



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



École Doctorale BioSE (Biologie-Santé-Environnement)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

**DOCTEUR DE L'UNIVERSITE DE LORRAINE**

Mention : « Sciences de la Vie et de la Santé »

**Axel R.S.X. DUBOIS**

# Convergent antibody signatures for the measles virus in transgenic rats expressing a human B cell IG repertoire

17<sup>th</sup> December 2015

Membres du jury:

**Prof Dr Sofia KOSSIDA**, Department of Molecular Bases of Human Disease, Institute of Human Genetics, CNRS – University of Montpellier 2 (France) - Rapporteur

**Prof Dr Alain VANDERPLASSCHEN**, Laboratory of Immunology-Vaccinology, Faculty of Veterinary Medicine, University of Liège (Belgium) - Rapporteur

**Prof Dr Claude P. MULLER**, Department of Immunology, Luxembourg Institute for Health/National Public Health Laboratory (Grand-Duchy of Luxembourg) - Examineur

**Dr Véronique VENARD**, Laboratory of Virology, Hôpital Adultes, University of Hospital de Nancy-Brabois (France) - Examineur

**Prof Dr Marcelo DE CARVALHO**, Laboratory of Immunology, Faculty of Medicine, University of Lorraine & University Hospital of Nancy (France) - Membre invité

---

Department of Infection and Immunity, Luxembourg Institute for Health (LIH), 29 rue Henry Koch,  
L-4354 Esch-Sur-Alzette, Grand Duchy of Luxembourg

# Convergent antibody signatures for the measles virus in transgenic rats expressing a human B cell IG repertoire

Dissertation Submitted In Partial Fulfillment of the Requirements for the Degree of Doctor of  
Philosophy



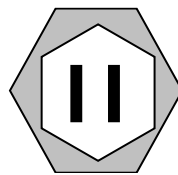
**AXEL R.S.X. DUBOIS**

September 2015

Supervisor:

Professor and Head of Laboratory

**Prof. Dr. Claude P. Muller**



**Department of Infection and Immunity**

WHO European Regional Reference Laboratory for Measles and Rubella  
WHO Collaborating Centre for Reference and Research on Measles Infections  
Luxembourg Institute for Health (LIH)

And

Laboratoire National de Santé, Luxembourg

In loving memory of Nathalie RODENBOUR,

## ***Acknowledgements***

I am proud to present here parts of 4 years of hard work and dedication. A large part of the labor performed during those years could not be included. I have lived through the development of this project, beginning with the original proposal through its development into multiple PhD's and a whole new field of activity for our lab. It was a challenge I was happy to take on. I hope for these data to largely benefit the department and other PhD students.

First, I would like to express my deepest gratitude to my supervisor and mentor, Professor Dr. Claude P. Muller, for his excellent guidance, encouragements, thrilling scientific discussions, and for providing me with an excellent atmosphere for doing research. Claude is an amazingly hard worker and has been a driving force in the development of biological research in the Grand Duchy of Luxembourg. Working in his laboratory was an intense experience. Thank you Claude for giving me the opportunity to work on this great project and to live my passion for Science.

All of my appreciation to my co-promotor at the Université de Lorraine, Dr. Véronique Venard, for her encouragements and caring. I enjoyed the -too few- moments spent with you.

I had the great chance and honor to be surrounded by great colleagues. I am indebted to all of you for your help and making my stay at the department of Immunology an unforgettable moment. I wish I could name each and every one of you here. I am in particular beholden to the members of the B cell team, Jean-Philippe Bürckert and Josiane Kirpach. Jean-Philippe, your help and informatics skills were an invaluable add-on to this work. Thanks to all the post-docs, technicians and bioinformaticians for sharing their experience and expertise with me. In particular, thanks to Sophie Farinelle and Stéphanie Willieme for your help of all moments. Special mention to my dear friends Dr. Sara Vernocchi and Sebastien de Landtsheer. The staff of the Pasteur institute in Vientiane has also played a significant role in this project and has all of my gratitude.

This thesis would not have been possible without the financial support from the Aides à la Formation-Recherche and the Fonds National de la Recherche du Luxembourg.

Thanks to my family, brothers, sisters, nieces and nephews, and especially my parents for their unconditional support. I cherish you.

Thanks also to all the opponents to this project. You made me stronger.

Last but not least, I want express my appreciation to my better half, Vincent. You have been more than patient during these 4 years. Thanks for always showing me the colors.

# *Abstract*

The development of a specific antibody response is driven by the proliferation of selected B cell clones and the affinity maturation of their immunoglobulin (IG) repertoire by somatic hypermutation. The complementarity-determining region 3 (CDR3) of the variable domain of the immunoglobulin heavy chain (VH) is the major antigen binding domain and determinant of antigen-specificity of the antibody molecule. Using next-generation sequencing, we tested whether immunization resulted in the generation and accumulation of similar VH CDR3 sequences that are antigen-specific and shared across individuals, i.e. public VH CDR3. Such public "antigen-specific signatures" can potentially be used to retrospectively reconstruct past antigenic challenges. We investigated the response of the IG repertoire to measles virus (MV)-derived antigens in transgenic rats with fully functional human IG heavy and light chain loci (OmniRat™). First, OmniRats were immunized with an inactivated MV. Lymph nodes were used to generate a hybridoma library and MV-specific hybridoma clones. Using an Ion Torrent next-generation sequencing platform, we analyzed the IG repertoire in the organs, in the hybridoma library and the hybridoma clones. In the hybridoma clones, 23 MV-specific CDR3 amino acid sequences ("signatures") were identified. We next dissected the IgG repertoires of the lymphoid organs of OmniRats immunized with a viral vector carrying the MV haemagglutinin (H) and fusion (F) glycoproteins, as well as various other MV-unrelated antigens. We demonstrated a strong public immune response within the different groups of rats characterized by convergent VH CDR3 amino acid sequences in the animals that received the same vaccine. These clusters of CDR3 represent complex antigen-specific VH CDR3 signatures. Some signatures for the MV H and F proteins were identical or highly similar to the VH CDR3 of MV-specific hybridoma clones, confirming that these immunoglobulins are not only antigen-triggered but also antigen-specific. In a last set of experiments, OmniRats were immunized by DNA vaccination against the haemagglutinin protein of the most genetically distant strains of the MV in order to identify MV strain-specific signatures from their B cell IG repertoire. We are now transposing this concept to human studies by performing measles infection and vaccination follow ups to test whether MV-specific CDR3 signatures can also be found in peripheral blood B cells. In the context of the MV eradication campaign of the World Health Organization, new epidemiological tools that enable to distinguish between immunized and infected individuals would be valuable assets.

# *Thesis summary*

The ability of the humoral immune system to mount a specific response to an almost infinite number of antigens relies on a diverse repertoire of antigen-specific receptors at the surface of B cells, the surface immunoglobulins (IG) or B cell receptors (BcR). The diversity of the IG repertoire (also called BcR repertoire) is generated during B cell ontogeny by the step-wise and uneven junction of Variable (V), Diversity (D; heavy chain only), and Joining (J) genes that encode for the variable domains of the IG heavy (IGH) and light chains. At the core of the V-(D)-J junction on the rearranged IGH locus is the highly variable Complementary Determining Region 3 (CDR3). The VH CDR3 encodes for the region at the center of the antigen-binding site of the IG molecule and is the major determinant of its antigen-specificity. Further diversity in the IG repertoire is added by somatic hypermutation (SHM) of the IG loci following antigen-recognition and B cell clonal expansion in secondary lymphoid organs. The CDR3 is preferentially affected by replacement mutations that slightly change the conformation of the antigen-binding sites of the IG molecule and can affect its specificity toward the antigen. B cells with a BcR of increased specificity are preferentially selected to expand, resulting in affinity maturation of the IG. Antigen-specific B cells ultimately differentiate into plasma cells that gain the bone marrow where they long-lastly produce antigen-specific IG, or in memory B cells that allow a faster IG response upon subsequent exposure to the same antigen. The IG repertoire is constantly modeled by antigens and contains the record of past antigen exposures of the individual.

The astronomical number of possible BcR largely exceeds the total number of B cells in the human body. The diversity of the IG repertoire is difficult to assess using tradition techniques e.g. B cell immortalization and Sanger sequencing. The recent advances in Next-Generation Sequencing (NGS) technologies make it now possible to study a large number of BcR in a sample at a high-depth. The investigation of the IG repertoire is of fundamental interest and has found a wide range of applications in fundamental and clinical immunology. The study of its shaping by particular antigen can notably enable us to gain better insight into the pathophysiology of certain infections and help to develop improved therapies or vaccines. The IG repertoire can also be a source of novel biomarkers for particular infections or immunizations. Recent studies indeed observed convergent IG signatures in groups of patients recovering from acute dengue infection as well as in vaccinees with a conjugate polysaccharide vaccine or against H1N1 influenza virus. These signatures may be used for the sequence-based monitoring of vaccines or infectious diseases, or help in the diagnosis or prognostic of particular conditions.

Although effective vaccines are available, measles remains one of the leading causes of childhood mortality worldwide and is undergoing resurgence in many countries. Measles is a highly contagious disease caused by a Morbillivirus, the Measles Virus (MV). There exist 8 clades of MV further subdivided into 24 genotypes (or strains). Infection with any strain of MV confers cross-protection to all others, and it is so far impossible to retrospectively determine the strain responsible for the immunity. Some studies nevertheless suggest certain variability in the antigenic properties of the different strains. For instance, sera from immune donors differ in their ability to neutralize *in vitro* different MV strains. It is therefore to anticipate that different MV strains differentially impact the IG repertoire and eventually leave on it a specific IG signature. In the context of the MV eradication campaign of the World Health Organization (WHO), new epidemiological tools that enable to distinguish between immunized and infected individuals would be valuable assets.

The analysis of the human IG repertoire to provide evidence of convergent BcR evolution in individuals exposed to the same antigens and identify antigen-specific signatures directly in human subjects can be particularly challenging. Human studies are largely restricted to peripheral blood where only a small amount of the IG repertoire diversity can be sampled, and it is difficult to predict when the B cells of interest are sufficiently expanded in the blood to be detected. Large numbers of participants enrolled in extensive longitudinal studies are required to detect convergent IG responses and identify reliable IG signatures. Human studies are also limited to certain antigens. Here we propose an innovative approach that allows a more flexible identification of human antigen-specific VH CDR3 signatures, notably for the different MV strains. This approach relies on transgenic rats carrying human IG loci and expressing a diversified repertoire of chimeric human/rats IG with fully human idiotypes (OmniRat™). Upon antigen challenges, these animals mount antigen-specific IG responses at levels similar to wild type animals and have been used to produce human monoclonal antibodies (mAbs) against diverse antigens. In order to investigate the IG response to the MV and determine whether MV strain-specific IG signatures can be found in the IG repertoire, OmniRats were immunized against various targets including MV antigens. We dissected the post-vaccinal IgG repertoires in diverse lymphoid organs using the Ion Torrent PGM™ NGS platform. Our work was built into 3 subsequent steps whose results converged toward the identification of VH CDR3 signatures specific for the MV haemagglutinin (H) protein.

In a first set of experiments (Chapter 3), OmniRats were immunized with MV antigens and their regional lymph nodes used to produce MV-specific hybridoma clones. We analyzed in parallel the IgG repertoire in the lymph nodes and in the bulk hybridoma library generated by polyethylene glycol (PEG)-mediated



fusion of lymphocytes to myeloma cells. Using the sequences of the MV clones as templates, we showed that the fusion is biased towards rare antigen-specific B cells in the lymph nodes that become enriched by orders of magnitude in the hybridoma library. We also gained deep insight into the dynamics of the IG response and showed that MV-specific VH CDR3 amino acid sequences are encoded by a multitude of different V, D and J genes. Moreover we generated a set of IG sequences of known MV-specificity (i.e. MV-specific IGH signatures) to guide us in our subsequent experiments. In a concomitant set of experiments (Chapter 4), groups of OmniRats were immunized against various antigens, including a recombinant Modified Vaccinia Virus encoding for the MV H and Fusion (F) glycoproteins (MVA-MV H/F). We observed that the humoral immune response in the rats exposed to the same antigens converged towards the expression of B cells with highly similar VH CDR3. We implemented a method to identify antigen-specific VH CDR3 signatures without previous selection for antigen-specific cells. The VH CDR3 sequences of 3 MV clones generated in previous experiments (Chapter 3) were found among the signatures that we identified in the group of MVA-MV H/F immunized animals. These observations restrict the specificity of the hybridoma clones to the MV H or F proteins and demonstrate the ability of our method to select for antigen-specific IG signatures. In a last set of experiments (Chapter 5), OmniRats were immunized against the H protein of 2 genetically distant strains of the MV by genetic immunization. Our preliminary results indicate that the same phenomenon of convergent evolution occurs in these rats. Highly similar VH CDR3 to the same 3 MV-specific hybridomas were found in the bone marrow samples of the MV immunized rats, independently of the H strain. This indicates that these VH CDR3 are strong markers for the MV H protein, but do not enable to differentiate between the vaccine strain of MV and the D11 genotype. We also identified clusters of homologous VH CDR3 present only in the animals immunized with MV clade A or D11 genotype. The antibodies encoding for these CDR3 are likely directed against immunodominant epitopes differing between the two MV strains.

We studied for the first time at unprecedented depth the IG response to various antigens in the lymphoid organs of an innovative animal model with a human IG repertoire. A set of putative fully-human VH CDR3 signatures has been created for MV antigens as well as other bacterial or chemically-defined antigens. These VH CDR3 signatures need now to be validated in human subjects. This can notably be achieved by screening them against human IG datasets from MV infected and vaccinated donors. If validated in human subjects, these signatures can be used as biomarkers for MV infection or vaccination. Potentially, transgenic humanized animals can be used to identify IG signatures for a large number of antigens.

I believe that in the near future the IG repertoire will become a source of markers for various infectious diseases or vaccines, in an approach that has the potential to retrospectively reconstruct the immune history of an individual.

## *Résumé en français*

Le système immunitaire humoral a la capacité de réagir spécifiquement envers un nombre presque infini d'antigènes grâce à un répertoire très diversifié de récepteurs spécifiques situés à la surface des lymphocytes B : les immunoglobulines (IG) ou récepteurs des lymphocytes B (B cell receptors – BcR). La diversité du répertoire des immunoglobulines (répertoire IG, aussi appelé répertoire BcR) est générée durant l'ontogenèse des lymphocytes B par la jonction séquentielle et irrégulière des gènes Variable (V), Diversité (D ; chaîne lourde uniquement) et Jonction (J), qui encodent les domaines variables (VH et VL) des chaînes lourdes (IGH) et légères (IGK ou IGL) de l'immunoglobuline. Au cœur de la jonction V-(D)-J sur le locus IGH réarrangé se trouve la très variable *région déterminant la complémentarité 3* (Complementary Determining Region 3 – CDR3). La région CDR3 de la chaîne lourde (VH CDR3) encode pour la région de l'anticorps au cœur du site de reconnaissance pour l'antigène et est le déterminant majeur de la spécificité antigénique de l'immunoglobuline. La diversité du répertoire IG est encore augmentée par hypermutation somatique des régions codant pour les domaines VH et VL suite à la reconnaissance d'un antigène et à l'expansion clonale des lymphocytes B dans les organes lymphoïdes secondaires. La région CDR3 est principalement affectée par des mutations qui changent légèrement la conformation des sites de liaison à l'antigène de l'immunoglobuline et peuvent améliorer sa spécificité. Les lymphocytes B dont l'immunoglobuline est d'une spécificité plus grande envers l'antigène sont sélectionnés préférentiellement (maturation d'affinité des immunoglobulines). Les lymphocytes B spécifiques à l'antigène finissent par se différencier en cellules plasmiques (ou plasmocytes) qui migrent vers la moelle osseuse où elles produisent sur de longues périodes des IG spécifiques à l'antigène, ou en cellules B mémoire qui permettent de produire une réponse IG rapide lors des expositions subséquentes au même antigène. Le répertoire IG est constamment modelé par des antigènes et porte l'historique des expositions antigéniques passées de l'individu.

Le nombre astronomique d'IG possibles dépasse largement le nombre de lymphocytes B dans le corps humain. La diversité du répertoire IG est difficile à estimer à l'aide des techniques traditionnelles, telles l'immortalisation de lymphocytes B (production d'hybridomes) et techniques traditionnelles de séquençage (Sanger). Cependant, les avancées récentes en matière de séquençage de dernière génération (Next-Generation Sequencing – NGS) permettent désormais l'étude en profondeur d'un grand nombre de BcR différents dans le même échantillon. L'étude du répertoire IG est d'un intérêt fondamental et a trouvé un large éventail d'applications en immunologie clinique comme fondamentale.

L'analyse de son façonnage par des antigènes spécifiques peut en particulier nous apporter une meilleure compréhension de la pathophysiologie de certaines infections et permettre le développement de thérapies ou de vaccins plus efficaces. Potentiellement, le répertoire IG peut également être une source de nouveaux biomarqueurs pour des infections ou vaccinations particulières. Cela suppose que des molécules d'IG identiques ou similaires soient impliquées chez différents individus exposés au même antigène, ou en d'autres mots que la réponse IG soit au moins partiellement publique/partagée. Des études récentes ont observé des réponses humorales convergentes au sein de groupes de patients se remettant d'une infection aiguë de dengue, ainsi que chez des personnes immunisées avec un vaccin polysaccharidique conjugué ou contre le virus de la grippe H1N1. Ainsi, l'analyse en profondeur du répertoire IG afin de détecter une réponse publique peut être utilisée comme outil pour le monitoring de vaccins et maladies infectieuses, ou aider dans le diagnostic ou pronostique de certaines pathologies.

Bien que des vaccins efficaces soient disponibles, la rougeole reste une des causes principales de mortalité infantile dans le monde, et est en train de connaître une résurgence dans de nombreux pays. La rougeole est une maladie hautement contagieuse causée par un Morbillivirus, le virus de la rougeole (Measles Virus – MV). Il existe 8 sous-groupes génétiques ou clades (A-H) de MV, elles-mêmes divisées en 24 souches ou génotypes. L'infection par n'importe laquelle des souches de MV confère une protection contre toutes les autres. Il est donc à ce jour impossible de déterminer rétrospectivement la souche à l'origine de l'immunité. Certaines études ont néanmoins démontré une certaine variabilité dans les propriétés antigéniques des différentes souches. Par exemple, les sérums de donneurs immunisés (vaccination ou infection naturelle par une souche sauvage) diffèrent dans leur capacité à neutraliser diverses souches *in vitro*. Il est par conséquent supposé que les différentes souches de MV impactent le répertoire IG de manière différenciée, et y laissent une signature spécifique qui permettrait de déterminer rétrospectivement quelle souche a été à l'origine de l'immunité. Dans le contexte de la campagne d'éradication du MV de l'Organisation Mondiale de la Santé (OMS), de nouveaux outils épidémiologiques qui permettraient de distinguer individus immunisés et infecté seraient des atouts de valeur.

Bien que l'existence d'une réponse immunitaire publique, convergente et spécifique à l'antigène ait pu être démontrée chez l'humain, son étude directe sur des sujets humains est limitée par divers obstacles. Les études chez l'humain sont essentiellement limitées au sang périphérique, où seule une petite partie de la diversité du répertoire IG peut être observée, et sont limitées à certains antigènes. Il est également difficile de prédire quand les lymphocytes B d'intérêt seront suffisamment fréquents dans le sang pour être

détectés. Des études longitudinales extensives sur un grand nombre de sujets sont nécessaires à la détection d'une réponse IG commune et à l'indentification de signatures IG fiables. Nous proposons ici une approche innovante qui permet une identification plus flexible de signatures VH CDR3 humaines spécifiques à l'antigène pour les différentes souches du MV. Cette approche fait usage de rats transgéniques porteurs de loci IG humains et qui expriment un répertoire diversifié d'IG chimériques humain/rat aux idiotypes pleinement humains (OmniRats). Confrontés à un antigène, ces animaux montent une réponse IG spécifique à l'antigène similaire à celle d'animaux de type sauvage, et ont été utilisés pour produire des anticorps monoclonaux humains (mAbs) contre divers antigènes. Afin d'explorer la réponse humoral au MV et de déterminer si des signatures spécifiques à une souche spécifique de MV peuvent être trouvées dans le répertoire IG, des OmniRats ont été immunisés avec divers antigènes, dont des antigènes du MV. Nous avons disséqué le répertoire IG (IgG) après immunisation dans divers organes lymphoïdes à l'aide de la plateforme NGS Ion Torrent PGM. Notre étude s'est construite en trois étapes successives dont les résultats ont convergé vers l'identification de signatures VH CDR3 humaines spécifiques à la protéine hémagglutinine (H) du MV.

Lors d'une première série d'expériences (Chapitre 3), des OmniRats ont été immunisés avec des antigènes du MV entier et leurs ganglions lymphatiques régionaux utilisés pour produire des clones d'hybridomes MV-spécifiques. Nous avons analysé en parallèle le répertoire IgG dans le ganglion lymphatique et dans la librairie d'hybridomes générée par fusion au polyéthylène glycol (PEG) de lymphocytes à des cellules de myélome. En prenant les séquences des clones MV comme base, nous avons montré que la fusion est biaisée en faveur de population de lymphocytes B rares dans le ganglion lymphatique qui se retrouvent fortement enrichies dans la librairie d'hybridomes. Nous avons également acquis une vue en profondeur de la dynamique de la réponse IG et montré que les séquences d'acides aminés VH CDR3 sont encodées par une multitude de gènes V, D et J. Nous avons généré un ensemble de séquences IG à la spécificité MV avérée, c'est-à-dire des signatures IG MV-spécifiques à même de nous guider dans les expériences suivantes.

Dans une série d'expériences concomitantes (chapitre 4), des groupes d'OmniRats ont été vaccinés contre divers antigènes, dont un virus de la vaccine modifié (modified vaccinia Ankara virus - MVA) encodant les glycoprotéines de surface hémagglutinine H et fusion F du virus de la rougeole (MVA-MV H/F). Nous avons développé une méthode qui permet d'identifier des signatures VH CDR3 spécifiques à l'antigène sans enrichissement préalable en cellules spécifiques. Nous avons effectivement observé que la réponse immunitaire humorale chez ces rats converge vers l'expression de lymphocytes B avec des

séquences VH CDR3 très homologues parmi les seuls animaux immunisés avec un antigène particulier. Ces signatures VH CDR3 étaient encodées par de multiples V(D)J recombinaisons chez le même individu ou des individus différents, par un processus de recombinaisons convergentes. Les séquences VH CDR3 de 3 clones MV-spécifique générés précédemment (Chapitre 3) ont été trouvées parmi les signatures que nous avons identifiées dans le groupe d'animaux vaccinés avec le MVA-MV H/F. Ces observations restreignent la spécificité des clones d'hybridome aux protéines MV H ou F, et démontrent la capacité de notre méthode de sélectionner des signatures IG spécifiques aux antigènes.

Dans une dernière série d'expériences (Chapitre 5), des OmniRats ont été vaccinés contre la protéine H de 2 souches génétiquement distantes du MV. Nos résultats préliminaires indiquent que le même phénomène d'évolution convergente se manifeste chez ces rats. De façon intéressante, des VH CDR3 présentant une forte homologie avec le même 3 hybridomes spécifiques au MV ont été trouvés dans les échantillons de moelle osseuse de tous les rats vaccinés, indépendamment de la souche de rougeole. Cela indique que ces VH CDR3 sont des marqueurs fiables de la protéine MV H mais ne permettent pas de différencier une souche vaccinale du MV et le génotype D11. Nous avons également identifié des groupes de VH CDR3 fortement homologues présent seulement chez les rat immunisés avec la souche vaccinale A ou le génotype D11. The anticorps associés à ces CDR3 sont certainement dirigés contre des épitopes immunodominants variant entre les deux souches.

Une série de signatures VH CDR3 pleinement humaines a été établie pour des antigènes du MV ainsi que d'autres antigènes bactériens ou chimiques définis. Ces signatures doivent maintenant être validées chez des sujets humains. Nous confrontons actuellement nos signatures à des bases de données de séquences d'IG provenant de donneurs infectés par ou vaccinés contre le MV.

Si validées chez des sujets humains, ces signatures peuvent être utilisées comme biomarqueurs pour la vaccination ou l'infection par le MV. Potentiellement, les animaux transgéniques humanisés peuvent servir à l'identification de signatures IG pour un très grand nombre d'antigènes. Je suis convaincu que dans le futur proche, le répertoire IG deviendra une source de marqueurs pour diverses infections et vaccins, dans une approche qui pourra potentiellement permettre de reconstruire l'histoire immunitaire d'un individu. Les résultats présentés dans cette thèse élargissent également notre connaissance de la réponse IG aux antigènes viraux dans un modèle animal innovateur avec un répertoire de lymphocyte B humain.

# Table of contents

<i>Acknowledgements</i> .....	1
<i>Abstract</i> .....	2
<i>Thesis summary</i> .....	3
<i>Résumé en français</i> .....	7
<i>Table of contents</i> .....	11
<i>List of abbreviations</i> .....	16
<i>List of tables</i> .....	18
<i>List of figures</i> .....	20
<b>Chapter 1. General introduction</b> .....	<b>22</b>
<b>1.1 Immunoglobulins and diversity of the IG repertoire</b> .....	<b>23</b>
1.1.1 Basics of B cell biology.....	23
1.1.2 Structure of an antibody molecule.....	25
1.1.3 Somatic recombination and combinatorial diversity.....	27
1.1.4 Somatic hypermutation and class-switch recombination.....	31
<b>1.2 Analyzing the IG repertoire: new perspectives with NGS</b> .....	<b>34</b>
1.2.1 Conventional techniques for the analysis of the IG repertoire .....	34
1.2.2 Next-generation sequencing: Ion Torrent technology.....	35
1.2.3 Bioinformatics tools for the analysis of IG sequences.....	38
<b>1.3 NGS applications on IG repertoire analysis</b> .....	<b>40</b>
1.3.1 IG repertoire development, composition and diversity .....	40
1.3.2 Modeling of the IG repertoire by infection and vaccination.....	41
<b>1.4 Innovative techniques for the development of therapeutic antibodies</b> .....	<b>44</b>
1.4.1 Antibody production and discovery using NGS technologies .....	44
1.4.2 Transgenic animals expressing human immunoglobulins .....	46

<b>1.5</b>	<b><i>Measles Virus</i></b> .....	<b>49</b>
1.5.1	Virus structure and molecular epidemiology .....	49
1.5.2	Pathogenesis and immune response.....	51
1.5.3	Measles vaccines and eradication efforts .....	54
<b>Chapter 2.</b>	<b>Aims</b> .....	<b>57</b>
<b>Chapter 3.</b>	<b>Development of MV-specific human monoclonal antibodies in transgenic humanized rats</b> .....	<b>59</b>
<b>3.1</b>	<b><i>Abstract</i></b> .....	<b>60</b>
<b>3.2</b>	<b><i>Introduction</i></b> .....	<b>61</b>
<b>3.3</b>	<b><i>Materials and methods</i></b> .....	<b>64</b>
3.3.1	Immunization and generation of hybridoma clones.....	64
3.3.2	Anti-MV IgG ELISA.....	64
3.3.3	Sample preparation, amplification and Ion Torrent Deep Sequencing.....	65
3.3.4	Quality Control and annotation of the deep sequencing results .....	65
<b>3.4</b>	<b><i>Results</i></b> .....	<b>68</b>
3.4.1	Generation and identification of MV-specific hybridoma clones .....	68
3.4.2	Analysis of the IGH repertoire in lymph nodes and hybridoma library by deep-sequencing.....	71
3.4.3	Overlap between sequences encoding identical VH CDR3 amino acid sequences .....	75
3.4.4	MV-associated CDR3 are highly enriched in the MV-HB compared to the MV-LN.....	77
3.4.5	The clonally expanded B cell populations in the MV-LN do not fuse efficiently.....	79
3.4.6	MV CDR3 are associated with different V-(D)-J rearrangements .....	83
<b>3.5</b>	<b><i>Discussion</i></b> .....	<b>88</b>
<b>3.6</b>	<b><i>Acknowledgements</i></b> .....	<b>91</b>



<b>Chapter 4.</b>	<b>Convergent antibody signatures for the MV H and F proteins in transgenic rats expressing a human B cell IG repertoire .....</b>	<b>92</b>
<b>4.1</b>	<b><i>Abstract</i> .....</b>	<b>93</b>
<b>4.2</b>	<b><i>Introduction</i>.....</b>	<b>94</b>
<b>4.3</b>	<b><i>Materials and methods</i>.....</b>	<b>96</b>
4.3.1	Animals and immunizations .....	96
4.3.2	Generation of the Benzo[ $\alpha$ ]Pyrene constructs .....	98
4.3.3	Modified Vaccinia Ankara viruses propagation and purification.....	98
4.3.4	Anti-MV, -MVA, -TT and -B[ $\alpha$ ]P IgG ELISA.....	99
4.3.5	Sample preparation, amplification and Ion Torrent Deep Sequencing.....	99
4.3.6	Data processing.....	100
<b>4.4</b>	<b><i>Results</i>.....</b>	<b>102</b>
4.4.1	Antigen-specific antibody response and next-generation sequencing of the IGH repertoire of immunized OmniRat.....	102
4.4.2	Transgenic rats use sets of identical CDR3 sequences in response to the same antigenic challenges .....	108
4.4.3	Highly similar CDR3 sequences are produced in response to the same antigenic challenge.....	113
4.4.4	Identification of putative antigen-specific human CDR3 signatures.....	115
4.4.5	The set of putative VH CDR3 signatures for the MV H and F proteins contains previously described MV-specific sequences.....	121
4.4.6	Different V-(D)-J rearrangements result in CDR3 signatures.....	122
<b>4.5</b>	<b><i>Discussion</i>.....</b>	<b>123</b>
<b>4.6</b>	<b><i>Acknowledgement</i>.....</b>	<b>125</b>

---

<b>Chapter 5.</b>	<b>DNA vaccination of transgenic humanized IG rats against the MV H protein of two phylogenetically distant strains.....</b>	<b>126</b>
<b>5.1</b>	<b><i>Abstract</i>.....</b>	<b>127</b>
<b>5.2</b>	<b><i>Introduction</i>.....</b>	<b>128</b>
<b>5.3</b>	<b><i>Materials and methods</i>.....</b>	<b>131</b>
5.3.1	Production of DNA vaccines against MV haemagglutinin of clade 1 and clade D11.....	131
5.3.2	Animals and immunizations .....	131
5.3.3	Anti-MV IgG ELISA.....	132
5.3.4	Sample preparation, amplification and Ion Torrent Deep Sequencing.....	132
5.3.5	Quality control and sequence annotation.....	134
5.3.6	Identification of VH CDR3 signatures.....	135
<b>5.4</b>	<b><i>Results</i>.....</b>	<b>136</b>
5.4.1	The MV-specific antibody response is higher in humanized transgenic rats immunized with the wild type H protein .....	136
5.4.2	NGS of the rearranged IGH repertoire.....	140
5.4.3	Convergent IG signatures for conserved epitopes on the MV H protein	142
5.4.4	Strain specific signatures for phylogenetically distant MV strains.....	146
<b>5.5</b>	<b><i>Discussion</i>.....</b>	<b>149</b>
<b>5.6</b>	<b><i>Acknowledgements</i>.....</b>	<b>151</b>

<b>Chapter 6.</b>	<b>General discussion.....</b>	<b>152</b>
6.1	<i>Technical issues of NGS technologies.....</i>	<i>154</i>
6.2	<i>Transgenic IG humanized animals: innovative organisms for the study of the B cell IG repertoire.....</i>	<i>158</i>
6.3	<i>Insight into the hybridoma fusion process.....</i>	<i>160</i>
6.4	<i>Immune history using NGS-derived signatures.....</i>	<i>161</i>
<b>Chapter 7.</b>	<b>Concluding remarks &amp; future perspectives .....</b>	<b>166</b>
	<i>References.....</i>	<i>168</i>

## *List of abbreviations*

AA	amino acid	GALT:	gut-associated lymphoid tissue
BALT:	bronchial-associated lymphoid tissue	HCV:	Hepatitis Virus C
BcR:	B cell receptor (IG+ CD79A/CD79B coreceptors)	Hib:	<i>Haemophilus influenzae</i>
BM:	bone marrow	HIV:	Human Immunodeficiency Virus
bp:	base pair	HTS:	high-throughput sequencing
cDNA:	complementary DNA	IG:	immunoglobulin
CDR:	complementary determining region	IGH:	immunoglobulin heavy
CH:	constant domain of the immunoglobulin heavy chain	IGL:	immunoglobulin lambda
CL:	constant domain of the immunoglobulin light chain	IGK:	immunoglobulin kappa
CSR:	class-switch recombination	IMGT:	IMGT®, the international ImMunoGeneTics Information System®
D:	diversity gene	IP:	intraperitoneal
DNA:	deoxyribonucleic acid	ISFET:	Ion-sensitive field effector transistor
ELISA:	enzyme-linked immunosorbent assay	J:	joining gene
ELISPOT:	enzyme-linked immunospot	LC-MS/MS:	liquid chromatography-tandem mass spectrometry
EPT:	endpoint titers	LPS:	lipopolysaccharide
ES:	embryonic stem	MACS:	magnetically activated cell sorting
Fab:	fragment antigen-binding	MALT:	mucosa-associated lymphoid tissue
FACS:	fluorescence-activated cell sorting	MenC:	Meningococcal type C polysaccharides
Fc:	fragment crystallizable	MID:	multiplex identifier
FR:	framework region		

---

mRNA:	messenger ribonucleic acid	PEG:	polyethylene glycol
MV CDR3:	CDR3 amino acid sequence identical to a MV-specific hybridoma clone	PFU:	plaque-forming unit
MV:	measles virus	PGM™:	Personal Genome Machine
MVH-A:	haemagglutinin of the MV clade A	P nt:	palindromic nucleotide
MV- D11:	haemagglutinin of the MV genotype D11	RAG:	recombination activating gene
MVH-HB:	hybridoma library derived from MV-immunized animals	RSS:	recombination signaling sequences
MV-LN:	lymph nodes from MV-immunized animals	RT:	reverse transcription
NGS:	next generation sequencing	SHM:	somatic hypermutation
NI-LN:	lymph node from unimmunized animal	sIG:	surface immunoglobulin
N nt:	nontemplated nucleotide	SP:	spleen
nt:	nucleotide	ssDNA:	single-stranded DNA
PBMC:	peripheral blood mononuclear cell	TdT:	terminal deoxynucleotidyl transferase
PCR:	polymerase chain reaction	V:	variable gene
		VH:	variable domain of the immunoglobulin heavy chain
		VL:	variable domain of the immunoglobulin light chain

# *List of tables*

## **Chapter 1: Introduction**

**Table 1:** NGS studies of the IG repertoire applied to vaccines or infection diseases in human

## **Chapter 3: Development of MV-specific human monoclonal antibodies in transgenic humanized rats**

**Table 1:** Primer list

**Table 2:** Characteristics of the IGH chimeric human-rat gamma chain of the 23 MV-specific hybridoma clones sequenced by Sanger sequencing

**Table 3:** Ion Torrent™ output - read numbers before and after quality controls

**Table 4:** V and J gene usage – Coefficient of determination  $R^2$

**Table 5:** Overlap of sequence reads encoding identical CDR3 AA sequences

**Table 6:** Frequency of sequences encoding a MV-specific CDR3 AA in MV-LN and MV-HB

## **Chapter 4: Convergent antibody signatures for the MV H and F proteins in transgenic rats expressing a human B cell IG repertoire**

**Table 1:** OmniRats involved in the study

**Table 2:** Primer list

**Table 3:** Number of mismatches allowed by CDR3 length

**Table 4:** Reciprocal endpoint titers (1:) at sacrifice

**Table 5:** Ion Torrent read counts before and after quality controls – Spleen

**Table 6:** Ion Torrent read counts before and after quality controls – Bone Marrow

**Table 7:** Unique CDR3 AA sequences shared between rats in the same immunization group only

**Table 8:** List of selected CDR3 shared by rats in the MVA-MV H/F, the MVA and both groups combined

**Table 9:** Highly similar CDR3 are found with increased frequencies in the different groups of rats and can be grouped into antigen-specific signatures

**Table 10:** Putative human VH CDR3 signatures for the MV H or F glycoproteins

**Table 11:** Sequence logos for the putative MV-signatures

**Table 12:** Sequence logos for the most prominent putative TT-signatures

## **Chapter 5: DNA vaccination of transgenic humanized IG rats against the MV H protein of phylogenetically distant strains**

**Table 1:** Primer list

**Table 2:** Endpoint titers

**Table 3:** Ion Torrent read counts

**Table 4:** Convergent CDR3 for conserved epitopes on the MV H protein

**Table 5:** 84%-related CDR3 amino acid sequences to the MV CDR3 M

**Table 6:** Putative signatures for the H protein of the MV genotype A

**Table 7:** Putative signatures for the H protein of the MV genotype D11

# *List of figures*

## **Chapter 1: Introduction**

*Figure 1:* Clonal deletion and clonal selection

*Figure 2:* Structure of the IgG molecule

*Figure 3:* V, D and J genes are joined by somatic recombination

*Figure 4:* The 12/23 rule ensures proper rearrangement of V-region genes

*Figure 5:* B cells mature in the bone marrow from lymphoid progenitors

*Figure 6:* Schematic representation of the germinal center reaction

*Figure 7:* Ion Torrent workflow

*Figure 8:* OmniRat's constructs and transgenics

*Figure 9:* Schematic representation of the measles virus and its genome

*Figure 10:* Global distribution of measles genotypes and measles incidence in 2009

*Figure 11:* Basic pathogenesis of MV infection

## **Chapter 3: Development of MV-specific human monoclonal antibodies in transgenic humanized rats**

*Figure 1:* Study design

*Figure 2:* MV ELISA

*Figure 3:* IGHV and IGHJ gene usage in the NI-LNs, MV-LN and MV-HB

*Figure 4:* IGH V and J combinations in the MV-HB, MV-LN and NI-LNs

*Figure 5:* Comparison of the frequencies of the nt sequences for each CDR3 AA sequences from the MV-HB and the MV-LN

*Figure 6:* Frequency distribution of the clonotypes from the MV-LN (A) and MV-HB (B)



**Figure 7:** Comparison of the frequencies of each clonotypes in the MV-HB and MV-LN

**Figure 8:** Different V-(D)-J rearrangements, V genes and J genes are associated with identical MV CDR3 in the MV-HB

**Figure 9:** Identical MV CDR3 are associated with different V-(D)-J rearrangements

**Figure 10:** Frequency (%) of the different V genes associated with each MV CDR3

**Figure 11:** B cell network before and after fusion

#### **Chapter 4: Convergent antibody signatures for the MV H and F proteins in transgenic rats expressing a human B cell IG repertoire**

**Figure 1:** Immunization scheme

**Figure 2:** ELISAs

**Figure 3:** V gene usage (%) in the spleen (A) and bone marrow (B)

**Figure 4:** Overlap of sequence reads encoding identical CDR3 AA sequences

**Figure 5:** Immunized OmniRats express closely related VH CDR3 AA sequences

**Figure 6:** OmniRats express an overlapping repertoire of highly similar CDR3

**Figure 7:** Different V-(D)-J rearrangements result in CDR3 signatures

#### **Chapter 5: DNA vaccination of transgenic humanized IG rats against the MV H protein of phylogenetically distant strains**

**Figure 1:** Immunization scheme

**Figure 2:** Library preparation according to the UID method

**Figure 3:** MV hemagglutinin genes from MV clade A and genotype D11

**Figure 4:** Anti-MV IgG ELISAs

# **Chapter 1.**

# **General introduction**

# 1.1 Immunoglobulins and diversity of the IG repertoire

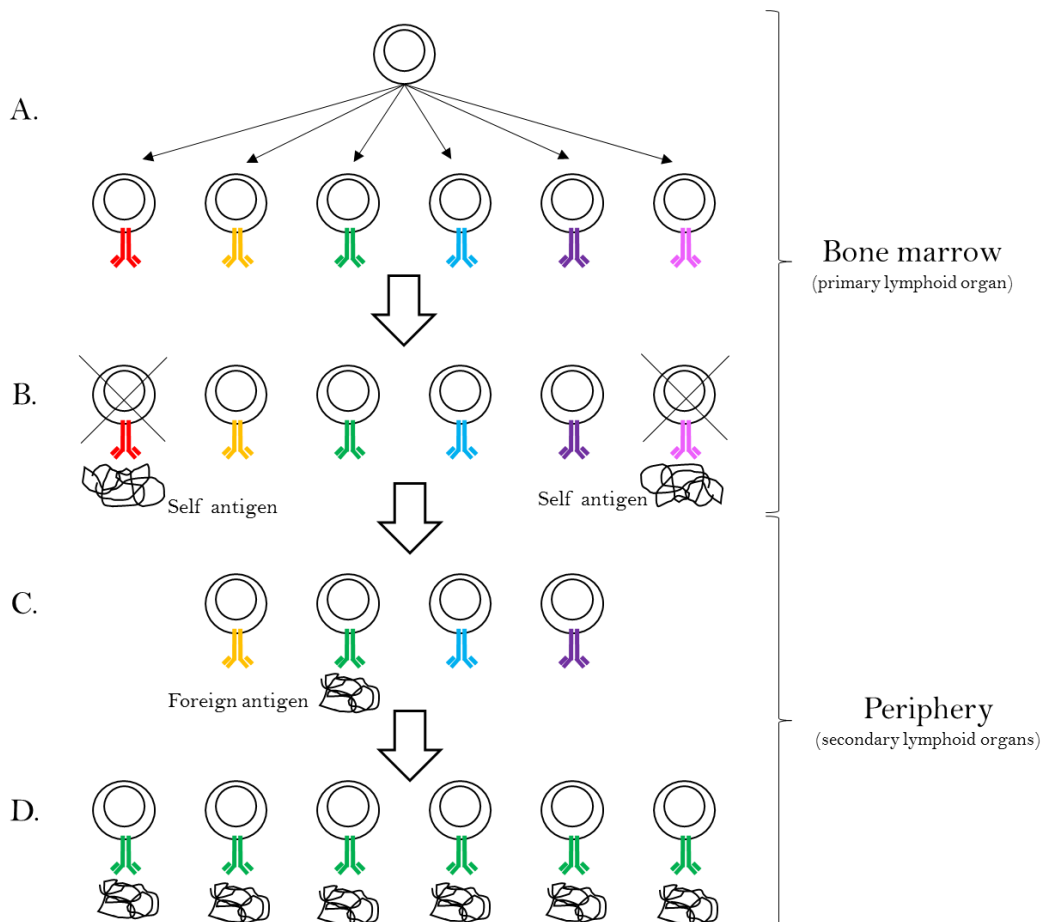
## 1.1.1 Basics of B cell biology

The immune system protects organisms from infection with layered defenses of increasing specificity. In higher organisms, the immune system is divided into innate and adaptive immune systems. The innate immune system provides a first line of defense against many microorganisms and an immediate, but non-specific response. Its components are found in multicellular organisms throughout the animal kingdom and, from an evolutionary perspective, it is the oldest mechanism of defense against pathogens (2-6). However, pathogens have developed ways to escape the innate immune system and, as life forms became more complex, a new system of defense emerged.

The adaptive immune system arose around 500 million years ago in jawed vertebrates (gnathostomata) (7-9). This system is more specific and, unlike the innate immune system, is able to develop an immunological memory. The system is based mainly on lymphocytes: the B cells and the T cells, both having different functions. The adaptive immune system is capable of recognizing and responding to an almost infinite number of antigens thanks to a very large repertoire of specific cell surface receptors on the lymphocytes membrane. Each lymphocyte receptor only reacts and binds to one specific antigen. The general structure of these receptors consists in a variable domain involved in antigen recognition and a constant domain delivering the signal to the cell after antigen-binding. The immunoglobulins (IG) are the antigen-recognition molecules of the B cells. They serve as cell receptors on the surface of B cells (B cell receptor; BcR) and can be secreted as antibody by the terminally differentiated B cells, the plasma cells. The major contribution of B cells to immunity is the secretion of antibodies, whose main biological function is to bind to pathogens (or their products) and ease their elimination from the organism.

In mammals, B cells originate in the bone marrow from undifferentiated hematopoietic stem cells that have the ability to develop into lymphoid or myeloid stem cells (10). The lymphoid stem cells are the common progenitor of the B and T cells. While B cells mature in the bone marrow, T cells further develop in the thymus. B cell development proceeds through several stages controlled by signals from the stromal cells. A single B cell progenitor gives rise to a large number of B cells, each expressing an IG with a different specificity thanks to a complex series of gene rearrangements (Fig. 1A) (see 1.1.3). Before leaving the bone marrow, immature B cells are subject to selection for self-tolerance, and cells that react

to self-antigens are removed from the germline repertoire (clonal deletion; Fig. 1B). Mature B cells enter the bloodstream and migrate to the peripheral lymphoid organs. B cells circulate continually from the blood into the peripheral lymphoid organs, and back to the blood via the lymphatic vessels. The millions of B cells in the body collectively carry millions of different antigen receptor specificities. This constitutes the IG repertoire of an individual, which is continuously modelled by antigen challenges. In case of an infection, B cells that recognize the antigen are arrested in the lymphoid tissue (Fig. 1C). Antigen-reactive B cells are activated to divide and proliferate (clonal selection; Fig. 1D). Ultimately, some progenies of the antigen-reactive B cells will differentiate into plasma cells and memory cells (primary immune response). The role of the plasma cells is the production of antibodies. Upon later exposure to the same antigen, the circulating antibodies provide an immediate protection until the memory cells are activated (secondary immune response) and repeat the cycle of division and differentiation.

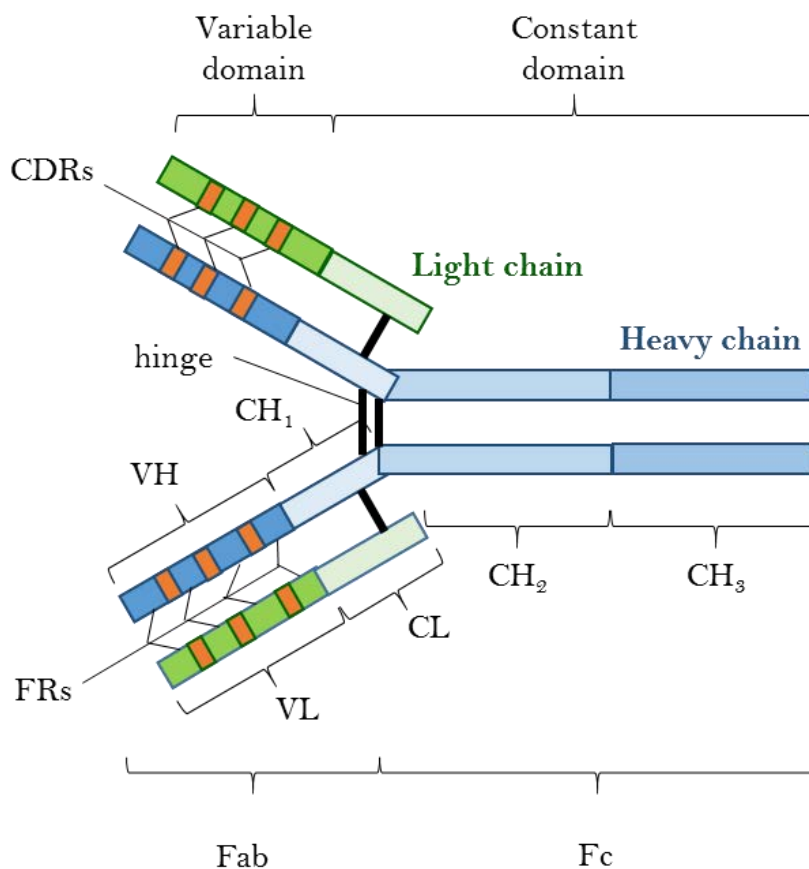


**Figure 1:** Clonal deletion and clonal selection (1)

(A) During development in the bone marrow, progenitor cells give rise to large number of immature B cells, each bearing an antigen-receptor with a unique specificity. (B) Immature B cells that react with self-antigens are eliminated by clonal deletion. (C) Mature antigen-naïve B cells enter the bloodstream and migrate to secondary lymphoid organs where they may encounter their specific antigen. (D) When a B cell meets its specific antigen, it is activated to proliferate and differentiate to form a clone of effector cells.

## 1.1.2 Structure of an antibody molecule

Antibodies are Y-shaped heterotetramer glycoproteins consisting of 2 identical heavy (IGH) and 2 identical light chains (Fig. 2) (11). The 2 IGH chains are linked by disulfide bonds, and each IGH is also linked to a light chain by a disulfide bond. Both the IGH and light chains are composed of constant and variable domains. The structure of the constant domain of the IGH defines the isotypes of the antibody. The variable domains are involved in the antigen-recognition. There exist 2 types of light chain, kappa ( $\kappa$ ; IGK) and lambda ( $\lambda$ ; IGL).



**Figure 2:** Structure of the IgG molecule

The antibody molecule is formed of 2 type of polypeptide chains, the heavy (in blue) and light (in green) chains, linked by disulfide bounds (black lines). The variable domains (V-domains) of both the light and heavy chains compose the V region of the antibody, and the constant domains (C-domains) of the heavy chains make up the C region. The 2 domains of the light chains are termed VL and CL while the 4 heavy chain domains are termed VH, CH1, CH2 and CH3 (A.N. for the IgG). The VH and VL domains are linked together and the CH1 and CL domains are paired. Based on papain digestion, the antibody can be divided into 3 pieces: 2 Fabs (Fragment antigen binding) and 1 Fc (Fragment crystallizable) fragments (12). The Fabs, corresponding to the “arms” of the antibody molecule, are involved in the antigen-binding. They are composed of the complete light chains paired with the VH and CH1 domains of the heavy chains. The Fc corresponds to the paired CH2 and CH3 domains and is the part of the antibody that interacts with cellular effector functions. The hinge region that links the Fab and Fc portions is flexible and allows independent movement of the 2 Fabs arms. The Fc fragment and hinge regions differ in antibodies of different isotypes, determining their functional properties.

There exist five types of IGH C region ( $C\mu$ ,  $C\delta$ ,  $C\epsilon$ ,  $C\gamma$ ,  $C\alpha$ ) to which correspond 5 main types of antibodies, or isotypes (IgM, IgD, IgE, IgG, IgA). Antibodies from different isotypes have different functional properties. The differences in the sequence of the various CH cause the isotypes to have different features. For instance, the length of the hinge region, the number of C domains, or the number of disulfide bonds vary between isotypes. The isotype greatly influence the location of antibodies and thus the different IG are adapted to different body compartments.

IgM is the first isotype to be expressed and is associated with the primary humoral immune response (13). IgM antibodies have lower affinity towards the antigen than the other isotypes because they are produced, in most of the cases, before the B cells have undergone affinity maturation. Because of this lower affinity, IgM antibodies tend to be more polyreactive. Their avidity for the antigen is however high because IgM molecules form pentamers and their 10 antigen-binding sites can bind simultaneously (14). Their pentameric structure usually confines IgM to the blood and makes them potent in activating the complement system. IgD antibodies are produced by alternative splicing of the same mRNA molecule as IgMs, and can be found together on the surface of B cells (15). One role of IgD is to regulate B cell fate at specific development stages.

The other isotypes, IgG, IgA and IgE, are generated once the B cell meets its antigen. They are smaller molecules than IgM and can diffuse more easily to the different tissues (1). IgG dominates the humoral immune response and is the principal isotype both in the blood and extracellular fluids (16). It has the longest half-life of all antibodies. IgG molecules are always monomeric and their smaller size enables them to diffuse easily in the tissues. IgG are the sole antibodies to be transported across the placenta. IgG opsonize pathogens, resulting in phagocytosis by specialized cells. In humans, the IgG class is subdivided into 4 subgroups named according to their relative abundance in the blood of healthy individuals: IgG1 to IgG4. IgA are the principal isotypes in secretions, especially in the mucous layer of the respiratory and intestinal tracts, and therefore dominate the humoral mucosal immunity (17). They are divided into 2 subclasses: IgA1 and IgA2 (18). IgE are monomeric antibodies present only at very low levels in the blood and extracellular fluids (19, 20). IgE bind with great avidity to mast cells receptors triggering the release of chemical mediators. They are associated with allergic reactions and one of their major role is the protection against parasites (21, 22).

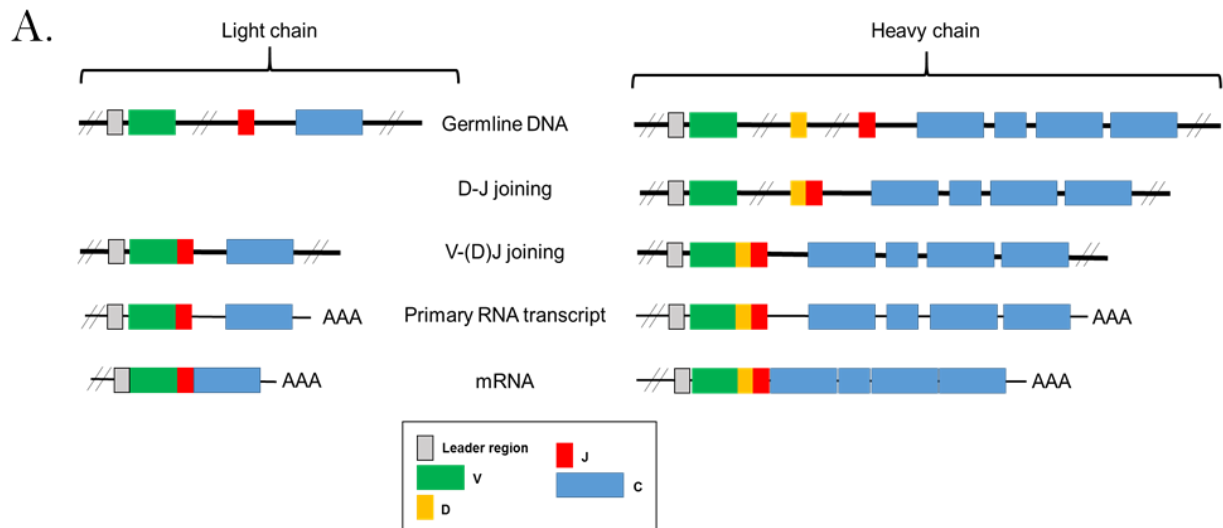
Based on the nucleotide sequence homology between different antibody molecules, the V-domain of both heavy and light chains can be divided into different regions (Fig. 2). Three regions, named

Complementary Determining Regions (CDRs), are particularly variable in sequence while the 4 in-between regions, termed Framework Regions, are relatively invariant (23). When the VH and VL domains pair together, the CDRs of each are brought together and form the antibody-binding site or paratope. Consequently, the CDRs of both VH and VL domains contribute to the antigen-specificity (24). The CDR3 is the most variable part of the VH region and is the major determinant of antigen-specificity (25). The paratopes of any given antibody can specifically recognize a particular epitope on foreign molecules. This epitope can either be linear or conformational (26).

The complete collection of antibodies with different specificities of an individual is defined as the antibody repertoire (also called IG repertoire or BcR repertoire). In humans, the theoretical size of this repertoire is of at least  $10^{11}$  different antibody molecules (1). Its actual size is however limited by the total number of B cells of the organism and their fluctuations in response to antigens. There are 3 main processes involved in the generation of this highly diverse and incredibly large IG repertoire: combinatorial diversity, junctional diversity and somatic hypermutation (1, 27). The firsts are direct consequences of DNA rearrangements that occur during B cell ontogeny in the bone marrow and during which different V region genes selected from a relatively small repertoire of gene sequences available on the germline DNA are combined. The last process is a mutational process that occurs on rearranged V region exons in mature B cells only.

### 1.1.3 Somatic recombination and combinatorial diversity

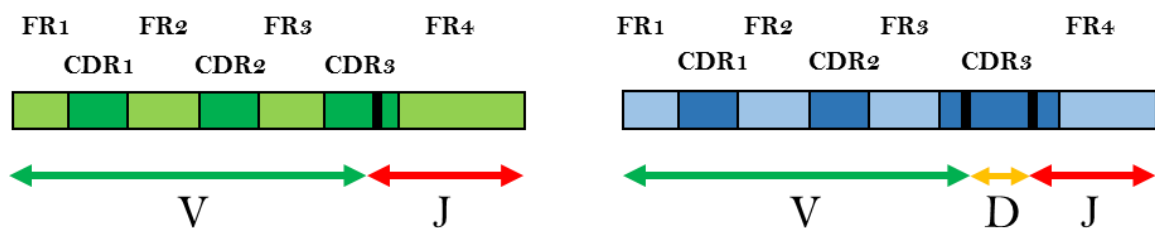
The IG heavy and light chains are encoded by a separate multigene family, and the individual V and C domains are each encoded by independent elements (28). While the constant domain is encoded by individual exons, the V regions are generated by an ordered series of gene rearrangement events during B cell ontogeny in a process called somatic recombination (Fig. 3A). The IGH V regions are encoded by 3 genes: IGHV (Variable), IGHD (Diversity) and IGHJ (Joining) genes. The light chain V region exon is formed by the joining of a Variable (IGKV or IGLV) to a Joining (IGKJ or IGLJ) gene (29, 30). The uneven junction of the different genes chain give rise to the CDR3, the core of the antigen-binding site. The CDR1 and CDR2 are on the other hand directly encoded in the IGHV and IGKV or IGLV genes (Fig. 3B).



**B.**

IG kappa or lambda variable region

IG heavy variable region



**Figure 3:** V, D and J genes are joined by somatic recombination

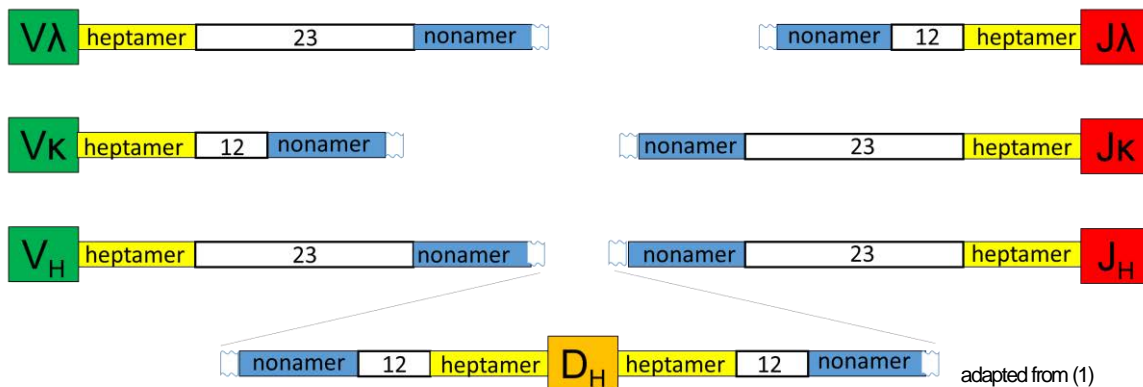
The IGH V regions are encoded by 3 genes: IGHV (Variable), IGHD (Diversity) and IGHJ (Joining) genes (29, 30). The ligation of these segments occurs in 2 stages: an IGHD gene is first joined to an IGHJ and, secondly, an IGHJ gene rearranges to the newly formed IGHD-IGHJ (31). The IGHC region gene is joined to the rearranged V-D-J exon by RNA splicing after transcription (32). In humans, the IGH locus is located on chromosome 14q32.33 (30, 33). For the production of an IG light chain, the entire V region exon is formed by the joining of an IGKV to an IGKJ gene or an IGLV to an IGLJ only. Similarly to the heavy chain, a C region gene will be linked by RNA splicing to the rearranged V region sequence. The kappa locus is located on the chromosome 2p11.2 and the lambda locus on chromosome 22q11.2 (34). For both the heavy and light chains, only the genes located on the same chromosome rearrange together. The rearrangement process takes place only on one chromosome at a time (31).

Each gene contributing to the generation of the VH and VL regions is present in multiple copies on germline DNA. There are 38–46 functional IGHV genes, 23 IGHD genes, 6 IGHJ genes and 9 IGHC genes that can be combined into a functional IG heavy chain, 34–38 functional IGKV, 6 IGKJ and one IGKC gene that can be rearranged into functional IG kappa light chains, and 29–33 IGLV, 4–5 IGLJ and 4–5 IGLC functional genes that can be rearranged into IG lambda light chains. Consequently, there are approximately 320 different possible light chains and ~6,300 possible heavy chain V-D-J combinations. This is the first aspect of the process of combinatorial diversity. The second source of combinatorial



diversity comes from the pairing of different heavy and light chains, as both light and heavy chains contribute to antibody diversity. Their pairing generates  $>2 \times 10^6$  different antibodies of distinct specificities.

The mechanism of somatic recombination of the V-region genes are similar for the heavy and light chain genes (35). The process requires ubiquitous DNA cleavage and repair enzymes as well as specialized enzymes. A specific mechanism, known as the 12/23 rule, ensures the proper rearrangement of the V-region genes (Fig. 4) (29, 35). There are conserved non-coding flanking regions in the heavy and light chain V, D and J genes adjacent to the places where recombination takes place (Fig. 4)(36, 37). These regions, called recombination signal sequences (RSS), consist in a heptamer sequence contiguous to the coding sequence of the gene followed by a spacer of 12 or 23 bp and a conserved nonamer (38, 39). The recombination can only link genes flanked by a 12mer-spaced RSS to a 23mer-spaced RSS. On the IGH locus, the IGHV and IGHJ genes are flanked by 23-spaced RSS while the IGHD gene is flanked by 12mer-spaced RSS, ensuring a correct V-D-J rearrangement (29). Similarly, on the light chain loci, the V genes are flanked by 23-spaced RSS and J genes by 12-spaced RSS (37).



