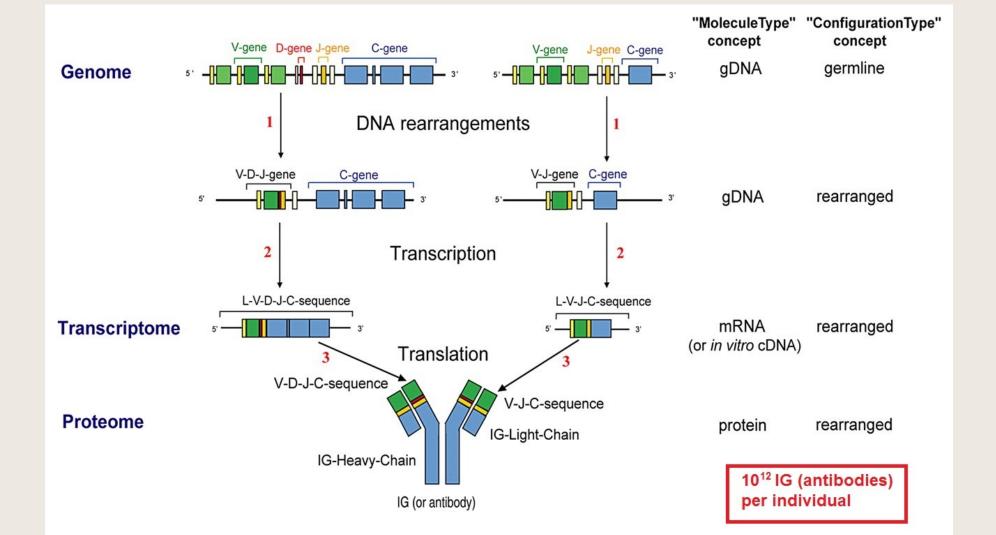
'IMGTStatClonotype': An R package with integrated web tool for pairwise evaluation and visualization of IMGT clonotype diversity and expression from IMGT/HighV-QUEST output

Safa Aouinti^{1,2}, Patrice Duroux¹, Véronique Giudicelli¹, Dhafer Malouche², Sofia Kossida¹, Marie-Paule Lefranc¹

¹IMGT[®], the international ImMunoGeneTics information system[®], Laboratoire d'ImmunoGénétique Moléculaire (LIGM), Institut de Génétique Humaine (IGH), UPR CNRS 1142, Montpellier University and CNRS, Montpellier (France) ²Higher 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^{12} 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].



'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

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).

ю		ING I 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	nb of '1	Total nb of 'More than 1'	Total	Sequences file ('1 copy')
н	omsap 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,w	95.14	425	GJNZTB402H8X9K length=425	1	0		<u>Sequences</u> file
	2 157-mid5	Homsap IGHV1-2*02 F	Homsap IGHD6-6*01 F	Homsap IGHJ6*03 F	22 AA	ARERVGRSIAARRAPDYYYMDV	c,w	97.92	4/6	GJNZTB402H4DK W_length=426	1	0		<u>Sequences</u> file
	3 305-mid5	Homsap IGHV1-2*02 F	Homsap IGHD3-22*01 F	Homsap IGHJ4*02 F	21 AA	ARGPYHRPTYYYDSSGYYGDY	c,w	96.15	374	<u>GJNZTB402FSHFL</u> <u>length=374</u>	1	0		<u>Sequences</u> file

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].



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].

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.

User outputs

<u>'IMGTStatClonotype' analysis modules</u>

- 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.
- Multiples 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 userfriendly Shiny web interface, allowing use of the IMGT/StatClonotype tool, in users' own browser.

