Doctoral Thesis in Biotechnology

On the analysis of antibody repertoires

AUGUST JERNBOM FALK

Stockholm, Sweden 2023
On the analysis of antibody repertoires

AUGUST JERNBOM FALK

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Doctor of Philosophy on Friday the 10th of November 2023, at 09:30 in Air & Fire, SciLifeLab, Tomtebodavägen 23A, Stockholm.

Doctoral Thesis in Biotechnology
KTH Royal Institute of Technology
Stockholm, Sweden 2023
Till mormor,
som lärt mig värdet av bildning.
Till morfar,
som lärt mig att fiska.
Abstract

The antibody repertoire is the ensemble of antibodies found in an individual at a given time. It displays high heterogeneity between individuals while being both largely temporally stable within an individual and rapidly responsive to immunological challenge. As distinct collections of antibodies within the repertoire contribute to the function and malfunction of the immune system, studying the many aspects of the antibody repertoire can give increased knowledge on antibody-mediated pathogen defense and autoimmune conditions.

There are several emergent techniques for assessing different properties of the antibody repertoire as well as determining distinct antibodies of interest in health and disease. The studies presented in this thesis use planar and bead-based arrays to investigate parts of the antibody repertoire consisting of antibodies against SARS-CoV-2 proteins in serological studies, as well as autoantibodies against the large collection of antigens in the Human Protein Atlas. **Paper I** explores the autoantibody repertoires of patients with psychosis using planar arrays of 42,000 antigens followed by targeted bead arrays and identifies associations to specific symptoms. **Paper II** defines the baseline serological characteristics of a longitudinal cohort using a then recently developed multiplex serological assay and gives an early description of COVID-19 symptomatology. **Paper III** investigates the four-month persistence and antigen diversity of antibodies against SARS-CoV-2 following infection. This work is continued in **Paper IV** which examines the persistence of the humoral and cellular response to infection and their protective effect against reinfection. **Paper V** connects these parts by exploring the autoantibody repertoire of this longitudinal cohort and identifying new-onset autoantibodies emerging at infection using arrays of human and viral antigens. It associates three new-onset autoantibodies to post-COVID-19 symptoms and demonstrates sequence similarity between human and viral epitopes, which may indicate molecular mimicry.

Antibody repertoires are heterogeneous and multifaceted, requiring several methods for full comprehension. The present investigation encompasses the analysis of one facet using antigen arrays and contributes to knowledge on disease-associated autoantibody repertoires as well as the prevalence and persistence of the serological and autoantibody response emerging after viral infection. This work represents a small step towards the goal of understanding the full repertoire complexity. Emergent large-scale techniques combined with the herein described analysis are together poised to identify clinically relevant antigens and advance knowledge on the diversity and heterogeneity of the antibody repertoire.
Sammanfattning

Antikroppsrepertoaren utgörs av den samling av antikroppar som återfinns i en individ vid ett givet tillfälle. Den uppvissar stor heterogenitet mellan individer samtidigt som den inom en individ både är övervägande stabil över tid och snabbt föränderlig vid immunologiska händelser. Eftersom avgränsade samlingar av antikroppar inom repertoaren bidrar till immunförsvarets funktion och dysfunktion är det av stor vikt att studera de många olika aspekterna av antikroppsrepertoaren för att öka förståelsen av både det försvaret mot patogen och de autoimmuna tillstånd som tillkommer genom antikroppars verkan.

Det finns flera banbrytande tekniker som utvecklats för att undersöka olika aspekter av antikroppsrepertoaren samt identifiera särskilda antikroppar som kan bidra till kunskap om friska och sjuka tillstånd. De forskningsarbeten som presenteras i den här avhandlingen använde sig av analysmetoder som grundar sig på plana ytor och mikrosfärer för att undersöka olika delar av antikroppsrepertoaren. Dessa delar bestod av antikroppar mot proteiner hos SARS-CoV-2 som undersöktes i serologiska arbeten, samt autoantikroppar mot den stora samling av antigen som finns i HPA – atlenas över människans proteiner.


Antikroppsrepertoarer är heterogena och mångfasetterade och kräver därför flera metoder för full förståelse. De forskningsarbeten som förs fram i den här avhandlingen omfattar analys av en fasett med hjälp av antigenbaserade analysmetoder. Dessa arbeten bidrar till kunskap om sjukdomskopplade autoantikroppsrepertoarer samt förekomsten och varaktigheten hos det serologiska svaret och autoantikroppar efter virusinfektion. Arbetet representerar ett litet steg mot det slutgiltiga målet att förstå helheten av repertoarens komplexitet. Banbrytande storskaliga tekniker i kombination med den analys som beskrivs i den här avhandlingen har stor potential att identifiera kliniskt relevanta antigen och ge ökad kunskap om antikroppsrepertoarens mångfald och heterogenitet.
Thesis defense

This thesis will be defended on the 10th of November 2023 at 09:30 in the room Air & Fire at SciLifeLab, Tomtebodavägen 23A, Stockholm, for the degree of Teknologie doktor, Doctor of Philosophy (PhD) in Biotechnology.

Respondent
August Jernbom Falk, MSc in Biotechnology
Division of Affinity Proteomics, Department of Protein Science,
KTH Royal Institute of Technology, SciLifeLab, Stockholm, Sweden

Faculty opponent
Prof. Albert J.R. Heck
Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, the Netherlands

Evaluation committee
Doc. Caroline Grönwall
Division of Rheumatology, Department of Medicine,
Karolinska Institutet, Stockholm, Sweden

Doc. Christopher Sundling
Division of Infectious Diseases, Department of Medicine,
Karolinska Institutet, Stockholm, Sweden

Prof. Lukas Käll
Department of Gene Technology,
KTH Royal Institute of Technology, SciLifeLab, Stockholm, Sweden

Chairman
Prof. Jochen M. Schwenk
Division of Affinity Proteomics, Department of Protein Science,
KTH Royal Institute of Technology, SciLifeLab, Stockholm, Sweden

Respondent’s main supervisor
Prof. Peter Nilsson
Division of Affinity Proteomics, Department of Protein Science,
KTH Royal Institute of Technology, SciLifeLab, Stockholm, Sweden

Respondent’s co-supervisor
Dr. Anna Månberg
Division of Affinity Proteomics, Department of Protein Science,
KTH Royal Institute of Technology, SciLifeLab, Stockholm, Sweden
List of publications and manuscripts

The presented thesis is based on the following five papers. All papers are included in the appendices of the thesis.

**Paper I**  
*Autoantibody profiles associated with clinical features in psychotic disorders*  
*Translational Psychiatry* **11** (2021). doi:10.1038/s41398-021-01596-0

**Paper II**  
*SARS-CoV-2 exposure, symptoms and seroprevalence in healthcare workers in Sweden*  

**Paper III**  
*SARS-CoV-2 induces a durable and antigen specific humoral immunity after asymptomatic to mild COVID-19 infection*  

**Paper IV**  
*Robust humoral and cellular immune responses and low risk for reinfection at least 8 months following asymptomatic to mild COVID-19*  

**Paper V**  
*Prevalent and persistent new-onset autoantibodies in mild to severe COVID-19*  
*Manuscript in preparation*

# These authors contributed equally to this work.
Respondent’s contribution to the included work

**Paper I** Main responsible for experimental work and data processing, analysis, and visualization. Joint responsible for interpretation and manuscript writing.

**Paper II** Shared main responsible for processing and analysis of serological data. Main responsible for data integration and visualization. Contributed to interpretation and manuscript drafting.

**Paper III** Shared main responsible for processing and analysis of serological data. Main responsible for data integration and visualization. Joint responsible for interpretation and manuscript writing.

**Paper IV** Shared main responsible for processing and analysis of serological data, data integration, and visualization. Contributed to interpretation and manuscript drafting.

**Paper V** Main responsible for study design, experimental work, and data processing, analysis, and visualization. Main responsible for interpretation and manuscript writing.
Related publications not included in the thesis


**Array-Based Profiling of Proteins and Autoantibody Repertoires in CSF,**


**Systematic evaluation of SARS-CoV-2 antigens enables a highly specific and sensitive multiplex serological COVID-19 assay**


**High Amounts of SARS-CoV-2 Precede Sickness Among Asymptomatic Health Care Workers**


**Antibodies to SARS-CoV-2 and risk of past or future sick leave**

**An evaluation of a FluoroSpot assay as a diagnostic tool to determine SARS-CoV-2 specific T cell responses**


**Multiomics Profiling of Alzheimer's Disease Serum for the Identification of Autoantibody Biomarkers**

*Journal of Proteome Research* (2021). doi:10.1021/acs.jproteome.1c00630


**Long-term SARS-CoV-2-specific and cross-reactive cellular immune responses correlate with humoral responses, disease severity, and symptomatology**


**Comparative serum proteomic analysis of a selected protein panel in individuals with schizophrenia and bipolar disorder and the impact of genetic risk burden on serum proteomic profiles**

*Translational Psychiatry* **12** (2022). doi:10.1038/s41398-022-02228-x
Contents

ABSTRACT ........................................................................................................................................ I
SAMMANFATTNING .................................................................................................................. II
THESIS DEFENSE ................................................................................................................... III
LIST OF PUBLICATIONS AND MANUSCRIPTS ................................................................. IV
RESPONDENT'S CONTRIBUTION TO THE INCLUDED WORK ........................................ V
RELATED PUBLICATIONS NOT INCLUDED IN THE THESIS .................................. VI
CONTENTS .................................................................................................................................. VIII

CHAPTER I: THE ANTIBODY REPERTOIRE ............................................................................ 1
  SEROLOGY – THE STUDY OF ANTIBODIES .............................................................................. 1
  FOUNDATIONS OF ANTIBODY REPERTOIRE DIVERSITY ...................................................... 2
    Antibody structure .................................................................................................................... 2
    B cell development .................................................................................................................. 4
  REGULATING THE DIVERSITY OF THE ANTIBODY REPERTOIRE .......................................... 6
    Induction of tolerance ............................................................................................................. 6
    Natural antibodies .................................................................................................................. 7
    Autoantibodies ....................................................................................................................... 8
  TOWARDS ANALYSIS OF THE ANTIBODY REPERTOIRE ....................................................... 10

CHAPTER II: METHODS FOR ANALYSIS
OF THE ANTIBODY REPERTOIRE ...................................................................................... 13
  ANTIGEN ARRAYS .................................................................................................................. 13
  ANTIGEN DISPLAY TECHNOLOGIES ...................................................................................... 20
  B CELL RECEPTOR SEQUENCING ......................................................................................... 22
  MASS SPECTROMETRY .......................................................................................................... 24
  CLONALITY AND ANTIGEN DIVERSITY OF THE ANTIBODY REPERTOIRE ......................... 27
CHAPTER III: ANALYSIS OF ANTIBODY REPERTOIRE DATA ...................... 29

HETEROGENEITY OF ANTIBODY REPERTOIRES ................................................................. 29

COLLECTION OF ANTIBODY REPERTOIRE DATASETS ..................................................... 31

Study design .......................................................................................................................... 31

Method-dependent background signal levels .......................................................... 33

DATA SPARSITY IN ANTIBODY REPERTOIRES ............................................................... 34

TRANSFORMATION OF ANTIBODY REPERTOIRE DATA ................................................... 35

Binarization .......................................................................................................................... 35

Stratification ......................................................................................................................... 37

QUANTIFICATION OF ANTIBODIES ..................................................................................... 38

ASSOCIATIONS WITH THE ANTIBODY REPERTOIRE ....................................................... 40

CHAPTER IV: PRESENT INVESTIGATION ............................................................................. 43

PAPER I: AUTOANTIBODY PROFILES ASSOCIATED WITH CLINICAL FEATURES IN
PSYCHOTIC DISORDERS ....................................................................................................... 45

PAPER II: SARS-CoV-2 EXPOSURE, SYMPTOMS AND SEROPREVALENCE IN
HEALTHCARE WORKERS IN SWEDEN ............................................................................... 48

PAPER III: SARS-CoV-2 INDUCES A DURABLE AND ANTIGEN SPECIFIC HUMORAL
IMMUNITY AFTER ASYMPTOMATIC TO MILD COVID-19 INFECTION .............................. 51

PAPER IV: ROBUST HUMORAL AND CELLULAR IMMUNE RESPONSES AND LOW RISK
FOR REINFECTION AT LEAST 8 MONTHS FOLLOWING ASYMPTOMATIC TO MILD
COVID-19 ................................................................................................................................ 53

PAPER V: PREVALENT AND PERSISTENT NEW-ONSET AUTOANTIBODIES IN MILD TO
SEVERE COVID-19 ............................................................................................................... 55

CONCLUDING REMARKS ................................................................................................. 60

ACKNOWLEDGEMENTS ............................................................................................... 62

REFERENCES ..................................................................................................................... 66
Chapter I

The antibody repertoire

The antibody repertoire can be regarded as one part of a memory which has registered the immunological events throughout the life of an individual. It consists of the collection of different antibodies which is unique and largely stable within an individual, but also able to rapidly adapt in response to infection. Reading the full extent of this memory would give understanding of an individual’s immunological protection and weaknesses and improved knowledge of the biological processes that underlie them.

Analysis of the antibody repertoire begins with the biological mechanisms generating repertoire diversity. In this chapter I introduce these processes, beginning with the earliest studies of antibodies.

Serology – the study of antibodies

Since its inception at the end of the 19th century, the field of serology has experienced a change of scope and direction. The original definition of serology reflected the literal meaning of the word – the study of (proteins in) blood serum. This connotation has later been shifted to other fields of study, whereas serology has been focused on the study of a single group of proteins, antibodies, while simultaneously expanding its scope to other body fluids, somewhat resulting in a misnomer.

The contradictory terminology might stem from the earliest work leading to the discovery of antibodies. Pioneering work around 1900 focused on the antitoxic and lytic properties of serum, closely tying these concepts to serology. In their 1890 publication, Emil von Behring and Kitasato
Shibasaburō reported that they had developed sera with “antitoxic properties” for tetanus and diphtheria by immunization of animals in a process they termed “serum therapy” (Von Behring & Kitasato, 1890). Further work on serum was carried out by Paul Ehrlich who in 1891 abandoned the term “antitoxin” and instead coined the term “Antikörper”, the German word for antibody, in a paper where he determined that animals exposed to two different toxins produced two different Antikörper (Ehrlich, 1891). This provided proof for the “lock-and-key” component of his 1900 side chain theory, which has striking similarities (and differences) to our current understanding of antibody secretion. Von Behring and Ehrlich continued the development of serum therapy, resulting in the optimization of diphtheria antisera and efficient treatment, winning Von Behring the first Nobel Prize in Medicine or Physiology in 1901.