**Figure 4:** The 12/23 rule ensures proper rearrangement of V-region genes

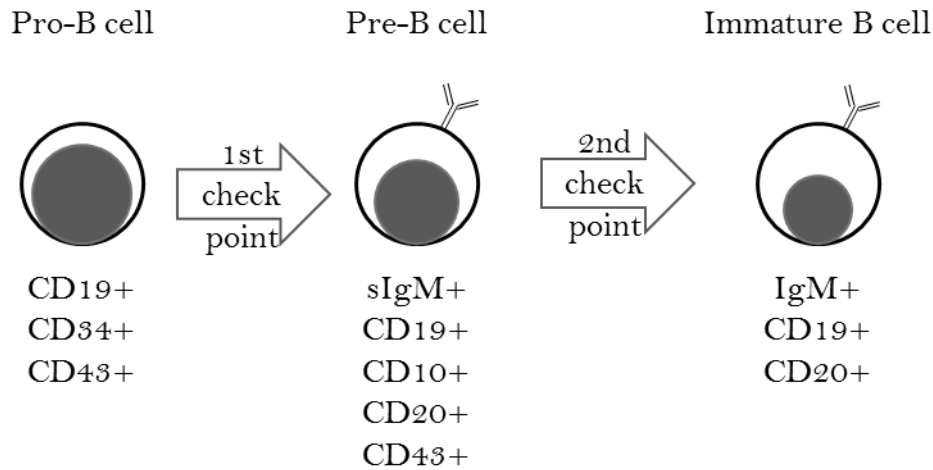
*Conserved non-coding sequences are flanking the heavy and light chain V, D and J genes. These regions, called recombination signal sequences (RSS), consist in a heptamer sequence contiguous to coding sequence of the gene followed by a spacer of 12 or 23 bp and a conserved nonamer. The recombination can only link genes flanked by a 12mer-spaced RSS to a 23mer-spaced RSS.*

At the core of the V-(D)-J recombination process are 2 endonucleases encoded by recombination-activation genes that are normally only expressed in developing lymphocytes, the RAG-1 and RAG-2 (40, 41). RAG-1 and RAG-2 are united in a heterodimeric RAG-complex that cleaves at the 3'-end of the heptamer causing a double stranded-DNA break and the formation of 2 DNA ends: a hairpin loop on the

coding end and a blunt end on the freed heptamer (42). Several complex events further increase the diversity in the junction region (35, 43). First, the hairpin is reopened at a random position by an enzymatic complex, the Artemis/DNA-dependent protein kinase (Artemis/DNA-PK) (44). In many cases the opening is not centered and results in single stranded-DNA 3'-overhangs. These extra bases are referred to as palindromic nucleotides (P nt) due to the palindromic nature of the sequence produced (35). The Artemis/DNA-PK has also an exonuclease activity and P nt may or not be excised already at this step. Next, the Terminal deoxynucleotidyl Transferase (TdT) adds nontemplated nucleotides (N nt) to the 3'-end (35, 43). Bases can subsequently be removed from the coding ends (including any P nt or N nt) by exonucleases before the two 3'-overhangs are joined by non-homologous end joining. Non-pairing bases are subsequently removed by repair enzymes and DNA polymerases insert additional nucleotides as needed to fill the gap and make the 2 ends compatible for the final ligation by DNA ligase IV (45). Consequently, the length of the CDR3 can vary substantially and, in the VH region, can be even shorter than the smallest germline IGHD gene. Based on the IG sequence, it can be difficult, if not impossible, to retrospectively identify the germline IGHD that contributed to the CDR3 because of the excision of most of its nucleotides. As this process of addition and deletion is random, the reading frame may be disrupted and result in non-coding sequences.

The rearrangement of the IGH genes occurs in the earliest B cell lineage stage, the pro-B cell stage: first the IGHD-IGHJ joining on both chromosomes (early pro-B cell stage), then IGHV to IGHD-IGHJ pairing on one chromosome at a time (late pro-B cell stage). A successful IGH V-(D)-J rearrangement leads to the expression of IGH  $\mu$ , or pre-B cell receptor, at the surface of large pre-B cells that are actively dividing. The large pre-B cells then stop dividing and become small pre-B cell in which the light chain genes start to rearrange on one chromosome at a time. The survival of the developing B cells depends on the productive rearrangement of their IGH and light chain genes and a large number are lost because of unproductive rearrangements (1<sup>st</sup> check point; Fig. 5) (46). Because of the uneven joining mechanism, the chances of generating out-of-frame sequences are high. If the junction of the IGHV to IGHD-IGHJ or of the light chain genes leads to an unproductive sequence, the rearrangement will start on the other chromosome, giving the B cell additional chances to survive. When the process of V-(D)-J recombination has resulted in a functional IG gene, the process is blocked for the other chromosome and a B cell cannot express any other variable region (47). This process is called allelic exclusion and ensures that each B cell produces antibodies with only one type of variable chain (48). The cells that fail to productively rearrange their IG genes undergo apoptosis. Furthermore, for the light chain, isotypic exclusion is the phenomenon in which only one of the 4 light chain loci is expressed in each individual B cell (49, 50). Upon successful

IG rearrangement, the cells become immature B cells and are subject to selection for self-tolerance (2<sup>nd</sup> check point; Fig. 5) (51). Self-reactive immature B cells can be rescued by further IG rearrangements of the genes through receptor editing (52). The immature B cells that survive this selection become mature B cells that express IgD in addition to IgM at their surface.



**Figure 5:** B cells mature in the bone marrow from lymphoid progenitors

The first developmental stage is the pro-B cell stage. They express the antigen associated with stem cells (CD34 and CD117), as well as antigen specific of the B cell lineage (CD19 and CD22). At the next stage, the pre-B cell stage, the synthesis of IG starts and IgM heavy chain become detectable in the cytoplasm. The pre-B cells that successfully rearranged their IG loci become immature B cells. The immature B cells express IgM molecules on their surface and no longer express the pre-B cell receptor. The subsequent phases of the differentiation are controlled by the antigen. At this stage, the cells are subject to selection for self-tolerance (51). B cells that have the ability to be stimulated by self-antigens are eliminated or inactivated (53). Immature B cells that bind to ubiquitous self-surface molecules e.g. Major Histocompatibility Complex (MHC) undergo apoptosis (clonal deletion)(54). Immature B cells that bind to soluble self-antigens lose their ability to respond to the antigen and migrate to the periphery as anergic B cells (55). The immature B cells that survive this selection become mature B cells that express IgD in addition to IgM at their surface.

### 1.1.4 Somatic hypermutation and class-switch recombination

The mechanisms described above take place during B cell ontogeny in the bone marrow. When B cells have successfully rearranged their IG light and heavy chain loci and express a functional BcR, they leave the bone marrow and enter the periphery. The mature but antigen-naïve B cells migrate through the peripheral blood stream to secondary lymphoid tissues: the lymph nodes, the spleen, the gut-associated lymphoid tissues (GALT; tonsils, appendix and Peyer's patches), the bronchial-associated lymphoid tissue (BALT), and the mucosal-associated lymphoid tissue (MALT). When a B cell encounters its specific antigen in a secondary lymphoid organ, the cell is activated by the binding of the antigen to the BcR. Protein antigens normally require additional signals from helper T cells to activate specific B cells. These antigens are referred as T-dependent (1, 56). Non-protein antigens (e.g. lipids or polysaccharides)

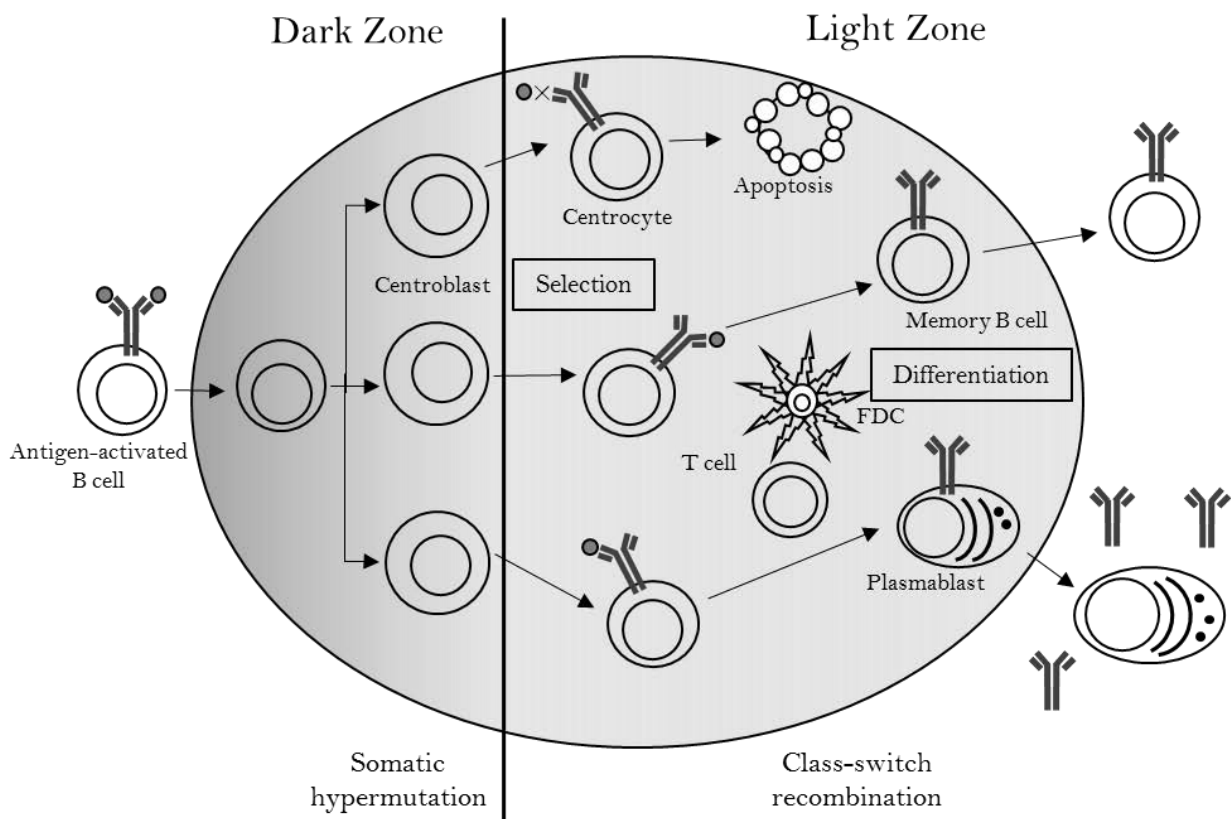
that do not require participation of T cells are called T-independent. Most T-independent antigens are able to cross-link surface receptors because they expose multiple identical epitopes. T cell-dependent responses usually result in high affinity antibodies of different isotypes, while T-independent antigens induce essentially IgM of lower affinity (57, 58).

Upon activation with a T-dependent antigen at the interface between the T cell area and the primary follicle, antigen-stimulated B cells adopt different fates (59). The mechanisms conditioning the fate decision are still poorly understood (60). Some activated B cells, under the influence of cytokines, will move to extrafollicular zones of the lymphoid tissue, proliferate, increase their amount of antibody production and differentiate into short-lived plasma cells. Most of these short-lived plasma cells produce IgM antibodies and will migrate to the bone marrow approximately 2 to 3 weeks after antigen exposure. Others, still under the influence of particular cytokines, will gain the center of the primary follicle and proliferate at high rate to establish the germinal center (Fig. 6) (61, 62).

Germinal centers appear within 4 to 7 days after antigen exposure upon the migration and rapid proliferation of some activated B cells at the center of the primary follicles (62, 63). Each germinal center arises from solely one or only a limited number of proliferative activated B cells called centroblasts. This process is defined as B cell clonal expansion and occurs in the dark zone of the germinal center (Figure 6). The progenies of the centroblasts are smaller cells called centrocytes, which are responsible for the characteristic structure of the germinal center (63). The centroblasts are confined to the basal dark zone of the germinal center. The centrocytes migrate to the adjacent light zone where they are in contact with abundant follicular dendritic cells that continuously display antigens (63).

The process of somatic hypermutation (SHM) takes place in the dark zone of the germinal center (64-66). Stimulation of the B cells by armed T cells induce the expression of the enzyme activation-induced cytidine deaminase, involved in both SHM and class-switch recombination (CSR) processes (67-72). Centroblasts accumulate point mutations in their rearranged IG heavy and light chain genes at a very high rate (about 1bp per  $10^3$  per cell division), especially at certain hotspots of mutation (73, 74). Replacement mutations tend to be clustered into the CDRs, while silent mutations that preserved the antibody structure, occur essentially in the FRs (75). These mutations can change the antibody affinity and lead to antibody of increased affinity. Only the centrocytes that bind the antigen displayed by the follicular dendritic cells with high affinity are selected to survive and, will be preferentially selected to mature into antibody-secreting cells (affinity maturation) (76, 77). This antigen signal is required to prevent them from apoptosis. The selection process thus depends on binding competition to follicular

dendritic cells between centrocytes (63). Some centrocytes that have received the survival signal migrate to the basal zone of the light zone where they may undergo CSR (78, 79). CSR will replace the default IGH C $\mu$  gene by another type of IGHC gene and change the effector function of the immunoglobulin (80). Thanks to this process, the same VH can be associated with different IGHC genes in the course of an immune response, and every B cell with a variable region of desired specificity can produce the adapted antibody isotype to protect all body compartments. Others centrocytes will differentiate into long-lived plasma cells, exit the germinal center, and migrate to the bone marrow where they can secrete antibodies for extensive periods of time. Some other centrocytes do not differentiate into plasma cells but become memory B cells (81). These cells have the ability to survive for long periods without antigen stimulation and are able to produce a more rapid antibody response upon subsequent antibody exposures. If B cells do not encounter their specific antigen, they either die or leave the secondary lymphoid organ via lymphatic vessels before returning to the blood circulation.



**Figure 6:** Schematic representation of the germinal center reaction (82)

Based on its histological appearance, the germinal center is divided in 2 compartments, the dark and light zones (83). The centroblasts are confined to the basal dark zone of the germinal center. Centroblasts actively divide in the dark zone and their IG loci undergo somatic hypermutation. The progenies of the centroblasts, the centrocytes, migrate to the adjacent light zone (63). Centrocytes that bind to the antigen with higher affinity are selected to survive and will differentiate into plasma cells or memory B cells. Class-switch recombination takes place in the light zone (78, 79).

## **1.2 Analyzing the IG repertoire: new perspectives with NGS**

### **1.2.1 Conventional techniques for the analysis of the IG repertoire**

The information lying in the IG repertoire is extremely precious for both theoretical and applied immunology. The frequencies of different B cells populations and the diversity of the immunoglobulin molecules can be studied at different biological levels, and different techniques to investigate the IG repertoire have been established over the years (84, 85).

B cell receptors can be investigated at the protein level. Flow-cytometry is a widely used tool to study the dynamic and function of the different B cell subsets in physiological and pathological conditions (86-94). This technique allows to track and sort cells with particular phenotypes or reactivity. Thus antigen-specific B cells can be stained with fluorescent antigens or epitopes (95-99). Enzyme-linked immunosorbent assays (ELISA) and enzyme-linked immunospot technique (ELISPOT) are also being used to dissect humoral immune responses and the modulations of the repertoire of antigen-specific B cells in human (100-102). Recently, techniques based on micro-wells loaded with single lymphocytes have been developed to interrogate the specificity of a large number of cells (103, 104). For instance the PANAMA-blot technology, a semi-quantitative immunoblot technique, is used to analyze the reaction of immunoglobulins in sera against complex protein mixtures (105). Proteomic tools can be advantageously combined to NGS technologies for the identification of high-affinity antigen-specific antibodies from immunized animals or human donors. For instance, specific antibodies against Hepatitis B and Cytomegalovirus have recently been identified directly from human samples by combining liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of serum antibodies to high-throughput sequencing of the IG repertoire (106-108). Similar analyses have also been performed in rabbits to determine the antibody composition of the polyclonal serum response after immunization (109).

Investigating the immunoglobulins at the DNA sequence level provide information that are essential for understanding the underlying variability of the B cell IG repertoire. Junction length spectratyping is a cheap and straightforward way to visualize oligoclonal or clonal expansion of lymphocytes. Spectratypes (the results of the approach) provide a picture of the IG repertoire diversity and show variations between individuals and over time, in different physiological or pathological conditions. The technique was

initially developed to analyze the murine repertoire of T cell receptors (TR) after immunization (110). Since then, several groups have applied this technique to the study of the TR and IG repertoire. Numerous studies in various areas of T and B cell research are based on junction length spectratyping, e.g. immune cell development and fetal repertoire, immune response to vaccination or infection, autoimmunity, immunosenescence, oncology, and transplantation (111-120).

Early techniques to determine the expression of various V gene subgroups relied on RNA colony blots of hybridomas probed with IGHV-subgroup-specific probes (121, 122). Using specific primers, methods based on real-time quantitative PCR have been developed (123). This approach has been essentially and extensively used to monitor B cell leukemias and detect minimal residual disease (124-126). Before the advent of next-generation sequencing technologies, Sanger sequencing has been the method of reference to gain insight into the IG repertoire at the sequence level. The studies using this method are however limited to a relatively small set of sorted single cells and/or immortalized B cells (127). For about 3 decades, Sanger sequencing has dominated the field of DNA sequencing and is still used by many laboratories worldwide. However, in the first decade of this century, the emergence of next-generation (or high-throughput) DNA sequencing revolutionized the field of genomics.

### **1.2.2 Next-generation sequencing: Ion Torrent technology**

Next-generation sequencing (NGS) technologies are one of the major breakthrough in molecular biology. Each sequencing run provide millions of sequences allowing to get a broad and detailed picture of classical problems in biology. For example, NGS offers now the possibility to produce antibodies without the extensive screening steps of conventional antibody discovery technologies, or enables to investigate the complexity and the dynamics of the adaptive immune response elicited by pathogens or vaccination. Providing a global, representative view of the IG repertoire and estimating the IG diversity effectively present used to be extremely challenging tasks, but the advances in sequencing technologies allow now the study of representative subsets of B cells at unprecedented depth.

In the past decade several NGS platforms have been developed and each one possess different strengths, bottlenecks and characteristics (128-130). They differ in terms of accuracy, read length and number of generated reads, hence not all technologies are suitable for the study of the IG repertoire.

The Ion Personal Genome Machine (PGM) was launched in 2010. This platform has a benchtop format and is very versatile in its applications. Chips come in different formats, allowing to obtain an average of 100,000 (314 chip), 1,600,000 (316 chip) or 4,000,000 (318 chip) reads per run. The technique regularly

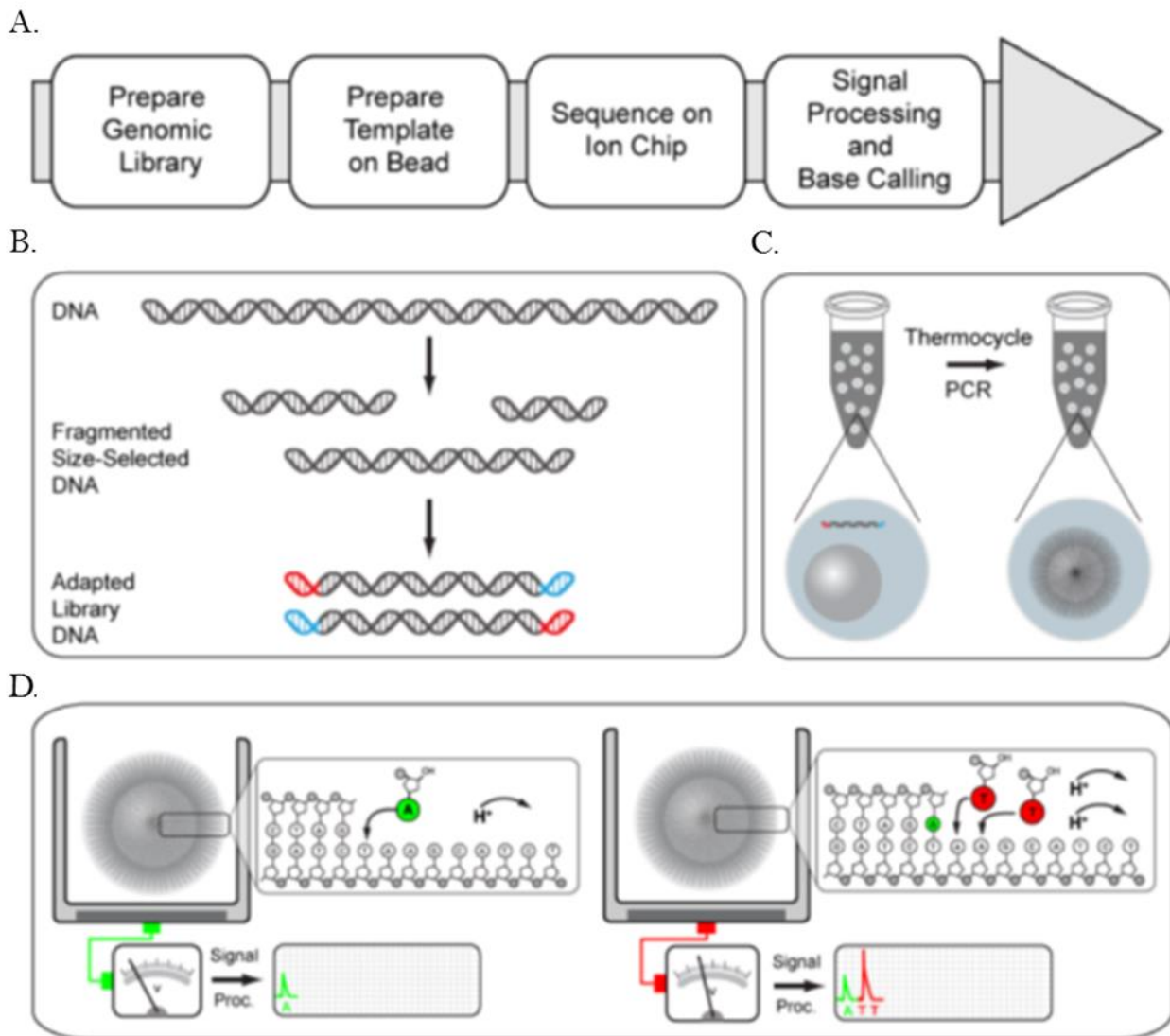
improved in terms of read length abilities (128). Originally of 100 bp, the read length was subsequently increased to 200 bp and is presently of 400 bp. Because of the recently extended read length and the relative low cost of the platform, the Ion Torrent PGM platforms has only now begun to be used in IG repertoire research (131).

The high throughput sequencing requires a DNA sample fairly homogeneous in length. Therefore only amplicons of similar size can be sequenced in a single run and a step of fragmentation and/or size selection is necessary when sequencing genomic DNA. Adapters, named A and P, are required at both ends of the target sequence (Fig. 7B). These adapters, together with multiplex identifiers (MIDs), can be either added by ligation during the library preparation or directly added during the PCR using extended primers. MIDs are nucleotide sequences of 8 to 10 bp at the beginning of each reads (between the adapter and the target sequence) that allow to differentiate between samples and therefore to sequence different samples during the same sequencing run. The DNA fragments or amplicons are then captured by hybridization onto beads covered with oligonucleotides complementary to the Adapter A (Fig. 7C). It is essential that the DNA library is sufficiently diluted to ensure that each bead carries a single DNA fragment. An amplification of the library is performed by emulsion PCR (Fig. 7C). An emulsion PCR is an elegant way to amplify multiple DNA fragments in a single PCR reaction. Oil is added to the mix and upon shaking micelles containing a single bead are formed and work as millions micro-reactors for the PCR reaction. After a normal thermal cycling each beads is covered by thousands of copies of the DNA fragment originally present. The template-loaded beads are then enriched and purified. Next, the DNA fragments are denatured to ssDNA and enriched beads are primed for sequencing by annealing a sequencing primer complementary to the adapter P.

Subsequently, beads are loaded onto a chip composed of millions of wells and specially designed to detect pH changes within each well thanks to an ion-sensitive field-effect transistor (ISFET) (Fig. 7D). The Ion Torrent system uses a semiconductor sequencing approach. It detects hydrogen ions released during DNA synthesis by quantifying pH changes with a semiconductor sensor (132, 133). Nucleotides are added sequentially in a specific order. If a nucleotide is incorporated, a proton is released and the consecutive slight pH change detected by the ISEFT. The measured shift in pH is proportional to the number of nucleotides incorporated (Fig. 7D). Between each addition, the unbound nucleotides are washed away. The process of sequential nucleotide addition is repeated hundreds of times per sequencing run.



The major source of errors for this technique are insertions and deletions (indels) that occur especially in regions containing homopolymers. This effect increases with the length of the homopolymer because of a loss of resolution and eventual saturation of the pH detector. Substitution errors occur with very low frequencies. The overall error rate is of about 1-1.5% but the reported rates in studies range from 0.46% to 2.4% (129, 134, 135).



**Figure 7:** Ion Torrent workflow

(A) General workflow of the Ion Torrent PGM platform. (B) Amplification primers are added at both ends of the sequencing DNA target, either by ligation or during PCR amplification. (C) Amplicons are captured onto beads and amplified by emulsion PCR, ensuring that each bead in the mixture is covered with millions of identical copies of the sequencing target. (D) Beads are loaded onto a sequencing chip composed of millions of wells. Each well can only accommodate for a single bead and acts as a micro-sensor that detects the slight pH changes when nucleotides are incorporated.

### 1.2.3 Bioinformatics tools for the analysis of IG sequences

The results of a NGS run are thousands to millions of raw nucleotides reads. The raw sequences have to be cleaned and annotated before any biological information can be gained. There exists no “gold standard” in the bioinformatics processing of the NGS data and the strategies differ between laboratories. However, the analytic pipeline followed by NGS IG sequences can be basically decomposed into 3 steps: filtering and quality check, annotation of the IG genes, and exploiting of the IG features.

The first analytical steps are common to all amplicon-sequencing experiments. Several samples can be multiplexed in a single NGS run by the incorporation of a string of known nucleotides (MIDs) in the primer or linkers used in library preparations. Based on unique MID sequences, the sequencing reads can be retrospectively assigned to the proper sample. While multiplexing reduces the overall per sample cost of an experiment, it also negatively affects the sequencing depth, and therefore the global coverage of the individual IG repertoire. All of the NGS platforms assign a Quality Score to each nucleotide position, and a threshold in the overall Quality Scores of a read is usually set to exclude sequences of poor quality. NGS platforms, such as the Ion Torrent PGM, can directly conduct these filtering based on user defined criteria. The sequences of the primers used for amplification of the rearranged IG sequences are usually trimmed to avoid introduction of point mutations. Because of the high variability of IG sequences, the primers used may indeed not have a perfect match to the gene.

Next, IG sequences need to be annotated. Several established algorithms exist for germline V-(D)-J assignment and annotation of the IG gene. The leading platforms are IMGT/HighV-QUEST, IgBlast, SoDa and iHMMune-align but faster and more precise algorithms are continuously developed (85, 136-141). Of major interest, IMGT®, the international ImMunoGeneTics information system® has been created in 1989 by Prof. Marie-Paule Lefranc (142). This website offers large databases of immune receptor gene sequences from a large number of species and extensive documentation on immune cell receptor repertoires ([www.imgt.org](http://www.imgt.org)). The IMGT/HighV-QUEST enables the standardized analysis of up to 500,000 IG sequences obtained by NGS and is a major tool for IG repertoire research (137). Briefly, the annotation programs aligns the IG sequences obtained by NGS to a reference set of germline V, D and J genes. The closest germline gene is assigned to the sequence. The different regions of the VH domain (FR and CDR regions) are annotated based on this alignment and a standardized numbering system (IMGT® numbering). The positions of somatic mutations in the sequence are also determined by comparison to the closest germline sequence. However, high level of somatic hypermutation (or error rate) can potentially lead to the erroneous identification of the original germline gene. Because of

accumulated point mutations, a sequence can indeed become closer to an unrelated germline sequence than to its actual germline gene. Because of the uneven junction of the V-(D)-J genes, the N nucleotides addition and hypermutation, the original IGHD gene is difficult to reliably assign. The annotation of the different V, D and J and CDR3 AA sequences is crucial to group together IG sequences that originate from the same clonally expanded B cell. The assumption is that sequences belong to the same clone if they use the same V-(D)-J genes and have a high level of similarity in their CDR3 sequence. While this approach is conceptually accepted and applied by many groups to determine the clonality of the sample, there is no consensus on the minimum identity level to be applied on the CDR3 (generally about 80% nt identity).

For every single IG sequence, “annotation programs” provide a wide number of information: V(D)J gene usages; mutation rates for the different VH regions and types of mutation; CDRs length and composition; and many other metrics. Thus the IMGT/HighV-QUEST output contains eleven files which provide results (143). Data mining and visualization of the results constitute probably the main brain-teaser and challenge for the researchers investigating the IG repertoire. Some tools have been developed to make use of the data generated by IMGT/HighV-QUEST. For instance the IgAT program, based on Microsoft Excel, facilitates the extraction and visualization of data (144). However, IgAT only accommodates for small number of sequences and rapidly became obsolete. IMGT/HighV-QUEST also provides some tools for data mining and visualization (140). Different groups have developed programs that allow the grouping of the sequences into clonotypes and the creation of IG lineage trees (145, 146). The affinity maturation of antibodies can indeed be explored by examining the IG gene sequence of the responding B cells. These approaches have been useful for the study of the mutation events that for instance lead to the development of antigen-specific antibodies or have been applied to the study of lymphoproliferative diseases and multiple sclerosis (145, 147-150). Vertex cluster plots are also used to group sequences into clonotypes and visualize the phenomena of clonal expansion and SHM (Chapter 3) (117, 151, 152).

## 1.3 NGS applications on IG repertoire analysis

### 1.3.1 IG repertoire development, composition and diversity

NGS analysis of the IG repertoire has found a wide range of basic and applied immunology applications (153-157). Beside humans, NGS has been applied to the study of the IG repertoire of a wide range of species, notably fishes (zebrafish, rainbow trout), mice, rabbits, cattle, macaques and chickens (117, 158-163). Although the field is still in its infancy, numerous studies have contributed to a better understanding of the humoral arm of the immune system, both in physiological and various pathological conditions.

NGS has enabled to get deep insights into the development of the IG repertoire. Investigation of the antibody repertoire in zebrafish at different developmental time-points provided interesting insight into how the adaptive immune system achieves diversity (164). Very recently, the IGH repertoire in fetuses has been assessed at various weeks of gestation (165). Results suggested that IG SHM and CSR occur already during intrauterine life. More accessible, investigation of human cord blood samples revealed unique features of the immature naive IG repertoire in comparison to the mature antigen-experienced one of human subjects and gave insight into the maturation processes (166). While the gene usage and VH CDR3 diversities were similar to the adult IgM repertoire, the usage of the different genes and rearrangements as well as the junctional diversity and level of somatic hypermutation differed significantly (166).

Providing a global, representative, view of the IG repertoire in humans and estimating the proportion of the diversity effectively used are extremely challenging tasks. The theoretical number of the different antibodies largely exceeds the number of B cells present at any time in an individual. Even if considering the VH CDR3 as the only source of antibody diversity, and considering that this region has a length between 4 and 20 amino acids, the number of possible antibodies is still  $10^{15}$ - $10^{16}$  times higher than the number of B cells in the organism. Several NGS studies have attempted to estimate the total diversity of the human antibody repertoire in the blood in physiological conditions (167-170). For instance, the VH CDR3 diversity in the blood of human adults was estimated to be of about 3 to  $9 \times 10^6$  (168) and the human IgM repertoire in the blood was found to be composed of at least  $3.5 \times 10^{10}$  different antibody molecules (170). Drawing general conclusions on the composition of the human naïve IG repertoire can be hazardous because of the multiple parameters that can influence its composition, such as the age of the donor or the genetic background. Furthermore, humans are naturally exposed to numerous antigens that continuously model the repertoire and previous antigen-exposures or diseases can impact the

repertoire composition for long periods of time. Analyses of IG germline sequences revealed a high degree of allelic diversity in the human IG locus (171-174). This observation further complicates the proper annotation of the V-(D)-J genes and the comparisons between repertoires. Analyses of IG genes revealed preferential usages of certain IGHV, IGKV and IGHL genes and rearrangements with D (IGH only) and J genes in the natural adult human repertoire (168-171, 175, 176).

In order to describe the global antibody repertoire, model organisms such as zebrafishes have been used (158, 177). Zebrafishes have fewer possible V-(D)-J combinations than mice or humans, and possess about  $3 \times 10^5$  antibody producing B cells. These animals were found to use between 50 and 86% of the possible V-(D)-J combinations, with some correlation of V-(D)-J patterns between individuals. Interestingly, identical heavy chain sequences were found in different individuals, suggesting a convergence towards the expression of the same antibodies (158).

NGS also offers the opportunity to study the B cell IG repertoire at different anatomical sites, and the repertoire of the different IG isotypes. Thus, beside IgM and IgG, the IgA and IgE repertoires in the blood have also been investigated (178-180). The human IG repertoire in paraffin-stored lymph nodes and a tonsil have been assessed, revealing clonal expansion concordant with lymphadenopathies (181, 182). Analysis of the IgG repertoire in cerebrospinal fluid in the context of multiple sclerosis gave evidence of an exchange of a selected few IgG-expressing B cells between the periphery and the central nervous system where they undergo clonal expansion and diversification (183). NGS offers a novel approach to examine the mucosal B cell system, notably the intestinal IgA repertoire (184, 185). Interestingly, analysis of the murine mucosal B cell IG repertoires revealed that B cell development and selection may take place in the *lamina propria* in adult mice (186).

### **1.3.2 Modeling of the IG repertoire by infection and vaccination**

A large number of studies have aimed to characterize the IG repertoire upon particular vaccination or natural infections using low-throughput (e.g. single cell sorting, cloning...) techniques. Although limited to the investigation of only small numbers of cells and/or donors, low-throughput methods enabled to gain important insights into the shaping of the IG repertoire for a variety of vaccines or infections. These studies relied notably on CDR3 length spectratyping, single-cell sorting followed or not by immortalization, and Sanger sequencing of rearranged IG sequences. For instance, investigation of the IG repertoire of patients with hepatitis C virus (HCV) infection showed that the presence of fewer clones

(pauciclinality) in the peripheral memory B cell population is a distinctive feature between donors with resolving or chronic HCV infection (187). Others have shown that B cell diversity decreases in old age, which may explain why elderly people respond less effectively to some immune challenges (111, 188).

NGS methods enable to investigate the complexity and the dynamics of the adaptive immune response elicited by pathogens or vaccination at unprecedented depth. NGS has been specifically applied to the study of several vaccines or infections (Table 5).

**Table 1:** NGS studies of the IG repertoire applied to vaccines or infection diseases in human

<b>Vaccines</b>	
Influenza virus (Live attenuated or trivalent inactivated)	(152, 178, 189-192)
Tetanus Toxoid	(190, 193)
23-valent pneumococcal polysaccharide vaccine	(178)
Hib-MenC* polysaccharide-protein conjugate vaccine	(194)
Malaria vaccine (TLRs adjuvants)	(195)
<b>Natural infections</b>	
Human Immunodeficiency Virus (HIV)	(131, 145, 196-200)
Influenza virus	(201)
Dengue Virus	(202)

\*Hib: *Haemophilus influenzae* type B; MenC: meningococcal serogroup C bacteria

The application of NGS to vaccine studies is of special interest as it can aid the development and assessment of new vaccines and vaccination schemes (157). Studies investigating the IG repertoire after vaccination with the trivalent inactivated influenza vaccine or the live attenuated influenza vaccine found more abundant IgG lineages in donors immunized with the live vaccine (189). Similarly, dissecting the humoral immune responses to particular diseases can lead to a better understanding of the pathophysiology of these infections and help to improve or design treatments (156). For instance, intermediate IG sequences in the maturation pathway leading to broadly neutralizing antibodies (bnAbs) in HIV patients can be identified by NGS, and could help to design innovative vaccine immunogens (198-200, 203, 204). It was indeed observed that bnAbs against HIV are highly mutated and highly divergent from the closest corresponding IG germline genes. Also, naïve B cells expressing germline precursors of bnAbs may not bind efficiently enough to their epitopes to initiate an immune response (205). It was thus hypothesized that HIV-1 evolved to display epitopes that cannot be bound by germline non-mutated antibodies (203). A possibility to counter this immune-evasion strategy is first to use primary immunogens to activate naïve B cells expressing germline precursors of bnAbs and second to lead these towards maturation pathways resulting in bnAbs (203). NGS is also helpful to investigate the phenomenon of immunosenescence and understand the reasons of the higher susceptibility of elderly

individuals to infectious disease and of their lower response to some vaccines. It was notably shown that elderly individuals display a lower degree of diversity in their IgM and IgA repertoires than young individuals upon vaccination, and that donors over 50 years-old have a reduced CSR ability (152, 188, 206).

At the core of the post-vaccine or post-infection studies is the identification of antigen-specific IG sequences. Indeed, in order to fully benefit from NGS data it is crucial to develop strategies to distinguish the antigen-specific from the total IG repertoire. One way is to search for sequences with a high degree of identity to IG sequences of known specificity (152, 197, 198, 200, 201). Another approach to identify “putative” antigen-specific sequences is to investigate what is referred to as the convergent IG repertoire. This consists in comparing the IG repertoires of patients affected by a particular condition, or donors immunized against the same antigen(s), and identify the shared sequences (192, 194, 202). The sequences shared between the immunized individuals are assumed to be antigen-driven i.e. public antigen-specific signatures. At the root of this project, we hypothesized that the immune response in different individuals immunized against the same antigens converges and that shared VH CDR3 can be identified using NGS technologies. Since then IG convergence has been demonstrated using NGS in 3 longitudinal studies (192, 194, 202). IGH amino acid CDR3 sequences have been observed with increased frequency in patients recovering from acute dengue infection as compared to healthy or convalescent individuals (202). The specificity for dengue virus antigens of these convergent VH CDR3 signatures has however not been investigated further. Later the same year analysis of the IGH repertoire of donors immunized with a conjugate vaccine containing *Haemophilus influenzae* type B (Hib) and group C meningococcal (MenC) polysaccharides revealed that different individuals were using convergent VH CDR3 sequences identical or highly similar to known Hib-specific or TT-specific sequences (194). Convergent IGH rearrangements were also identified among different people in response to vaccination and infection with the H1N1 2009 influenza strain (192). Together these studies demonstrated that particular infections or vaccinations can induce specific IGH signatures. The signatures can potentially be used as markers for particular antigens and for the sequence-based monitoring of infectious diseases or vaccination campaigns.

## 1.4 Innovative techniques for the development of therapeutic antibodies

### 1.4.1 Antibody production and discovery using NGS technologies

The identification of monoclonal antibodies (mAbs) with desired characteristics, whenever they were generated *in vivo* (hybridoma) or *in vitro* (e.g. phage display), presents some drawbacks. Since its first description by Köhler and Milstein, the hybridoma technology has been extensively used with great success to generate high-affinity and highly specific mAbs (207). However, although hundred-thousands or millions of antigen-specific B cells are generated during a typical immunization process, only 10-1,000s antigen-specific hybridoma clones can be generated from a single experiment because of the low efficiency of the process (208). Thus less than 1% of the repertoire of antigen-specific B cells can be assessed, and the majority of the antigen-specific B cells potentially bearing receptors with desired characteristics is lost. Phage-display and other *in vitro* technologies require multiple sequential rounds of screening of the combinatorial libraries during which antigen-binding clones are enriched. However, the expression of some antibodies with high affinity often negatively affects the growth of the cells that express them (154). Following an initial enrichment by the first rounds of screening, these cells are progressively overgrown by cells encoding antibodies of lower quality. As a result, antibodies with high-affinity and high-specificity can be lost during the screening process (154). NGS analysis of combinatorial libraries after one or 2 rounds of screening have been used to recover clones expressing high-affinity antibodies that would have been missed by the classical approach (209). The authors followed the enrichment of CDR3 sequences between rounds of screening, cloned the high frequency CDR3 sequences and found out these were antigen-specific. These antigen-specific sequences would not have been ultimately discovered.

The large amount of IG sequences obtained by NGS offers now the possibility to produce antibodies without the extensive screening steps of conventional antibody discovery technologies. While the use of traditional technologies often leads to the repeated identification of the same clones, sequence-based identification allows more effective and elaborate testing as it avoids this redundancy (210). This approach offers the possibility to directly select and clone into expression vector IG sequences of interest, for instance those found with high-frequency in the sample or present in different individuals upon the same antigenic challenge. Several methods have been developed to exploit IG repertoire data generated



by NGS for antibody discovery. In a study performed in immunized mice, target-specific antibodies were successfully generated by pairing IG heavy to IG light chain sequences based on their frequencies in the IG repertoire of bone marrow plasma cells (159). Similarly, others showed that highly frequent IGHV genes in the spleen of immunized mice are encoding for antigen-binding antibodies (211). More recently, broadly neutralizing HIV-1-specific antibodies were produced thanks to a sophisticated VH:VL pairing strategy relying on phylogenetic analysis of the VH and VL sequences (197, 200).

In order to successfully generate antibodies *in vitro* by cloning of IG sequences in expression vectors, it is required to know the sequence of both the light and the heavy chains. For long, VH:VL pairing was only possible after single-cell sorting and Sanger sequencing of the individual IGHV, IGKV and IGLV genes. Because of its technical tediousness and its cost, this approach is only applicable to a limited number of cells. The development of methods enabling the sequencing of paired VH:VL domains is therefore one of the main challenge in antibody repertoire research. Recently, sophisticated methods have been developed to combine single-cell IGHV, IGKV and IGLV genes amplification to NGS technologies. These techniques allowed to determine the endogenous VH:VL pairing from relatively large numbers of B cells (190, 212, 213). The first reported method relied on single-cell sorting of B cells and amplification of the rearranged VH and VL domains of each cell independently using uniquely barcoded primers before pooling for deep sequencing (213). VH and VL amplicons could then be paired based on matching barcodes. The second method consisted in the single-cell sorting of B cells into *subnanoliter* volume wells, followed by the lysis of the cells and the capture of mRNA molecules onto polydT beads (190). Amplicons of linked VH:VL segments were then generated by reverse-transcription and overlap-extension PCR (“emulsion VH:VL linkage PCR”). Sequencing was performed on an Illumina MiSeq platform. This approach enabled to interrogate more than  $5 \times 10^4$  B cells per experiment. This year, an approach derived from the previous but by-passing the single-cell sorting into plates enabled the high-throughput analysis of VH:VL pairs from more than  $2 \times 10^6$  B cells (214). Using an axisymmetric flow-focusing device, individual B cells were sequestered into emulsion droplets containing lysis buffer and oligodT-covered beads to capture mRNA molecules. Linked VH:VL amplicons were then generated by overlap-extension RT-PCR and nested PCR amplification. The resulting ~850 bp VH:VL products were sequenced on an Illumina MiSeq platform. Because of the read length limitation of the NGS platform, 3 sequencing reactions (from the VH-side of the amplicons, from the VL-side, and a bidirectional sequencing) followed by *in silico* assembly of the reads were required to determine the sequence of complete linked VH:VL amplicons (214).

## 1.4.2 Transgenic animals expressing human immunoglobulins

The hybridoma technology, described by Köhler and Milstein in 1975, opened up the possibility to produce high-affinity and high-specificity mAbs with clinical applications from immunized laboratory animals (207). In 1986, the first therapeutic rodent mAbs (muromonab-CD3) was approved by the US Food and Drug Administration to prevent rejections of organ transplants (215). However, the therapeutic use of rodent mAbs in humans is severely impaired by their immunogenicity (216, 217). Therefore, the development of human mAbs as novel therapeutic agents rapidly became a major research interest. The rarity of the antigen-specific cells in the peripheral blood combined to the low fusion efficiency of these cells makes difficult the generation of mAbs directly from human B cells. Furthermore, the use of human subjects is largely restricted to certain antigens. To counter these pitfalls, different approaches have been developed (218–222). The humanization of rodent mAbs by manipulation of the V region sequences appeared to be a successful alternative. This process is however very tedious and needs to be carried out on a case-by-case basis (223, 224). Additional technologies were also developed to directly isolate *in vitro* synthetic mAbs from libraries of human and synthetic IG sequences (e.g. phage display technology) (225, 226). As mentioned in the last section, NGS offers now the possibility to directly identify antigen-specific IG sequences *in vivo*. Another alternative relies on the immunization of genetically engineered animals expressing repertoires of human antibody genes to derive antigen-specific hybridoma clones (221). In contrast with antibody engineering technologies, which involve the modification and optimization of individual sequences, transgenic animals generate affinity-matured antibodies that do not require further optimization and can be moved more rapidly into clinical trials (219). Thus, fully human antibodies from transgenic animals account for an increasing number of new therapeutics (219, 221).

Several strategies have been employed to generate humanized transgenic rodents (219, 221, 222). Generally, they rely on the genomic integration of BAC or YAC carrying IG gene loci by oocyte microinjection or transfection of embryonic stem (ES) cells. The endogenous IG loci of the recipient animals are silenced or deleted by gene targeting in ES cells or zing finger technology via DNA microinjection.

The size of the human IG loci constitutes a serious challenge to the transgenic antibody technology. Because the cloning vectors used for their creation could only accommodate small inserts (“microgene constructs”), the first IG transgenic animals carried only a limited set of human IG genes (227–229). Human antigen-specific mAbs were nevertheless successfully produced with these animals (228, 229).

The ability to generate antibodies against a variety of antigens with only a fraction of the natural human IG repertoire came as a surprise. This observation elegantly demonstrated that antibodies with a broad range of specificity can be created even with a limited combinatorial diversity, and highlighted the fundamental roles played by the junctional diversity (i.e. CDR3) and the SHM processes in antigen-specificity (218, 230). Similarly, a mouse with only a single human IGHV gene and 3 mouse IGKV genes was found to mount specific immune responses against a variety of T-dependent antigens, demonstrating that the diversity in the VH CDR3 is sufficient for most antibody diversity (25). The assembly of larger constructs in YAC or BAC by homologous overlap later enabled the integration of much larger human IG heavy and IG light chain regions into the mouse genome (221). Moreover, the introduction of the largest fraction of the human IG repertoire has been achieved by microcell-mediated chromosome transfer (231-234). With this approach a single human chromosome, or a fragment, is added to the mouse genome and replicates independently during cell division. To date there exist several rodent models that express complete or almost complete human heavy and light chain repertoires (reviewed in (222)). However, despite the fact that the IG transgene undergo normal V-(D)-J rearrangements, CSR and SHM, the expression of human antibodies in transgenic animals with fully human IG transloci appears to be suboptimal (235). The suboptimal performance of these transgenic mice, in respect of antibody yield and immune response, is attributed to the imperfect interaction of the human constant region of membrane IG with the endogenous rodent cellular signaling machinery (236). In this context, Open Monoclonal Technology Inc. recently developed a platform for the enhanced production of human antibodies based on a transgenic humanized rat strain, the OmniRat™ (235, 237). These animals carry a chimeric human/rat IGH locus, where 22 human IGHV, all (27) IGHD and all (6 functional) IGHJ genes in germline configuration are linked to the rat IGHC genes, together with fully human IG kappa locus (12 IGKV, 5 IGKJ and 1 IGKC) and IG lambda locus (16 IGLV, 5 IGLJ and 5 IGLC) (Fig. 8). The endogenous rat IG loci have been silenced using zinc finger nucleases. They express a diversified repertoire of chimeric antibodies with human idiotypes at levels similar to those of wild type animals and have been used to generate high-affinity chimeric mAbs (235). In this thesis, we analyzed the shaping of the IG repertoire of OmniRats immunized with various antigen, with emphasis on the Measles Virus (MV).

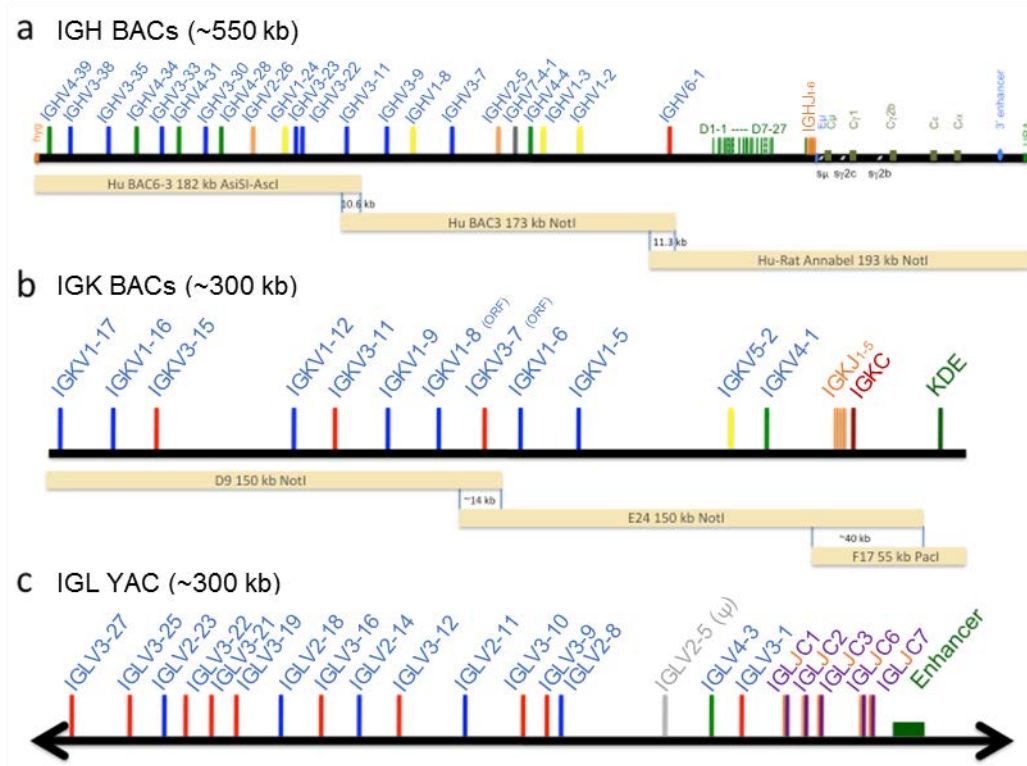


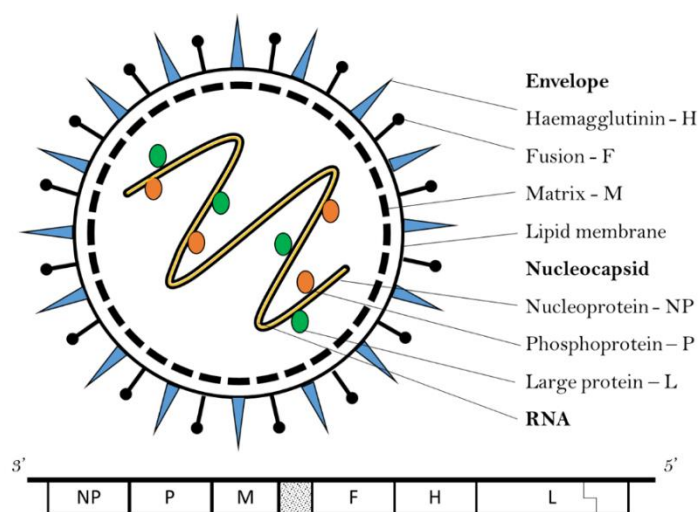
Figure 8: OmniRat's construct and transgenics (235)

## 1.5 Measles Virus

Measles is a highly contagious disease caused by the Measles Virus (MV), a morbillivirus (Paramyxoviridae) transmitted via aerosols or by direct contact with contaminated respiratory secretions (238). Although effective vaccines are available, measles remains a major cause of childhood morbidity and mortality, especially in developing countries (239, 240). In 2014, there were still around 115,000 measles-related deaths globally, mostly children under the age of five. The MV is phylogenetically most closely related to Rinderpest Virus and measles probably evolved from a morbillivirus infecting cattle (241, 242). Humans are the only natural reservoir of MV (238).

### 1.5.1 Virus structure and molecular epidemiology

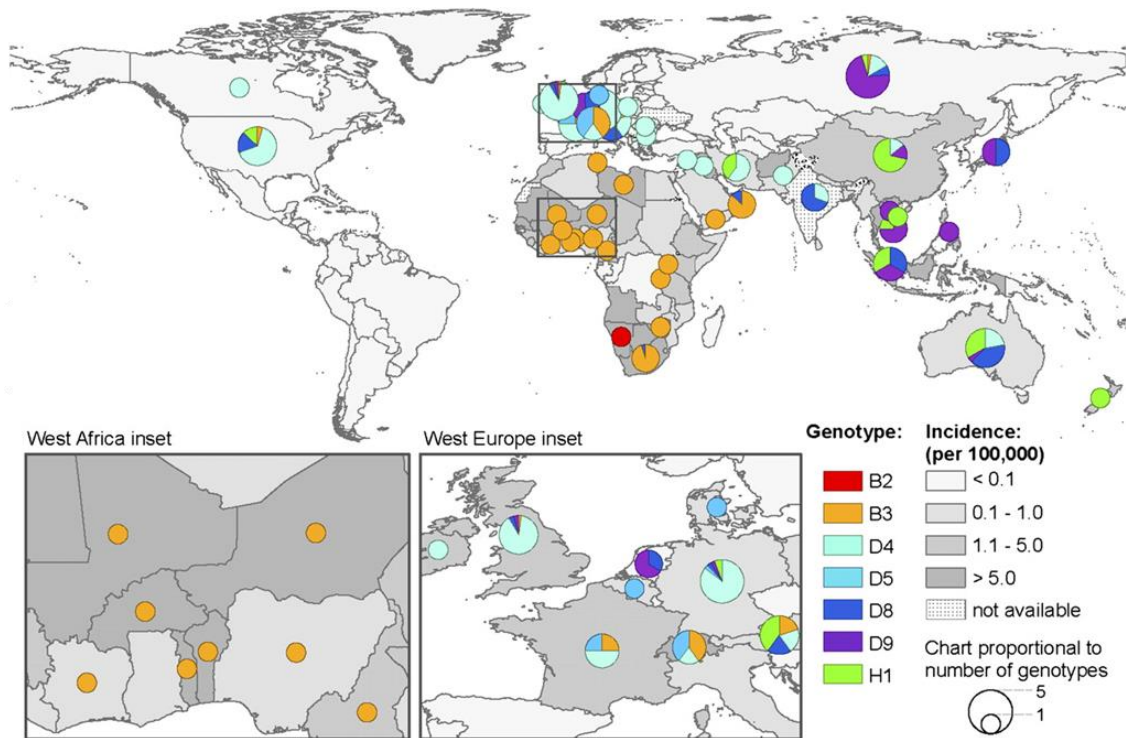
MV is a pleomorphic virus ranging in diameter from 120 to 300 nm (Fig. 9) (243, 244). The lipid envelop is derived from the plasma membrane of the host cell and is traversed by 2 viral trans-membrane glycoproteins: haemagglutinin (H) and fusion (F). The matrix (M) protein lines the inner surface of the virion and surrounds the nucleocapsid core that typically contains several RNA strands (245). The MV genome is tightly encapsidated by the nucleocapsid (N) protein that has a coiled helical structure. The genome is also associated with two other proteins, the polymerase (L) and a polymerase cofactor (phosphoprotein, P), which together with NP form the ribonucleoprotein (RNP) complex. The negative sense RNA genome (15,894 nt) contains six genes: nucleoprotein (NP), phospho (P), matrix (M), fusion (F), haemagglutinin (H) and the large (L) (238). Each transcription unit is flanked by a short leader and trailer sequence, containing the genomic and antigenomic promoters.



**Figure 9:** Schematic representation of the Measles Virus and its genome (238)

The NP mRNA is the first to be transcribed, and the NP protein (525 AA) is the most abundant protein in the virion and the infected cell. NP polymers entirely cover the RNA genome and protect it from nucleases. The NP protein associates with the P/L polymerase complex to form the RNP, and is therefore required for transcription and replication. The NP protein is the first target of the immune response but anti-N antibodies do not have the ability to neutralize the virus (246). The P protein (507 AA) links L to NP to form the replicase complex (247). The P gene encodes for the P protein as well as 3 non-structural proteins (C, V and R proteins) that are generated by alternative initiation of protein translation, mRNA editing and protein truncation (248-250). V and C proteins are involved in the regulation of transcription and translation, as well as in immune evasion mechanisms (251-254). The M protein (335 AA) links the RNP complex with H and F during assembly. It plays a central role in the maturation, stabilization and budding of the virion (255). The F protein (550 AA) is a trans-membrane protein composed of 2 subunits (F1 and F2) generated by enzymatic cleavage of an inactive precursor (F0) (256). The F protein associates with the H protein to mediate membrane fusion and virus entry (238, 241). The H protein (617 AA) is expressed as a disulfide-linked homodimer at the surface of the virus and the infected cell (257). The C-terminal part of the H is involved in receptor binding, while its N-terminal part acts as membrane anchor.

MV is divided into 8 clades A-H which are further sub-divided into 24 genotypes (A, B1-B3, C1 and C2, D1-D11, E, F, G1-G3, H1 and H2) (258). The genotype A is the vaccine strain of MV and does not naturally circulate. Some genotypes are found in one geographic region and others are co-circulating (Fig. 10). Several genotypes (B1, D1, E, F, G1) seem to be extinct or inactive (259). MV genotyping, based on the C-terminal of the NP, is an important tool for epidemiological surveillance, to document chains of transmission and control elimination efforts. The complete H sequence should be obtained if a new genotype is suspected but is rarely performed on clinical samples (258). Laboratory-based surveillance is performed through the world by the World Health Organization Measles and Rubella Laboratory network. Currently, identifying a MV strain is only possible from samples taken during the acute phase of the disease and it seems impossible to retrospectively determine the MV strain that has caused infection and immunity.



**Figure 10:** Global distribution of measles genotypes and measles incidence in 2009 (260)

*MV* genotypes are indicated by colored circles, whose size is proportional to the number of genotypes reported for the indicated areas. Grey-scale indicates *MV* incidence rates. In countries where *MV* is endemic (e.g. West Africa), co-circulation of different strains is uncommon. In countries where *MV* has been eliminated (e.g. Americas), small outbreaks are caused by few imported *MV* strains. In countries with performant *MV* control (e.g. European countries) but with an increasing number of susceptible individuals, *MV* re-introduction leads to large outbreaks associated with single genotypes (258).

### 1.5.2 Pathogenesis and immune response

Two main cellular receptors are identified for *MV*. The signaling lymphocyte activation molecule SLAM (CD150), which is expressed on immune cells among others, acts as the principal *MV* cellular receptor accounting for its lymphotropism and immunosuppressive nature. Both vaccine and wild type strains are able to bind to this receptor, while vaccine and laboratory adapted *MV* strains use the ubiquitous complement regulatory molecule CD46 receptor (261). For natural infection, the distribution of SLAM explains most but not all aspects of *MV* pathogenesis (262). A yet unidentified epithelial receptor molecule (epR) has been recently located at the basolateral side of polarized cells (263, 264), supporting the model of wild type *MV* entering the host at the alveolar level, through the infection of macrophages and dendritic cells (265).

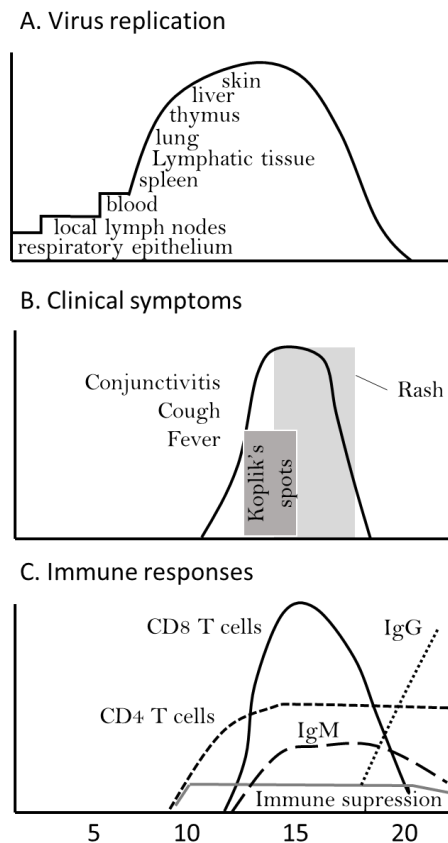
The virus proliferates regionally in the bronchus-associated lymphoid tissue (BALT) and local lymph nodes, which leads to a primary viremia (Fig. 11A). Lymphocytes, monocytes, macrophages, etc are infected. A second viremia takes place 5 to 7 days post-exposure when the virus spreads to various organs:

lymph nodes but also kidneys, gastrointestinal tract, liver, skin, ... (241). Leukocytes are thus key components of MV pathogenesis, as they are targets of the immunosuppressive effect of the MV and they disseminate the virus throughout the body (266-268).

The first non-specific symptoms are fever, cough, coryza and conjunctivitis, and are followed by the appearance of the Koplik's spots, the first pathognomonic signs (Fig. 11B). A characteristic macropapular rash follows the prodrome period of 2-3 days and lasts 3-5 days. The onset of the rash corresponds to the activation of the MV-specific humoral and cellular immunity and initiation of virus clearance. In uncomplicated cases virus clearance occurs 7-10 days after the onset of the rash. Natural MV infections trigger a lifelong immunity (269).

The MV induces a long-lasting immunosuppression that starts with the onset of the clinical symptoms and persist for weeks (270, 271). During this period, affected patients are more susceptible to secondary bacterial and viral infections. These can lead to severe complications (pneumonia, diarrhea, otitis media, encephalitis ...) that are responsible for the high measles mortality rate in developing countries (272). Measles case fatality rates vary depending on various factors including age at infection, nutritional status, and access to healthcare. Before the advent of measles vaccines, the measles case fatality rate was as high as 30%. It has now dropped to <0.5% in industrialized countries but can rise up to 5% to 10% in developing countries, even reaching up to 25% during emergency setting (e.g. refugee camps) (273). Rarely (1 per 100,000 cases), measles can cause a fatal degenerative disease of the central nervous system called subacute sclerosing panencephalitis (SSPE).





