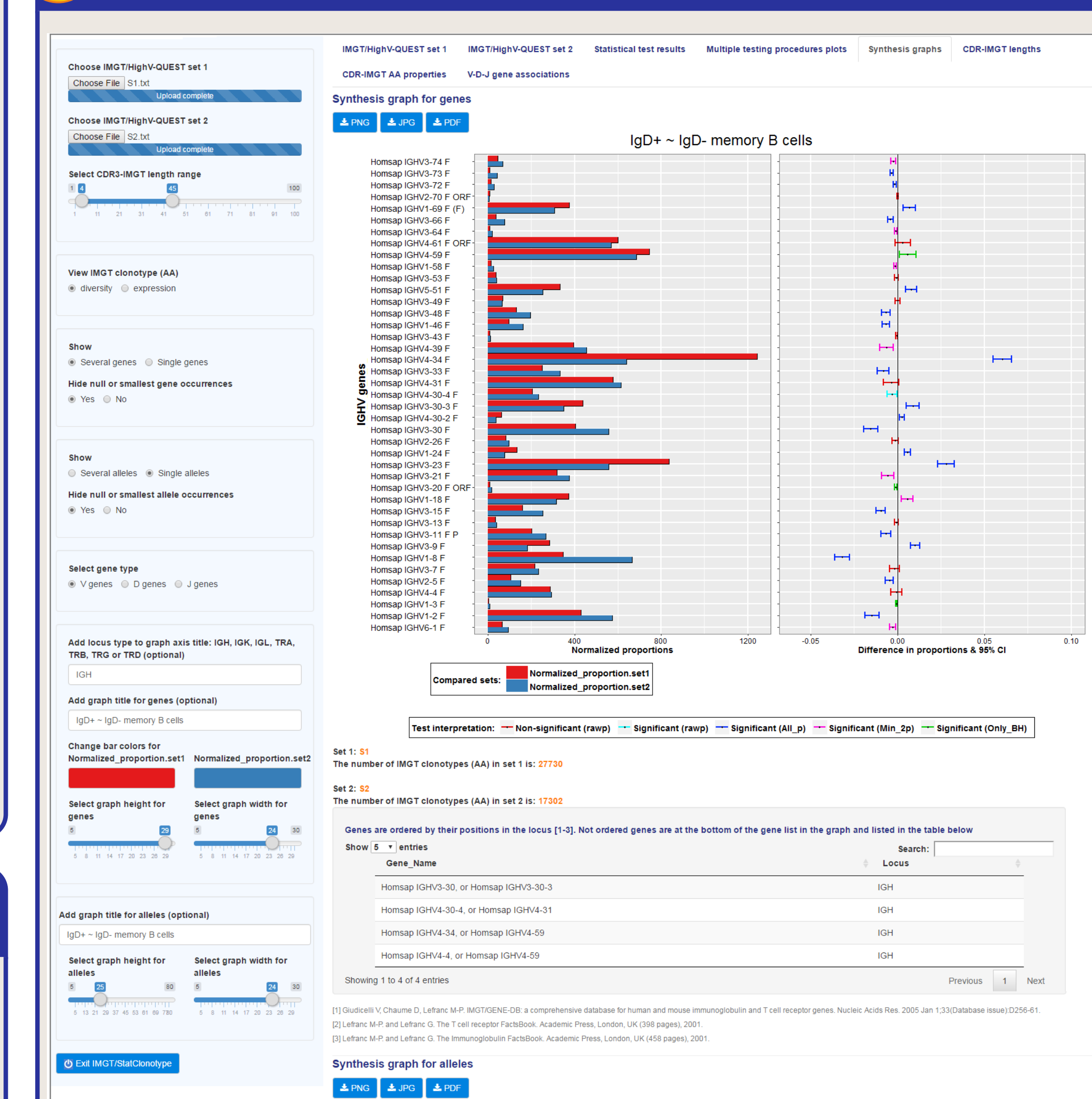
## IMGT/HighV-QUEST for NGS analysis

IMGT/HighV-QUEST [4-7], the first web portal for next generation sequencing (NGS) analysis of IG and TR, provides the identification of the variable (V), diversity (D) and joining (J) genes and alleles, analysis of the V-(D)-J junction and characterization of the 'IMGT clonotype (AA)' (AA for amino acid).

