DOCTORAL THESIS

# Epitope-level Antibody Profiling for the Development of Diagnostics and Therapeutics Using Next Generation Phage Display

Annika Rähni

TALLINNA TEHNIKAÜLIKOOL
TALLINN UNIVERSITY OF TECHNOLOGY
TALLINN 2026

TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
67/2026

# Epitope-level Antibody Profiling for the Development of Diagnostics and Therapeutics Using Next Generation Phage Display

ANNIKA RÄHNI

TAL TECH PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
School of Science
Department of Chemistry and Biotechnology
This dissertation was accepted for the defence of the Doctor of Philosophy in Gene Technology on 17/09/2026

**Supervisor**: Principal Researcher Kaia Palm, PhD
Protobios LLC
Associate Professor
Department of Chemistry and Biotechnology
Tallinn University of Technology
Tallinn, Estonia

**Co-supervisor**: Professor Tõnis Timmusk, PhD
Department of Chemistry and Biotechnology
Tallinn University of Technology
Tallinn, Estonia

**Opponents**: Professor Mario Milco D'Elios, PhD
Department of Molecular and Developmental Medicine
University of Siena
Siena, Italy

Associate Professor Kaur Alasoo, PhD
Faculty of Science and Technology
Institute of Computer Science
University of Tartu
Tartu, Estonia

**Defence of the thesis**: 09/10/2026, Tallinn

**Declaration:**
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for doctoral or equivalent academic degree.

Annika Rähni

---
signature


Copyright: Annika Rähni, 2026
ISSN 2585-6898 (paperback)
ISBN 978-9916-80-573-2 (paperback)
ISSN 2585-6901 (PDF)
ISBN 978-9916-80-570-1 (PDF)
https://doi.org/10.23658/taltech.67/2026
Printed by Koopia Niini & Rauam

Rähni, A. (2026). *Epitope-level Antibody Profiling for the Development of Diagnostics and Therapeutics Using Next Generation Phage Display* [TalTech Press]. https://doi.org/10.23658/taltech.67/2026

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
67/2026

# Antikehaepitoopide kaardistamine diagnostika ja teraapia arenduseks järgmise põlvkonna faagikuvamise meetodiga

ANNIKA RÄHNI

TAL TECH
KIRJASTUS

# Contents

## List of publications

This thesis is based on the following research publications:

I **Rähni, A.**, Jaago, M., Sadam, H., Pupina, N., Pihlak, A., Tuvikene, J., Annuk, M., Mägi, A., Timmusk, T., Ghaemmaghami, A. M., & Palm, K. (2022). Melanoma-specific antigen-associated antitumor antibody reactivity as an immune-related biomarker for targeted immunotherapies. *Communications Medicine*, 2, 48. https://doi.org/10.1038/s43856-022-00114-7

II Jaago, M.*, **Rähni, A.***, Pupina, N., Pihlak, A., Sadam, H., Tuvikene, J., Avarlaid, A., Planken, A., Planken, M., Haring, L., Vasar, E., Baćević, M., Lambert, F., Kalso, E., Pussinen, P., Tienari, P. J., Vaheri, A., Lindholm, D., Timmusk, T., … Palm, K. (2022). Differential patterns of cross-reactive antibody response against SARS-CoV-2 spike protein detected for chronically ill and healthy COVID-19 naïve individuals. Scientific Reports, 12(1), 16817. https://doi.org/10.1038/s41598-022-20849-6

III Pupina, N., Avarlaid, A., Sadam, H., Pihlak, A., Jaago, M., Tuvikene, J., **Rähni, A.**, Planken, A., Planken, M., Kalso, E., Tienari, P. J., Nieminen, J. K., Seppänen, M. R. J., Vaheri, A., Lindholm, D., Sinisalo, J., Pussinen, P., Timmusk, T., & Palm, K. (2022). Immune response to a conserved enteroviral epitope of the major capsid VP1 protein is associated with lower risk of cardiovascular disease. EBioMedicine, 76, 103835. https://doi.org/10.1016/j.ebiom.2022.103835

IV Sadam, H.*, Mustonen, L.*, **Rähni, A**.*, Nieminen, J. K., Toots, M., Uusväli, M., Pihlak, A., Tienari, P. J., Harno, H., Palm, K.**, & Kalso, E.** (2026). Comprehensive antigen profiling predicts post-surgical neuropathic pain in women treated for breast cancer. Scientific Reports, 16(1), 12511. https://doi.org/10.1038/s41598-026-41637-6

*, ** Equal contribution

## Author's contribution to the publications

Contribution to the publications in this thesis are:

I Author contributed to the conceptualisation of the study, performed data analysis of the main and supplementary results, interpreted and visualised the data, and contributed to the writing and revision of the manuscript. MVA, MVA with competing agent and ELISA experiments were performed by others.

II Author contributed to the conceptualisation of the study and experimental design, performed MVA for COVID-19 patient samples and validation experiments with SARS-CoV-2 spike subunits (dot-ELISA), performed data analysis of the main and supplementary results, interpreted and visualised the data, and contributed to the writing and revision of the manuscript. MVA for main cohort samples and other ELISA experiments were performed by others.

III Author performed data analysis and visualisation of a part of the supplementary results, interpreted and verified the data of the whole manuscript, and contributed to the revision of the manuscript. Data analysis of the main results, and MVA and ELISA experiments were performed by others.

IV Author contributed to the conceptualisation of the study and experimental design, performed data analysis of the main and supplementary results, interpreted and visualised the data, and contributed to the writing and revision of the manuscript. MVA and ELISA experiments were performed by others.

**Other associated publications:**

Sadam, H., Maruste, R., **Rähni, A.**, Toots, M., Piirsalu, M., Pihlas, L., Ausman, H., Sirp, A., Pihlak, A., Laasfeld, T., Rull, K., Salumets, A., Antson, A., Tillmann, V., Aleksejeva, A., Rööp, T., Štšepetova, J., Sepp, E., Mändar, R., & Palm, K. (2026). Maternal antibodies shape infant immune response development in an epitope-specific manner. eBioMedicine, 123, 106056. https://doi.org/10.1016/j.ebiom.2025.106056

Jaago, M., Pupina, N., **Rähni, A.**, Pihlak, A., Sadam, H., Vrana, N. E., Sinisalo, J., Pussinen, P., & Palm, K. (2022). Antibody response to oral biofilm is a biomarker for acute coronary syndrome in periodontal disease. Communications Biology, 5(1), 205. https://doi.org/10.1038/s42003-022-03122-4

Sadam, H., Pihlak, A., Jaago, M., Pupina, N., **Rähni, A.**, Toots, M., Vaheri, A., Nieminen, J. K., Siuko, M., Tienari, P. J., & Palm, K. (2021). Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients. EBioMedicine, 64, 103211. https://doi.org/10.1016/j.ebiom.2021.103211

# Introduction

The human immune system is dynamic, develops, matures and evolves throughout life. Infections and diseases shape the unique immune fingerprint that is the human antibody and immune cell repertoire. Immune memory allows for immune reactivity to pathogenic and environmental antigens years after initial encounter. Individual immune memory patterns are markedly influenced by the genetic heterogeneity of the genes in the HLA locus, hypervariability and somatic hypermutation processes taking place during immune response maturation into highly antigen-specific responses and by the experiences and encounters with pathogens. Furthermore, immune responses to previously encountered antigens can alter antibody responses to unrelated new pathogens though heterologous immunity and immune imprinting mechanisms. Specific antibody patterns can be associated with the development of chronic diseases and used as biomarkers to predict disease. Antibody signatures are individual-specific and can be harnessed to design individual-specific therapies or preventative measures. In recent years, machine learning approaches have become widely used to precisely map epitope targets and used as a tool for developing new diagnostics, vaccines and drug design. Since these approaches rely on reference databases and datasets of high-quality experimental data, methods that map B cell epitopes in a high throughput manner have great potential in enabling the discovery of therapeutically relevant antigen targets and advancing personalised diagnostic efforts.