**Figure 11:** Basic pathogenesis of MV infection (238)

In immunocompetent hosts, MV infection triggers an effective innate response followed by an adaptive cell-mediated and humoral immunity (Fig. 11C). Thus, MV induces both a significant generalized immune suppression, which is responsible for most of the mortality associated with measles, as well as antiviral mechanisms that ultimately result in life-long protection (274, 275). While infection is thought to end with full viral clearance, some authors report the persistence of viral RNA in various organs and in PBMC of asymptomatic patients which have been either infected or vaccinated up to years before (276-279). It is unknown if this persistence has a role in the life-long immunity conferred by MV infection and vaccination.

MV infection induces both CD4+ and CD8+ T cells. CD8+ cells are activated during the prodrome phase (Fig. 11C). A CD4+ Th1 response occurs during the acute phase and is essential for virus clearance. A CD4+ Th2 response profile is typical of the convalescent phase and promotes the induction of MV-specific antibodies. Anti-MV IgMs are detectable in the blood already 72h after the onset of the rash and last for ~2 months (Fig. 11C) (280). IgA and IgG become detectable a few days after the onset of the rash, and IgG typically reach their maximal concentration in the blood 3-4 weeks later. IgG are mainly directed against the abundant N protein, followed by the H, F and M proteins (246, 281, 282). MV-

neutralizing antibodies are exclusively directed against the MV H (90%) and F (10%) transmembrane proteins, and mostly recognize conformational epitopes (283-285). IgG1 and IgG3 are predominant during the acute phase of the disease, whereas IgG1 and IgG4 dominates the convalescent phase and long-term immunity.

The MV is considered to be antigenically relatively stable and serologically monotypic, antibodies produced against one genotype of measles being cross-protective with all other genotypes (286). However, sera from vaccinees and from naturally infected individuals differ in their ability to neutralize wild type MV strains (283, 287, 288). Similarly, MV strains can be discriminated by mAbs (289, 290). These observations suggest a certain variation in antigenic properties of the different strains, and it is therefore anticipated that the IG repertoire is also differently modelled by diverse MV strains.

### 1.5.3 Measles vaccines and eradication efforts

The isolation of MV in culture from the blood of an affected child in 1954, the so-called Edmonston isolate, rapidly led to the development of both inactivated and live attenuated virus vaccines. In the early sixties, formalin and tween-ether inactivated vaccines were licensed for the American market (291). These vaccines however triggered moderate levels of neutralizing antibodies, and antibody titers were declining rapidly (292). Furthermore, vaccinated individuals were at risk of developing a more severe form of the disease, called atypical measles, upon infection with wildtype measles (293). These problems, combined with the concomitant development of more effective live attenuated vaccines, led to their withdrawal in 1967 (294).

Several live attenuated measles vaccines have been developed by adaptation of laboratory MV isolates, originally propagated in human or simian cells to cells of nonsusceptible hosts (295-298). The Edmonston B strain was the first live attenuated measles vaccine to be developed by serial passages of the Edmonston strain in chick embryo fibroblasts, and was first licensed in 1963 (299). This vaccine was however associated with a high frequency of fever and rash in vaccinated children (300). Further passage in chick embryo fibroblast led to the development of more attenuated vaccines (301, 302). Most available attenuated vaccines are still derived from the original Edmonston isolate: Schwarz and Edmonston Zagreb vaccines are widely used worldwide, whereas the Moraten strain is primarily used in the United States. All MV vaccines strains belong to genotype A, and few differences (nucleotide sequence divergence, immunogenicity...) have been described among the different MV vaccine strains regardless

of the geographic origin of the parent virus or the passage history (301). The genotype A has now ceased to be circulating. The vaccine genotype is therefore distinct from all currently circulating MV strains.

Live attenuated vaccines are inactivated by light and heat, and a cold chain must be maintained, which complicates the logistics during MV immunization campaigns. Relatively stable in their lyophilized form, the reconstituted live attenuated vaccines rapidly loses infectivity when kept un-refrigerated (303).

Although attenuated MV strains have a lower ability to replicate *in vivo*, they induce humoral and cellular immune responses qualitatively similar to that induced by wildtype strains, although titers after vaccination are lower and wane more rapidly than after natural infection (304–309). Antibodies become detectable in the blood 12–15 days after vaccination and peak at 1–3 months (294). Both maternally acquired antibodies and immunological immaturity of infants reduce the efficacy of measles vaccination and hinder the effective immunization of young children. In areas where measles remains prevalent, measles primo-vaccination is routinely performed at 9 months, whereas in areas with low measles incidence, vaccination is recommended at 12 months old. A high level of population immunity (>95% immunity coverage) is required to prevent endemic MV transmission, and a second vaccine dose is necessary to protect individuals that did not respond to the first immunization (310, 311). This second boost is usually performed at the age of five or six years, or at a later age during supplementary immunization activities (SIAs). In many countries, multivalent live attenuated vaccines to protect against measles, mumps, rubella, and sometimes varicella, are routinely employed (312).

MV-containing vaccines have greatly contributed to a dramatic reduction in the number of measles-related deaths worldwide. In the prevaccine era, over 130 million cases and 2.5 million measles-related death occurred annually (313). In the early 1980s, around 4 million cases of measles were declared per year. This number had been reduced to under 120,000 in 2014 thanks to large scale vaccination efforts, led in part by the Measles and Rubella Initiative, a common initiative by various international organisations since 2001. This translated in the reduction of the worldwide measles mortality by 79% between 2000 and 2014 alone. The American continents have been free of indigenous measles case since 2002. This trend is however declining due to reduced political and social commitment. Unprotected children are getting more numerous, causing large outbreaks. In 2014, an estimated 20.6 million infants did not receive the first routine measles vaccination and 314 children died of measles each day.

MV infection of immune individuals causes only a mild form of the disease or can remain asymptomatic. Evidence has been put forward that the MV might survive and circulate in fully vaccinated populations (314). Immune individuals in these populations have demonstrated asymptomatic secondary immune

responses without any new MV vaccine inoculation and in the absence of clinical measles in the population (315, 316). It has been speculated that individuals prone to develop asymptomatic secondary immune responses, those with low MV-specific antibody titers, might support viral transmission (317). The very absence of clinical symptoms makes the documentation of such cases difficult. No conclusive evidence of viral transmission between asymptomatic immune individuals has of yet been published. Nevertheless, if confirmed, this process could have important epidemiological consequences and even make the eradication of the virus impossible (318-320).

# **Chapter 2. Aims**

In the context of the measles elimination efforts of the World Health Organization (WHO), new epidemiological tools enabling the retrospective identification of the MV strain that caused infection and immunity would be of particular interest. Although some variations in the antigenic properties between the different MV genotypes have been reported, it is presently impossible to determine *a posteriori* the strain of the virus responsible for the immunity in an individuals as all MV clades form a single serotype.

In this thesis, we aimed at identifying MV-specific CDR3 sequences that could be successfully used in tracing back the person's immune history. For this purpose we immunized an animal model with a fully human IG repertoire (OmniRats) using different types of vaccine (adjuvanted MV antigens, MVA vector, and DNA vaccination) and different immunization routes (footpads, IP, IM). We decreased the variety of MV antigens used for immunization in a stepwise manner (whole sonicated MV, recombinant MVA expressing H and F, and H only via DNA vaccination). We needed to verify whether the measles virus is able to trigger a public response, and whether we could identify conserved and strain specific signatures in the IG repertoire of OmniRats using next-generation sequencing technologies. First we produced a hybridoma library and identified VH CDR3 sequences of known antigen specificity from the MV-specific hybridoma clones. These sequences can validate the antigen specificity of shared immunoglobulins. Next, the IG repertoire of immunized rats was sequenced using the Ion Torrent platform in order to identify Measles virus signatures within the VH CDR3 repertoire. The immunization of OmniRats with two phylogenetically distant MV clades was subsequently performed to investigate whether they differentially impact the IG repertoire. This project aimed at developing an approach to identify antigen-specific signatures without previous enrichment in antigen-specific B cells.

The cross comparison of the different studies presented in this thesis allowed to identify IG signatures for a vaccine and a wild type strain from the public IG repertoire of humanized rats. These signatures can potentially be used as markers for particular antigens and for the sequence-based monitoring of infectious diseases or vaccination campaigns.

## Chapter 3.

# Development of MV-specific human monoclonal antibodies in transgenic humanized rats

*High-resolution analysis of the B cell IG repertoire before and after polyethylene glycol fusion reveals preferential fusion of rare-antigen specific B cells. Human Antibodies. Dubois et al., in press*

## 3.1 Abstract

The hybridoma technology is one of the most important advances in clinical immunology. However, little is known about the differences between the antibodies produced during the *in vivo* immune response and those recovered in hybridoma libraries. Here, we investigate a potential fusion bias inherent to the hybridoma production process. Transgenic rats carrying human IG heavy and light chain loci were immunized with measles virus (MV) to generate human mAbs. Using high-throughput sequencing of IGH mRNA, we compared the IGH repertoire of lymph nodes and the derived hybridoma library using the sequences of the MV-specific hybridoma clones as a reference set with known specificity. We observed that large clonotypes from the lymph nodes were not represented in the hybridoma library, but low-frequency B cell populations became highly enriched and most hybridoma clones were derived from these. Our data also showed that identical VH CDR3 evolved from diverse V-(D)-J rearrangements, indicating convergence of different B cells subpopulations towards expression of antibodies with similar paratopes. In conclusion, the efficient generation of mAbs results from a fusion process highly selective for rare antigen-specific B cells rather than *in vivo* expanded populations. Antibodies of particular interest may therefore be missed during classical hybridoma production.



## 3.2 Introduction

Since its first description by Köhler and Milstein, the hybridoma technology has been extensively used with great success to generate highly specific and high-affinity mAbs (207, 321). However, although millions of antigen-specific B cells are generated during an immunization, orders of magnitude less antigen-specific hybridoma clones can typically be generated from a single experiment because of the low efficiency of the process (208). As a result, only less than 1% of the repertoire of antigen-specific B cells can be tested for antigen-reactivity, and the majority of the antigen-specific B cells are lost.

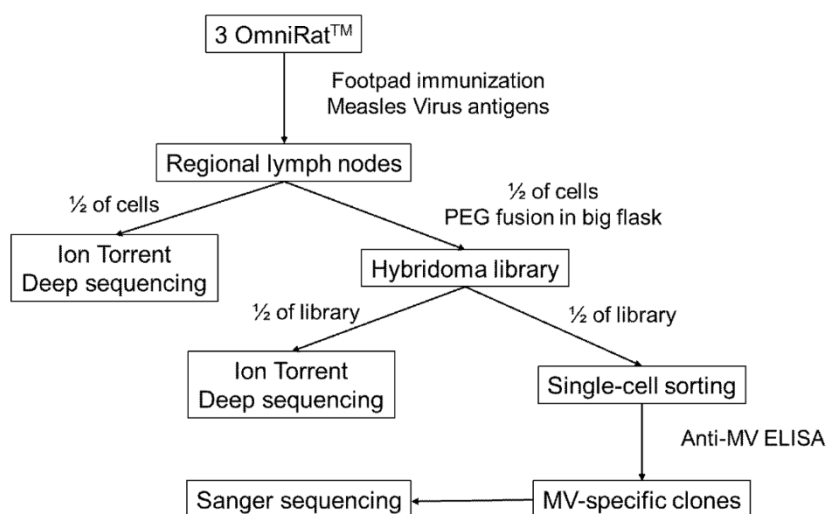
The immortalization of antibody-secreting cells by fusion to myeloma cells is the most critical step in the generation of mAbs. Fusion efficiencies between  $5 \times 10^{-6}$  for the traditional polyethylene glycol (PEG) fusion and up to  $4 \times 10^{-3}$  for electrofusion methods have been reported (322). Despite these low fusion efficiencies the generation of antigen-specific hybridoma cells is successful and does not seem to critically undermine the production of mAbs, probably because antigen-stimulated B cells are more susceptible than naïve resting B cell to fusion with myeloma cell. Köhler and Milstein already suspected that fusion to myeloma cells may involve a “restricted fraction of the spleen cell population, probably cells committed to antibody production” (207). Some early studies suggested that antibody-secreting hybridomas are the results of fusion between plasma and/or blast cells and myeloma cells (323-325). Although it is an important question, even today little is known about the cells that are susceptible to fusion and to what extent B cells with the desired antigen-specificity are enriched or lost upon fusion.

Antigen-specificity is encoded by the variable domain of the IG heavy (IGH) and IG light kappa (IGK) or lambda (IGL) chains. The complementary determining region 3 (CDR3) of the variable domain of the IGH chain (VH) encodes the principal antigen recognition site and is the most important determinant of the antigen specificity of an antibody (25). The VH CDR3 results from the V-(D)-J junction. Following antigen recognition, antigen-reactive naïve B cells start to proliferate in lymph nodes while their V domain nucleotide sequences undergoes DNA somatic hypermutation.

The use of therapeutics rodent mAbs is limited because of their immunogenicity (216) while human mAbs are well tolerated but difficult to produce (326, 327). Chimeric and humanized antibodies have been developed, but they still retain non-human sequences. The development of mAbs directly from human material is complicated by the dearth of antigen-specific cells in the blood, and a poor fusion efficiency paired with a low and unstable IG production of human hybridomas (322). To overcome these and other limitations, genetically engineered animals expressing a human antibody repertoire have been

engineered (218, 219, 328-331). In this context, Open Monoclonal Technology Inc. recently developed a platform for the enhanced production of human antibodies based on transgenic humanized rats, the OmniRat™ (235). These animals carry a chimeric human/rat IGH locus, where 22 human IGHV, the 27 IGHD and the 6 functional IGHJ are linked to the rat IGHC, together with the complete IGK and IGL loci (235, 331, 332). They produce no rat but human/rat chimeric IG, with serum IgM and IgG levels similar to those of wild-type rats. The rat IGHC region can be easily replaced *in vitro* by a human IGHC region to produce fully human mAbs of therapeutic interest. These rats have been used to efficiently generate high-affinity humanized mAbs (235).

Here, we immunized these human IG transgenic rats with Measles Virus (MV) antigen via footpad immunization and used the regional lymph nodes to generate a hybridoma library by PEG-mediated fusion with myeloma cells (Fig.1). A total of 290 MV-specific hybridoma clones were obtained and their VH loci sequenced. The B cell IgG repertoires in the lymph nodes and in the corresponding hybridoma library were investigated at the level of the VH V-(D)-J sequences with an Ion Torrent Personal Genome Machine (PGM™) Deep Sequencing platform. The VH sequences of the hybridoma clones served as a reference set with known specificity to investigate the inherent bias of PEG-fusion. As expected the bulk of B cells in the lymph node did not fuse. We found that only B cells that were exceptionally rare in the lymph node were highly susceptible to fusion and that virtually all antigen-specific hybridoma clones were derived from these rare B cell populations. This highly selective fusion bias seems to be critical for the success of the hybridoma technology.



**Figure 1:** Study design

A group of 3 humanized IG transgenic rats (OmniRat™, Open Monoclonal Technology Inc., Pablo Alto, USA) were immunized with MV antigens. At sacrifice inguinal and popliteal lymph nodes were collected. The lymphocytes from the 3 immunized OmniRats were pooled. Half of the pooled cells were fused to myeloma cells to produce hybridoma cells while the other half was further processed for analysis of the IG repertoire by deep sequencing (MV-LN sample). After 10 days of growth, half of the bulk hybridoma library was processed for deep sequencing (MV-HB sample). With the other half, single hybridoma clones were sorted in 96-well plates and the supernatant was tested for MV-specific IgG by indirect ELISA. The VH loci of the MV-positive clones were sequenced individually by Sanger sequencing.

## 3.3 Materials and methods

### 3.3.1 Immunization and generation of hybridoma clones

Humanized IG transgenic rats (OmniRat™, Open Monoclonal Technology Inc., Pablo Alto, USA) were developed and bred as recently described (235, 331-333). Seven animals were maintained on a 12h light/dark cycle with *ad libitum* access to food and water. All animal procedures were in accordance with the guidelines described in the Guide for the Care and Use of Laboratory Animals (334).

Four unimmunized rats were used as controls. Three rats were immunized with 40ug of MV antigen (Measles grade 2 Antigen, Microbix Biosystems, Mississauga, CA) in Ribi adjuvant into each of the hind footpads twice a week during four weeks, followed by a ninth injection without adjuvant 3 days before sacrifice. Half of the pooled lymphocytes of the popliteal and inguinal lymph nodes were lysed with Buffer RLT (Qiagen, Venlo, NL) and processed for analysis of the IG repertoire by deep sequencing. The other half was fused to mouse P3X63Ag8.653 myeloma cells by PEG (335) and grown in selection medium. After 7 days, dead cells were removed by density gradient centrifugation and hybridomas returned into culture in growth medium for 3 days. Half of the bulk hybridoma library was then lysed in Buffer RLT and processed for deep sequencing. The other half was stored in medium/DMSO at -80°C in aliquots until FACS cloning of single hybridoma cells into thirty 96-well plates. After expansion of the clones, supernatants were tested by ELISA. RNA was extracted from the antigen-specific clones (RNASpin mini kit, GE Healthcare, Buckinghamshire, UK) and cDNA was prepared using oligodT and Promega RT (Promega, Southampton, UK) at 42°C for 1h. PCRs using VH leader primers with rat  $\gamma$ CH2 primer (Table 1) and GoTaq Green Master mix PCR (Promega) were performed as described (235). The products of the expected size were gel purified, Sanger-sequenced and annotated as described below.

### 3.3.2 Anti-MV IgG ELISA

Anti-MV IgG antibody levels were determined in serum of the rats, in the supernatants of the hybridoma libraries and of the hybridoma clones. 384-well microtiter plates (Greiner bio-one, Wemmel, BE) were coated overnight at 4°C with 0.25  $\mu$ g of MV Ag in carbonate buffer (100 mM, pH9.6). Free binding sites were saturated with 1% BSA in TBS at room temperature for 2h. Serial dilution of the sera or hybridoma supernatants were added for 90 min at 37°C, and developed with alkaline phosphatase (AP)-conjugated goat anti-rat IgG (1/750 dilution, ImTec Diagnostics, Antwerp, BE) and the appropriate substrate at 405 nm. Endpoint titers (EPT) were determined as the serum or supernatant dilutions corresponding to

5 times the background. The cut-off value for specificity of the hybridoma clones was set at 0.2 OD corresponding to 2.5 times the average OD of the background ( $0.08 \pm 0.0067$ ). The positive clones were also tested for anti-BSA IgG and no reactivity was found. IgG isotype subtypes were tested using subtype specific AP-conjugated goat anti-rat IgGs (IgG1 1/750 dilution, IgG2b, 1/500 dilution, ImTec Diagnostics).

### 3.3.3 Sample preparation, amplification and Ion Torrent Deep Sequencing

Total RNA was extracted from the lymphocytes and hybridoma cells (RNeasy midi kit, Qiagen) and mRNA was enriched by paramagnetic separation ( $\mu$ MACS mRNA Isolation kit, Miltenyi Biotech, Leiden, NL). cDNA was prepared using dT18 primers and Superscript III RT (Invitrogen, Merelbeke, BE) at 50°C for 80 min. The VH domains were amplified by PCR for 35 cycles in two 25  $\mu$ l reactions using a set of primers designed to cover all the IGHV genes carried by the OmniRat<sup>TM</sup>, a rat CH primer (Table 1) and Q5 Hot Start High Fidelity polymerase (New England Biolabs, Frankfurt-Main, DE). 50  $\mu$ l of the PCR products were loaded on a 2% agarose gel, amplicons were size selected (QIAQuick Gel Extraction Kit, Qiagen) and quantified (Quant-iT PicoGreen dsDNA, Invitrogen). Quality was checked with a Bioanalyzer (High Sensitivity DNA, Agilent Technologies, Diegem, BE). Libraries were pooled in equimolar concentrations and sequenced on a 318<sup>TM</sup> Chip v2 (Life Technologies, Gent, BE) using multiplex identifiers (MIDs) with the Ion OneTouch<sup>TM</sup> Template OT2 400 Kit (Life Technologies) and the Ion PGM<sup>TM</sup> Sequencing 400 Kit (Life Technologies) on the Ion Torrent Ion PGM<sup>TM</sup> System (Life Technologies). Lymphocytes from the lymph nodes of unimmunized OmniRats were similarly treated and used as negative controls.

### 3.3.4 Quality Control and annotation of the deep sequencing results

Only sequencing reads ranging from 240 to 300 nt, with an unambiguous MID (no mismatches allowed), primers at both ends (2 mismatches allowed) and a quality score of >20 in at least 80% of the bases in the sequence were included in the analysis. The primer sequences were trimmed and the sequences were collapsed to unique nucleotide (nt) sequences. Sequences were next submitted to the IMGT/HighV-QUEST webserver allowing for indels correction (IMGT, [www.imgt.org](http://www.imgt.org); MP. Lefranc, Montpellier, France) to annotate the different V gene regions, and in particular the CDR3 (143). V, D, and J genes for

the in-frame productive sequences were subsequently assigned using a local installation of IgBlast (336), including only the genes present in the genome of the OmniRat<sup>TM</sup> as a reference set. Only sequences with an unambiguously assigned V and J gene (Blast E-value  $< 10^{-9}$ ) were considered for further analysis. Outputs from IMGT/HighV-QUEST and IgBlast were cross-compared, and reads still containing indels were discarded. Indel-corrected sequences were subsequently collapsed to unique nt sequences over the gapped-nt V-(D)-J sequence as provided by IMGT/HighV-QUEST. As expected because of the blunt-end ligation of the Ion Torrent adaptors to the amplicons at the sequencing library preparation stage, the reverse and forward reads were equally represented in all samples and at all steps of the analysis (not shown). As identical observations were made on forward and reverse reads, both were combined. All analytical scripts were written in python and R. Circos diagrams displaying were generated using the Circos Table Viewer v0.63-9 8 (<http://circos.ca>). B cell IG network plots were produced based on gapped nt-sequences and hamming distance of 1, with a script written in R, similarly to Bashford-Rogers *et al.* (337). The Ion Torrent platform is particularly prone to indels. Therefore, to determine whether our cleaning and annotation pipeline effectively removed most of the indels introduced by sequencing, we analyzed the IgG datasets of two mice hybridoma clones. The indel rates before quality control (output reads) were of 1.48% and 1.05% of all base pairs, which is in line with or below the rates previously reported (128, 129, 338). After quality control, the indel rates dropped to 0.43% and 0.49%. After the indel correction by IMGT/HighV-QUEST, we calculated an indel rate of 0.08% and 0.22% amongst all bases. After annotation with IgBlast, and considering only the productive sequences, we were able to generate datasets that had removed indels as detected by IMGT (140). Sequences are available on NCBI's Sequence Read Archive (SRA, BioProject ID PRJNA297405).

**Table 1:** Primer list

	Name	Primer sequence
Sanger sequencing	VH1 Leader	5'-atggactggacctggaggatcc-3'
	VH2 Leader	5'-tccacgctcctgctgtgac-3'
	VH3 Leader	5'-atggagtttgggctgagctgg-3'
	VH4 Leader	5'-tgaaacacctgtggttcttcc-3'
	VH6 Leader	5'-tcattcttggccgtgctgg-3'
	VH4-39 Leader	5'-tggagtggattgggagt-3'
	Rat $\gamma$ CH2	5'-gggaagatgaagacagatg-3'
Ion Torrent sequencing	VH1-FR2	5'-ctggacaagggcttgagtgg-3'
	VH2-FR2	5'-caggaaggccctggagtgg-3'
	VH3-FR2	5'-caggaaggggctggagtgg-3'
	VH4-FR2	5'-caggaagggactggagtgg-3'
	VH6-FR2	5'-catcgagaggccttgagtgg-3'
	Rat IgG CH1	5'-ggatagacagatggggctgttgtt-3'

## 3.4 Results

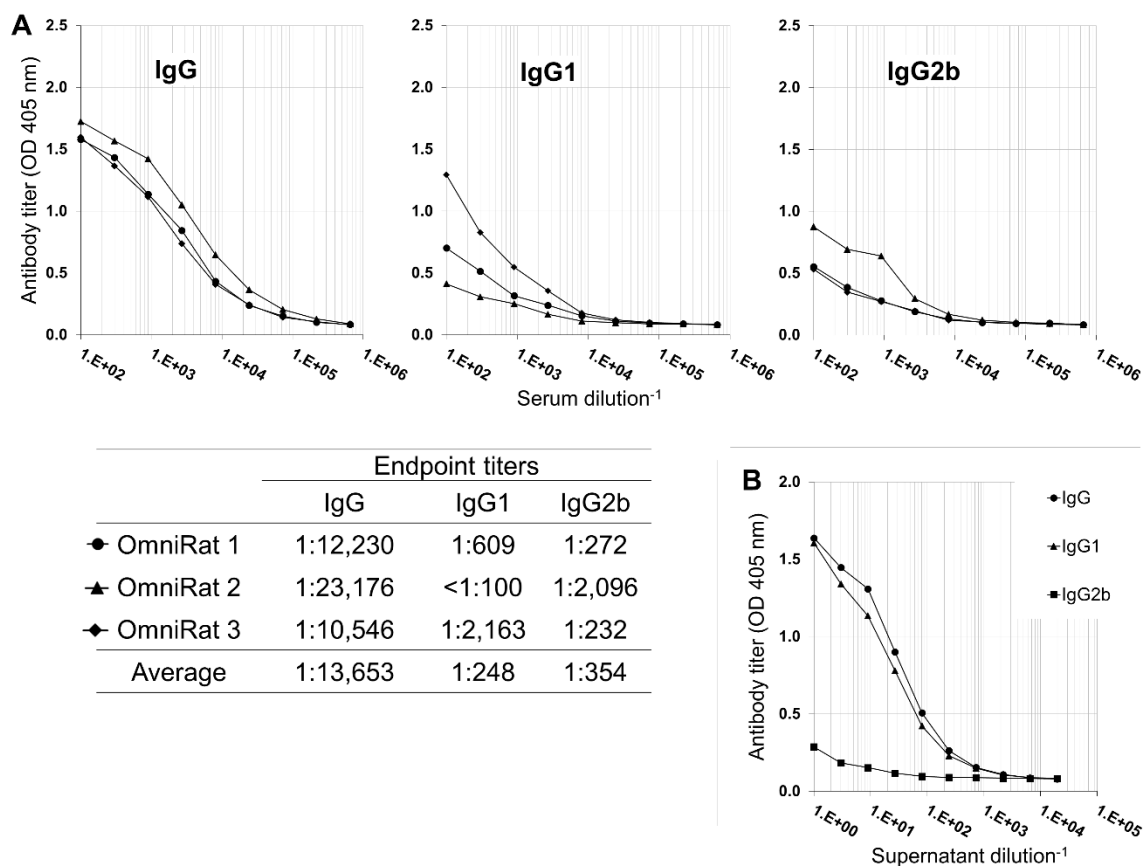
### 3.4.1 Generation and identification of MV-specific hybridoma clones

Three OmniRats were immunized with MV antigens and their IgG levels were measured in sera by ELISA. At sacrifice, similar antigen-specific IgG levels were detected in the sera of the MV-immunized rats (EPT IgG: 1:12,230, 1:23,176 and 1:10,546; average 1:13,653) although proportions of IgG1 (EPT IgG1: 1:609, <1:100 and 1:2,163; average 1:248) and IgG2b (EPT IgG2: 1:272, 1:2,096 and 1:232; average 1:354) isotype subclasses varied between animals (Fig.2A). Sera of the 4 control rats were negative.

The lymphocytes from the popliteal and inguinal lymph nodes of the three MV-immunized animals were extracted and pooled. Half of the pooled lymphocytes were used to produce a hybridoma library by PEG-mediated fusion to myeloma cells, while the other half were processed for the analysis of the IG repertoire on the Ion Torrent PGM™ platform. Based on the numbers of lymphocytes and myeloma cells used for fusion and the number of hybridomas in the library, we calculated that one cell in about  $3 \times 10^3$  cells underwent fusion (fusion efficiency of  $3.125 \times 10^{-4}$ ). Anti-MV IgGs in the hybridoma library were mainly of the IgG1 isotype subclass (EPT IgG: 1:158; EPT IgG1: 1:91) (Fig.2B).

Half of the bulk hybridoma library was then processed for deep sequencing, and the other half stored in aliquots before single-cell sorting into 96-well plates. A total of 2,912 single cells of the hybridoma library were cloned. 290 clones produced MV-specific IgG antibodies. The IGH loci of these positive clones were Sanger-sequenced, annotated and grouped according to identical amino acid (AA) sequences within their CDR3 region (Table 2). A total of 213 IGH sequences were obtained, coding for 23 different CDR3 AA sequences (designated A to W, and referred to as MV CDR3) evolving from 20 different V-(D)-J rearrangements. The vast majority of the sequences with an identical CDR3 AA were also identical at the nucleotide (nt) level over the whole IGH region. Three MV CDR3 (L, M and N) shared the same V-(D)-J recombination and had a high percent of identity (>96,1% nt and 94,1% AA identity), indicating they are clonally related. Similarly, the MV CDR3 V and W shared the same V-(D)-J recombination and a high level of similarity with each other (89.7% nt identity, 76.9% AA identity).





**Figure 2: MV ELISA**

*Anti-MV IgG, IgG1 and IgG2b levels in the sera (A) and in the supernatants of the hybridoma library (B) were measured by indirect ELISA. The serum dilution<sup>-1</sup> (1:100 to 1:656100) or supernatant dilution<sup>-1</sup> (1:1 to 1:19683) was plotted against binding measured by adsorption at 405 nm. Individual sera were titrated by three-fold dilutions to determine endpoint titers (EPT; A). Similar anti-MV IgG antibody levels were detected in the three animals (average EPT: 1:13,653). The immune response of the different rats included varying proportions of both IgG1 (average EPT: 1:248) and IgG2b (average EPT: 1:354) isotype subclasses. The antibodies in the hybridoma supernatant were largely of the IgG1 subtype (EPT: 1:91; B).*

**Table 2:** Characteristics of the IGH chimeric human-rat gamma chain of the 23 MV-specific hybridoma clones sequenced by Sanger sequencing

ID	CDR3 AA sequence	Frequency	IGHV	IGHD <sup>a</sup>	IGHJ	CDR-IMGT lengths (AA) <sup>b</sup>	V germline nt identity (%) <sup>c</sup>	IgG subclass <sup>d</sup>
A	AHRRDSGGFDY	1 (0.5%)	V2-5	D6-19	J4	[10.7.11]	99.7	rat $\gamma$ 2b
B	AKDMGITVVRGAGDYYYGMDV	5 (2.3%)	V3-9	D3-10	J6	[8.8.22]	98.7	rat $\gamma$ 1
C	AKEGGYYSFDY	1 (0.5%)	V3-23	D3-22	J4	[8.8.11]	98.3	rat $\gamma$ 2b
D	AKLYSSGLDY	1 (0.5%)	V3-9	D6-19	J4	[8.8.10]	99.3	rat $\gamma$ 2b
E	ARDIRGYRSSWYLFDY	1 (0.5%)	V3-11	D6-13	J4	[8.8.16]	98	rat $\gamma$ 2b
F	AREEGLAAAALDY	1 (0.5%)	V4-28	D6-13	J4	[9.7.13]	93.9	rat $\gamma$ 2b
G	AREGTGYSSGWCFDY	4 (1.9%)	V1-8	D6-19	J4	[8.8.15]	98.9	rat $\gamma$ 1
H	ARGGLPFDC	1 (0.5%)	V1-3	D1-26	J4	[8.8.9]	96.6	rat $\gamma$ 1
I	ARGRDSSGWYPTRYLQH	1 (0.5%)	V4-34	D6-19	J1	[8.7.17]	98.6	rat $\gamma$ 2b
J	ARGRGYSAYGPFDY	1 (0.5%)	V6-1	D5-12	J4	[10.9.14]	99.7	rat $\gamma$ 2b
K	ARHDSGYDWSYWFDF	6 (2.8%)	V4-39	D5-12	J2	[10.7.16]	96.7	rat $\gamma$ 2b
L	ARHMTYYYYGSGSPNFDY	1 (0.5%)	V4-39	D3-10	J4	[5*.7.17]	97.6	rat $\gamma$ 1
M	ARHRTFYYGSGSPNFDY	2 (0.9%)	V4-39	D3-10	J4	[10.7.17]	98.3	rat $\gamma$ 1
N	ARHRTYYYYGSGSPNFDY	77 (36.2%)	V4-39	D3-10	J4	[10.7.17]	98.7	rat $\gamma$ 1
O	ARHYSGSYGY	1 (0.5%)	V1-2	D1-26	J4	[8.8.10]	98.6	rat $\gamma$ 1
P	ARILWFGESGDY	2 (0.9%)	V2-26	D3-10	J4	[10.7.12]	98.7	rat $\gamma$ 2b
Q	ARLSWYFDL	6 (2.8%)	V4-39	n.a.	J2	[10.7.9]	99	rat $\gamma$ 1
R	ARPWGTYRHTFFDY	7 (3.3%)	V4-28	D3-16	J4	[9.7.14]	93.8	rat $\gamma$ 2b
S	ARQSYGAFDI	1 (0.5%)	V1-2	D1-26	J3	[8.8.10]	99	rat $\gamma$ 1
T	ARRRGYSYGLFDY	1 (0.5%)	V1-8	D5-18	J4	[8.8.13]	99.7	rat $\gamma$ 1
U	ARYFDWLFDY	5 (2.3%)	V3-33	D3-9	J4	[8.8.10]	99.3	rat $\gamma$ 1
V	ASIAAAGTDAFDI	86 (40.4%)	V4-31	D6-13	J3	[10.7.13]	97.3	rat $\gamma$ 1
W	ASIAAAGTDIYDT	1 (0.5%)	V4-31	D6-13	J3	[10.7.13]	97.3	rat $\gamma$ 1
		213 (100%)						

<sup>a</sup> From IMGT/HighV-QUEST; n.a.: non-assign.

<sup>b</sup> Average CDR3 AA length (median): 13.4  $\pm$  3.3 AA (13).

<sup>c</sup> Mutation rate over FR1-FR3 region: 1.93%  $\pm$  1.60%; Mutation rate over CDR2-FR3 region: 2.43%  $\pm$  1.57%.

<sup>d</sup> Based on the nt sequence of the rat IGHC.

\*partial

### 3.4.2 Analysis of the IGH repertoire in lymph nodes and hybridoma library by deep-sequencing

The IgG repertoires of the hybridoma library (MV-HB) and the lymph nodes from immunized animals (MV-LN) and non-immunized animals (NI-LN 1 to 4) were sequenced on an Ion Torrent PGM™ deep sequencing platform. The number of output reads ranged from 1,038,952 to 1,610,062 reads per sample (7,785,825 reads in total; Table 3). After quality control, reads were annotated using IgBlast and IMGT/HighV-QUEST webserver, for insertions/deletions (indels) correction. Unproductive reads (sequences containing a stop-codon and/or out-of-frame V-(D)-J sequences) and reads with ambiguously assigned V or J genes were discarded. The above quality control and indel cleaning process was tested using IGH sequence datasets from two unrelated hybridoma clones obtained with the same deep sequencing platform. We confirmed that our procedure removed all indels from these datasets and concluded that our annotated reads of the lymph nodes and hybridoma libraries are essentially free from indels (see Materials and Methods). All further analyses were performed on the filtered dataset of non-redundant nt sequences (i.e. counting only unique sequences over the V-(D)-J region as annotated by IMGT/HighV-QUEST; subsequently referred as the set of nt sequences). The read length for the set of nt sequences ranged from 196 to 254 nt (218 nt on average; from the FR2 region of the V to the end of the J). Analyzing the non-redundant nt reads is a more conservative approach than considering all the reads since it is less influenced by a potential PCR amplification bias and by the levels of mRNA expression per B cell.

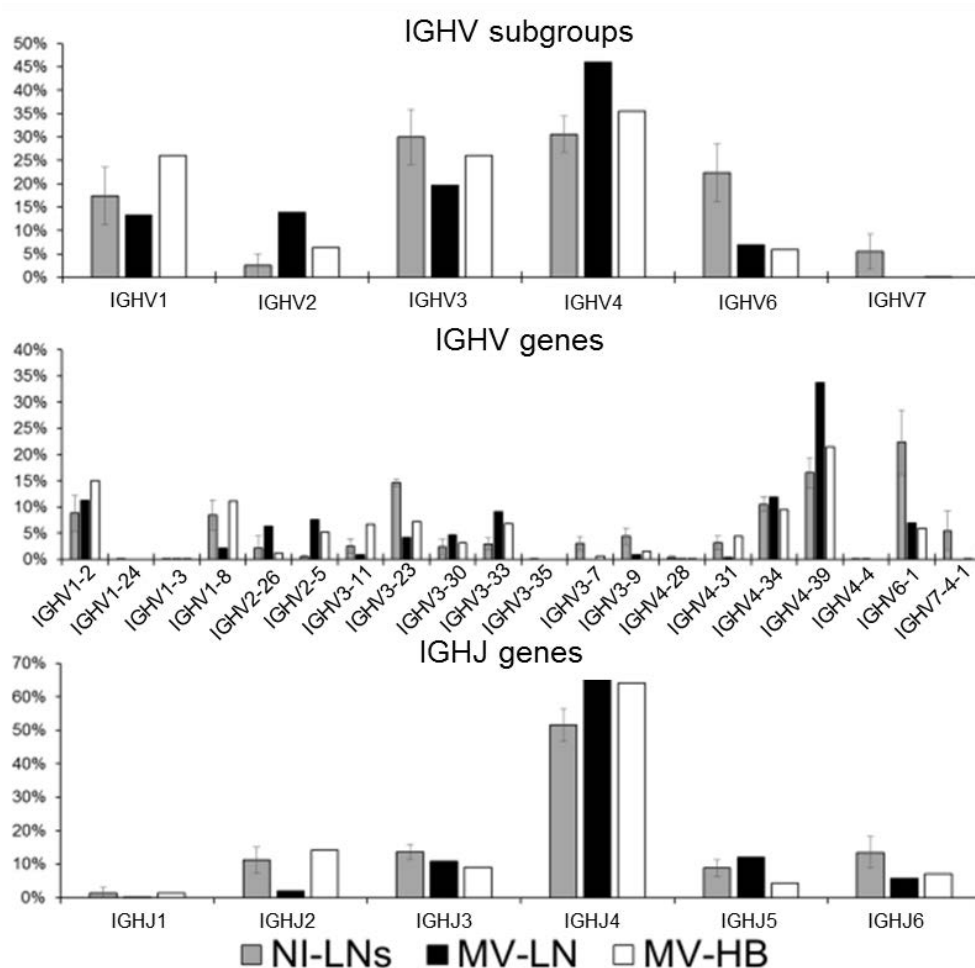
The NI-LNs had up to 8 times more unique nt sequences than the MV-LN (Table 3). Similarly, the number of unique CDR3 AA sequences in the NI-LNs was much higher than in the MV-LN. Thus, the diversity of the IgG repertoire was higher in lymph nodes from non-immunized compared to immunized animals, reflecting an oligoclonal response in the immunized animals.

**Table 3:** Ion Torrent™ output - read numbers before and after quality controls

	MV-LN	MV-HB	NI-LN1	NI-LN2	NI-LN3	NI-LN4
Output reads	1,038,952	1,228,131	1,610,621	1,400,877	1,062,250	1,444,994
Filtered set Non-redundant <sup>a</sup>	14,642	24,042	181,929	149,425	110,059	122,187
Unique CDR3 amino acid	1,744	2,051	11,032	9,990	8,324	6,471

<sup>a</sup> The filtered dataset contain the unique nucleotide sequences after quality control, annotated as productive and with unambiguously assigned V and J gene.

We further analyzed the V and J gene usage in the MV-HB, the MV-LN and the NI-LNs (Fig.3). We identified 20 of the 22 IGHV gene present in the OmniRat's genome. In MV-HB the frequency of certain V genes (IGHV1-8, IGHV3-11, IGHV3-23, IGHV4-31) was favored to the detriment of others (IGHV2-26, IGHV3-30, IGHV3-33, IGHV4-39). Similarly, J2 was used more frequently than J5. When the frequencies of V gene subgroups were compared between samples (Table 4A), there was a very high correlation between all the NI-LNs ( $R^2$  range: 0.77-0.91), as well as between the MV-HB and MV-LN ( $R^2=0.72$ ). However, the correlation between the MV-LN and the NI-LNs was highly perturbed by the MV immunization ( $R^2$  range: 0.23-0.36). J gene usage in all animals was similar and not altered by immunization ( $R^2$  range: 0.86 – 0.99) (Table 4B). Interestingly, the hybridoma fusion process clearly favored certain VJ pairs (Fig. 4 and Table 4C), dramatically reducing the correlation of VJ frequencies between MV-HB and MV-LN ( $R^2=0.42$ ). This is indicative of a strong shift of VJ genes usage introduced by the hybridoma fusion process.



**Figure 3:** IGHV and IGHJ gene usage in the NI-LNs, MV-LN and MV-HB

*Standard deviation (SD) for the NI-LNs combined are shown.*

**Table 4:** IGHV and IGHJ gene usage – Coefficient of determination  $R^2$ 

A	V genes					
	MV-HB	MV-LN	NI-LN1	NI-LN2	NI-LN3	NI-LN4
MV-HB	1	0.72*	0.40*	0.26 <sup>Δ</sup>	0.52*	0.41*
MV-LN		1	0.24 <sup>Δ</sup>	0.23 <sup>Δ</sup>	0.43*	0.36*
NI-LN1			1	0.87*	0.87*	0.77*
NI-LN2				1	0.84*	0.84*
NI-LN3					1	0.91*
NI-LN4						1

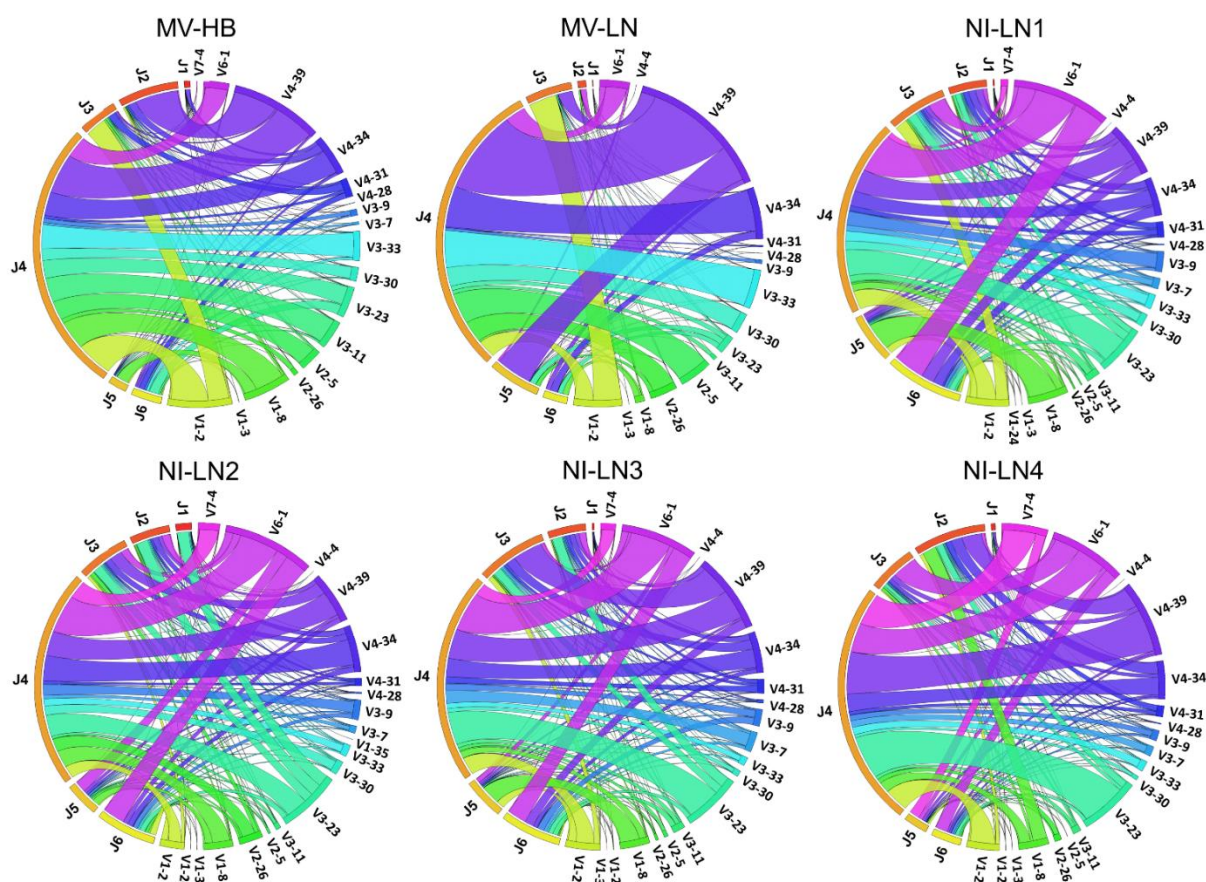
B	J genes					
	MV-HB	MV-LN	NI-LN1	NI-LN2	NI-LN3	NI-LN4
MV-HB	1	0.93*	0.88*	0.97*	0.92*	0.99*
MV-LN		1	0.91*	0.96*	0.94*	0.91*
NI-LN1			1	0.96*	0.98*	0.86*
NI-LN2				1	0.98*	0.94*
NI-LN3					1	0.91*
NI-LN4						1

C	V and J combination					
	MV-HB	MV-LN	NI-LN1	NI-LN2	NI-LN3	NI-LN4
MV-HB	1	0.42*	0.34*	0.31*	0.31*	0.40*
MV-LN		1	0.20*	0.28*	0.25*	0.25*
NI-LN1			1	0.67*	0.73*	0.51*
NI-LN2				1	0.74*	0.60*
NI-LN3					1	0.58*
NI-LN4						1

Significance F test: \*:  $p \leq 0.01$ ; <sup>Δ</sup>:  $0.01 < p \leq 0.05$

Color code $R^2$	0.00	0.25	0.50	0.75	1.00
------------------	------	------	------	------	------



**Figure 4:** IGH V and J combinations in the MV-HB, MV-LN and NI-LNs

*Circos diagrams displaying linked V and J genes used in the NI-LN2, MV-LN and MV-HB were generated using the Circos Table Viewer v0.63-98. Within each plot, each V and J combination is represented by a colored ribbon. The color code indicates the V gene used while its width is proportional to the frequency of the V-J combination in the nt dataset.*

### 3.4.3 Overlap between sequences encoding identical VH CDR3 amino acid sequences

Next, we assessed the “functional” overlap by comparing samples at the level of VH CDR3, independently of the associated V and J genes. Table 5 describes the numbers of non-redundant CDR3 AA sequences shared (A) and the percentage of reads encoding for these shared CDR3 in the nt sequence set (B). The overlap between the NI-LNs was very high compared to their overlap with MV-LN. The frequencies of the reads encoding a shared CDR3 between NI-LNs reached up to 33.9% of the nt sequence set (NI-LN2 vs NI-LN1). On the other hand, except for one single CDR3 sequence, there was no overlap between NI-LNs and MV-LN. Thus the CDR3 repertoires in the lymph node of unimmunized animals are similar to each other but very different from the immunized animals. This suggests that the majority of the CDR3 detected in the MV-LN are a result of the MV immunization, demonstrating that the MV immunization via footpad inoculation is a very efficient process. Since our analyses focused on IGH of class-switched B cells the high overlapping repertoire between lymph nodes of different naïve animals may be partially a reflection of their common housing environment.

Surprisingly, only a very limited number ( $n=116$ ) of unique CDR3 AA sequences were shared between the MV-HB ( $116/2,051=5.66\%$ ) and MV-LN ( $116/1,744=6.65\%$ ). In other words, the majority of the CDR3 observed in the MV-LN are not found in the MV-HB and *vice versa*. This indicates that B cells coding for low frequency MV-LN CDR3 are preferentially fused, as most of the unique CDR3 in the MV-HB are not represented enough in the MV-LN to be detected by our sequencing approach. However, the cumulative frequencies of this limited number of shared AA CDR3 in the set of nt sequences were high, accounting for 51.6% in MV-HB and 37.6% in the MV-LN. Therefore, these CDR3 are important components of the IGH repertoire of both MV samples. This set of shared CDR3 is actually composed of two sets: the sequences identical to a MV CDR3, increased in MV-HB and those encoding for other CDR3, decreased in MV-HB.

**Table 5:** Overlap of sequence reads encoding identical CDR3 AA sequences

		A. Numbers of shared unique CDR3 AA					
		Sample 2					
		MV-HB	MV-LN	NI-LN1	NI-LN2	NI-LN3	NI-LN4
Sample 1	MV-HB	2,051	116	0	0	0	0
	MV-LN		1,744	1	0	0	0
	NI-LN1			11,032	324	220	160
	NI-LN2				9,990	283	180
	NI-LN3					8,324	168
	NI-LN4						6,471
		B. Percentage of the nt sequence dataset encoding for shared CDR3 AA					
		Sample 2					
		MV-HB	MV-LN	NI-LN1	NI-LN2	NI-LN3	NI-LN4
Sample 1	MV-HB	100%	51.58%	0%	0%	0%	0%
	MV-LN	37.56%	100%	0.01%	0%	0%	0%
	NI-LN1	0%	0.002%	100%	30.20%	26.78%	13.28%
	NI-LN2	0%	0%	33.87%	100%	24.19%	5.01%
	NI-LN3	0%	0%	26.59%	33.54%	100%	17.74%
	NI-LN4	0%	0%	24.76%	31.77%	24.12%	100%



### 3.4.4 MV-associated CDR3 are highly enriched in the MV-HB compared to the MV-LN

We retrieved the MV CDR3 from all datasets. Not a single MV CDR3 was found in any of the NI-LNs. All the MV CDR3 except W were found in the MV-HB but only half (12/23) were also found in the MV-LN (Table 6). Only 9.7% of the nt sequences in the MV-LN dataset encoded a MV CDR3, while 50.8% of sequences in the MV-HB dataset encoded MV CDR3 AA sequences. This corresponds, on average, to a five-fold enrichment of sequences encoding a MV-specific CDR3 in the hybridoma library. Our results further show that the rare MV CDR3 in the LN are highly enriched upon fusion (Table 6). More than 85% (20/23) of the MV CDR3 are derived from these very rare MV-LN CDR3, with a frequency below 0.02%. The maximal enrichment for all CDR3 that we were able to directly calculate was about 796-fold (MV CDR3 N). On the other hand, in the lymph nodes, 98.8% of the MV CDR3 encoding sequences are attributed to only 3 MV CDR3 (U: 55.6%, S: 25.5%, U: 17.7%). Those ranked among the 20 top CDR3 in the MV-LN, while all other MV CDR3 were detected at only very low frequencies (<0.02%). Thus only these 3 MV CDR3 are encoded by B cells that have already expanded in the lymph nodes in response to the immunization, while the majority of the activated cells that undergo fusion appear to be very poorly represented in the *in vivo* immune response.

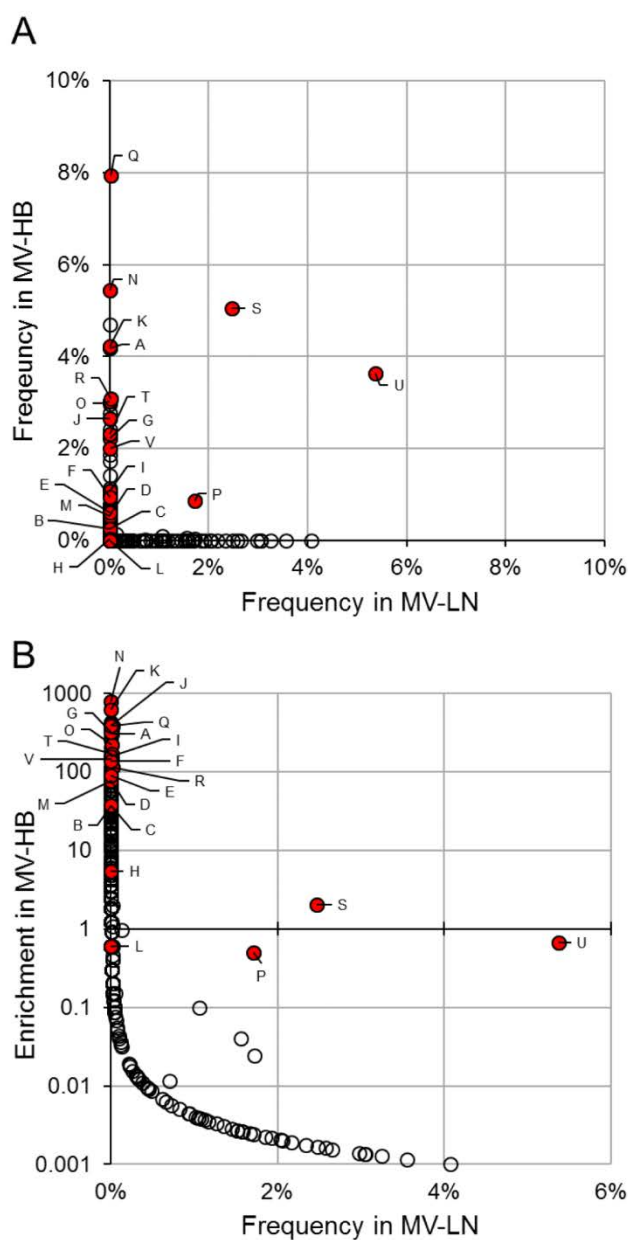
**Table 6:** Frequency of sequences encoding a MV-specific CDR3 AA in MV-LN and MV-HB

ID	Lymph node			Hybridoma library			
	Count (n)	Frequency (%)	Rank	Count (n)	Frequency (%)	Rank	Enrichment <sup>a</sup>
A	2	0.014%	166	1,005	4.180%	6	306
B	0	<0.007%	-	60	0.250%	50	>37
C	0	<0.007%	-	62	0.258%	49	>38
D	0	<0.007%	-	130	0.541%	34	>79
E	0	<0.007%	-	146	0.607%	32	>89
F	0	<0.007%	-	228	0.948%	23	>139
G	1	0.007%	386	531	2.209%	15	323
H	0	<0.007%	-	9	0.037%	103	>5
I	0	<0.007%	-	257	1.069%	21	>157
J	0	<0.007%	-	635	2.641%	12	>387
K	0	<0.007%	-	1,013	4.213%	5	>617
L	0	<0.007%	-	1	0.004%	671	>0.61
M	0	<0.007%	-	125	0.520%	37	>76
N	1	0.007%	386	1,307	5.436%	2	796
O	2	0.014%	166	729	3.032%	9	222
P	251	1.714%	19	205	0.853%	24	0.50
Q	3	0.020%	108	1,906	7.928%	1	387
R	4	0.027%	83	742	3.086%	8	113
S	361	2.466%	11	1,214	5.049%	3	2
T	2	0.014%	166	555	2.308%	14	169
U	787	5.375%	1	871	3.623%	7	0.67
V	2	0.014%	166	479	1.992%	16	146
W	0	-	-	0	-	-	-
	1,416	9.67%		12,210	50.79%		5.25

<sup>a</sup> Whenever a CDR3 is not found in the lymph node, we assume a frequency of <0.007%. Values are rounded to the next digit.

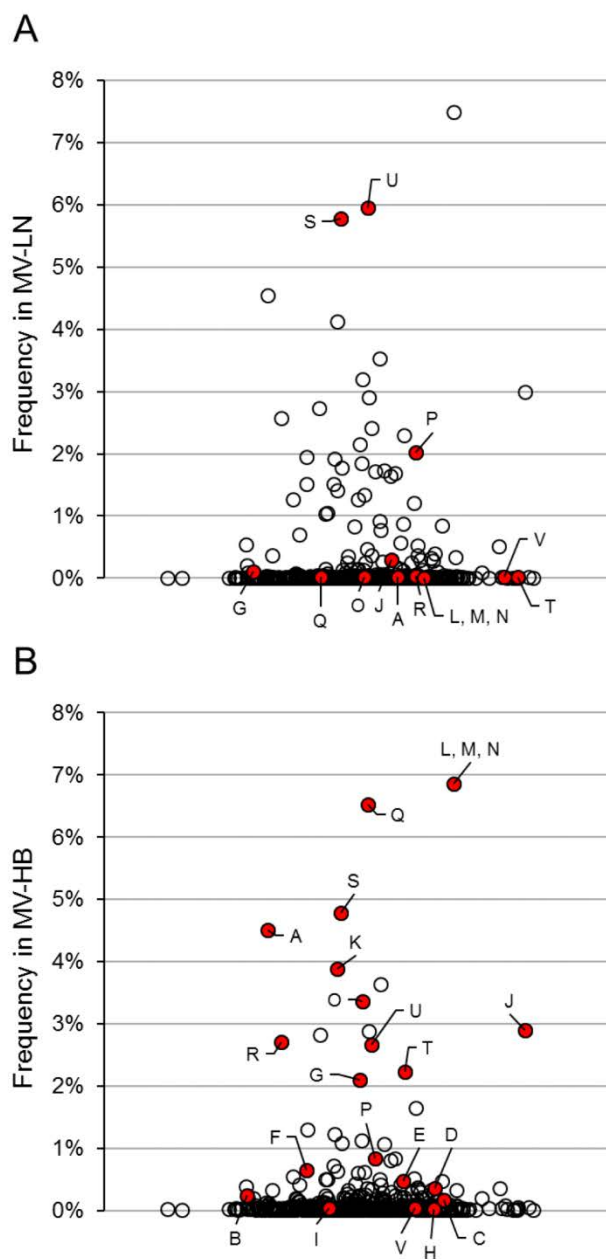
### 3.4.5 The clonally expanded B cell populations in the MV-LN do not fuse efficiently

When the frequencies of all CDR3 AA sequences in the MV-HB and MV-LN are plotted against each other it becomes clear that essentially only very rare MV-LN CDR3 are recovered in the MV-HB (Fig. 5A). The lower the frequencies in the MV-LN the higher their enrichment in the MV-HB (Fig. 5B). With the exception of the MV CDR3 P, S and U, all the CDR3 that were expanded *in vivo* were largely undetected in the MV-HB. These findings indicate that the rare activated B cells in the LN are preferentially fused whereas the cells that already underwent clonal expansion *in vivo* appear to be less prone to fusion. During clonal expansion, affinity maturation includes replacement mutations (amino acid changes) that slightly change the CDR3 AA sequence and therefore the paratope topography, resulting in increased antigen affinity of the antibody. We grouped all the CDR3 nt sequences that were associated with identical V and J genes into clusters of sequences with at least 85% nt identity, corresponding to B cell populations with similar paratopes and originating from common ancestors i.e. clonotypes. Clonotypes were named according to the letter code of the MV CDR3 clones that they contained. Except from the 3 clones P, S and U, the expanded clonotypes in the MV-LN were not related to any of the MV-specific hybridoma clones (Fig.6A). These correspond to 3 clonotypes out of the 15 with a frequency over 2% in the MV-LN. On the other hand, 13 MV-specific hybridoma clones were found within large clusters in the MV-HB (Fig. 6B). In other words, out of the 14 clonotypes with a frequency >2% in the MV-HB, only 3 did not contain sequences of the MV-specific hybridoma clones. We compared the frequency of each clonotype in the lymph nodes and in the hybridoma library (Fig. 7A). Only 72 clusters of 446 in the lymph nodes and of 1,184 in the hybridoma library were shared, and included respectively 47.3% and 50.4% of their IGH nt sequences. More generally, the most frequent clonotypes in the MV-LN were absent in the MV-HB, except for the three clusters containing the sequences of the MV CDR3 P, S, and U. On the other hand, the very rare MV-LN clones were enriched in the MV-HB (Fig. 7B). These results extend our previous observation to low frequency B cell populations with 85% homology to MV-specific CDR3 sequences. This further confirms that antigen-activated B cells which have *not* (yet) undergone clonal expansion are more prone to fusion and therefore over represented in the hybridoma library.



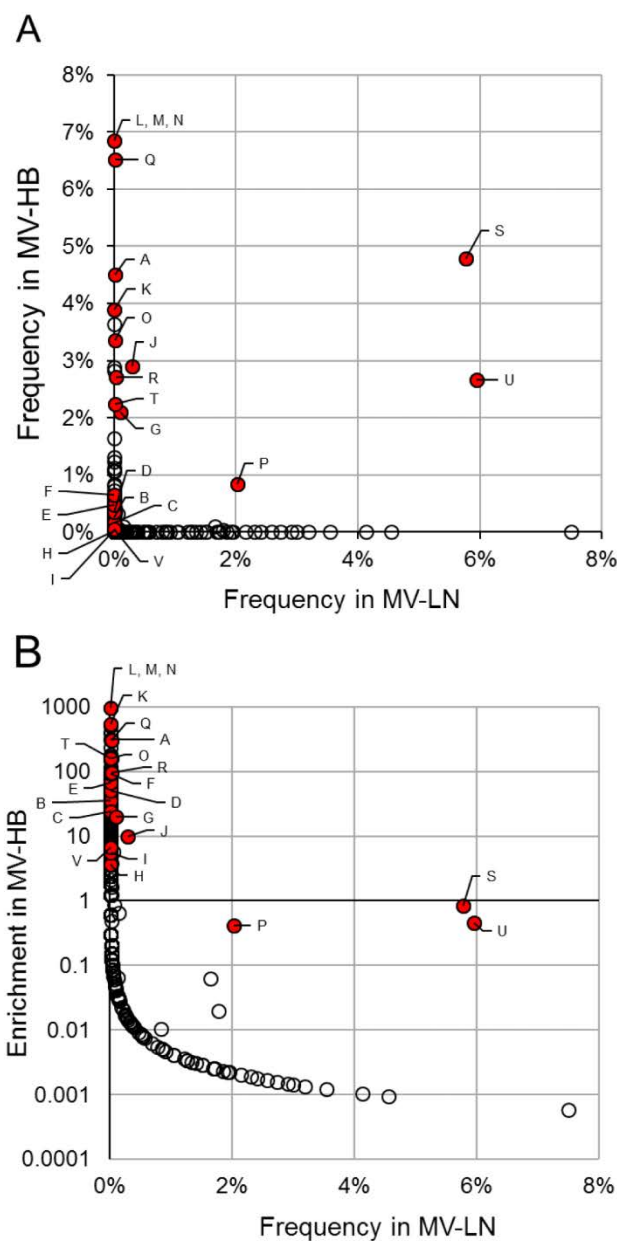
**Figure 5:** Comparison of the frequencies of the nt sequences for each CDR3 AA sequences from the MV-HB and the MV-LN

(A) Relative frequencies are expressed in percent of all CDR3 sequences, based on the non-redundant sequence read counts i.e. on the number of variants nucleotide reads carrying the same CDR3 AA sequence. (B) Same data as (A) expressed in enrichment i.e. ratio between MV-HB/MV-LN. All CDR3 AA sequence counts were assumed to be at least 1. Red dots correspond to the CDR3 of the MV hybridoma clones.