Exp. ID	IMGT clonotype (AA) definition	IMGT clonotype (AA) representative sequence	AA	IMGT clonotype (AA) sequence ID	Total No. of clones	Total No. of reads	IMGT clonotype (AA) frequency (%)
1-137-m145	Homolog IGKV1-2*02 F	Homolog IGHJ2-2*03 F	IGKV1-2*02 F	IGHJ2-2*03 F	423	1014	0.000000
1-137-m145	Homolog IGKV1-2*02 F	Homolog IGHJ2-2*03 F	IGKV1-2*02 F	IGHJ2-2*03 F	1	1	0.000000
1-137-m145	Homolog IGKV1-2*02 F	Homolog IGHJ2-2*03 F	IGKV1-2*02 F	IGHJ2-2*03 F	1	1	0.000000

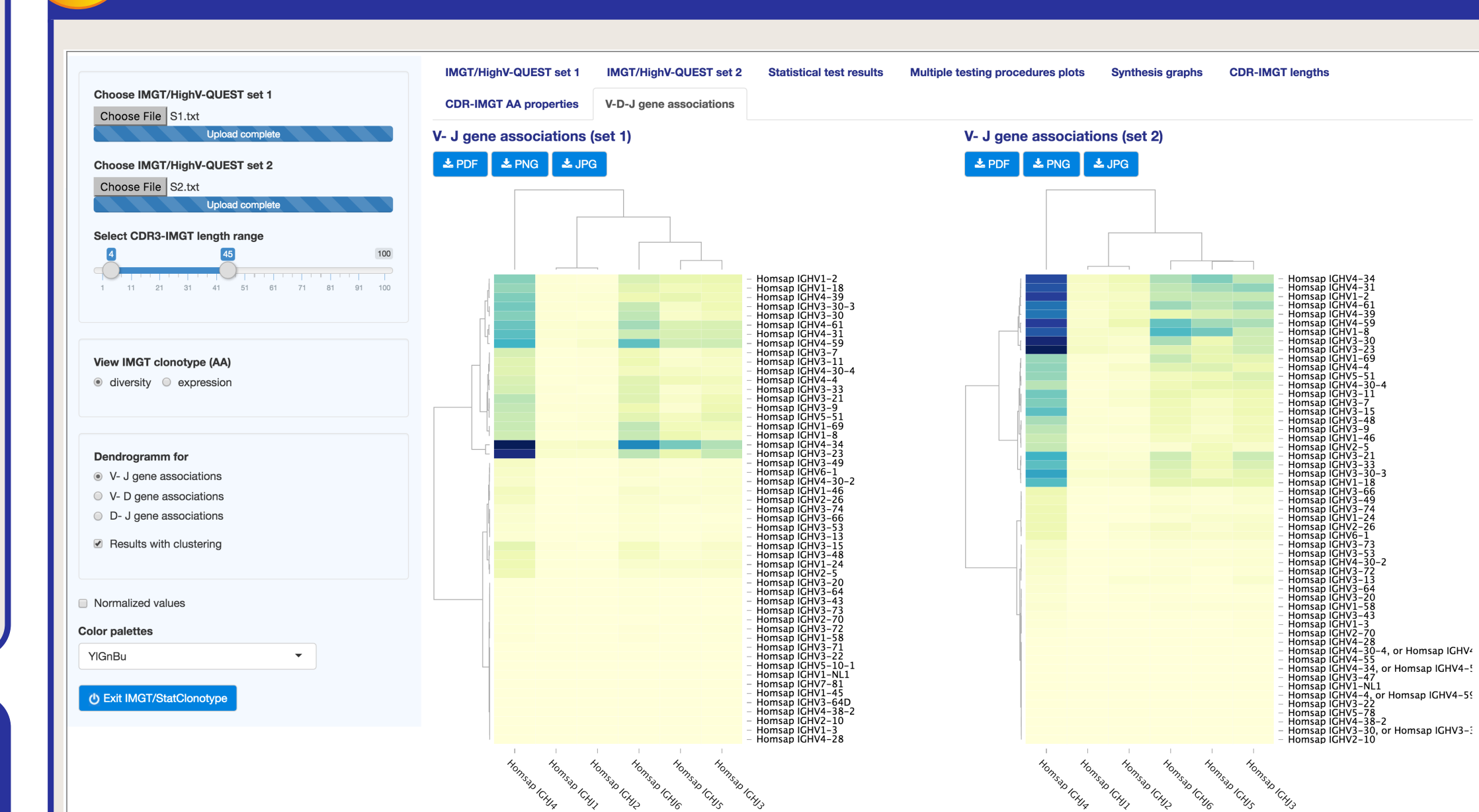
IMGT clonotype (AA) is defined by a unique V-(D)-J rearrangement (IMGT genes and alleles determined at the nucleotide level), conserved anchors (C104, W or F 118), and a unique complementarity determining region 3 (CDR3-IMGT) AA in frame junction [5].