Synthesis graphs

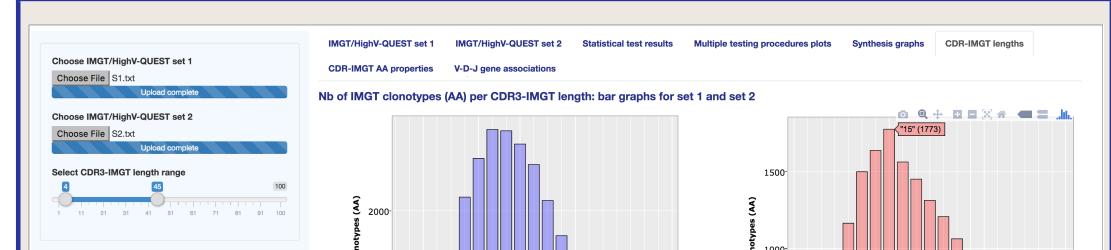
	IMGT/HighV-QUEST set 1 IMGT/HighV-QUEST set 2 Statistical test results Multiple testing procedures plots Synthesis graphs CDR-IMGT lengths
oose IMGT/HighV-QUEST set 1	IMGT/HighV-QUEST set 1 IMGT/HighV-QUEST set 2 Statistical test results Multiple testing procedures plots Synthesis graphs CDR-IMGT lengths CDR-IMGT AA properties V-D-J gene associations
hoose File S1.txt Upload complete	Synthesis graph for genes
oose IMGT/HighV-QUEST set 2	
hoose File S2.txt Upload complete	IgD+ ~ IgD- memory B cells
lect CDR3-IMGT length range	Homsap IGHV3-74 F Homsap IGHV3-73 F H
4 45 100	Homsap IGHV3-72 F Homsap IGHV2-70 F ORF
11 21 31 41 51 61 71 81 91 100	Homsap IGHV1-69 F (F) - Homsap IGHV3-66 F - Homsap IGHV3-66 F - Homsap IGHV3-64 F - Homsap IGHV4-61 F ORF - Homsap IGHV4-61 F
	Homsap IGHV4-51 F C
ew IMGT clonotype (AA) diversity O expression	Homsap IGHV3-53 F - Homsap IGHV5-51 F - Homsap IGHV3-49 F - H
	Homsap IGHV3-48 F Homsap IGHV1-46 F H
low	Homsap IGHV3-43 F Homsap IGHV4-39 F Homsap IGHV4
Several genes 💿 Single genes	Homsap IGHV4-34 F - Homsap IGHV3-33 F - Homsap IGHV4-31 F - Homsap
de null or smallest gene occurrences Yes 💿 No	Homsap IGHV4-30-4 F
	Homsap IGHV3-30-3 F
	Homsap IGHV3-30 F
ow Several alleles Single alleles 	Homsap IGHV1-24 F H
de null or smallest allele occurrences	Homsap IGHV3-21 F - Homsap IGHV3-20 F ORF - HOMSAP IGHV3- HOMSAP IGHV3-20 F ORF - HOMSAP IGHV3- HOMSAP IGHV3- HOMSAP IGHV3- HO
Yes 🔘 No	Homsap IGHV1-18 F - Homsap IGHV3-15 F - Homsap
	Hollisapi IGHV3-15 F Homsap IGHV3-13 F Homsap IGHV3-17 FP Homsap IGHV3-9 F Homsap IGHV1-8 F
lect gene type	Homsap IGHV3-9 F - F Homsap IGHV1-8 F - F Homsap IGHV3-7 F - F
V genes 💿 D genes 💿 J genes	Homsap IGHV2-7 F - Homsap IGHV2-4 F - Homsap IGHV2-4 F - Homsap IGHV2-6 F - Homsap IGHV2-