**Figure 6:** Frequency distribution of the clonotypes from the MV-LN (A) and MV-HB (B)

*CDR3 nt sequences associated with the same V and J genes, and with at least 85% sequence identity were grouped into clusters representing B cell populations originating from a common ancestor i.e. clonotypes. The frequency of each clonotype corresponds to the cumulative frequency of the IG nt sequences in cluster. The clusters containing a MV CDR3 or an  $\geq 85\%$  identity-related CDR3 are shown in red and named accordingly.*



**Figure 7:** Comparison of the frequencies of each clonotypes in the MV-HB and MV-LN

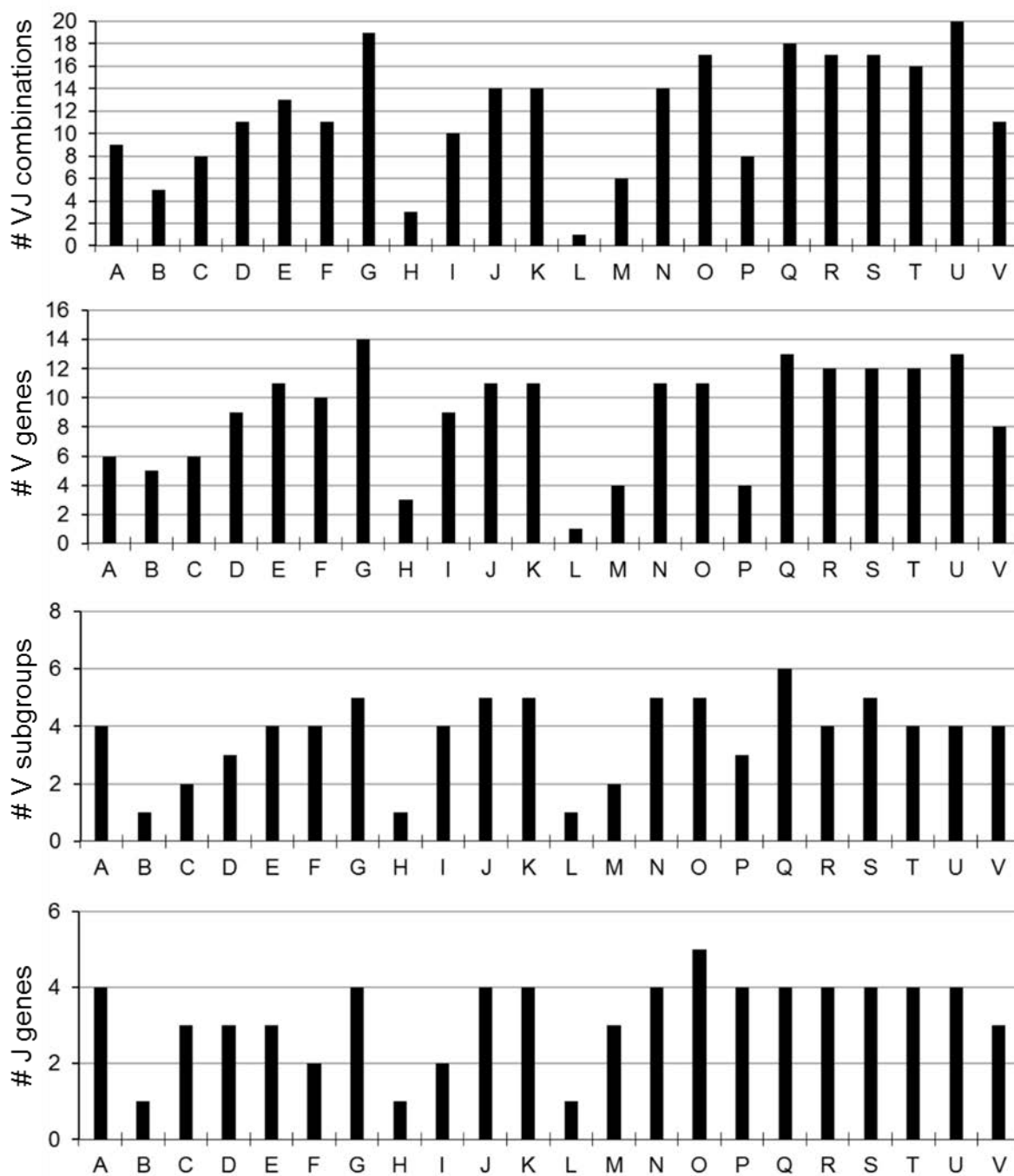
(A) The relative frequency of each clonotype corresponds to the cumulative frequency of the IG nt sequences in the cluster. (B) Same data as (A) expressed in enrichment i.e. ratio between MV-HB/MV-LN. All CDR3 amino acid sequence counts were assumed to be at least 1. Clusters containing a MV CDR3 or an  $\geq 85\%$  identity-related CDR3 are colored in red and named accordingly.

### 3.4.6 MV CDR3 are associated with different V-(D)-J rearrangements

In both the MV-HB and the MV-LN, we observed that CDR3 identical at the AA level were associated with different V and J gene rearrangements. The more frequent the CDR3 were, the more V and J genes encoded for them. Because of the enrichment of rare antigen-specific sequences, it was possible to analyze the V and J gene usage of MV CDR3 encoding sequences at high depth. In the MV-HB, each MV CDR3 was encoded by up to 14 V genes, 6 V subgroups, 5 J genes and 20 V-(D)-J rearrangements (Fig.8). This suggests that different cell populations produce antibodies with similar paratopes, indicating a convergence towards the same antigen specificity, as the VH CDR3 is its major determinant. The V and J gene rearrangements used by the MV-specific hybridoma clones (Table 2) were, in virtually all cases (except for MV CDR3 F and R), the most prominent among the sequences sharing the same MV CDR3 (Fig.9). Other V and J rearrangements were present, but at much lower frequencies in both the MV-HB and MV-LN.

Identical CDR3 were predominantly associated with V genes belonging to the same V subgroups. For instance, 90.98% of the V gene associated with MV CDR3 Q in the MV-HB belonged to the V4-39 (73.87%), V4-34 (13.17%) and V4-31 (3.93%) genes (Fig.10). This suggest that V genes belonging to a certain subgroup (e.g. V4 for the MV CDR3 Q) are favored, possibly because they encode for components that also play a role in the efficient binding to the antigen (CDR1 and CDR2).

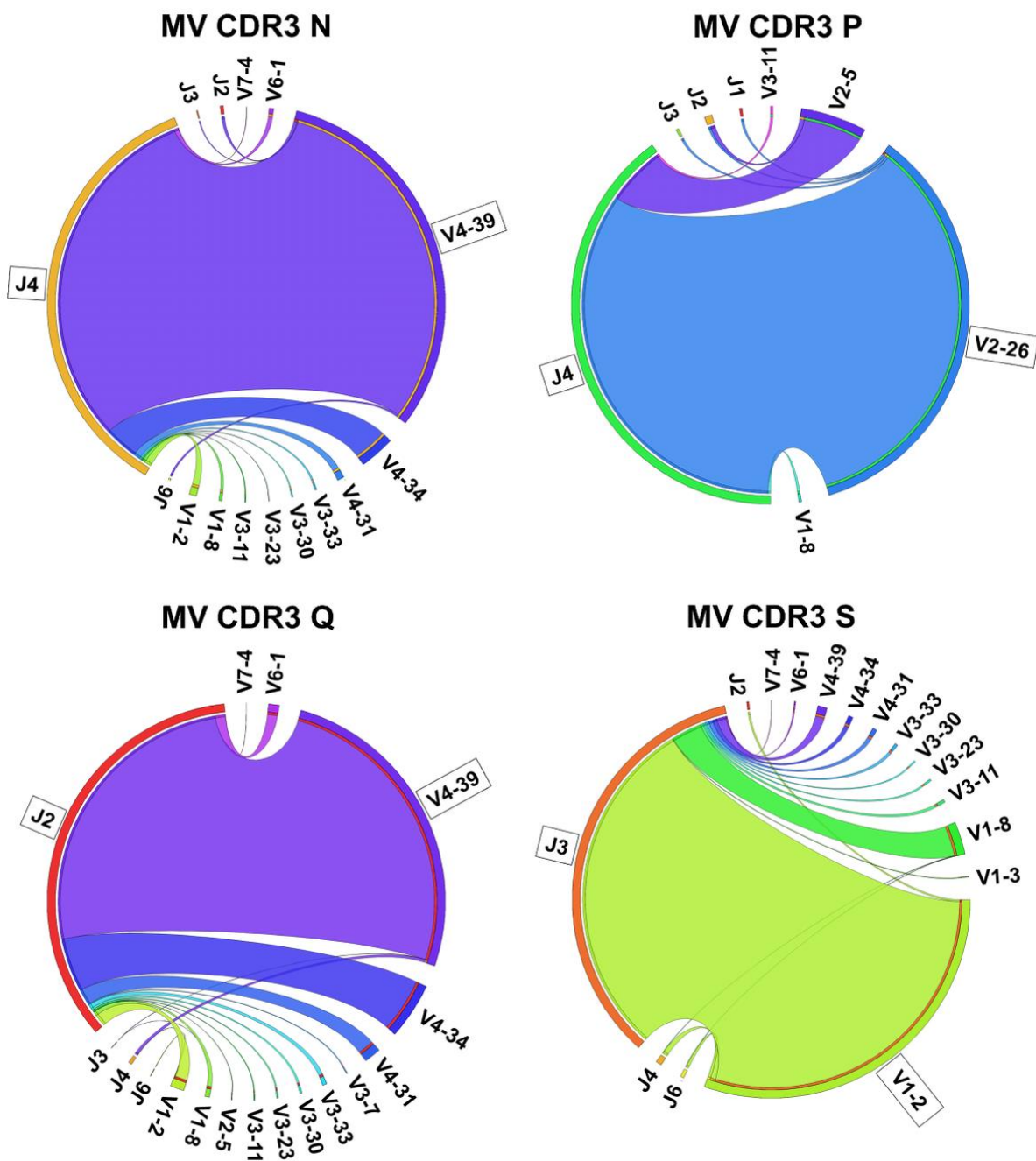
Figure 11 shows the full diversity of the V-(D)-J segments (represented as vertices) and sequences that differ by a single nucleotide are connected. The nt sequences encoding for a MV CDR3 AA sequence have the same color. Clonal expansions of hypermutating B cells appear in these graphs as large clusters of connected vertices. In the MV-LN, MV CDR3 appear mostly as single unconnected vertices (Fig. 11A), corresponding to their low frequency. Only the MV CDR3 P, S and U are involved in large clusters, indicating that they underwent clonal expansions and somatic hypermutation. In contrast, in the MV-HB, most of the MV CDR3 are part of big clusters, indicating that their V-(D)-J segments underwent somatic hypermutation (Fig. 11B). However, in both the MV-LN and the MV-HB, MV CDR3 are encoded by different V and J genes and rearrangements, therefore not all vertices of the same color code are connected. This suggests that different subpopulations of activated B cells with different V-(D)-J rearrangements converged towards a similar antigen-specificity (i.e. a common CDR3). In the NI-LNs, MV CDR3 are totally absent (Fig. 11C).



**Figure 8:** Different V-(D)-J rearrangements, V genes and J genes are associated with identical MV CDR3 in the MV-HB

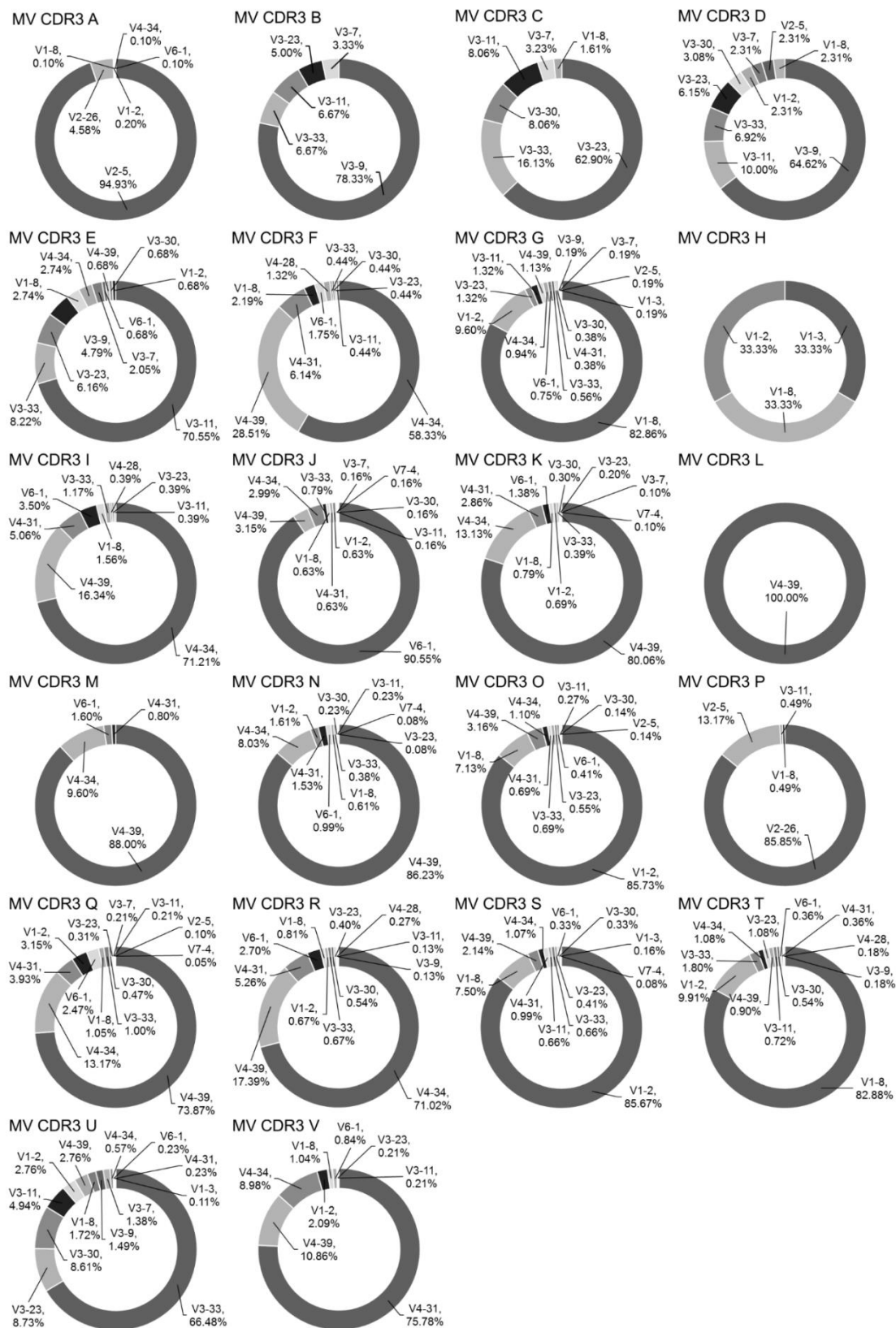
*Number of V-(D)-J rearrangements, V genes, V subgroups and J genes associated with a MV CDR3 in the MV-HB.*



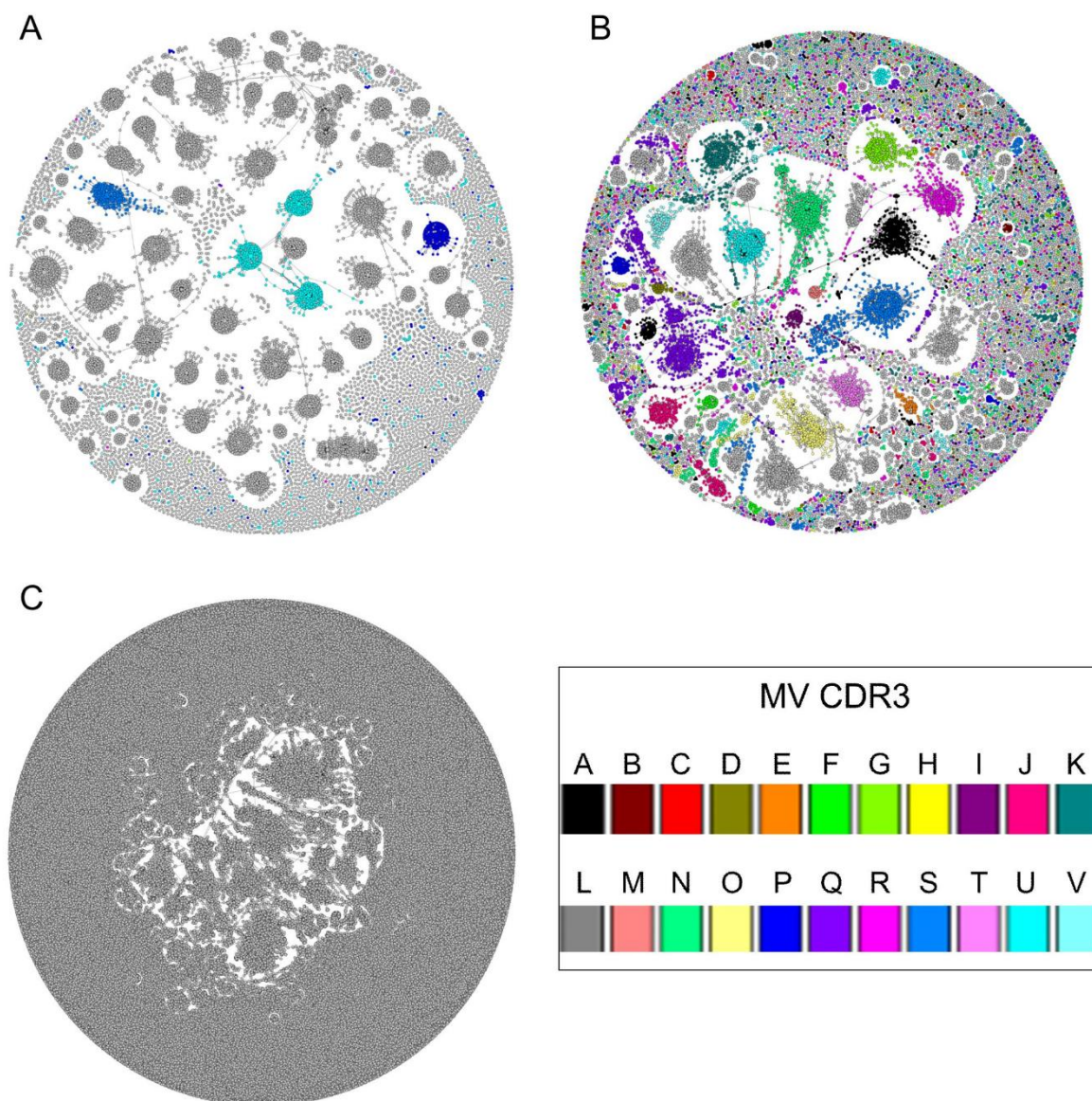


**Figure 9:** Identical MV CDR3 are associated with different V-(D)-J rearrangements

Circos diagrams displaying linked V and J genes associated with the MV CDR3 N ( $n=1,307$ ), P ( $n=205$ ), Q ( $n=1,906$ ) and S ( $n=1,214$ ) in the MV-HB were generated using the Circos Table Viewer. Within each plot, each V and J combination is represented by a colored ribbon. The color code indicates the V gene used while its width is proportional to the frequency of the combination in the nt dataset. The V and J genes used by the MV-specific hybridoma clones are boxed.



**Figure 10:** Frequency (%) of the different V genes associated with each MV CDR3



**Figure 11:** B cell network before and after fusion

*Diversity of IGHV-(D)-J sequence reads in the MV-LN (A), MV-HB (B) and NI-LN3 (C). Sequences differing by a single nucleotide are connected. Each CDR3 AA sequence of the 23 MV clones is associated with a different color.*

## 3.5 Discussion

Over the last decade, an increasing interest has been devoted to the production of human mAbs as new therapeutic agents (339). The hybridoma technology allows the production of antigen-specific antibodies in large quantities. However, little is known about the differences between the antibodies produced during the *in vivo* immune response and those successfully recovered in the hybridoma libraries. We present here the first high-resolution analysis of the IGH repertoire during immunization, and before and after PEG-fusion using high-throughput sequencing. A unique model of rats (OmniRat™, Open Monoclonal Technology Inc.) expressing a large human IGH repertoire was immunized with Measles Virus antigens. We analyzed both the composition of the IGH repertoire in the lymph nodes and in the derived hybridoma library. Our data shows that the fusion process is highly selective for antigen-specific and antigen-stimulated B cells found at very low frequencies in the regional lymph nodes, while most of the clonally expanded lymph node B cells are not present in the hybridoma library. This in-depth analysis of the IGH repertoire also suggests a convergent evolution of multiple different V-(D)-J rearrangements towards identical VH CDR3 amino acid sequences in this humanized rat model.

Recently, deep sequencing technologies have been applied to study the B cell IG repertoire and used to monitor clonal expansions triggered by different antigens (154, 157). While next-generation sequencing provides the technology to investigate the IGH repertoire at great depth, the interpretation of the sequences may be flawed by a PCR amplification bias and clonal expansion after fusion. To overcome this problem we have taken a conservative approach by collapsing the data to non-redundant i.e. unique nt sequence reads. Nevertheless, analyzes performed on both the total (redundant) and non-redundant dataset resulted in similar conclusions (not shown) because of similar technical bias across the different samples. The comparisons of lymph nodes and hybridoma library are also complicated by nt insertions and deletions (indels) that are inherent to Ion Torrent sequencing. Reads with indels in the V and J gene were eliminated by using the indels correction algorithm of IMGT/HighV-QUEST and by analyzing only the productive sequences with unambiguously assigned V and J genes by IgBlast. The efficacy of this approach to remove all reads with indels was confirmed with the IGH sequences of two unrelated hybridoma clones.

Upon footpad immunization with an inactivated Measles Virus, the OmniRat™ mounted a specific immune response involving most V and J genes and a wide diversity of rearrangements of human

V-(D)-J genes that had undergone extensive somatic hypermutation and efficient class-switching to IgG1 and IgG2b. As expected immunization resulted in a dramatic shift in the composition of the IGH repertoire. Compared to the lymph nodes of unimmunized rats, MV-LN displayed a much lower diversity of both full length IGH nt sequences and CDR3 AA sequences, indicating a successful oligoclonal response. While the unimmunized animals utilized similar V and J genes and V-(D)-J rearrangements, MV immunization resulted in a different usage of the V-(D)-J genes mainly due to a strong shift in V gene usage. Finally, no CDR3 AA sequence was shared between the MV-LN (or the MV-HB) and the NI-LNs. In contrast, all unimmunized rats showed a high overlapping usage of identical CDR3, counting for up to 33.9% of the IGH nt sequences. These observations demonstrate that IGH repertoire of the immunized animals differentiate towards a conspicuously antigen-specific repertoire that can be characterized at the level of IGH sequences (152, 340).

If PEG-fusion was a random process, the IGH repertoire of the hybridoma library should be representative of the lymph node repertoire. However, only a limited number of identical CDR3 AA sequences were shared between both samples, suggesting that the hybridoma repertoire only partially reflects the B cell IG repertoire in the lymphoid organs. To investigate at high depth the bias introduced by the fusion process, we retrieved the CDR3 sequences of the MV-specific hybridoma clones (MV CDR3) and calculated the proportion of IGH reads encoding for these CDR3 in both MV-LN and MV-HB. The cumulative frequency of the MV CDR3 in the MV-HB corresponded to ~50% of the IGH nt reads. In the MV-LN, most of the MV CDR3 were found at very low frequency or even undetectable. If they accounted all together for about 10% of its IGH repertoire, this was for the vast majority due to only 3 MV CDR3 (P, S and U) that were clearly expanded in the MV-LN. Excluding these 3 MV CDR3 and assuming that each MV-CDR3 were present at least once in the MV-LN, the average enrichment in MV CDR3 encoding IGH sequences in the MV-HB was at least 355-fold. Considering how many MV CDR3 were not detectable at all in the MV-LN the effective enrichment may be even orders of magnitude higher. In general, the lower the frequency of a CDR3 in the MV-LN, the higher its enrichment in the MV-HB. Conversely, the most frequent MV-LN CDR3 were essentially not detected in the MV-HB.

We further extended our analysis to clonal populations of B cells by grouping the IGH sequences based on V-(D)-J rearrangements and >85% identity within their CDR3 nt sequences (341). Also in this case, essentially only the low-frequency clones in the MV-LN were recovered in the MV-HB, whereas the high-frequency clonotypes were not found. Again only 3 clonotypes that were expanded in the MV-LN had the same V-(D)-J rearrangements and the same CDR3 as MV-specific hybridomas (P, S and U),

and were recovered in the MV-HB. This further confirms that expanded B cells subpopulations are much less susceptible to fusion. Thus the fusion process results in a dramatic enrichment of low abundant clones, while the *in vivo* expanded B cell populations, likely to be “antigen-committed” or antigen-reactive, are largely lost during the process.

In-depth analysis of the IGH sequences revealed that CDR3 sequences can be associated with several V and J genes or subgroups. These observations are indicative of a highly selective pressure towards identical (or similar) CDR3 independently of the associated V-(D)-J rearrangements. Converging recombination of different V and J genes towards the same CDR3 AA sequences have been shown recently to take place in humans, both within and across individuals (194, 202). The complete absence of MV CDR3 in any of the NI-LNs confirms that the high diversity in V and J genes and their rearrangements developed in the MV-LN as a result of immunization. Thus, they clearly represent B cells that were “antigen-committed” irrespective of whether all of them would further expand and become antigen-reactive or not. The high diversity of gene usage could be partially an artifact as high levels of somatic hypermutation could result in an improper assignment of the germline V gene, especially for V genes belonging to the same subgroup that have a high level of nt identity between them. Also, sequencing errors may result in misassignment but substitution errors are rare for the Ion Torrent sequencing and the quality control measures limit their potential impact.

The hybridoma technology is well established, routinely used worldwide to generate mAbs, and is one of the most important advances in clinical immunology of the last century. Recently, deep sequencing has been presented as an alternative technique for the production of mAbs by cloning the most frequent IG sequences into expression vectors, hence avoiding the extensive screening and cloning of hybridoma libraries (210, 211). While the use of traditional technologies often leads to the repeated identification of the same clones, sequence-based identification would avoid this redundancy (210). High-throughput sequencing methods can also be used to recover mAbs that would be lost using traditional techniques, as shown for phage display (209). Here, we showed that the B cell IG repertoire in the hybridoma library does not reflect the parental repertoire of the lymphoid organ. Therefore, the construction of antibodies based on sequences of large expanded clones in the lymph nodes would result in a totally different selection of antibodies (159, 208, 211). MV immunization had a clear impact on the IGH repertoire of the MV-LN (e.g. reduced CDR3 diversity compared to NI-LNs; skewed VJ distribution and low CDR3 overlap with NI-LNs). Consequently, we assume that most of the expanded B cell populations are MV-specific and cloning the most frequent IGH sequences from the MV-LN may



be an alternative way to produce mAbs that are relevant in the *in vivo* immune response. Current knowledge of the IG repertoire and the antibodies involved in infectious diseases or other malignant conditions has largely relied on hybridoma production and Sanger sequencing. Our data suggests that this is most likely influenced by the described fusion bias, thus painting an incomplete picture of the repertoire.

In conclusion, our findings demonstrate that clonally expanded B cells in the lymph node have only a low propensity to undergo fusion. In contrast, rare but “antigen-committed” and antigen-reactive B cell populations are highly susceptible to fusion and are efficiently recovered in hybridoma libraries. While this fusion bias results in the efficient enrichment of antigen-specific B cells by several order of magnitude and is the key to effective generation of mAbs, relevant mAbs with potential applications may be lost during the process. The enrichment in low-frequency antigen-specific variants allowed deep insights into the dynamics of the B cell IG repertoire upon immunization, highlighting the recruitment of a large proportion of the B cells in the lymph node to a convergent evolution towards “antigen-committed” antibodies with similar paratopes.

## 3.6 Acknowledgements

We thank R. Buelow from Open Monoclonal Technology, Inc. (Palo Alto, CA, USA) for providing the OmniRat™, and for critical discussion and comments on the manuscript. The generation of the hybridoma library and cloning were performed by J. Kenney at Antibody Solutions Inc. (Sunnyvale, CA, USA). We also thank M. Bruggemann and M. Osborn from Recombinant Antibody Technology Ltd. (Babraham Research Campus, Cambridge, UK) for sequencing the MV-specific hybridoma clones and their comments on the manuscript. We thank A. K. Wienecke-Baldacchino for her contribution to the bioinformatics analyses.

## **Chapter 4.**

# **Convergent antibody signatures for the MV H and F proteins in transgenic rats expressing a human B cell IG repertoire**

*Manuscript in preparation*



## 4.1 Abstract

B cells are able to respond specifically to an almost infinite number of foreign antigens thanks to a very large repertoire of immunoglobulins (IG). The IG repertoire is continuously modelled by antigen challenges, and previous antigen-exposures or diseases can impact the repertoire composition for long periods of time. Next-generation sequencing (NGS) offers now new possibilities to investigate the IG repertoire at unprecedented depth. Decrypting the signatures leave by specific pathogens using NGS would open the doors to an approach that could be used to retrospectively reconstruct the immune history of an individual. Here, we investigate whether immunization results in the accumulation of antigen-specific IG signatures is a model of transgenic rats carrying human germline heavy and light chain antibody loci (OmniRat™). We aimed to ascertain fully human antigen-specific signatures without preselecting for antigen-specific B cells. Thirty-two transgenic humanized rats were immunized with different antigens: tetanus toxoid (TT), benzo[*a*]pyrene coupled to TT (BaP-TT), a wild type modified vaccinia Ankara virus (MVA) vector and a recombinant MVA expressing the haemagglutinin (H) and fusion (F) glycoproteins of the measles virus (MVA-MV H/F). Using the Ion Torrent Deep Sequencing platform, the immunoglobulin heavy chain (IGH) repertoire in the spleen and bone marrow was analyzed. We demonstrated a strong public immune response within the different groups of rats characterized by convergent VH CDR3 amino acid sequences with a high level of similarity expanded only in the animals that received the same vaccination. Thus 18 VH CDR3 signatures for the measles virus H or F glycoproteins were identified from the spleen and 37 from the bone marrow samples. Some signatures were identical or highly similar to the VH CDR3 of MV-specific hybridoma clones. These fully human VH CDR3 signatures can be used as template in human studies.

## 4.2 Introduction

The ability of the humoral immune system to respond specifically to an almost infinite number of antigens relies on a large repertoire of immunoglobulin (IG) with different specificities. The diversity of the IG repertoire is generated during B cell ontogeny by the stepwise and uneven junction of multiple V (variable), D (diversity; heavy chain only) and J (joining) genes that encode for the variable region of the IG heavy (IGH) and light chain proteins.(28) Nucleotides are deleted and randomly inserted at the joining sites, resulting in an enormous diversity of unique antigen receptors. In the presence of antigen and T cell help, the antigen-specific naive B cells are activated to divide and produce many identical progeny. During the expansion in secondary lymphoid organs that follows antigen recognition, the IG loci undergo extensive DNA somatic mutations, mainly in the CDR3 region (Complement Determining Region 3). The CDR3 of the variable domain of the IG heavy chain (VH), located at the junction of V-(D)-J genes, encodes for the core of the antibody binding-site and is considered to be the major determinant of antigen-specificity (25). The process of somatic hypermutation gives rise to mutant IG molecules on the surface of B cells and cells expressing mutant IG receptors with an increased antigen affinity are preferentially selected to mature into antibody-secreting cells. Furthermore, those with the greatest affinity will eventually differentiate into long-lived memory B cells, leading to an enhanced response to this antigen in subsequent encounters. The sum of the IG with different antigen-specificity in an individual is called the IG repertoire.

The size of the IG repertoire makes its complexity and dynamics difficult to globally assess using conventional sequence analysis. Next-generation sequencing (NGS) technology allows the investigation of complex sequencing targets at unprecedented depth and is a useful tool for understanding IG repertoire dynamics in response to immune challenges (191, 192), natural infections (202), or in lymphoid malignancies (168). However, to investigate the development of antibodies specific to a given antigen starting from NGS datasets, it is of fundamental importance to distinguish and extract the relevant IG sequences from the complete repertoire. This deconvolution of the B cell response and discrimination of antigen-specific sequences has been notably achieved by searching for IG sequences homologous to those previously described and for which the antigen-specificity is known (152, 197). An alternative *a priori* approach consists in identifying the IG sequences shared between multiple individuals following a recent exposure to a common antigen.

It is still largely uncertain if the IG immune response tends to converge towards identical IG sequences in different individuals exposed to the same antigen challenge. Longitudinal studies in which multiple samples were taken from the same donors during the course of the immune response have recently shown that a public IG response may exist in humans. Convergent VH CDR3 amino acid sequences were thus identified in patients recovering from acute dengue infection and in vaccinees against the pandemic influenza H1N1 and tetanus toxoid-conjugated bacterial polysaccharides (192, 194, 202) The extent of the antigen-specific public IG repertoire (i.e. what is shared between individuals) in comparison to the antigen-specific private IG repertoire (i.e. what is unique to an individual) is important to assess further. Public IG sequences can indeed be valuable tools for tracking infections, tracing back immunological histories of individuals, or potentially for being used as measure of vaccine immunogenicity.

To date there exist several rodent models that express complete or almost complete IGH and light chain repertoires (222), and fully human antibodies from these transgenic animals account for an increasing number of new therapeutics (219, 221). These animals also offer the opportunity to investigate the human antigen-specific IG repertoire in a flexible way. Here we used this unique model of transgenic rats (OmniRat™) and developed a method to identify human antigen-specific VH CDR3 signatures from their IG repertoire without sorting for antigen-specific B cells. We used the Ion Torrent PGM platform to analyze the IgG IGH repertoire in the spleen and bone marrow of rats immunized with various antigens, the Measles Virus (MV) haemagglutinin (H) and fusion (F) glycoproteins being used as a reference antigen. We demonstrated a strong public IG response within the different groups of rats characterized by convergent highly similar VH CDR3 amino acid sequences in the animals that received the same vaccination. Thus several fully human VH CDR3 signatures for the MV H or F proteins were identified. We validated our approach by retrieving CDR3 sequences from hybridoma clones with known MV-specificity in our dataset of selected signatures. These fully human VH CDR3 signatures constitute interesting templates for the screening of human IGH datasets generated by NGS.

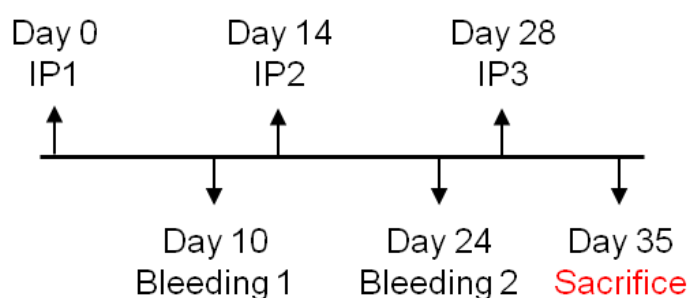
## 4.3 Materials and methods

### 4.3.1 Animals and immunizations

Humanized IG transgenic rats (OmniRat™, Open Monoclonal Technology Inc., Pablo Alto, USA) were developed and bred as recently described (235, 331-333). These animals carry a chimeric human/rat IGH locus, where 22 human IGHV, all IGHD and all IGHJ in germline configuration are linked to the rat IGHC genes, together with fully human IG light chain loci. The endogenous rat IG loci have been silenced using zinc finger nucleases. They express a diversified repertoire of chimeric antibodies with human idiotypes at levels similar to those of wild type animals and have been used to generate high-affinity chimeric mAbs (220).

Thirty-two animals (17 males, 15 females; average age =  $17.8 \pm 1.4$  weeks old; Table 1) were maintained on a 12h light/dark cycle with *ad libitum* access to food and water. All animal procedures were in compliance with the rules described in the Guide for the Care and Use of Laboratory Animals (334).

Animals were separated into 6 groups of 4 to 6 individuals (Table 1). Animals received 3 intraperitoneal injections (IP) at 2-weeks intervals and were sacrificed 7 days after the last injection (Fig. 1). Each injection contained 100ug of tetanus toxoid (TT group; Serum Institute of India) or of a benzo[*a*]pyrene-TT construct (BaP-TT group) (342) with 330ug of aluminum hydroxide (Alhydrogel 2%, Brenntag, Frederikssund, DK). Other rats were injected with  $10^7$  p.f.u. of a recombinant MVA expressing the H and F glycoproteins of the MV (MVA-MV H/F group) or the MVA viral vector only (MVA group) without adjuvant. The control animals received either 330ug of aluminum hydroxide (ALUM group) or were left untouched (NEG group). Antigen-specific IgG responses were monitored by ELISA 10 days after the 2 first immunizations and at sacrifice.



**Figure 1:** Immunization scheme

**Table 1:** OmniRats involved in the study

Treatment	Rat ID	Gender	Age in weeks at start
1 <sup>st</sup> cohort			
MVA-MV H/F N=6	1	F	19
	2	F	19
	3	F	19
	4	M	19
	5	M	19
	6	M	19
MVA N=6	7	F	22
	8	F	19
	9	F	15
	10	M	15
	11	M	19
	12	M	19
2 <sup>nd</sup> cohort			
BaP-TT N=5	13	M	17
	14	M	16
	15	M	17
	16	F	18
	17	F	18
TT N=4	18	M	16
	19	M	17
	20	F	16
	21	F	18
ALUM N=6	22	M	17
	23	M	16
	24	M	17
	25	F	17
	26	F	17
	27	F	18
3 <sup>rd</sup> cohort			
NEG N=5	28	F	18
	29	F	18
	30	M	18
	31	M	18
	32	M	18

### 4.3.2 Generation of the Benzo[a]Pyrene constructs

B[a]P was coupled to ovalbumin (OVA, Sigma-Aldrich, Bornem, Belgium) and purified TT (Serum Institute of India) as described (342). This hapten-conjugate was shown to be highly immunogenic in mice and to induce strong antibody responses against B[a]P (342-344).

### 4.3.3 Modified Vaccinia Ankara viruses propagation and purification

The modified vaccinia Ankara (MVA) and the recombinant MVA carrying MV H and F proteins of the Edmonston strain (MV vaccine strain, clade A; MVA-MV H/F) viruses (kind gift from Dr B. Moss, NIAID, National Institutes of Health, Bethesda, USA) were propagated on BHK-21 cells (ATTC® CCL-10™) as described (345-347). Briefly, T175 cell culture flasks (Greiner) at 90% confluence were infected at a MOI of 0.05 (~1.1 10<sup>6</sup> PFU per flask) with sonicated virus. After incubation for 2-3 days at 37°C, the infected cells were harvested, suspended in 10 mM Tris-HCl pH 9.0, lysed by repeated freezing and thawing cycles, sonicated and homogenized with a dounce homogenizer (5x5 times on ice). The crude stock preparations were then purified by ultracentrifugation through a 36% sucrose cushion at 30,000g for 60 minutes at 4°C. Purified virus material was resuspended in 1 mM Tris-HCl pH 9.0 and the viral titer determined by immunostaining. Serial dilutions of virus stock (10<sup>-4</sup> to 10<sup>-9</sup>) were used to infect nearly confluent BHK-21 cells in 6-well culture plates. 24 hours after infection, the cells were fixed with acetone/methanol (1:1 mixture) and stained using a polyclonal vaccinia virus antiserum (Polyclonal rabbit anti-vaccinia antibody, IgG fraction; AbD Serotec, Hamburg, DE), a secondary peroxidase-conjugated antibody (horseradish peroxidase-conjugated polyclonal goat anti-rabbit antibody, IgG H+L; Dianova, Hamburg, DE) and appropriate substrate (O-Dianisidine, Sigma-Aldrich). The generation of the recombinant MVA-MV H/F is described elsewhere (348, 349).

Although MVA viruses are unable to replicate in most mammalian cells, replication of viral DNA is not impaired in these cells and viral proteins are synthesized (350). Replication-deficient MVAs were safely used during the smallpox eradication campaign and have been shown to express recombinant genes at levels similar to replication-competent strains (351). MVAs induce B and T cell responses against various recombinant proteins in animals following intramuscular or intranasal immunizations (352, 353). More particularly, cynomolgus macaques vaccinated with the recombinant MVA-MV H/F used in this study developed MV neutralizing antibodies and MV specific T cell response (349).

#### 4.3.4 Anti-MV, -MVA, -TT and -B[a]P IgG ELISA

Antigen-specific IgG antibody levels in sera were determined in 384-well microtiter plates (Greiner bio-one, Wemmel, BE) coated overnight at 4°C with either 250 ng of MV antigen (Measles grade 2 antigen, Microbix Biosystems, Mississauga, CA), 2.5E+05 PFU of sonicated MVA (~314 ng), 187.5 ng of TT or 0.25  $\mu$ M of B[a]P-OVA in carbonate buffer (100 mM, pH9.6). Free binding sites were saturated with 1% bovine serum albumin (BSA) in Tris-buffered saline at room temperature for 2h. Serial dilution of the sera were added for 90 min at 37°C, and developed with alkaline phosphatase-conjugated goat anti-rat IgG (1/750 dilution, ImTec Diagnostics, Antwerp, BE) and the appropriate substrate at 405 nm. Endpoint titers (EPT) were determined as the serum dilutions corresponding to 5 times the background.

#### 4.3.5 Sample preparation, amplification and Ion Torrent Deep Sequencing

Lymphocytes were isolated from spleen and bone marrow by gradient centrifugation (Ficoll Paque PLUS; GE Healthcare Life Sciences, Bornem, BE), total RNA extracted (RNeasy midi kit, Qiagen, Venlo, NL) and mRNA enriched by paramagnetic separation ( $\mu$ MACS mRNA Isolation kit, Miltenyi Biotech, Leiden, NL). cDNA was prepared using dT18 primers and Superscript III RT (Invitrogen, Merelbeke, BE) at 50°C for 80 min. Recombined IGH fragments were subsequently amplified by PCR for 35 cycles using primers designed to cover all the IGHV genes carried by the OmniRat, a rat IGHC primer (Table 2) and Q5 Hot Start High Fidelity polymerase (New England Biolabs, Frankfurt-Main, DE). Amplicons were size selected on a 2% agarose gel (QIAQuick Gel Extraction Kit, Qiagen) and quantified (Quant-iT PicoGreen dsDNA, Invitrogen). Quality was checked with a Bioanalyzer (High Sensitivity DNA, Agilent Technologies, Diegem, BE). Libraries were pooled in equimolar concentrations and sequenced on a 318™ Chip v2 using multiple identifiers (MIDs) with the Ion OneTouch™ Template OT2 400 Kit and the Ion PGM Sequencing 400 Kit (Life Technologies, Gent, BE) on the Ion Torrent Ion Personal Genome Machine (PGM™) System (Life Technologies). Four randomly-selected samples were run per chip. Precautions were taken to avoid sample cross-contaminations, with each sample being processed individually and each step followed by thorough decontamination of working surfaces and instruments (DNA Remover, Minerva Biolabs, Berlin, DE). A maximum of 8 randomly-selected samples were processed in one day.

**Table 2:** Primer list

Name	Primer sequence
VH1-FR2	5'-ctggacaagggcttgagtgg-3'
VH2-FR2	5'-caggaaagggccttgagtgg-3'
VH3-FR2	5'-caggggaaggggcttgagtgg-3'
VH4-FR2	5'-caggggaagggacttgagtgg-3'
VH6-FR2	5'-catcgagagggccttgagtgg-3'
Rat IgG CH1	5'-ggatagacagatggggctgttgg-3'

## 4.3.6 Data processing

### 4.3.6.1 Quality control and sequence annotation

Only reads where a MID could be unambiguously assigned (no mismatches allowed), primers identified at both ends (2 mismatches allowed) and from which more than 85% of the bases had a quality score  $>25$  were considered for further analysis. The MIDs and primer sequences were trimmed, and sequences were subsequently collapsed to unique nucleotide sequences. Sequences were next submitted to the IMGT/HighV-QUEST webserver v1.4.0beta (IMGT, [www.imgt.org](http://www.imgt.org); MP. Lefranc, Montpellier, France) allowing for indels correction to annotate the different V gene regions and to delineate the CDR3 (143). Variable (V), Diverse (D), and Joining (J) genes for the in-frame productive sequences were subsequently assigned using a local installation of IgBlast (336), including only the genes present in the genome of the OmniRat<sup>TM</sup> as a reference set. Only sequences with an unambiguously assigned V and J gene (Blast E-value  $<10^{-3}$ ) were considered for further analysis. Outputs from IMGT/HighV-QUEST and IgBlast were cross-compared, and reads still containing indels were discarded.

### 4.3.6.2 Setting of the CDR3 identity threshold to assess public immune responses

All CDR3 amino acid sequences were extracted from IMGT datasets and sorted in descending order of frequency in a separate file for each rat. The 200 most frequent CDR3 (top 200) of each rat were then compared to all CDR3 from all 32 rats. Using a modified version of `get_close_matches` (Python `difflib` library), the identity of each CDR3 pair was calculated as Levenshtein distance. This allowed us to find close matches of CDR3 based on a specified threshold of similarity (e.g. 80%, 85%, 90%, 95% and 100%) and return these as a ratio from 0 to 1. As a result, a matrix was generated displaying the percentage of the top CDR3 that are present, or have a close match to all CDR3 present in another individual. The matrices were then analyzed in R and heat maps were generated using `ph heatmap` with hierarchical



clustering enabled. The heat maps were analyzed visually for vaccination based grouping of rats and the optimal similarity threshold was selected (84%).

#### 4.3.6.3 Identification of VH CDR3 signatures

Only VH CDR3 longer than 4 AA were included in the analysis. Based on the threshold selected in 4.3.6.2., all CDR3 that shared a minimum of 84% amino acid homology (Table 3) and that differed in length by a maximum of 1 AA were considered as relatives. To reduce the number of cross-comparisons, only CDR3 with an 84%-relative in at least 2 rats of the same immunization groups were included (preselection). For each CDR3 the cumulative frequency (based on full length nt sequence frequency) of all its 84%-relatives per rat was calculated. A CDR3 was considered "antigen-driven" if its relative cumulative frequency was significantly increased in a group of animals immunized with this particular antigen in comparison to the other vaccination groups (ANOVA 1 on log-transformed frequencies and Tukey-Kramer method;  $p < 0.05$ ). Over-represented CDR3 sequences were grouped again based on 84%-homology and sequence logos were generated using WebLogo 2.8.2. (<http://weblogo.berkeley.edu>) (354). All analytical scripts were written in Python 2.7 and R. Circos diagrams were generated using the Circos Table Viewer v0.63-9 (<http://circos.ca>) (355). Vertex cluster networks were produced based on gapped nt-sequences or AA-sequences and hamming distances of 1 or 2, with a script written in R, similarly to Bashford-Rogers et al.(337)

**Table 3:** Number of mismatches allowed by CDR3 length

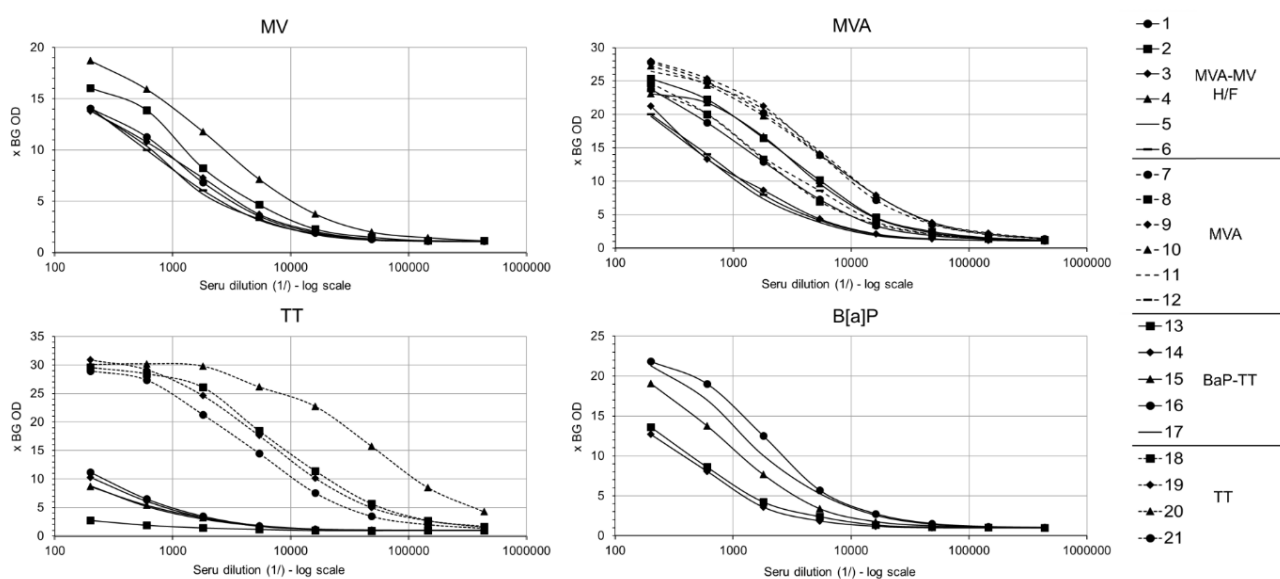
CDR3 length (AA)	Mismatches allowed	AA identity level
4-6	0	100%
7-12	1	85.7-91.7%
13-18	2	84.6-88.9%
19-24	3	84.2-87.5%
>24	4	Min. 84.0%

## 4.4 Results

### 4.4.1 Antigen-specific antibody response and next-generation sequencing of the IGH repertoire of immunized OmniRat

In order to identify MV-specific sequences from the total IGH repertoire without sorting for antigen-specific cells, thirty-two OmniRats were vaccinated with four different antigens and separated into 3 cohorts (Table 1). The shared IGH sequences within these different vaccination groups allow to discriminate the antigen-specific sequences as overlap/shared IGH between different vaccination groups. In the first cohort, 6 rats were immunized with a recombinant modified vaccinia Ankara (MVA) virus carrying the measles virus haemagglutinin and fusion proteins (MVA-MV H/F group) and 6 rats with the naked MVA vector (MVA group). The MVA group is a control for the MVA-MV H/F group and allows the identification of MV H and F specific sequences, the MV H/F glycoproteins being our antigen of interest. Combining the MVA-immunized animals with the rats challenged with the recombinant MVA allows to identify sequences for immunodominant epitopes for the MVA backbone. A second cohort of rats was immunized with either tetanus toxoid (n=4, TT group), a B[a]P-TT conjugate vaccine (n=5, BaP-TT group) (342) or the adjuvant only (n=6, ALUM group). The rats in the third cohort were left untouched as controls (n=5, NEG group). Seven days after the last injection, rats were sacrificed and antigen-specific IgG antibody levels measured in sera by ELISA. A specific immune response was found in all the immunized animals, while the mock immunized (ALUM group) and control (NEG group) animals showed no detectable antigen-specific antibodies (Fig. 2 & Table 4).

Rearranged IGH genes (IgG) were amplified from mRNA extracted from lymphocytes of the spleen (SP) and the bone marrow (BM), and sequenced on the Ion Torrent PGM platform. A total of 37,473,982 reads for the spleen samples (range: 850,298-1,879,372 per animal) and 50,279,455 reads for the bone marrow samples (range: 1,042,474-2,392,897 per animal) were obtained (Table 5 & Table 6). After quality control and annotation, only the productive non-redundant sequences with an unambiguously assigned V and J gene were retained for analysis. On average,  $86,619 \pm 24,441$  IGH sequences for SP and  $205,506 \pm 54,596$  IGH sequences for BM were analyzed for each individual. The rats expressed a diverse IGH repertoire, including varying frequencies of all V and J genes, except for IGHV3-22-2 which was not found in the spleen samples (Fig.3). All possible V-(D)-J rearrangements were found in the SP and BM of animals in all the groups (not shown). There was no obvious bias in V, J genes or V-(D)-J rearrangements usage after immunization with any of the antigens (Figure 3).



**Figure 2:** ELISAs

Antigen-specific IgG in the sera were measured by indirect ELISA. The serum dilution-1 (1:100 to 1:656,100) was plotted against binding measured by adsorption at 405 nm and expressed as multiples of the average optical density of the background ( $x$  BG OD). A specific IgG immune response was observed in all the immunized animals. Anti-MV IgG were found in all the 6 rats from the MVA-MV H/F group (average EPT= 1:4,116) but at lower levels than anti-MVA IgG (average EPT= 1:6,635). Anti-MVA IgG levels were higher in the rats that received the MVA vector only (MVA group; average EPT= 1:22,687). Anti-TT IgG levels were high in the sera of TT-immunized animal (average EPT= 1:61,670; range: 386,350 to 36,354). These levels were however low for the animals in the BaP-TT group, probably because their antibodies are directed against a denatured form of TT (average EPT= 1:530). Hapten-specific IgG were detected in all rats immunized with the conjugate vaccine (average EPT= 1:2,691), as we showed previously in mice (342-344).

**Table 4:** Reciprocal endpoint titers (1:) at sacrifice

Treatment <sup>a</sup>	Rat ID	Antigen					
		BaP	TT	MV	MVA		
MVA-MV H/F N=6	1			3,811	11,645		
	2			5,059	15,375		
	3	<1:200	<1:200	4,109	4,862		
	4			12,217	15,079		
	5			2,839	4,214		
	6			3,122	4,596		
	Average			<1:200	<1:200	4,116	6,835
MVA N=6	7				35,601		
	8				11,584		
	9	<1:200	<1:200	<1:200	39,275		
	10				38,438		
	11				39,478		
	12				13,653		
	Average				<1:200	<1:200	<1:200
BaP-TT N=5	13	1,587	<1:200				
	14	1,422	1,059				
	15	4,049	815	<1:200	<1:200		
	16	8,021	1,193				
	17	6,543	701				
	Average	2,691	530 <sup>b</sup>			<1:200	<1:200
TT N=4	18		70,333				
	19	<1:200	48,667	<1:200	<1:200		
	20		386,350				
	21		36,354				
	Average		<1:200			61,670	<1:200

<sup>a</sup> no reactivity over background was found in all the control rats (negative and alum groups)

<sup>b</sup> For the serum with no detectable specific antibody (Rat #13), EPT was set to 1:200 (highest serum concentration tested)

**Table 5:** Ion Torrent read counts before and after quality controls – Spleen

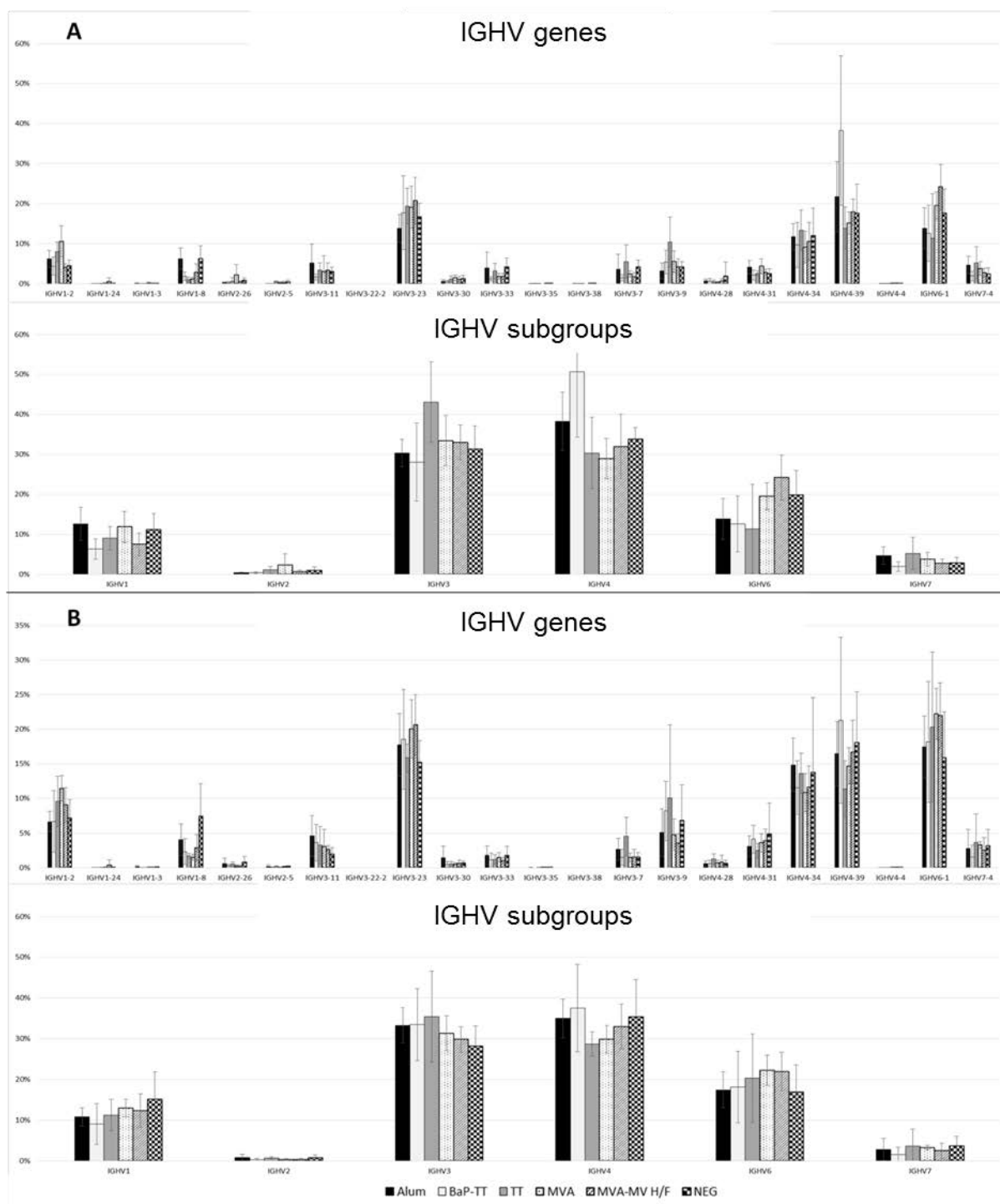
		1	2	3	4
	ID#	Output sequence reads	Filtered nt sequences total dataset	Filtered nt sequences Non-redundant dataset	Unique CDR3 AA
MVA-MV H/F N=6	1	1,365,752	377,046	75,632	4,007
	2	1,072,907	375,003	87,632	5,680
	3	1,573,425	451,893	97,070	5,050
	4	1,579,703	433,549	105,168	5,609
	5	1,110,541	202,493	53,913	3,536
	6	1,073,485	309,199	68,542	4,469
	Total	7,775,813	2,149,183	487,957	28,351
	Average $\pm$ SD	1,295,969 $\pm$ 243,309	358,197 $\pm$ 91,366	81,326 $\pm$ 18,992	4,725 $\pm$ 871
MVA N=6	7	929,621	219,719	75,643	4,333
	8	995,002	398,167	91,300	5,084
	9	937,392	269,346	65,840	3,288
	10	1,084,088	309,205	129,977	7,232
	11	998,174	305,988	105,000	4,980
	12	1,198,061	475,341	129,219	5,691
	Total	6,142,338	1,977,766	596,979	30,608
	Average $\pm$ SD	1,023,723 $\pm$ 101,807	329,628 $\pm$ 92,325	99,497 $\pm$ 26,876	5,101 $\pm$ 1,325
BaP-TT N=5	13	1,177,010	328,269	94,218	3,191
	14	990,929	361,511	91,500	5,838
	15	989,221	308,410	80,886	5,538
	16	1,126,831	379,904	89,527	4,399
	17	1,646,696	303,863	50,234	2,379
	Total	5,930,687	1,681,957	406,365	21,345
	Average $\pm$ SD	1,186,137 $\pm$ 270,461	336,391 $\pm$ 33,284	81,273 $\pm$ 1,8055	4,269 $\pm$ 1,485
TT N=4	18	1,184,407	403,988	105,527	3,912
	19	1,119,688	325,011	65,630	4,228
	20	1,036,147	304,799	54,763	4,471
	21	1,380,809	469,132	137,371	6,756
	Total	4,721,051	1,502,930	363,291	19,367
	Average $\pm$ SD	1,180,263 $\pm$ 146,827	375,733 $\pm$ 75,556	90,823 $\pm$ 37,938	4,842 $\pm$ 1,297
Alum N=6	22	1,431,185	426,569	65,363	3,586
	23	1,879,372	607,120	85,192	7,202
	24	1,004,133	276,970	78,375	8,013
	25	896,017	354,473	68,058	6,602
	26	879,714	218,371	46,409	3,946
	27	850,298	379,796	86,623	4,365
	Total	6,940,719	2,263,299	430,020	33,714
	Average $\pm$ SD	1,156,787 $\pm$ 414,592	377,217 $\pm$ 134,979	71,670 $\pm$ 15,115	5,619 $\pm$ 1,882
NEG N=5	28	929,883	380,288	60,509	5,619
	29	1,121,113	294,331	100,287	5,885
	30	938,306	401,837	106,170	6,422
	31	1,380,088	453,771	135,281	6,809
	32	1,593,984	415,810	84,948	4,026
	Total	5,963,374	1,946,037	487,195	28,761
	Average $\pm$ SD	1,192,675 $\pm$ 289,420	389,207 $\pm$ 59,407	97,439 $\pm$ 27,551	5,752 $\pm$ 1,070
	TOTAL N=32	37,473,982	11,521,172	2,771,807	162,146
	AVERAGE $\pm$ SD	1,171,062 $\pm$ 262,032	360,037 $\pm$ 85,479	86,619 $\pm$ 24,441	5,067 $\pm$ 1,356

SD: standard deviation

**Table 6:** Ion Torrent read counts before and after quality controls – Bone Marrow.

		1	2	3	4
	ID#	Output sequence reads	Filtered nt sequences total dataset	Filtered nt sequences Non-redundant dataset	Unique CDR3 AA
MVA-MV H/F N=6	1	2,076,987	750,479	231,768	8,681
	2	1,481,070	445,901	170,578	6,293
	3	1,741,791	769,381	240,546	6,998
	4	1,586,929	481,206	201,353	9,972
	5	1,624,366	1,393,889	222,345	7,726
	6	1,363,857	640,985	273,434	9,499
Total		9,875,000	4,481,841	1,340,024	49,169
Average $\pm$ SD		1,645,833 $\pm$ 247,261	746,974 $\pm$ 343,955	223,337 $\pm$ 35,056	8,195 $\pm$ 1,440
MVA N=6	7	1,414,364	1,197,246	271,704	12,548
	8	1,483,660	488,394	146,135	5,581
	9	1,459,257	733,717	286,023	11,929
	10	1,190,683	500,964	289,061	16,903
	11	1,836,660	571,923	268,596	11,350
	12	1,508,353	460,629	223,288	10,623
Total		8,892,977	3,952,873	1,484,807	68,934
Average $\pm$ SD		1,482,163 $\pm$ 208,057	658,812 $\pm$ 281,534	247,468 $\pm$ 54,961	11,489 $\pm$ 3,641
BaP-TT N=5	13	1,350,107	662,337	212,322	6,697
	14	1,230,306	365,306	150,224	7,761
	15	1,668,762	1,450,300	237,258	7,626
	16	1,355,296	320,423	94,597	4,533
	17	2,392,897	838,162	187,141	4,704
Total		7,997,368	3,636,528	881,542	31,321
Average $\pm$ SD		1,599,474 $\pm$ 472,338	727,306 $\pm$ 457,118	176,308 $\pm$ 55,871	6,264 $\pm$ 1,558
TT N=4	18	1,611,933	754,979	195,858	6,024
	19	1,180,402	422,103	183,613	9,331
	20	2,004,175	1,098,947	246,754	8,276
	21	1,182,325	364,112	132,746	7,581
Total		5,978,835	2,640,141	758,971	31,212
Average $\pm$ SD		1,494,709 $\pm$ 395,672	660,035 $\pm$ 339,529	189,743 $\pm$ 46,811	7,803 $\pm$ 1,387
Alum N=6	22	1,871,915	717,181	275,722	9,583
	23	1,563,310	1,355,530	175,924	6,950
	24	1,470,656	681,554	236,822	10,897
	25	2,071,339	628,644	126,829	4,401
	26	1,791,049	668,390	174,429	6,215
	27	1,113,360	338,503	129,640	5,003
Total		9,881,629	4,389,802	1,119,366	43,049
Average $\pm$ SD		1,646,938 $\pm$ 338,902	731,634 $\pm$ 334,999	186,561 $\pm$ 59,235	7,175 $\pm$ 2,570
NEG N=5	28	1,676,194	490,493	122,135	4,316
	29	1,669,493	662,836	289,893	8,777
	30	1,716,450	554,220	196,485	7,695
	31	1,042,474	508,654	199,276	10,268
	32	1,549,035	580,259	183,694	6,310
Total		7,653,646	2,796,462	991,483	37,366
Average $\pm$ SD		1,530,729 $\pm$ 280,016	559,292 $\pm$ 67,978	198,297 $\pm$ 60,017	7,473 $\pm$ 2,285
TOTAL N=32		50,279,455	21,897,647	6,576,193	261,051
AVERAGE $\pm$ SD		1,571,233 $\pm$ 307,073	684,301 $\pm$ 304,787	205,506 $\pm$ 54,596	8,158 $\pm$ 2,778

SD: standard deviation



**Figure 3:** V gene usage (%) in the spleen (A) and bone marrow (B)

Standard deviations (SD) for the animals in the same group combined are shown. Variations within the animals in the same groups were marked for some V genes, resulting in large SD. Preferential usage of a certain V or J gene did not appear as a criterion to differentiate animals from different immunization groups.