## 3 Synthesis graphs



Synthesis graphs are for IG or TR genes and alleles.

## 6 V-J, V-D and D-J gene associations



IG or TR V-J, V-D and D-J gene associations per set.

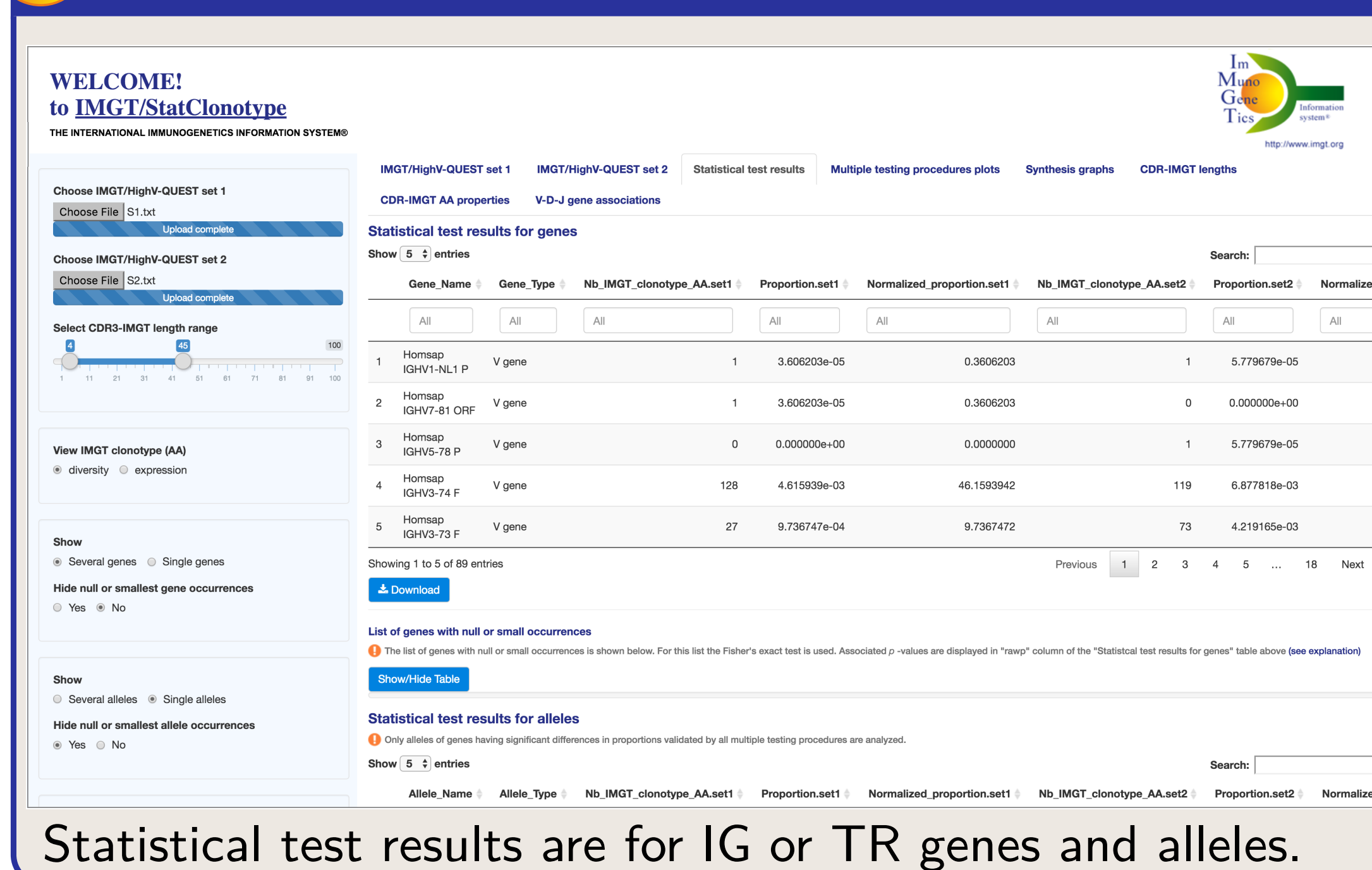
## Conclusion

High throughput IG and TR repertoire immunoprofiles are of prime importance in vaccination, cancer, autoimmunity and lymphoproliferative disorders. IMGT/StatClonotype is generic [8] and suitable for detecting significant changes in IG and TR immunoprofiles in protective (vaccination, cancers and infections) or pathogenic (autoimmunity and lymphoproliferative disorders) immune responses.