	Homsap IGHV1-3 F Homsap IGHV1-2 F
d locus type to graph axis title: IGH, IGK, IGL, TRA,	Homsap IGHV6-1 F
B, TRG or TRD (optional)	Normalized proportions Difference in proportions & 95% Cl
GH	Compared sets: Normalized_proportion.set1 Normalized_proportion.set2
d graph title for genes (optional)	
gD+ ~ IgD- memory B cells	Test interpretation: 🕶 Non-significant (rawp) 😁 Significant (rawp) 🕶 Significant (All_p) 😁 Significant (Min_2p) 🔤 Significant (Only_BH)
ange bar colors for rmalized_proportion.set1 Normalized_proportion.set2	2 Set 1: S1 The number of IMGT clonotypes (AA) in set 1 is: 27730
	Set 2: 52
lect graph height for Select graph width for	The number of IMGT clonotypes (AA) in set 2 is: 17302
nes genes 29 5 24 30	Genes are ordered by their positions in the locus [1-3]. Not ordered genes are at the bottom of the gene list in the graph and listed in the table below
8 11 14 17 20 23 28 29 5 8 11 14 17 20 23 28 29	Show 5 v entries Search: Search: Locus
	Homsap IGHV3-30, or Homsap IGHV3-30-3
	Homsap IGHV4-30-4, or Homsap IGHV4-31 IGH
graph title for alleles (optional))+ ~ IgD- memory B cells	Homsap IGHV4-34, or Homsap IGHV4-59 IGH
lect graph height for Select graph width for	Homsap IGHV4-4, or Homsap IGHV4-59 IGH
eles alleles	Showing 1 to 4 of 4 entries Previous 1 Next
11 21 29 37 45 53 81 69 780 5 8 11 14 17 20 23 26 29	 [1] Giudicelli V, Chaume D, Lefranc M-P. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res. 2005 Jan 1;33(Database issue):D256-61. [2] Lefranc M-P. and Lefranc G. The T cell receptor FactsBook. Academic Press, London, UK (398 pages), 2001. [3] Lefranc M-P. and Lefranc G. The Immunoglobulin FactsBook. Academic Press, London, UK (458 pages), 2001.
Exit IMGT/StatClonotype	Synthesis graph for alleles