In these studies, we used mimotope variation analysis, a high throughput phage display technology, to describe the IgG-mediated immune response across healthy and chronic disease cohorts, together with statistical data analysis and sequence similarity detection tools to link antibody responses to and predict potential antigens from both human and microbial proteomes. We demonstrate that this approach allows for the delineation of health and disease-specific immune responses and predicts B cell epitopes connected to several chronic diseases, including cancer and cardiovascular disease. Furthermore, MVA-derived epitope-specific antibody patterns show promise as biomarkers to predict immunotherapy response in cancer patients and convey potential disease risk in cardiovascular and neuropathic pain cohorts. These studies offer support for epitope-specific antibody profiling as a powerful tool for identifying predictive and diagnostic biomarkers across diverse clinical contexts.

## Abbreviations

| | |
|---|---|
| ADA | Anti-drug antibody |
| ADCC | Antibody-dependent cytotoxicity |
| APC | Antigen presenting cell |
| AUC | Area under the curve |
| BCE | B cell epitope |
| BCR | B cell receptor |
| BLAST | Basic local alignment search tool |
| C1q | Complement component 1q |
| CD4+ | Cluster of differentiation protein 4 expressing cell |
| CD8+ | Cluster of differentiation protein 8 expressing cell |
| CMV | Cytomegalovirus |
| CoD | Curse of dimensionality |
| COVID-19 | Coronavirus disease 2019 |
| CPSNP | Post-surgical neuropathic pain |
| CSI | Cosine similarity index |
| CTA | Cancer testis antigen |
| CTL | Cytotoxic lymphocyte |
| CTLA-4 | Cytotoxic T-lymphocyte-associated protein 4 |
| CVD | Cardiovascular disease |
| DC | Dendritic cell |
| EBV | Epstein-Barr virus |
| ELISA | Enzyme-linked immunosorbent assay |
| FDR | False discovery rate |
| Fc | Antibody effector domain |
| GRAB | Germline-encoded amino-acid-binding |
| HA | Hemagglutinin |
| HCoV | Human coronavirus |
| HLA | Human leukocyte antigen |
| HMM | Hidden Markov model |
| HSV-1 | Human herpesvirus 1 |
| HSV-2 | Human herpesvirus 2 |
| IC | Immune checkpoint |
| ICI | Immune checkpoint inhibitor |
| IEDB | Immune epitope database |
| Ig | Immunoglobulin |
| IMUNE | Identifying motifs using next-generation sequencing experiments |
| mAb | Monoclonal antibody |
| MAST | Motif alignment and search tool |
| MEME | Multiple expectation-maximisation for motif elicitation |

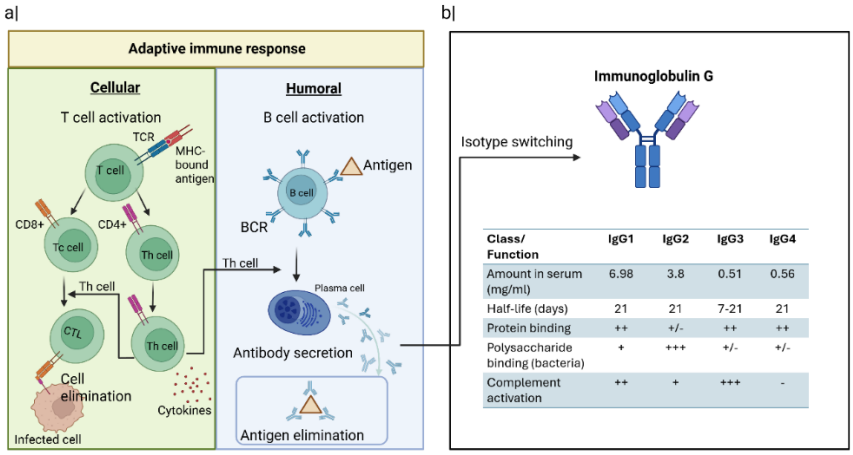
| | |
|---|---|
| MHC | Major histocompatibility complex |
| MI | Myocardial infarction |
| ML | Machine learning |
| mRNA | Messenger ribonucleic acid |
| MS | Mass spectroscopy |
| MVA | Mimotope variation analysis |
| NK | Natural killer cell |
| PD-1/PD-L1 | Programmed cell death/ ligand 1 |
| PDB | Protein data bank |
| PhipSeq | Phage immunoprecipitation sequencing |
| PI | Protrusion index |
| PIWAS | Protein-based immunome wide association studies |
| PTM | Post-translational modification |
| RF | Random forest |
| ROC | Receiver operating characteristic analysis |
| RSA | Relative surface accessibility |
| SARS-CoV-2 | Severe acute respiratory syncytial virus 2 |
| SERA | Serum epitope repertoire analysis |
| SHM | Somatic hypermutation |
| SPEXS2 | Sequence pattern exhaustive search 2 |
| Tc | Cytotoxic T cell |
| TCE | T cell epitope |
| TCR | T cell receptor |
| Th | Helper T cell |

# 1 Introduction to the immune system

At any given time, in the body of a healthy human adult, there can be as many as two trillion (2x $10^{12}$) immune cells (Sender et al., 2023), whose responsibility is to protect the organism from invading pathogens (e.g. viruses, bacteria, parasites), prevent tumour cell proliferation (immune surveillance) (Y. Li et al., 2021) and together with the endocrine- and nervous system keep optimal conditions for the continuation of the organism (homeostasis) (Zefferino et al., 2020). Immune cells are distributed throughout the body to provide local surveillance for potential invaders in the skin, mucosal epithelial barriers (gastro-intestinal and respiratory tract) and blood, however, nearly ~79% of all immune cells reside in the bone marrow and the lymphatic system (Sender et al., 2023). The immune system can conceptionally be split into two distinctive but interconnected systems – innate and adaptive (R. Wang et al., 2024). Synergistic mechanisms of the innate and adaptive immune responses, which ensure the protection of organism against foreign pathogens and help keep homeostasis, are strictly controlled. Potentially harmful foreign agents (antigens) are recognised based on their specific molecular patterns that activate cells of both the innate and adaptive immune system, which work in harmony to detect, target and neutralise the offending agents. Innate immune system includes cells that have the "innate" ability to detect specific evolutionarily conserved alarm patterns (belonging to tissue damage or pathogens) and then eliminate the threat or recruit help from the adaptive immune system (S. Carpenter & O'Neill, 2024). Cells of the adaptive immune system need activation signals from innate immunity, and once adaptive immune response is activated it generates a very specific, high affinity response that gets stored as memory to be recalled at a later date upon encountering the same pathogen (R. Wang et al., 2024).

