EMBnet.journal
Bioinformatics in Action

Tran et al. (2021) EMBnet.journal 26(Suppl A), e953
http://dx.doi.org/10.14806/ej.26.A.953

In silico characterisation of the gene repertoires of immunoglobulins and T cell receptors of the various inbred laboratory strains of *Mus musculus*

Anna Tran✉, **Géraldine Folch, Véronique Giudicelli, Marie-Paule Lefranc, Sofia Kossida**

IMGT®, The International ImMunoGeneTics Information System®, Institute of Human Genetics (IGH), Scientific Research National Center (CNRS), University of Montpellier, Montpellier, France

Competing interests: AT none; GF none; VG none; MPL none; SK none

The laboratory mouse is the most widely used animal model in the life sciences for the study of disease and human development. Mouse strains are known for their differences in the adaptive immune response, but the genomic repertoires of genes that code for antigen receptors, immunoglobulins or antibodies (IG) and T cell receptors (TR) are far from having been fully and precisely sequenced and/or characterised in each strain despite the existence of Mouse Genome Informatics resources dedicated to the species.

IG (proteins composed of two heavy chains or IGH, and two light chains IGK or IGL) and TR (composed of chains alpha and beta, or chains gamma and delta) are encoded by four types of genes, variable (V), diversity (D), joining (J), constant (C) belonging to multigene families and are very polymorphic. The synthesis of these molecules results from complex mechanisms, including rearrangements of the V, D and J genes at the DNA level, the mechanisms of N-diversity and, for IGs, of somatic hypermutations. These mechanisms are at the origin of an extreme diversity of IG and TR (potentially more than 2.1012 IG and 2.1012 different TR per individual) and the effectiveness of the adaptive immune system.

Knowing and understanding the organisation of these repertoires in the different strains is therefore essential for understanding the reactions of the adaptive immune system and for the choice of mouse models in biology. For example, on IGH locus, the most widely used inbred strains C57BL/6 and BALB/c have only a few sequences in common, which means that their IGH locus are probably a mosaic of very disparate genes. It is highly probable that the same holds true for the loci of other inbred strains of mice. It is important to document this diversity to understand the variation within as well between strain models of antibody-mediated diseases, among other things.

IMGT®, the international ImMunoGeneTics information system®[1], is a unique source of knowledge in immunogenetics and immunoinformatics and is recognised as the international reference. IMGT® engages in the precise and detailed characterisation of the IG and TR loci by mouse strain according to IMGT® standards to establish their genomic repertoires and allow their comparison. The work carried out during thisthesis aims to design and develop and/or adapt high-performance software tools and a methodology which implement the standards and carry out the annotation of the loci IG and TR of the mouse strains with a "Gold standard" quality (equivalent to the manual annotation). This will allow enrichment of IMGT® databases and implementation of strain-specific research and analysis in IMGT® software tools.

Acknowledgements

We thank all members of the IMGT® team for their expertise and constant motivation. IMGT® was funded in part by the BIOMED1 (BIOCT930038), Biotechnology BIOTECH2 (BIO4CT960037), 5th PCRDT Quality of Life and Management of Living Resources (QLG2 2000 01287), and 6th PCRDT Information Science and Technology (ImmunoGrid, FP6 IST 028069) programmes of the European Union (EU). IMGT® received financial support from the GIS IBiSA, the Agence Nationale de la Recherche (ANR) Labex MabImprove (ANR 10 LABX 53 01), the Région Occitanie Languedoc Roussillon (Grand Plateau Technique pour la Recherche (GPTR), BioCampus Montpellier. IMGT® is currently supported by the Centre National de la Recherche Scientifique (CNRS), the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI), the University of Montpellier, and the French Infrastructure Institut Français de Bioinformatique (IFB) ANR 11 INBS 0013. IMGT® is a registered trademark of CNRS. IMGT® is member of the International Medical Informatics Association (IMIA) and a member of the Global Alliance for Genomics and Health (GA4GH). This work was granted access to the High Performance Computing (HPC) resources of Meso@LR and of Centre Informatique National de l'Enseignement Supérieur (CINES) and to Très Grand Centre de Calcul (TGCC) of the Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA) under the allocation 036029 (2010

[1]http://www.imgt.org

2021) made by GENCI (Grand Equipement National de Calcul Intensif).

References

1. RS Sellers, CB Clifford, PM Treuting, and Cory Brayton. Immunological variation between inbred laboratory mouse strains: points to consider in phenotyping genetically immunomodiffed mice. Veterinary pathology, **49**(1):32-43, 2012. http://dx.doi.org/10.1177%2F0300985811429314
2. Cynthia L Smith, Judith A Blake, James A Kadin, Joel E Richardson, Carol J Bult, and Mouse Genome Database Group. Mouse genome database (mgd)-2018: knowledgebase for the laboratory mouse. Nucleic acids research, **46**(D1):D836{D842, 2018. http://dx.doi.org/10.1093/nar/gkx1006
3. Corey T Watson, Justin T Kos, William S Gibson, Leah Newman, Gintaras Deikus, Christian E Busse, Melissa L Smith, Katherine JL Jackson, and Andrew M Collins. A comparison of immunoglobulin ighv, ighd and ighj genes in wild-derived and classical inbred mouse strains. Immunology and cell biology, **97**(10):888-901, 2019. http://dx.doi.org/10.1111/imcb.12288
4. Marie-Paule Lefranc, Véronique Giudicelli, Patrice Duroux, Joumana Jabado-Michaloud, Géraldine Folch, Safa Aouinti, Emilie Carillon, Hugo Duvergey, Amélie Houles, Typhaine Paysan-Lafosse, *et al.* Imgt®, the international immunogenetics information system® 25 years on. Nucleic acids research, **43**(D1):D413–D422,2015 http://dx.doi.org/10.1093/nar/gku1056