Statistical test results

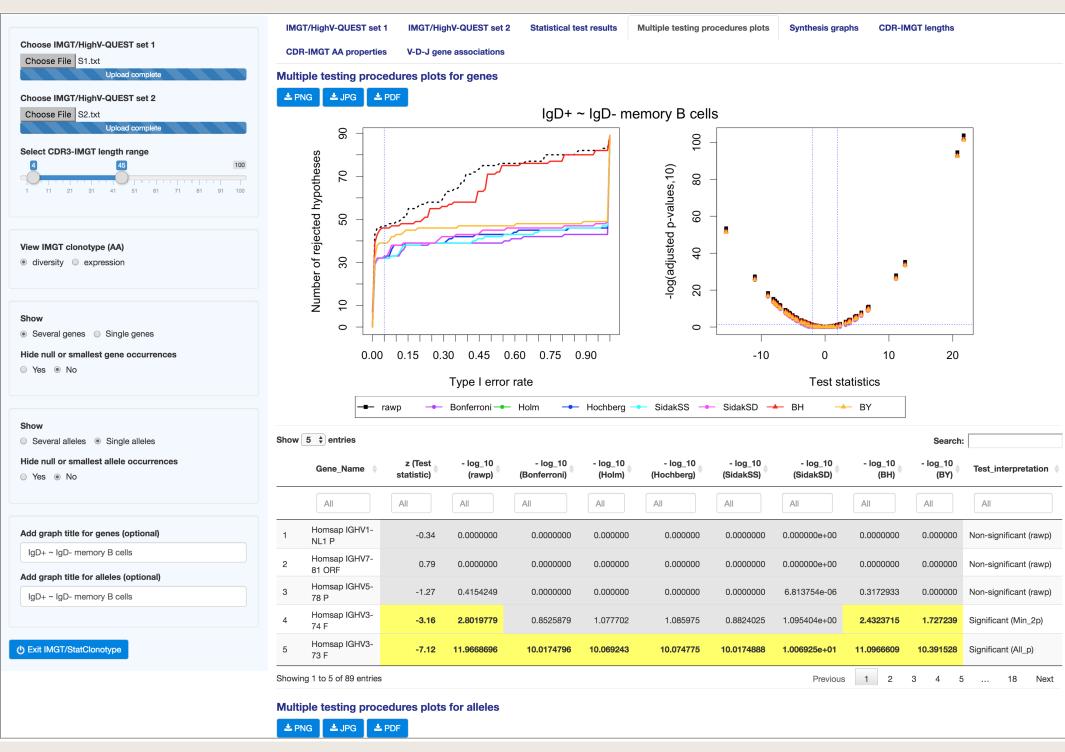
VELCOME! 0 <u>IMGT/StatClonotype</u> he international immunogenetics information system®								formation stem®		
	IMGT/HighV-QUEST set 1 IMGT/HighV-QUEST set 2	Statistical test	results Mu	Itiple testing procedures plots	Synthesis graphs	CDR-IMGT I	engths			
Choose IMGT/HighV-QUEST set 1 Choose File S1.txt	CDR-IMGT AA properties V-D-J gene associations									
Upload complete	Statistical test results for genes									
Choose IMGT/HighV-QUEST set 2	Show 5 + entries							Search:		
Choose File S2.txt Upload complete	Gene_Name Gene_Type Nb_IMGT_clonotyp	e_AA.set1 🔶 P	roportion.set1 🗧	Normalized_proportion.set1 \$	Nb_IMGT_clonoty	/pe_AA.set2	Proportion.set2 🔶	Normalize		
elect CDR3-IMGT length range	All All		All	All	All		All	All		
4 45 100 1 11 21 31 41 51 61 71 81 91 100	1 Homsap V gene IGHV1-NL1 P	1	3.606203e-05	0.3606203		1	5.779679e-05			
	2 Homsap V gene IGHV7-81 ORF	1	3.606203e-05	0.3606203		0	0.000000e+00			
iew IMGT clonotype (AA)	3 Homsap V gene IGHV5-78 P	0	0.000000e+00	0.0000000		1	5.779679e-05			
diversity O expression	4 Homsap V gene IGHV3-74 F	128	4.615939e-03	46.1593942		119	6.877818e-03			
how	5 Homsap V gene IGHV3-73 F	27	9.736747e-04	9.7367472		73	4.219165e-03			
Several genes ○ Single genes lide null or smallest gene occurrences	Showing 1 to 5 of 89 entries				Previous 1	2 3	4 5 1	18 Next		
	List of genes with null or small occurrences The list of genes with null or small occurrences is shown below. For th	is list the Fisher's e	act test is used. A	ssociated <i>p</i> -values are displayed in "rawn"	column of the "Statist	cal test results for	genes" table above (see	explanation)		
how Several alleles Single alleles 	Show/Hide Table				Jan State					
null or smallest allele occurrences	Statistical test results for alleles Only alleles of genes having significant differences in proportions validated by all multiple testing procedures are analyzed.									
) Yes 🔘 No	Show 5 + entries	acco by an multiple	cound brocedules	, alo andiy200.						