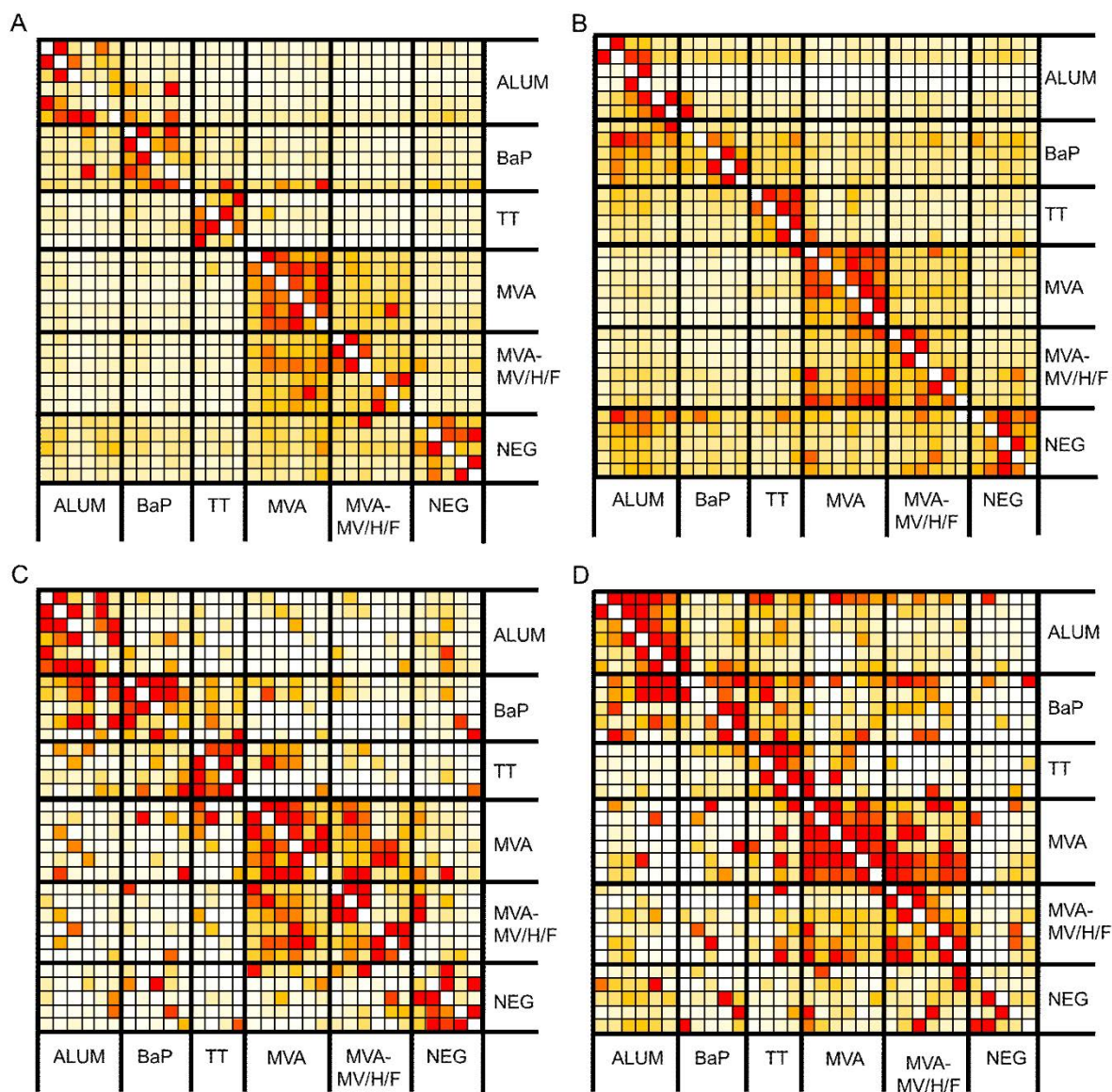
#### 4.4.2 Transgenic rats use sets of identical CDR3 sequences in response to the same antigenic challenges

We first investigated whether rats in the same group mount a convergent antibody response by assessing their overlap in terms of VH CDR3 amino acid sequences (subsequently referred to as CDR3). All unique CDR3 of each rat were cross-compared to the unique CDR3 of all other rats, SP and BM being considered separately. Figure 4 shows the overlap between rats in terms of shared CDR3 in BM (A) and SP (B), as well as their frequencies in the non-redundant dataset of nucleotide sequences (BM: C; SP: D). The overlap was higher, in most of the cases, between rats in the same immunization group compared to animals from other groups. The rats of the MVA and the MVA-MV H/F groups also showed a much higher overlap between them compared to the other groups (Fig. 4). Thus, the VH CDR3 repertoires in the bone marrow and the spleen of animals that received the same immunization are more similar to each other than to unimmunized rats or animals immunized with other antigens. This indicates that these rats are using similar CDR3 in response to the same immunization, i.e. are mounting convergent responses.

If a CDR3 is present in a group of rat but absent from all the others, it is likely that this sequence has been induced by the immunization. Although many VH CDR3 sequences are shared between at least two individuals, very rarely a specific CDR3 was present in all animals that received the same vaccination (Table 7). However, CDR3 shared by rats in the same group were highly similar, differing by essentially only one or two amino acid, both in SP and BM. Table 8 displays a list of selected CDR3 shared between rats in the MVA-MV H/F group, the MVA group and in both groups combined, homologous CDR3 being colored similarly. Groups of related CDR3 were found to be shared between animals in the MVA-MV H/F group only, suggesting that these belong to antibodies directed against the MV H or F glycoproteins (e.g. CDR3 related to ARIVGATTEFDY, ARHRTYYYYGSGSPLFDY and ASGSSGWYSSFDY). Other groups of CDR3 were found in the MVA immunized animals or in the two group combined (e.g. TRDRTGDWYFDL, AIRYNWNDGFD), strongly indicating that these are triggered by MVA antigens. Thus, immunization does not only trigger the expression of strictly identical CDR3 sequences, but clusters of highly similar CDR3 that represent complex “antigen-specific” signatures and a certain number of mismatches in the CDR3 sequence needs to be allowed in order to cover all the animals in a vaccination group only (Fig.5). The CDR3 region being known has the major antigen-binding determinant of the antibody molecule (25), we assumed that an overlapping CDR3



usage is indicative of a functional overlap i.e. similar antigen-specificity of the antibodies. These public antigen-driven CDR3 may therefore constitute antigen-specific signatures (192, 194, 202).



**Figure 4:** Overlap of sequence reads encoding identical CDR3 AA sequences

*When comparing the AA CDR3 sequences from all the rats in all the groups, it appeared clear that the animals have more unique CDR3 in common with animals from the same vaccination group than with animals from other groups (A: BM; B: SP). The same observation was made when taking the copy numbers (i.e. full-length non-redundant nt IGH sequences per CDR3 AA) into consideration (C: BM; D: SP). Each line and column corresponds to one animal. Color code ranges from light yellow (minimum number of CDR3 shared) to dark red (maximum number of CDR3 shared) on a per line basis.*

**Table 7:** Unique CDR3 AA sequences shared between rats in the same immunization group only

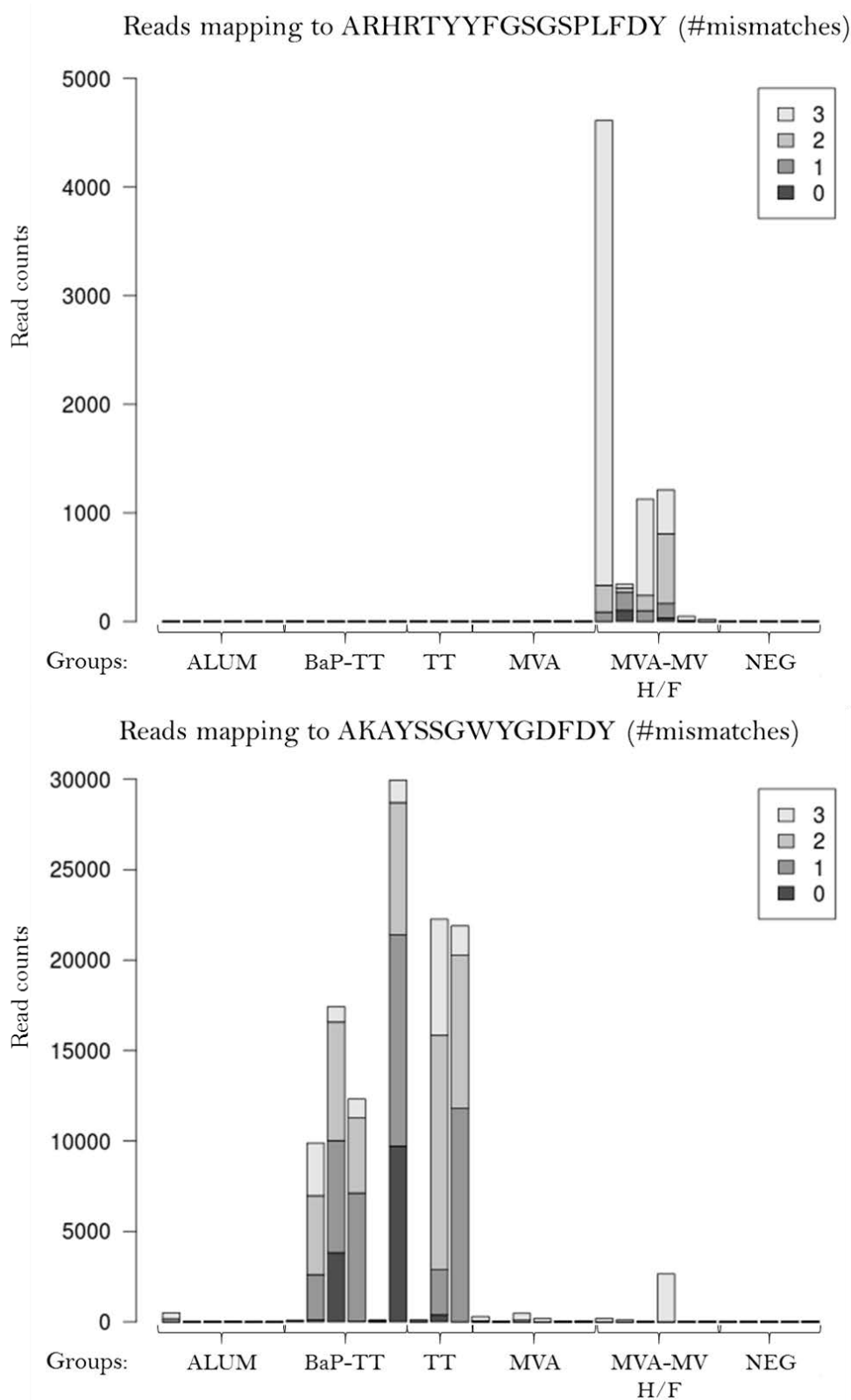
SPLEEN													
GROUP	Not shared	# CDR3 shared by x animals in the group											
		≥2	2	3	4	5	6	7	8	9	10	11	12
ALUM	26,549	2,122	2,008	100	12	2	0						
BaP-TT	19,246	197	182	15	0	0							
TT	17,044	337	307	23	7								
MVA	26,130	776	660	86	23	7	0						
MVA-MV H/F	22,883	382	335	40	7	0	0						
NEG	26,229	225	212	13	0	0							
BaP-TT + TT	/	167	123	29	11	2	2	0	0	0			
MVA + MVA-MV H/F	/	715	413	136	66	38	17	18	13	9	4	1	0

BONE MARROW													
GROUP	Not shared	# CDR3 shared by x animals in the group											
		≥2	2	3	4	5	6	7	8	9	10	11	12
ALUM	37,632	1,103	917	150	29	7	0						
BaP-TT	27,167	596	510	74	12	0							
TT	27,317	1,078	988	75	15								
MVA	61,139	1,332	1,123	168	40	1	0						
MVA-MV H/F	43,778	742	686	51	4	0	1						
NEG	33,675	495	455	37	3	0							
BaP-TT + TT	/	396	258	101	24	11	0	1	1	0			
MVA + MVA-MV H/F	/	1,516	924	316	102	72	35	29	15	8	8	4	3

**Table 8:** List of selected CDR3 shared by rats in the MVA-MV H/F (/6 rats), the MVA (/6) and both groups combined (/12)

SPLEEN					
MVA-MV H/F		MVA		MVA and MVA-MV H/F combined	
ARIVGATTEFDY	4/6	TRDRTGDWYFDL	5/6	AKGDYGDPPYFD	11/12
ARAVGATTEFDY	3/6	ARDRTGNWYFDL	5/6	AKGDYGDPPYFDH	10/12
ARIVGATTEFD	3/6	ARDRTGYWYFDL	4/6	VKGDYGDPPYFDY	10/12
ARAVGATTEFD	3/6	ARDRTGDWYFDF	4/6	TKGDYGDPPYFDY	10/12
ARVVGATTEFDY	3/6	ARDRTGDWYFDF	3/6	AKGNYGDPPYFDY	9/12
ARSVGATTEFDY	2/6	SRDRTGDWYFDL	3/6	AKGDYGDPPYLDY	9/12
VRAVGATTEFDY	2/6	ARDRTGGWYFDL	3/6	AKGDYGDPPYFDL	9/12
ARIVGATTDFDY	2/6	ARGRTGDWYFDL	3/6	AKGGYGDPPYFDY	9/12
ARHRTYYYGSGSPLFDY	4/6	ARDWTGDWYFDL	3/6	ARGDYGDPPYFDY	9/12
ARHRTFYFGSGSPLFDY	3/6	ARDRTGDWCFDL	3/6	AKGDYSDPPYFDY	8/12
ARHRTYYYGSGSPLFD	3/6	ARNRTGDWYFDL	3/6	AKGDYGDPCYFDY	8/12
ARHRTFYFGSGSPLFDY	2/6	ARDRTGDWYFDI	3/6	AKGDYGDPPYFEY	8/12
ARHRTFYFGSGSPLFDY	2/6	ARDLTGNWYFDL	3/6	AKGDYGDPPYFNY	8/12
ARHRTFYFGSGTPPFDY	2/6	TRDLTGYWYFDL	3/6	AKGDYGDPPYFDF	8/12
ARHRTYYYGSGSPRFDY	2/6	ARDLTGSWYFDL	3/6	AKGDYGDPPYFDF	8/12
ASGSSGWYSSFDY	3/6	ALRYNWNDDGLDY	4/6	AKGDYDDPPYFDY	8/12
ARGTSGWYSSFDY	3/6	ALRYNWNDDGLDY	3/6	AKGDYGDPPYFDS	8/12
ARGSSGWYSSFDY	2/6	AIRYNWNDDGLDY	2/6	AIRYNWNDDGFD	10/12
ARGSSAWYSSFDY	2/6	ALRYNWNDDGFDY	2/6	AIRYNWNDDGFDY	9/12
ARGSGDWYSSFDY	2/6	ALRYNWNNGFDY	2/6	PLRYNWNDDGFDY	8/12
VRGSSGWYSSFDS	2/6	ALRYNWNDDGFDY	2/6	ALRYNWNDDGFDY	8/12
ALRYNWNNEGFY	4/6	ALRYNWNDDGFDY	2/6	VIRYNWNDDGFDY	7/12
PVRYNWNDDGFDY	3/6	ALRYNWNDDGFDY	2/6	ATRYNWNDDGFDY	6/12
AVRYNWNDDCFDY	2/6	AREGVAVALYYYYGMDV	5/6	AIRYNWNDDGFD	6/12
BONE MARROW					
MVA-MV H/F		MVA		MVA and MVA-MV H/F combined	
ARIVGATTEFDY	6/6	TRDLTGDWYFDL	4/6	TKGDYGDPPYFDY	12/12
ARIVGATTDFDY	4/6	ARDKTGDWYFDL	3/6	AKGDYGDPPYFDL	11/12
ARVVGATTEFDY	4/6	ARDETDWYFDL	3/6	AKGDYGDLYYFDY	11/12
ARAVGATTEFDY	3/6	ARDLTGDWYFDF	3/6	AKGDYGDPPYFDS	11/12
ARIVGASTDFDY	2/6	ARDLTGSWYFDL	3/6	AKGDYGDPPYFDF	10/12
ARVVGATTDFDY	2/6	ARDLTGHWYFDL	3/6	ARGDYGDPPYFDY	10/12
ARIVGASTEFDY	2/6	ARDITGNWYFDL	3/6	AKGDYGDPPYLDY	10/12
ARAVGASTEFDY	2/6	ARDVTGYWYFDL	3/6	AKGDYGDPPYFDH	10/12
ARIVGASTEFDY	2/6	TRDKTGDWYFDL	2/6	AKGDYGDPPYFDF	10/12
ARIVGASNEFDY	2/6	TRDLTGYWYFDL	2/6	AKGGYGDPPYFDY	9/12
ARIVGDSTEFDY	2/6	VRDLTGDWYFDL	2/6	AKGDYGDTPYFDY	9/12
ARGVGATTEFDY	2/6	ARDWTGDWYFDL	2/6	VKGDYGDPPYFDY	9/12
ARHRTYYYGSGSPLFDY	4/6	ARDLTGDWYFDI	2/6	AKGYYGDPPYFDY	9/12
ARHRTFYFGSGSPLFDY	4/6	GRDLTGYWYFDL	2/6	AKGDYGDPPYFGY	9/12
ARHRTYYYGSGSPPFDY	3/6	ARDVTGDWYFDL	2/6	AKVDYGDPPYFDY	8/12
ARHRTYYYGSGSPPFDY	2/6	ARDTTGDWYFDL	2/6	AKGDYGDPPYFDC	8/12
ARHRTHYFGSGSPLFDY	2/6	GRDRTGDWYFDL	2/6	AKGDYGDPPYSDY	8/12
ARHRTFYFGSGSPLFDY	2/6	TRDKTGYWYFDL	2/6	AKGDYGDPPYFDI	8/12
ARHRTYYYGSGSPPRFDY	2/6	TRDRTGNWYFDL	2/6	AKGDYGDPPHYFDY	8/12
ARHRTFYFGSGSPPRFDY	2/6	TRDRTGYWYFDL	2/6	AKGDYCDPPYFDY	8/12
ARHRTHYFGSGSPPFDY	2/6	ERDLTGYWYFDL	2/6	AKGDYGDPPYCFDY	8/12
ARHRTYYYGSGSPPFDY	2/6	ERDRTGDWYFDL	2/6	AKGDYGYPPYFDY	8/12
ARHRTFYFGSGSPPRFDY	2/6	TRDRTGDWYFDF	2/6	SKGDYGDPPYFDY	8/12
ARHRTFYFGSGSPPRFDY	2/6	ARDLIGDWYFDL	2/6	AKGDYGDPPYFDY	7/12
ARHRTYYYGSGSPPFDY	2/6	ARDITGDWYFDL	2/6	AKGDYGDPPYFEY	7/12
ARHRTYYYGSGSPPRFDY	2/6	ARDLTGDWYLDL	2/6	AKGDYGDSYFDY	7/12
ARGSSGWYSSFDF	3/6	AIRYNWNDDYFDY	3/6	AKGDYGEPPYFDY	7/12
ARGSSGWYSSFDY	3/6	ALRYNWNDDGFDV	2/6	AEGDYGDPPYFDY	7/12
ARGSSGWYSSFDY	3/6	AMRYNWNDDGFD	2/6	AKGDYGGPPYFDY	7/12
ARGSSGWYSSFDY	3/6	AIRYNWNDDGFNY	2/6	AKGDYDDPPYFDY	7/12
ARGRSAWYSSFDC	2/6	ALRYNWNDDYFDY	2/6	AKGEYGDPPYFDY	7/12
ARGRSGWYSSFDY	2/6	AREGVAVALYYYYGMDV	5/6	AKGDCGDPYFDY	7/12
ARGTSGWYSSFDY	2/6			AKGDYGDPPYFDY	7/12
ARGSSGWYSSFDC	2/6			AIRYNWNDDGFDY	12/12
SRGSSDWYSSFDY	2/6			ALRYNWNDDGFDY	12/12
ARGSSGWYSSFDQ	2/6			AIRYNWNDDGFDY	10/12
ARGRSGWYSSYDY	2/6			ALRYNWNDDGFDY	9/12
SRGSSGWYSSFDY	2/6			ANRYNWNDDGFDY	8/12
ARVSSGWYSSFDY	2/6			TIRYNWNDDGFDY	7/12
ARGSSDWYSSFDC	2/6			VLRYNWNDDGFDY	6/12
ARGSSAWYSSFDC	2/6			ALRYNWNDDGLDY	6/12
ARGSSGWYSSFDS	2/6			AIRYNWNDDGFDY	6/12
ARGRSDWYSSFDY	2/6			VIRYNWNDDGFDY	6/12
ASGSSGWYSSFDY	2/6				
ARGSSGWYSSLDY	2/6				
AVRYNWNNEGFY	3/6				
AVRYNWNDDCFDY	2/6				
EVRYNWNDDGFDY	2/6				
VKRYNWNDDGFDY	2/6				



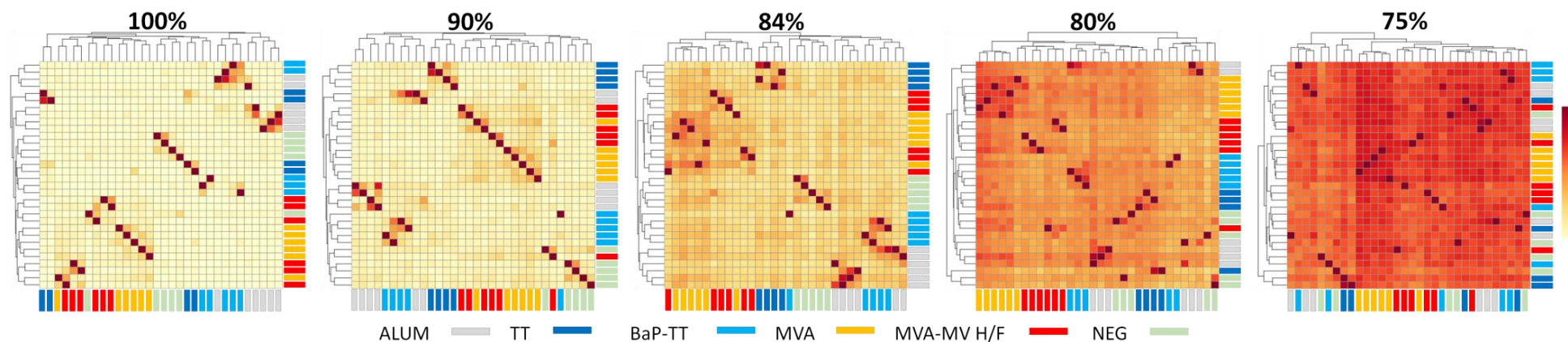
**Figure 5:** Immunized OmniRats express closely related VH CDR3 amino acid sequences

*Selected CDR3 were screen over the bone marrow datasets allowing 0, 1, 2 or 3 amino acid mismatches. Identical VH CDR3 are rarely present in all the animals in an immunization group. However, CDR3 differing by only few amino acids can be found in all or almost all of them.*

### 4.4.3 Highly similar CDR3 sequences are produced in response to the same antigenic challenge

In order to identify antigen-specific signatures, it is necessary to determine a threshold of CDR3 sequence identity that enable to maximize the differences between the different vaccination groups, and allow to cluster together the IGH sequences produced upon the same immune challenge. The most interesting threshold is the one for which inter-group distances (i.e. between animals from different immunization groups) are highest while intra-group distances (i.e. between rats in the same immunization group) are lowest. The top 200 CDR3 of each animal were compared with all CDR3 from all the rats using a Levenshtein distance algorithm. Heat maps showing the percentage of homologous CDR3 shared with any of the other animals were created using different homology thresholds from 100% (identity) to 50% homology. (Fig. 6). Heat maps and hierarchical clustering for 100%, 90%, 84%, 80% and 75% sequence identity levels are shown in Figure 5 for the BM samples. Between 100% and 90% identity thresholds, animals of the same group tended to cluster together but the intra-group distances were still large (Fig. 6; right panels). When lowering the identity threshold, intra-group distances decreased less rapidly than the inter-group distances until approximately 80% identity. Below the 80% identity threshold, inter-group distances increased dramatically and clustering of animals by vaccination group was lost (Fig. 6; left panels). The threshold with the highest ratio between inter-group distances and intra-group distances was found at 84% sequence homology. At this level of identity, the rats immunized with the same antigen cluster together almost perfectly (Fig. 6). It indicates that at this level of identity most of the overlap in CDR3 usage is between the rats of the same group, and therefore it should be used to identify CDR3 that are antigen-driven.





**Figure 6:** OmniRats express an overlapping repertoire of highly similar CDR3

*We tested whether the overlap in CDR3 usage between the rats in the same vaccination group can be optimized by varying the identity cut-off in the CDR3 amino acid sequence homology. The 200 most frequent CDR3 amino acid sequences were for each rat cross-compared to all CDR3 from all other rats. The heat maps display the overlap between rats in terms of percentages of the top 200 CDR3 that are present, or have a close match to all CDR3 of another individual (100%: dark red; 0%: light yellow). At 100% identity, low numbers of CDR3 were shared between the rats. While the shared CDR3 were principally between animals in the same immunization group (high inter-group distances), the variations in the overlap between the animals in the same group were pronounced (high intra-group distances). When lowering the identity cut off, the intra-group decreased less quickly than the inter-group distances until approximately 80% identity. Below this level, intra- and inter- groups distances are both too low to differentiate between the different group of rats i.e. the overlap between the rats became too high to distinct antigen-specific sequences. The ratio between intra- and inter-group distances became optimal at 84% identity cut off. At this level, rats in the same group shared more CDR3 between them than with animals from other groups. Together these observations suggest that rats immunized with the same antigen mount a public immune response characterized by the expression of highly similar CDR3 amino acid sequences and that 84% identify cut off represents the optimal threshold to select for antigen-specific CDR3 signatures.*

#### 4.4.4 Identification of putative antigen-specific human CDR3 signatures

We have shown that transgenic humanized rats mount a public immune response characterized by the expression of CDR3 amino acid sequences with at least 84% identity. We next tested whether human antigen-specific VH CDR3 signatures can be identified from the public repertoire without preselecting for antigen-specific B cells. We developed a method based on sequence cross-comparisons that selects for highly similar CDR3 with levels of expression significantly higher in the animals of a particular vaccination group only.

The CDR3 that shared a minimum of 84% amino acid homology and that differed in length by a maximum of 1 AA were considered “relatives”. The SP and BM datasets were analyzed separately. Sequence cross-comparisons were first performed within groups to select for the CDR3 with a relative in at least 2 rats of the same group (preselection). A total of 25,129 unique CDR3 in the bone marrow and of 34,385 in the spleen were preselected in the MVA-MV H/F group. Only these sets of unique CDR3 were matched against all unique CDR3 from all the 32 rats (SP: 162,146; BM: 261,051) to limit the total number of sequence cross-comparisons.

The results of the sequence cross-comparisons are presented as a frequency matrix. In the matrix, the frequency of each unique CDR3 consists in the summed frequencies of all its relative sequences (cumulative frequency) in the repertoire for each individual separately. According to our definition, a CDR3 is considered as antigen-specific if the average cumulated frequencies of its  $\geq 84\%$  amino acid sequence homologues (relatives) are significantly increased in the group of animals immunized with this particular antigen in comparison to the other groups. A first round of screening was performed to identify the unique CDR3 with significantly different levels of expression between the different groups of rats (ANOVA 1 on log-transformed frequencies plus one;  $p < 0.05$ ). As much as 1,722 CDR3 in BM and 1,999 CDR3 in SP had significantly different expression levels in the MVA-MV H/F group, which represent about 6% of the preselected CDR3 for this group. This round was followed by a post-hoc multiple comparison test to locate the intergroup differences (Tukey-Kramer method;  $p < 0.05$ ). Table 3 shows the numbers of over-expressed CDR3 that meet our definition of antigen-specificity identified in the different groups of rats. For the group of MVA-MV H/F immunized animals, 145 CDR3 in BM and 110 CDR3 in SP met our definition.

Next, for each immunization group, these CDR3 were again compared between each other and grouped into clusters based on 84% homology. This enabled us to identify various numbers of footprints, i.e. CDR3 signatures, in the different group of rats (Table 9). The largest number of VH CDR3 signatures were found in the groups of rats immunized with TT (SP: n = 291; BM: n = 392), followed by the group of rats immunized with the MVA vector (SP: n = 98; BM: n = 137). The large number of TT signatures can be explained by the smaller number of rats in the group. We also identified an overall of 18 putative CDR3 signatures for the MV surface glycoproteins H/F in the SP and 37 in the BM, as well as 9 CDR3 signatures for B[ $\alpha$ ]P in the SP and 14 in the BM. Surprisingly, as much as 78 IG signatures were identified in the group of unimmunized individuals. These CDR3 are likely to be specific for common environmental antigens. The VH CDR3 signatures and sequence logos of the MV H or F glycoproteins are listed in Table 10 and 11. Although SP and BM were processed and analyzed as independent libraries, similar signatures were identified in both tissues, but with different ranking. For instance, the most prominent CDR3 signature identified from the SPs and BMs of the animals immunized with the MVA-MV H/F were identical (ARHRTYYYYGSGSPLFDY) (Tables 10 & 11). This was also the case for the principal TT signatures (e.g. AKAYTSGWYVDF, ARGGSYYHYYYGMD) (Table 12). This indicates that some of the expanded B cell clones in the SP have differentiated into plasma cells and migrated to the BM.



**Table 9:** Highly similar CDR3 amino acid sequences are found with increased frequencies in the different groups of rats and can be grouped into antigen-specific signatures

SPLEEN					
GROUPS	Antigen	Over- represented CDR3 (n)	Clustering		Putative signatures (n)
			Singlets* (n)	Clusters (n)	
BaP-TT	B[a]P	11	8	1	9
MVA-MV H/F	MV H and F	110	11	7	18
MVA	MVA	702	46	52	98
TT	TT	1879	167	124	291
ALUM	Aluminum	149	42	25	67
NEG	Background	358	42	36	78
BONE MARROW					
GROUPS	Antigen	Over- represented CDR3 (n)	Clustering		Putative signatures (n)
			Singlets* (n)	Clusters (n)	
BaP-TT	B[a]P	17	13	1	14
MVA-MV H/F	MV H and F	145	26	11	37
MVA	MVA	1014	82	55	137
TT	TT	1769	229	163	392
ALUM	Aluminum	37	11	5	16
NEG	Background	247	29	17	46

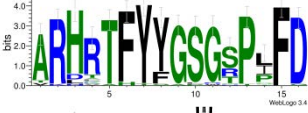
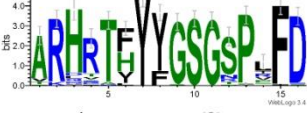
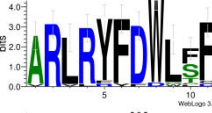



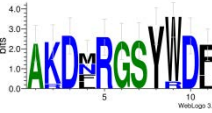



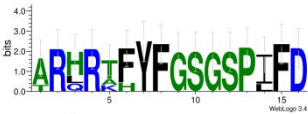

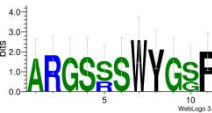




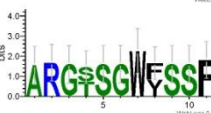
\* Singlets: CDR3 that do not enter a cluster

**Table 10:** Putative human VH CDR3 signatures for the MV H or F glycoproteins

SPLEEN			BONE MARROW		
	Representative sequence	Cluster size		Representative sequence	Cluster size
SP-HF1	ARHRTFY YGSGSPLFDY*	39	BM-HF1	ARHRTYYYGSGSPLFDY*	64
SP-HF2	ARLRYFDWLFFD	22	BM-HF2	AKDESSGWPFDP	15
SP-HF3	ARGSSGWYSAFD	14	BM-HF3	ARGSSAWYSAFD	8
SP-HF4	AKDNRGSYWDFFD	10	BM-HF4	ARGEGATQYYYYGMDV	6
SP-HF5	ARHVTHYYGSGTTPRFD*	6	BM-HF5	ATGVGVIGDFDV	6
SP-HF6	ARLRTFYFGSGSPIFDY	5	BM-HF6	ARHRTFFYYGSGSPLFDY	5
SP-HF7	ARGSSSWYGSFD	3	BM-HF7	ARHVTFYFGSGRPLFDY	5
SP-HF8	AIRYNWYEGFDF	1	BM-HF8	ARGEGATTDFDY	3
SP-HF9	AKEGDTAMVTGFDY	1	BM-HF9	ARQORTHYYGSGIPFDY	3
SP-HF10	AKEGSSNWWYWFYD	1	BM-HF10	AREYSSAWYDAFDI	2
SP-HF11	ARDENTYYYGSGSPLFDY	1	BM-HF11	ARGSSGWFFSFD	2
SP-HF12	ARDTSGWSDY	1	BM-HF12	AKHVTHYYGSGSPLFDY*	1
SP-HF13	ARGNSAWYGSFDY	1	BM-HF13	AQDNSSGWPFDY	1
SP-HF14	ARGRGYDILTGYSNFDY	1	BM-HF14	ARAGATTTFDY	1
SP-HF15	ARHATHYYGSGSPLFDF	1	BM-HF15	ARDGYSSGWNDGFDY	1
SP-HF16	ARHKTFYYGSGSPAFDI	1	BM-HF16	AREYSSAWYGWFPD	1
SP-HF17	ARLRYFDWLPFFDY	1	BM-HF17	ARGGGDPDLDY	1
SP-HF18	ARRFGSFD	1	BM-HF18	ARGSSAVLFFDY	1
			BM-HF19	ARGSSSWYGSFDS	1
			BM-HF20	ARGSSVWYKGFDF	1
			BM-HF21	ARHRTFTMVRGVPVFDY*	1
			BM-HF22	ARHRTFY YGSGSYYGFDY*	1
			BM-HF23	ARHRTFY YSSGSSPAFDY*	1
			BM-HF24	ARHRTYYYGSGTTPVVRP*	1
			BM-HF25	ARHVTFTLVRGAPSFYD	1
			BM-HF26	ARLVVGTTFTEFDY	1
			BM-HF27	ARQSGGGLNWFPD	1
			BM-HF28	ARRRGFDVLTGYSNFDY	1
			BM-HF29	ARSLWELTGDY	1
			BM-HF30	ARSVGATTELDY	1
			BM-HF31	ARVRFDILTGYSPPFDY	1
			BM-HF32	ASHSSDGFD	1
			BM-HF33	ATYGSGSPRFDP	1
			BM-HF34	GRGSSGWYDHFDY	1
			BM-HF35	GSGSPRFDP	1
			BM-HF36	SRHSELTGDY	1
			BM-HF37	TRPYCSGTSCYLDY	1





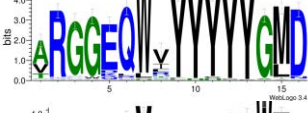
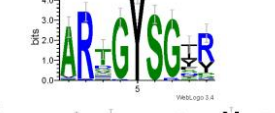


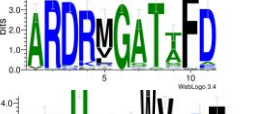


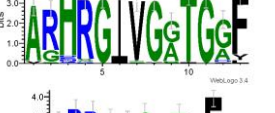
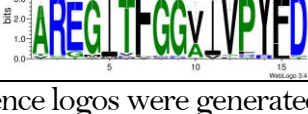

\*Similar to the CDR3 of MV-specific hybridoma clones

**Table 11:** Sequence logos for the putative MV-signatures

SPLEEN		BONE MARROW	
sequence	Cluster size	Sequence	Cluster size
	39		64
	22		15
	14		8
	10		6
	6		6
	5		5
	3		5
			3
			3
			2
			2

Sequence logos were generated using the python package WebLogo 3.4

**Table 12:** Sequence logos for the most prominent putative TT-signatures

SPLEEN		BONE MARROW	
sequence	Cluster size	Sequence	Cluster size
	252		149
	109		107
	87		51
	61		48
	53		45
	51		45
	44		43

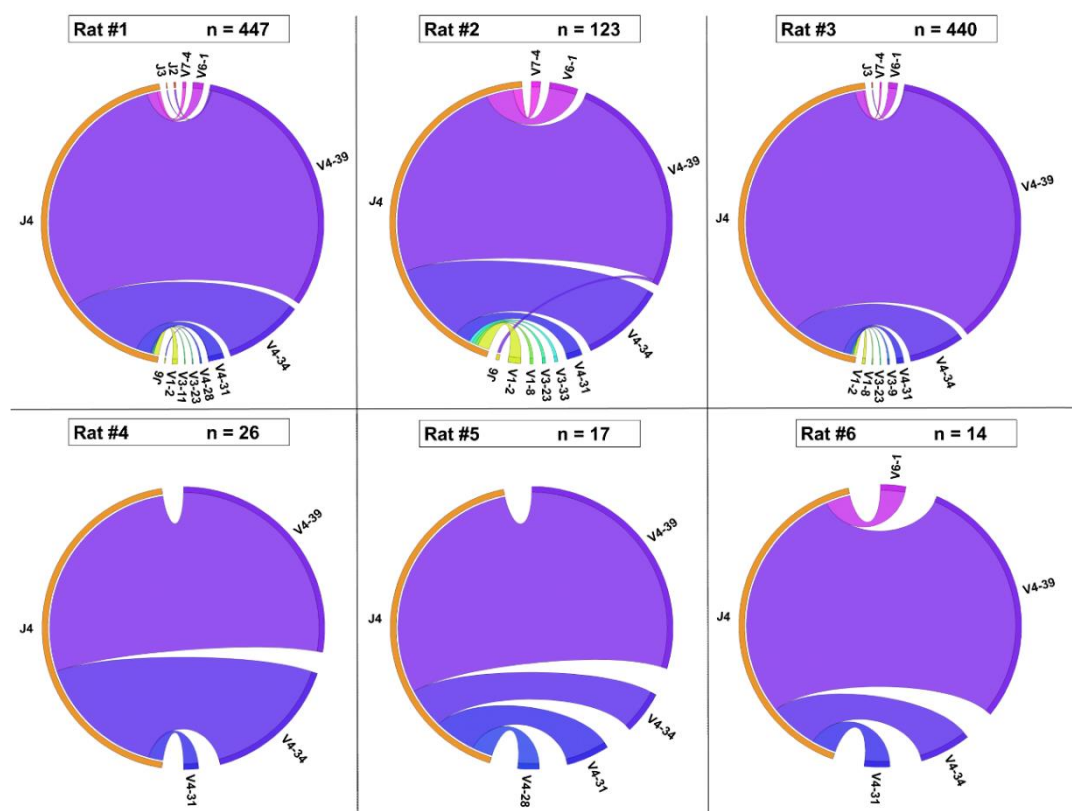
Sequence logos were generated using the python package WebLogo 3.4

#### **4.4.5 The set of putative VH CDR3 signatures for the MV H and F proteins contains previously described MV-specific sequences.**

To verify whether our approach effectively selects for antigen-specific CDR3, we used as template the CDR3 from MV-specific hybridoma clones (Dubois et al., in press). These clones were generated by traditional polyethylene glycol fusion to myeloma cells of lymph node cells from OmniRats immunized against MV antigens (whole inactivated MV of the Edmonston strain). The CDR3 sequences of 3 MV-specific clones (ARHMTYYYYGSGSPNFDY, ARHRTFYYGSGSPNFDY, and ARHRTYYYYGSGSPNFDY) corresponded to the most prominent signatures identified in both the SP (ARHRTFYYGSGSPLFDY) and the BM (ARHRTYYYYGSGSPLFDY) of the rats immunized with the MVA-MV H/F. The same V-J gene recombination as used by the hybridoma clones (V4-39/J4) was the most frequently observed. The presence of known MV-specific CDR3 among the putative signatures for the MV H or F proteins confirms that these rats mount convergent antigen-specific IG responses, and that our approach selects for antigen-specific VH CDR3 signatures. Some of the putative MV H/F signatures were also highly similar to CDR3 found in high frequency in the lymph nodes used for the generation of the MV-specific clones (not shown).

#### 4.4.6 Different V-(D)-J rearrangements result in CDR3 signatures

Multiple V and J genes encoded for the same CDR3 signature both within and across individuals. For instance, the most prominent CDR3 signature in the BM of the rats immunized with the MVA-MV H/F (BM-HF1, representative: ARHRTFYGGSGSPLFDY) was associated with 9 different V genes, 4 J genes and 12 V-(D)-J rearrangements (Fig 7.). However, a certain V-J combination was always predominantly used, and was, in most of the cases, shared by all the rats in the group. The MV H/F signature ARHRTFYGGSGSPLFDY was for instance predominantly formed by the recombination of the IGHV4-39 and IGHJ4 genes in all the rats in the immunization group. Other rearrangements, essentially involving V genes belonging to the same subgroup as the most prominently used V gene, were present but at lower frequencies. More variability was found in the J gene usage. The TT signature AKAYTSGWYVDFDL was largely associated with IGHV3-9 in all TT-immunized animals but the associated J gene varied: some rats preferentially using IGHJ2, and others IGHJ4 (not shown).



**Figure 7:** Different V-(D)-J rearrangements result in CDR3 signatures

*In all MVA-MV H/F animals, the signature BM-HF1 (representative: ARHRTFYGGSGSPLFDY) resulted from the recombination of different V and J genes. The signature was principally encoded by IGHV4-39 and IGHJ4.*

## 4.5 Discussion

Slightly different IGH V-(D)-J gene rearrangements can lead to highly similar, or even identical, junctional amino acid sequences in the same or different individual, probably because of a convergent evolution towards antibodies specific for common antigens. This phenomenon is called convergent recombination and has been illustrated essentially for the TR repertoire (356-358). While NGS studies showed that a significant part of the TR $\beta$  is shared between individuals, the IG repertoire is assumed to be essentially private (168, 358, 359). Public VH CDR3 sequences have however been identified in the IG repertoire of immunized and infected patients (192, 194, 202), open up the possibility of using the public repertoire for the sequence based-monitoring of diseases or immunizations. Here, we analyzed for the first time and high depth the modeling of the bone marrow and spleen IG repertoires by diverse antigens in transgenic rats with human IG loci. We aimed to determine whether immunization with bacterial, viral and chemically-defined antigens triggers convergent rearrangements in these rats and if antigen-specific IG signatures can be identified from their public IGH repertoire.

Despite the fact that the IG transgene undergoes normal V-(D)-J recombination, class-switching and somatic hypermutation, the expression of human antibodies in transgenic IG animals appears to be suboptimal (233). The suboptimal performance of these transgenic mice, in respect of antibody yield and immune response, is attributed to the imperfect interaction of the human constant region of membrane IG with the endogenous rodent cellular signaling machinery (234). In this context, Open Monoclonal Technology Inc. recently developed a platform for the enhanced production of human antibodies based on a transgenic humanized rat strain, the OmniRat<sup>TM</sup> (233, 235). Upon repeated IP immunizations, the rats expressed a diversified repertoire of class-switched antibodies that had undergone somatic hypermutation. While rats in the same group shared more identical VH CDR3 AA sequences between them than with individuals from other groups, indicating a convergence of the IG response to immunization, very few CDR3 were shared between all rats in a vaccination group only. However, we observed that rats in the same groups expressed VH CDR3 that were differing by only a few amino acids. In order to identify the threshold that would optimize the overlap within the groups of rats, we performed CDR3 cross-comparisons and gradually lowered the level of sequence identity (Fig.5). At 84% amino acid identity, the overlap inside the groups of rat was maximized compared to the overlap between the groups. This suggested that lowering the sequence identity threshold to 84% selects for antigen-driven sequences, and confirmed that the public IG response do not only involve identical but highly similar CDR3 sequences. Based on this threshold, we developed an approach that performs cross-comparisons

of all CDR3 sequences in all the groups to identify potential antigen-specific CDR3 signatures without pre-selecting for antigen-specific B cells. We selected for public CDR3 sequences that represent clusters of 84%-related CDR3 with significantly increased frequencies in the IG repertoires of rats in a particular group only (Table 9). We can reasonably assume that these CDR3 are antigen-driven i.e. expanded in reaction to the antigen-stimulation (194, 202).

Investigations in human subjects have shown that the VH CDR3 repertoire after TT immunization is very diverse, with only few CDR3 sequences shared between individuals (193, 360). Recently, some authors observed known (or putative) TT-specific CDR3 AA sequences shared in the blood of donors immunized with a TT-containing vaccine, demonstrating a public response (194, 214, 361, 362). Our observations in transgenic rats suggest that TT proteins also trigger convergent IG responses in our model. The immune response to TT has been extensively studied and a large number of IG sequences are available on public sequence databases. We blasted our identified TT signatures over NCBI database and found that one of these signatures was highly similar to the VH CDR3 of previously described TT-specific hybridoma clones derived from the blood of a human immune donors (363). Further investigation of the full length sequences revealed that these CDR3 were associated with the same V and J genes in both the OmniRats and the human hybridoma. A hapten-specific IgG immune response was detected by ELISA in all the B[ $\alpha$ ]P-immunized OmniRats. Only a small number of signatures (Table 9; SP= 9, BM=14) were identified from their IG repertoire, which is in line with the small size of the B[ $\alpha$ ]P molecule. However, linkage of B[ $\alpha$ ]P to TT proteins may have altered the antigenic properties of the later, and these signatures can potentially be for epitopes on the denatured TT protein rather than hapten-specific.

Although the MVA-MV H/F group acts as a control for the MVA backbone, a large number of signatures were still identified from the MVA group (SP: n=98 ; BM: n=137). We interpret this as a sign of an immune response essentially directed against the recombinant proteins in the MVA-MV H/F group. We identified 18 and 37 VH CDR3 signatures for the MV H or F glycoproteins from the spleen and the bone marrow samples, respectively. In both organs, the largest cluster of CDR3 was similar to the CDR3 sequences of MV-specific hybridoma clones derived from the lymph nodes of MV-immunized OmniRats (Dubois *et al.*, in press). This is a strong proof of concept that these rats mount convergent antigen-specific IG responses.

The IG repertoire is continuously modeled by antigen challenges and records all pathogen exposures of the individual. Recent studies in human subjects demonstrated the existence of a public IG repertoire



from which antigen-specific signatures can be identified using NGS (192, 194, 202). The CDR3 identified from public repertoire can potentially be used as markers of exposure to certain antigens, opening up new opportunities for the sequence-based diagnosis or monitoring of particular diseases (202). Here we showed that a phenomenon of convergent evolution leads to the over-expression of highly similar CDR3 in groups of immunized rats. We developed a bioinformatics pipeline that enables the identification of these convergent VH CDR3 signatures. Further confirmation of antigen-specificity should be performed. If the signatures are validated in human subjects, the public IG repertoire of these rats can potentially be source of biomarkers for a wide range of antigens. Decrypting the signatures left by specific pathogens would open the doors to an approach that could be used to retrospectively reconstruct the immune history of an individual. These investigations should demonstrate whether pathogen exposure in human can be assessed from IG repertoires using public antigen-specific signatures identified in transgenic rats with human IG loci.

## 4.6 Acknowledgement

We are extremely grateful to R. Buelow from Open Monoclonal Technology Inc. (Palo Alto, CA, USA) for providing the OmniRats. We thank Dr B. Moss, NIAID, National Institutes of Health (Bethesda, USA) for kindly providing the MVA and the recombinant MVA viruses. We also thank S. Farinelle and E. Charpentier for their technical assistance in the immunization procedures and virus culture.

## **Chapter 5.**

**DNA vaccination of transgenic**

**humanized IG rats against the**

**MV H protein of two**

**phylogenetically distant strains**

*Manuscript in preparation*

## 5.1 Abstract

Measles virus (MV) infection is undergoing resurgence and remains one of the leading causes of death among young children worldwide despite the availability of an effective vaccine. All MV strains confer cross-protection and it is to date not possible to retrospectively differentiate infected from vaccinated individuals based on their serology. However there is a certain variation in antigenic properties of the different MV strains, suggesting that the IG repertoire is differently modelled by diverse strains. Here, we immunized transgenic rats expressing human IG loci against the H proteins of two genetically distant strains of MV using a new technique of nanospheres-DNA immunization. All immunized rats mounted a MV-specific immune response of IgG antibodies with fully human idiotypes. Over 15 million of IG transcripts were obtained from the bone marrows using the Ion Torrent PGM platform. Our results indicate that the rats mount a convergent immune response against the MV H protein. We found VH CDR3 highly similar to the MV-specific hybridoma clones in the IG repertoire of two-thirds of the MV immunized rats, independently of the H strain. This observation further confirmed that a phenomenon of convergent evolution lead to the expression of antigen-specific antibodies in these rats, as we showed for protein-based immunizations. These VH CDR3 are strong markers for the MV H protein. We also found strain specific signature that would allow to discriminate between a MV vaccine strain and a wild type virus. Our data show that NGS technologies can be used to reconstruct retrospectively the immune history of a patient.

## 5.2 Introduction

The ability of B cells to respond to an almost infinite number of different antigens relies on the high polymorphism of their receptor and effector molecule, the immunoglobulin (IG). During B cell ontogeny in the bone marrow, genes encoding the variable region of the IG heavy and light chain proteins are formed by stepwise rearrangement of variable (V), diversity (D), and joining (J) genes (somatic V-(D)-J recombination) that are selected from a repertoire of multiple genes on the IG loci (1, 28). Nucleotides are deleted and randomly inserted at the joining sites, resulting in an enormous diversity of unique antigen receptors. The highly variable CDR3 of the variable domain of the IGH chain (VH) is formed by the uneven junction of the V-(D)-J genes and encodes for the principal determinant for the antigen-specificity of the antibody (25). Further diversity in the IG repertoire is added by somatic hypermutation that follow clonal expansion in germinal centers. During this process, the CDR3 are especially affected by replacement mutations that alter the affinity towards the antigen and leads to affinity maturation of the antibody (64).

Measles is a highly contagious disease caused by the Measles Virus (MV), a morbillivirus transmitted via aerosols or by direct contact with contaminated respiratory secretions (274). Despite the recent significant improvement in MV control and eradication, measles remains a major cause of childhood morbidity and mortality (239, 240). MV is divided into 8 clades A-H which are further sub-divided into 24 genotypes (258). Currently, identifying a MV strain is only possible in samples taken during the acute phase of the disease and it does not seem possible to retrospectively determine the MV strain that has caused infection and immunity. MV-neutralizing antibodies are exclusively directed against the MV H (90%) and F (10%) trans-membrane proteins, and mostly recognize conformational epitopes (283-285). The MV is considered to be relatively stable antigenically and serologically monotypic, antibodies produced against one genotype of measles being cross-protective with all other genotypes (286). However, sera from vaccinees and from naturally infected individuals differ in their ability to neutralize wild type MV strains (283, 287, 288). Similarly, MV strains can be discriminated by mAbs (289, 290). These observations suggest a certain variation in antigenic properties of the different strains, suggesting that the IG repertoire is also differently modelled by diverse MV strains.

Different V-(D)-J rearrangements can result in the same junctional sequence, in a phenomenon called convergent recombination. This process is assumed to be rare and unlikely in humans and the existence of an antigen-specific public VH CDR3 repertoire generated by convergent recombination in different

individuals is still uncertain (155, 170). Public "stereotyped" IG CDR3 have however been observed in humans with different lymphoproliferative malignancies, suggesting selection by common or similar epitopes (364). Recently, translational studies performed in human observed highly similar VH CDR3 amino acid sequences in different patients in the acute phase of dengue or donors immunized with influenza or bacterial polysaccharide vaccines (192, 194, 202). These observations highlighted that a phenomenon of convergent evolution toward antibodies with similar specificity exists in humans. Conserved VH CDR3 sequences can be used as markers for particular infections or vaccinations. However, ethical, experimental and technical limitations makes difficult the identification of IG CDR3 signatures in patients. Human studies are largely limited to the study of peripheral blood, where a limited part of the IG repertoire diversity can be assessed by next-generation sequencing. Furthermore, these studies are restricted to antigens for which an approved vaccine exists and depend on the availability of donors affected with the studied condition.

DNA vaccines consist in formulations of DNA plasmids encoding for an antigen of interest (365-368). The plasmids usually contain a strong promoter to drive the *in vivo* transcription and translation of the DNA or cDNA sequence. Upon administration, either subcutaneously, intramuscularly or mucosally, transfected cells express the foreign protein in its natural conformation and with appropriate post-translational modifications. DNA vaccines proved to be effective in animals, where they trigger good cellular and humoral immune responses. To date, prophylactic and therapeutic DNA vaccines are available on the veterinary market for use in horses (West Nile virus), salmonids (infectious hematopoietic necrosis virus), swine (against the growth hormone-releasing hormone for the treatment of *Mycoplasma hyopneumoniae* and the porcine reproductive and respiratory syndrome virus) and dogs (for melanoma treatment) (369-371). DNA vaccines are however poorly immunogenic in humans (372). Despite that over hundred phase I clinical trials have been conducted, no human DNA vaccine has been approved yet (367, 372). A possibility to increase DNA vaccine efficacy is to inject them in coated nanoparticles that protects plasmids from degradation and increase the phagocytic uptake by antigen-presenting cells (373-375).

Here, we immunized transgenic rats expressing human immunoglobulin heavy and light chain loci against the MV H protein of 2 phylogenetically distant MV strains by DNA vaccination. These rats express a diversified repertoire of human/rat chimeric antibodies with human idiotypes and have been previously used to generate MV-specific hybridoma clones and high-affinity mAbs for various antigens (235) (Dubois *et al.*, 2015; in press). Using the Ion Torrent PGM platform, we sequenced the IG

repertoire in the bone marrows using an approach that allowed to uniquely label each mRNA molecule present in the sample. Over 15 million IG transcripts have been generated. We investigated how the IG repertoire in these rats is modeled by the H protein of a vaccine strain (clade A; Edmonston) and a wild type stain (genotype D11). We observed highly similar VH CDR3 to MV-specific hybridoma clones in the rats in the 2 immunization groups. These VH CDR3 constitute signatures for the MV H protein independently of the MV strain. Further in-depth analysis will determine whether strain-specific signatures can be identified from the IG repertoire. We also found strain specific signature that would allow to discriminate between a MV vaccine strain and a wild type virus. Our data show that NGS technologies can be used to reconstruct retrospectively the immune history of a patient.

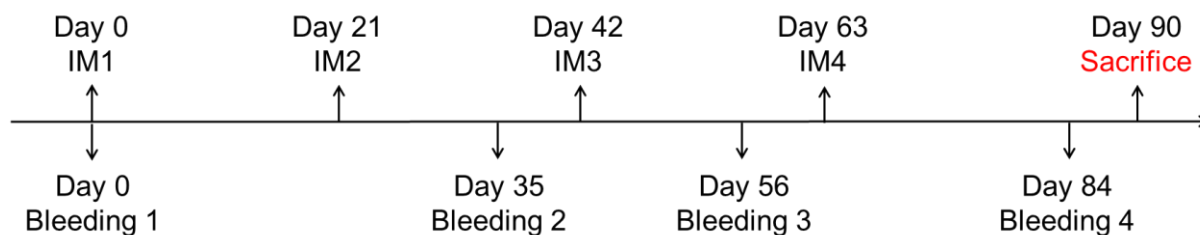
## 5.3 Materials and methods

### 5.3.1 Production of DNA vaccines against MV haemagglutinin of clade 1 and clade D11

Synthesis of the nucleotide sequences and cloning in immunization plasmid pVAX was performed by ATG:biosynthetics (Merzhausen, DE) (374). pVAX confers kanamycin resistance and protein expression under the control of the CMV promoter in mammal cells. Plasmids were amplified in DH5 $\alpha$  *Escherichia coli*, and purified using the EndoFree plasmid purification kit (Qiagen, Courtaboeuf, FR). The quality control of each plasmid was monitored by electrophoresis on an agarose gel to confirm that the supercoiled form predominated. Analyses of restriction enzyme profiles were also performed to check the plasmid integrity. Plasmid concentrations were determined by UV spectrophotometry and fluorometry (374, 376).

### 5.3.2 Animals and immunizations

Humanized transgenic rats (OmniRat<sup>TM</sup>, Open Monoclonal Technology Inc., Pablo Alto, USA) were developed and bred as recently described (235, 331-333). These animals carry a chimeric human/rat IGH locus, where 22 human IGHV, all IGHD and all IGHJ genes in in germline configuration are linked to the rat IGHC genes, together with human IG kappa locus (12 IGKV, 5 IGKJ and 1 IGKC) and IG lambda locus (16 IGLV, 5 IGLJ and 5 IGLC) (344). The endogenous rat IG loci have been silenced using zinc finger nucleases (354). Thirty-two animals (14 males, 18 females) were maintained on a 12h light/dark cycle with *ad libitum* access to food and water. Rats were separated into 4 groups of 8 individuals. Animals in the first group were immunized against the MV H protein from the clade A (group MVH-A) and rats in the second group against the MV H protein of the genotype D11 (group MVH-D11) (258, 377). The third group of rats received the adjuvanted synthetic vector only (pVAX group) and animals in the fourth group received no immunization (NEG group). Nanospheres were assembled using the ICANtibodies<sup>TM</sup> technology as described (374). Animals received 4 intramuscular injections (IM) of nanospheres at 20-days intervals and were sacrificed 27 days after the last injection (Fig. 2). All injections contained the same amount of DNA formulated with the synthetic vector. All immunizations were performed by In-Cell-Art Inc., Nantes, France. For each animal, the anti-MV IgG immune response was monitored by specific ELISA at days 0, 35, 56 and 84. All animal procedures were in compliance with the rules described in the Guide for the Care and Use of Laboratory Animals (334).