The early medical breakthroughs of serology were progressed during the 20th century in parallel with advances in biochemistry, leading to drastic progress in the comprehension of the immune system, the emergence of the field of immunology, and further development of vaccine technologies, diagnostic tests, and monoclonal antibody therapies. A central contributor to these advances has been the characterization of antibody structure, diversity, and function.

**Foundations of antibody repertoire diversity**

Antibodies may be one of the most widely recognized molecules with their roughly Y-shaped structure. The tops of the arms contain the variable domains which comprise most of the astounding diversity of antibodies. This diversity enables the major function of antibodies: the specific recognition and subsequent elimination of invading pathogens.

**Antibody structure**

Antibodies are large glycoproteins of approximately 150 kDa whose structure may be described in terms of structural domains or functional regions. Structurally, mammalian antibodies are composed of two identical heterodimers composed of one heavy (H) and one light (L) chain, which are linked by disulfide bonds (Figure 1). Each chain is composed of
The antibody repertoire constant (C) and variable (V) domains. The secondary structure of each domain, the “immunoglobulin fold”, consists of two β-sheets linked by a disulfide bridge. The light chain is composed of one variable (VL) and one constant domain (CL), while the heavy chain has one variable (VH) and, depending on heavy chain type, three or four constant domains (CH1, CH2, ...) that form the trunk of the Y-shaped molecule. In humans, the five main heavy chain types are γ, μ, α, δ, and ε, which yield the antibody isotypes IgG, IgM, IgA, IgD, and IgE, respectively (Janeway et al., 2001).

Functionally, the trunk of the antibody forms the fraction crystallizable (Fc) region and is responsible for mediating the various effects of antibody binding to other components of the immune system, and for membrane attachment in the B cell receptor complex. The arms of the antibody are made up of the VH, CH1, VL, and CL, collectively called the fraction antigen binding (Fab) (Figure 1). In IgG, IgA, and IgD, the Fabs are tethered to the Fc by flexible domains, allowing binding of antigens with variable spacing (Harris et al., 1992; Silverton et al., 1977). The VH and VL of the Fab make up the fraction variable (Fv), which houses the largest portion of antibody structural diversity. The variability is greatest in the three hypervariable regions, or complementarity-determining regions (CDR), contained in

---

**Figure 1 | Foundations of antibody repertoire diversity.** The diversity of the antibody repertoire is based on the structure of antibodies and the recombination and hypermutation of the heavy and light chain variable domains. To the right, recombination of V, D, and J gene segments of the heavy (blue) and light (green) chains generates large combinatorial diversity as well as junctional diversity. SHM further increases antibody diversity during affinity maturation. To the left, a schematic of antibody structure is shown.
each variable domain. These CDRs consist of amino acid loops which interact with antigens and provide the affinity of the antibody (Janeway et al., 2001).

B cell development

The theoretical upper limit of antibody diversity has been estimated to exceed $10^{15}$ unique antibody clones (Briney et al., 2019; Schroeder, 2006). This striking diversity is generated during B cell maturation in the central and peripheral lymphoid organs through two intricate processes of somatic recombination and mutation: V(D)J recombination and somatic hypermutation (SHM).

V(D)J recombination is a central process in the early development of B cells in the bone marrow which generates the initial antibody repertoire diversity. This process relies on rearrangement of the multiple copies of the gene segments encoding the heavy and light chain variable domains. The light chain V domain consists of two gene segments. The first, variable (V\(_L\)) segment encodes 95-101 amino acids, thus making up most of the V domain. The second, joining (J\(_L\)) segment is much shorter, encoding up to 13 amino acids. The heavy chain consists of three gene segments. Like the light chain, a V\(_H\) and a J\(_H\) segment constitute each end of the V\(_H\) domain. In addition, these are joined by a diversity (D) segment. This recombination takes place in two steps. First, a D segment and a J\(_H\) gene segment are joined. The resulting DJ segment is joined to a V\(_H\) gene segment, creating the complete VDJ exon, encoding the full V\(_H\) domain. In both heavy and light chains genes, the rearranged V domain is joined to a C domain through RNA splicing (Figure 1) (Janeway et al., 2001).

The antibody diversity that is created by recombination of gene copies is termed combinatorial diversity and is dependent on the number of copies of each segment. Light chains are encoded at two loci, \(\kappa\) and \(\lambda\). The \(\kappa\) light chain locus contains approximately 40 functional V\(_L\) segments and 5 J\(_L\) segments, which in combination yields 200 possible V\(_{\kappa}\) domains. Correspondingly, the \(\lambda\) light chain locus contains approximately 30 functional V\(_L\) segments and 4 J\(_L\) segments, and thus 120 possible V\(_{\lambda}\) domains. Therefore, there are around 320 possible germline encoded combinations yielding functional V\(_{\lambda}\) domains. Heavy chains are encoded at
The antibody repertoire contains approximately 65 functional $V_H$ segments, 27 $D$ segments, and 6 $J_H$ segments, combinatorically resulting in around 11,000 possible $V_H$ domains. In theory, any heavy and light Ig chain can be combined, giving an estimated germline encoded antibody diversity of 3.5 million possible combinations. However, this combinatorial diversity is smaller than this estimate in practice due to non-uniform rates of gene segment usage and incompatibility of some heavy and light chain rearrangements (Janeway et al., 2001).

In addition to combinatorial diversity, junctional diversity contributes to the diversity of the naïve B cell repertoire. Estimates place total naïve diversity at around $10^{13}$ distinct B cell receptors (Calis & Rosenberg, 2014). Junctional diversity stems from the process joining the $V$, $D$, and $J$ gene segments which involves addition and deletion of nucleotides at the junctions. The number of added nucleotides is random, which results in frameshifts and loss of functionality in about two out of three joins. Junctional diversity affects CDR3, which contains the $V$-$D$ junction in light chains, or the $V$-$D$-$J$ junction in heavy chains. Therefore, CDR3 has a considerably higher diversity than CDR1 and 2, particularly in the heavy chain where CDR3 contains two junctions (Janeway et al., 2001). For this reason, bulk sequencing of heavy chain CDR3 has been used as a proxy for estimating total antibody repertoire size (Briney et al., 2019; Calis & Rosenberg, 2014).

The antibody repertoire contained in naïve B cells may be further diversified by somatic hypermutation. When the B cell receptor of a naïve B cell binds a target epitope with sufficient affinity, the antigen is endocytosed, digested, and resulting peptides are displayed on major histocompatibility complex (MHC) class II on the B cell surface. Recognition of this MHC class II:peptide complex by an activated T cell in the peripheral lymphoid organs results in B cell activation and proliferation and the formation of a primary focus of clonal expansion. This produces the first phase of the humoral immune response with early IgM-secreting plasma cells. After several days, some proliferating B and T cells migrate into a primary lymphoid follicle where continued proliferation eventually leads to the formation of a germinal center. Here, the germinal center reaction takes place during intense B cell proliferation, composed of somatic hypermutation leading to affinity maturation, as well as class
switching. Somatic hypermutation introduces point mutations in the V domains of the heavy and light chains at a rate of approximately $10^3$ base pairs per cell division, which corresponds to around 1 amino acid substitution in the V domains at every second B cell division. Accumulating mutations that confer increased affinity to the antigen (primarily in the CDRs) result in B cell selection and expansion, ultimately yielding plasma cells producing affinity matured and class-switched antibodies with enhanced affinity to the target antigen (Janeway et al., 2001).

**Regulating the diversity of the antibody repertoire**

The combinatorial processes generating the vast diversity of the naïve antibody repertoire invariably generate B cell receptors recognizing constituent “self” components of the host, such as proteins, DNA, or lipids. Indeed, it has been shown that more than half of early immature B cells display autoreactive B cell receptors (Wardemann et al., 2003). Left unchecked, these autoreactive B cells might develop into plasma cells secreting large amounts of autoreactive antibodies, often called autoantibodies. These autoantibodies may have the capacity to destroy host tissues and disrupt cell signaling, which is observed in many autoimmune diseases. To reduce self-recognition and autoimmunity, there are several mechanisms for culling autoreactive B cells.

**Induction of tolerance**

During B cell development, a substantial portion of immature autoreactive B cells is neutralized by receptor editing or clonal deletion. These mechanisms are carried out in the bone marrow and are therefore termed central tolerance. The mechanism of receptor editing provides a possibility for strongly or multivalently autoreactive immature B cells to continue rearrangement of their light chain despite already arriving at a productive but autoreactive gene combination. Failing this, clonal deletion occurs, where the B cell clone is eliminated by apoptosis (Gay et al., 1993; Nemazee & Bürki, 1989; Tiegs et al., 1993).

Although these processes of central tolerance purge immature B cells with strong affinity or avidity to self-molecules presented in the bone marrow,
B cells targeting other self-antigens migrate to the periphery. There, peripheral tolerance mechanisms contribute to further culling potentially autoimmune B cells. Autoreactive B cells may be eliminated by anergy, which is a state of unresponsiveness to the target antigen. Anergy results from encountering the target antigen and launching an initial antigen response, while failing to receive the required co-stimulation for a sustained response, such as T cell help. Eventually, anergic B cells die by neglect (Cambier et al., 2007).

Escaping anergy, autoreactive B cells in the periphery may be eliminated by entrapment and apoptosis. Upon encountering their target antigen and displaying MHC:antigen peptide, B cells stop in the T cell zone of secondary lymphoid tissues. For autoreactive B cells, no activated cognate T cells are available, as T cells recognizing MHC:self-peptide have been culled during normal T cell development. Lacking T cell help, the antigen-stimulated autoreactive B cell cannot enter a primary lymphoid follicle and is eliminated by apoptosis (Parham & Janeway, 2009).

**Natural antibodies**

Despite the mechanisms for induction of tolerance, a subpopulation of B cells is noted for its polyreactive and therefore autoreactive affinities. These B-1 cells follow a separate developmental pathway generating reduced clonal diversity, largely stemming from reduced junctional diversity (Parham & Janeway, 2009).

Antibodies secreted by B-1 cells are part of an antibody subset with restricted diversity, termed natural antibodies. This antibody subset is characterized by polyreactivity, binding a wide range of bacterial components. As B-1 cells are activated in a T cell-independent manner, they can respond rapidly to general inflammatory and pathogen stimuli in an innate-like manner with swift production of short-lived IgM-secreting plasma cells. In addition to pathogen recognition, the polyreactive B cell receptors can recognize self-antigens and aid in homeostasis by clearance of apoptotic cells (Baumgarth, 2011). However, their poly- and self-reactivity and T cell-independent activation requires strict regulation to prevent autoimmunity. Therefore, dysregulation of B-1 cells is a suspected source of autoantibodies (Lee et al., 2020).
Autoantibodies

Failure of the mechanisms preventing activation of autoreactive B cells may result in loss of tolerance, autoantibody production, and autoimmune disease. In line with the importance of T cells in classical B cell activation, breakdown of T cell tolerance seems to be critical in many autoimmune diseases due to the presence of autoantibodies that have undergone class switching and somatic hypermutation (Parham & Janeway, 2009). However, knowledge on the mechanisms of autoantibody formation remains incomplete, though environmental triggers in combination with genetic susceptibility, particularly in the MHC genes, is a prevalent theory.

Environmental triggers include trauma, radiation, and infection. Infection has long been suspected to be involved in autoantibody formation. Molecular mimicry is one possibility, where a pathogen-derived antigen triggers a normal B cell response. In some unfortunate cases, the epitope of the pathogen has a highly similar counterpart among host epitopes, triggering an autoimmune response (Münz et al., 2009). This has been observed in, e.g., multiple sclerosis (Epstein-Barr virus) (Tengvall et al., 2019), and rheumatic fever (Streptococcus pyogenes infection) (Parham & Janeway, 2009). Furthermore, bystander activation and epitope spreading may contribute to emergence of autoantibodies in viral infections. These mechanisms are based on bystander destruction of tissues upon infection, whereby self-antigens are processed by antigen-presenting cells and presented to autoreactive T cells. Further tissue damage based on the autoreactivity rather than the viral infection may cause the process to repeat, activating further autoreactive T cells (Münz et al., 2009). Epitope spreading may also occur due to physical linkage of antigens. For instance, in systemic lupus erythematosus (SLE), a single autoreactive T cell clone targeting a single component of the nucleosome may activate B cells targeting the same component. In addition, the T cell may activate B cells targeting another component, e.g., nucleosomal DNA, which is physically linked to nucleosomal proteins. This occurs as the B cell endocytoses and digests the whole nucleosome on B cell receptor binding, and is implicated in the spread of the disease (Parham & Janeway, 2009).

Autoantibodies are not the only source of autoimmune diseases, although they are one of the best understood due to their relative accessibility for
study. Among the plethora of diseases of known autoimmune origin, a wide range of autoantibodies have been described, although the target antigens have been described at highly variable levels ranging from tissue lysates, via cellular components, to individual proteins. The target of the autoantibody reflects the presentation of the disease, which can be roughly divided into three categories. Autoantibodies that target antigens which are expressed in many tissues of the body cause systemic autoimmune diseases. These include SLE, where DNA-associated components such as histones, ribosomes, and DNA are targeted, and Goodpasture’s syndrome, where the antigen is type IV collagen. On the other side of the spectrum, autoantibodies targeting tissue-specific antigens may cause destruction of specific organs, such as Hashimoto’s disease (thyroid gland), autoimmune hemolytic anemia (red blood cells), and rheumatoid arthritis (joints) (Parham & Janeway, 2009). The present understanding of target antigens varies considerably. In rheumatoid arthritis, anti-citrullinated protein antibodies (ACPA) target various citrullinated proteins. In vasculitis, anti-neutrophil cytoplasmic antibodies (ANCA) target antigens in the cytoplasm of neutrophils, such as myeloperoxidase (MPO) and proteinase 3 (PR3). In systemic autoimmune diseases such as SLE, Sjögren syndrome, and scleroderma, various antinuclear antibodies (ANA) such as anti-Ro, La, or Sm antibodies target different components of the nucleus.