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

4 CDR-IMGT length distribution



2 Multiple testing procedures



Multiple testing procedures plots are for genes and alleles



IMGT/HighV-QUEST set 1

IMGT/HighV-QUEST set 1

IMGT/HighV-QUEST set 1

IMGT/HighV-QUEST set 2

Statistical test results

Multiple testing procedures plots

Synthesis graphs

CDR-IMGT lengths

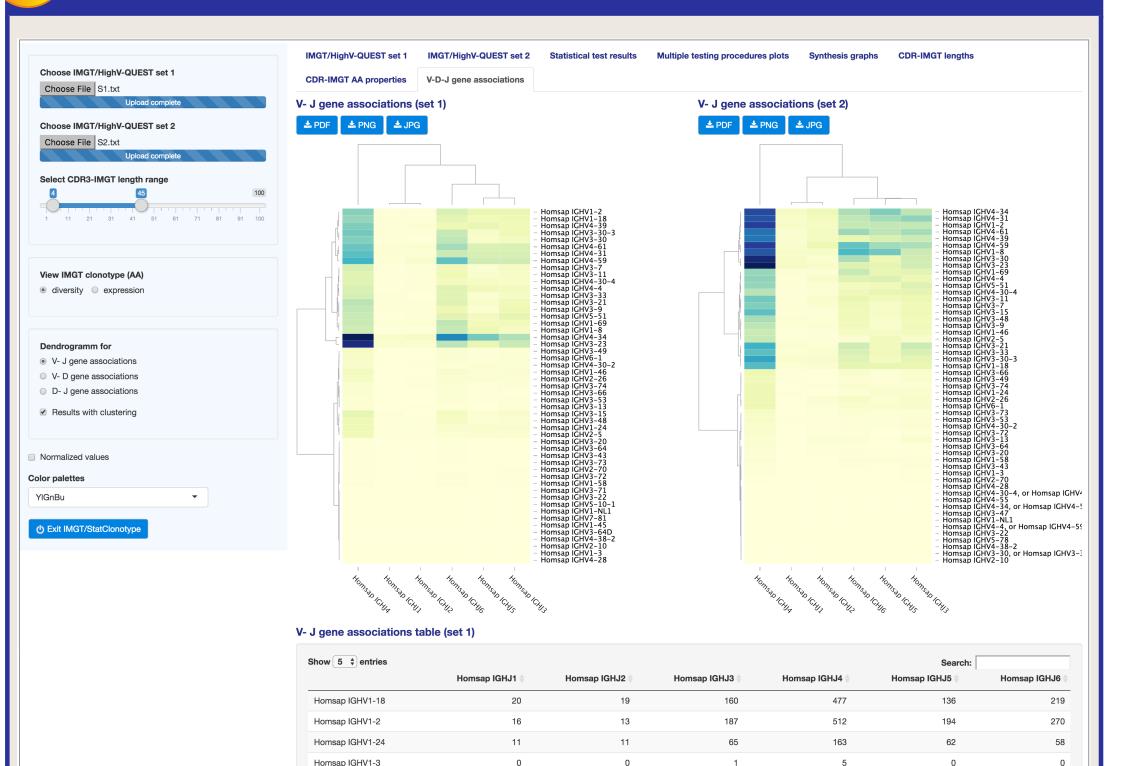
V-D-J gene associations

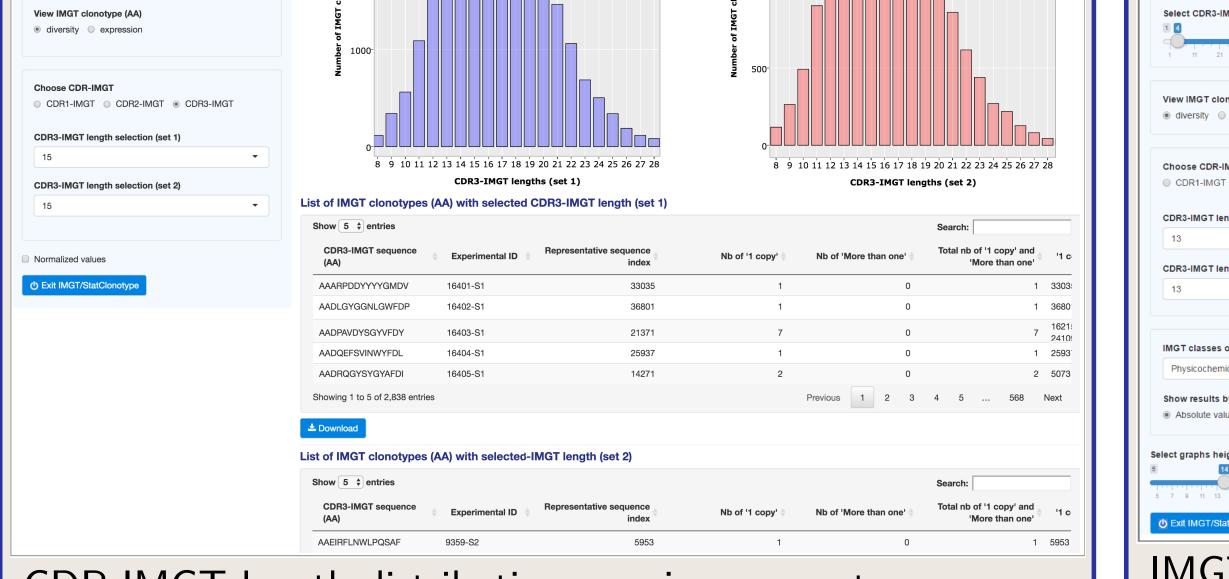
Frequencies of IMGT amino acid Physicochemical classes at CDR3-IMGT positions (set 1)