## 1.1 Collaboration of innate and adaptive response creates immune memory

Antigen presentation links innate immunity to adaptive immunity and enables the formation of immune memory. During the activation of an immune response to an extracellular pathogen or other foreign antigen, antigen presenting cells (APCs) of the innate immune system detect the antigen, process it and present it to the B and T cells of the adaptive immune system (Elsen & J, 2011). T cell receptors (TCRs) bind antigens presented by major histocompatibility complex (MHC) I and II molecules that can be found on cell surfaces, including APCs (van den Elsen et al., 2004). MHC I molecules are expressed by all nucleated cells (including T, B, dendritic (DC) and natural killer (NK) cells) and present intracellular antigens to CD8+ T cells, whereas MHC II molecules are limited to professional APCs (including DCs, macrophages and B cells) and mainly present extracellular antigens to CD4+ T cells (Elsen & J, 2011). Signals from APCs and co-stimulatory molecules (cytokines, chemokines integrins and metabolites) instruct T cell differentiation into cytotoxic (CTL) and helper T (Th) cells, which mediate antigen-specific cellular immune responses (Hwang et al., 2020). Furthermore, MHC I-antigen complexes bind CD8+ T cell receptors, which activates T cells and leads to the elimination of antigen-presenting target cells, whereas MHC II molecules bind and activate CD4+ T cells and support their differentiation into T helper cells (R. Wang et al., 2024). Together, these pathways coordinate the initiation and specialization of adaptive cellular immunity (**Figure 1a**).




| Class/ Function | IgG1 | IgG2 | IgG3 | IgG4 |
|---|---|---|---|---|
| Amount in serum (mg/ml) | 6.98 | 3.8 | 0.51 | 0.56 |
| Half-life (days) | 21 | 21 | 7-21 | 21 |
| Protein binding | ++ | +/- | ++ | ++ |
| Polysaccharide binding (bacteria) | + | +++ | +/- | +/- |
| Complement activation | ++ | + | +++ | - |

***Figure 1. Generation of adaptive immune response.*** *Both T and B cell responses are part of the adaptive immune response and generated to eliminate targeted pathogens, infected cells or neutralise antigens. Activated T cells can differentiate into cytotoxic CD8+ T cells (Tc) and helper CD4+ T cells (Th). Th cells stimulate Tc cells to differentiate into cytotoxic lymphocytes (CTL) which can recognise infected cells by their surface markers and facilitate death signals. Th cells secrete cytokines which act as co-stimulatory signals for B cells to differentiate into plasma cells, which in turn secrete immunoglobulins (Igs) that can target and bind antigens. Immunoglobulins undergo isotype switching from primary IgM molecules to other isotypes, including IgG molecules. Immunoglobulin G molecule is one of the most abundant proteins in human blood. Out of four subclasses, IgG1 is the most abundant subtype. IgG2 is the second most abundant IgG subtype and has strong polysaccharide binding properties, which among other antigens, are found on bacterial membranes. IgG3 and IgG4 are the two least abundant subclasses in healthy adults, however IgG3 has an important effector function as a strong binder of complement protein C1q therefore acting as a complement activator. Table modified from (Vidarsson et al., 2014). TCR – T cell receptor, BCR – B cell receptor. Figure created with BioRender.com.*

B cells are a type of APC, which bind antigens through B cell receptors (BCRs), internalise and process the antigen into signals presented by MHC II molecules similarly to other APCs, and elicit a humoral immune response through the production of different antigen-specific immunoglobulin molecules (Akkaya et al., 2020). Immunoglobulin (Ig) molecules also known as antibodies, mediate the effects of the adaptive B cell responses and antigen elimination. B cells can produce five antibody isotypes, altogether 9 different subclasses of IgM, IgG (1-4), IgA (1-2), IgE and IgD molecules (Monroy-Iglesias et al., 2022). Different classes of Ig-s have different functionalities due to the structure of their Fc (effector) domains, which bind Fc receptors present on immune cells, and complement factors (Vidarsson et al., 2014). Fc domains facilitate effector functions of antibodies, including complement activation (by binding complement component C1q), antibody-dependent cell cytotoxicity (ADCC), antibody-dependent cell phagocytosis and complement-dependent cytotoxicity (L. L. Lu et al., 2018). IgM molecule is the first antibody isotype that is produced during primary immune response (J. Liu et al., 2019). Class switching recombination is the process wherein the constant heavy chain of an antigen activated IgM producing B cell switches to producing other forms of Ig molecules (IgG, IgA, IgE, IgD) (J.-C. Liu et al., 2024). IgG is the most prevalent Ig type in human blood (20% of total plasma proteins), with a diverse role of effector functions, including

activation of innate immunity (complement, neutrophils, macrophages, NK and DC cells), neutralisation of toxins/microorganisms/virulence factors, targeting pathogens for opsonisation and ADCC (L. L. Lu et al., 2018). IgG molecules can be further divided into four subclasses, which all have slightly different roles in the adaptive immune response (**Figure 1b**) (Vidarsson et al., 2014). In addition, antibodies can have neutralising or non-neutralising effects. Non-neutralising antibodies have a broad spectrum of different functions, whereas neutralising antibodies directly block pathogen entry into host cells or neutralise bioactive antigens (Chandler et al., 2023). These factors make IgG molecules one of the most versatile targets for vaccine- and immunotherapy design and development of clinical diagnostics (L. L. Lu et al., 2018; Monroy-Iglesias et al., 2022).

The generation of B cell immunity can be influenced by T cells. In T cell dependent immune response, B cells form conjugates with activated CD4+ Th cells through MHC II-TCR complexes. Co-stimulatory signals from Th cells (including cytokines) help B cells differentiate into antibody producing plasma and memory cells and facilitate IgM-type antibody isotype switching together with affinity maturation of Ig-s (**Figure 1**) (R. Wang et al., 2024). In T cell independent response antibody low-affinity IgM molecules are generated and class switching is limited and generation on of B cell memory is impaired (R. Wang et al., 2024). Conversely, studies have shown that lasting (half-life ~50 days) plasma cells can be generated in a T cell independent manner against antigens like bacterial polysaccharides that can be found on bacterial pathogens (Bortnick & Allman, 2013).

After each activation and effective neutralisation of antigen, the balance between immune cells and pro-inflammatory molecules is restored through the induced apoptosis of effector T lymphocytes. However, a subpopulation of memory T and B cells and antibody producing plasma cells are retained to unleash a pathogen-specific response in the future (Akkaya et al., 2020; Hwang et al., 2020). Previous antigenic exposure via either through infection or vaccination can influence effective protection and neutralisation of new potentially related antigens. Known as immune imprinting (or original antigenic sin) (Francis, 1960), the immune system has a preference to recall existing memory cells if they are generated against similar enough antigens, rather than mount a de novo immune response (Wrynla et al., 2025).

## 1.2 Generation of individual antigen and antibody repertoires

Antigen recognition and therefore elimination of antigen parent (molecule, cell, pathogen), relies on the antigen binding a matching TCR or BCR to elicit an immune response (R. Wang et al., 2024). Given the vast number of potentially pathogenic microorganisms and their extensive proteomes (Bartlett et al., 2022; Ye et al., 2022), the number of different TCR, BCR and MHC-binding combinations is broad, which the immune system has evolved to overcome. Specifically, the TCR and BCR genes decode diverse gene segments for the immune cell receptors and undergo somatic recombination (V(D)J recombination), wherein randomly chosen V, D and J gene segments are combined to generate a repertoire of theoretically $10^{12}$ -$10^{18}$ different receptors (Teraguchi et al., 2020). These recombination processes can potentially generate $10^{13}$ different antigen-recognising antibodies, which is more than all $10^{11}$ B cells present in the human body (X. Liu & Wu, 2018). In addition, the human leukocyte antigen (HLA) locus, which decodes MHC molecules for antigen presenting and undergoes combinatorial gene segment selection, is considered as the most polymorphic region in the human genome (Robinson et al., 2017) and the high variability between individuals

contributes to the generation of individual-specific immune response to the same antigens (Palmer & Norman, 2023). Although recombination mechanisms in the MHC, BCR and TCR genes allow for the recognition of diverse antigens (including self-proteins), potentially autoreactive B and T cells are eliminated in the bone marrow or thymus through negative selection processes or turned anergic (unreactive, unable to activate) before entering the peripheral lymph and blood circulation (Levine et al., 2000; Perera et al., 2013; Taams & Wauben, 2000). T cells themselves curb potentially harmful B cell memory, since B cells are dependent on mature Th cell costimulatory signals to mature into antibody producing plasmablasts and to undergo antibody isotype switching (Thomas et al., 2006). Furthermore, somatic hypermutation processes take place in B cells upon antigen-binding and cell maturation introducing single-nucleotide mutations that along with antibody isotype switching processes contribute to the generation of higher affinity antibody binding responses and diversify individual antibody repertoire (X. Liu & Wu, 2018; Seo & Choi, 2025).