Autoantibodies may also target specific cell surface receptors and cause disease by interfering with cell signaling. This is for instance seen in Graves’ disease, myasthenia gravis, and anti-N-methyl-D-aspartate receptor (NMDAR) encephalitis. In Graves’ disease, autoantibodies targeting the thyroid-stimulating hormone receptor cause overactivation of the thyroid gland and excessive production of thyroid hormones, which disrupts metabolism. In myasthenia gravis, autoantibodies targeting the acetylcholine receptor causes internalization and degradation of acetylcholine receptors at the neuromuscular junction, leading to reduced receptor density causing reduced neuromuscular signaling capacity, resulting in muscle weakness (Parham & Janeway, 2009). In anti-NMDAR encephalitis, antibodies in cerebrospinal fluid that target the NMDA receptor interfere with NMDA signaling and cause symptoms of psychosis (Pollak et al., 2020).
Towards analysis of the antibody repertoire

The analysis of antibodies has been of great importance for modern immunology. Serological assays have been developed for monitoring viral outbreaks such as Ebola, Zika, and SARS-CoV-2 (Amanat et al., 2020; Ayoub et al., 2017; Balmaseda et al., 2017; Hober et al., 2021), evaluation of vaccine trials (Wilde et al., 1999), and detection of autoantibodies serving as clinical biomarkers of autoimmune diseases (Dalmau & Bataller, 2007; Schellekens et al., 2000; Schmidt & Zillikens, 2013). The vast majority of serological studies and nearly all clinical serological tests have been performed in a singleplex format, analyzing antibodies towards one antigen or one antigen mixture at a time. The enzyme-linked immunosorbent assay, ELISA (Engvall & Perlmann, 1972), is perhaps the most prominent example and remains the gold standard in many settings, including clinical diagnostics. In vaccine trials, development of neutralizing antibodies has long served as the correlate of protection and target outcome, which typically is evaluated using cell-based neutralization assays or cell-free pseudoneutralization assays (Plotkin, 2008; Rappuoli, 2007). Although singleplex serological assays have propelled the study of antibody responses in infection and autoimmune conditions, it has become apparent that multiplex serological methods for simultaneous analysis of multiple antibody clones or affinities provide valuable opportunities for the analysis of the multifaceted antibody response.

In a 2005 editorial, Dillner (2005) highlighted the potential of emerging technologies for multiplex antibody analysis for several areas of medical research such as autoantibody signatures of autoimmune diseases, and profiling of emerging antibodies in cancers. Dillner proposed the term “serolomics” for this field of research and stressed its potential in the analysis of broad spectra of microbial antigens to which a person might have been exposed for monitoring and risk assessment of future disease. He exemplified this with the multiplex serology presented by Waterboer et al. (2005) which was capable of analyzing antibodies towards up to 100 antigens in the variable and oncogenic human papillomaviruses (HPV), increasing sensitivity and enabling association of distinct antibodies with cancer. This and other studies highlight how simultaneous analysis of
several antibodies may directly contribute to improved diagnostics. In SARS-CoV-2 infection, analysis of just three viral antigens enhanced sensitivity and specificity of a serological assay over that of other assays available at the time which relied on single antigens (Hober et al., 2021). In SLE, a panel of 26 autoantibodies improved the diagnostic accuracy compared to established serological tests and reflected presented symptoms, which may be of prognostic use for involvement of different organs and stratification of patients for personalized treatment (Lewis et al., 2018).

Antibody repertoires paint a complex picture of the antibody landscape, and several complementary methods are required to get a broad view of the land. Several reviews have highlighted different aspects of the analysis. The paper by Georgiou and colleagues (Wine et al., 2015) focused on the need for analysis of antibody repertoires for the resolution of biological questions. The authors suggested using multiple orthogonal approaches for profiling of the antibody repertoire such as affinity-based assays, functional assays, and immunoglobulin mass spectrometry, alongside sequencing of the B cell receptor repertoire. With these methods, they suggest that it might be possible to elucidate the size of the serological antibody repertoire, the concentration distribution of circulating antibodies, the immunological contributions of antibody isotypes, effects of post-translational modifications, and the significance of the polyreactive antibody repertoire. Other reviews have supported the concept of “systems serology” or “antibodyomics” (Ackerman et al., 2017; Arnold & Chung, 2018; Loos et al., 2020). In this concept, multiple high-throughput and multiplex techniques are suggested to be used in parallel on the same sample material to collect a high dimensional dataset. Applying machine learning techniques to such datasets has been suggested to provide a novel venue for disentangling the complex immune responses following infection or vaccination and has been exemplified in the context of human immunodeficiency virus (HIV) infection (Alter et al., 2018). Moritz et al. (2020) suggested supplementing this view with “autoantigenomics”, an autoantigen-centric view of affinity-based untargeted antibody profiles.
It is evident that antibody repertoires remain a trove of knowledge for medicine, immunology, and biology, and has yet to be fully explored. Combining large-scale data on different modalities of antibody repertoires is poised to advance understanding of antibody repertoires in health and disease and the immunological processes that govern them. In the following chapter, I will present some current methods that enable such investigations.
Chapter II

Methods for analysis of the antibody repertoire

Experimental methods for analysis of the antibody repertoire may investigate different levels of antibody repertoire diversity providing different layers of information. Sequencing can reveal the clonality of the B cell repertoire, either in bulk as separate V<sub>L</sub> and V<sub>H</sub> chains, or in single cell as complete Fabs. Mass spectrometry is better situated to investigate the clonality of the humoral antibody repertoire and may also be used for identification of target antigens. Affinity-based methods address the antigen diversity of the antibody repertoire and have been developed in many variants which may be roughly divided into antigen arrays and antigen display technologies.

In this chapter, I give an overview of selected methods and conclude with considering their strengths and weaknesses in analyzing different facets of the antibody repertoire.

Antigen arrays

The work presented within this thesis was performed using antigen arrays. Antigen arrays come in two main formats: planar arrays and bead-based arrays. These methods for analysis of antibody repertoires are closely related to the DNA arrays developed and used in the 1990s (Schena et al., 1995) and the subsequent protein microarrays based on cDNA library overexpression (Büssow et al., 1998; Cahill, 2000; Lueking et al., 1999), as well as the bead-based antibody arrays later utilized for affinity proteomics (Gupta et al., 2016; Häggmark et al., 2013; Kingsmore, 2006; Schwenk et
al., 2008). Both array types consist of a collection of antigens which are attached to a solid support (Figure 2). For planar arrays, this support consists of functionalized glass slides (Nilsson et al., 2005; Robinson et al., 2002) or cellulose membranes (Davies et al., 2005), while the support of bead-based arrays consists of polymeric microspheres (Fulton et al., 1997). The main function of the support is the preservation of antigen identity, i.e., knowing which antigen is which after array production. On planar arrays, antigen identity is preserved using the physical address, or location, of the antigen on the glass slide. On bead-based arrays, antigen identity is preserved using the bead identity established with varying concentrations of embedded dyes and sometimes bead size. For antibody repertoire analysis, the antigen collections typically consist of proteins, protein fragments, or peptides (Ayoglu et al., 2013; Gnjatic et al., 2010; Sokolove et al., 2012), but may also consist of other biomolecules such as glycans (Yu et al., 2014) or lipids (Kanter et al., 2006). Samples are often derived from human blood, but other body fluids are also used, such as cerebrospinal fluid (CSF), saliva, and milk, as well as lavage fluids, and tissue lysates (Boustani et al., 2022; Häggmark et al., 2015; Hu et al., 2011; Just et al., 2020; McGuire et al., 2021; Querol et al., 2013).

In a typical antigen array analysis, the samples to be analyzed are incubated on the array to allow binding of antibodies in the sample to antigens on the solid support. Sample identity is preserved by keeping samples separate. On planar arrays, samples are kept on separate slides or different slide sections using a slide mask, while bead-based arrays keep samples in separate wells of a microtiter plate (e.g., 96-well or 384-well plates) with aliquots of the array added to the wells. Excess sample is washed away and any human antibodies remaining on the array are taken to stem from the analyzed sample. These antibodies are typically detected using a secondary detection antibody with affinity towards the intended Ig subtype, commonly anti-human IgG, which has been labeled with a fluorophore such as R-phycoerythrin (R-PE) or an enzyme of a chemiluminescence system such as horseradish peroxidase. The amount of fluorescence at each array address gives a relative quantification of the amount of antibody bound to each antigen.

A major challenge in antigen array construction is the production of antigens. As protein and peptide antigens are major categories used
antigen arrays, I will discuss these further. There are several approaches to protein and peptide antigen production which may be divided into 4 main categories: prior expression, prior synthesis, in situ synthesis, and in situ expression.

Prior expression of antigens relies on a set of proteins or peptides expressed in cell-based systems and arranged in antigen collections. Antigens are acquired from these collections and chemically attached to the solid support. Expression is performed using a cell-based expression system, e.g., *Escherichia coli*, yeast, or human cell lines. The choice of expression system will affect splicing and glycosylation and other post-translational modifications such as citrullination, as well as any expressed purification tags. These processes and modifications affect the surface of the antigen and thus the epitopes displayed (Ayoglu et al., 2013; Tokmakov et al., 2012). Therefore, it is important to bear the expression and purification systems in mind when considering the opportunities and limitations of an antigen array. The number of immobilized antigens is constrained by the size of the antigen library and the array technology used and typically ranges from hundreds to tens of thousands of antigens. Expressed antigens can be immobilized on planar or bead-based arrays. Arrays in this category include the HuProt array (planar array containing 21 000 proteins expressed in yeast cells, CDI Laboratories), the Human ProtoArray (planar array containing 9400 proteins expressed in insect cells, Invitrogen, discontinued), the i-Ome Protein Array (planar array containing 1800 proteins expressed in insect cells, Sengenics), ImmunoINSIGHTS (bead arrays containing subsets of hundreds of proteins from a library of 8000, Oncimmune), the Human Protein Atlas antigen arrays (planar and bead arrays containing up to 42 000 protein fragments expressed in *E. coli*, accessible via the Autoimmunity and Serology Profiling unit at SciLifeLab), and several other non-commercial viral arrays and autoantigen arrays produced by academic research groups.

The Human Protein Atlas antigen arrays were used in **Paper I and V**. These arrays utilize the collection of 42 000 protein fragments produced in the Human Protein Atlas for production and verification of polyclonal antibodies. The fragments represent 18 000 human proteins and were designed to have low sequence similarity to other human proteins (Berglund et al., 2008; Lindskog et al., 2005). They are expressed in *E. coli*
together with a hexahistidyl-albumin-binding protein (His6ABP) tag which enables purification as well as immunopotentiation for production of polyclonal antibodies (Nilsson et al., 2005). The mean length of the fragments is 80 aa (5th and 95th percentiles 26 and 140 aa, respectively), and they together cover approximately 40% of the amino acid residues of the human proteome. In addition to immunization and purification of antibodies, the protein fragments have been arranged on planar arrays in several assemblies ranging from 384 to all 42 000 antigens for verification of antibodies and analysis of the autoantibody repertoire. Furthermore, bead arrays of up to 380 selected protein fragments are used for high-throughput targeted analysis of the autoantibody repertoire (Ayoglu, Schwenk, et al., 2016; Häggmark-Månberg et al., 2016; Sjöberg et al., 2012).

Prior synthesis of antigens is a similar approach to prior expression in that an antigen collection is established prior to array production. However, in this approach peptides are synthetized in cell-free systems, which typically yields short to medium length peptides without posttranslational modifications. The resulting antigen arrays are typically suited for detection of antibodies targeting linear epitopes and especially for epitope mapping. In epitope mapping, a library of tiled peptides is designed, e.g.,
with a peptide length of 15 amino acid residues and a lateral shift of 5 residues. This tiled design allows for determination of the targeted epitope, down to single amino acids if the lateral shift is reduced to a single amino acid. Peptide libraries are available from several vendors and chemical coupling to the array substrate is often performed in-house. This array type was used in Paper V.

Most if not all commercially available peptide arrays instead use the in situ synthesis approach, likely due to ease of production. The array properties are like those of antigen arrays with prior synthesis of antigens, although the array format is exclusively planar owing to the production process. Array contents may be pre-determined or made to order, with commercially available arrays including PEPperCHIP (PEPperPRINT), and PepStar™ and PepSpots (JPT Peptide Technologies). Furthermore, array synthesizers can be bought by researchers for in-house production of arrays.

Production of antigen arrays using in situ expression of antigens relies on immobilization of genetic material which is translated to protein or peptide in situ. The Nucleic Acid-Programmable Protein Array (NAPPA) technology is a major member of this category. In NAPPA, a library of complementary DNA (cDNA) corresponding to the antigens of interest is constructed and immobilized on glass slides together with a capture antibody. The capture antibody targets the protein tag which is expressed in fusion with the antigens of interest. Before analysis, the antigen-tag constructs are expressed in situ using a cell-free expression system. Expressed antigen-tag constructs are immobilized on the array by the capture antibody (Anderson et al., 2011; Ramachandran et al., 2004; Ramachandran et al., 2008; Sibani & LaBaer, 2011). Expressed protein may be denatured on the array for detection of denatured epitopes (Wang et al., 2013).

Each antigen array format carries advantages and disadvantages. These are dependent on the combination of the properties of each specific array, and I will not attempt to give an exhaustive description of them here. I will, however, give the general differences between planar and bead-based arrays. These array types differ substantially in their physical properties due to their different modes of preserving antigen identity as described previously. This also generates differences in multiplexity and sample throughput.
On planar arrays, it is generally possible to analyze single samples per array, giving a low sample throughput. The low throughput is, however, compensated for by the large multiplexing capacity of planar arrays. As antigen identity is preserved by its physical address, the limiting factor is the density with which antigens can be arrayed onto the glass surface, the area of the surface, and the access to antigens. As we have seen, antigen access can vary substantially depending on array production method. The glass surface area is largely standardized to that of a standard microscope slide, i.e., 25×75 mm (1×3 in). This leaves the array density, which varies with production method. For arrays with prior expression or synthesis of antigens, current production methods enable arraying of approximately 40 000 features on a microscope slide in ~100 µm antigen spots (Jeong et al., 2012; Sjöberg et al., 2016). For arrays with in situ synthesis of antigens, 2.1 million 13×13 µm features have been arrayed (Forsström et al., 2014; Zandian, Forsström, et al., 2017). Some of the enormous multiplexing capacity of planar arrays can be traded for increased sample throughput by printing many arrays with smaller surface area on a single microscope slide. These sub-arrays are held separated by array masks, allowing several samples to be analyzed on each antigen array. In addition, sample throughput can be further increased in screening studies by analyzing several samples in combination on a sub-array. This approach was used in Paper I and V and enabled the identification of antigen targets from 32 individuals using 8 arrays. Furthermore, antigen representation on planar arrays may have different properties owing to their production method. To increase the epitope coverage, different planar array platforms may be combined (Henjes et al., 2014).