## IMGT/StatClonotype Shiny web tool

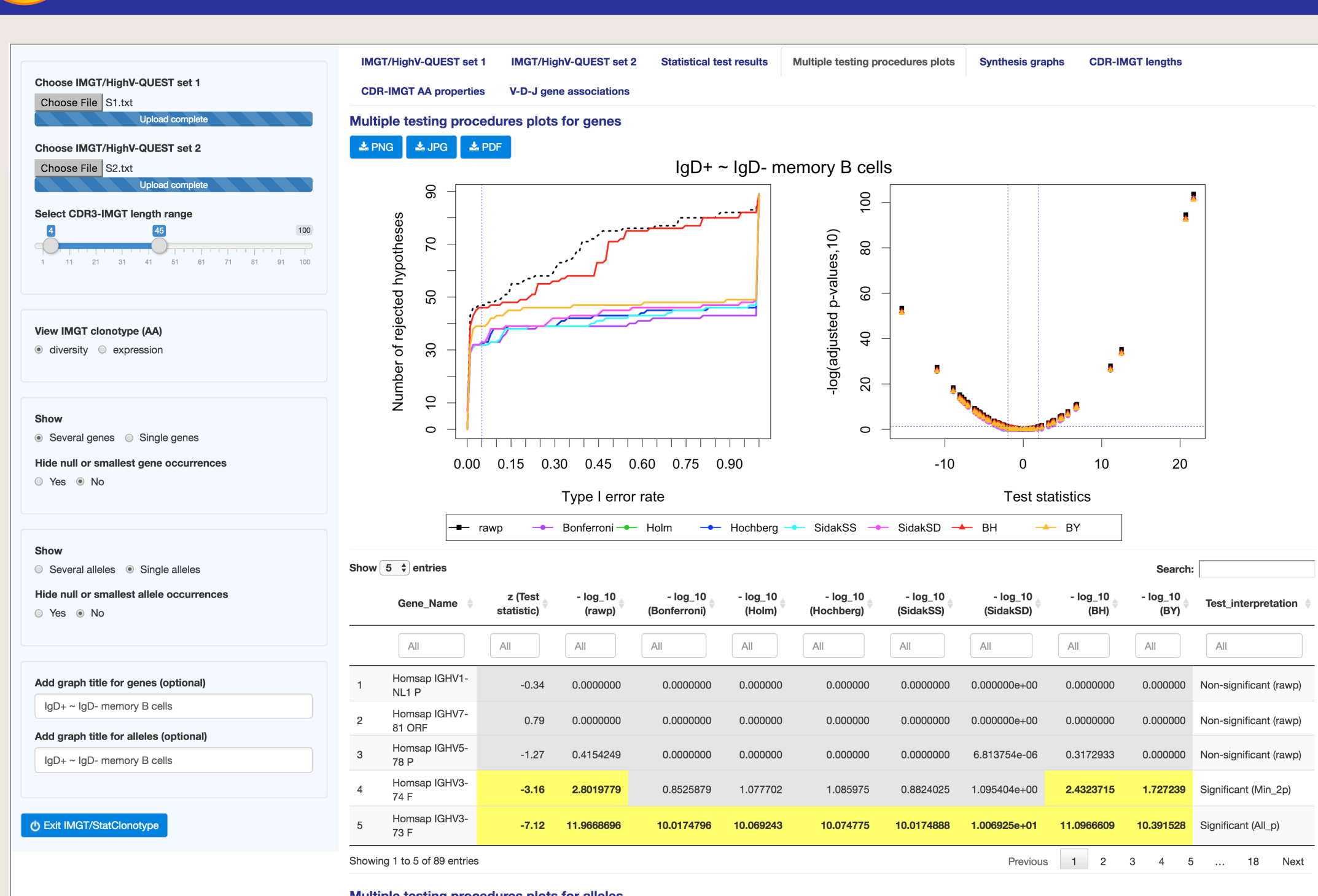
IMGT/StatClonotype, is an IMGT® [1] tool for statistical analysis of sets from IMGT/HighV-QUEST output [4-7]. IMGT/StatClonotype uses a generic statistical procedure [8] for identifying significant changes in IG and TR differences of proportions of IMGT clonotypes (AA) diversity and expression [6]. It uses the IMGT gene and allele nomenclature based on IMGT-ONTOLOGY [3] and IMGT standards in immunoinformatics [2]. IMGT/StatClonotype, performs pairwise comparison of sets from IMGT/HighV-QUEST output through a user-friendly web interface implemented using Shiny framework [10] in users' own browser.