**Figure 1:** Immunization scheme

### 5.3.3 Anti-MV IgG ELISA

Antigen-specific IgG antibody levels in the sera were determined in 384-well microtiter plates (Greiner bio-one, Wemmel, BE) coated overnight at 4°C with 250 ng of MV antigen (Measles grade 2 antigen, Microbix Biosystems, Mississauga, CA). Free binding sites were saturated with 1% bovine serum albumin (BSA) in Tris-buffered saline at room temperature for 2h. Serial dilution of the sera were added for 90 min at 37°C, and developed with alkaline phosphatase (AP)-conjugated goat anti-rat IgG (1/750 dilution, ImTec Diagnostics, Antwerp, BE) and the appropriate substrate at 405 nm. Endpoint titers (EPT) were determined as the serum or supernatant dilutions corresponding to 5 times the background.

### 5.3.4 Sample preparation, amplification and Ion Torrent Deep Sequencing

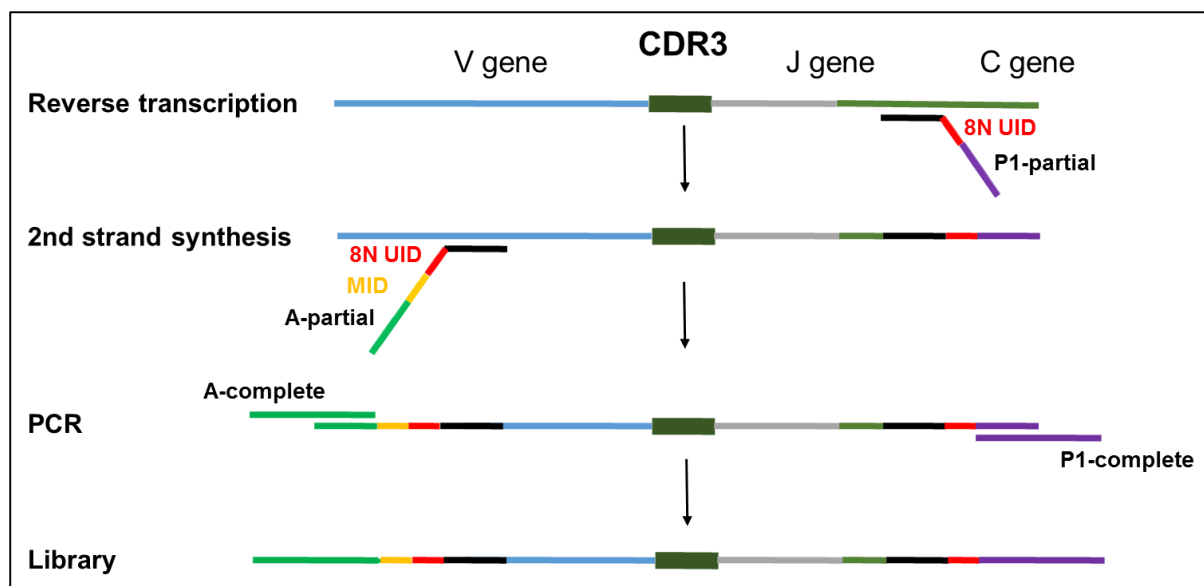
Lymphocytes were isolated from the bone marrow by gradient centrifugation (Ficoll Paque PLUS; GE Healthcare Life Sciences, Bornem, BE), total RNA extracted (RNeasy midi kit, Qiagen, Venlo, NL) and mRNA enriched by paramagnetic separation ( $\mu$ MACS mRNA Isolation kit, Miltenyi Biotec, Leiden, NL). Sequencing libraries were prepared using a method derived from Vollmers et al. where each IG mRNA molecule originally present in the sample is uniquely labelled (Fig. 2) (189). cDNA was prepared at 55°C for 80 min using Superscript III RT (Invitrogen, Merelbeke, BE) and primer for the rat IgG extended by 8 random nt (UID) and partial sequence of the P1 adapter for Ion Torrent sequencing. Second-strand synthesis was performed using Phusion High-Fidelity DNA polymerase (New England Biolabs, Frankfurt-Main, DE) and primers covering all V genes carried by the OmniRat containing 8 random nt (UID) and partial sequence of the A adapter for Ion Torrent sequencing (Table 1) (98 °C for 2 min, 57 °C for 2 min, and 72 °C for 10 min). Double-stranded cDNA was purified twice (Agencourt AMPure XP; Beckman Coulter) and subsequently amplified by PCR for 20 cycles using primers



completing the adapter sequences and Q5 Hot Start High Fidelity polymerase (New England Biolabs). Amplicons were purified (Agencourt AMPure XP; Beckman Coulter) and quantified (Quant-iT PicoGreen dsDNA, Invitrogen). Quality was checked with a Bioanalyzer (High Sensitivity DNA, Agilent Technologies, Diegem, BE). Libraries were pooled in equimolar concentrations and sequenced on a 318™ Chip v2 using multiple identifiers (MIDs) with the Ion OneTouch™ Template OT2 400 Kit and the Ion PGM Sequencing 400 Kit (Life Technologies, Gent, BE) on the Ion Torrent Ion Personal Genome Machine (PGM™) System (Life Technologies). Four randomly selected samples were sequenced on a chip.

**Table 1:** Primer list

Name	Primer sequence
First strand primer (C region rat)	
pP1_8N_IgG CH1	5'-ctatgggcagtcggtgatNNNNNNNNNggatagacagatggggctgtt-3'
Second strand primers (V region human)	
pA_MID_8N_VH1-FR2	5'-gcgtgtctccgactcag- <b>MID</b> -NNNNNNNNNctggacaagggcttgagtgg-3'
pA_MID_8N_VH2-FR2	5'-gcgtgtctccgactcag- <b>MID</b> -NNNNNNNNNcaggaaagggccttgagtgg-3'
pA_MID_8N_VH3-FR2	5'-gcgtgtctccgactcag- <b>MID</b> -NNNNNNNNNcagggaagggcttgagtgg-3'
pA_MID_8N_VH4-FR2	5'-gcgtgtctccgactcag- <b>MID</b> -NNNNNNNNNcagggaagggacttgagtgg-3'
pA_MID_8N_VH6-FR2	5'-gcgtgtctccgactcag- <b>MID</b> -NNNNNNNNNcatcgagaggccttgagtgg-3'
Amplification primers	
amp_A	5'-ccatctcatcctgcgtgtctccgactcag-3'
amp_P1	5'-cctctctatgggcagtcggtgat-3'



**Figure 2:** Library preparation according to the UID method (189)

*IGH mRNA molecules were reverse transcribed using a primer specific for the constant region (IgG-specific) extended with a string of 8 random nucleotides (UIDs) and a partial P sequencing adapter. The second strand synthesis was performed using a pool of 5 V gene-specific primers (binding to FR2) extended with 8 random nucleotides, a sample specific MID and a partial A sequencing adapter. IGH sequences were next amplified by PCR using primers binding to the partial adapter sequences. After Ion Torrent sequencing, raw nt reads with identical UIDs are grouped. For each UID group, a consensus sequence is built, allowing for error-corrections.*

### 5.3.5 Quality control and sequence annotation

Reads from each rats were identified based on a perfect MID match (no mismatches allowed). Only the reads with a UID tag, primers identified at both ends (2 mismatches allowed) and from which more than 85% of the bases had a quality score above 25 were considered for further analysis. For each rats, sequences with identical UIDs were grouped and a consensus nucleotide sequence was built. As a result, to each IG nucleotide sequence in the dataset corresponds a single mRNA molecule in the sample. This approach corrects for the potential preferential amplification of particular IG sequences and allows to correct for PCR and sequencing errors (e.g. substitutions, indels...) for the sequences with a consensus sequence.

The primer sequences were trimmed. Variable (V), Diverse (D), and Joining (J) genes were assigned using a local installation of IgBlast (336) with only the genes present on the OmniRat's genome used as reference. Only the productive sequences for which a V and J gene were unambiguously assigned (BLAST E-value  $>10^{-9}$ ) were analyzed further. The clean dataset of sequences was subsequently submitted to the IMGT/HighV-QUEST webserver (IMGT, [www.imgt.org](http://www.imgt.org); Marie-Paule Lefranc,

Montpellier, France) for annotation of the different V gene regions (143). The results presented here are for the forward and reverse reads combined.

### **5.3.6 Identification of VH CDR3 signatures**

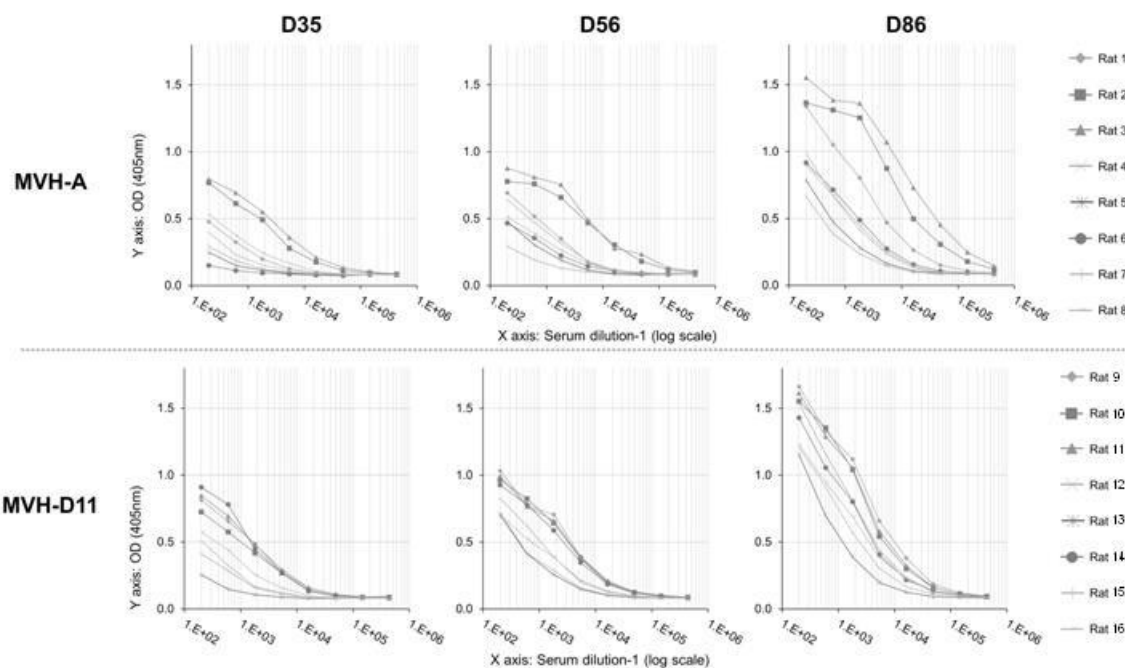
The VH CDR3 amino acid sequences of all animals were combined and made non-redundant. For the clustering analysis, a local version of CD-HIT was used ([http://weizhong-lab.ucsd.edu/cdhit\\_suite](http://weizhong-lab.ucsd.edu/cdhit_suite); L. Weizhong, San Diego, USA) (24, 25). CD-HIT allows to group sequences into clusters based on user-defined sequence identity threshold. Each sequence is assigned to one and only one cluster. The sequences were clustered at 80% identity and at 85% identity. No length difference was allowed.

## 5.4 Results

### 5.4.1 The MV-specific antibody response is higher in humanized transgenic rats immunized with the wild type H protein

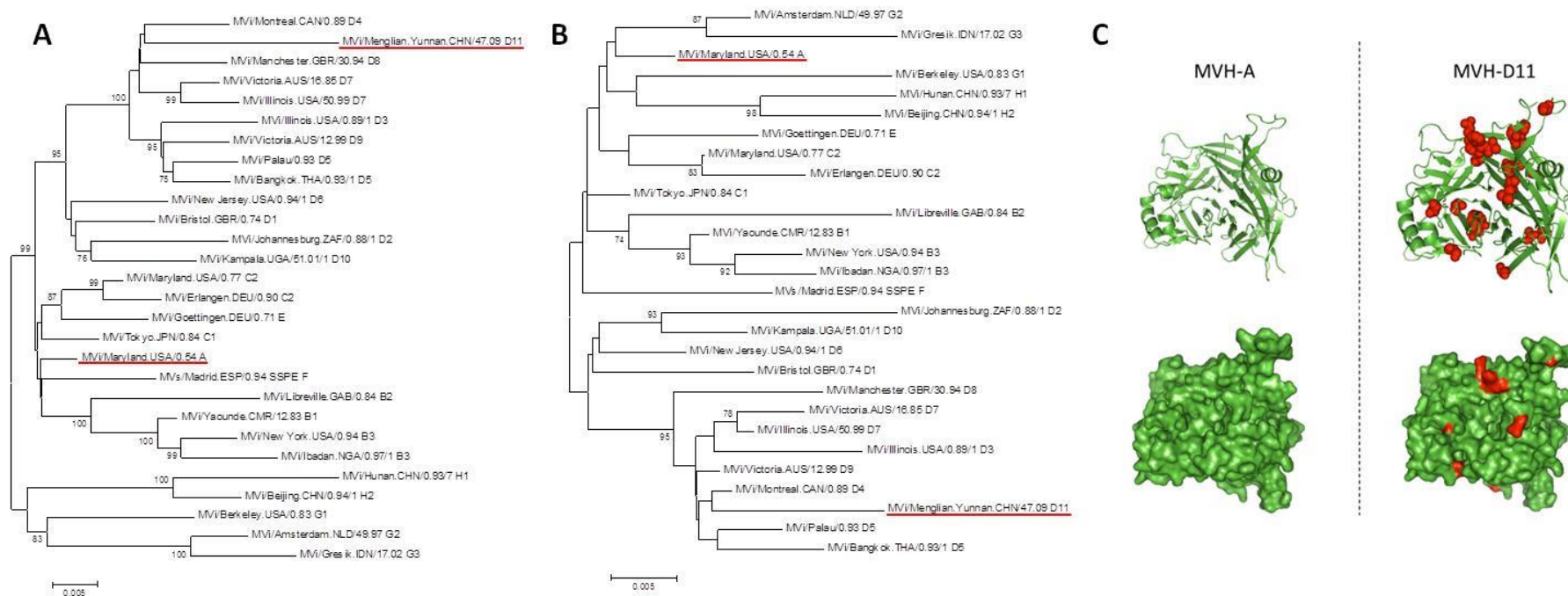
Rats were immunized against 2 genetically distant strains of MV H proteins using a new technique of nanospheres-DNA immunization. The H genes of the MV Edmonston Vaccine strain (MVH-A) and of the recently described wild type genotype D11 (MVH-D11) were chosen for our study. The H gene of genotype D11 is the second most distant from clade A on nucleotide basis and the most distant on the amino acid level (Fig. 3A & 3B). The amino acid differences between the strains are principally located in the outer surface of the H protein (Fig. 3C). Thirty-two rats with a human B cell IG repertoire were separated into 4 groups of 8 rats each. The first group of rats was immunized with a DNA vaccine containing the H proteins of the MVH-A strain. The second group received a DNA vaccine containing the H protein of the MVH-D11 strain. Control animals received empty DNA vectors with all adjuvants (pVAX group) or were left untouched (NEG group). The anti-MV IgG immune response was monitored by specific ELISA at days 35, 56 and 84 (Fig. 4). A specific immune response was found in all the immunized animals while the mock immunized animals (pVAX group) and unimmunized animals (NEG group) showed no detectable antigen-specific antibodies (Fig. 4, Table 2). The IgG immune response was stronger and more homogenous in the rats immunized against the wild type H protein (MVH-D11) at all time-points (Fig. 4). Two weeks after the second immunization (day 35), 7 rats out of 8 had detectable amount of anti-MV IgGs, while anti-MV IG were detected in only 4 rats of the MHV-A group (Table 2). At day 56 (2 weeks after the third injection), anti-MV IgG levels increased in all the rats except one in the MHV-A group (Fig. 4, Table 2). These levels were on average higher in the MVH-D11 animals (average EPT: 1:552) than in the MVH-A rats (1:1,485). While the IgG levels in the MVH-D11 group were relatively homogeneous (EPT range: 1:5,115 to 1:593), they varied largely between rats in the MVH-A group (<1:200 to 1:9,210). After the fourth injection (day 86), 2 subgroups could be identified in the MVH-D11 group based on the level of the anti-MV IgG response (Fig. 4, Table 2). Three rats had EPT below 1:10,000 (1:11,344 to 1:15,061) and 5 rats were above this threshold (1:3,965 to 1:5,365). The average anti-MV IgG level at day 84 was >2.5 times higher for the MVH-D11 group (average EPT: 1:5,014) than for the MVH-A group (1:1,913). Surprisingly, the anti-MV IgG levels measured for the MVH-A immunized animals varied largely between them, with animals being very high and others low responders (EPT range: 1: 565 to 1:69,384). The highest levels of anti-MV IgGs,

independently of the immunization group, were measured in 2 of these rats (EPT: 1:69,384 and 1:31,132). The other individuals in the MVH-A group had however much lower MV-specific IgG levels (EPT: 1:8,535 to 1:565)(Fig.4, Table 2).



**Figure 4:** Anti-MV IgG ELISAs

*Anti-MV IgG in the sera of the rats immunized against the H protein of the clade A (MVH-A; top panel) or of genotype D11 (MVH-D11; bottom panel) were measured by indirect ELISA. The serum dilution<sup>1</sup> (1:100 to 1:656100) was plotted against binding measured by adsorption at 405 nm. The IgG immune response was on average stronger and more homogenous in the rats immunized against the wild type H protein.*



**Figure 3:** MV haemagglutinin genes from MV clade A and genotype D11

*D11 is one of the most phylogenetically distant strain to the vaccine A strain. (A&B) The phylogenetic trees based on the full nucleotide (1,854 nt) (A) and amino acid (617 AA) (B) sequences of H gene were generated using MEGA6 (Kimura 2-parameter model). The branch length reflects the evolutionary rate of individual WHO H gene reference sequences. (C) The cartoon models and surface presentation of MV H proteins were generated by using PyMOL (<http://pymol.sourceforge.net>). Amino acid differences to MV H clade A are colored in red for MV H genotype D11. Most differences are located on the surface of the H protein and result in slight structural changes.*

**Table 2:** Endpoint titers

Treatment	Endpoint titers		
	D35	D56	D86
MVH-A N=8	1:389	1:1,326	1:8,535
	1:3,270	1:8,836	1:31,132
	1:4,566	1:9,210	1:69,385
	<1:200 <sup>b</sup>	<1:200 <sup>b</sup>	1:565
	<1:200 <sup>b</sup>	1:341	1:997
	<1:200 <sup>b</sup>	1:379	1:3,134
	<1:200 <sup>b</sup>	1:1,028	1:1,913
	1:524	1:510	1:2,475
Average	1:320	1:552	1:1,913
MVH-D11 N=8	1:3,180	1:5,115	1:15,061
	1:2,107	1:4,863	1:11,344
	1:3,277	1:4,976	1:12,631
	1:220	1:1,659	1:6,907
	<1:200 <sup>b</sup>	1:593	1:1,719
	1:2,657	1:4,356	1:5,365
	1:428	1:619	1:3,965
	1:819	1:1,539	1:5,081
Average	1:549	1:1,485	1:5,014

n.d. : no reactivity over background was detected in the control rats (pVAX and NEG groups)

<sup>b</sup> For sera with no detectable antibody, EPT was set to 1:200 (highest serum concentration tested)

### 5.4.2 NGS of the rearranged IGH repertoire

Rearranged IGH sequences (IgG) were amplified from mRNA extracted from lymphocytes of the bone marrow (BM) and sequenced on the Ion Torrent PGM platform. After quality control, reads were sorted based on individual barcodes (MIDs). Reads with identical UID combinations were next grouped together, and a consensus sequence was built (Table 3). Most of the reads were uniquely labeled, suggesting that there is a minimal amplification bias in our datasets (not shown). Beside correcting for this bias, the so-called UID approach also enabled to correct for sequence errors introduced during PCR and sequencing. The IGH rearrangements were subsequently annotated by the IMGT/HighV-QUEST webserver and IgBlast. Only the productive sequences with an unambiguously assigned V and J gene were considered. A total of 15,557,443 filtered nucleotide sequences was obtained (Table 3). The VH CDR3 amino acid sequences were extracted and collapsed to non-redundant sequences (47,455 unique CDR3). The CD-HIT program was used to group CDR3 in clusters of sequences with 80% and 85% homology. Each sequence is uniquely assigned to a single cluster. A total of 30,944 and 24,976 clusters were obtained for the 80% and the 85% identity levels, respectively. From these, 26,973 (87%) and 20,033 (80%) contained sequences derived from a single rat only, indicating that the VH CDR3 repertoires of the rats is essentially private.



**Table 3:** Ion Torrent read counts

GROUP	Rat ID	1 QC reads*	2 Filtered nt sequences**	3 Unique CDR3 AA
MVH-A N=8	1	321,825	228,024	14,352
	2	629,053	481,088	34,501
	3	591,963	480,108	32,671
	4	396,440	334,392	29,857
	5	872,422	681,323	54,117
	6	1,041,263	576,833	37,055
	7	518,522	410,782	27,084
	8	662,610	500,621	36,766
Total		5,034,098	3,693,171	266,403
Average $\pm$ stdev		629,262 $\pm$ 236,752	461,646 $\pm$ 139,961	33,300 $\pm$ 11,154
MVH-D11 N=8	9	554,643	411,946	42,505
	10	748,014	588,875	42,505
	11	940,279	774,391	49,364
	12	1,052,895	868,023	63,364
	13	643,068	227,555	19,530
	14	769,356	301,980	29,472
	15	569,294	466,967	39,459
	16	513,509	361,459	29,871
Total		5,791,058	4,001,196	316,070
Average $\pm$ stdev		723,882 $\pm$ 193,115	500,149 $\pm$ 226,812	39,509 $\pm$ 13,506
pVAX N=8	17	531,345	404,841	49,364
	18	578,467	422,229	63,364
	19	546,871	450,554	19,530
	20	633,439	498,036	29,472
	21	618,934	330,900	39,459
	22	869,287	706,160	29,871
	23	1,002,611	842,609	79,006
	24	455,937	335,322	24,340
Total		5,236,891	3,990,651	334,406
Average $\pm$ stdev		654,611 $\pm$ 185,543	498,831 $\pm$ 182,498	41,801 $\pm$ 20,726
NEG N=8	25	414,772	325,283	23,020
	26	680,754	513,213	30,571
	27	477,359	398,858	27,138
	28	738,161	580,913	31,839
	29	735,617	522,554	28,266
	30	763,944	535,528	24,326
	31	863,379	721,191	47,772
	32	405,091	274,885	20,728
Total		5,079,077	3,872,425	233,660
Average $\pm$ stdev		634,885 $\pm$ 176,413	484,053 $\pm$ 144,823	29,207 $\pm$ 83,389
TOTAL N=32		21,141,124	15,557,443	1,150,539
Average $\pm$ stdev		660,660 $\pm$ 193,251	486,170 $\pm$ 168,951	35,954 $\pm$ 14,416

\* number of nucleotide reads after Quality Check

\*\* number of nucleotide reads after building of the consensus sequences based on identical UID

### 5.4.3 Convergent IG signatures for conserved epitopes on the MV H protein

In order to identify the public response for the MV H protein, we analyzed clusters composed of sequences belonging to several immunized rats and that do not contain any sequence from the mock or unimmunized animals. When grouping the MVH-A and MVH-D11 rats together, a total 1,185 clusters at 80% homology and 1,011 at 85% homology contained sequences of at least two MV H immunized rats. Both at 80% and 85% homology, 78% of these clusters were shared between exactly two rats (921 clusters at 80% homology and 789 at 85% homology). No clusters included CDR3 sequences from all 16 immunized rats, and at most 9 rats (80%-homology) and 8 rats (85% homology) had CDR3 in a cluster. We found 5 and 3 clusters shared between at least half of the immunized rats at 80% and 85% homology levels, respectively (Table 4, next page). Among the CDR3 sequences shared by the highest number of animals we identified the CDR3 of 3 previously described MV specific hybridoma clones (Dubois et al, in press). Identical signatures had also been previously observed in the bone marrow and the spleen of OmniRats immunized with a Modified Vaccinia Ankara virus encoding for the MV H and F glycoprotein (MVA-MV H/F) (Dubois *et al.*, manuscript in preparation).

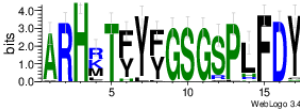
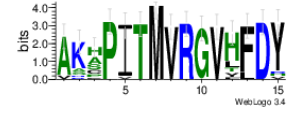
We used as template the CDR3 amino acid sequences of these MV-specific hybridoma clones. Interestingly, we found related CDR3 sequences (> 84% identity) in the bone marrow of 11 out of the 16 MV immunized rats independently of the H strain used for immunization (68.75%; Table 5). This indicates that these VH CDR3 are signatures of the MV H protein and that antibodies encoding for these CDR3 are targeting a conserved epitope between the vaccine and the wild type D11 strain.

**Table 5:** 84%-related CDR3 amino acid sequences to the MV CDR3 M\*

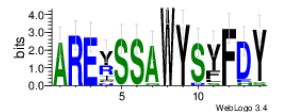
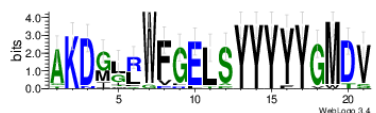
GROUP	Rat ID	Unique 84%-related CDR3	# of sequences
MVH-A (6/8)	#2	3	282
	#3	2	1024
	#4	3	77
	#5	1	143
	#6	26	6513
	#7	1	560
	MVH-D11 (5/8)	#10	26
#11		1	2
#14		12	1308
#15		6	559
#16		3	68

\* MV CDR3 L, M and N returned similar counts

**Table 4:** Convergent CDR3 for conserved epitopes on the MV H protein

Cluster 80%																		
Unique CDR3	Sequence Logo	MV A								MV D11								Num. of rats
		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	
ARHKTFYFGSGSPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	258	.	9	
ARHKTFYYGSGSPLFDC		.	.	.	.	.	.	.	.	.	.	.	.	.	.	37		.
ARHKTFYYGSGSPLFDS		.	.	.	.	.	.	.	.	.	.	.	.	.	.	54		.
ARHKTFYYGSGSPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	225		.
ARHKTHYYGSGSPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		4
ARHKTYFYGSGRPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	9		.
ARHKTYYYGSGSPHFDY		.	.	.	.	.	.	225	.	.	.	.	.	.	.	.		.
ARHKTYYYGSGSPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	228		.
ARHLTFLFGSGSPRFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		5
ARHLTFYFGSGSPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	3		.
ARHLTFYFGSGSPRFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		35
ARHMTFYFGSGSPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	163		.
ARHMTFYFGSGSPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	17		.
ARHMTTHYYGSGSPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	4		.
ARHMTYYFGSGRPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	143		.
ARHMTYYFGSGLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	4		.
ARHMTYYFGSGSPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	573		.
ARHMTYYFGSGSPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	1043		.
ARHRTFLLGSGSPLFDY		.	68	.	.	.	.	.	.	.	.	.	.	.	.	.		.
ARHRTFYFGSGSPHFDY		.	.	8	.	.	.	.	.	.	.	.	.	.	.	.		.
ARHRTFYFGSGSPLFDY	.	1066	.	.	.	.	.	.	.	.	.	.	5	.	.	.		
ARHRTFYFGSGSPLFDY	.	.	.	.	.	.	34	.	.	.	.	.	172	.	.	.		
ARHRTFYFGSGSPPFDY	.	.	.	.	.	.	.	683	.	.	.	.	.	.	.	.		
ARHRTYYFGSGCPLFDY	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.		
ARHRTYYFGSGSPHFDY	.	.	60	.	.	.	.	.	.	.	.	.	.	.	.	.		
ARHRTYYFGSGSPLFDY	.	.	.	.	.	.	.	.	.	.	.	.	36	.	148	22		
ARHRTYYFGSGSPLFDY	.	.	.	.	.	.	.	.	.	2	.	.	527	.	.	54		
ARHYTFYYGSGRPLFDY	.	.	.	.	.	.	.	.	.	.	.	303	.	.	.	.		
VRHRTFYFGSGSPLFDY	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
sum		0	1136	68	0	0	0	259	683	0	2	0	303	741	0	2936	119	
AAHPITMVRGVHFDY		.	.	.	.	.	.	.	.	.	.	.	23	.	.	.	9	
AAHPITMVRGVHFDY		.	.	.	.	.	.	.	.	.	.	.	.	22	.	.		.
AKAPITMVRGVHFDH		.	.	.	.	.	.	.	.	161	.	.	.	.	.	.		.
AKAPITMVRGVHFDV		.	.	.	.	.	.	.	.	24	.	.	.	.	.	.		.
AKAPITMVRGVHFDY		.	.	.	.	.	.	.	.	1173	.	.	.	.	.	.		.
AKAPITMVRGVLFYD		.	.	.	.	.	.	.	.	2100	.	.	.	.	.	.		.
AKCPITMVRGVLFYD		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		26
AKCPITMVRGVTFYD		.	.	.	.	.	.	.	.	.	.	.	.	.	.	33		.
AKFPITMVRGVYFDY	.	299	.	.	.	.	.	.	.	.	.	.	.	.	.	.		

AKHPITMVRGVFLDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	81	
AKHPITMVRGVTLDY		.	.	.	.	.	.	.	.	.	.	.	.	.	19	.	
AKPPITMVRGVHFDY		304	220	.	.	26	.	.	212	.	.	.	101	.	.	.	
AKSPITMVRGVHFDY		.	.	.	.	.	.	.	.	.	.	15	4	.	.	.	
AKSPITMVRGVHLDY		.	.	.	.	.	.	.	.	.	.	318	.	.	.	.	
AKSPITMVRGVYFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	79	
ANSPITMVRGVHFDY		.	.	.	.	.	.	.	.	.	.	.	6	.	.	.	
ARAPITMVRGVFLDY		.	.	.	.	.	.	.	.	113	.	.	.	.	.	.	
ASHPITMVRGVHFDY		.	.	.	.	.	.	.	.	.	.	.	34	.	.	.	
ASHPITMVRGVHFDY		.	.	.	.	.	.	.	.	.	.	.	17	.	.	.	
VKCPITMVRGVHFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	581	
VKPPITMVRGVHFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
sum		304	519	0	0	615	0	0	212	3570	0	0	332	207	0	633	186
AKDGLLWFGELSYYYYYGMDV		.	.	.	.	.	.	.	.	.	.	.	26	.	.	.	.
AKDGLLWFGELSYYYYYVWTS		.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
AKDGLRWFGELAYYYYYGMDV		.	.	.	.	.	.	.	.	.	.	.	138	.	.	.	.
AKDGLRWFGELSYYYYYGMDV		1014	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AKDGLRWFGELSYYYYYGMTS		5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AKDGLSWFGELSYYYYYGMDV		.	.	.	.	.	.	.	.	.	.	.	30	.	.	.	.
AKDGWFGELLTPYYYYGMDV		.	.	.	.	.	.	.	.	11	.	.	.	.	.	.	.
AKDIGRWFGELSYYYYYGMDV		.	.	.	.	.	.	115	.	.	.	.	.	.	.	.	.
AKDIGRWVFGELSYYYYYGMDV		.	.	.	.	.	.	43	.	.	.	.	.	.	.	.	.
AKDMGLWFGELSYFYFGMDV		.	.	.	.	.	.	.	.	.	.	.	4	.	.	.	.
AKDMGRWFGELSYYYYYGMDV		.	.	.	24	.	.	.	.	.	.	.	.	.	.	.	.
AKDMILWFGELSYYYYYGMDV		.	.	.	.	.	.	.	1138	.	.	.	.	.	.	.	.
AKDMILWVRELSYYYYGMDV		.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.
AKDMTLWFGELSYYYYYGMDV		.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.
AKNGLRWFGELSYYYYYGMDV		4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SKDIGLWFGELSYYYYYGMDV		.	.	.	.	.	.	12	.	.	.	.	.	.	.	.	.
TKDGLRWFGELSYYYYYGMDV		113	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sum		1136	0	24	2	0	170	0	1140	11	0	0	4	194	0	0	0
ARESSAWYSYFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	.	204	.
AREISSAWYSFFDY		.	.	.	.	.	.	.	.	41	.	.	.	.	.	.	.
AREISSAWYSYFDY		697	.	.	.	.	.	.	.	166	.	.	.	.	.	.	.
ARENSSAWYSLFDY		.	.	.	.	.	.	.	.	542	.	.	.	.	.	.	.
ARERSSAWYSFFDY		.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ARERSSAWYSYFDY		.	730	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ARERSSAWYSYFDY		1554	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ARERSSAWYSYFEY		3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AREYSSAWYDYFDY		.	.	.	.	.	.	.	.	2289	.	.	.	.	.	.	.
AREYSSAWYSFFDY		.	.	.	.	997	.	.	.	.	.	27	.	.	.	.	.
AREYSSAWYSYFDS		.	.	.	.	.	.	.	.	171	.	.	.	.	.	.	.
AREYSSAWYSYFDY		.	.	.	.	139	.	.	.	1701	356	.	407	.	.	.	.
AREYSSAWYSYFHY		.	.	.	.	.	.	.	.	598	.	.	.	.	.	.	.
AREYSSGWYSYFDY		.	.	.	.	.	.	.	.	606	.	.	85	.	.	.	.
sum		2254	732	0	0	997	139	0	0	5364	1105	0	519	0	0	204	0
ASKTTVTTRFDY		451	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ASLTSVTTRFDY		.	.	.	.	.	.	.	.	30	.	.	.	.	.	.	.



8

8

8

ASLTTVTTRFAY		.	.	.	.	.	.	.	26	.	48	.	.	2475	.	.	.
ASLTTVTTRFDS		.	.	.	.	.	.	.	.	90	.	.	.	.	.	.	.
ASLTTVTTRFDY		.	.	.	.	.	.	.	60	872	116	.	.	.	.	.	.
ASMTVTTRFAY		.	.	.	.	.	.	.	.	.	.	.	.	525	.	.	.
ASRATVTTRFDY		.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
ASRTTVTTRFAY		.	.	.	64	.	.	.	.	.	.	.	.	.	.	.	.
ASRTTVTTRFDC		.	.	.	.	.	.	.	.	.	.	.	.	.	18	.	.
ASRTTVTTRFDS		.	.	.	81	.	.	193	.	.	.	.	.	.	.	.	.
ASRTTVTTRFDY		295	.	.	1716	.	.	.	.	.	.	.	.	33	.	.	.
ASRTTVTTRFGY		.	.	.	16	.	.	.	.	.	.	.	.	.	.	.	.
ASRTTVTTRFH Y		.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.
ASRTTVTTRSDY		.	.	.	11	.	.	.	.	.	.	.	.	.	.	.	.
sum		746	0	0	1894	0	0	193	86	992	164	0	0	300	51	0	0
<b>Cluster 85%</b>																	
CDR3 sequence	Sequence Logo	Count MV A							Count MV D11							Num. of rats	
ARECSSAWYSYFDY		.	.	.	.	.	.	.	.	.	.	.	.	.	204	.	
AREISSAWYSFFDY		.	.	.	.	.	.	.	.	41	.	.	.	.	.	.	
AREISSAWYSYFDY		697	.	.	.	.	.	.	.	166	.	.	.	.	.	.	
ARENSSAWYSLFDY		.	.	.	.	.	.	.	.	542	.	.	.	.	.	.	
ARERSSAWYSFFDY		.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	
ARERSSAWYSSFDY		.	730	.	.	.	.	.	.	.	.	.	.	.	.	.	
ARERSSAWYSYFDY		1554	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
ARERSSAWYSYFEY		3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
AREYSSAWYDYFDY		.	.	.	.	.	.	.	.	2289	.	.	.	.	.	.	
AREYSSAWYSFFDY		.	.	.	.	997	.	.	.	.	.	27	.	.	.	.	
AREYSSAWYSYFDS		.	.	.	.	.	.	.	.	171	.	.	.	.	.	.	
AREYSSAWYSYFDY		.	.	.	.	139	.	.	.	1701	356	.	407	.	.	.	
AREYSSAWYSYFHY		.	.	.	.	.	.	.	.	598	.	.	.	.	.	.	
AREYSSGWYSYFDY	.	.	.	.	.	.	.	.	606	.	.	85	.	.	.		
sum	2254	732	0	0	997	139	0	0	5364	1105	0	519	0	0	204	0	
ARHKTYYYGSGSPHFDY		.	.	.	.	.	.	225	.	.	.	.	.	.	.		
ARHMTYYYGSGSPLFDY		.	.	.	.	.	.	.	.	.	.	.	.	1043	.		
ARHRTYYYGSGSPHFDY		.	.	60	.	.	.	.	.	.	.	.	.	.	.		
ARHRTYYYGSGSPLFDY		.	.	.	.	.	.	.	.	2	.	.	527	.	54		
ARHRTYYYGSGSPFDY		.	.	.	.	.	.	.	4419	.	.	.	.	.	.		
ARHYTYYYGSGSPVFDY		.	.	.	.	.	.	.	.	.	.	4	.	.	.		
sum	0	.	60	.	.	.	225	4419	.	2	.	4	527	1043	54		
ARGDSSSSFFDY		.	.	.	.	.	.	.	.	.	.	.	30	.			
ARGDSSSSFYFDY		.	159	.	9	.	.	45	.	.	.	236	.	64	.	72	
ARGDTSSSFFDY		.	.	.	.	.	.	.	.	94	.	.	.	35	.	.	
TRGDSSSSFFDY		.	.	.	.	.	.	.	.	.	.	.	.	4	.	.	
sum	0	159	0	9	0	0	45	0	94	0	236	0	64	69	72	0	

Another signature identified in the DNA immunized animals (representative AREYSSAWYSYFDY) was identical to a signature previously found in the bone marrow and spleen of OmniRats immunized with a recombinant modified Ankara virus encoding for the MV H and F protein (Chapter 4). Although this signature did not correspond to MV-specific clones, it can be assumed these are also specific for a conserved epitope on the MV H protein.

#### **5.4.4 Strain specific signatures for phylogenetically distant MV strains**

To assess whether the H protein of two different strains of the MV can differentially model the IGH repertoire of immunized animals, we analyzed only CDR3 clusters composed by sequences belonging to at least half of the animals within the same vaccination group. In the MVH-A group, we obtained 3 80%-homology and 4 85%-homology clusters (Table 6). To control whether these CDR3 are clade A specific, we screened them over IG datasets from OmniRats that were immunized with the same MV vaccine strain. CDR3 closely related to these in one of the selected clusters (representative sequence ARIIVGATEAFDI) were found in the lymph nodes of OmniRats immunized with an inactivated MV as well as in the hybridoma library derived from these animals. We also found similar sequences to this signature in the spleen and the bone marrow of OmniRats immunized with a recombinant Modified Vaccinia Ankara virus encoding the MV H and F glycoproteins. Similarly, there were 4 80%-homology and 2 85%-homology clusters for the MVH-D vaccination group (Table 7). No homologous CDR3 sequences were found in the IG datasets from OmniRats immunized with MV clade A antigens.

**Table 6:** Putative signatures for the H protein of the MV genotype A

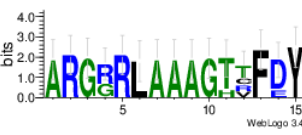
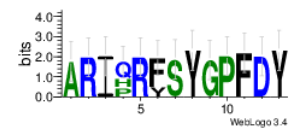
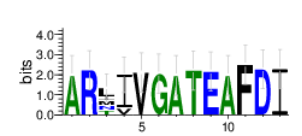

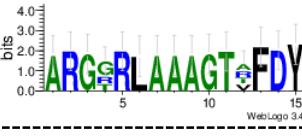



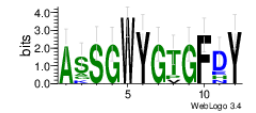

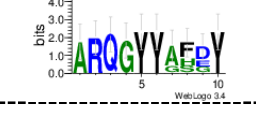
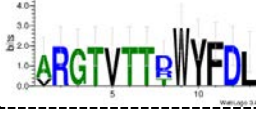

Cluster 80%										
Unique CDR3	Sequence Logo	MVH-A								Num. of rats
		#1	#2	#3	#4	#5	#6	#7	#8	
ARGGRLAAAGTRFDY		.	.	.	.	209	.	.	.	4
ARGGRLAAAGTVFDY		.	287	.	.	.	.	.	.	
ARGRRLAAAGHCFDY		.	.	.	.	.	.	18	.	
ARGRRLAAAGTTFDY		.	.	.	14	.	.	.	.	
ARGRRLAAAGTTFEY		.	.	.	.	.	.	107	.	
sum		0	287	0	14	209	0	125	0	
ARIHRFSYGPFDY		.	.	.	.	.	.	16	.	4
ARIPRFSYGPFDY		.	.	15	.	.	.	20	13	
ARIQRFSYGPFDY		91	.	.	.	.	.	.	.	
ARIQRYSYGPFDY		1.7	.	.	.	.	.	.	.	
sum		198	0	15	0	0	0	36	13	
ARIIVGATEAFDI		.	.	.	.	.	.	.	74	4
ARLIVGATEAFDI		.	.	.	13	.	.	.	.	
ARLVVGATEAFDI		.	.	.	28	.	.	.	.	
ARMIVGATEAFDI		.	.	.	.	.	99	.	.	
ARNIVGATEAFDI		.	.	.	.	.	.	.	89	
ARTIVGATEAFDI		.	.	36	.	.	.	.	.	
sum		0	0	36	41	0	99	0	163	
Cluster 85%										
Unique CDR3	Sequence Logo	MVH-A								Num. of rats
		#1	#2	#3	#4	#5	#6	#7	#8	
ARDITIVRGVILNWFDP		.	.	.	.	.	.	.	14	4
ARDITLVRGVILNWFDP		.	.	.	.	.	.	.	11	
ARDITMVRGVIIINWFDP		.	.	3	.	.	.	.	.	
ARDITVVRGVIIENWFDP		.	.	.	12	.	.	.	.	
ARDVTIVRGVIIINWFDP		.	.	.	.	.	45	.	.	
sum		0	0	3	12	0	45	0	26	
ARGGRLAAAGTRFDY		.	.	.	.	209	.	.	.	4
ARGGRLAAAGTVFDY		.	287	.	.	.	.	.	.	
ARGRRLAAAGTTFDY		.	.	.	14	.	.	.	.	
ARGTRLAAAGTAFDY		.	.	.	.	.	.	115	.	
sum		0	287	0	14	209	0	115	0	
ARIHRFSYGPFDY		.	.	.	.	.	.	16	.	4
ARIPRFSYGPFDY		.	.	15	.	.	.	20	13	
ARIQRFSYGPFDY		91	.	.	.	.	.	.	.	
sum		91	0	15	0	0	0	36	13	
ARIIVGATEAFDI		.	.	.	.	.	.	.	74	4
ARLIVGATEAFDI		.	.	.	13	.	.	.	.	
ARMIVGATEAFDI		.	.	.	.	.	99	.	.	
ARNIVGATEAFDI		.	.	.	.	.	.	.	89	
ARTIVGATEAFDI		.	.	36	.	.	.	.	.	
sum		0	0	36	13	0	99	0	163	

Table 7: Putative signatures for the H protein of the MV genotype D11

Cluster 80%										
Unique CDR3	Sequence Logo	MVH-D11								Num. of rats
		#9	#10	#11	#12	#13	#14	#15	#16	
AKDRAAAATRYFDY		.	.	.	29	.	.	.	.	4
AKDRAAAGTRYFDN		.	.	.	.	57	.	.	.	
AKDRAAAGTRYLDY		.	.	.	.	28	.	.	.	
AKDRAAAITRYFDY		.	246	.	.	319	.	.	.	
AKDRAAASTRYFDY		.	.	.	.	77	.	.	.	
AKDRAAATIRYFDY		.	.	.	.	55	.	.	.	
AKDRAAATTRYFDN		.	.	.	19	.	.	.	.	
AKDRAAATTRYFDY		.	.	.	.	433	.	.	.	
AKDRAAAVIRYFDY		.	.	.	.	6	.	.	.	
AKDRAAAVTRYFDY		.	.	13	297	1116	.	.	.	
AKDRAAAVVRYFDY		.	.	.	.	64	.	.	.	
AKDRAATSTRYFDY		.	.	.	.	30	.	.	.	
AKDRAPAVTRYFDY		.	.	.	.	26	.	.	.	
AKDRTAAVTRYFDY	.	.	.	.	1	.	.	.		
sum		0	246	13	344	2212	0	0	0	
AASGWYGTGFYD		.	.	.	.	19	.	.	4	
AGSGWYGTGFYD		.	.	.	.	93	.	.		
AKSGWYGPGFYD		.	.	179	.	.	.	.		
ASSGWYGTGFYD		.	.	.	.	118	.	.		
ASSGWYGTGFYD		.	.	.	.	9	.	.		
ASSGWYGTGFHY		.	.	.	.	50	.	.		
ASSGWYGTGFNY		.	.	.	.	8	.	.		
ASSGWYGVGFYD		.	.	.	.	.	41	6		
sum		0	0	179	0	296	41	0	6	
AKAIAAAGFDY		.	.	.	.	17	.	.	4	
AKALAAAGFDY		.	.	.	.	.	1	.		
AKARAVAGFDY		.	.	.	.	.	.	52		
ASGIAAAGFDY		.	.	128	.	.	.	.		
sum		0	0	128	0	17	1	52	0	
ARQGYAFYD		.	.	.	.	132	.	.	4	
ARQGYAHYD		.	.	.	.	.	18	.		
ARQGYASDY		.	.	.	.	.	.	9		
ARQGYGFGY		62	.	.	.	.	.	.		
sum		62	0	0	0	132	18	9	0	
Cluster 85%										
Unique CDR3	Sequence Logo	MVH-D11								Num. of rats
		#9	#10	#11	#12	#13	#14	#15	#16	
ARGTVTTDWFYDL		.	.	416	.	9	9	.	.	5
ARGTVTTRWFYDL		.	.	.	.	.	.	11		
ARGTVTTTWFYDL		.	.	.	.	.	.	80	.	
VRGTVTTDWFYDL		.	14	.	.	.	.	.	.	
sum		0	14	416	0	9	9	91	0	
AKDRAAAATRYFDY		.	.	.	29	.	.	.	4	
AKDRAAAGTRYFDN		.	.	.	.	57	.	.		
AKDRAAAGTRYLDY		.	.	.	.	28	.	.		
AKDRAAAITRYFDY		.	246	.	.	319	.	.		
AKDRAAASTRYFDY		.	.	.	.	77	.	.		
AKDRAAATIRYFDY		.	.	.	.	55	.	.		
AKDRAAATTRYFDN		.	.	.	19	.	.	.		
AKDRAAATTRYFDY		.	.	.	.	433	.	.		
AKDRAAAVIRYFDY		.	.	.	.	6	.	.		
AKDRAAAVTRYFDY		.	.	13	297	1116	.	.		
AKDRAAAVVRYFDY		.	.	.	.	64	.	.		
AKDRAATSTRYFDY		.	.	.	.	30	.	.		
AKDRAPAVTRYFDY		.	.	.	.	26	.	.		
AKDRTAAVTRYFDY	.	.	.	.	1	.	.			
sum		0	246	13	344	2212	0	0	0	



## 5.5 Discussion

Measles virus (MV) infection remains one of the leading causes of death among young children worldwide despite the availability of an effective vaccine (239, 240). In Western countries the vaccine coverage is high (>95% of the population), controlling the endogenous circulation of wild-type MV strains. However, in the last decade Measles is undergoing resurgence. Vaccine coverage tend to decrease in the West and occasional outbreaks occur essentially due to imported MV strains from developing countries (258). MV is serologically monotypic and an interesting tool in MV eradication would be to dispose of markers to tract back the different strains of the virus present in any given population. Next-generation sequencing (NGS) technologies offer new tools for surveying infectious diseases (191, 192). The vast number of sequences obtained in a single experiments allow to have a broad picture of the immune response elicited by both vaccination and ongoing infections. Recent studies in human subjects demonstrated the existence of a public IG repertoire from which antigen-specific signatures can be identify using NGS (192, 194, 202). The CDR3 identified from public repertoire can potentially be used as markers of exposure to certain antigens, opening up new opportunities for the sequence-based diagnosis or monitoring of particular diseases (202).

However, ethical, experimental and technical limitations makes difficult the identification of IG CDR3 signatures in patients. To overcome the obstacles of investigating human subjects, here we immunized transgenic rats carrying a chimeric human/rat IGH locus, where 22 human IGHV, all IGHD and all IGJ genes linked to rat IGHC regions, together with the complete IGK and IGL loci (235). We used a new technique of nanospheres-DNA immunization because it allows to express the antigen *in vivo* in its natural conformation and with appropriate post-translational modifications (365-368). We injected plasmids coding for the MV H protein of 2 phylogenetically distant MV strains: the MV Edmonston Vaccine strain (MVH-A) and the wild type genotype D11 (MVH-D11) (258, 377). The H protein of genotype D11 is the most distant from the MVH-A on the amino acid level, and differences are principally located in the outer surface of the protein (Fig. 1). Thirty-two OmniRats with a human B cell IG repertoire were separated into 4 groups of 8 rats each. The first group of rats was immunized with MVH-A, and the second group with MVH-D11. Control animals received empty DNA vectors (Mock) or were left untouched (NEG). All immunized rats mounted a MV-specific immune response of IgG antibodies with fully human idiotypes. Over 15 million of IG transcripts were obtained from the bone marrows using the Ion Torrent PGM platform. We aimed to investigate whether these two phylogenetically distant strains were able to induce two distinct convergent responses.

The immune response for the D11 strain was more homogeneous and on average stronger than the one observed in OmniRats immunized with the MVH-A. These observations suggest that the immunogenicity of the H protein of the vaccine strains is weaker and less predictable than the wild type H protein. Similarly, the MV-specific antibody levels induced in humans by vaccination are lower than those induced by wildtype infections (378). Some of the factors that confer the reported lower immunogenicity of the vaccine MV strain as compared to wild type strains may be encoded on the haemagglutinin gene (283-285, 379).

CDR3 AA sequences were clustered based on two different levels of amino acid identity. Only a small number of clusters were composed of sequences originating from at least half of the animals immunized against the MV H protein. However, we found VH CDR3 highly similar to three previously described MV-specific hybridoma clones in the IG repertoire of two-thirds of the MV immunized rats, independently of the H strain (Dubois *et al.*, in press). This observation further confirms that a phenomenon of convergent evolution lead to the expression of antigen-specific antibodies in these rats, as we showed for protein-based immunizations (Chapter 4) These VH CDR3 are signatures for the MV H protein and we conclude that the associated immunoglobulins target an epitope conserved among MV strains. All MV strains confer cross-protection and it is to date not possible to retrospectively differentiate infected from vaccinated individuals. However there is a certain variation in antigenic properties of the different MV strains, suggesting that the IG repertoire is differently modelled by diverse strains (283, 287, 288). We were able to identify few strain specific signature for both clad A and clade D11. One of the signature that emerge from our pipeline was found in OmniRats immunized with the H protein of MVH-A. Closely related CDR3 were found in the spleen and the bone marrow of OmniRats immunized with a Modified Vaccinia Ankara virus encoding the MV H and F glycoproteins. Similar CDR3 were also found in the lymph nodes of OmniRats immunized with an inactivated MV and in the hybridoma library derived from these animals. This suggests that the signature is actually clade A specific. On the other hand, no signature identified for the D11 strain was found in those datasets, suggesting that the CDR3 we isolated are possibly recognizing epitopes unique to the D11 strain.

The magnitude of antigen-specific public VH CDR3 repertoires generated by convergent recombination is still uncertain and assume to be relatively limited (155, 170). Our data obtained in OmniRats suggest that different animals can converge towards the expression of antibodies with similar CDR3 that target conserved epitopes of Measles H protein. At the same time, the IG repertoire can be also differentially modelled by phylogenetically distant MV strains. MV is divided into 8 clades which are further sub-

divided into 24 genotypes (258). Currently, identifying a MV strain is only possible in samples taken during the acute phase of the disease. It is therefore not possible to retrospectively determine the MV strain that has caused immunity or to differentiate infected from vaccinated individuals. Screening for MV-specific signatures over the human dataset could be useful to monitor vaccine coverage or to detect asymptomatic circulation of wild type strains in highly immunized populations. If signatures were available for all MV genotypes, diagnosis of an ongoing infection could be achieved with no direct isolation of the virus and genotyping.

The IG repertoire is continuously modeled by antigen challenges and records all pathogen exposures of the individual. Decrypting the signatures leave by specific pathogens would open the doors to an approach that could be used to retrospectively reconstruct the immune history of an individual.

## 5.6 Acknowledgements

We are grateful to R. Buelow from Open Monoclonal Technology Inc. (Palo Alto, CA, USA) for providing the OmniRats. We thank B. Pitard and G. Besin from In-Cell-Art (Nantes, France) for performing the DNA immunization procedures, and to S. Farinelle for her excellent technical assistance.

# **Chapter 6. General discussion**

B cells are able to recognize an almost infinite number of antigens, and confer the organism an efficient protection against pathogens. This ability relies on a very diverse repertoire of immunoglobulins (IG) with different specificities that are generated during B cell ontogeny by the stepwise rearrangement of genes. The IG are theoretically able to recognize any portion on the surface of an antigen, however not all epitopes are equally immunogenic and the immune response is preferentially targeted towards certain regions of the antigen, i.e. immunodominant epitopes (379, 380). The structure of these epitopes only allows a few different IG molecules to effectively bind and the best binders are preferentially selected.

Previous studies in our laboratory investigated the immunogenicity of the haemagglutinin (H) protein of the measles virus, finding that mainly two distinct regions on the MV H protein harbor immunodominant epitopes that are relatively conserved between MV genotypes (379). As only a limited fraction of the complete IG repertoire have the ability to recognize immunodominant epitopes, our working hypothesis was that the IG response would converge in different individuals and that antigen-specific footprints can be identified from the IG repertoire using NGS technologies. The identification and the creation of a dataset of MV-specific signatures with fully human features was the final purpose of this thesis. A unique model of rats expressing human IG loci were immunized using different types of vaccine (adjuvanted MV antigens, MVA vector, and DNA vaccination) and different immunization routes (footpads, IP, IM). We decreased the variety of MV antigens used for immunization in a stepwise manner (whole sonicated MV, recombinant MVA expressing H and F, and H only via DNA vaccination).

In Chapter 3, we first generated a set of IG sequences of known MV-specificity (i.e. MV-specific IGH signatures) to guide us in our subsequent experiments. For this purpose, MV-specific hybridoma clones were derived from the regional lymph nodes of rats immunized with adjuvanted MV antigens. A total of 293 MV-specific hybridoma clones characterized by 23 different VH CDR3 amino acid sequences and V-(D)-J gene usage were produced. Screening the B cells repertoire of different lymphoid organs for the sequences of the MV clones, we showed that the hybridoma generation is biased towards rare antigen-specific B cells in the lymph nodes that become enriched by orders of magnitude in the hybridoma library. On the other hand, the vast majority of the *in vivo* expanded populations were lost.

In Chapter 4, we immunized thirty-two OmniRats against various antigens to determine if antigen-specific IG signatures can be identify from their IG repertoire with no previous enrichment in antigen-specific B cells. We reduced the investigated number of MV antigens to the transmembrane H and F

glycoproteins using a MVA vector. The immunized rats mounted a convergent IG response and putative IG CDR3 signatures for the MV H and F proteins were identified from their public CDR3 repertoire. The most prominent signature in the spleen and bone marrow IG datasets were similar to the VH CDR3 of hybridoma clones we generated in Chapter 3, supporting that these sequences are antigen-specific. This validates our hypothesis that the MV leave specific signatures in the IG repertoire.

In Chapter 5, OmniRats were immunized against the H protein of two genetically distant strains of the MV. Analysis of the bone marrow IG transcripts revealed that same phenomenon of convergent evolution occurred in these rats. VH CDR3 with a high level of sequence identity to the same MV-specific hybridomas were found in the bone marrow samples of the MV H immunized rats, independently of the H strain. This indicates that these VH CDR3 are strong markers for the MV H protein, targeting a conserved immunodominant epitope on the MV H glycoprotein of both a vaccine clade A and a wild type D11 strains.

Here, we first discuss the advantages and the drawbacks of the methodologies used throughout the thesis. A set of putative fully-human VH CDR3 signatures has been created for MV antigens as well as other bacterial or chemically-defined antigens. These signatures can potentially be used as markers for particular antigens and for the sequence-based monitoring of infectious diseases or vaccination campaigns.

## **6.1 Technical issues of NGS technologies**

Before the advent of next-generation sequencing technologies, Sanger sequencing has been the method of reference to gain insight into the IG repertoire at the sequence level. The studies using this method are however limited to a relatively small set of sorted single cells and/or immortalized B cells (127). Beside the production of clinically-relevant antibodies, B cell cloning techniques followed by Sanger sequencing enabled the study of the features of high-affinity neutralizing antibodies or antibodies associated with autoimmune diseases and helped to understand the mechanisms leading to their production (154, 333, 381-390).

Next-generation sequencing (NGS) offers now new possibilities to investigate the IG diversity and dynamics at unprecedented depth. IG repertoire analysis using NGS technologies is an extremely challenging field of research. In order to make biological sense out of the large amount of information

generated, appropriate and innovative data mining and visualization tools as well as extensive bioinformatics processing are required.

### **Sequence errors and bioinformatics analyses of IG reads**

The result of a NGS run consists in various and often very high numbers of raw nucleotide sequences. After some quality controls inherent to the NGS platform, these sequences are aligned to reference germline sequences for annotation of the different genes. Various software, for instance the IMGT/HighV-QUEST or IgBlast, can perform this annotation and assign the closest germline gene based on nucleotide identity to reference sequences. A nucleotide divergence in an IG sequence in comparison to the germline sequence can either reflect a biological variation (e.g. somatic hypermutation) or be a technical artifact i.e. an error introduced by the observation technique. The accuracy of the nucleotide sequences obtained by NGS is not only important to ensure an accurate identification of the germline genes in affinity matured IG, but also to properly locate and define the somatic hypermutation. Therefore, like for any biological technique, it is important to understand the potential source of sequence errors in order to implement measures that limit or correct them. Not all research questions are however similarly affected by sequence errors. For instance, sequencing errors are less problematic if the purpose of the study is focused on the CDR3 or on V-(D)-J genes usage, as in this thesis. However, a detailed analysis of the somatic hypermutation process requires higher sequence accuracy and more thorough corrective measures. There exist two sources of sequence errors: those introduced during the preparation of the sequencing library and those inherent to the NGS platform (85, 154).

First, errors can arise at the sample preparation stage during the reverse transcription and/or the PCR amplification. The sequence variability between the different IG genes combined to the length restriction imposed by the NGS platform often requires performing a so-called multiplex PCR to produce IG amplicons. In these PCRs, multiple V primers, each specific for a V gene subgroup, are combined to usually one consensus J primer or to one or several C primers specific to a particular isotype (only if mRNA is used as starting material). V gene primers usually anneal to the conserved regions of IG: leader V region, FR1, FR2 or FR3. In this thesis, we used a combination of five different forward primers binding specifically to the FR2 of all human IGHV genes present on the OmniRat's genome. A drawback of this approach is that different V primers may differ in their binding efficiencies, resulting in the over-representation of certain V genes at the expense of others. The preferential amplification of a V gene, or a certain V-(D)-J rearrangements, may be falsely interpreted as a clonal expansion. Using a 5'RACE PCR approach when starting from mRNA circumvents the need of V gene-specific primer sets, but results in

longer IG amplicons that may be above the sequencing length range of the NGS platform (131, 391). A recently reported unbiased quantitative method to assess the IG repertoire relied on the emulsion PCR amplification of individual IG sequences (206). Another alternative consists in repetitively sequencing the same sample using different primer sets and comparing the results between each other (151, 158). This approach substantially increases the experimental costs and computational power required for the analysis. A more conservative approach to counter the potential amplification bias used in this thesis consists in collapsing the reads to non-redundant nucleotide sequences.

Nucleotide misincorporations during PCR cannot be distinguished from base-calling errors introduced during sequencing. Another mistake that can occur during PCR amplification is template switching. This results in the joining of fragments encoded by different DNA templates and in the generation of “sequence chimeras”. Generally, annotation programs cannot unambiguously assign a germline V gene to these sequence chimeras or provide artificially high rates of somatic hypermutation. Here, IG sequences with ambiguously assigned IGHV and IGHJ genes have been excluded from the analysis.

For sequencing the IG repertoire, it is mandatory to use NGS platforms that can provide long reads. The longer the reads, the more information can be obtained from the immunoglobulin sequence. Because of the high IG sequence variability, it is indeed impossible to *in silico* assemble short reads into a complete IG sequence. The leading platforms for the study of the IG repertoire are the Roche 454, Illumina and Ion Torrent, each of them being able to generate reads sufficiently long to cover a large part of the rearranged IG gene. While the errors generated by Illumina sequencing are principally nucleotide substitution errors, the 454 and the Ion Torrent technologies are particularly prone to insertion/deletion errors (indels). Particular computational approaches have been developed to account for these errors (200, 392). To determine whether our cleaning and annotation pipeline effectively removed most of the indels introduced by sequencing, we analyzed the IgG datasets of two mice hybridoma clones. After the indel correction by IMGT/HighV-QUEST, annotation with IgBlast, and considering only the productive sequences, we were able to generate IGH datasets with no detectable indels by IMGT (140).

Recently, a method based on DNA barcoding enabled to achieve high sequencing accuracies (189). In our laboratory, sequencing libraries are now processed using a modified protocol of the method. This method has been used to prepare the sequencing libraries presented in Chapter 5. The approach relies on the PCR amplification of IG transcripts with primers extended by a short string of random nucleotides. The number of possible different combinations of nucleotides within these strings exceeding the number



of target sequences in the sample, each mRNA molecule originally present is statistically uniquely labeled with an identifier sequence (“UID”). Sequences with the same UID can then be combined and sequencing errors corrected by building consensus sequences during data processing. This approach is nevertheless at the expense of sequencing depth and coverage. This approach also enables to take the PCR bias (e.g. by over-amplification with certain V primers) into consideration.

The development of innovative bioinformatics tools to analyze large IG sequence datasets is a major challenge and is at the core of any IG repertoire study (84, 85, 154). Many groups have developed their own tools and pipelines to process their data, from the cleaning of the reads to the visualization of the results (392, 393). The lack of standardization in the analytic techniques between the different laboratories, as well as the absence of central databases or shared computational resources, are real obstacles to the field. Meta-analyses of data generated in different laboratories with different platforms and cross-validation of results obtained within one laboratory remain a challenging task.