In contrast to planar arrays, bead-based arrays have a high sample throughput but lower multiplexity. As sample identities are preserved by physical separation in microtiter plates, the throughput is dependent on the microtiter plate format. Common analyzers for bead-based arrays, e.g., the FlexMap3D®, are compatible with 96- or 384-well formats. Antigen identities are allocated to bead identities, with 500 identities being commercially available (Luminex xMAP® Microspheres). Furthermore, research groups have constructed 1728 bead identities, which may increase the multiplexity of bead-based arrays (Holm et al., 2012; Slaastad et al., 2011).
Methods for analysis of the antibody repertoire

Considering their advantages and disadvantages, planar and bead-based antigen arrays might be utilized to their fullest potential when combined in a stepwise analytical design. Using this approach, a limited number of carefully selected samples from individuals with the highest relevance to the research question are analyzed on highly multiplex planar antigen arrays. To further increase sample throughput and maximize the number of screened antibody repertoires, samples may be combined prior to analysis on the array. However, the number of combined samples should be limited, as each sample dilutes the constituent antibody species of the other samples. In the experience of our lab, four samples can be combined while still detecting antibodies generating relatively low signal levels. With individual or combined samples subjected to a broad screening, a list of antibody hits is generated. A suitable number of antigens is selected from this screening list based on criteria tailored to the experiment, such as highest signal levels, moderate to high signals in several samples, and moderate to high signals and biological relevance of the antigen. The antigens on this list are collected and coupled to microspheres, yielding a bead-based antigen array. This array is used for screening of a larger number of samples, ranging from hundreds to thousands of samples. It is worth noting that the translation of antibody detection between array formats has a higher success rate if the same antigen representations are used in both array formats. In our lab, this approach is often employed when using the Human Protein Atlas arrays for autoantibody profiling, as exemplified in several publications as well as Paper I and V (Ayoglu et al., 2013; Ayoglu, Schwenk, et al., 2016; Häggmark-Månberg et al., 2016; Häggmark et al., 2015; Henjes et al., 2014; Just et al., 2020; Mikus et al., 2019; Pin et al., 2017; Pin et al., 2021; Zandian, Forsström, et al., 2017; Zandian, Wingard, et al., 2017).

Aside from their use for analysis of antibody repertoires, antigen arrays have been used for validation of monoclonal and polyclonal antibodies and other affinity binders, as well as characterization of protein-protein interactions, protein-DNA interactions, small molecule interactions, and enzyme substrates (Azevedo et al., 2018; Buus et al., 2012; Chen et al., 2013; Cox et al., 2015; Forsström et al., 2014; Hu et al., 2009; Jeong et al., 2012; Sjöberg et al., 2012; Venkataraman et al., 2018).
Antigen display technologies

Display technologies for analysis of antibody repertoires have been developed from the phage display technology first described by Smith (1985). Since their inception, phage display technologies have been used for directed evolution of peptides, antibodies, and other affinity molecules for affinity maturation (McCafferty et al., 1990; Nord et al., 1997; Parmley & Smith, 1988; Scott & Smith, 1990; Smith, 1985).

The central principle of display technologies is the fusion of viral structural proteins with a peptide or protein of interest (Figure 2). This fusion product is designed on the genetic level and cloned into the display vector, e.g., a bacteriophage. The display vector translates and displays the protein on its surface. This connection between genotype and phenotype constitutes the preservation of antigen identity in display technologies. In relevant applications, it allows for affinity-based enrichment of vectors along with their genetic material. Selected vectors may be expanded and subject to further rounds of enrichment for, e.g., directed evolution or increased selection stringency. Enriched vectors are sequenced and antigen identities are obtained (Li, 2000; Smith & Petrenko, 1997).

In addition to directed evolution, phage display has been utilized for characterization of autoantibodies. Early approaches often relied on the expression of cDNA libraries derived from diseased tissue, which has enabled the discovery of autoantigens in, e.g., prostate cancer (Wang et al., 2005). Several display technologies for analysis of antibody repertoires have been developed to address the limitations of cDNA phage display, such as out of frame translation, limited antigen length, limited posttranslational modifications, skewed library representation, and incomplete coverage of the proteome. Here, I describe three recent display technologies for antibody repertoire analysis: phage immunoprecipitation sequencing (PhIP-Seq), molecular indexing of proteins by self-assembly (MIPSA), and rapid extracellular antigen profiling (REAP).

PhIP-Seq was first described by Larman et al. (2011) for the detection of autoantigens and has subsequently been adapted for viral and microbial serology (Schubert et al., 2019; Shrock et al., 2020; Vogl et al., 2021; Xu et al., 2015). With PhIP-Seq, phage display for analysis of antibody
Methods for analysis of the antibody repertoire

repertoires took a step forward from cDNA libraries to rationally designed proteome-wide peptide libraries. In the disseminating publication, Larman et al. (2011) presented a synthetic library representing all known and predicted open reading frame (ORF) sequences of the human genome at that time. This library consists of tiled 36-amino acid residue peptides with a lateral shift of 7 residues. The library is produced on DNA microarrays and cloned into T7 phage. The resulting phage library can be propagated for renewal and aliquots used for serological analysis by immunoprecipitation. Phage are incubated with sera to be analyzed in 96-plate wells. Any resulting phage-antibody complexes are pulled down with the addition of magnetic microspheres with coupled protein A and G. Unbound phage are washed away and the phage inserts of enriched phage are amplified. Sample barcodes are introduced in each well using sample-specific primers which map sample identity to primer sequence. This allows for the combination of enriched tile DNA from different sample wells and single readout using next-generation sequencing.

While PhIP-Seq represented a major step forward, the technique has limitations. Antigen size is limited to ~90 amino acid residues (Mohan et al., 2018), which limits the amount of conformational epitopes displayed. Although secondary structure might be observed, it might not represent the native folding state. In addition, epitopes consisting of interacting but linearly distant parts of the protein are not represented. Furthermore, disulfide bridges and posttranslational modifications are typically not representative of the native state as the T7 phage particles are produced in the cytoplasm of E.coli (Mohan et al., 2018).

To address these limitations, the molecular-display technology molecular indexing of proteins by self-assembly (MIPSA) was developed by Credle et al. (2022) with proof-of-concept shown by detection of autoantibodies in COVID-19. The MIPSA system relies on a specialized vector and the HaloTag covalent labelling system (Los et al., 2008). In brief, an ORF library is expressed in a cell-free system using the MIPSA vector. This vector contains a barcode which is used to preserve antigen identity, and a HaloTag fusion protein for labelling. The barcode is located upstream of the ribosome-binding site and is not translated. In vitro transcription results in RNA for translation. Reverse transcription of the upstream barcode using a HaloLigand-labelled primer yields a DNA barcode tag.
Upon translation of the HaloTag-ORF fusion protein, the DNA barcode tag covalently binds the freshly synthesized protein within the translation complex. In the disseminating publication, 11 000 full-length proteins were expressed and tagged using this system. This protein library was used for immunoprecipitation using patient sera and microspheres with coupled protein A and G, as previously described for PhIP-Seq. While MIPSA may become a powerful tool for analysis of antibody repertoires there are few publications on the method at the time of writing, possibly due to its recent dissemination.

As extracellular proteins are exposed to the immune system, they are of particular interest in antibody repertoire analysis. However, these proteins are often difficult to express as they often have specific processing requirements such as removal of signal peptide, formation of disulfide bonds, and post-translational modifications (Wang et al., 2022). To address these challenges, Wang et al. (2021) developed rapid extracellular antigen profiling (REAP). REAP is a yeast display technique developed for discovery of autoantibodies targeting the extracellular proteome. In brief, 2688 extracellular or secreted proteins are expressed on the surface of yeast cells using a yeast display vector. The vector contains a DNA barcode which preserves antigen identity. Immunoprecipitation is performed similarly to PhIP-Seq and MIPSA, although patient serum IgG is purified prior to incubation with yeast, and pull-down is performed with anti-human IgG Fc antibodies. Barcode readout is performed using next-generation sequencing with indexed primers preserving sample identity. Like MIPSA, REAP has not been adopted by independent research groups. Again, this may be due to its recent dissemination. Although Wang et al. (2022) note that a portion of proteins were not successfully expressed and some were incorrectly folded, the promise of faithful representation of a large portion of the extracellular proteome may render REAP a valuable addition to the repertoire of display technologies.

**B cell receptor sequencing**

As the antibody repertoire stems from the diversity of B cell receptors, much effort has been dedicated to understanding the B cell receptor repertoire. Although the B cell receptor repertoire may not fully reflect the
humoral antibody repertoire, I briefly discuss B cell receptor sequencing and emerging methods connecting sequenced B cell receptor repertoires to antigen specificity.

The role of B cell receptors in the generation of antibody diversity is discussed in Chapter I. Briefly, the heavy and light chains of Ig undergo V(D)J recombination in developing B cells, creating a remarkable diversity of B cell receptors, which has been estimated to exceed $10^{13}$ unique B cell receptors prior to immune exposure (Calis & Rosenberg, 2014). This diversity is further increased by somatic hypermutation and has been estimated to reach $10^{18}$ unique antibody sequences (Briney et al., 2019).

Bulk B cell receptor sequencing has long been used to study the evolution of the B cell receptor repertoire (Loos et al., 2020). Although the pairing of recombined heavy and light chains is lost in bulk analysis, several B cell receptor repertoire features are discernible, e.g., V(D)J segment usage frequencies, CDR3 properties, somatic hypermutation patterns, class switching patterns, and clonal lineage analysis (Chaudhary & Wesemann, 2018). These analyses have led to awareness of overall size of repertoires, dynamics, and diversity (Loos et al., 2020). Interestingly, large inter-individual heterogeneity has been observed in bulk B cell receptor sequencing (Kräutler et al., 2020).

Single cell B cell receptor sequencing has been realized with advances in next-generation sequencing. This has substantially advanced the study of B cell receptor repertoires by giving a truer representation of B cell receptor repertoires and enabling the study of the co-evolution of heavy and light Ig chains (Chaudhary & Wesemann, 2018).

For the assessment of antibody binding, functional studies of B cell receptor repertoires may be the most interesting approach. Before the development of single cell B cell receptor sequencing, methods for assessment of functionality had limited throughput, e.g., functional screening of B cell hybridoma (Buchacher et al., 1994), flow cytometry of B cells and fluorescently labelled antigens (Scheid et al., 2009), and B cell culture (Walker et al., 2009). Single cell B cell receptor sequencing has enabled characterization of antibody affinity by, e.g., production of recombinant monoclonal antibodies (Setliff et al., 2018), and yeast display of Fab fragments (Wang et al., 2018). Further increasing throughput,
Setliff et al. (2019) developed a method for identification of antigen specificity of B cell receptor sequences with paired heavy and light chains. This method is based on probing of B cells using DNA barcoded antigens followed by single cell sequencing for identification of antigen and B cell receptor sequence. This may aid in identification of potentially therapeutic patient-derived antibodies (Setliff et al., 2019; Shiakolas et al., 2021) as well as yield insights in the clonality of antigen-specific responses.

Mass spectrometry

Mass spectrometry-based methods have been developed for analysis of antibodies. The various methods can yield a wide range of information, such as antibody clonality of untargeted antibody repertoires or of antigen-directed species, sequences of variable domains, and identification of unknown antigens. Many mass spectrometry-based methods focus on other antibody characteristics than antibody affinity, making mass spectrometry an orthogonal approach to purely affinity-based methods. Here, I give a brief overview of mass spectrometry and selected applications.

Mass spectrometry is based on the determination of mass to charge ratio of analyzed molecules. For the analysis of proteins, this can occur in four modalities: bottom-up, top-down, middle-down, and intact mass analysis. Bottom-up is the most common approach. It relies on the predictable fragmentation of peptides using a specific fragmentation method, typically digestion by trypsin. The resulting peptides are separated using liquid chromatography, ionized, and introduced to the first mass analyzer where mass over charge ratios are registered. Depending on the mode of operation, single or multiple precursor ions are selected for fragmentation by, e.g., collision-induced dissociation, and detected in a second mass analyzer. The recorded mass spectra are compared to simulated mass spectra which allows for the identification of constituent proteins (Aebersold & Mann, 2016). However, simulated mass spectra are not available for mature antibodies owing to their enormous diversity (De Graaf et al., 2022). Intact mass analysis, on the other hand, does not rely on the fragmentation of proteins. Instead, intact proteins or protein subunits are analyzed by LC-MS. This can yield information on
Methods for analysis of the antibody repertoire

proteoforms, e.g., point mutations, post-translational modifications, or antibody clones (De Graaf et al., 2022). Top-down and middle-down mass spectrometry represent middle paths. In these modalities, full-length or specifically cleaved proteins are separated by liquid chromatography and injected in the first mass analyzer. Ions are fragmented, typically using electron-based techniques, and analyzed in the second mass analyzer. These modalities may be useful in de novo antibody sequencing (De Graaf et al., 2022).

Antibody clonality may be assessed in biofluids by a method based on native mass analysis (Bondt, Hoek, et al., 2021). Here, IgG is purified from donor samples by bead-based pull-down. Using the highly IgG1-specific Ig degrading enzyme (IgdE), complete Fab fragments are cleaved from IgG1, leaving the Fc fragment attached to the bead (Spoerry et al., 2016). Notably, this increases binder concentration twofold as each Ig contains two Fab fragments and increases mass homogeneity by eliminating the heterogeneous N-glycosylation sites of the Fc. Eluted Fab molecules are analyzed by LC-MS together with spiked-in monoclonal IgG1 of known sequences allowing for absolute quantification of donor Fab molecules. In the original publication, Bondt, Hoek, et al. (2021) used this method to detect individual clones of concentrations down to 0.05 µg/ml and could show that IgG1 Fab repertoires were dominated in terms of abundance by 30-500 clones, a surprisingly small number compared to the predicted diversity of B cell receptors. Furthermore, Bondt, Hoek, et al. (2021) showed that antibody repertoires were highly individual specific and stable over time, although responsive to immunological events such as sepsis, which triggered the emergence of novel clones of high abundance. This technique has been adapted for assessing clonality and origin of serum and breastmilk IgA (Bondt, Dingess, et al., 2021; Dingess et al., 2022) and the antigen-directed antibody repertoire in SARS-CoV-2 infection (Van Rijswijck et al., 2022).