Individual antibody repertoire is in part determined by the antigens an individual encounters during their lifetime. For example, global seroprevalence of *Helicobacter pylori*, a bacteria associated with the development of gastric cancer, is approximately 42-46% (Y.-C. Chen et al., 2024). However, in the Baltic states of Estonia, Latvia and Lithuania, the probability of contracting this bacterium is higher and seroprevalence in these states ranges from, 69% (Thjodleifsson et al., 2007), 63.1% (Jonaitis et al., 2025) to 79.2% (Leja et al., 2012), respectively. Clinical characteristics can also influence the immune repertoire. For example, it has been documented that the proportion and number of T and B cell clones decrease with age and the proportion of CD19+ B cells in older men is significantly lower compared to women (Wilcoxon $p = 0.0032$) (Márquez et al., 2020). In addition, a study conducted among UK Biobank participants showed that women have higher antibody titres against Epstein-Barr virus and human papillomavirus (Mentzer et al., 2022). Therefore, immune repertoire is unique to everyone, shaped by genetics, previous antigen exposure, gender and age.

## 1.3 Immune mechanisms against self-proteins and neoantigens

The immune system has a role in keeping homeostasis of the organism, including eliminating self-reactive cells (autoimmunity), which might otherwise contribute to the development of autoimmune disease if left unchecked. In addition to professional antigen-presenting cells, all human cells express MHC class I molecules on their surface, which present peptides derived from clearing up cell metabolism products and processed fragments of self-proteins (Mei et al., 2020). During T cell development, clones that bind self-antigens are eliminated or turned anergic to circumvent pathological autoreactivity (Kenison et al., 2024). However, autoimmune diseases can arise by mechanisms of molecular mimicry, wherein antibodies are generated, or immune cells are activated against foreign or pathogen antigens that share sequence homology or structural similarities with host proteins (Manjula et al., 1985). Today immune-based therapeutic treatment options often take advantage of the immune regulatory mechanisms designed to protect against autoimmunity. For example, therapies eliciting antigen-specific immune tolerance have been proposed as a potential therapeutic strategy to modulate harmful immune responses in autoimmune disease (Kenison et al., 2024). Conversely, the strategy of blocking immune checkpoint processes is routinely used as anti-tumour therapy in the clinical setting. Immune checkpoint (IC) molecules, including Programmed Cell Death Protein-1/ Programmed Death-Ligand 1 (PD-1/PD-L1) and Cytotoxic T-

lymphocyte-associated protein 4 (CTLA-4), have evolved to curb autoimmunity and tissue damage, but are also important in tumorigenesis, wherein cancer cells use immune checkpoint proteins and receptor signalling systems to mask from and evade antitumour immune detection (Arafat Hossain, 2024). Increased counts of PD-1 positive cells are a marker of metastatic cancer, increased disease progression and decreased patient survival (Muenst et al., 2013). IC inhibitors (ICIs) have shown promising results in treating otherwise treatment-resistant cancers by activating T cell and T cell-dependent immune responses (Arafat Hossain, 2024). During tumorigenesis, spontaneous somatic mutations can give rise to a plethora of mutated proteins, that vary between individuals and can therefore harbour individual-specific neoantigens (Jhunjhunwala et al., 2021). Public (i.e. shared) neoantigens are recurring mutations in known common oncogenes that can be used as universal targets of precision immunotherapies (Klebanoff & Wolchok, 2017; Zhang, Lu, et al., 2021), whereas individualised neoepitopes require development of personalised neoantigen-specific immunotherapy (Jhunjhunwala et al., 2021). One of the known hallmarks of cancer is cell dedifferentiation – a process, where cancer cells revert to a precursor cell like state, which can elicit the expression of otherwise developmentally- and tissue-restricted proteins termed cancer-testis antigens (CTAs) (Fratta et al., 2011). Therefore, CTAs present good biomarkers and targets for antitumor therapeutics. Neoantigens can also rise in processes where proteins are post-translationally modified (PTM) in an abnormal manner that the immune system detects as foreign. For example, citrullination is a PTM where arginine is deiminated to citrulline. Anti-citrulline antibodies are primarily associated with rheumatoid arthritis and its progression (Loh et al., 2024). Aberrant glycosylation patterns are associated with tumour progression, metastasis and immune evasion (Pinho & Reis, 2015). Disialoganglioside GD2, a tumour-associated glycolipid highly expressed in neuroblastoma, is the target of monoclonal antibody therapies like dinutuximab to induce tumour cell death with minimal impact on normal tissues (X.-Y. Liu et al., 2022). In addition, targeting cryptic epitopes, that are revealed upon abnormal folding of proteins (for example due to mutations in diseases like cancer), can represent an alternative therapeutic strategy with fewer side effects than some standard-of-care anti-cancer antibodies, which may show cross-reactivity with self-proteins (H. Xu et al., 2025).

# 2 Bioinformatical prediction and experimental validation of epitopes

## 2.1 Characterisation of epitopes

Epitopes i.e. antigenic determinants are specific molecular patterns on proteins, carbohydrates, lipids and nucleic acids that can bind receptors of T (TCRs) or B (BCRs) cells and activate the adaptive immune response (Commins et al., 2011; de Wit et al., 2025; Dowds et al., 2014; J. Li et al., 2025; Stern et al., 2024).

Epitope recognition and therefore the specific structure of epitopes differ substantially between T and B cells. T cell epitopes (TCEs) are linear protein-derived peptides, which are presented by APCs on either MHC I or MHC II complexes to the CD8+ T cytotoxic or CD4+ helper T cells, respectively. Depending on the MHC class, which are different in structure and peptide size allowance, the length of linear epitopes varies from 9-11 (MHC I) to 9-22 (MHC II) amino acid residues. Whereas TCRs can only bind continuous protein-derived epitopes presented on MHC complexes, BCRs can also bind conformational protein-derived epitopes, non-protein structures and other free soluble antigens. However, in case of T cell dependent B cell response, which is required for the generation of memory B cells and antibodies, the target epitopes remain identical. (Sanchez-Trincado et al., 2017)

B-cell epitopes (BCEs) can be described as linear (continuous) or conformational (discontinuous). Linear epitopes decode consecutive amino acid residues of the antigen whereas conformational epitopes can decode spatially separated amino acid residues or short consecutive amino acid sequences that may not follow the linear sequence of the antigen (Ras-Carmona et al., 2022; Sanchez-Trincado et al., 2017). The ratio of linear to conformational B cell epitopes has been under speculation for decades, with different sources reporting that >90% of B cell epitopes are discontinuous (Carroll et al., 2024), but the proportion has been difficult to estimate partly probably due to being influenced by the specific context, antigenic background (Barlow et al., 1986; EL-Manzalawy & Honavar, 2010; Van Regenmortel, 1996, 2009) and technological restrictions (De Leon et al., 2024). Systematic overview of B cell epitopes has shown that they primarily consist of ~15 residues (also reported 4-12 amino acids (Buus et al., 2012; Tang et al., 1988), 5-25 amino acids (Kringelum et al., 2013)), with an average of 5 amino acid linear stretches, containing predominantly hydrophobic amino acids in the centre and charged residues in the flanking areas (Kringelum et al., 2013).