Show/Hide Table

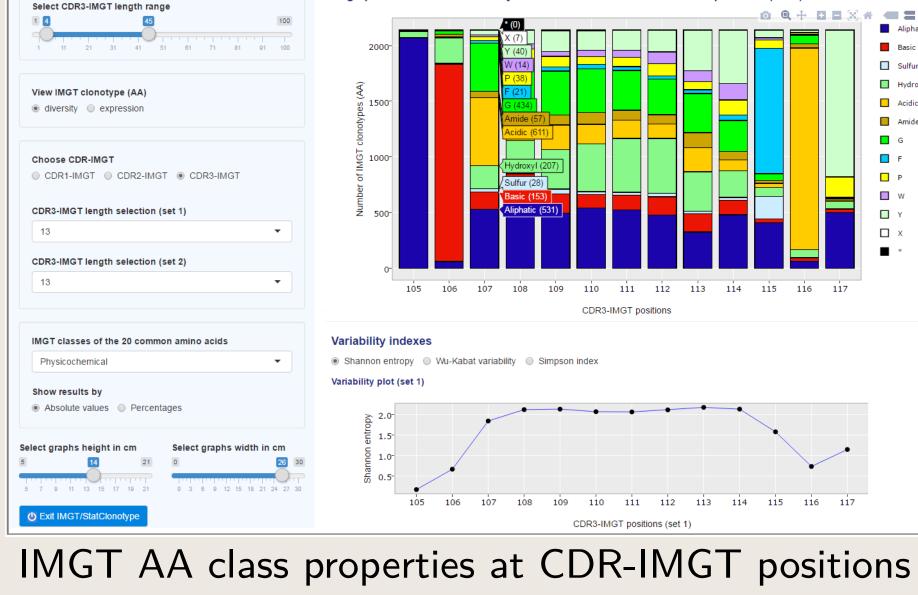
Synthesis graphs are for IG or TR genes and alleles.

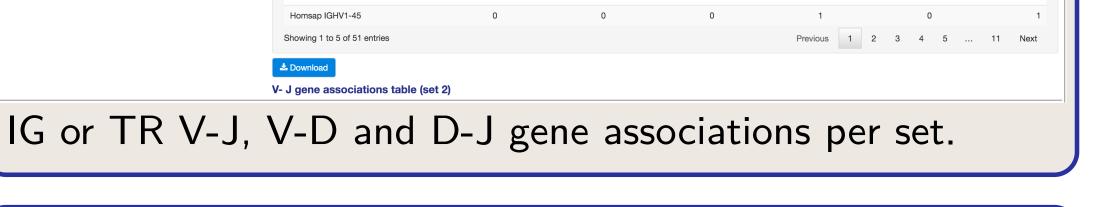
V-J, V-D and D-J gene associations





CDR-IMGT length distribution are given 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.

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



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