Identification of antibody sequences from human biofluids may yield insights in antibody responses after vaccination and infection and in autoimmune disorders, as well as lead to identification of therapeutic antibodies. Antibody sequencing has been attempted by several research groups but has proven challenging. The very diversity that enables directed affinity also impedes traditional bottom-up mass spectrometry
approaches, which rely on reference in silico protein fragmentation, which is not possible for mature antibodies. To this end, proteogenomic approaches have been developed where donor-specific reference B cell receptor sequence databases are established from isolated donor B cells (De Graaf et al., 2022; Lavinder et al., 2015). While this approach has proven powerful, its availability and throughput are limited by the use of dual technologies, particularly as single cell B cell receptor sequencing is required for pairing of heavy and light Ig chains. In addition, B cells in the periphery are typically not fully representative of the humoral antibody repertoire, which may result in skewed antibody detection due to missing B cell sequences. With this in mind, methods for de novo antibody sequencing have been developed. Bondt, Hoek, et al. (2021) demonstrated de novo sequencing of a single antibody from a donor with sepsis using integrated bottom-up and middle-down approaches. Furthermore, a software tool for mapping of bottom-up antibody mass spectrometry data to user-defined templates has been developed by Schulte et al. (2022).

Mass spectrometry in combination with immunoprecipitation may be used for validation or identification of target antigens of antibodies. This approach has been applied in the context of disorders with neuronal antibodies (Boronat et al., 2013; Do et al., 2019; Scharf et al., 2018; van Coevorden-Hameete et al., 2019) and other autoimmune diseases such as systemic sclerosis (Bossuyt et al., 2023; Vulsteke et al., 2023). The method is based on pull-down of proteins from complex samples such as cell or tissue lysate using human samples such as blood serum or cerebrospinal fluid containing the putative autoantibody. Alternatively, antigen binding may be performed on tissue sections followed by extraction of immunocomplexes (Scharf et al., 2018). Captured protein is eluted and typically separated by SDS-PAGE followed by on-gel tryptic digestion. Differentially expressed protein bands are excised from the gel and analyzed on a mass spectrometer. Identified putative autoantigens are typically further investigated using complementary methods, e.g., cell-based assays or bead-based arrays.
Clonality and antigen diversity of the antibody repertoire

The methods for analysis of the antibody repertoire presented in this chapter provide data on complementary aspects of the antibody repertoire. In general, affinity-based methods can provide information on the antigen diversity of the repertoire, or more accurately, the ability of the antibody repertoire to bind the arrayed antigens under the present experimental conditions. This information may be used for studies pertaining to repertoire affinity, such as evaluation of antigens for serological tests, population monitoring of pathogen exposure, development of autoantibody biomarkers for stratification of patient groups, or autoantibody discovery for formulation of hypotheses for disease mechanisms. Affinity-based methods may also be used for evaluation of produced affinity binders, such as monitoring specificity of monoclonal antibodies. Affinity-based methods are, however, limited to their presented antigens and experimental conditions in their view of the antibody repertoire.

B cell receptor sequencing on the other hand can provide information on the size and individual evolution of the B cell receptor repertoire which, for instance, may provide orthogonal information on the etiology of an autoimmune disease. However, the B cell receptor repertoire typically does not fully represent the humoral antibody repertoire, both as circulating B cells may be inactive, and as antibody-producing plasma cells may be sequestered in bone marrow niches rendering them unavailable for blood sampling.

Mass spectrometry-based methods are versatile and can provide information on the clonal diversity of the humoral antibody repertoire with or without affinity-based enrichment on antigens, as well as determine bound antigens and yield sequences of individual antibodies when combined with affinity reagents. Profiling repertoire diversity without the involvement of antigen can give a wide view of the clonality of the antibody repertoire, currently down to clones with an abundance of 0.05 µg/ml, and is poised to provide increased knowledge on the properties of the circulating antibody repertoire in health and disease. However,
determining the antigens of the clonally defined antibody repertoire remains arduous.

The antibody repertoire is complex, and no single method can capture all its aspects. Therefore, our current understanding is likely to be significantly advanced by combining methods and integrating data, which I discuss in the Concluding remarks. In this chapter, I have not touched upon methods for characterizing the functionality of antibodies using, e.g., competition assays, cell-based assays, viral neutralization assays, or animal models. While the functionality of antibodies is an important property for the downstream effects of individual antibodies, I have here focused on methods for characterizing repertoires of antibodies.
Chapter III

Analysis of antibody repertoire data

Current experimental methods for the analysis of antibody repertoires generate data in the range of hundreds to tens of thousands of antibodies in single to thousands of individuals. The characteristics of these data may vary considerably depending on the experimental method and study design. Therefore, the analysis of antibody repertoire data typically requires custom strategies that are adapted to these factors and the biological, medical, or technological aims of the study.

In this chapter, I discuss several aspects of the analysis of antibody repertoire data with a focus on the methods and strategies I have used for analysis of data from planar and bead-based protein arrays.

Heterogeneity of antibody repertoires

Large interindividual heterogeneity is a prominent feature of many datasets of antibody repertoires. This heterogeneity is found across several studies and experimental platforms, such as B cell receptor sequencing, mass spectrometry-based Fab profiling, display technologies, and antigen arrays, and has been dubbed “antibody fingerprints” or “repertoire signatures”. The diversity of antibody repertoires can be assessed in terms of clonality, target antigens, or a combination.

Clonal heterogeneity refers to the number of analyzed antibodies with different Fab sequences, primarily in the hypervariable regions. As described in Chapter II, this may be assessed with mass spectrometry-based methods or B cell receptor sequencing, although the peripheral B cell receptor repertoire might not reflect the circulating antibody repertoire.
Using mass spectrometry-based Fab profiling, Bondt, Hoek, et al. (2021) showed that the global antibody repertoire was dominated by surprisingly few antibody clones, although there may have been additional undetected clones with very low abundance. While the set of high-abundant clones was highly specific to each individual and displayed high stability over time, antibody repertoires were also shown to be dynamic and responsive to immunological events, as illustrated by the emergence of previously undetected clones during sepsis. Similar findings in B cell receptor repertoires have been reported by Kräutler et al. (2020) and Briney et al. (2019).

Antigenic heterogeneity refers to the number of antigens towards which antibodies are detected. Such profiling relies on affinity binding using antigen arrays or display-based systems, and the collected antigens are typically autoantigens. Like clonal heterogeneity, antigenic heterogeneity has been observed by our group and others across several studies in health (Neiman et al., 2019; Tebani et al., 2020) and across multiple and diverse diseases such as neuropsychiatric disorders (Just et al., 2021; Zandian, Wingard, et al., 2017), multiple sclerosis (Ayoglu et al., 2013; Ayoglu, Mitsios, et al., 2016; Häggmark et al., 2013), systemic lupus erythematosus (Frostegård et al., 2018; Hu et al., 2015), systemic sclerosis (Geroldinger-Simić et al., 2023), Sjögren’s syndrome (Longobardi et al., 2020), myositis (Zaenker et al., 2020), cancers (Cui et al., 2021; Qiu et al., 2021; Zhang et al., 2020), and COVID-19 (Chang et al., 2021; Jaycox et al., 2023; Wang et al., 2021). Again, observed antibodies were highly specific to each individual and displayed high temporal stability. Furthermore, clonal heterogeneity can be assessed in antibody species targeting a specific antigen. For instance, Van Rijswijck et al. (2022) enriched antibodies targeting the SARS-CoV-2 Spike protein, followed by mass spectrometry-based Fab profiling. Clonal heterogeneity was observed also within antigen-directed antibody species.

Although interindividual heterogeneity of antibody repertoires is consistently observed across studies, there are also reports on subsets of antibodies that are commonly observed in healthy individuals. These antibodies are varying called “shared”, “common”, or “public” antibodies. A portion of public antibodies have been hypothesized to be natural autoantibodies (Elkon & Casali, 2008), which mainly are IgM and whose
affinity is determined by germline V(D)J genes and may provide a type of
innate protection against pathogens, aid in clearance of intercellular
debris, and be involved in B cell regulation (Holodick et al., 2017). Another
portion of public antibodies may stem from cross-reactivity between
foreign proteins from infectious agents and self-proteins. In a meta-
analysis, Shome et al. (2022) found 21 autoantibodies that were commonly
reported in screening of healthy individuals using proteome microarrays
and that displayed sequence similarity to respiratory and common viruses.

In summary, interindividual heterogeneity of antibody repertoires is
consistently observed across studies, cohorts, and analytical methods.
Accordingly, this has also been true for the datasets generated in my
research. In my experience, this feature of antibody data, particularly
autoantibody data, results in specific analytical challenges which are
discussed in the following sections.

Collection of antibody repertoire datasets

The collection of data is the foundation on which the following data
analysis rests. In this dependency, there are several aspects that need to be
considered, such as technological, medical, biological, and demographic
aspects, the aims and purpose of the study, as well as practical aspects such
as logistics and cost. The choice of strategy for data analysis should reflect
the data collection, such as longitudinal or cross-sectional study, the
number of analytes and multiple comparisons effects, and data size and
applicability of large-data methods.

Study design

The study design is a major aspect that should be considered in data
analysis. Most studies on antibody repertoires are cross-sectional or case-
control studies, with each participant sampled and measured once. The
relative abundance of these single time-point studies is likely coupled to
such cohorts being more common, which is connected to their relative ease
of collection (Schulz & Grimes, 2002). Cross-sectional and case-control
serological studies may prove powerful, with certain antibodies having
differential expression in group comparisons. Using this strategy, our
group and others have discovered novel autoantibodies in, e.g., multiple sclerosis (Ayoglu, Mitsios, et al., 2016), first-episode psychosis (Zandian, Wingard, et al., 2017), systemic lupus erythematosus (Hu et al., 2015), Sjögren’s syndrome (Longobardi et al., 2020), COVID-19 (Bastard et al., 2020), as well as in Paper I and V.

As illustrated by these studies, public antibodies with high prevalence in the case group can be detected in case-control or cross-sectional studies. However, detection of antibody sets for characterization of patient subgroups has proven difficult with these study designs. This might be connected to the large intraindividual heterogeneity of antibody repertoires, which may be regarded as highly variable baseline levels of antibodies in health and disease. Aside from the strongest group differences, untangling these baseline levels from antibody levels associated with disease is challenging without longitudinal measurements. Therefore, the analysis of antibody repertoires benefits from longitudinal study designs. Furthermore, longitudinal studies can provide insight into antibody emergence and dynamics which are important parameters in many settings, such as infectious diseases, post-viral syndromes, and vaccine studies.

Using a longitudinal study design, antibody trajectories can be constructed for each individual. By designating a sample as baseline (preferably pre-morbid), intraindividual fold changes can be obtained which allow for comparisons of antibody levels with respect to an event of interest, e.g., infection, disease onset, or treatment. With this measure, emergent new-onset antibodies can be defined, which aids in accounting for individual background levels and may carry larger sensitivity for the detection of patient subgroups.

Longitudinal cohorts are however less common than single time-point cohorts, in particular prospective longitudinal cohorts (Grimes & Schulz, 2002). As a result, there are relatively few examples of longitudinal studies of antibody repertoires, although some have been conducted following the COVID-19 pandemic (Chang et al., 2021; Jaycox et al., 2023).
Method-dependent background signal levels

The technological platform used for the experimental analysis of antibody repertoires will affect the resulting data and its analysis. In protein arrays, background signal levels are a major factor. This signal reflects the inherent ambient readout of the system and may for instance involve non-specific binding, scattering effects, or analyte impurities. Tag-specific background signal is an important category for antigen arrays. This results from protein tags used for, e.g., purification, that are present on the arrayed antigens. An antibody pre-blocking step using a buffer containing free tag constructs may be necessary in the experimental procedure to reduce such background signal. This strategy was used in Paper I and V, to reduce signals from antibodies binding the His₆ABP tag of the Human Protein Atlas antigens (Nilsson et al., 2005). These anti-His₆ABP antibodies occur frequently, as ABP is derived from a streptococcal protein (Larsson et al., 2000; Nygren et al., 1988). In addition, adjustment for background signals may be performed computationally. Depending on the type of background signal and the aims of the analysis it may require different treatment ranging from no compensation for background levels to extensive normalization.

In a protein array analysis, the background level may be accepted if it is consistent across samples and assays and only relative results are required. Studies consisting of a single assay and a consistently sampled cohort might not require normalization. However, introduction of a control cohort often requires normalization, as the control individuals may have been sampled under different circumstances and at a different time which may affect signal levels. Furthermore, combination of data from different arrays often requires normalization due to batch effects.

In molecular biology assays, a common approach for normalization is subtraction of background levels. In bead arrays, this may be measured using a negative control bead, which has a coupled analyte with an expected negative signal. This may for instance be a bead with a ubiquitous tag present, such as biotin in the case of peptides coupled using biotin-neutravidin chemistry, or a bare bead (exposed to the coupling chemistry without addition of an antigen). In planar arrays, background subtraction can be performed by selecting a region immediately surrounding the
antigen spot and measuring the fluorescence intensity for a relative quantification of the local background level.

Lacking a suitable representation of the background level it can be estimated. In exploratory studies of antibody repertoires, most assayed individuals do not have antibodies against most of the arrayed antigens, resulting data sparsity as discussed below. This can be exploited to generate individual, antigen, or overall estimates of background signal levels by a summary statistic such as the median or mean. These noise estimates may be subtracted or used for distribution-based normalization such as ordinary or robust Z-scoring.

Data sparsity in antibody repertoires

Depending on the antigen set and cohort used, experimental analysis of antibody repertoires may yield data containing mostly signals at background or noise levels. Such data are called sparse data and arise in systems where there are few pairwise interactions (Davis, 2006).

For the analysis of antibody repertoires, this translates to studies where there are few interactions between the set of arrayed antigens and the set of assayed individuals, i.e., where most individuals do not have most antibodies. This situation occurs frequently in antibody screening studies, where hundreds to tens of thousands of potential antigens are assessed in hundreds to thousands of individuals. The interindividual heterogeneity of antibody repertoires results in a large proportion of features (antigens) displaying signals above background levels in only single samples, resulting in sparse datasets. Other study settings where there is an expected high prevalence of the examined antibodies may not yield sparse data, e.g., studies of a focused set of previously described autoantibodies in a cohort of individuals with the associated autoimmune disease (Nielen et al., 2004), or studies of a limited set of anti-viral antibodies in a cohort of individuals with the associated viral infection (Solastie et al., 2023).