## 1 Statistical test results



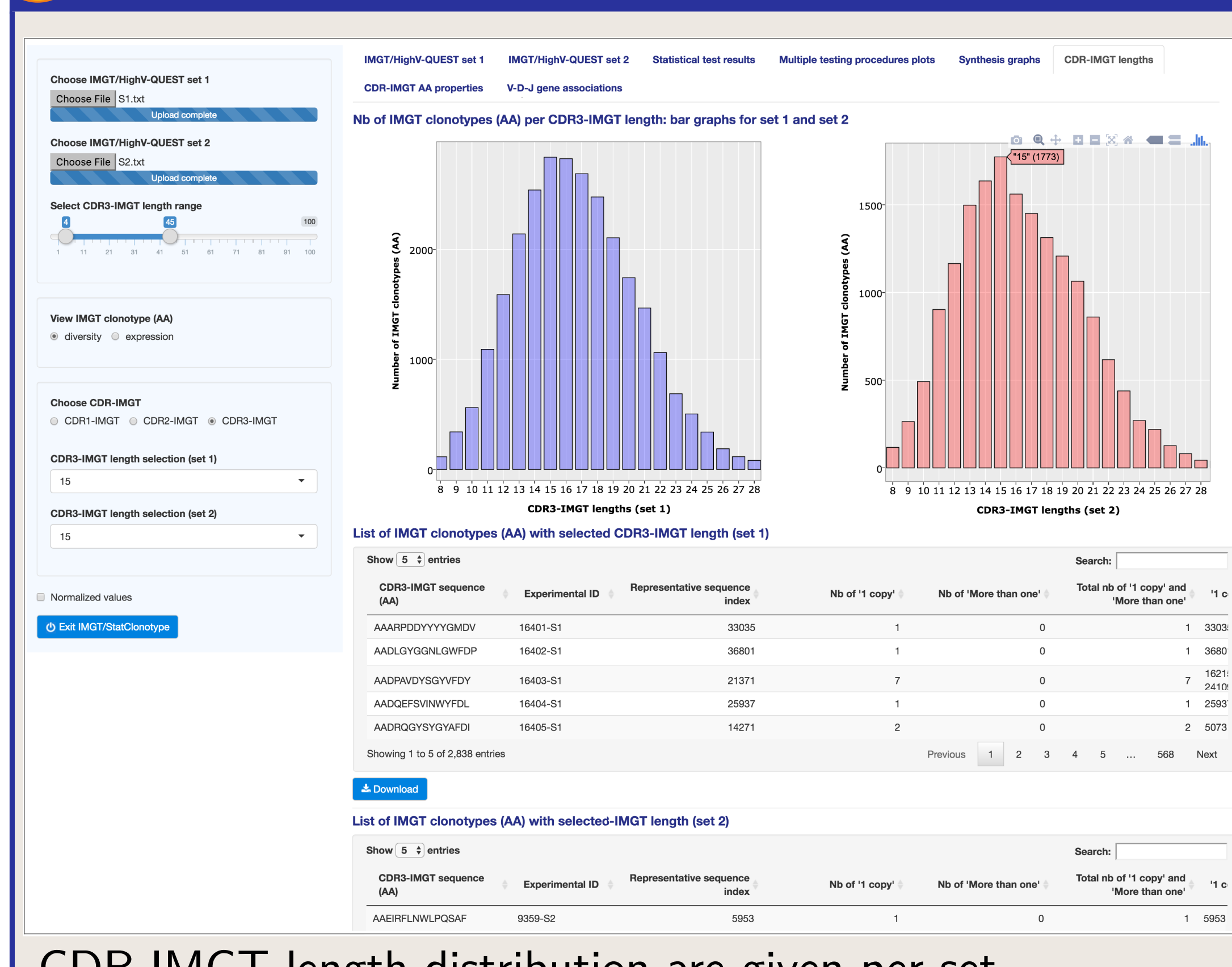
Statistical test results are for IG or TR genes and alleles.