The human immune system is hypothetically capable of generating an immune response to virtually unlimited variety of molecular structures (natural or synthetic) due to the hypervariability of the HLA locus and genetic recombination processes in the BCR and TCR genes, which can create a high number of combinations of epitope-binding receptors and antibodies (Lossius et al., 2016). During B cell maturation each cell that survives negative selection will be carefully evaluated and not only are cells with highest BCR-antigen affinity selected, but also the specificity is further improved by somatic hypermutation (SHM) processes. SHM ensures that the generated antibody response is epitope-specific and capable of binding non-covalently to their target epitopes with high affinity (Merkenschlager et al., 2025). Broad antibody eliciting epitopes can contain germline-encoded amino-acid-binding (GRAB) motifs, that ascertain the elicitation of a population of antibodies against the same epitope across individuals, populations and even species (i.e. public immunodominant epitopes) (E. L. Shrock et al., 2023). Different

epitopes of the same antigen can elicit the generation of neutralising and non-neutralising antibodies (Dugan et al., 2020). Vaccines against neutralising antibody eliciting epitopes are most effective (Guthmiller et al., 2022; Z. Hu et al., 2021). In addition, antibodies generated against immunodominant epitopes of related or structurally similar antigens may cross-react, either to convey continued protection or instead interfere with the generation of effective epitope-specific response through immune imprinting (Lima et al., 2022; Wrynla et al., 2025). Due to genetic diversity of human population, antigenic determinants from the same pathogen or antigen may elicit protective antibody responses in some individuals but not in others, partly due to HLA-restricted epitopes (Bian et al., 2024). In conclusion, the development process of new effective vaccines and therapeutics may need to consider the genetic heterogeneity, immune background and individual immune repertoires highlighting the importance of delineating epitope-specific immune responses at the individual level.

## 2.2 Epitope prediction approaches

Identification of specific BCEs is of great importance for clinical and biotechnological applications, including vaccine design, disease diagnostics and development of new antibody-based therapeutics (Crooke et al., 2020; J. Li et al., 2025; Natarajan et al., 2021; Sadam et al., 2021). There are numerous different methods and approaches to predict B cell epitopes. One of the earliest approaches used to predict BCEs were based on the physicochemical characteristics of amino acid residues. Over 188 physical properties of amino acid residues have been summarised into 10 Kidera factors that can be used as basis for describing the physicochemical properties of peptide sequences (Kidera et al., 1985). Properties including hydrophilicity, hydrophobicity (Sweet & Eisenberg, 1983), polarity (Radzicka & Wolfenden, 1988), flexibility and protrusion index (PI) (Thornton et al., 1986) of a given protein sequence give indication of surface accessibility and potential antigenicity of amino acid sequences and therefore their potential to bind antibodies or BCRs (Eisenberg et al., 1984; Hopp & Woods, 1981; Karplus & Schulz, 1985; Kyte & Doolittle, 1982). Relative surface accessibility (RSA) and secondary structure are features that can be computed with online tools like NetSurfP and NetSurfP-3.0 (Høie et al., 2022; Petersen et al., 2009). Methods to describe different physicochemical properties of residues are widely used as basis in BCE prediction pipelines to this day. A single parameter method developed by Kolaskar and Tongaonkar in 1990, which uses the surface distribution of hydrophobic amino acids leucine, cysteine and valine predicts antigenicity with 75% accuracy (Kolaskar & Tongaonkar, 1990). Already in the 1990s certain minimal amino acid motifs, essential for T cell and MHC binding, were described and the first tools using experimentally generated data like SYFPEITHI were created (Rammensee et al., 1999). Furthermore, for BCEs it has been shown that certain public epitopes (recognised across different individuals) contain some critical residues, including enriched lysine enrichment at epitope borders (E. L. Shrock et al., 2023). Abnormal glycosylation patterns can also be predictors of epitope sequences, wherein their presence or absence creates antigenic determinants recognisable by the immune system (Alves et al., 2021; Pon et al., 2020; Sikora et al., 2021). Conversely, added glycans on amino acid residues can influence the surface accessibility of BCEs and thus mask them from antibodies (Wintjens et al., 2020).

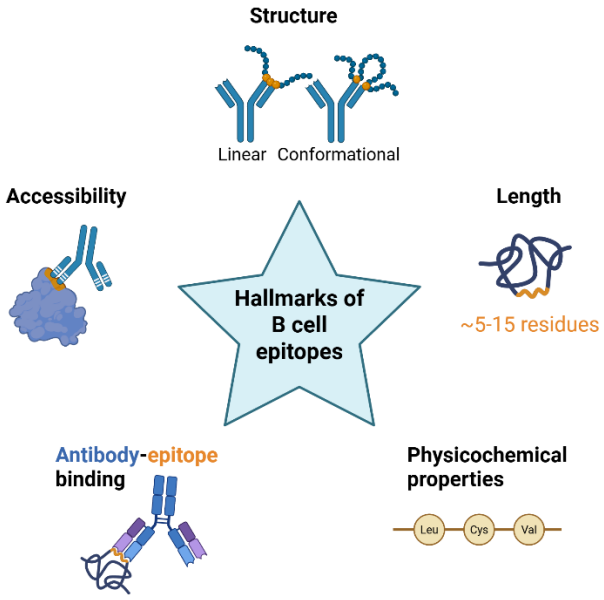



***Figure 2. Hallmarks of B cell epitopes****. B cell epitopes are antigenic determinants that bind B cell receptors and antibodies. BCEs exist in two structurally different forms – linear and conformational. The length of BCEs varies from 5 to 15 amino acid residues. BCEs have distinct physicochemical properties, including enrichment of Lysine at specific epitope borders. Antibodies bind antigen epitopes through their paratope. BCEs must be accessible for antibody or BCR binding. Leu – leucine, Cys – cysteine, Val – valine. Figure created with BioRender.com.*

Sequence-based epitope prediction methods were one of the first computational strategies that incorporated primary protein sequences with knowledge of the physicochemical properties of amino acids to determine antigenicity. These include methods that apply a sliding window approach, which considers the properties of each amino acid and its neighbouring residues in a potential antigen sequence to calculate averaged scores across peptide segments, indicating whether each residue has a high or low probability of being a part of an epitope (EL-Manzalawy & Honavar, 2010; Jespersen et al., 2017).

Sequence based prediction methods have been furthered along by the development of reference protein and epitope databases. As one of the reference resources in the epitope-mapping field, Immune Epitope Database (IEDB) has curated and hosts several epitope prediction tools since 2004 (Vita et al., 2024). IEDB enables linear BCE prediction from primary protein sequence by using 5 different amino acid propensity scales, including hydrophilicity prediction (Parker et al., 1986), beta turn prediction (Chou & Fasman, 1978), surface accessibility scale (Emini et al., 1985), flexibility scale (Karplus & Schulz, 1985) and antigenicity scale (Kolaskar & Tongaonkar, 1990). In addition to facilitating epitope prediction tools, IEDB has incorporated data from more than 25,000 publications into one database that houses over 1.6 million immune epitopes (both continuous and discontinuous sequences) from experimental B cell, T cell and MHC assays (Vita et al., 2024). IEDB is regularly used as a reference database to benchmark new algorithms that aim to predict new epitopes (F. Liu et al., 2024; Ras-Carmona et al., 2022). Other smaller datasets of experimentally determined B cell epitopes exist – for

example, Bcipep is a smaller database that includes links to other repositories including Protein Data Bank (PDB) (Burley et al., 2025) and Swiss-Prot/Uniprot (The UniProt Consortium, 2025) but is limited in size and focuses on B cell epitopes (Saha et al., 2005). Bioinformatical approaches to BCE predictions have also greatly benefitted from experimental protein structure databases like PDB and most recently from breakthroughs in solving protein structures with bioinformatical predictions, which enable the use of predicted structures of ca 200 million proteins and other biomolecules with AlphaFold 3 model (Abramson et al., 2024).