Analysis of sparse data may require specialized methods. Methods based on direct comparison of group means or medians are often unsuitable, as the mean or median signal of groups is dominated by background levels. Although there are specialized methods and models for the analysis of
Analysis of antibody repertoire data

sparse data, such as hurdle models, zero-inflated models, and dimensionality reduction techniques (e.g., UMAP, PCA), I have found that the data sets I have been working with have contained too few non-zero observations for applicability of these methods. Instead, I have largely employed data binarization to facilitate analysis.

Transformation of antibody repertoire data

Binarization

Sparsity is observed in many published studies on antibody repertoires, and binarization can be a powerful tool for analysis of antibody repertoire data. However, loss of data resolution as data is transformed from continuous to binary is an inherent feature of binarization and a potential drawback. In addition, binarization comes with the non-trivial task of selecting appropriate cutoffs. Chang et al. (2021) had a case-control study design, where the detection threshold for each autoantibody was set to 5 or 3 times the standard deviation above the mean of healthy controls. Wang et al. (2021) had a similar approach, using various integer cutoffs throughout their analysis, often a method-specific score exceeding 2. Depending on the normalization strategy, constant cutoffs such as these may be suitable for binarization (Figure 3a). In addition, there are strategies for variable cutoff selection (Figure 3b). The choice of cutoff selection strategy is dependent on the aim and next steps of the analysis.

In **Paper I**, we conducted a cross-sectional study of autoantibody repertoires in psychotic disorders. Here, we applied a method for selection of adaptive antigen-specific cutoffs for binarization of the autoantibody response to characterize autoantibody reactivity.

In brief, the binarization method relied on the assumption of data sparsity; most individuals do not have most autoantibodies. Autoantibody levels were normalized in each sample using the robust Z-score to account for sample-specific background levels. For each autoantibody, the cutoff was defined as the local minimum adjacent to the global maximum of the kernel density estimate of the robust Z-score. While this method of cutoff selection for binarization provides a systematic approach, it requires
manual selection of the parameters for the kernel density estimation and definition of the local minimum.

With the resulting binarized autoantibody responses, we were able to assess the prevalence of autoantibodies and perform a stepwise selection based on prevalence criteria. Among the selected autoantibodies, we identified six autoantibodies associated with patient subgroups having specific symptoms of psychosis.

Figure 3 | Binarization of antibody repertoire data. Binarization may be conducted using (a) a static cutoff for all antigens or (b) a variable cutoff based on the distribution of (normalized) antigen-derived signals, here shown for a single antigen. Regardless of method, the antibody repertoire is binarized (c).
Binarized antibody responses that reach the applied cutoff are commonly referred to as “reactivities” (Figure 3c). This term tends to carry the connotation that a positive response (reactivity) denotes the detection of the antibody of interest, while a negative response (non-reactive) denotes no detection. While binarization itself can be useful, the concept of reactivity may inflate confidence in the accuracy of the cutoff even when this is not warranted. All downstream group comparisons based on binarization will be influenced by the level of the binarization cutoff and may impact main conclusions of an associative study. The effect may be mitigated by applying a sensitivity analysis resembling the one performed in Paper I, but the cutoff influence will remain and still require consideration by the investigators.

Based on this, I believe efforts are needed to progress from binarization strategies to analysis of continuous data, even though binarization currently may serve as a useful tool for the initial phases of analysis. In Paper V, I utilized the longitudinal study design and refrained from binarization with the aim of reducing the effect of arbitrary cutoffs on downstream analysis. With larger datasets and adaptation of the models, it is possible that zero-inflated models, hurdle models, or UMAP might be successfully deployed for analysis of antibody repertoire datasets in the coming years.

Stratification

Aside from binarization, stratification may be used to divide antibody responses into several categories. In Paper V, we conducted a prospective study of autoantibodies emerging with COVID-19. The longitudinal study design enabled further methods for autoantibody stratification using the fold change of autoantibodies at disease onset compared to baseline samples. By applying PAM (partitioning around medoids) clustering, we stratified autoantibodies as stable (persistent), transient, and delayed new-onset autoantibodies emerging with COVID-19, and those that did not, resulting in the binary categories new-onset and non-new-onset autoantibodies.

The PAM clustering approach for autoantibody stratification utilized the fold change at the time of infection and the first follow-up sample, relative
to the pre-infectious baseline sample. As we were interested in both the size and the timing of the autoantibody response, we used the Euclidean and cosine distance metrics. If the fold change at the two time points is laid out on a two-dimensional plane, Euclidean distance represents the closeness of magnitude of data, while the cosine distance represents the angular closeness, which then represents the closeness in time. Multiplying the two distance metrics yields a combined measure which accounts for both size and timing of autoantibody responses.

With autoantibody responses stratified, associations of the most prevalent new-onset autoantibodies with symptoms remaining for several months after COVID-19 and severity of acute COVID-19 were made using variants of logistic regression.

Quantification of antibodies

Absolute measurement of biological data may increase reproducibility and facilitate comparisons between data sets and studies. Typically, biochemical assays include the construction of standard curves for conversion of recorded data to, e.g., concentration. However, absolute measurement of antigen-binding antibodies is hampered by the intrinsic properties of antibody affinity.

The affinity of an antibody to its target epitope depends on the physicochemical properties of its paratope, as outlined in Chapter I. In an affinity-based multiplexed serological assay, the measured value is typically proportional to the number of antibody-antigen complexes formed at each array location. In a simple scenario, with a single (monoclonal) antibody species and a single antigen, this number is determined by the affinity or dissociation constant of the affinity complex, the concentrations of the reactants and product, as well as the rates of complex formation and incubation time. In this scenario, only the concentration of antibody is variable, giving a signal proportional to this parameter. Using the same monoclonal antibody as the one measured, a standard curve may be constructed for absolute quantification of antibody concentration.
In a complex sample containing a diverse mixture of antibodies as well as other proteins, the signal is additionally influenced by several other parameters. Importantly, there may be several different antibody species targeting the same antigen in the sample, for instance resulting from a polyclonal immune response. As these antibodies have different paratopes and may bind different epitopes of the same antigen, their affinity constants may differ several orders of magnitude (Gorris & Soukka, 2022). Since the exact mix of antibodies is not known in this scenario, it is not possible to construct a standard curve for absolute quantification of antibody concentration, partly due to the differing affinity constants. In addition, the different antibody species might competitively bind the antigen, further affecting the relation between signal and antibody concentration. Furthermore, multiplex affinity-based serological experiments are typically performed at a single or sometimes two sample dilutions and concentrations of arrayed antigens. This may result in coverage of different parts of the signal range for different antibodies with different concentrations and affinity constants, affecting the linearity of signal readout. In addition, non-native antigen representations may contribute to unspecific binding by antibodies that do not bind the native antigen of interest, further complicating the construction of standard curves.

Despite these complicating factors, reports using standard curves of monoclonal antibodies for quantification in complex samples do exist in the literature (Bartsch et al., 2021). However, they might be employed as a standardizing measure for calibration across assays and estimation of the limits of quantification of an assay (Yman et al., 2019). Standardization may also be carried out using an established international standard such as the WHO first International Standard for anti-SARS-CoV-2 immunoglobulin (NIBSC code 20/136). This standard consists of a lyophilized sample pool arbitrarily defined as equivalent to 1000 binding antibody units (BAU)/ml which is intended for establishment of a conversion factor or local reference preparations or both (Knezevic et al., 2022; Kristiansen et al., 2021). Although this standardization does not yield absolute measures, it in theory gives comparable relative measures between assays. However, substantial discrepancies in antibody levels have been observed despite conversion to BAU/ml, raising concerns on its
reliability with current usage (Infantino et al., 2021; Perkmann et al., 2021; Saker et al., 2022).

Antibody quantification may be performed using mass spectrometry. Spiking in known amounts of monoclonal antibodies of known masses before purification of IgG from blood plasma samples enables determination of concentration of detected clones in full antibody repertoires (Bondt, Hoek, et al., 2021). Determining the concentration of antibody clones targeting a specific antigen may be more difficult. If internal standards are spiked in after antigen-specific enrichment, the enrichment yield is not accounted for. This yield may vary but has been reported to be 85% for a monoclonal antibody (Bondt, Hoek, et al., 2021). On the other hand, if antigen-directed standards are used and spiked in before enrichment on the target antigen, the affinity constants of different clones and the monoclonal antibody may give rise to challenges in defining absolute concentration. However, in enrichment procedures with high yield, this effect may be diminished by the large excess of antigen over antibody amount.

Associations with the antibody repertoire

In this chapter, I have discussed several aspects of antibody repertoire data with emphasis on the antigen diversity of autoantibody repertoires as detected by the array-based methods I have used for the papers of this thesis. Antibody repertoires detected with these (and other) methods are typically heterogeneous which often results in sparse data, particularly in autoantibody screening studies. Sparse data can be challenging to work with, and binarization or stratification can be a useful tool to transform the data for identification of antibodies of interest. A longitudinal study design is a powerful approach to identify potentially interesting antibodies as it ameliorates the inter-individual heterogeneity by enabling intra-individual comparisons and provides a temporal perspective on the antibody repertoire.

In many studies, the detection of antibodies of interest is only the first step of the analysis. Depending on the aim and scope of the study, there are often hypotheses on the association between antibodies and other variables, such as infection, autoimmune disease, or disease symptoms.
There are numerous statistical tests and models which may be considered in these analyses, and I do not elaborate on these here. I do, however, note that finding novel autoantibodies with a clear connection to clinical parameters remains a challenging task. Autoantibodies detected in screening studies are often found in small subsets of the cohort under study. Therefore, large cohorts are required to reach sufficient statistical power for significant associations. This is especially challenging in exploratory screening studies, where there often are several variables for association with the often many detected autoantibodies. Running multiple statistical tests in an exploratory manner may require correction of the false discovery rate (FDR) to avoid false positive findings, effectively making the threshold for statistical significance more stringent with each subsequent test. However, there is no consensus on the correct use of FDR correction in exploratory studies. In some studies, statistical measures of significance, e.g., $p$ values, that stem from multiple exploratory comparisons are used as a ranking tool. This may be appropriate if the aim is feature selection and no claims of statistically supported conclusions are made. On the other hand, there are statistical methods tailored for feature selection that may be better suited for the task. Another approach is detailed a priori consideration of the hypotheses and associated and possible pre-registration of the study, mirroring the setup of clinical trials. Proponents consider this practice as reducing the need for FDR correction, although it may not be compatible with the exploratory nature of such studies. Ultimately, it is crucial to move beyond the exploratory stage and validate findings with orthogonal methods to establish biologically or clinically significant associations. This is, however, not an easy task which often requires considerable resources and time, and which has had limited space within the present investigation.
Chapter IV

Present investigation

The work presented in this thesis revolves around analysis of patterns in antibody repertoires for generation of clinical and biological insights and hypotheses. This work follows two distinct but related areas of serology: autoantibody profiling and SARS-CoV-2 serology.

Autoantibody profiling was employed in Paper I and Paper V (Table 1). As discussed in Chapter II, analysis of autoantibody repertoires may be performed using several technologies alone or in combination. In the present work, we have used a combined design using both planar arrays and bead arrays. This design utilizes the high multiplexity of the planar arrays for autoantibody discovery and antigen selection, and the high throughput of bead arrays for screening of large cohorts. In the discovery phase, planar arrays were used to screen carefully selected samples. Here, 24 to 32 samples were combined in groups of 4 and each group was screened on a planar array containing 42 000 protein fragments produced in the Human Protein Atlas, presented in Chapter II. In Paper V, the range of antigens covered in the screening phase was extended with the Human Protein Atlas Secretome array, which contains 1500 full-length secreted or membrane-bound proteins. After applying the background corrections and normalizations discussed in Chapter III, the target antigens of detected autoantibodies with the highest levels or that were corroborated by previous evidence were selected for inclusion on a bead array. This array was used in the following screening phase to assay the entire cohort.

SARS-CoV-2 serology was performed in Paper II-V using an assay described in Hober et al. (2021). During its development, one hundred SARS-CoV-2 antigen representations were evaluated on an increasing set
of post-infectious and pre-pandemic samples, culminating with the evaluation of five select antigens on 2400 samples.

After acquisition, data were transformed using the methods outlined in Chapter III, followed by data analysis. This phase is complex as the analysis often progresses iteratively and new questions and hypotheses might emerge. Complementary methods may be used to further explore emerging findings. Discussions with collaborating clinicians are vital for interpretation and direction of data analysis.

The present investigation has been conducted in a cross-sectional cohort of patients with psychosis (Paper I), and a longitudinal and prospective cohort of patients and healthcare workers with COVID-19 (Paper II-V), as summarized in Table 1. In Paper I, we explored the autoantibody repertoires of patients with psychosis for identification of connections to clinical features. In Paper II, we investigated the SARS-CoV-2 seroprevalence in healthcare workers early in the COVID-19 pandemic and discovered connections to occupational risk factors and COVID-19 symptoms. In Paper III, we examined the antigen specificity and four-month persistence of the serological response to SARS-CoV-2, and the connection to viral neutralization. We conducted an eight-month follow-up in Paper IV, where we investigated the persistence of the cellular and humoral response and their protective effect against reinfection. In Paper V, we combined the approaches of the previous papers on a subset of the cohort. Here, we determined the prevalence of autoantibodies emerging after SARS-CoV-2 infection and evaluated their persistence during one year after infection. In addition, we explored connections to symptoms post-COVID-19 and used peptide-based epitope mapping to identify sequence similarity of human and viral antigens which may indicate molecular mimicry.

Table 1 | Overview of cohorts and arrays used in the present investigation.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Paper</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COVID-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep 2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan 2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep 2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein fragment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-length protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peptide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bead-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein fragment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-length protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

44
Paper I: Autoantibody profiles associated with clinical features in psychotic disorders

Autoantibodies are known to be pathologically involved in a portion of neuropsychiatric disorders. Notably, antibodies targeting the N-methyl-D-aspartate (NMDA) receptor may cause anti-NMDA receptor encephalitis. However, autoantibody screening studies are scarce in neuropsychiatric disorders and additional autoantigens associated with psychopathology likely remain to be discovered, which has been demonstrated in a previous comparative study in first-episode psychosis patients and healthy controls (Zandian, Wingard, et al., 2017). Broader screening in larger patient groups is positioned to identify autoantibodies associated with clinical features which could give insights in pathophysiology and inform patient stratification and choice of treatment.