### **Genomic DNA versus mRNA**

The IG repertoire can be analyzed starting either from genomic DNA (gDNA) or messenger RNA (mRNA). Clonal expansion is more easily estimated by using gDNA because sequence reads are proportional to the number of gDNA templates in the sample i.e. the number of individual B cells. On the other hand, the use of mRNA tends to bias the view towards antigen-producing B cells. As one cell has only one rearranged IG gene sequence, the gDNA templates are in low abundance, while multiple copies of the same mRNA molecule are available and makes easier the amplification of all rearranged receptor genes from the sample. Although gDNA avoids the potential errors introduced by the reverse-transcription step, it requires the use of multiplex V primer sets and is therefore subject to potential amplification biases. Moreover, amplification of productive IG rearrangements from gDNA will compete with the abortive or non-productive IG rearrangements on the other chromosome, resulting in a mixture of productive and undesired non-productive amplicons. Also, because the IGHC is post-transcriptionally linked to the IGH V-(D)-J region, it is not possible to study a particular isotype when starting from gDNA from a bulk of B cells. To do so, prior enrichment (e.g. by FACS or MACS) in B cells producing antibodies of a certain isotype is required. Conversely, mRNA reverse-transcription or PCR amplification with IGHC-specific primers enable to distinguish between the different B cell isotypes. Therefore, we prepared libraries starting from mRNA in order to study the IG repertoire of class-switched B cells.

### Sample source

IG repertoire studies in humans are essentially limited to the only readily accessible source of B cells: peripheral blood. However, only a limited proportion of the total B cells of an individual circulates (about 2% of the total number of B cells), hence this sample provides only a narrow view of the humoral immune response (154, 203). It is also difficult to determine the proper time-point(s) when the antigen-specific B cells of interest are sufficiently expanded and plasma cells are only transiently circulating in the blood. Therefore, it is crucial to ensure a sufficient sampling depth (i.e. to sample enough blood, and thus B cells) in order to cover enough of the IG repertoire to answer a particular biological question. Various B cell subsets can eventually be sorted using flow cytometry based on phenotypic features or using fluorescently-labeled antigen before preparation of NGS libraries (180, 394). The use of an animal model, allow to have direct access to the primary (the bone marrow containing the plasma cells) and secondary (spleen and lymph nodes where clonal expansion occurs) lymphoid organs, and therefor to sample a broader diversity of the IG repertoire. In humans, tonsils and other surgical waste tissues (e.g. spleens following splenectomy) may constitute other available sources of B cells but their availability is unpredictable. Furthermore, these samples essentially limit researchers to retrospective studies. Recently, human biopsies stored in paraffin have successfully been used to study the IG repertoire using NGS (182). Although largely neglected so far, paraffin-stored lymphoid tissues represent another source of samples to retrospectively study the IG repertoire in different pathological setting.

## **6.2 Transgenic IG humanized animals: innovative organisms for the study of the B cell IG repertoire**

The analysis of the human IG repertoire to provide evidence of convergent evolution in individuals exposed to the same antigens and identify antigen-specific signatures directly in human subjects can be particularly challenging. Human studies are largely restricted to peripheral blood where only a small amount of the IG repertoire diversity can be sampled, and it is difficult to predict when the B cells of interest are sufficiently expanded in the blood to be detected. Large numbers of participants enrolled in extensive longitudinal studies are required to detect convergent IG responses and identify reliable IG signatures. Human studies are also limited to certain antigens. Here we propose an innovative approach

that allows a more flexible identification of human antigen-specific VH CDR3 signatures, notably for the different MV strains. This approach relies on transgenic rats carrying human IG loci and expressing a diversified repertoire of chimeric human/rats IG with fully human idiotypes (OmniRat).

The basic mechanisms of V-(D)-J rearrangement, somatic hypermutation and class-switch recombination are conserved across species and explain the successful production of antigen-specific, class-switched and hypermutated antibodies by transgenic IG humanized animals (350). The uniqueness of the OmniRats comes from the linkage of the human IGH loci to germline-configured rat C regions. The interaction of the transgenic membrane IG with the endogenous rodent cellular signaling machinery is therefore optimized and allow these rats to produce a diversified repertoire of chimeric human/rat IG with level similar to wild type animals (235).

In this thesis, we studied the post-vaccinal IgG repertoire in transgenic rodents with human IG loci in order to have access to different lymphoid organs and to study more easily the immune response to the measles virus compared to human subjects. To investigate the IG response to the MV and determine whether MV strain-specific IG signatures can be found in the IG repertoire, OmniRats were immunized against various targets including MV antigens. To our knowledge we performed the first in-depth characterization of the modeling of the IG repertoire in transgenic IG animals. We showed that these rats mount a specific IgG response upon immunization with different immune challenges: viral, bacterial and hapten antigens. From their regional lymph nodes, we derived hybridoma clones producing MV-specific antibodies. Our work further highlights the potential use of these animals to produce human mAbs against a variety of antigens.

This thesis successfully demonstrated that upon exposure to the same antigens, OmniRats mount a convergent IgG response characterized by clusters of highly similar CDR3 amino acid sequences. In this animal model, this process has been shown for several different antigens and sets of VH CDR3 signatures with fully human features have been identified. Furthermore, we observed that OmniRats were able to produce shared CDR3 against conserved epitopes of the MV H protein upon immunization with three different immunization processes, namely the footpad injection of inactivated Measles virus, intraperitoneal injection of a recombinant viral vector encoding for the MV H and F proteins and intramuscular injection of a DNA vaccine of the MV H protein. These results suggest that OmniRats are a good animal model to identify convergent IGH signatures with fully human features, regardless the immunization protocol.

While the production of human antibodies with these animals is efficient, the OmniRats do not have the full IGHV gene diversity of humans. The size of the human IG loci constitutes indeed a serious challenge to the generation of transgenic IG animals (221). The 22 IGHV genes of these rats were selected because they are the more frequently used during the *in vivo* immune response in humans (Roland Buelow, personal communication). Nevertheless, it is possible that humans would preferentially use the IGHV genes lacking in these rats and signatures obtained in transgenic animals need to be validated in humans. However, our study showed that OmniRats are a valid and homogenous system to investigate the public response as we observed an extensive convergence across OmniRats. We cannot infer that in humans this phenomenon is equally strong across individuals because the immune response in patients is more complex and diverse. This is the reason why OmniRats are a valid tool to easily identify signatures against a given antigen; signatures that can be used for epidemiological studies or in diagnosis if the selected CDR3 sequences also occur in humans.

## 6.3 Insight into the hybridoma fusion process

In order to create a dataset of MV specific IGH sequences, OmniRats were immunized with the vaccine MV strain (clade A) and hybridoma clones were derived from the regional lymph nodes. Hybridoma library preparation was performed following a standard protocol by a company leader in the production on monoclonal antibodies for therapy and research. We analyzed the bias introduced on the composition of the IG repertoire during a standardized generation of hybridoma for clinical or research applications. To our knowledge, this is the first high-resolution analysis of the molecular dynamic of the IG repertoire during immunization and PEG-fusion using high-throughput sequencing of IG heavy chain transcripts.

Even today little is known about the cells that are susceptible to fuse and to what extent the fusion bias influences the detection of antigen-specific B cells. We compared the IG repertoire of the lymph nodes and of the hybridoma library and we showed that the rare antigen-specific cells are favored during fusion at the expense of clonally expanded B cell populations in the regional lymph nodes. Plasma and/or blast cells have been suggested to be the best fusion partners for myeloma cells (342, 400-402). Our observations conversely suggests that B cells advanced in the proliferation and differentiation stage lose their ability to fuse, and that B cells in an earlier developmental stage (or less advanced in the differentiation) may preferentially fuse to myeloma cells. We assume that the fusion susceptible B cells originate from ongoing germinal center reactions; they have received a proliferation signal but have not

yet expanded. Potential candidates may be dark zone germinal center B cells with a low IG mRNA expression (395).

Because of this bias, the hybridoma library did not reflect the composition of the lymph node, and clones only provide a narrow and skewed view of the IG response. The expanded B cells are lost during the immortalization process, and with them the antigen-specific antibodies they are likely to encode. Since the IG repertoire in the hybridoma library does not reflect the parental repertoire of the lymphoid organ, the design and construction of antibodies based on frequent IG sequences in the lymph nodes would result in a totally different selection of antibodies. Coupling traditional hybridoma generation to the NGS analysis of lymphoid organs used for fusion and the cloning of the expanded IG sequences into expression vectors should help to recover potentially interesting antigen-specific antibodies (210, 211). NGS analysis of combinatorial libraries after one or two rounds of screening have been used to recover clones expressing high-affinity antibodies that would have been missed by the classical approach (344).

Current knowledge of the B cell IG repertoire and the antibodies involved in infectious diseases or other malignant conditions has largely relied on hybridoma production and Sanger sequencing. Our data suggests that the standard PEG-fusion which is routinely used in biotech industry is most likely bias towards clonotypes under-represented in the *in vivo* immune response, thus painting an incomplete picture of the IG repertoire.

## 6.4 Immune history using NGS-derived signatures

The IG repertoire is continuously modelled by antigen challenges and previous antigen-exposures can impact the repertoire composition for long periods of time. Decrypting specific signatures left by antigens on the IG repertoire offers the opportunity to trace back the immunological history of individuals.

Some proteomic tools can be applied to the study of the repertoire of serum IG. Antibody reactivity profiles, quantified by densitometry (*PANAMA-blot technology*), can be compared between individuals to identify specific patterns of a pathology or clinical status. It has been used to characterize immunological changes in autoimmune diseases and chronic parasitic diseases (396, 397). Similarly, the antigen microarray chip technology also uses reactivity profiles of antibodies in sera to determine potential antigen-specific signatures (398, 399). Using this technology, it was possible to discriminate healthy individuals from diabetic patients and determine diabetes susceptibility or resistance patterns in mice (398, 399).

Next-generation sequencing (NGS) offers new possibilities to investigate the diversity and dynamics of the IG repertoire at very high depth. At the root of this project, we hypothesized that the immune response in different individuals immunized against the same antigens converges, and that shared IG sequences can be identified using NGS technologies.

In humans, it is thought that the IG repertoire is essentially private. While a large proportion of IG light chain CDR3 is shared between naive individuals (400), the overlap in VH CDR3 usage appears to be low (154). A study performed in homozygotic twins showed that although the V-(D)-J recombination frequencies are indistinguishable between twin pairs, the VH CDR3 repertoires display very little overlap (167). Authors concluded that even with a common genetic background, these individuals are likely to express different IG responses to common environmental exposure (167). Others have also shown that a very small fraction of the repertoire of unimmunized genetically different individual encodes for common VH CDR3 (169). Studies performed in immunocompromised mice engrafted with human CD34+ hematopoietic stem cells further support these observations. Authors showed that the human naive antibody repertoires of these animals display a similar preferential usage of some IG genes and IGHD-IGHJ pairings, demonstrating a genetic constraint imposed on the IG repertoire generation, but that the VH CDR3 repertoire was largely unique to the individual (401).

### **Tracking sequences of known antigen-specificity to identify public signatures**

One way to identify public antigen-specific sequences is to search for sequences with a high degree of homology to previously described ones (152, 197, 198, 200, 201). This approach was essentially implemented in the study of the humoral immune response to antigenically variable pathogens, such as HIV and influenza virus, in the search of broadly neutralizing antibodies (152, 197, 198, 200, 201). It was for instance shown that the IG repertoire of donors immunized with an influenza virus vaccine contained sequences similar to those of known antigen-specific cells (152, 190). By sequence comparison and lineage analysis, it was found that bnAbs of VRC01-class occur from a single B cell ancestor in multiple donors with HIV-1 infection (199, 402). These newly identified VRC01 VH sequences, when paired to VL sequences, were able to neutralize HIV-1 (199). Germline-precursors of VRC01 antibodies have also been observed in a cord blood IgM library (403).

Over the last decades, a very large number of hybridoma clones specific for a wide range of different antigens have been produced. From these hybridoma, the sequence encoding the antigen-specific antibody is known, sometimes with other functional information such as affinity or epitope-specificity. The “hybridoma repertoire” as a source of IG signature has been under-exploited. These antigen-specific

IG sequences could be advantageously combined to NGS investigations of the IG repertoire. It is however possible that many hybridomas derived from B cells belonging to the individual private repertoire, and hence do not encode for interesting antigen-specific signatures.

### **Convergent IG responses: a new source of markers for infectious diseases?**

Another approach, developed in this thesis, is to search within IG sequences datasets for sequences shared by individuals exposed to the same antigen challenge. By performing cross-comparisons between the IG sequences from different immunized donors or patients, shared IG features (VH CDR3, V-(D)-J rearrangements) have been identified.

In some lymphoproliferative conditions, identical public (stereotyped) CDR3 sequences can be found in different patients (364, 404–406). In this case, the IG sequence of malignant B cells can be used as biomarkers for diagnosis or the monitoring of the disease (e.g. detect minimal residual disease) in individual patients. NGS has thus here a direct clinical application.

Longitudinal studies in which multiple samples were taken from the same donors during the course of the immune response have recently highlighted public IG responses. Convergent VH CDR3 amino acid sequences were identified in patients recovering from acute dengue infection (202). The specificity for dengue virus antigens of these convergent VH CDR3 signatures has however not been investigated further. Interestingly, we observed that the identified “dengue-specific” signature (ARLDYYYYYGMDV, and all the one amino acid derivatives) is also overexpressed in patients in the acute phase of measles but not in the blood of healthy donors (unpublished observation). We assume that this particular sequence is rather vaccine-induced or specific for common epitopes between the dengue virus and MV. This observation highlights the need to include control groups composed of individuals immunized with unrelated antigens. Convergent IGH rearrangements were also identified among different people in response to vaccination and infection with the H1N1 2009 influenza strain (192). Analysis of the IG heavy chain repertoire of donors immunized with a conjugate vaccine containing *Haemophilus influenzae* type B (Hib) and group C meningococcal (MenC) polysaccharides showed that different individuals were using convergent VH CDR3 sequences identical or highly similar to known Hib-specific or TT-specific sequences (194).

The relative importance of the public IG response as compared to the private IG response is still unknown and is a key question to answer to determine the utility of convergent repertoire analysis. In this context, there is a need to develop strategies to determine the degree to which convergent sequences

are antigen-specific. Further longitudinal studies of the IG repertoire in human will likely contribute to identify novel antigen-specific signatures. However, convergent antigen-specific signatures in humans are difficult to demonstrate in a statistically significant manner due to the technical limitations that allow the study of only a limited fraction of the repertoire. This does not however mean that such a public response is nonexistent and that public signatures are not present in the IG repertoire.

### **Convergent IG signatures in transgenic humanized IG rats**

We demonstrated a strong public IG response within different groups of OmniRats characterized by convergent highly similar VH CDR3 amino acid sequences in the animals that received the same vaccination.

In chapter 4, we used the Ion Torrent PGM platform to analyze the IgG IGH repertoire in the spleen and bone marrow of rats immunized with a recombinant modified vaccinia Ankara (MVA) virus carrying the measles virus haemagglutinin and fusion proteins. In order to verify that a convergent response is a process normally occurring during the immune response, regardless of the antigen used for vaccination, we immunized a cohort of OmniRats with the bacterial tetanus toxoid (TT). TT has been described to trigger a diverse IG response in humans, with only few CDR3 sequences shared between individuals (193, 360). Recently, some authors however observed known (or putative) TT-specific CDR3 amino acid sequences shared in the blood of donors immunized with a TT-containing vaccine, demonstrating a public response (194, 214, 361, 362). Our observations in transgenic rats suggest that TT proteins also trigger convergent IG responses in our model. Comparison of the identified TT signatures to previously described IG sequences from human clones revealed that one of these signatures was highly similar to the VH CDR3 of previously described TT-specific hybridoma clone derived from the blood of a human immune donor (363). Investigation of the full length sequences revealed that these CDR3 were associated with the same IGHV and IGHJ genes in both the OmniRats and the human hybridoma. The unique sequences composing this validated signature contained common amino acids within the CDR3 region, while other positions were more likely to differ among individuals. Interestingly, those patterns were conserved between humans and OmniRats, suggesting that some key positions are crucial for an efficient binding to the antigen.

Several fully human VH CDR3 signatures for the MV H or F proteins have been identified in this study. In both the spleen and bone marrow of the rats immunized with a recombinant MVA expressing the MV H and F glycoproteins (MVA-MV H/F), the most prominent signature was similar the VH CDR3 of MV-specific hybridoma clones derived from the regional lymph nodes of OmniRats immunized



against whole MV antigens of the same genotype. Closely related CDR3 were retrieved in both lymphoid organs analyzed, indicating that those antibodies are relevant in the *in vivo* immune response, at least in our animal model. The same signature was also found in the bone marrow of rats immunized by DNA vaccination against the H protein of two phylogenetically distant MV strains (Chapter 5). We concluded that this CDR3 signature arises from IG targeting a conserved epitope on the MV H protein. Clusters of homologous VH CDR3 were also present only in the animals immunized with either the MV clade A or D11 genotype. The antibodies encoding for these CDR3 are likely directed against immunodominant epitopes differing between the two MV strains.

Although our IG signatures were strongly associated with the particular immunization, for most of them the antigen-specificity remains putative. Functional characterization of identified signatures can be achieved by cloning the most frequent VH sequence encoding for these signatures together with an IG light chain sequence into expression vectors. This would require to sequence IG light chain transcripts from the same samples and select the pairing light chain, based for instance on a statistical approach or eventually on public IG light chain CDR3 (159). For our MV-specific signature we used an indirect method to validate the specificity by comparing them to CDR3 of existent hybridoma clones of known-antigen specificity (192, 194).

In conclusion, this thesis demonstrated that MV triggers a public IG responses in transgenic rats expressing a large human IG repertoire and that MV strain-specific signatures can be identified by NGS. These fully human VH CDR3 signatures constitute interesting templates to screen over human datasets. If validated in human donors, these signatures can be used as biomarkers to differentiate MV vaccination from infection.

# **Chapter 7.**

# **Concluding remarks & future perspectives**

The analysis of the modulation of the IG repertoire using NGS technologies is a relatively new but very dynamic field of research. NGS analysis of the IG repertoire has found a wide range of applications in fundamental and clinical immunology. Notably, it contributes to a better understanding of B cell biology in particular pathological (e.g. infectious diseases, autoimmunity, lymphoproliferative disorders etc.) or physiological (e.g. vaccination, immunosenescence) conditions, with important repercussions on the research, diagnosis, prevention and treatment of diseases.

This thesis investigated the complexity of the IG repertoire in the context of MV vaccination to understand the dynamics of B cell selection in response to a specific antigen in an animal model with a human IG repertoire. We demonstrated that a process of convergence results in the production of antibodies with similar CDR3 sequences in different individuals.

We created a set of fully-human IGH signatures of known (hybridoma clones) and putative (public CDR3) antigen-specificity for the MV H protein, as well as other bacterial, viral and chemically-defined antigens. Our results also suggest that a wildtype and the vaccine strain of MV leave different footprints on the repertoire. The screening of the VH CDR3 signatures against human IG datasets from MV infected and vaccinated donors is ongoing. If validated in human subjects, these signatures can be used as biomarkers for MV infection or vaccination. Potentially, transgenic humanized animals can be used to identify IG signatures for a large number of antigens.

During the course of this PhD thesis a large number of human samples have been collected in Luxembourg and in Lao PDR. The IGH repertoires of PBMCs from donors in the acute phase of measles as well as from healthy volunteers before and after immunization with a MV containing vaccine have been assessed on a Roche 454 Junior deep sequencing platform. Furthermore, using our Ion Torrent deep sequencing platform, we sequenced the IgG, IgM and IgA repertoires of PBMCs from individuals that were initially negative for anti-MV IgG that were collected before and at different time-points after vaccination at the Luxembourg-Lao Friendship Laboratory in Vientiane. The analysis pipelines presented in this thesis are now being used on these datasets to identify expanded B cell populations and potential CDR3 signatures in humans. These experiments should determine the extent to which the results obtained in the OmniRat can be extrapolated to human.

# References

1. Janeway, C., K. P. Murphy, P. Travers, and M. Walport. 2008. *Janeway's immunobiology*. Garland Science.
2. Mallo, G. V., C. L. Kurz, C. Couillault, N. Pujol, S. Granjeaud, Y. Kohara, and J. J. Ewbank. 2002. Inducible antibacterial defense system in *C. elegans*. *Current biology: CB* 12: 1209-1214.
3. Basset, A., R. S. Khush, A. Braun, L. Gardan, F. Bocard, J. A. Hoffmann, and B. Lemaitre. 2000. The phytopathogenic bacteria *Erwinia carotovora* infects *Drosophila* and activates an immune response. *Proceedings of the National Academy of Sciences of the United States of America* 97: 3376-3381.
4. Muller, W. E., B. Blumbach, and I. M. Muller. 1999. Evolution of the innate and adaptive immune systems: relationships between potential immune molecules in the lowest metazoan phylum (Porifera) and those in vertebrates. *Transplantation* 68: 1215-1227.
5. Ronald, P. C., and B. Beutler. 2010. Plant and animal sensors of conserved microbial signatures. *Science* 330: 1061-1064.
6. Kennedy, A. D., and F. R. DeLeo. 2009. Neutrophil apoptosis and the resolution of infection. *Immunologic research* 43: 25-61.
7. Rolff, J. 2007. Why did the acquired immune system of vertebrates evolve? *Developmental and comparative immunology* 31: 476-482.
8. Rodriguez, R. M., A. Lopez-Vazquez, and C. Lopez-Larrea. 2012. Immune systems evolution. *Advances in experimental medicine and biology* 739: 237-251.
9. Danilova, N. 2012. The evolution of adaptive immunity. *Advances in experimental medicine and biology* 738: 218-235.
10. Blom, B., and H. Spits. 2006. Development of human lymphoid cells. *Annual review of immunology* 24: 287-320.
11. Schroeder, H. W., Jr., and L. Cavacini. 2010. Structure and function of immunoglobulins. *The Journal of allergy and clinical immunology* 125: S41-52.
12. Porter, R. R. 1959. The hydrolysis of rabbit  $\gamma$ -globulin and antibodies with crystalline papain. *The Biochemical journal* 73: 119-126.
13. Boes, M. 2000. Role of natural and immune IgM antibodies in immune responses. *Molecular immunology* 37: 1141-1149.
14. Davis, A. C., K. H. Roux, and M. J. Shulman. 1988. On the structure of polymeric IgM. *European journal of immunology* 18: 1001-1008.
15. Geisberger, R., M. Lamers, and G. Achatz. 2006. The riddle of the dual expression of IgM and IgD. *Immunology* 118: 429-437.
16. Papadea, C., and I. J. Check. 1989. Human immunoglobulin G and immunoglobulin G subclasses: biochemical, genetic, and clinical aspects. *Critical reviews in clinical laboratory sciences* 27: 27-58.
17. Macpherson, A. J., K. D. McCoy, F. E. Johansen, and P. Brandtzaeg. 2008. The immune geography of IgA induction and function. *Mucosal immunology* 1: 11-22.
18. Cerutti, A. 2008. The regulation of IgA class switching. *Nature reviews. Immunology* 8: 421-434.
19. Bennich, H. H., K. Ishizaka, S. G. Johansson, D. S. Rowe, D. R. Stanworth, and W. D. Terry. 1968. Immunoglobulin E: a new class of human immunoglobulin. *Immunology* 15: 323-324.
20. Corry, D. B., and F. Kheradmand. 1999. Induction and regulation of the IgE response. *Nature* 402: B18-23.
21. Burton, O. T., and H. C. Oettgen. 2011. Beyond immediate hypersensitivity: evolving roles for IgE antibodies in immune homeostasis and allergic diseases. *Immunological reviews* 242: 128-143.

22. Gounni, A. S., B. Lamkhioued, K. Ochiai, Y. Tanaka, E. Delaporte, A. Capron, J. P. Kinet, and M. Capron. 1994. High-affinity IgE receptor on eosinophils is involved in defence against parasites. *Nature* 367: 183-186.
23. Kabat, E. A., T. T. Wu, and H. Bilofsky. 1979. Evidence supporting somatic assembly of the DNA segments (minigenes), coding for the framework, and complementarity-determining segments of immunoglobulin variable regions. *The Journal of experimental medicine* 149: 1299-1313.
24. Chitarra, V., P. M. Alzari, G. A. Bentley, T. N. Bhat, J. L. Eisele, A. Houdusse, J. Lescar, H. Souchon, and R. J. Poljak. 1993. Three-dimensional structure of a heteroclitic antigen-antibody cross-reaction complex. *Proceedings of the National Academy of Sciences of the United States of America* 90: 7711-7715.
25. Xu, J. L., and M. M. Davis. 2000. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* 13: 37-45.
26. Davies, D. R., and G. H. Cohen. 1996. Interactions of protein antigens with antibodies. *Proceedings of the National Academy of Sciences of the United States of America* 93: 7-12.
27. Fanning, L. J., A. M. Connor, and G. E. Wu. 1996. Development of the immunoglobulin repertoire. *Clinical immunology and immunopathology* 79: 1-14.
28. Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature* 302: 575-581.
29. Early, P., H. Huang, M. Davis, K. Calame, and L. Hood. 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell* 19: 981-992.
30. Croce, C. M., M. Shander, J. Martinis, L. Cicurel, G. G. D'Ancona, T. W. Dolby, and H. Koprowski. 1979. Chromosomal location of the genes for human immunoglobulin heavy chains. *Proceedings of the National Academy of Sciences of the United States of America* 76: 3416-3419.
31. Alt, F. W., G. D. Yancopoulos, T. K. Blackwell, C. Wood, E. Thomas, M. Boss, R. Coffman, N. Rosenberg, S. Tonegawa, and D. Baltimore. 1984. Ordered rearrangement of immunoglobulin heavy chain variable region segments. *The EMBO journal* 3: 1209-1219.
32. Matsuda, F., E. K. Shin, Y. Hirabayashi, H. Nagaoka, M. C. Yoshida, S. Q. Zong, and T. Honjo. 1990. Organization of variable region segments of the human immunoglobulin heavy chain: duplication of the D5 cluster within the locus and interchromosomal translocation of variable region segments. *The EMBO journal* 9: 2501-2506.
33. McBride, O. W., J. Battey, G. F. Hollis, D. C. Swan, U. Siebenlist, and P. Leder. 1982. Localization of human variable and constant region immunoglobulin heavy chain genes on subtelomeric band q32 of chromosome 14. *Nucleic acids research* 10: 8155-8170.
34. McBride, O. W., P. A. Hieter, G. F. Hollis, D. Swan, M. C. Otey, and P. Leder. 1982. Chromosomal location of human kappa and lambda immunoglobulin light chain constant region genes. *The Journal of experimental medicine* 155: 1480-1490.
35. Fugmann, S. D., A. I. Lee, P. E. Shockett, I. J. Villey, and D. G. Schatz. 2000. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annual review of immunology* 18: 495-527.
36. Max, E. E., J. G. Seidman, and P. Leder. 1979. Sequences of five potential recombination sites encoded close to an immunoglobulin kappa constant region gene. *Proceedings of the National Academy of Sciences of the United States of America* 76: 3450-3454.
37. Akira, S., K. Okazaki, and H. Sakano. 1987. Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. *Science* 238: 1134-1138.
38. Ramsden, D. A., K. Baetz, and G. E. Wu. 1994. Conservation of sequence in recombination signal sequence spacers. *Nucleic acids research* 22: 1785-1796.
39. Lee, A. I., S. D. Fugmann, L. G. Cowell, L. M. Ptaszek, G. Kelsoe, and D. G. Schatz. 2003. A functional analysis of the spacer of V(D)J recombination signal sequences. *PLoS biology* 1: E1.

40. Oettinger, M. A., D. G. Schatz, C. Gorka, and D. Baltimore. 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* 248: 1517-1523.
41. Dudley, D. D., J. Chaudhuri, C. H. Bassing, and F. W. Alt. 2005. Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. *Advances in immunology* 86: 43-112.
42. van Gent, D. C., J. F. McBlane, D. A. Ramsden, M. J. Sadofsky, J. E. Hesse, and M. Gellert. 1996. Initiation of V(D)J recombinations in a cell-free system by RAG1 and RAG2 proteins. *Current topics in microbiology and immunology* 217: 1-10.
43. Lewis, S. M. 1994. The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. *Advances in immunology* 56: 27-150.
44. Ma, Y., U. Pannicke, K. Schwarz, and M. R. Lieber. 2002. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* 108: 781-794.
45. Critchlow, S. E., R. P. Bowater, and S. P. Jackson. 1997. Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Current biology: CB* 7: 588-598.
46. Buhl, A. M., D. Nemazee, J. C. Cambier, R. Rickert, and M. Hertz. 2000. B-cell antigen receptor competence regulates B-lymphocyte selection and survival. *Immunological reviews* 176: 141-153.
47. Daly, J., S. Licence, A. Nanou, G. Morgan, and I. L. Martensson. 2007. Transcription of productive and nonproductive VDJ-recombined alleles after IgH allelic exclusion. *The EMBO journal* 26: 4273-4282.
48. Bergman, Y. 1999. Allelic exclusion in B and T lymphopoiesis. *Seminars in immunology* 11: 319-328.
49. Wabl, M., and C. Steinberg. 1982. A theory of allelic and isotypic exclusion for immunoglobulin genes. *Proceedings of the National Academy of Sciences of the United States of America* 79: 6976-6978.
50. Nemazee, D. 2000. Receptor selection in B and T lymphocytes. *Annual review of immunology* 18: 19-51.
51. Meffre, E., and H. Wardemann. 2008. B-cell tolerance checkpoints in health and autoimmunity. *Current opinion in immunology* 20: 632-638.
52. Nemazee, D. 2000. Receptor editing in B cells. *Advances in immunology* 74: 89-126.
53. Hartley, S. B., M. P. Cooke, D. A. Fulcher, A. W. Harris, S. Cory, A. Basten, and C. C. Goodnow. 1993. Elimination of self-reactive B lymphocytes proceeds in two stages: arrested development and cell death. *Cell* 72: 325-335.
54. Hartley, S. B., J. Crosbie, R. Brink, A. B. Kantor, A. Basten, and C. C. Goodnow. 1991. Elimination from peripheral lymphoid tissues of self-reactive B lymphocytes recognizing membrane-bound antigens. *Nature* 353: 765-769.
55. Goodnow, C. C., J. Crosbie, S. Adelstein, T. B. Lavoie, S. J. Smith-Gill, R. A. Brink, H. Pritchard-Briscoe, J. S. Wotherspoon, R. H. Loblay, K. Raphael, and et al. 1988. Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice. *Nature* 334: 676-682.
56. Fu, Y. X., and D. D. Chaplin. 1999. Development and maturation of secondary lymphoid tissues. *Annual review of immunology* 17: 399-433.
57. Fagarasan, S., and T. Honjo. 2000. T-Independent immune response: new aspects of B cell biology. *Science* 290: 89-92.
58. Mond, J. J., Q. Vos, A. Lees, and C. M. Snapper. 1995. T cell independent antigens. *Current opinion in immunology* 7: 349-354.
59. Garside, P., E. Ingulli, R. R. Merica, J. G. Johnson, R. J. Noelle, and M. K. Jenkins. 1998. Visualization of specific B and T lymphocyte interactions in the lymph node. *Science* 281: 96-99.

60. Allen, C. D., T. Okada, and J. G. Cyster. 2007. Germinal-center organization and cellular dynamics. *Immunity* 27: 190-202.
61. Jacob, J., R. Kassir, and G. Kelsoe. 1991. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. I. The architecture and dynamics of responding cell populations. *The Journal of experimental medicine* 173: 1165-1175.
62. Liu, Y. J., G. D. Johnson, J. Gordon, and I. C. MacLennan. 1992. Germinal centres in T-cell-dependent antibody responses. *Immunology today* 13: 17-21.
63. MacLennan, I. C. 1994. Germinal centers. *Annual review of immunology* 12: 117-139.
64. Berek, C., A. Berger, and M. Apel. 1991. Maturation of the immune response in germinal centers. *Cell* 67: 1121-1129.
65. Jacob, J., G. Kelsoe, K. Rajewsky, and U. Weiss. 1991. Intraclonal generation of antibody mutants in germinal centres. *Nature* 354: 389-392.
66. Zhang, J., I. C. MacLennan, Y. J. Liu, and P. J. Lane. 1988. Is rapid proliferation in B centroblasts linked to somatic mutation in memory B cell clones? *Immunology letters* 18: 297-299.
67. Miga, A., S. Masters, M. Gonzalez, and R. J. Noelle. 2000. The role of CD40-CD154 interactions in the regulation of cell mediated immunity. *Immunological investigations* 29: 111-114.
68. Grewal, I. S., and R. A. Flavell. 1998. CD40 and CD154 in cell-mediated immunity. *Annual review of immunology* 16: 111-135.
69. Cattoretti, G., M. Buttner, R. Shakhovich, E. Kremmer, B. Alobeid, and G. Niedobitek. 2006. Nuclear and cytoplasmic AID in extrafollicular and germinal center B cells. *Blood* 107: 3967-3975.
70. Moldenhauer, G., S. W. Popov, B. Wotschke, S. Bruderlein, P. Riedl, N. Fissolo, R. Schirmbeck, O. Ritz, P. Moller, and F. Leithauser. 2006. AID expression identifies interfollicular large B cells as putative precursors of mature B-cell malignancies. *Blood* 107: 2470-2473.
71. Liu, Y. J., F. Malisan, O. de Bouteiller, C. Guret, S. Lebecque, J. Banchereau, F. C. Mills, E. E. Max, and H. Martinez-Valdez. 1996. Within germinal centers, isotype switching of immunoglobulin genes occurs after the onset of somatic mutation. *Immunity* 4: 241-250.
72. Martinez-Valdez, H., F. Malisan, O. de Bouteiller, C. Guret, J. Banchereau, and Y. J. Liu. 1995. Molecular evidence that in vivo isotype switching occurs within the germinal centers. *Annals of the New York Academy of Sciences* 764: 151-154.
73. Kim, S., M. Davis, E. Sinn, P. Patten, and L. Hood. 1981. Antibody diversity: somatic hypermutation of rearranged VH genes. *Cell* 27: 573-581.
74. Meyer, J., H. M. Jack, N. Ellis, and M. Wabl. 1986. High rate of somatic point mutation in vitro in and near the variable-region segment of an immunoglobulin heavy chain gene. *Proceedings of the National Academy of Sciences of the United States of America* 83: 6950-6953.
75. Lossos, I. S., R. Tibshirani, B. Narasimhan, and R. Levy. 2000. The inference of antigen selection on Ig genes. *Journal of immunology* 165: 5122-5126.
76. Tarlinton, D. M., A. Light, G. J. Nossal, and K. G. Smith. 1998. Affinity maturation of the primary response by V gene diversification. *Current topics in microbiology and immunology* 229: 71-83.
77. Tarlinton, D. M., and K. G. Smith. 2000. Dissecting affinity maturation: a model explaining selection of antibody-forming cells and memory B cells in the germinal centre. *Immunology today* 21: 436-441.
78. Kelsoe, G. 1996. Life and death in germinal centers (redux). *Immunity* 4: 107-111.
79. Kepler, T. B., and A. S. Perelson. 1993. Cyclic re-entry of germinal center B cells and the efficiency of affinity maturation. *Immunology today* 14: 412-415.
80. Stavnezer, J., J. E. Guikema, and C. E. Schrader. 2008. Mechanism and regulation of class switch recombination. *Annual review of immunology* 26: 261-292.

81. Kurosaki, T., K. Kometani, and W. Ise. 2015. Memory B cells. *Nature reviews. Immunology* 15: 149-159.
82. Klein, U., and R. Dalla-Favera. 2008. Germinal centres: role in B-cell physiology and malignancy. *Nature reviews. Immunology* 8: 22-33.
83. Röhlich, K. 1930. Beitrag zur Cytologie der Keimzentren der Lymphknoten. *Z Mikrosk Anat Forsch* 20: 287-297.
84. Six, A., M. E. Mariotti-Ferrandiz, W. Chaara, S. Magadan, H. P. Pham, M. P. Lefranc, T. Mora, V. Thomas-Vaslin, A. M. Walczak, and P. Boudinot. 2013. The past, present, and future of immune repertoire biology - the rise of next-generation repertoire analysis. *Frontiers in immunology* 4: 413.
85. Mehr, R., M. Sternberg-Simon, M. Michaeli, and Y. Pickman. 2012. Models and methods for analysis of lymphocyte repertoire generation, development, selection and evolution. *Immunology letters* 148: 11-22.
86. van Lochem, E. G., V. H. van der Velden, H. K. Wind, J. G. te Marvelde, N. A. Westerdal, and J. J. van Dongen. 2004. Immunophenotypic differentiation patterns of normal hematopoiesis in human bone marrow: reference patterns for age-related changes and disease-induced shifts. *Cytometry. Part B, Clinical cytometry* 60: 1-13.
87. de Masson, A., H. Le Buanec, and J. D. Bouaziz. 2014. Purification and immunophenotypic characterization of human B cells with regulatory functions. *Methods in molecular biology* 1190: 45-52.
88. Zheng, Z. J., and R. L. Xu. 2003. Immunophenotypic characterization of normal peripheral blood B lymphocyte by flow cytometry: reference for diagnosis of chronic B cell leukemia/lymphoma. *Zhongguo shi yan xue ye xue za zhi / Zhongguo bing li sheng li xue hui = Journal of experimental hematology / Chinese Association of Pathophysiology* 11: 398-404.
89. Paloczi, K., A. Batai, L. Gopcsa, R. Ezsi, and G. G. Petranyi. 1998. Immunophenotypic characterisation of cord blood B-lymphocytes. *Bone marrow transplantation* 22 Suppl 4: S89-91.
90. Llinas, L., A. Lazaro, J. de Salort, J. Matesanz-Isabel, J. Sintes, and P. Engel. 2011. Expression profiles of novel cell surface molecules on B-cell subsets and plasma cells as analyzed by flow cytometry. *Immunology letters* 134: 113-121.
91. Irish, J. M., D. K. Czerwinski, G. P. Nolan, and R. Levy. 2006. Kinetics of B cell receptor signaling in human B cell subsets mapped by phosphospecific flow cytometry. *Journal of immunology* 177: 1581-1589.
92. Moir, S., J. Ho, A. Malaspina, W. Wang, A. C. DiPoto, M. A. O'Shea, G. Roby, S. Kottlilil, J. Arthos, M. A. Proschan, T. W. Chun, and A. S. Fauci. 2008. Evidence for HIV-associated B cell exhaustion in a dysfunctional memory B cell compartment in HIV-infected viremic individuals. *The Journal of experimental medicine* 205: 1797-1805.
93. Shinall, S. M., M. Gonzalez-Fernandez, R. J. Noelle, and T. J. Waldschmidt. 2000. Identification of murine germinal center B cell subsets defined by the expression of surface isotypes and differentiation antigens. *Journal of immunology* 164: 5729-5738.
94. Demberg, T., E. Brocca-Cofano, P. Xiao, D. Venzon, D. Vargas-Inchaustegui, E. M. Lee, I. Kalisz, V. S. Kalyanaraman, J. Dipasquale, K. McKinnon, and M. Robert-Guroff. 2012. Dynamics of memory B-cell populations in blood, lymph nodes, and bone marrow during antiretroviral therapy and envelope boosting in simian immunodeficiency virus SIVmac251-infected rhesus macaques. *Journal of virology* 86: 12591-12604.
95. Moody, M. A., and B. F. Haynes. 2008. Antigen-specific B cell detection reagents: use and quality control. *Cytometry. Part A: the journal of the International Society for Analytical Cytology* 73: 1086-1092.
96. Franz, B., K. F. May, Jr., G. Dranoff, and K. Wucherpfennig. 2011. Ex vivo characterization and isolation of rare memory B cells with antigen tetramers. *Blood* 118: 348-357.



97. Kodituwakku, A. P., C. Jessup, H. Zola, and D. M. Robertson. 2003. Isolation of antigen-specific B cells. *Immunology and cell biology* 81: 163-170.
98. Townsend, S. E., C. C. Goodnow, and R. J. Cornall. 2001. Single epitope multiple staining to detect ultralow frequency B cells. *Journal of immunological methods* 249: 137-146.
99. Doucett, V. P., W. Gerhard, K. Owler, D. Curry, L. Brown, and N. Baumgarth. 2005. Enumeration and characterization of virus-specific B cells by multicolor flow cytometry. *Journal of immunological methods* 303: 40-52.
100. Leyendeckers, H., M. Odendahl, A. Lohndorf, J. Irsch, M. Spangfort, S. Miltenyi, N. Hunzelmann, M. Assenmacher, A. Radbruch, and J. Schmitz. 1999. Correlation analysis between frequencies of circulating antigen-specific IgG-bearing memory B cells and serum titers of antigen-specific IgG. *European journal of immunology* 29: 1406-1417.
101. Crotty, S., R. D. Aubert, J. Glidewell, and R. Ahmed. 2004. Tracking human antigen-specific memory B cells: a sensitive and generalized ELISPOT system. *Journal of immunological methods* 286: 111-122.
102. Pinna, D., D. Corti, D. Jarrossay, F. Sallusto, and A. Lanzavecchia. 2009. Clonal dissection of the human memory B-cell repertoire following infection and vaccination. *European journal of immunology* 39: 1260-1270.
103. Tokimitsu, Y., H. Kishi, S. Kondo, R. Honda, K. Tajiri, K. Motoki, T. Ozawa, S. Kadowaki, T. Obata, S. Fujiki, C. Tateno, H. Takaishi, K. Chayama, K. Yoshizato, E. Tamiya, T. Sugiyama, and A. Muraguchi. 2007. Single lymphocyte analysis with a microwell array chip. *Cytometry. Part A: the journal of the International Society for Analytical Cytology* 71: 1003-1010.
104. Park, S., J. Han, W. Kim, G. M. Lee, and H. S. Kim. 2011. Rapid selection of single cells with high antibody production rates by microwell array. *Journal of biotechnology* 156: 197-202.
105. Haury, M., A. Grandien, A. Sundblad, A. Coutinho, and A. Nobrega. 1994. Global analysis of antibody repertoires. 1. An immunoblot method for the quantitative screening of a large number of reactivities. *Scandinavian journal of immunology* 39: 79-87.
106. Boutz, D. R., A. P. Horton, Y. Wine, J. J. Lavinder, G. Georgiou, and E. M. Marcotte. 2014. Proteomic identification of monoclonal antibodies from serum. *Analytical chemistry* 86: 4758-4766.
107. Cheung, W. C., S. A. Beausoleil, X. Zhang, S. Sato, S. M. Schieferl, J. S. Wieler, J. G. Beaudet, R. K. Ramenani, L. Popova, M. J. Comb, J. Rush, and R. D. Polakiewicz. 2012. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nature biotechnology* 30: 447-452.
108. Sato, S., S. A. Beausoleil, L. Popova, J. G. Beaudet, R. K. Ramenani, X. Zhang, J. S. Wieler, S. M. Schieferl, W. C. Cheung, and R. D. Polakiewicz. 2012. Proteomics-directed cloning of circulating antiviral human monoclonal antibodies. *Nature biotechnology* 30: 1039-1043.
109. Wine, Y., D. R. Boutz, J. J. Lavinder, A. E. Miklos, R. A. Hughes, K. H. Hoi, S. T. Jung, A. P. Horton, E. M. Murrin, A. D. Ellington, E. M. Marcotte, and G. Georgiou. 2013. Molecular deconvolution of the monoclonal antibodies that comprise the polyclonal serum response. *Proceedings of the National Academy of Sciences* 110: 2993-2998.
110. Cochet, M., C. Pannetier, A. Regnault, S. Darche, C. Leclerc, and P. Kourilsky. 1992. Molecular detection and in vivo analysis of the specific T cell response to a protein antigen. *European journal of immunology* 22: 2639-2647.
111. Gibson, K. L., Y. C. Wu, Y. Barnett, O. Duggan, R. Vaughan, E. Kondeatis, B. O. Nilsson, A. Wikby, D. Kipling, and D. K. Dunn-Walters. 2009. B-cell diversity decreases in old age and is correlated with poor health status. *Aging cell* 8: 18-25.
112. Lim, A., V. Baron, L. Ferradini, M. Bonneville, P. Kourilsky, and C. Pannetier. 2002. Combination of MHC-peptide multimer-based T cell sorting with the Immunoscope permits sensitive ex vivo

- quantitation and follow-up of human CD8+ T cell immune responses. *Journal of immunological methods* 261: 177-194.
113. Ema, H., A. Cumano, and P. Kourilsky. 1997. TCR-beta repertoire development in the mouse embryo. *Journal of immunology* 159: 4227-4232.
  114. Arstila, T. P., A. Casrouge, V. Baron, J. Even, J. Kanellopoulos, and P. Kourilsky. 1999. A direct estimate of the human alphabeta T cell receptor diversity. *Science* 286: 958-961.
  115. Bernard, D., A. Six, L. Rigottier-Gois, S. Messiaen, S. Chilmonczyk, E. Quillet, P. Boudinot, and A. Benmansour. 2006. Phenotypic and functional similarity of gut intraepithelial and systemic T cells in a teleost fish. *Journal of immunology* 176: 3942-3949.
  116. Boudinot, P., D. Bernard, S. Boubekur, M. I. Thoulouze, M. Bremont, and A. Benmansour. 2004. The glycoprotein of a fish rhabdovirus profiles the virus-specific T-cell repertoire in rainbow trout. *The Journal of general virology* 85: 3099-3108.
  117. Castro, R., L. Jouneau, H. P. Pham, O. Bouchez, V. Giudicelli, M. P. Lefranc, E. Quillet, A. Benmansour, F. Cazals, A. Six, S. Fillatreau, O. Sunyer, and P. Boudinot. 2013. Teleost fish mount complex clonal IgM and IgT responses in spleen upon systemic viral infection. *PLoS pathogens* 9: e1003098.
  118. Verfuërth, S., K. Peggs, P. Vyas, L. Barnett, R. J. O'Reilly, and S. Mackinnon. 2000. Longitudinal monitoring of immune reconstitution by CDR3 size spectratyping after T-cell-depleted allogeneic bone marrow transplant and the effect of donor lymphocyte infusions on T-cell repertoire. *Blood* 95: 3990-3995.
  119. Cibotti, R., J. P. Cabaniols, C. Pannetier, C. Delarbre, I. Vergnon, J. M. Kanellopoulos, and P. Kourilsky. 1994. Public and private V beta T cell receptor repertoires against hen egg white lysozyme (HEL) in nontransgenic versus HEL transgenic mice. *The Journal of experimental medicine* 180: 861-872.
  120. Sourdive, D. J., K. Murali-Krishna, J. D. Altman, A. J. Zajac, J. K. Whitmire, C. Pannetier, P. Kourilsky, B. Evavold, A. Sette, and R. Ahmed. 1998. Conserved T cell receptor repertoire in primary and memory CD8 T cell responses to an acute viral infection. *The Journal of experimental medicine* 188: 71-82.
  121. Wu, G. E., and C. J. Paige. 1986. VH gene family utilization in colonies derived from B and pre-B cells detected by the RNA colony blot assay. *The EMBO journal* 5: 3475.
  122. Schulze, D. H., and G. Kelsoe. 1987. Genotypic analysis of B cell colonies by in situ hybridization. Stoichiometric expression of three VH families in adult C57BL/6 and BALB/c mice. *The Journal of experimental medicine* 166: 163-172.
  123. Pannetier, C., S. Delassus, S. Darce, C. Saucier, and P. Kourilsky. 1993. Quantitative titration of nucleic acids by enzymatic amplification reactions run to saturation. *Nucleic acids research* 21: 577-583.
  124. Flohr, T., A. Schrauder, G. Cazzaniga, R. Panzer-Grümayer, V. van der Velden, S. Fischer, M. Stanulla, G. Basso, F. Niggli, and B. Schäfer. 2008. Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia. *Leukemia* 22: 771-782.
  125. Van der Velden, V., A. Hochhaus, G. Cazzaniga, T. Szczepanski, J. Gabert, and J. Van Dongen. 2003. Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia* 17: 1013-1034.
  126. Verhagen, O., M. Willemse, W. Breunis, A. Wijkhuijs, D. Jacobs, S. Joosten, E. Van Wering, J. Van Dongen, and C. Van der Schoot. 2000. Application of germline IGH probes in real-time

- quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia. *Leukemia* 14: 1426-1435.
127. Corti, D., and A. Lanzavecchia. 2013. Broadly neutralizing antiviral antibodies. *Annual review of immunology* 31: 705-742.
128. El-Metwally, S., O. M. Ouda, and M. Helmy. 2014. Next-Generation Sequencing Platforms. In *Next Generation Sequencing Technologies and Challenges in Sequence Assembly*. SpringerBriefs in Systems Biology. 37-44.
129. Quail, M. A., M. Smith, P. Coupland, T. D. Otto, S. R. Harris, T. R. Connor, A. Bertoni, H. P. Swerdlow, and Y. Gu. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics* 13: 341.
130. Mardis, E. R. 2013. Next-generation sequencing platforms. *Annual review of analytical chemistry* 6: 287-303.
131. He, L., D. Sok, P. Azadnia, J. Hsueh, E. Landais, M. Simek, W. C. Koff, P. Poignard, D. R. Burton, and J. Zhu. 2014. Toward a more accurate view of human B-cell repertoire by next-generation sequencing; unbiased repertoire capture and single-molecule barcoding. *Scientific reports* 4: 6778.
132. Rusk, N. 2011. Torrents of sequence. *Nat Meth* 8: 44-44.
133. Rothberg, J. M., W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth, and J. Bustillo. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475: 348-352.
134. Liu, L., Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu, and M. Law. 2012. Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology* 2012: 251364.
135. El-Metwally, S., O. M. Ouda, and M. Helmy. 2014. First- and Next-Generations Sequencing Methods. In *Next Generation Sequencing Technologies and Challenges in Sequence Assembly*. SpringerBriefs in Systems Biology. 29-36.
136. Ye, J., N. Ma, T. L. Madden, and J. M. Ostell. 2013. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic acids research*: gkt382.
137. Alamyar, E., V. Giudicelli, S. Li, P. Duroux, and M.-P. Lefranc. 2012. IMGT/HighV-QUEST: the IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res* 8: 26.
138. Munshaw, S., and T. B. Kepler. 2010. SoDA2: a Hidden Markov Model approach for identification of immunoglobulin rearrangements. *Bioinformatics* 26: 867-872.
139. Gaëta, B. A., H. R. Malming, K. J. Jackson, M. E. Bain, P. Wilson, and A. M. Collins. 2007. iHMMune-align: hidden Markov model-based alignment and identification of germline genes in rearranged immunoglobulin gene sequences. *Bioinformatics* 23: 1580-1587.
140. Li, S., M. P. Lefranc, J. J. Miles, E. Alamyar, V. Giudicelli, P. Duroux, J. D. Freeman, V. D. Corbin, J. P. Scheerlinck, M. A. Frohman, P. U. Cameron, M. Plebanski, B. Loveland, S. R. Burrows, A. T. Papenfuss, and E. J. Gowans. 2013. IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. *Nature communications* 4: 2333.
141. Michaeli, M., M. Barak, L. Hazanov, H. Noga, and R. Mehr. 2013. Automated analysis of immunoglobulin genes from high-throughput sequencing: life without a template. *Journal of clinical bioinformatics* 3: 15.

142. Lefranc, M.-P., V. Giudicelli, C. Ginestoux, J. Jabado-Michaloud, G. Folch, F. Bellahcene, Y. Wu, E. Gemrot, X. Brochet, and J. m. Lane. 2009. IMGT®, the international ImMunoGeneTics information system®. *Nucleic acids research* 37: D1006-D1012.
143. Alamyar, E., P. Duroux, M. P. Lefranc, and V. Giudicelli. 2012. IMGT((R)) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods in molecular biology* 882: 569-604.
144. Rogosch, T., S. Kerzel, K. H. Hoi, Z. Zhang, R. F. Maier, G. C. Ippolito, and M. Zemlin. 2012. Immunoglobulin analysis tool: a novel tool for the analysis of human and mouse heavy and light chain transcripts. *Frontiers in immunology* 3: 176.
145. Sok, D., U. Laserson, J. Laserson, Y. Liu, F. Vigneault, J. P. Julien, B. Briney, A. Ramos, K. F. Saye, K. Le, A. Mahan, S. Wang, M. Kardar, G. Yaari, L. M. Walker, B. B. Simen, E. P. St John, P. Y. Chan-Hui, K. Swiderek, S. H. Kleinstein, G. Alter, M. S. Seaman, A. K. Chakraborty, D. Koller, I. A. Wilson, G. M. Church, D. R. Burton, and P. Poignard. 2013. The effects of somatic hypermutation on neutralization and binding in the PGT121 family of broadly neutralizing HIV antibodies. *PLoS pathogens* 9: e1003754.
146. Barak, M., N. S. Zuckerman, H. Edelman, R. Unger, and R. Mehr. 2008. IgTree: creating Immunoglobulin variable region gene lineage trees. *Journal of immunological methods* 338: 67-74.
147. Bergqvist, P., A. Stensson, L. Hazanov, A. Holmberg, J. Mattsson, R. Mehr, M. Bemark, and N. Y. Lycke. 2013. Re-utilization of germinal centers in multiple Peyer's patches results in highly synchronized, oligoclonal, and affinity-matured gut IgA responses. *Mucosal immunology* 6: 122-135.
148. Green, M. R., A. J. Gentles, R. V. Nair, J. M. Irish, S. Kihira, C. L. Liu, I. Kela, E. S. Hopmans, J. H. Myklebust, H. Ji, S. K. Plevritis, R. Levy, and A. A. Alizadeh. 2013. Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. *Blood* 121: 1604-1611.
149. Zuckerman, N. S., K. J. McCann, C. H. Ottensmeier, M. Barak, G. Shahaf, H. Edelman, D. Dunn-Walters, R. S. Abraham, F. K. Stevenson, and R. Mehr. 2010. Ig gene diversification and selection in follicular lymphoma, diffuse large B cell lymphoma and primary central nervous system lymphoma revealed by lineage tree and mutation analyses. *International immunology* 22: 875-887.
150. von Büdingen, H.-C., T. C. Kuo, M. Sirota, C. J. van Belle, L. Apeltsin, J. Glanville, B. A. Cree, P.-A. Gourraud, A. Schwartzburg, and G. Huerta. 2012. B cell exchange across the blood-brain barrier in multiple sclerosis. *The Journal of clinical investigation* 122: 4533.
151. Bashford-Rogers, R. J., A. L. Palser, B. J. Huntly, R. Rance, G. S. Vassiliou, G. A. Follows, and P. Kellam. 2013. Network properties derived from deep sequencing of human B-cell receptor repertoires delineate B-cell populations. *Genome research* 23: 1874-1884.
152. Jiang, N., J. He, J. A. Weinstein, L. Penland, S. Sasaki, X. S. He, C. L. Dekker, N. Y. Zheng, M. Huang, M. Sullivan, P. C. Wilson, H. B. Greenberg, M. M. Davis, D. S. Fisher, and S. R. Quake. 2013. Lineage structure of the human antibody repertoire in response to influenza vaccination. *Science translational medicine* 5: 171ra119.
153. Mathonet, P., and C. G. Ullman. 2013. The application of next generation sequencing to the understanding of antibody repertoires. *Frontiers in immunology* 4: 265.
154. Georgiou, G., G. C. Ippolito, J. Beausang, C. E. Busse, H. Wardemann, and S. R. Quake. 2014. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nature biotechnology* 32: 158-168.
155. Boyd, S. D., and S. A. Joshi. 2014. High-Throughput DNA Sequencing Analysis of Antibody Repertoires.
156. Calis, J. J., and B. R. Rosenberg. 2014. Characterizing immune repertoires by high throughput sequencing: strategies and applications. *Trends in immunology* 35: 581-590.

157. Galson, J. D., A. J. Pollard, J. Truck, and D. F. Kelly. 2014. Studying the antibody repertoire after vaccination: practical applications. *Trends Immunol* 35: 319-331.
158. Weinstein, J. A., N. Jiang, R. A. White, 3rd, D. S. Fisher, and S. R. Quake. 2009. High-throughput sequencing of the zebrafish antibody repertoire. *Science* 324: 807-810.
159. Reddy, S. T., X. Ge, A. E. Miklos, R. A. Hughes, S. H. Kang, K. H. Hoi, C. Chrysostomou, S. P. Hunicke-Smith, B. L. Iverson, P. W. Tucker, A. D. Ellington, and G. Georgiou. 2010. Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nature biotechnology* 28: 965-969.
160. Kodangattil, S., C. Huard, C. Ross, J. Li, H. Gao, A. Mascioni, S. Hodawadekar, S. Naik, J. Mindebartolo, A. Visintin, and J. C. Almagro. 2014. The functional repertoire of rabbit antibodies and antibody discovery via next-generation sequencing. *mAbs* 6: 628-636.
161. Larsen, P. A., and T. P. Smith. 2012. Application of circular consensus sequencing and network analysis to characterize the bovine IgG repertoire. *BMC immunology* 13: 52.
162. Sundling, C., Z. Zhang, G. E. Phad, Z. Sheng, Y. Wang, J. R. Mascola, Y. Li, R. T. Wyatt, L. Shapiro, and G. B. Karlsson Hedestam. 2014. Single-cell and deep sequencing of IgG-switched macaque B cells reveal a diverse Ig repertoire following immunization. *Journal of immunology* 192: 3637-3644.
163. Wu, L., K. Oficjalska, M. Lambert, B. J. Fennell, A. Darmanin-Sheehan, D. Ni Shuilleabhain, B. Autin, E. Cummins, L. Tchistiakova, L. Bloom, J. Paulsen, D. Gill, O. Cunningham, and W. J. Finlay. 2012. Fundamental characteristics of the immunoglobulin VH repertoire of chickens in comparison with those of humans, mice, and camelids. *Journal of immunology* 188: 322-333.
164. Jiang, N., J. A. Weinstein, L. Penland, R. A. White, 3rd, D. S. Fisher, and S. R. Quake. 2011. Determinism and stochasticity during maturation of the zebrafish antibody repertoire. *Proceedings of the National Academy of Sciences of the United States of America* 108: 5348-5353.
165. Rechavi, E., A. Lev, Y. N. Lee, A. J. Simon, Y. Yinon, S. Lipitz, N. Amariglio, B. Weisz, L. D. Notarangelo, and R. Somech. 2015. Timely and spatially regulated maturation of B and T cell repertoire during human fetal development. *Science translational medicine* 7: 276ra225.
166. Prabakaran, P., W. Chen, M. G. Singarayan, C. C. Stewart, E. Streaker, Y. Feng, and D. S. Dimitrov. 2012. Expressed antibody repertoires in human cord blood cells: 454 sequencing and IMGT/HighV-QUEST analysis of germline gene usage, junctional diversity, and somatic mutations. *Immunogenetics* 64: 337-350.
167. Glanville, J., T. C. Kuo, H. C. von Budingen, L. Guey, J. Berka, P. D. Sundar, G. Huerta, G. R. Mehta, J. R. Oksenberg, S. L. Hauser, D. R. Cox, A. Rajpal, and J. Pons. 2011. Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation. *Proceedings of the National Academy of Sciences of the United States of America* 108: 20066-20071.
168. Boyd, S. D., E. L. Marshall, J. D. Merker, J. M. Maniar, L. N. Zhang, B. Sahaf, C. D. Jones, B. B. Simen, B. Hanczaruk, K. D. Nguyen, K. C. Nadeau, M. Egholm, D. B. Miklos, J. L. Zehnder, and A. Z. Fire. 2009. Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Science translational medicine* 1: 12ra23.
169. Arnaout, R., W. Lee, P. Cahill, T. Honan, T. Sparrow, M. Weiland, C. Nusbaum, K. Rajewsky, and S. B. Koralov. 2011. High-resolution description of antibody heavy-chain repertoires in humans. *PLoS One* 6: e22365.
170. Glanville, J., W. Zhai, J. Berka, D. Telman, G. Huerta, G. R. Mehta, I. Ni, L. Mei, P. D. Sundar, G. M. Day, D. Cox, A. Rajpal, and J. Pons. 2009. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proceedings of the National Academy of Sciences of the United States of America* 106: 20216-20221.