Nowadays, prevailing methods to predict B cell epitopes couple data from high throughput experiments together with bioinformatical approaches to describe new potential antigenic targets. Different statistical machine learning (ML) algorithms that incorporate know features of BCEs and protein sequence (and/or structure) databases are widespread (**Table 1**). Numerous different approaches and subtypes of ML algorithms are broadly used including hidden Markov models (HMMs), random forests (RF) and others (Sanchez-Trincado et al., 2017). For example, a linear BCE prediction algorithm called LBCE-BERT has been developed that uses IEDB for benchmarking and combines traditional amino acid physicochemical properties with sequence features and machine learning (F. Liu et al., 2024). Ras-Carmona et al. used a sequence similarity-based method that utilises Basic Local Alignment Search Tool (BLAST) to predict possible new linear B cell epitopes by determining sequence commonalities of known BCEs from IEDB and showed that their method (BepiBlast) was more accurate than many other commonly used ML-based BCE prediction methods (Ras-Carmona et al., 2022). In addition, ElliPro is a web-tool that does not require model training and incorporates BLAST and conformational PDB structures to predict epitopes based on the predicted protein conformation and amino acid composition (Ponomarenko et al., 2008; R Taylor et al., 1983; Thornton et al., 1986). IEDB hosts both BepiPred 1.0 (Larsen et al., 2006) and BepiPred 2.0 (Jespersen et al., 2017), which predict linear or sequential BCEs from primary protein sequences by applying knowledge from known properties of BCEs and incorporating them with ML algorithms to estimate the probability of peptide sequences being epitopes. BepiPred 1.0 relies on amino acid propensity scales coupled with a HMM, whereas BepiPred 2.0 uses a RF algorithm trained on crystal structures to predict amino acids, which have a high probability of being a part of discontinuous epitopes. Building on linear epitope prediction tools, researcher have created neural network-based approaches to try predicting both linear and conformational epitopes together. For example, CALIBER is based on more traditional sliding window and supervised learning like RF methods, where each residue in a sequence gets a probability score of being a part of an epitope (Israeli & Louzoun, 2024). DiscoTope 3.0 is an online tool capable of predicting conformational BCEs by being trained on data from PDB crystalline structures or using AlphaFold predictions of monomeric protein conformational structures (Høie et al., 2024). These are some examples how ML-algorithms are used today to predict BCEs, with many more reported elsewhere (Sanchez-Trincado et al., 2017).

***Table 1. Examples of B cell epitope (BCE) prediction methods and tools.*** *Both linear and discontinuous epitope prediction methods using machine learning (ML) or other approaches are provided. Discont. – discontinuous epitope, RSA – relative surface accessibility, PI – protrusion index, BLAST – Basic Local Alignment Search Tool, HMM – Hidden Markov Model, RF – Random Forest, ESM-2 – Evolutionary Scale Modelling, BERT - Bidirectional encoder representations from transformers, IEDB – Immune Epitope Database, PDB – Protein Data Bank.*

| **Method** | **BepiBlast** | **BepiPred 2.0** | **CALIBER** | **DiscoTope-3.0** | **ElliPro** | **LBCE-BERT** |
|---|---|---|---|---|---|---|
| Predicted BCEs | Linear | Linear, discont. | Linear, discont. | Discont. | Linear, discont. | Linear |
| **Approaches** | | | | | | |
| Physico-chemical properties | | Determi-ned by different methods, including NetSurfP | Determi-ned by different methods, including NetSurfP and Kidera factors | Antigen sequence and RSA scores in addition to other factors | Determi-ned by different methods, including PI | Amino acid composi-tion and anti-genicity scales |
| BLAST | Basis of method | | | | Used for linear sequences | |
| ML models | None | HMM, RF | RF, ESM-2 | XGBoost | None | BERT, XGBoost |
| **Training/validation data** | | | | | | |
| Swiss-Prot /UniProt | | | | | Training (user input) | |
| IEDB | Training | Validation | Training / validation | | | Bench-marking |
| Bcipep | Validation | | | | | Bench-marking |
| PDB | | Training / validation | Training / validation | Training / validation | Training (user input) | |
| Alpha Fold | | | | Training / validation | | |

As previously described, many BCE prediction algorithms have used publicly available datasets like IEDB (Vita et al., 2024), Swiss-Prot/UniProt (The UniProt Consortium, 2025) or PDB (Burley et al., 2025) to train and validate BCE predicting models (Israeli & Louzoun, 2024; Jespersen et al., 2017; F. Liu et al., 2024; Ras-Carmona et al., 2022). Since many ML algorithms are dependent on experimental data or public databases to generate training and validation datasets, some caveats need to be considered. Publicly available datasets can be limited and biased towards overrepresented and specific sequences garnering more scientific interests (including sequences from pathogens like SARS-CoV-2, herpesviruses, influenza or HIV), while rare antigens, genetic haplotypes or proteins with challenging crystallisation properties (like membrane proteins) (E. P. Carpenter et al.,

2008) are systematically under-represented (Villanueva-Flores et al., 2025). To circumvent underrepresentation bias, many BCE prediction algorithms have used AlphaFold-predicted structures (Abramson et al., 2024; Høie et al., 2024). In addition, data with negative outcomes (true non-epitopes) are scarcely reported, which can skew models towards over-predicting and over-estimating epitopes (Villanueva-Flores et al., 2025).

B cell epitope prediction can contribute to rapid vaccine developments in time sensitive situations. At the height of the SARS-CoV-2 pandemic in 2020, Crook et al. combined different webtools including available protein databases and epitope prediction tools (such as BepiPred and DiscoTope) to predict potential new vaccine-candidate epitopes (both linear and conformational), identifying 41 TCEs and 6 BCEs from all 10 proteins of the SARS-CoV-2 proteome (Crooke et al., 2020). In addition, Brewpitopes is a multidisciplinary approach to BCE prediction, where linear epitope prediction, structural epitope prediction, target protein subcellular location, accessibility predictions and different glycosylation prediction tools are combined into a detailed pipeline (Farriol-Duran et al., 2023). However, independent benchmarking of nine state-of-the-art conformational BCE predicting models, including Bepipred-2.0 and ElliPro, has demonstrated poor performance (ROC AUC < 0.6) and the authors have cautioned researchers to use in silico prediction tools with reservations (Cia et al., 2023). To conclude, ML algorithms have transformed the field of B cell epitope prediction and generated valuable tools for disease diagnostics, vaccine development and drug design, but further advancements are still needed.

## 2.3 Experimental approaches to describe B cell epitopes

Bioinformatical B cell epitope prediction approaches depend on high-quality experimental data to generate reference epitope datasets upon which prediction algorithms can be built, benchmarked and/or validated against (**Figure 3**). Experimental approaches to epitope mapping, prediction and validation are diverse and comprise of very different methodological approaches based on antibody-antigen reactions. Depending on whether the aim is to delineate linear and/or conformational epitopes and whether the expected output is low or high throughput, BCEs can be defined using a wide range of different experimental methods, including mass spectroscopy (MS), nuclear magnet resonance (NMR), X-ray crystallography, cryo-electron microscopy (cryo-EM) and protein and peptide display-based approaches (Abbott et al., 2014; Ferguson et al., 2025; Schwarz et al., 2021; G. J. Xu et al., 2015). Delineation of conformational BCEs is usually dependent on structural assays like NMR, X-ray crystallography or cryo-EM, which have historically been low throughput, because the structure of each specific antibody-antigen complex must be resolved to identify the epitope sequence (Abbott et al., 2014; García-Nafría & Tate, 2021; Merino & Raunser, 2017). However, advances have been made to resolve polyclonal antibody responses using combination of cryo-EM and ML approaches (Ferguson et al., 2025). In addition, synthetic recombinant or native protein-based methods, where proteins maintain their native conformation can be used to assess antibody response to structural epitopes. Mammalian (Thalén et al., 2024) and yeast (Najar et al., 2017) cell surface display methods can be used for high throughput epitope mapping of structural epitopes of human or pathogen-derived protein antigens, including membrane-bound proteins such as G protein-coupled receptors (GPCRs), which are important therapeutic development targets (Huang et al., 2019). Delineation of linear BCEs can be performed with peptide-based approaches, including MS analysis, peptide

arrays, phage- or bacteria-displayed peptide libraries (Pantazes et al., 2016; G. J. Xu et al., 2015). Peptide display-based approaches trade resolution for high throughput making them practical for large-scale epitope screening assays and identification of binding sites compared to MS or other structural approaches (Matsumoto et al., 2025). MS-based tandem methods enable the detection of peptide epitopes together with post-translational modification (PTM) patterns at large scale (Geffen et al., 2023).