We first conducted an antigen discovery phase on planar arrays as described in Chapter II. The planar arrays were used to assay the autoantibody repertoire of 6 groups of individuals recruited in the 2010 second Australian National Survey of High Impact Psychosis (SHIP). Each group consisted of 4 unique individuals sharing specific symptoms of psychosis, and the 180 antigens with reactive antibodies in at least one group were included on the following bead array together with 200 additional antigens indicated in previous literature. Using this 380-plex bead array, we assayed the autoantibody repertoire of 461 individuals with psychosis recruited in SHIP. This repertoire was binarized using the adaptive cutoff method discussed in Chapter III.

Examining the autoantibody count, we observed that individuals with many autoantibodies tended to be female and to have family histories of obesity and psychiatric disorders. Conversely, we observed a trend for individuals with few autoantibodies to be treated with the atypical antipsychotic Clozapine. In line with the tendency of Clozapine to induce secondary antibody deficiency, this finding may suggest that autoantibody suppression could contribute to its clinical efficacy.

Next, we selected candidate autoantibodies for patient stratification using a stepwise selection procedure based on three clinically informed criteria. First, we required autoantibodies to be present in at least 10 individuals. Second, 85% of these individuals should present with a specific psychotic
symptom. Third, this symptom should be 25% more prevalent in individuals with the autoantibody than those without it. Finally, formal group comparisons were conducted on the resulting list of autoantibodies and adjusted for multiple comparisons.

Using this analysis approach, we found associations of six IgG autoantibodies with specific psychopathology (Figure 4a): anti-AP3B2, persecutory delusions; anti-TDO2, hallucinations; anti-CRYGN, insomnia; anti-APMAP, poor appetite, anti-OLFM1, better cognitive performance; and anti-WHAMMP3, anhedonia and dysphoria. Odds ratios (OR) and associations were obtained using logistic regression. An OR of 10 indicates that the odds of having a given feature (e.g., a specific symptom) is 10 times higher in the group with vs without the autoantibody. For further characterization, we investigated additional clinical and biological features in individuals with these autoantibodies (Figure 4b). Of the six symptom-associated autoantibodies, anti-AP3B2 and anti-TDO2 IgG are the most interesting.

We detected anti-AP3B2 IgG in 23 individuals with a record of persecutory delusions and mainly affective psychotic disorders. Furthermore, Gandal et al. (2018) have showed that mRNA levels of AP3B2 were disrupted in postmortem brain among several psychiatric disorders, and that the gene co-expression network containing AP3B2 was associated with synaptic changes across multiple diagnoses.

We found anti-TDO2 IgG in a different set of 23 individuals. All of these had experienced hallucinations, and they had an increased proportion of increased levels of the pro-inflammatory cytokines IL-6 and IL-4. Interestingly, TDO2 is the main rate-limiting enzyme of the neuromodulatory kynurenine metabolic pathway which has been shown to be altered in individuals with psychosis (Erhardt et al., 2017).

In summary, we found six autoantibodies that had novel associations with specific transdiagnostic symptoms. These findings may enable clinically relevant patient stratification and personalized treatment if replicated and further developed.
Figure 4 | Profiles of clinical and biological parameters in individuals with symptom-associated autoantibodies. Bar charts and forest plots display prevalence and odds ratios with 95% CI in each group, respectively. A Symptom profiles of the six symptom-associated autoantibodies. B Additional associated clinical and biological features of the six symptom-associated autoantibodies. Half diamonds and half squares indicate zero or infinite estimates of odds ratios, respectively. Pos and Neg denote seropositive and seronegative with respect to the antigen. Medians: DSCT score, 39; IL-4 and IL-6, <1 fg/mL; IL-8, 3.6 pg/mL; Cu, Zn-SOD, 333 ng/mL.
Paper II: SARS-CoV-2 exposure, symptoms and seroprevalence in healthcare workers in Sweden

In the first months following the emergence of the COVID-19 pandemic, knowledge was incomplete on the spread, exposure, risk factors and symptomatology of the disease, as well as most other factors. In particular, the occupational risk for front-line healthcare workers (HCW) under Swedish healthcare guidelines was not known. To investigate these parameters in the immediate and long term, the COMMUNITY study was launched in early 2020 at the Danderyd Hospital in Stockholm, Sweden. In parallel, we developed a multiplex and high-throughput bead-based SARS-CoV-2 serological assay with high sensitivity and specificity by evaluating over 100 antigens in up to 2400 samples (Hober et al., 2021). The final assay utilized two SARS-CoV-2 Spike protein antigens and one Nucleocapsid protein antigen. A reactivity threshold for each antigen was determined using a select set of 12 pre-pandemic negative control samples. By assessing several antibodies, as discussed in Chapter I, anti-SARS-CoV-2 IgG seropositivity was called when at least two antigens were reactive, which resulted in a sensitivity of 99.7% (95% CI 98.3-100%) and a specificity of 100% (99.8-100%) (Hober et al., 2021). Here, this assay was applied to the COMMUNITY cohort.

Among the 2149 recruited HCW, 19% (n=410) were anti-SARS-CoV-2 IgG seropositive. Notably, this is a higher proportion than in the general population of the Stockholm region during the same period, as well as in other contemporary reports on HCW seroprevalence. This may be explained by the personal protective equipment (PPE) practices at Danderyd Hospital at the time, where scarcity made reuse of PPE required, notably also between individuals.

Among the 410 seropositive HCW, 9% (n=37) reported experiencing no symptoms of COVID-19 since January 2020. At this time, PPE was only worn in contact with known or suspected COVID-19 patients and polymerase chain reaction (PCR) screening was not available for HCW. Together, this suggests that there were opportunities for SARS-CoV-2 transmission from HCW to patients.
Furthermore, HCW with patient contact had increased odds of seropositivity compared to HCW without any patient contact, regardless of whether the patient contact was in COVID-19 wards or not (OR [95% CI] 3.3 [2.2-5.3] and 2.3 [1.5-3.8] with COVID-19 patient contact and non-COVID-19 patient contact, respectively; Figure 5). This suggests that PPE practices were insufficient to protect against both known and unknown COVID-19 exposure. Notably, aerosol-filtering face masks were restricted to aerosol-generating procedures, e.g., intubation.

![Figure 5 | Association of occupational exposure with seroprevalence of anti-SARS-CoV-2 IgG. Odds ratios of seropositivity given patient contact, non-COVID-19 patient contact, or COVID-19 patient contact compared to no patient contact, and (under the horizontal black line) given COVID-19 patient contact compared to non-COVID-19 patient contact. Data are presented as odds ratios and 95% confidence intervals obtained using two-sided Fisher’s exact test.](image)

The most common symptoms in the seropositive HCW were headache (66%), malaise (65%), fever (57%), anosmia (53%), cough (52%) and ageusia (50%). While all these symptoms were associated with increased odds of seropositivity, anosmia and ageusia (loss of sense of smell and taste, respectively) had the highest odds ratios by far (OR [95% CI] 28.4 [20.6–39.5] and 19.2 [14.3–26.1], respectively; Figure 6), corroborating contemporary emerging evidence of a strong association between anosmia, ageusia, and COVID-19.

In conclusion, we provided new evidence early in the COVID-19 pandemic that the current PPE practices were insufficient for preventing SARS-CoV-2 transmission between HCW and patients in a hospital setting. Furthermore, we corroborated early evidence that anosmia and ageusia were highly indicative of COVID-19. Taken together, this could inform the implementation of practices for prevention of SARS-CoV-2 transmission, such as adequate PPE and recommendations of self-isolation based on self-assessed symptoms.
Figure 6 | Associations of prior symptoms with seroprevalence of anti-SARS-CoV-2 IgG. Odds ratios of seropositivity for individually reported symptoms. Odds ratios were calculated using two-sided Fisher’s exact test with n = 2149 independent individuals. No adjustment for multiple comparisons was applied. Data are presented as odds ratios and 95% confidence intervals.
Present investigation

Paper III: SARS-CoV-2 induces a durable and antigen specific humoral immunity after asymptomatic to mild COVID-19 infection

As the pandemic continued, we expanded our efforts to track the spread and effects of the virus. We and others were interested in the duration and antigen specificity of neutralizing humoral immunity after SARS-CoV-2 infection, which we assessed in the four-month follow-up of the COMMUNITY study.

At the four-month follow-up, 1965 HCW and 59 convalescent COVID-19 patients remained in the study. Among HCW, 96% (366/381) remained anti-Spike IgG positive, and 8% (134/1584) developed anti-Spike IgG for the first time. In contrast, anti-Nucleocapsid IgG remained in 80% (342/430) of HCW with previous anti-Nucleocapsid IgG. Additionally, 7% (113/1535) developed new anti-Nucleocapsid IgG. Antibody levels at follow-up remained at higher levels in convalescent COVID-19 patients and in symptomatic HCW.

A cell-based microneutralization assay (Varnaitė et al., 2020) was used to determine SARS-CoV-2 neutralization capacity in the 425 individuals who remained anti-Spike IgG positive at four-month follow-up (59 convalescent COVID-19 patients and 366 HCW) and in a subgroup of HCW who were anti-Spike IgG negative at four-month follow-up (197 HCW). Neutralization capacity indicates whether the serum or antibodies can prevent viral entry into cells, i.e., neutralize the virus. Among individuals remaining anti-Spike IgG positive at follow-up, 94% (401/425) showed neutralization capacity, which was not found in any of the anti-Spike IgG negative individuals (n=197). Among individuals who were anti-Spike IgG positive at study inclusion (n=381), follow-up levels were considerably lower in individuals without neutralization capacity (Figure 7a). However, we observed neutralization capacity across a wide range of anti-Spike IgG levels. Interestingly, the odds of showing neutralization capacity were almost ten times higher in symptomatic than asymptomatic HCW among those that remained anti-Spike IgG positive at follow-up (Figure 7b). Furthermore, individuals with anti-Nucleocapsid IgG alone did not display any neutralization capacity (Figure 7c).
In summary, we corroborated contemporary findings that neutralization capacity is closely connected to anti-Spike IgG levels, but not anti-Nucleocapsid IgG levels, as well as that anti-Spike IgG levels remain elevated for at least four months after infection. Therefore, serological assays for assessment of long-term immunity should be based on the Spike protein.

Figure 7 | Virus neutralization was confirmed in the vast majority of anti-Spike IgG positive samples and was associated to COVID-19 symptoms. A Among individuals who were anti-Spike IgG positive at study inclusion, anti-Spike IgG levels at four-month follow-up were significantly higher in individuals with than without SARS-CoV-2 neutralization capacity. Purple: Anti-Spike IgG positive individuals at four-month follow-up. Grey: Anti-Spike IgG negative individuals at four-month follow-up. B Among HCW who were anti-Spike IgG positive at both study inclusion and follow-up, neutralization capacity was significantly associated with COVID-19 symptoms prior to study inclusion. C Samples with high levels of anti-Nucleocapsid IgG alone did not show any neutralization capacity. Green: With neutralization capacity. Brown: No neutralization capacity. Circles (panel A and C): HCW. Triangles: Convalescent COVID-19 patients. AU: Arbitrary Units.
Paper IV: Robust humoral and cellular immune responses and low risk for reinfection at least 8 months following asymptomatic to mild COVID-19

As anti-Spike IgG was identified as persistent and connected to neutralizing capacity, we focused on this antigen in the eight-month follow-up of the COMMUNITY study. As the supply of vaccine doses in Sweden was still scarce, knowledge on the long-term humoral and cellular protection conferred by natural infection could be of importance for public health measures.

In total, 1884 HCW and 51 convalescent hospitalized COVID-19 patients participated in the eight-month follow-up study. All 51 patients and 96% (355/370) of the HCW who had been anti-Spike IgG positive at study inclusion remained positive at the eight-month follow-up. In line with a humoral response dependent on disease severity, the levels of anti-Spike IgG were twice as high in convalescent patients than in HCW eight months after seroconversion. Furthermore, HCW who seroconverted less than 4 months from sampling had moderately higher anti-Spike IgG levels than HCW who seroconverted 8 months or more from sampling, indicating that anti-Spike IgG levels remain relatively stable at least 8 months from infection (Figure 8a).

To investigate whether these patterns were reflected in the cellular immune response to SARS-CoV-2 infection, the full cohort of 1884 HCW and 51 patients was examined using an IFN-γ release assay. This assay is based on T-cell stimulation using a SARS-CoV-2 specific peptide pool (Mangsbo et al., 2021). The IFN-γ response was three times higher in patients than in HCW 8 months after infection, indicating that the cellular immune memory follows the same pattern of dependence on disease severity as the humoral persistence. Repeating this pattern, HCW who seroconverted less than 4 months from sampling had a moderately higher IFN-γ response than HCW who seroconverted 8 months or more from sampling (Figure 8b).

Next, we asked whether these persistent immune responses are protective against reinfection. To address this, we performed a 12-week SARS-CoV-2 quantitative PCR (qPCR) screening conducted weekly on 252 HCW who were anti-Spike IgG positive at study inclusion and on 48 HCW who were anti-Spike IgG negative. In the positive group, the cumulative incidence of
qPCR-confirmed infection was 1% (3/252; 0.13 per 100 weeks at risk), compared to 23% (11/48; 2.78 per 100 weeks at risk) in the negative group (Figure 9), yielding a protective effect of 95.2% (95% CI 81.9%–99.1%) among anti-Spike IgG positive HCW, and an incident rate ratio of 0.05 (95% CI 0.01–0.18).

In summary, we presented new data supporting a humoral and cellular immune response after asymptomatic, mild, and severe COVID-19 that lasts for at least 8 months. Furthermore, we showed that previous asymptomatic to mild SARS-CoV-2 infection with development of anti-Spike IgG is coupled to a markedly reduced risk of reinfection for at least 8 months. At a time with limited vaccine supplies, consideration of anti-Spike IgG serostatus could have guided COVID-19 vaccination policies.
Paper V: Prevalent and persistent new-onset autoantibodies in mild to severe COVID-19

During and after the COVID-19 pandemic, there has been increasing evidence of immune dysregulation in subgroups of infected individuals. Autoantibodies against type I interferons have been detected in 5–20% of severe COVID-19 cases and may affect the efficacy of therapeutic interventions (Knight et al., 2021; Puel et al., 2022). In addition, autoantibodies binding a wide range of antigens have been detected after COVID-19 and a portion has been shown to display functional binding indicative of detrimental effects on disease progression (Wang et al., 2021). To address the heterogeneity of antibody repertoires and facilitate associations to symptoms and clinical outcomes, longitudinal studies of emergent autoantibodies have been performed in hospitalized patients with COVID-19 (Chang et al., 2021; Jaycox et al., 2023). These studies of new-onset autoantibodies have, however, been constrained by the lack of pre-infectious samples and short follow-up times, and associations to symptomatology have been limited.