## 2 Multiple testing procedures



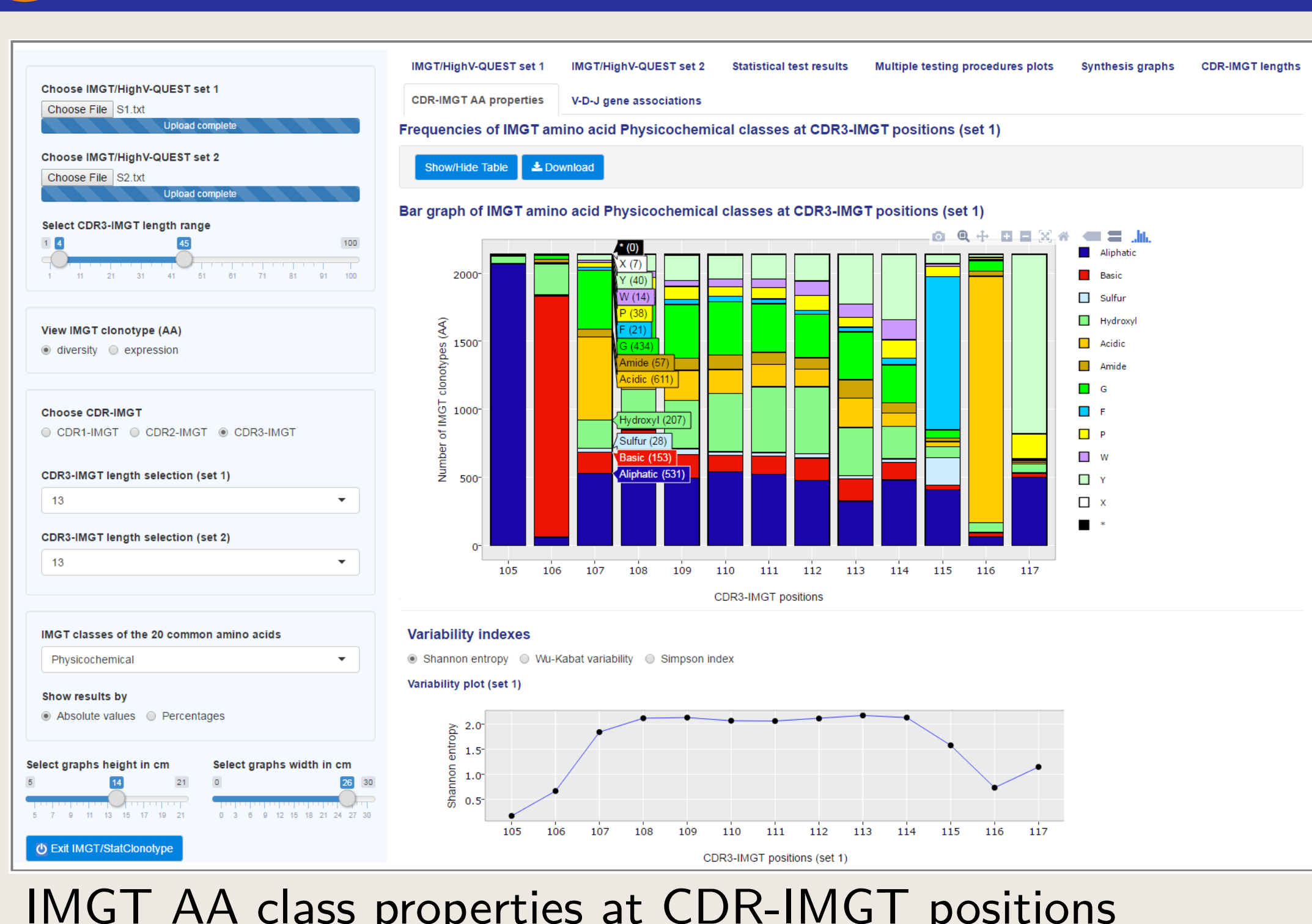
Multiple testing procedures plots are for genes and alleles

## 4 CDR-IMGT length distribution



CDR-IMGT length distribution are given per set.

## 5 CDR-IMGT AA class properties



IMGT AA class properties at CDR-IMGT positions

**References:** [1] Lefranc M-P et al. Nucleic Acids Res. 43:413-422, 2015. [2] Lefranc M-P, Front Immunol, 5:22, 2014. [3] Giudicelli V and Lefranc M-P, Front Genet, 3:79, 2012. [4] Alamyar E et al. Mol Biol 882:569-604, 2012. [5] Alamyar E et al. Immunome Res 8(1):26, 2012. [6] Li S et al. Nat. Commun. 4:2333, 2013. [7] Giudicelli V et al. Autoimmun Infec Dis 1(1), 2015. [8] Aouinti S et al. PloS ONE 10(11): e0142353, 2015. [9] Dudoit S, van der Laan MJ. Springer Series in Statistics; 2008. [10] Chang W et al. shiny: Web Application Framework for R (v.0.13.2) 2016.

**Acknowledgments:** this work was granted access to the HPC resources of HPC@LR and of CINES and TGCC-CEA under the allocation 036029-(2010-2016) made by GENCI.

IMGT® founder and executive director emeritus: Marie-Paule Lefranc (Marie-Paule.Lefranc@igh.cnrs.fr)

IMGT® director: Sofia Kossida (Sofia.Kossida@igh.cnrs.fr)

Bioinformatics manager: Véronique Giudicelli (Veronique.Giudicelli@igh.cnrs.fr)

Computer manager: Patrice Duroux (Patrice.Duroux@igh.cnrs.fr)



© Copyright 1989-2016 IMGT®, the international ImMunoGeneTics information system®