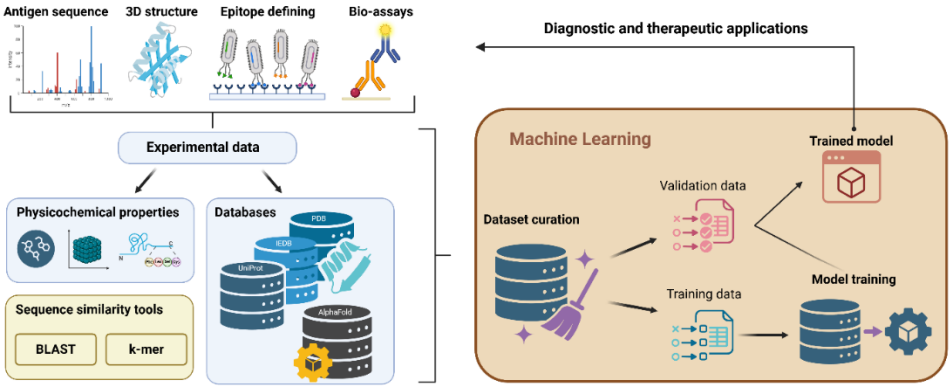



***Figure 3. Epitope prediction workflow.*** *Machine learning algorithms use data from high-throughput experiments or biological molecule databases to generate reference datasets (training and validation) for epitope prediction models. Biological experiments, including (but not limited to) mass-spectrometry, protein crystallisation techniques for determining protein conformational structures, phage and peptide display methods and data from individual bioassays are combined into open databases (UniProt, IEDB, PDB), that can be accessed by the scientific and public communities and data can be reused and repurposed. The AlphaFold database, housing high-quality structural protein models, is also used as reference dataset for the prediction of discontinuous epitopes. Physicochemical properties of each amino acid residue can be calculated using specific methods and incorporated when predicting new epitopes from peptide sequences. Pre-developed sequence similarity tools like BLAST can be also used as an input. When using public databases as a source of training data, data must first be cleaned and filtered to fit the needs of the expected model. Model performance can be evaluated against a subset of data that has been held out from model training. Ultimately, the predicted BCE epitopes serve as guidelines for experimental scientist who wish to use them for potential biomarker, therapeutic or biological tool development. 3D – three dimensional. Figure created with BioRender.com.*

Peptide microarrays, synthetic peptide libraries and peptide display assays with short overlapping peptide sequences have been widely used to discover and characterise antibody response to human and pathogen proteomes, including to describe antibody response at epitope-precision to SARS-CoV-2 (Schwarz et al., 2021), common human viruses (Bjornevik et al., 2022; Mina et al., 2019; G. J. Xu et al., 2015), autoantigens (Larman et al., 2011), major antigens of tick-borne pathogens (Tokarz et al., 2018), the human gut microbiota (Vogl et al., 2021), viruses that affect gut bacteria (Liebhoff et al., 2024), the entire human proteome (Zandian et al., 2017) and to describe epitope responses associated with the development of human disease (Jaago et al., 2022; Larman et al., 2013; Sadam et al., 2021). Peptide display libraries bind both neutralising and non-neutralising antibodies, therefore describing the full spectrum of antibody activity, although the functional consequences of antibody responses can be difficult to connect

to a specific biological outcome without additional experiments (Paull et al., 2019). In peptide-based BCE defining experiments, presented or displayed peptide libraries are used to mimic potential antigen sequences or structures. Therefore, library design is carefully considered since the number, length and diversity of peptides in the library determines the number of potential epitopes detected. There are advantages to using peptide display technology compared to peptide microarrays. Peptide arrays usually encode far fewer peptides compared to high throughput phage or bacterial display assays ($10^5$-$10^6$ vs $10^{10}$ peptides, respectively) (Larman et al., 2011; Legutki et al., 2014; Pantazes et al., 2016; Tokarz et al., 2018; Wu et al., 2019; Zandian et al., 2017).

Another consideration in library design is peptide length. When designing peptide libraries to study antibody response to a wide array of antigens, researchers have long used a tiling approach, where the primary sequences of a reference set of proteins (either of human or pathogen origin) is split into fixed length peptides (usually ranging from 12 to 60 amino acids residues) with a fixed overlap to achieve 2x coverage of the input proteome (Larman et al., 2011; Liebhoff et al., 2024; G. J. Xu et al., 2015). These short sequence peptide libraries can quickly amount to millions (even billions) of peptides (depending on the size and length of the input proteome), out of which only a fraction is reactive with antibodies from human samples (Liebhoff et al., 2024). The length of peptides in reference (displayed) libraries needs to be carefully considered since it will affect the number of potentially cross-reactive peptide-epitopes that are generated (Berglund et al., 2008). The length of displayed peptides should be optimised for high epitope resolution i.e. single epitope detection – peptides that are too long may contain multiple epitopes, which will potentially hinder delineation of single epitopes (Richer et al., 2015). A random 12-mer peptide library with the complexity of $10^{10}$ peptides has been shown to represent all possible 6 amino acid motifs with 99% confidence, generating sufficient diversity to discover immune response patterns to potentially any biological organism or protein antigen (Kamath et al., 2020). To reduce the cost of experimental approaches and expedite epitope defining numerous bioinformatical epitope prediction methods have been developed to date. To scale down the number of presented peptides in an assay, Liebhoff et al. have developed a ML algorithm called Dolphyn that computationally selects fewer peptides into the reference library from a classical tiling output (compressing the display library by 78%) to increase number of antibody binding (antigenic) peptides in the assay (Liebhoff et al., 2024).

The design of peptide display libraries can also vary due to the method by which the individual sequences of peptides are generated. Technologies based on phage immunoprecipitation sequencing (PhipSeq and MVA) use *E. coli* bacteriophages (like T7 or M13), which mimic antigen epitopes either through peptides derived from primary antigen protein sequences (Mohan et al., 2018; G. J. Xu et al., 2015) or by displaying randomised peptide sequences (Sadam et al., 2021; Wu et al., 2019). Antigen-derived protein or peptide arrays are biased towards detecting antibodies against structures and sequences defined by the pre-selected reference library, whereas randomised peptide libraries offer an alternative, potentially decoding antibody response to a hypothesis free range of environmental and commensal antigens (Jaago et al., 2022; Pantazes et al., 2016; Pupina et al., 2022; Sadam et al., 2021, 2021). Bacterial display technologies like SERA (Serum Epitope Repertoire Analysis) use randomised peptide sequences displayed on the coat proteins of *E. coli* to decode a broad spectrum of auto- and pathogen antigens (Pantazes et al., 2016). *E. coli* M13 phage-displayed randomised peptide libraries have been used to determine epitopes that elicit human anti-factor IX response in mice (Lozier

et al., 2005), to identify peptide sequences capable of binding *Bacillus anthracis* for more efficient detection (Sainath Rao et al., 2010), to develop new biomarkers for systemic lupus erythematosus (Wu et al., 2019) and to determine specific epitope sequences that bind antibodies following immunisation with specific antigens (J.-N. Li et al., 2019).