Combining our array platforms for autoantibody profiling and SARS-CoV-2 serology, we expanded our analysis in a subset of the COMMUNITY cohort. We selected 478 HCW and 47 hospitalized COVID-19 patients participating in the five visits during May 2020 to September 2021. In total, 389 individuals did not display anti-SARS-CoV-2 IgG at study inclusion but seroconverted during the study period, before vaccination, indicating that they were infected with SARS-CoV-2. In the discovery stage of the study, we investigated the autoantibody repertoire of 32 HCW and 16 hospitalized COVID-19 patients on our planar arrays. Based on these arrays and the literature, we selected 307 protein fragments and 56 secreted full-length proteins for inclusion on a bead array, which was used to investigate the autoantibody repertoire of the full set of 2532 samples.

New-onset autoantibodies were assessed in two steps. First, the autoantibody trajectories of the individuals with baseline seronegative samples were considered. The trajectories were defined by taking the fold change (FC) of the first and second seropositive samples relative to the last seronegative sample. These trajectories were stratified using partitioning around medoids (PAM) clustering and three categories of new-onset...
Autoantibodies with distinct dynamics were identified: acute, transient, and delayed new-onset autoantibodies (Figure 10a). These new-onset autoantibodies were found in 204 individuals and targeted a total of 187 antigens. The 12-month persistence of these new-onset autoantibodies was found to follow the antibody grouping, with 95% of stable new-onset autoantibodies and 60% in general persisting at FC ≥ 2 at 12 months after infection. The antibody repertoire was heterogeneous, as most individuals displayed single autoantibodies and the majority of new-onset autoantibodies were detected in single individuals (Figure 10b and c). However, 22 autoantibodies were detected in >1% of the cohort (>4 individuals).

In the second step, the 22 most prevalent new-onset autoantibodies were examined in the individuals that were seropositive at study inclusion. We built a model for classification of new-onset autoantibodies using the aggregated categories acute new onset (stable or transient) and delayed new-onset of the previously assessed individuals. This multinomial linear regression (MNL) model classified 98 autoantibody trajectories as acute new-onset and 56 as delayed new-onset in 79 individuals. Combining new-onset autoantibodies detected in individuals with seronegative and seropositive baseline samples, 43% of HCW (n=196/456) and 72% of hospitalized COVID-19 patients (n=33/46) displayed one or more of the 22 most prevalent new-onset autoantibodies.

After identifying the most prevalent new-onset autoantibodies, we asked whether they were associated with disease course. We found that 10 of the 22 most prevalent autoantibodies had significantly increased prevalence in the hospitalized COVID-19 patients than the HCW. Furthermore, the odds of neuropsychiatric symptoms post-COVID-19 were significantly increased in HCW with anti-CALU (calumenin), anti-MYO16 (unconventional myosin-XVI), or anti-SNURF (SNRPN upstream reading frame protein) IgG (OR = 2.8, 5.3, 2.5; p = 0.01, 0.03, 0.01, respectively).

Furthermore, we investigated what epitopes were targeted by the most prevalent new-onset autoantibodies. We performed epitope mapping of eight selected autoantibodies using a bead array of overlapping peptides and found main epitopes of six autoantibodies (Figure 11a). Using
validation cohorts of pre-pandemic healthy controls and patients with COVID-19 with neurological symptoms we could validate the main epitopes of anti-TRIM63 (E3 ubiquitin-protein ligase TRIM63), anti-SNURF, and anti-NPC1 (NPC intracellular cholesterol transporter 1) IgG (Figure 11b). Aligning the corresponding peptides to the SARS-CoV-2 proteins, we found sequence similarities between the main epitopes of the

Figure 10 | Prevalent and persistent new-onset autoantibodies emerge with COVID-19. a Persistence of new-onset autoantibodies across categories. Black lines represent new-onset autoantibody trajectories based on fold change in relation to seronegative baseline. Blue lines and shaded areas represent median and quartiles, respectively. b Distribution of new-onset autoantibodies among baseline seronegative individuals. Bars depict the number of individuals with the indicated number of new-onset autoantibodies. c Distribution of new-onset autoantibody prevalence. Bars depict the number of new-onset autoantibodies with the indicated prevalence. The 22 most prevalent new-onset autoantibodies (prevalence >1%) are indicated with gene names.
muscle proteins TRIM63 and CCDC63 (coiled-coil domain-containing protein 63) with the conserved fusion peptide of the SARS-COV-2 Spike glycoprotein, which may indicate molecular mimicry.

In summary, we showed that prevalent and persistent new-onset autoantibodies emerge in mild to severe COVID-19. Some autoantibodies were more prevalent in hospitalized patients with COVID-19 or were associated with neuropsychiatric symptoms post-COVID-19. We demonstrated that 10% of the study cohort developed anti-TRIM63 and anti-CCDC63 IgG and revealed their main epitopes using epitope mapping. These epitopes sequence similarities suggestive of molecular mimicry with the fusion peptide of the SARS-CoV-2 Spike glycoprotein, which is

Figure 11 | The main epitopes of new-onset autoantibodies are elevated in independent cohorts with severe COVID-19 and neuro-COVID. a Epitope mapping revealed the main epitopes of six new-onset autoantibodies. Lines depict epitope profiles at new-onset in individuals with the corresponding new-onset autoantibody. Gray lines depict the mean. The dashed line indicates the cutoff for classification as a main epitope (FC ≥ 2.52). b Antibodies against the main epitopes were elevated in an independent cohort of neuro-COVID patients compared to pre-pandemic HCs. Brackets indicate statistically significant difference to pre-pandemic HC (q-values ≤ 0.05 from Mann-Whitney U-test with Benjamini-Hochberg correction). The y-axis displays signal intensity on the pseudo-log10 scale.
essential for viral entry and the target of antibodies neutralizing several Coronaviruses. In contrast, there were no indications of molecular mimicry of the highly prevalent anti-SNURF IgG. This study shows that the autoantibody repertoire emerging with COVID-19 is complex and provides a powerful rationale for continued investigation of the described new-onset autoantibodies together with continued investigation of new-onset autoantibody repertoires in other infectious diseases.
Concluding remarks

In this thesis, I have aimed to describe the research on antibody repertoires that I have performed during my doctoral studies and to put it in perspective of other efforts to analyze the antibody repertoire. During these four years, my research has been based on composing and applying the antigen arrays produced within the group to medically and biologically interesting cohorts of individuals and exploring their antibody repertoires. The foundations of this work were laid by the development and deployment of protein arrays within the Human Protein Atlas project and would not have been possible without the efforts of many dedicated students, engineers, and researchers before and alongside me. I am incredibly grateful to have received the opportunity to build upon this work.

The arrays used here together constitute a unique tool for the analysis of antibody repertoires. Combining large-scale autoantigen discovery on planar arrays and high-throughput screening on bead arrays enables broad study of the antigen diversity of the autoantibody repertoire. This is, however, only one aspect of the repertoire, and an increased diversity of antigen representations is needed for a wider view. Other antigen representations of the human proteome such as full-length proteins or shorter peptides, on arrays or display systems, enable exploration of a larger spectrum of antibody targets. The range of biologically and medically relevant antigens is not limited to proteins but can be further extended to glycans, lipids, and oligonucleotides. Reaching outside the autoantibody repertoire, antigen diversity can be assessed using collections of components of allergens or pathogens. In the present investigation, I have only touched upon this aspect of the antibody repertoire in the studies on SARS-CoV-2 serology. There are currently efforts focused on expanding our capability to analyze the pathogen-directed antibody repertoire within the scope of pandemic preparedness.
Antigen diversity is only one aspect of the antibody repertoire, and other analytical modalities are needed to complete the picture. Emerging mass spectrometry and B cell receptor sequencing-based technologies can be used to investigate the clonality of antibody repertoires. Applying them to repertoires enriched on antigens of interest is positioned to yield biological insights towards understanding the nature of the antibody response to pathogens and potential development of autoantibodies. With increased analytical depth, using these techniques in an antigen-agnostic setting may reveal repertoire-wide temporal dynamics of antibody repertoires between B cell populations and antibody isotypes.

With these recently described methods, the analysis of antibody repertoires seems poised to make substantial advancements in the coming years. In my opinion, the greatest potential lies within the analysis of a carefully selected set of longitudinally sampled individuals in health and across events of interest such as vaccination or infection. Multiple analytical modalities could be combined to render a highly detailed picture. Integration of the resulting data will be a challenging task and considerable efforts should be directed towards adapting and developing methods for analysis of the heterogeneous antibody repertoires. These advances may not only result in increased knowledge on the biology and medical implications of antibody repertoires but may also be used to mine the repertoire for potentially therapeutic antibodies.

Even with successful integration of the many facets of the antibody repertoire, future work will need to continue taking repertoire heterogeneity into account. Heterogeneity has been found across methods and settings and seems to be a fundamental property of the antibody repertoire, which also aligns with the combinatorial and stochastic processes generating clonal repertoire diversity. However, this property makes identification of autoantibodies with clinical relevance challenging. Tailored computational approaches, perhaps adapted from neighboring fields with sparse data, need to be developed to address this challenge and enable further exploration of autoantibody repertoires across diseases.

In conclusion, the analysis of antibody repertoires stands on the verge of a new era of integrative approaches. The work presented in this thesis constitutes a small part of these advances, which together may increase our understanding of the antibody repertoire in health and disease.
Acknowledgements

This thesis and the work it summarizes would not have been possible without the support of so many people, and I give my thanks to every one of you. I would like to acknowledge all the people who are willing to contribute to science by participating in studies and donating samples. Your contributions are essential to our research. I would also like to acknowledge Vetenskapsrådet, Märta Lundqvists stiftelse, and Olle Engkvists stiftelse, for funding this work.


Many people have guided me and provided new and thought-provoking perspectives during my doctoral studies. Two have been particularly important. To Jochen, thank you for asking difficult, insightful, and constructive questions. I greatly appreciate our discussions, and I hope they will continue. And thank you for the great travel company! Till Jean, tack för våra språkluncher under det senaste året. Våra samtal har gett mig värdefulla perspektiv på vetenskapen och livet, och jag uppskattar dem verkligen. Du är en stor källa till inspiration.
All members of the Nilsson lab and Schwenk lab, together known as PAPP, thank you for making up this fantastic team and for soaring through the highs and powering through the lows of research together.

Lovisa, jag är så glad att du stannade i gruppen trots att jag var din handledare! Tack för alla roliga stunder i och utanför labbet och för att du peppat mig hela vägen in i kaklet. Jennie, tack för att du tog dig an mig i labbet under exjobbet (och långt senare) med otroligt tålamod, pepp, och humor. Sára, thanks for all the good times at and outside of work. It has been great having you as an office neighbor and spontaneously discussing plots and data. Soon it is your turn, you have got this! Sofia, tack för allt bollande av kod, data, och paket. Din glädje och optimism lyfter hela gruppen. Ceke, tack för att du introducerade mig till dataanalysens (ibland) underbara värld. Dina stora system för hantering av alla våra data är helt magiska. Elisa, thank you for our discussions on autoimmunity and autoantibodies as well as terminology! It has been great having another autoantibody person to exchange ideas with. Eni, tack för att du håller koll på allt i labbet. Ingenting skulle fungera utan dig. Shaghayegh, tack för all god persisk mat. Annika och Leo, tack för att ni gjorde Busan och Seoul ännu roligare och hjälpte mig att slappna av från arbetet med avhandlingen. Sarah, tack för alla skratt under långa dagar på labbet. Matilda och Maryam, tack för ni fortfarande visar mig var saker finns, trots att jag borde veta nu. Ronald, tack för alla diskussioner om plana arrayer och normalisering. Utan din arrayexpertis hade den här avhandlingen inte varit möjlig. Claudia, thanks for being a great head of facility. William, det är så kul att du sökte dig till PAPP efter kexjobbet. Chus, thank you for being so kind and caring.

During my time in PAPP many fantastic people have left the groups, including Mun-Gwan, Cecilia M, Cecilia ET, Ragna, Maja, David J, Julia, Tea, Maria, Sanna, Philippa, Jamil, Julius, Katerina, Pablo, Alfred, Axelina, and many more. Thank you for contributing to the fantastic atmosphere of PAPP.

Outside of PAPP, there are many people at SciLifeLab who have enriched my time as a doctoral student in various ways. Past and present people of Alfa 2, particularly my fellow doctoral students Christian, Emmie, Sanem, Casper, Nicolai, Max, David K, Andreas, and many others, thank you for making our floor such a great place to work. My fellow
doctoral students across Sweden at MedBioInfo, thank you for contributing to our fun and inspiring community at courses and annual meetings.

Jochen and Jacob, thank you for your valuable comments and advice at the half-time seminar. Paul, thank you for your insightful feedback during the formal advance review. It has greatly improved the thesis. Sophia, thank you for diskussionen om affinitet och kvantifiering.

During my doctoral studies, I have had the privilege of collaborating with many brilliant scientists from other universities. Oliver, thank you for a great collaboration in the SHIP autoantibody project in the early days of my doctoral studies. Thank you for answering my endless questions and guiding me through the process of writing my first paper. Charlotte T och Sebastian, tack för ett gott samarbete och ert stora engagemang. Det har varit roligt och värdefullt att bidra till COMMUNITY-projektet. Janet och Annica, tack för många bra diskussioner och ert stora engagemang. Sophie, det ska bli spännande att se om vi kan ta autoantikroppar och kynureniner vidare.

Jag har den stora turen att ha många människor omkring mig som har stöttat mig innan och under den här resan. Utan er hade jag aldrig klarat detta. Ett extra stort tack till alla vänner som har förgyllt min vardag med umgänge och skratt, och som finns kvar även när livet fört oss ifrån varandra ett tag. Ni vet vilka ni är!

References


Antibody-based assay discriminates Zika virus infection from other flaviviruses. *Proceedings of the National Academy of Sciences*, 114(31), 8384-8389. doi: 10.1073/pnas.1704984114


References


Lee, S., Ko, Y., & Kim, T. J. (2020). Homeostasis and regulation of autoreactive B cells. *Cellular & Molecular Immunology*, 17(6), 561-569. doi: 10.1038/s41423-020-0445-4


71

References