171. Boyd, S. D., B. A. Gaeta, K. J. Jackson, A. Z. Fire, E. L. Marshall, J. D. Merker, J. M. Maniar, L. N. Zhang, B. Sahaf, C. D. Jones, B. B. Simen, B. Hanczaruk, K. D. Nguyen, K. C. Nadeau, M. Egholm, D. B. Miklos, J. L. Zehnder, and A. M. Collins. 2010. Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *Journal of immunology* 184: 6986-6992.
172. Wang, Y., K. J. Jackson, B. Gäeta, W. Pomat, P. Siba, W. A. Sewell, and A. M. Collins. 2011. Genomic screening by 454 pyrosequencing identifies a new human IGHV gene and sixteen other new IGHV allelic variants. *Immunogenetics* 63: 259-265.
173. Kidd, M. J., Z. Chen, Y. Wang, K. J. Jackson, L. Zhang, S. D. Boyd, A. Z. Fire, M. M. Tanaka, B. A. Gaëta, and A. M. Collins. 2012. The inference of phased haplotypes for the immunoglobulin H chain V region gene loci by analysis of VDJ gene rearrangements. *The Journal of Immunology* 188: 1333-1340.
174. Watson, C. T., K. M. Steinberg, J. Huddleston, R. L. Warren, M. Malig, J. Schein, A. J. Willsey, J. B. Joy, J. K. Scott, T. A. Graves, R. K. Wilson, R. A. Holt, E. E. Eichler, and F. Breden. 2013. Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. *American journal of human genetics* 92: 530-546.
175. Larimore, K., M. W. McCormick, H. S. Robins, and P. D. Greenberg. 2012. Shaping of human germline IgH repertoires revealed by deep sequencing. *Journal of immunology* 189: 3221-3230.
176. Briney, B. S., J. R. Willis, M. D. Hicar, J. W. Thomas, 2nd, and J. E. Crowe, Jr. 2012. Frequency and genetic characterization of V(DD)J recombinants in the human peripheral blood antibody repertoire. *Immunology* 137: 56-64.
177. Ben-Hamo, R., and S. Efroni. 2011. The whole-organism heavy chain B cell repertoire from Zebrafish self-organizes into distinct network features. *BMC systems biology* 5: 27.
178. Wu, Y. C., D. Kipling, and D. K. Dunn-Walters. 2012. Age-Related Changes in Human Peripheral Blood IGH Repertoire Following Vaccination. *Frontiers in immunology* 3: 193.
179. Wang, Y., K. J. Jackson, Z. Chen, B. A. Gaeta, P. M. Siba, W. Pomat, E. Walpole, J. Rimmer, W. A. Sewell, and A. M. Collins. 2011. IgE sequences in individuals living in an area of endemic parasitism show little mutational evidence of antigen selection. *Scandinavian journal of immunology* 73: 496-504.
180. Wu, Y.-C., D. Kipling, H. S. Leong, V. Martin, A. A. Ademokun, and D. K. Dunn-Walters. 2010. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. *Blood* 116: 1070-1078.
181. Tabibian-Keissar, H., G. Schiby, N. Azogui-Rosenthal, H. Hazanov, A. S. Rakovsky, M. Michaeli, K. Rosenblatt, R. Mehr, and I. Barshack. 2013. [B lymphocyte clonal evolution of human reactive lymph nodes revealed by lineage tree analysis]. *Harefuah* 152: 330-333, 369.
182. Tabibian-Keissar, H., G. Schibby, M. Michaeli, A. Rakovsky-Shapira, N. Azogui-Rosenthal, D. K. Dunn-Walters, K. Rosenblatt, R. Mehr, and I. Barshack. 2013. PCR amplification and high throughput sequencing of immunoglobulin heavy chain genes from formalin-fixed paraffin-embedded human biopsies. *Experimental and molecular pathology* 94: 182-187.
183. von Budingen, H. C., T. C. Kuo, M. Sirota, C. J. van Belle, L. Apeltsin, J. Glanville, B. A. Cree, P. A. Gourraud, A. Schwartzburg, G. Huerta, D. Telman, P. D. Sundar, T. Casey, D. R. Cox, and S. L. Hauser. 2012. B cell exchange across the blood-brain barrier in multiple sclerosis. *J Clin Invest* 122: 4533-4543.
184. Lindner, C., B. Wahl, L. Fohse, S. Suerbaum, A. J. Macpherson, I. Prinz, and O. Pabst. 2012. Age, microbiota, and T cells shape diverse individual IgA repertoires in the intestine. *The Journal of experimental medicine* 209: 365-377.

185. Pabst, O., H. Hazanov, and R. Mehr. 2015. Old questions, new tools: does next-generation sequencing hold the key to unraveling intestinal B-cell responses? *Mucosal immunology* 8: 29-37.
186. Wesemann, D. R., A. J. Portuguese, R. M. Meyers, M. P. Gallagher, K. Cluff-Jones, J. M. Magee, R. A. Panchakshari, S. J. Rodig, T. B. Kepler, and F. W. Alt. 2013. Microbial colonization influences early B-lineage development in the gut lamina propria. *Nature* 501: 112-115.
187. Racanelli, V., C. Brunetti, V. De Re, L. Caggiari, M. De Zorzi, P. Leone, F. Perosa, A. Vacca, and F. Dammacco. 2011. Antibody V(h) repertoire differences between resolving and chronically evolving hepatitis C virus infections. *PLoS One* 6: e25606.
188. Ademokun, A., Y. C. Wu, V. Martin, R. Mitra, U. Sack, H. Baxendale, D. Kipling, and D. K. Dunn-Walters. 2011. Vaccination-induced changes in human B-cell repertoire and pneumococcal IgM and IgA antibody at different ages. *Aging cell* 10: 922-930.
189. Vollmers, C., R. V. Sit, J. A. Weinstein, C. L. Dekker, and S. R. Quake. 2013. Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 110: 13463-13468.
190. DeKosky, B. J., G. C. Ippolito, R. P. Deschner, J. J. Lavinder, Y. Wine, B. M. Rawlings, N. Varadarajan, C. Giesecke, T. Dorner, S. F. Andrews, P. C. Wilson, S. P. Hunicke-Smith, C. G. Willson, A. D. Ellington, and G. Georgiou. 2013. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nature biotechnology* 31: 166-169.
191. Laserson, U., F. Vigneault, D. Gadala-Maria, G. Yaari, M. Uduman, J. A. Vander Heiden, W. Kelton, S. Taek Jung, Y. Liu, J. Laserson, R. Chari, J. H. Lee, I. Bachelet, B. Hickey, E. Lieberman-Aiden, B. Hanczaruk, B. B. Simen, M. Egholm, D. Koller, G. Georgiou, S. H. Kleinstein, and G. M. Church. 2014. High-resolution antibody dynamics of vaccine-induced immune responses. *Proceedings of the National Academy of Sciences of the United States of America* 111: 4928-4933.
192. Jackson, K. J., Y. Liu, K. M. Roskin, J. Glanville, R. A. Hoh, K. Seo, E. L. Marshall, T. C. Gurley, M. A. Moody, B. F. Haynes, E. B. Walter, H. X. Liao, R. A. Albrecht, A. Garcia-Sastre, J. Chaparro-Riggers, A. Rajpal, J. Pons, B. B. Simen, B. Hanczaruk, C. L. Dekker, J. Laserson, D. Koller, M. M. Davis, A. Z. Fire, and S. D. Boyd. 2014. Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. *Cell host & microbe* 16: 105-114.
193. Lavinder, J. J., Y. Wine, C. Giesecke, G. C. Ippolito, A. P. Horton, O. I. Lungu, K. H. Hoi, B. J. DeKosky, E. M. Murrin, M. M. Wirth, A. D. Ellington, T. Dorner, E. M. Marcotte, D. R. Boutz, and G. Georgiou. 2014. Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proceedings of the National Academy of Sciences of the United States of America* 111: 2259-2264.
194. Truck, J., M. N. Ramasamy, J. D. Galson, R. Rance, J. Parkhill, G. Lunter, A. J. Pollard, and D. F. Kelly. 2015. Identification of antigen-specific B cell receptor sequences using public repertoire analysis. *Journal of immunology* 194: 252-261.
195. Wiley, S. R., V. S. Raman, A. Desbien, H. R. Bailor, R. Bhardwaj, A. R. Shakri, S. G. Reed, C. E. Chitnis, and D. Carter. 2011. Targeting TLRs expands the antibody repertoire in response to a malaria vaccine. *Science translational medicine* 3: 93ra69.
196. Zhu, J., S. O'Dell, G. Ofek, M. Pancera, X. Wu, B. Zhang, Z. Zhang, N. C. S. Program, J. C. Mullikin, M. Simek, D. R. Burton, W. C. Koff, L. Shapiro, J. R. Mascola, and P. D. Kwong. 2012. Somatic Populations of PGT135-137 HIV-1-Neutralizing Antibodies Identified by 454 Pyrosequencing and Bioinformatics. *Frontiers in microbiology* 3: 315.
197. Zhu, J., X. Wu, B. Zhang, K. McKee, S. O'Dell, C. Soto, T. Zhou, J. P. Casazza, N. C. S. Program, J. C. Mullikin, P. D. Kwong, J. R. Mascola, and L. Shapiro. 2013. De novo identification of VRC01

- class HIV-1-neutralizing antibodies by next-generation sequencing of B-cell transcripts. *Proceedings of the National Academy of Sciences of the United States of America* 110: E4088-4097.
198. Wu, X., T. Zhou, J. Zhu, B. Zhang, I. Georgiev, C. Wang, X. Chen, N. S. Longo, M. Louder, K. McKee, S. O'Dell, S. Peretto, S. D. Schmidt, W. Shi, L. Wu, Y. Yang, Z. Y. Yang, Z. Yang, Z. Zhang, M. Bonsignori, J. A. Crump, S. H. Kapiga, N. E. Sam, B. F. Haynes, M. Simek, D. R. Burton, W. C. Koff, N. A. Doria-Rose, M. Connors, N. C. S. Program, J. C. Mullikin, G. J. Nabel, M. Roederer, L. Shapiro, P. D. Kwong, and J. R. Mascola. 2011. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* 333: 1593-1602.
199. Zhou, T., J. Zhu, X. Wu, S. Moquin, B. Zhang, P. Acharya, I. S. Georgiev, H. R. Altae-Tran, G. Y. Chuang, M. G. Joyce, Y. Do Kwon, N. S. Longo, M. K. Louder, T. Luongo, K. McKee, C. A. Schramm, J. Skinner, Y. Yang, Z. Yang, Z. Zhang, A. Zheng, M. Bonsignori, B. F. Haynes, J. F. Scheid, M. C. Nussenzweig, M. Simek, D. R. Burton, W. C. Koff, N. C. S. Program, J. C. Mullikin, M. Connors, L. Shapiro, G. J. Nabel, J. R. Mascola, and P. D. Kwong. 2013. Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-class antibodies. *Immunity* 39: 245-258.
200. Zhu, J., G. Ofek, Y. Yang, B. Zhang, M. K. Louder, G. Lu, K. McKee, M. Pancera, J. Skinner, Z. Zhang, R. Parks, J. Eudailey, K. E. Lloyd, J. Blinn, S. M. Alam, B. F. Haynes, M. Simek, D. R. Burton, W. C. Koff, N. C. S. Program, J. C. Mullikin, J. R. Mascola, L. Shapiro, and P. D. Kwong. 2013. Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proceedings of the National Academy of Sciences of the United States of America* 110: 6470-6475.
201. Krause, J. C., T. Tsibane, T. M. Tumpey, C. J. Huffman, B. S. Briney, S. A. Smith, C. F. Basler, and J. E. Crowe, Jr. 2011. Epitope-specific human influenza antibody repertoires diversify by B cell intracлонаl sequence divergence and interclonal convergence. *Journal of immunology* 187: 3704-3711.
202. Parameswaran, P., Y. Liu, K. M. Roskin, K. K. Jackson, V. P. Dixit, J. Y. Lee, K. L. Artilles, S. Zompi, M. J. Vargas, B. B. Simen, B. Hanczaruk, K. R. McGowan, M. A. Tariq, N. Pourmand, D. Koller, A. Balmaseda, S. D. Boyd, E. Harris, and A. Z. Fire. 2013. Convergent antibody signatures in human dengue. *Cell host & microbe* 13: 691-700.
203. Dimitrov, D. S. 2010. Therapeutic antibodies, vaccines and antibodyomes. *mAbs* 2: 347-356.
204. Xiao, X., W. Chen, Y. Feng, and D. S. Dimitrov. 2009. Maturation Pathways of Cross-Reactive HIV-1 Neutralizing Antibodies. *Viruses* 1: 802-817.
205. Xiao, X., W. Chen, Y. Feng, Z. Zhu, P. Prabakaran, Y. Wang, M. Y. Zhang, N. S. Longo, and D. S. Dimitrov. 2009. Germline-like predecessors of broadly neutralizing antibodies lack measurable binding to HIV-1 envelope glycoproteins: implications for evasion of immune responses and design of vaccine immunogens. *Biochemical and biophysical research communications* 390: 404-409.
206. Rubelt, F., V. Sievert, F. Knaust, C. Diener, T. S. Lim, K. Skriner, E. Klipp, R. Reinhardt, H. Lehrach, and Z. Konthur. 2012. Onset of immune senescence defined by unbiased pyrosequencing of human immunoglobulin mRNA repertoires. *PloS one* 7: e49774.
207. Kohler, G., and C. Milstein. 1976. Derivation of specific antibody-producing tissue culture and tumor lines by cell fusion. *European journal of immunology* 6: 511-519.
208. Naso, M. F., J. Lu, and T. Panavas. 2014. Deep sequencing approaches to antibody discovery. *Current drug discovery technologies* 11: 85-95.
209. Ravn, U., F. Gueneau, L. Baerlocher, M. Osteras, M. Desmurs, P. Malinge, G. Magistrelli, L. Farinelli, M. H. Kosco-Vilbois, and N. Fischer. 2010. By-passing in vitro screening—next generation sequencing technologies applied to antibody display and in silico candidate selection. *Nucleic acids research* 38: e193.



210. Fischer, N. 2011. Sequencing antibody repertoires: the next generation. *mAbs* 3: 17-20.
211. Saggy, I., Y. Wine, L. Shefet-Carasso, L. Nahary, G. Georgiou, and I. Benhar. 2012. Antibody isolation from immunized animals: comparison of phage display and antibody discovery via V gene repertoire mining. *Protein engineering, design & selection : PEDS* 25: 539-549.
212. Meijer, P. J., P. S. Andersen, M. Haahr Hansen, L. Steinaa, A. Jensen, J. Lantto, M. B. Oleksiewicz, K. Tengbjerg, T. R. Poulsen, V. W. Coljee, S. Bregenholt, J. S. Haurum, and L. S. Nielsen. 2006. Isolation of human antibody repertoires with preservation of the natural heavy and light chain pairing. *Journal of molecular biology* 358: 764-772.
213. Busse, C. E., I. Czogiel, P. Braun, P. F. Arndt, and H. Wardemann. 2014. Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes. *European journal of immunology* 44: 597-603.
214. DeKosky, B. J., T. Kojima, A. Rodin, W. Charab, G. C. Ippolito, A. D. Ellington, and G. Georgiou. 2015. In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nature medicine* 21: 86-91.
215. Group, O. M. T. S. 1985. A randomized clinical trial of OKT3 monoclonal antibody for acute rejection of cadaveric renal transplants. *N Engl J Med* 313: 337-342.
216. Pendley, C., A. Schantz, and C. Wagner. 2003. Immunogenicity of therapeutic monoclonal antibodies. *Current opinion in molecular therapeutics* 5: 172-179.
217. Kuus-Reichel, K., L. S. Grauer, L. M. Karavodin, C. Knott, M. Krusemeier, and N. E. Kay. 1994. Will immunogenicity limit the use, efficacy, and future development of therapeutic monoclonal antibodies? *Clinical and diagnostic laboratory immunology* 1: 365-372.
218. Lonberg, N. 2005. Human antibodies from transgenic animals. *Nature biotechnology* 23: 1117-1125.
219. Lonberg, N. 2008. Human monoclonal antibodies from transgenic mice. *Handbook of experimental pharmacology*: 69-97.
220. Lonberg, N. 2008. Fully human antibodies from transgenic mouse and phage display platforms. *Current opinion in immunology* 20: 450-459.
221. Bruggemann, M., M. J. Osborn, B. Ma, J. Hayre, S. Avis, B. Lundstrom, and R. Buelow. 2014. Human Antibody Production in Transgenic Animals. *Archivum immunologiae et therapiae experimentalis*.
222. Green, L. L. 2014. Transgenic mouse strains as platforms for the successful discovery and development of human therapeutic monoclonal antibodies. *Current drug discovery technologies* 11: 74-84.
223. Morrison, S. L., M. J. Johnson, L. A. Herzenberg, and V. T. Oi. 1984. Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proceedings of the National Academy of Sciences of the United States of America* 81: 6851-6855.
224. Jones, P. T., P. H. Dear, J. Foote, M. S. Neuberger, and G. Winter. 1986. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* 321: 522-525.
225. Smith, G. P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228: 1315-1317.
226. McCafferty, J., A. D. Griffiths, G. Winter, and D. J. Chiswell. 1990. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* 348: 552-554.
227. Bruggemann, M., H. M. Caskey, C. Teale, H. Waldmann, G. T. Williams, M. A. Surani, and M. S. Neuberger. 1989. A repertoire of monoclonal antibodies with human heavy chains from transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* 86: 6709-6713.

228. Lonberg, N., L. D. Taylor, F. A. Harding, M. Trounstein, K. M. Higgins, S. R. Schramm, C. C. Kuo, R. Mashayekh, K. Wymore, J. G. McCabe, and et al. 1994. Antigen-specific human antibodies from mice comprising four distinct genetic modifications. *Nature* 368: 856-859.
229. Green, L. L., M. C. Hardy, C. E. Maynard-Currie, H. Tsuda, D. M. Louie, M. J. Mendez, H. Abderrahim, M. Noguchi, D. H. Smith, Y. Zeng, N. E. David, H. Sasai, D. Garza, D. G. Brenner, J. F. Hales, R. P. McGuinness, D. J. Capon, S. Klapholz, and A. Jakobovits. 1994. Antigen-specific human monoclonal antibodies from mice engineered with human Ig heavy and light chain YACs. *Nature genetics* 7: 13-21.
230. Taylor, L. D., C. E. Carmack, D. Huszar, K. M. Higgins, R. Mashayekh, G. Sequar, S. R. Schramm, C. C. Kuo, S. L. O'Donnell, R. M. Kay, and et al. 1994. Human immunoglobulin transgenes undergo rearrangement, somatic mutation and class switching in mice that lack endogenous IgM. *International immunology* 6: 579-591.
231. Tomizuka, K., H. Yoshida, H. Uejima, H. Kugoh, K. Sato, A. Ohguma, M. Hayasaka, K. Hanaoka, M. Oshimura, and I. Ishida. 1997. Functional expression and germline transmission of a human chromosome fragment in chimaeric mice. *Nature genetics* 16: 133-143.
232. Tomizuka, K., T. Shinohara, H. Yoshida, H. Uejima, A. Ohguma, S. Tanaka, K. Sato, M. Oshimura, and I. Ishida. 2000. Double trans-chromosomal mice: maintenance of two individual human chromosome fragments containing Ig heavy and kappa loci and expression of fully human antibodies. *Proceedings of the National Academy of Sciences of the United States of America* 97: 722-727.
233. Kuroiwa, Y., K. Tomizuka, T. Shinohara, Y. Kazuki, H. Yoshida, A. Ohguma, T. Yamamoto, S. Tanaka, M. Oshimura, and I. Ishida. 2000. Manipulation of human minichromosomes to carry greater than megabase-sized chromosome inserts. *Nature biotechnology* 18: 1086-1090.
234. Popov, A. V., X. Zou, J. Xian, I. C. Nicholson, and M. Bruggemann. 1999. A human immunoglobulin lambda locus is similarly well expressed in mice and humans. *The Journal of experimental medicine* 189: 1611-1620.
235. Osborn, M. J., B. Ma, S. Avis, A. Binnie, J. Dilley, X. Yang, K. Lindquist, S. Menoret, A. L. Iscache, L. H. Ouisse, A. Rajpal, I. Anegon, M. S. Neuberger, R. Buelow, and M. Bruggemann. 2013. High-affinity IgG antibodies develop naturally in Ig-knockout rats carrying germline human IgH/Igkappa/Iglambda loci bearing the rat CH region. *Journal of immunology* 190: 1481-1490.
236. Pruzina, S., G. T. Williams, G. Kaneva, S. L. Davies, A. Martin-Lopez, M. Bruggemann, S. M. Vieira, S. A. Jeffs, Q. J. Sattentau, and M. S. Neuberger. 2011. Human monoclonal antibodies to HIV-1 gp140 from mice bearing YAC-based human immunoglobulin transloci. *Protein engineering, design & selection: PEDS* 24: 791-799.
237. Ma, B., M. J. Osborn, S. Avis, L. H. Ouisse, S. Menoret, I. Anegon, R. Buelow, and M. Bruggemann. 2013. Human antibody expression in transgenic rats: comparison of chimeric IgH loci with human VH, D and JH but bearing different rat C-gene regions. *Journal of immunological methods* 400-401: 78-86.
238. Griffin, D. E., and M. B. Oldstone. 2008. *Measles: history and basic biology*. Springer Science & Business Media.
239. Perry, R. T., M. Gacic-Dobo, A. Dabbagh, M. N. Mulders, P. M. Strebel, J.-M. Okwo-Bele, P. A. Rota, and J. L. Goodson. 2014. Global control and regional elimination of measles, 2000-2012. *MMWR. Morbidity and mortality weekly report* 63: 103-107.
240. Bank, W. 2014. Levels and trends in child mortality : estimates developed by the UN Inter-agency Group for child Mortality Estimation (IGME) - report 2014. . World Bank Group., Washington, DC.
241. Moss, W. J., and D. E. Griffin. 2006. Global measles elimination. *Nature reviews. Microbiology* 4: 900-908.

242. Norrby, E., J. Kovamees, M. Blixenkroné-Møller, B. Sharma, and C. Orvell. 1992. Humanized animal viruses with special reference to the primate adaptation of morbillivirus. *Veterinary microbiology* 33: 275-286.
243. Armstrong, M. A., K. B. Fraser, E. Dermott, and P. V. Shirodaria. 1982. Immunoelectron microscopic studies on haemagglutinin and haemolysin of measles virus in infected HEp2 cells. *The Journal of general virology* 59: 187-192.
244. Casali, P., J. G. Sissons, R. S. Fujinami, and M. B. Oldstone. 1981. Purification of measles virus glycoproteins and their integration into artificial lipid membranes. *The Journal of general virology* 54: 161-171.
245. Rager, M., S. Vongpunsawad, W. P. Duprex, and R. Cattaneo. 2002. Polyploid measles virus with hexameric genome length. *The EMBO journal* 21: 2364-2372.
246. Graves, M., D. E. Griffin, R. T. Johnson, R. L. Hirsch, I. L. de Soriano, S. Roedenbeck, and A. Vaisberg. 1984. Development of antibody to measles virus polypeptides during complicated and uncomplicated measles virus infections. *Journal of virology* 49: 409-412.
247. Curran, J., R. Boeck, N. Lin-Marq, A. Lupas, and D. Kolakofsky. 1995. Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils. *Virology* 214: 139-149.
248. Bellini, W. J., G. Englund, S. Rozenblatt, H. Arnheiter, and C. D. Richardson. 1985. Measles virus P gene codes for two proteins. *Journal of virology* 53: 908-919.
249. Cattaneo, R., K. Kaelin, K. Baczkó, and M. A. Billeter. 1989. Measles virus editing provides an additional cysteine-rich protein. *Cell* 56: 759-764.
250. Liston, P., and D. J. Briedis. 1995. Ribosomal frameshifting during translation of measles virus P protein mRNA is capable of directing synthesis of a unique protein. *Journal of virology* 69: 6742-6750.
251. Palosaari, H., J. P. Parisien, J. J. Rodriguez, C. M. Ulane, and C. M. Horvath. 2003. STAT protein interference and suppression of cytokine signal transduction by measles virus V protein. *Journal of virology* 77: 7635-7644.
252. Reutter, G. L., C. Cortese-Grogan, J. Wilson, and S. A. Moyer. 2001. Mutations in the measles virus C protein that up regulate viral RNA synthesis. *Virology* 285: 100-109.
253. Shaffer, J. A., W. J. Bellini, and P. A. Rota. 2003. The C protein of measles virus inhibits the type I interferon response. *Virology* 315: 389-397.
254. Takeuchi, K., S. I. Kadota, M. Takeda, N. Miyajima, and K. Nagata. 2003. Measles virus V protein blocks interferon (IFN)-alpha/beta but not IFN-gamma signaling by inhibiting STAT1 and STAT2 phosphorylation. *FEBS letters* 545: 177-182.
255. Hirano, A., M. Ayata, A. H. Wang, and T. C. Wong. 1993. Functional analysis of matrix proteins expressed from cloned genes of measles virus variants that cause subacute sclerosing panencephalitis reveals a common defect in nucleocapsid binding. *Journal of virology* 67: 1848-1853.
256. Wild, T. F., and R. Buckland. 1997. Inhibition of measles virus infection and fusion with peptides corresponding to the leucine zipper region of the fusion protein. *The Journal of general virology* 78 (Pt 1): 107-111.
257. Hu, A., J. Kovamees, and E. Norrby. 1994. Intracellular processing and antigenic maturation of measles virus hemagglutinin protein. *Archives of virology* 136: 239-253.
258. Rota, P. A., D. A. Featherstone, and W. J. Bellini. 2009. Molecular epidemiology of measles virus. *Current topics in microbiology and immunology* 330: 129-150.
259. Rota, P. A., and W. J. Bellini. 2003. Update on the global distribution of genotypes of wild type measles viruses. *The Journal of infectious diseases* 187 Suppl 1: S270-276.

260. Rota, P. A., K. Brown, A. Mankertz, S. Santibanez, S. Shulga, C. P. Muller, J. M. Hubschen, M. Siqueira, J. Beirnes, H. Ahmed, H. Triki, S. Al-Busaidy, A. Dosseh, C. Byabamazima, S. Smit, C. Akoua-Koffi, J. Bwogi, H. Bukenya, N. Wairagkar, N. Ramamurty, P. Incomserb, S. Pattamadilok, Y. Jee, W. Lim, W. Xu, K. Komase, M. Takeda, T. Tran, C. Castillo-Solorzano, P. Chenoweth, D. Brown, M. N. Mulders, W. J. Bellini, and D. Featherstone. 2011. Global distribution of measles genotypes and measles molecular epidemiology. *The Journal of infectious diseases* 204 Suppl 1: S514-523.
261. Yanagi, Y., M. Takeda, S. Ohno, and T. Hashiguchi. 2009. Measles virus receptors. *Current topics in microbiology and immunology* 329: 13-30.
262. Navaratnarajah, C. K., V. H. Leonard, and R. Cattaneo. 2009. Measles virus glycoprotein complex assembly, receptor attachment, and cell entry. *Current topics in microbiology and immunology* 329: 59-76.
263. Leonard, V. H., P. L. Sinn, G. Hodge, T. Miest, P. Devaux, N. Oezguen, W. Braun, P. B. McCray, Jr., M. B. McChesney, and R. Cattaneo. 2008. Measles virus blind to its epithelial cell receptor remains virulent in rhesus monkeys but cannot cross the airway epithelium and is not shed. *J Clin Invest* 118: 2448-2458.
264. Ludlow, M., L. J. Rennick, S. Sarlang, G. Skibinski, S. McQuaid, T. Moore, R. L. de Swart, and W. P. Duprex. 2010. Wild-type measles virus infection of primary epithelial cells occurs via the basolateral surface without syncytium formation or release of infectious virus. *The Journal of general virology* 91: 971-979.
265. Lemon, K., R. D. de Vries, A. W. Mesman, S. McQuaid, G. van Amerongen, S. Yuksel, M. Ludlow, L. J. Rennick, T. Kuiken, B. K. Rima, T. B. Geijtenbeek, A. D. Osterhaus, W. P. Duprex, and R. L. de Swart. 2011. Early target cells of measles virus after aerosol infection of non-human primates. *PLoS pathogens* 7: e1001263.
266. Bolt, G., K. Berg, and M. Blixenkrone-Moller. 2002. Measles virus-induced modulation of host-cell gene expression. *The Journal of general virology* 83: 1157-1165.
267. Borrow, P., and M. B. Oldstone. 1995. Measles virus-mono-nuclear cell interactions. *Current topics in microbiology and immunology* 191: 85-100.
268. Griffin, D. E., B. J. Ward, and L. M. Esolen. 1994. Pathogenesis of measles virus infection: an hypothesis for altered immune responses. *The Journal of infectious diseases* 170 Suppl 1: S24-31.
269. Black, F. L., and L. Rosen. 1962. Patterns of measles antibodies in residents of Tahiti and their stability in the absence of re-exposure. *Journal of immunology* 88: 725-731.
270. Hirsch, R. L., D. E. Griffin, R. T. Johnson, S. J. Cooper, I. Lindo de Soriano, S. Roedenbeck, and A. Vaisberg. 1984. Cellular immune responses during complicated and uncomplicated measles virus infections of man. *Clinical immunology and immunopathology* 31: 1-12.
271. Marie, J. C., J. Kehren, M. C. Trescol-Biemont, A. Evlashev, H. Valentin, T. Walzer, R. Tedone, B. Loveland, J. F. Nicolas, C. Roubourdin-Combe, and B. Horvat. 2001. Mechanism of measles virus-induced suppression of inflammatory immune responses. *Immunity* 14: 69-79.
272. Duke, T., and C. S. Mgone. 2003. Measles: not just another viral exanthem. *Lancet* 361: 763-773.
273. Butler, D. 2015. Measles by the numbers: A race to eradication. *Nature* 518: 148-149.
274. Griffin, D. E., and M. B. Oldstone. 2009. Measles. History and basic biology. Introduction. *Current topics in microbiology and immunology* 329: 1.
275. Griffin, D. E., and M. M. Oldstone. 2009. Measles. Pathogenesis and control. Introduction. *Current topics in microbiology and immunology* 330: 1.
276. Anlar, B., A. Ayhan, H. Hotta, M. Itoh, D. Engin, S. Barun, and O. Koseoglu. 2002. Measles virus RNA in tonsils of asymptomatic children. *Journal of paediatrics and child health* 38: 424-425.

277. Katayama, Y., K. Kohso, A. Nishimura, Y. Tatsuno, M. Homma, and H. Hotta. 1998. Detection of measles virus mRNA from autopsied human tissues. *Journal of clinical microbiology* 36: 299-301.
278. Kawashima, H., T. Mori, K. Takekuma, A. Hoshika, M. Hata, and T. Nakayama. 1996. Polymerase chain reaction detection of the hemagglutinin gene from an attenuated measles vaccine strain in the peripheral mononuclear cells of children with autoimmune hepatitis. *Archives of virology* 141: 877-884.
279. Sonoda, S., and T. Nakayama. 2001. Detection of measles virus genome in lymphocytes from asymptomatic healthy children. *Journal of medical virology* 65: 381-387.
280. Nanche, D. 2009. Human immunology of measles virus infection. *Current topics in microbiology and immunology* 330: 151-171.
281. Norrby, E., C. Orvell, B. Vandvik, and J. D. Cherry. 1981. Antibodies against measles virus polypeptides in different disease conditions. *Infection and immunity* 34: 718-724.
282. Stephenson, J. R., and V. ter Meulen. 1979. Antigenic relationships between measles and canine distemper viruses: comparison of immune response in animals and humans to individual virus-specific polypeptides. *Proceedings of the National Academy of Sciences of the United States of America* 76: 6601-6605.
283. de Swart, R. L., S. Yuksel, C. N. Langerijs, C. P. Muller, and A. D. Osterhaus. 2009. Depletion of measles virus glycoprotein-specific antibodies from human sera reveals genotype-specific neutralizing antibodies. *The Journal of general virology* 90: 2982-2989.
284. Bouche, F. B., O. T. Ertl, and C. P. Muller. 2002. Neutralizing B cell response in measles. *Viral immunology* 15: 451-471.
285. de Swart, R. L., S. Yuksel, and A. D. Osterhaus. 2005. Relative contributions of measles virus hemagglutinin- and fusion protein-specific serum antibodies to virus neutralization. *Journal of virology* 79: 11547-11551.
286. Tahara, M., Y. Ito, M. A. Brindley, X. Ma, J. He, S. Xu, H. Fukuhara, K. Sakai, K. Komase, P. A. Rota, R. K. Plemper, K. Maenaka, and M. Takeda. 2013. Functional and structural characterization of neutralizing epitopes of measles virus hemagglutinin protein. *Journal of virology* 87: 666-675.
287. Klingele, M., H. K. Hartter, F. Adu, W. Ammerlaan, W. Ikusika, and C. P. Muller. 2000. Resistance of recent measles virus wild-type isolates to antibody-mediated neutralization by vaccinees with antibody. *Journal of medical virology* 62: 91-98.
288. Tamin, A., P. A. Rota, Z. D. Wang, J. L. Heath, L. J. Anderson, and W. J. Bellini. 1994. Antigenic analysis of current wild type and vaccine strains of measles virus. *The Journal of infectious diseases* 170: 795-801.
289. Santibanez, S., S. Niewiesk, A. Heider, J. Schneider-Schaulies, G. A. Berbers, A. Zimmermann, A. Halenius, A. Wolbert, I. Deitemeier, A. Tischer, and H. Hengel. 2005. Probing neutralizing-antibody responses against emerging measles viruses (MVs): immune selection of MV by H protein-specific antibodies? *The Journal of general virology* 86: 365-374.
290. Schrag, S. J., P. A. Rota, and W. J. Bellini. 1999. Spontaneous mutation rate of measles virus: direct estimation based on mutations conferring monoclonal antibody resistance. *Journal of virology* 73: 51-54.
291. Feldman, H. A., A. Novack, and J. Warren. 1962. Inactivated measles virus vaccine. II. Prevention of natural and experimental measles with the vaccine. *JAMA: the journal of the American Medical Association* 179: 391-397.
292. Fulginiti, V. A., and C. H. Kempe. 1963. Measles Exposure among Vaccine Recipients. Response to Measles Exposure and Antibody Persistence among Recipients of Measles Vaccines. *American journal of diseases of children* 106: 450-461.

293. Nader, P. R., and R. J. Warren. 1968. Reported neurologic disorders following live measles vaccine. *Pediatrics* 41: 997-1001.
294. Griffin, D. E., and C. H. Pan. 2009. Measles: old vaccines, new vaccines. *Current topics in microbiology and immunology* 330: 191-212.
295. Enders, J. F., S. L. Katz, M. V. Milovanovic, and A. Holloway. 1960. Studies on an attenuated measles-virus vaccine. I. Development and preparations of the vaccine: technics for assay of effects of vaccination. *The New England journal of medicine* 263: 153-159.
296. Katz, S. L., M. V. Milovanovic, and J. F. Enders. 1958. Propagation of measles virus in cultures of chick embryo cells. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine* 97: 23-29.
297. Milovanovic, M. V., J. F. Enders, and A. Mitus. 1957. Cultivation of measles virus in human amnion cells and in developing chick embryo. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine* 95: 120-127.
298. Schwarz, A. J., and L. W. Zirbel. 1959. Propagation of measles virus in non-primate tissue culture. I. Propagation in bovine kidney tissue culture. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine* 102: 711-714.
299. Enders, J. F., S. L. Katz, and A. Holloway. 1962. Development of attenuated measles-virus vaccines. A summary of recent investigation. *American journal of diseases of children* 103: 335-340.
300. Katz, S. L., J. F. Enders, and A. Holloway. 1960. Studies on an attenuated measles-virus vaccine. II. Clinical, virologic and immunologic effects of vaccine in institutionalized children. *The New England journal of medicine* 263: 159-161.
301. Rota, J. S., Z. D. Wang, P. A. Rota, and W. J. Bellini. 1994. Comparison of sequences of the H, F, and N coding genes of measles virus vaccine strains. *Virus research* 31: 317-330.
302. Schwarz, A. J. 1962. Preliminary tests of a highly attenuated measles vaccine. *American journal of diseases of children* 103: 386-389.
303. Melnick, J. L. 1996. Thermostability of poliovirus and measles vaccines. *Developments in biological standardization* 87: 155-160.
304. Auwaerter, P. G., P. A. Rota, W. R. Elkins, R. J. Adams, T. DeLozier, Y. Shi, W. J. Bellini, B. R. Murphy, and D. E. Griffin. 1999. Measles virus infection in rhesus macaques: altered immune responses and comparison of the virulence of six different virus strains. *The Journal of infectious diseases* 180: 950-958.
305. van Binnendijk, R. S., R. W. van der Heijden, G. van Amerongen, F. G. UytdeHaag, and A. D. Osterhaus. 1994. Viral replication and development of specific immunity in macaques after infection with different measles virus strains. *The Journal of infectious diseases* 170: 443-448.
306. Ovsyannikova, I. G., K. C. Reid, R. M. Jacobson, A. L. Oberg, G. G. Klee, and G. A. Poland. 2003. Cytokine production patterns and antibody response to measles vaccine. *Vaccine* 21: 3946-3953.
307. Dhiman, N., I. G. Ovsyannikova, N. A. Pinsky, R. A. Vierkant, S. J. Jacobsen, R. M. Jacobson, and G. A. Poland. 2003. Lack of association between transporter associated with antigen processing (TAP) and HLA-DM gene polymorphisms and antibody levels following measles vaccination. *European journal of immunogenetics : official journal of the British Society for Histocompatibility and Immunogenetics* 30: 195-200.
308. Christenson, B., and M. Bottiger. 1994. Measles antibody: comparison of long-term vaccination titres, early vaccination titres and naturally acquired immunity to and booster effects on the measles virus. *Vaccine* 12: 129-133.
309. Nanche, D., M. Garenne, C. Rae, M. Manchester, R. Buchta, S. K. Brodine, and M. B. Oldstone. 2004. Decrease in measles virus-specific CD4 T cell memory in vaccinated subjects. *The Journal of infectious diseases* 190: 1387-1395.

310. Gay, N. J. 2004. The theory of measles elimination: implications for the design of elimination strategies. *The Journal of infectious diseases* 189 Suppl 1: S27-35.
311. Chakravarti, A. 2005. Measles control: current trends & recommendations. *The Indian journal of medical research* 121: 73-76.
312. Krugman, S., G. Muriel, and V. J. Fontana. 1971. Combined live measles, mumps, rubella vaccine. Immunological response. *American journal of diseases of children* 121: 380-381.
313. Clements, C. J., M. Strassburg, F. T. Cutts, and C. Trel. 1992. The epidemiology of measles. *World health statistics quarterly. Rapport trimestriel de statistiques sanitaires mondiales* 45: 285-291.
314. Gustafson, T. L., A. W. Lievens, P. A. Brunell, R. G. Moellenberg, C. M. Buttery, and L. M. Schulster. 1987. Measles outbreak in a fully immunized secondary-school population. *The New England journal of medicine* 316: 771-774.
315. Pedersen, I. R., C. H. Mordhorst, G. Glikmann, and H. von Magnus. 1989. Subclinical measles infection in vaccinated seropositive individuals in arctic Greenland. *Vaccine* 7: 345-348.
316. Huiss, S., B. Damien, F. Schneider, and C. P. Muller. 1997. Characteristics of asymptomatic secondary immune responses to measles virus in late convalescent donors. *Clinical and experimental immunology* 109: 416-420.
317. Damien, B., S. Huiss, F. Schneider, and C. P. Muller. 1998. Estimated susceptibility to asymptomatic secondary immune response against measles in late convalescent and vaccinated persons. *Journal of medical virology* 56: 85-90.
318. Lievano, F. A., M. J. Papania, R. F. Helfand, R. Harpaz, L. Walls, R. S. Katz, I. Williams, Y. S. Villamarzo, P. A. Rota, and W. J. Bellini. 2004. Lack of evidence of measles virus shedding in people with inapparent measles virus infections. *The Journal of infectious diseases* 189 Suppl 1: S165-170.
319. Mossong, J., D. J. Nokes, W. J. Edmunds, M. J. Cox, S. Ratnam, and C. P. Muller. 1999. Modeling the impact of subclinical measles transmission in vaccinated populations with waning immunity. *American journal of epidemiology* 150: 1238-1249.
320. Sanders, R., A. Dabbagh, and D. Featherstone. 2011. Risk analysis for measles reintroduction after global certification of eradication. *The Journal of infectious diseases* 204 Suppl 1: S71-77.
321. Nissim, A., and Y. Chernajovsky. 2008. Historical development of monoclonal antibody therapeutics. *Handbook of experimental pharmacology*: 3-18.
322. Yu, X., P. A. McGraw, F. S. House, and J. E. Crowe, Jr. 2008. An optimized electrofusion-based protocol for generating virus-specific human monoclonal antibodies. *Journal of immunological methods* 336: 142-151.
323. Stahli, C., T. Staehelin, V. Miggiano, J. Schmidt, and P. Haring. 1980. High frequencies of antigen-specific hybridomas: dependence on immunization parameters and prediction by spleen cell analysis. *Journal of immunological methods* 32: 297-304.
324. Andersson, J., and F. Melchers. 1978. The antibody repertoire of hybrid cell lines obtained by fusion of X63-AG8 myeloma cells with mitogen-activated B-cell blasts. *Current topics in microbiology and immunology* 81: 130-139.
325. Kohler, G., and M. J. Shulman. 1978. Cellular and molecular restrictions of the lymphocyte fusion. *Current topics in microbiology and immunology* 81: 143-148.
326. Chan, A. C., and P. J. Carter. 2010. Therapeutic antibodies for autoimmunity and inflammation. *Nature reviews. Immunology* 10: 301-316.
327. Enever, C., T. Batuwangala, C. Plummer, and A. Sepp. 2009. Next generation immunotherapeutics—honing the magic bullet. *Current opinion in biotechnology* 20: 405-411.
328. Bruggemann, M., N. P. Davies, and I. R. Rosewell. 1993. Designer mice: the production of human antibody repertoires in transgenic animals. *The Year in immunology* 7: 33-40.

329. Bruggemann, M., and M. J. Taussig. 1997. Production of human antibody repertoires in transgenic mice. *Current opinion in biotechnology* 8: 455-458.
330. Jakobovits, A., R. G. Amado, X. Yang, L. Roskos, and G. Schwab. 2007. From XenoMouse technology to panitumumab, the first fully human antibody product from transgenic mice. *Nature biotechnology* 25: 1134-1143.
331. Geurts, A. M., G. J. Cost, Y. Freyvert, B. Zeitler, J. C. Miller, V. M. Choi, S. S. Jenkins, A. Wood, X. Cui, X. Meng, A. Vincent, S. Lam, M. Michalkiewicz, R. Schilling, J. Foeckler, S. Kalloway, H. Weiler, S. Menoret, I. Anegon, G. D. Davis, L. Zhang, E. J. Rebar, P. D. Gregory, F. D. Urnov, H. J. Jacob, and R. Buelow. 2009. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* 325: 433.
332. Menoret, S., A. L. Iscache, L. Tesson, S. Remy, C. Usal, M. J. Osborn, G. J. Cost, M. Bruggemann, R. Buelow, and I. Anegon. 2010. Characterization of immunoglobulin heavy chain knockout rats. *European journal of immunology* 40: 2932-2941.
333. Muellenbeck, M. F., B. Ueberheide, B. Amulic, A. Epp, D. Fenyó, C. E. Busse, M. Esen, M. Theisen, B. Mordmüller, and H. Wardemann. 2013. Atypical and classical memory B cells produce *Plasmodium falciparum* neutralizing antibodies. *The Journal of experimental medicine* 210: 389-399.
334. National Research Council (U.S.). Committee for the Update of the Guide for the Care and Use of Laboratory Animals., Institute for Laboratory Animal Research (U.S.), and National Academies Press (U.S.). 2011. Guide for the care and use of laboratory animals. 8th ed. National Academies Press., Washington, D.C. xxv, 220 p.
335. Kishiro, Y., M. Kagawa, I. Naito, and Y. Sado. 1995. A novel method of preparing rat-monoclonal antibody-producing hybridomas by using rat medial iliac lymph node cells. *Cell structure and function* 20: 151-156.
336. Ye, J., N. Ma, T. L. Madden, and J. M. Ostell. 2013. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic acids research* 41: W34-40.
337. Bashford-Rogers, R., A. Palser, B. Huntly, R. Rance, G. Vassiliou, G. Follows, and P. Kellam. 2013. Network properties derived from deep sequencing of the human B-cell receptor repertoires delineates B-cell populations. *Genome research* 23: 1874-1884.
338. Bragg, L. M., G. Stone, M. K. Butler, P. Hugenholtz, and G. W. Tyson. 2013. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS computational biology* 9: e1003031.
339. Ecker, D. M., S. D. Jones, and H. L. Levine. 2015. The therapeutic monoclonal antibody market. *mAbs* 7: 9-14.
340. Xiao, M., P. Prabakaran, W. Chen, B. Kessing, and D. S. Dimitrov. 2013. Deep sequencing and Circos analyses of antibody libraries reveal antigen-driven selection of Ig VH genes during HIV-1 infection. *Experimental and molecular pathology* 95: 357-363.
341. Cortina-Ceballos, B., E. E. Godoy-Lozano, H. Samano-Sanchez, A. Aguilar-Salgado, C. Velasco-Herrera Mdel, C. Vargas-Chavez, D. Velazquez-Ramirez, G. Romero, J. Moreno, J. Tellez-Sosa, and J. Martinez-Barnette. 2015. Reconstructing and mining the B cell repertoire with ImmuneDiversity. *mAbs* 7: 516-524.
342. Grova, N., E. J. Prodhomme, M. T. Schellenberger, S. Farinelle, and C. P. Muller. 2009. Modulation of carcinogen bioavailability by immunisation with benzo[a]pyrene-conjugate vaccines. *Vaccine* 27: 4142-4151.
343. Schellenberger, M. T., N. Grova, S. Willieme, S. Farinelle, E. J. Prodhomme, and C. P. Muller. 2009. Modulation of benzo[a]pyrene induced immunotoxicity in mice actively immunized with a B[a]P-diphtheria toxoid conjugate. *Toxicology and applied pharmacology* 240: 37-45.



344. Schellenberger, M. T., N. Grova, S. Farinelle, S. Willieme, D. Revets, and C. P. Muller. 2012. Immunogenicity of a promiscuous T cell epitope peptide based conjugate vaccine against benzo[a]pyrene: redirecting antibodies to the hapten. *PLoS One* 7: e38329.
345. Carroll, M. W., and B. Moss. 1997. Host range and cytopathogenicity of the highly attenuated MVA strain of vaccinia virus: propagation and generation of recombinant viruses in a nonhuman mammalian cell line. *Virology* 238: 198-211.
346. Drexler, I., K. Heller, B. Wahren, V. Erfle, and G. Sutter. 1998. Highly attenuated modified vaccinia virus Ankara replicates in baby hamster kidney cells, a potential host for virus propagation, but not in various human transformed and primary cells. *The Journal of general virology* 79 ( Pt 2): 347-352.
347. Staib, C., and G. Sutter. 2003. Live viral vectors: vaccinia virus. *Methods in molecular medicine* 87: 51-68.
348. Earl, P. L., B. Moss, L. S. Wyatt, and M. W. Carroll. 2001. Generation of recombinant vaccinia viruses. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* Chapter 5: Unit 5.13.
349. Stittelaar, K. J., L. S. Wyatt, R. L. de Swart, H. W. Vos, J. Groen, G. van Amerongen, R. S. van Binnendijk, S. Rozenblatt, B. Moss, and A. D. Osterhaus. 2000. Protective immunity in macaques vaccinated with a modified vaccinia virus Ankara-based measles virus vaccine in the presence of passively acquired antibodies. *Journal of virology* 74: 4236-4243.
350. Sutter, G., and B. Moss. 1992. Nonreplicating vaccinia vector efficiently expresses recombinant genes. *Proceedings of the National Academy of Sciences of the United States of America* 89: 10847-10851.
351. Moss, B. 1996. Genetically engineered poxviruses for recombinant gene expression, vaccination, and safety. *Proceedings of the National Academy of Sciences of the United States of America* 93: 11341-11348.
352. Schneider, J., S. C. Gilbert, T. J. Blanchard, T. Hanke, K. J. Robson, C. M. Hannan, M. Becker, R. Sinden, G. L. Smith, and A. V. Hill. 1998. Enhanced immunogenicity for CD8+ T cell induction and complete protective efficacy of malaria DNA vaccination by boosting with modified vaccinia virus Ankara. *Nature medicine* 4: 397-402.
353. Moss, B., M. W. Carroll, L. S. Wyatt, J. R. Bennink, V. M. Hirsch, S. Goldstein, W. R. Elkins, T. R. Fuerst, J. D. Lifson, M. Piatak, N. P. Restifo, W. Overwijk, R. Chamberlain, S. A. Rosenberg, and G. Sutter. 1996. Host range restricted, non-replicating vaccinia virus vectors as vaccine candidates. *Advances in experimental medicine and biology* 397: 7-13.
354. Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner. 2004. WebLogo: a sequence logo generator. *Genome research* 14: 1188-1190.
355. Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra. 2009. Circos: an information aesthetic for comparative genomics. *Genome research* 19: 1639-1645.
356. Venturi, V., K. Kedzierska, D. A. Price, P. C. Doherty, D. C. Douek, S. J. Turner, and M. P. Davenport. 2006. Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proceedings of the National Academy of Sciences of the United States of America* 103: 18691-18696.
357. Quigley, M. F., H. Y. Greenaway, V. Venturi, R. Lindsay, K. M. Quinn, R. A. Seder, D. C. Douek, M. P. Davenport, and D. A. Price. 2010. Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. *Proceedings of the National Academy of Sciences of the United States of America* 107: 19414-19419.
358. Venturi, V., M. F. Quigley, H. Y. Greenaway, P. C. Ng, Z. S. Ende, T. McIntosh, T. E. Asher, J. R. Almeida, S. Levy, D. A. Price, M. P. Davenport, and D. C. Douek. 2011. A mechanism for TCR

- sharing between T cell subsets and individuals revealed by pyrosequencing. *Journal of immunology* 186: 4285-4294.
359. Warren, R. L., J. D. Freeman, T. Zeng, G. Choe, S. Munro, R. Moore, J. R. Webb, and R. A. Holt. 2011. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome research* 21: 790-797.
360. Poulsen, T. R., P. J. Meijer, A. Jensen, L. S. Nielsen, and P. S. Andersen. 2007. Kinetic, affinity, and diversity limits of human polyclonal antibody responses against tetanus toxoid. *Journal of immunology* 179: 3841-3850.
361. Frolich, D., C. Giesecke, H. E. Mei, K. Reiter, C. Daridon, P. E. Lipsky, and T. Dorner. 2010. Secondary immunization generates clonally related antigen-specific plasma cells and memory B cells. *Journal of immunology* 185: 3103-3110.
362. Poulsen, T. R., A. Jensen, J. S. Haurum, and P. S. Andersen. 2011. Limits for antibody affinity maturation and repertoire diversification in hypervaccinated humans. *Journal of immunology* 187: 4229-4235.
363. de Kruif, J., A. Kramer, T. Visser, C. Clements, R. Nijhuis, F. Cox, V. van der Zande, R. Smit, D. Pinto, M. Throsby, and T. Logtenberg. 2009. Human immunoglobulin repertoires against tetanus toxoid contain a large and diverse fraction of high-affinity promiscuous V(H) genes. *Journal of molecular biology* 387: 548-558.
364. Darzentas, N., and K. Stamatopoulos. 2013. Stereotyped B cell receptors in B cell leukemias and lymphomas. *Methods in molecular biology* 971: 135-148.
365. Liu, M. A. 2003. DNA vaccines: a review. *Journal of internal medicine* 253: 402-410.
366. Rajcani, J., T. Mosko, and I. Rezuchova. 2005. Current developments in viral DNA vaccines: shall they solve the unsolved? *Reviews in medical virology* 15: 303-325.
367. Li, L., F. Saade, and N. Petrovsky. 2012. The future of human DNA vaccines. *Journal of biotechnology* 162: 171-182.
368. Ferraro, B., M. P. Morrow, N. A. Hutnick, T. H. Shin, C. E. Lucke, and D. B. Weiner. 2011. Clinical applications of DNA vaccines: current progress. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 53: 296-302.
369. Davidson, A. H., J. L. Traub-Dargatz, R. M. Rodeheaver, E. N. Ostlund, D. D. Pedersen, R. G. Moorhead, J. B. Stricklin, R. D. Dewell, S. D. Roach, R. E. Long, S. J. Albers, R. J. Callan, and M. D. Salman. 2005. Immunologic responses to West Nile virus in vaccinated and clinically affected horses. *Journal of the American Veterinary Medical Association* 226: 240-245.
370. Garver, K. A., S. E. LaPatra, and G. Kurath. 2005. Efficacy of an infectious hematopoietic necrosis (IHN) virus DNA vaccine in Chinook *Oncorhynchus tshawytscha* and sockeye *O. nerka* salmon. *Diseases of aquatic organisms* 64: 13-22.
371. Thacker, E. L., D. J. Holtkamp, A. S. Khan, P. A. Brown, and R. Draghia-Akli. 2006. Plasmid-mediated growth hormone-releasing hormone efficacy in reducing disease associated with *Mycoplasma hyopneumoniae* and porcine reproductive and respiratory syndrome virus infection. *Journal of animal science* 84: 733-742.
372. Klinman, D. M., S. Klaschik, D. Tross, H. Shirota, and F. Steinhagen. 2010. FDA guidance on prophylactic DNA vaccines: analysis and recommendations. *Vaccine* 28: 2801-2805.
373. Xiang, S. D., C. Selomulya, J. Ho, V. Apostolopoulos, and M. Plebanski. 2010. Delivery of DNA vaccines: an overview on the use of biodegradable polymeric and magnetic nanoparticles. *Wiley interdisciplinary reviews. Nanomedicine and nanobiotechnology* 2: 205-218.

374. Arnaoty, A., V. Gouilleux-Gruart, S. Casteret, B. Pitard, Y. Bigot, and T. Lecomte. 2013. Reliability of the nanopheres-DNA immunization technology to produce polyclonal antibodies directed against human neogenic proteins. *Molecular genetics and genomics: MGG* 288: 347-363.
375. Shah, M. A., S. U. Khan, Z. Ali, H. Yang, K. Liu, and L. Mao. 2014. Applications of nanoparticles for DNA based rabies vaccine. *Journal of nanoscience and nanotechnology* 14: 881-891.
376. Rengarajan, K., S. M. Cristol, M. Mehta, and J. M. Nickerson. 2002. Quantifying DNA concentrations using fluorometry: a comparison of fluorophores. *Molecular vision* 8: 416-421.
377. Zhang, Y., Z. Ding, H. Wang, L. Li, Y. Pang, K. E. Brown, S. Xu, Z. Zhu, P. A. Rota, and D. Featherstone. 2010. New measles virus genotype associated with outbreak, China. *Emerging infectious diseases* 16: 943.
378. Itoh, M., Y. Okuno, and H. Hotta. 2002. Comparative analysis of titers of antibody against measles virus in sera of vaccinated and naturally infected Japanese individuals of different age groups. *Journal of clinical microbiology* 40: 1733-1738.
379. Ertl, O. T., D. C. Wenz, F. B. Bouche, G. A. Berbers, and C. P. Muller. 2003. Immunodominant domains of the Measles virus hemagglutinin protein eliciting a neutralizing human B cell response. *Archives of virology* 148: 2195-2206.
380. Berzofsky, J. A. 1985. Intrinsic and extrinsic factors in protein antigenic structure. *Science* 229: 932-940.
381. Corti, D., J. Voss, S. J. Gamblin, G. Codoni, A. Macagno, D. Jarrossay, S. G. Vachieri, D. Pinna, A. Minola, F. Vanzetta, C. Silacci, B. M. Fernandez-Rodriguez, G. Agatic, S. Bianchi, I. Giacchetto-Sasselli, L. Calder, F. Sallusto, P. Collins, L. F. Haire, N. Temperton, J. P. Langedijk, J. J. Skehel, and A. Lanzavecchia. 2011. A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* 333: 850-856.
382. Wardemann, H., S. Yurasov, A. Schaefer, J. W. Young, E. Meffre, and M. C. Nussenzweig. 2003. Predominant autoantibody production by early human B cell precursors. *Science* 301: 1374-1377.
383. Amara, K., J. Steen, F. Murray, H. Morbach, B. M. Fernandez-Rodriguez, V. Joshua, M. Engstrom, O. Snir, L. Israelsson, A. I. Catrina, H. Wardemann, D. Corti, E. Meffre, L. Klareskog, and V. Malmstrom. 2013. Monoclonal IgG antibodies generated from joint-derived B cells of RA patients have a strong bias toward citrullinated autoantigen recognition. *The Journal of experimental medicine* 210: 445-455.
384. Traggiai, E., S. Becker, K. Subbarao, L. Kolesnikova, Y. Uematsu, M. R. Gismondo, B. R. Murphy, R. Rappuoli, and A. Lanzavecchia. 2004. An efficient method to make human monoclonal antibodies from memory B cells: potent neutralization of SARS coronavirus. *Nature medicine* 10: 871-875.
385. Wrammert, J., K. Smith, J. Miller, W. A. Langley, K. Kokko, C. Larsen, N. Y. Zheng, I. Mays, L. Garman, C. Helms, J. James, G. M. Air, J. D. Capra, R. Ahmed, and P. C. Wilson. 2008. Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* 453: 667-671.
386. Yu, X., T. Tsibane, P. A. McGraw, F. S. House, C. J. Keefer, M. D. Hicar, T. M. Tumpey, C. Pappas, L. A. Perrone, O. Martinez, J. Stevens, I. A. Wilson, P. V. Aguilar, E. L. Altschuler, C. F. Basler, and J. E. Crowe, Jr. 2008. Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature* 455: 532-536.
387. Walker, L. M., S. K. Phogat, P. Y. Chan-Hui, D. Wagner, P. Phung, J. L. Goss, T. Wrin, M. D. Simek, S. Fling, J. L. Mitcham, J. K. Lehrman, F. H. Priddy, O. A. Olsen, S. M. Frey, P. W. Hammond, G. P. I. Protocol, S. Kaminsky, T. Zamb, M. Moyle, W. C. Koff, P. Poirnard, and D. R. Burton. 2009. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* 326: 285-289.

388. Scheid, J. F., H. Mouquet, N. Feldhahn, M. S. Seaman, K. Velinzon, J. Pietzsch, R. G. Ott, R. M. Anthony, H. Zebroski, A. Hurley, A. Phogat, B. Chakrabarti, Y. Li, M. Connors, F. Pereyra, B. D. Walker, H. Wardemann, D. Ho, R. T. Wyatt, J. R. Mascola, J. V. Ravetch, and M. C. Nussenzweig. 2009. Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* 458: 636-640.
389. Kwong, P. D., and J. R. Mascola. 2012. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity* 37: 412-425.
390. Haynes, B. F., G. Kelsoe, S. C. Harrison, and T. B. Kepler. 2012. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nature biotechnology* 30: 423-433.
391. Ozawa, T., H. Kishi, and A. Muraguchi. 2006. Amplification and analysis of cDNA generated from a single cell by 5'-RACE: application to isolation of antibody heavy and light chain variable gene sequences from single B cells. *BioTechniques* 40: 469-470, 472, 474 passim.
392. Michaeli, M., H. Noga, H. Tabibian-Keissar, I. Barshack, and R. Mehr. 2012. Automated cleaning and pre-processing of immunoglobulin gene sequences from high-throughput sequencing. *Frontiers in immunology* 3: 386.
393. HAZANOV, H., M. MICHAELI, G. LAVY-SHAHAF, and R. MEHR. 2014. INFORMATIC TOOLS FOR IMMUNOGLOBULIN GENE SEQUENCE ANALYSIS. *Comparative Immunoglobulin Genetics*: 223.
394. Mroczek, E. S., G. C. Ippolito, T. Rogosch, K. H. Hoi, T. A. Hwangpo, M. G. Brand, Y. Zhuang, C. R. Liu, D. A. Schneider, M. Zemlin, E. E. Brown, G. Georgiou, and H. W. Schroeder, Jr. 2014. Differences in the composition of the human antibody repertoire by B cell subsets in the blood. *Frontiers in immunology* 5: 96.
395. Caron, G., S. Le Gallou, T. Lamy, K. Tarte, and T. Fest. 2009. CXCR4 expression functionally discriminates centroblasts versus centrocytes within human germinal center B cells. *Journal of immunology* 182: 7595-7602.
396. Stahl, D., S. Lacroix-Desmazes, L. Mouthon, S. V. Kaveri, and M. D. Kazatchkine. 2000. Analysis of human self-reactive antibody repertoires by quantitative immunoblotting. *Journal of immunological methods* 240: 1-14.
397. Vaz, N. M. 2000. Natural immunoglobulins (contribution to a debate on biomedical education). *Memórias do Instituto Oswaldo Cruz* 95: 59-62.
398. Quintana, F. J., P. H. Hagedorn, G. Elizur, Y. Merbl, E. Domany, and I. R. Cohen. 2004. Functional immunomics: microarray analysis of IgG autoantibody repertoires predicts the future response of mice to induced diabetes. *Proceedings of the National Academy of Sciences of the United States of America* 101 Suppl 2: 14615-14621.
399. Quintana, F. J., G. Getz, G. Hed, E. Domany, and I. R. Cohen. 2003. Cluster analysis of human autoantibody reactivities in health and in type 1 diabetes mellitus: a bio-informatic approach to immune complexity. *Journal of autoimmunity* 21: 65-75.
400. Hoi, K. H., and G. C. Ippolito. 2013. Intrinsic bias and public rearrangements in the human immunoglobulin Vlambda light chain repertoire. *Genes and immunity* 14: 271-276.
401. Ippolito, G. C., K. H. Hoi, S. T. Reddy, S. M. Carroll, X. Ge, T. Rogosch, M. Zemlin, L. D. Shultz, A. D. Ellington, and C. L. VanDenBerg. 2012. Antibody repertoires in humanized NOD-scid-IL2R  $\gamma$  null mice and human B cells reveals human-like diversification and tolerance checkpoints in the mouse. *PloS one* 7: e35497.
402. Chen, W., P. Prabakaran, Z. Zhu, Y. Feng, E. D. Streaker, and D. S. Dimitrov. 2012. Characterization of human IgG repertoires in an acute HIV-1 infection. *Experimental and molecular pathology* 93: 399-407.

- 
403. Prabakaran, P., W. Chen, and D. S. Dimitrov. 2014. The Antibody Germline/Maturation Hypothesis, Elicitation of Broadly Neutralizing Antibodies Against HIV-1 and Cord Blood IgM Repertoires. *Frontiers in immunology* 5: 398.
  404. Sutton, L. A., E. Kostareli, E. Stalika, A. Tsiftaris, A. Anagnostopoulos, N. Darzentas, R. Rosenquist, and K. Stamatopoulos. 2013. Temporal dynamics of clonal evolution in chronic lymphocytic leukemia with stereotyped IGHV4-34/IGKV2-30 antigen receptors: longitudinal immunogenetic evidence. *Molecular medicine* 19: 230-236.
  405. Darzentas, N., and K. Stamatopoulos. 2013. The significance of stereotyped B-cell receptors in chronic lymphocytic leukemia. *Hematology/oncology clinics of North America* 27: 237-250.
  406. Messmer, B. T., E. Albesiano, D. G. Efremov, F. Ghiotto, S. L. Allen, J. Kolitz, R. Foa, R. N. Damle, F. Fais, D. Messmer, K. R. Rai, M. Ferrarini, and N. Chiorazzi. 2004. Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *The Journal of experimental medicine* 200: 519-525.