Peptide display libraries in essence mimic linear epitope sequences with varied lengths of defined amino acid residues. Antigen-derived peptide libraries usually capture linear BCEs if the displayed peptides are too short (7-12 residues) to fold into the same conformational structure present in the original antigen (D. Hu & Irving, 2023). Randomised peptide sequences allow for hypothesis free discovery of antibody-binding reactive peptides that describe both linear and conformational B cell epitopes, since randomised peptide display libraries contain sufficient peptide diversity and may also display mimics of structural epitopes by harbouring linear stretches of conformational epitopes (Kamath et al., 2020). For example, antibodies generated by immunising rabbits with the *Staphylococcus aureus* FnBPA-A protein bound 12-mer peptides from a randomised *E. coli* M13 phage display library that shared no more than three continuous amino acid residues with the primary sequence of FnBPA-A (J.-N. Li et al., 2019).

## 2.4 Computational strategies to delineate epitopes from mass-sequencing experiments

Nobel prize laureate in Physiology or Medicine, Dr. Sydney Brenner once coined the phrase “Low input, high throughput, no output”, to criticise the generation of vast amounts of data, without tools, methods or skills required to translate data into useful biological insight (Friedberg, 2008). This critique remains valid in the field of high throughput peptide array- and display-based sequencing methods, which generate millions of reads per each analysed sample (Jaago et al., 2022; Pantazes et al., 2016; G. J. Xu et al., 2015). These extensive datasets create a challenge for downstream data analysis that could be described as the curse of dimensionality (CoD) – a situation where the number of detected features exceed the number of samples analysed, which can reduce the value of individual data points, and increase computational burden and the risk of overfitting due to noise (Altman & Krzywinski, 2018). To mitigate CoD effects, while preserving biologically meaningful amino acid sequence patterns, high throughput peptide display methods benefit from peptide grouping or clustering algorithms, that identify similar peptides to reveal statistically enriched recognition patterns and combine the abundance of detected immune response into fewer representative features (Elbre, 2017; Pantazes et al., 2016).

Antibody-binding peptide sequences from high throughput experiments can be condensed into representative motif sequences to downscale the amount of information processed in downstream analyses. There are several different motif generation tools that look for repetitive and characteristic structures (patterns) in sequences, including MEME (Multiple Expectation-maximisation for Motif Elicitation) (Bailey & Elkan, 1994), IMUNE (Identifying Motifs Using Next-generation sequencing Experiments) (Pantazes et al., 2016) and SPEXS2 (Sequence Pattern EXhaustive Search 2) (Elbre, 2017). These algorithms can use experimentally or bioinformatically defined peptide sequences as input to describe representative epitopes as motif sequences that retain the most frequently detected and discriminative amino acids together with their relative location within each peptide. MEME has been shown to be less effective in defining disease-specific epitope sequences, more computationally expensive, 4 times slower and limited

in the number of input peptides compared to IMUNE (Pantazes et al., 2016). SPEXS2 is a statistical enrichment-based motif discovery tool, that is optimised for large datasets (including high-throughput phage display) (Elbre, 2017). In addition, custom motif generation tools utilising machine learning models like random forest have been integrated into full pipelines to define sample-specific epitopes (Ashkenazy et al., 2021).

To confer biological insight, each generated epitope sequence (peptide-mimic or representative motif) is further analysed to associate the abundance of antibody response to epitope mimics with the immune response to specific epitopes of autoantigens, pathogen proteins, allergens or other biologicals (Jaago et al., 2022; Pantazes et al., 2016; Sadam et al., 2021; G. J. Xu et al., 2015). The annotation strategy used to delineate potential true or genuine antigens as the source of immune response, can depend on 1) design of the peptide/antigen library used to identify epitope-mimics in the experiment (random sequence versus antigen-derived sequence libraries) and 2) type of the detected epitope sequences (continuous peptides or representative motif sequences). Of note, whereas experimentally derived epitope-mimicking peptide sequences are continuous, representative motif sequences can contain wildcard positions i.e. positions, where the exact amino acid residue is not conserved and can vary (Elbre, 2017; Pantazes et al., 2016). Annotating peptide epitopes in pre-defined or antigen-derived library experiments is relatively straightforward – each peptide in the reference library has been tiled from the primary amino acid sequence of a known reference protein and the copy number of each peptide in the input library is known (Larman et al., 2011; G. J. Xu et al., 2015). To annotate B cell epitopes derived from library experiments using randomised peptides with unknown origin, the defined epitope sequences can be queried against sequences found in curated databases, including human or pathogen proteins in Swiss-Prot/UniProt (The UniProt Consortium, 2025) or experimentally tested epitope sequences of pathogens, allergens or autoantigens in IEDB (Vita et al., 2024). To generate hypotheses about the origin of determined epitope sequences, continuous peptides and discontinuous motifs (with wildcard positions) can be queried against known databases with regular expression (regex) methods, which help detect sequence similarity in reference proteins, help identify candidates for epitope origins and can be used to screen large databases (Kolakowski et al., 1992). However, regex-based approaches cannot estimate whether the observed sequence similarity is biologically meaningful or merely due to random chance without additional experimental validation. Short predicted epitopes (e.g. 4–5 amino acids) may yield numerous matches from a database search (Bastas et al., 2008), most of which are likely random alignments or may indicate potential cross-reactivity between antigens (McClain, 2017). To reduce the number of random matches, sequence alignment and search tools based on probabilistic scoring matrices can be used. Freely available sequence similarity algorithms include Motif Alignment and Search Tool (MAST) (Bailey & Gribskov, 1998) and BLAST (Altschul et al., 1990) that can query representative motifs and peptide sequences against reference protein databases. MAST uses position specific scoring matrices created during motif prediction in MEME (containing the full amino acid probability characterisation of each motif) to search for matches within a reference database of proteins (Bastas et al., 2008). Regular BLAST uses statistical scoring functions based on substitution matrices such as Blocks Substitution Matrix (BLOSUM) and Point Accepted Mutation (PAM) matrix, which estimate the probability of amino acid substitutions during evolution. Of note, BLOSUM62 is the default scoring matrix in BLAST, while PAM30 is often preferred for aligning short or closely related peptide sequences

(Altschul, 1991; Altschul et al., 2005). While MAST estimates motif match scores, BLAST calculates sequence alignment scores, but both MAST and BLAST evaluate the chance of observing each match/alignment by calculating an E-value, which indicates whether the number of motif matches (in MAST) or sequence alignments (in BLAST) is greater or equal to the number of match observations or alignment score values that would occur by chance in a reference database of a given size. Specifically, a higher E-value indicates frequently occurring random matches, whereas a lower E-value indicates that a match is unlikely to occur randomly. (Altschul et al., 2005; Bailey & Elkan, 1994) A caveat for annotating B cell epitopes with either tool is that short identical sequences, showing 100% match between reference and query, generate high alignment score E-values, because the probability of random matches for shorter sequences is higher (Altschul et al., 1990; Boratyn et al., 2012). Since linear B cell epitopes are relatively short, on average spanning 4-12 amino acid residues (Buus et al., 2012), querying epitope sequences from high throughput experiments against extensive reference databases with BLAST may fail to achieve statistical significance. To circumvent and query large databases with high-throughput peptide assay-generated epitope datasets simultaneously several customised workflows have been reported. For example, Epitope-Mediated Antigen Prediction (E-MAP) determines cognate antigens in a hypothesis free approach by combining motif definition with MEME and motif query with MAST into one pipeline (Bastas et al., 2008). Protein-based Immunome Wide Association Study (PIWAS) couples randomised bacterial display library platform SERA (Kamath et al., 2020) with Kmer-Tiling of Protein Epitopes (K-TOPE) (Paull et al., 2019) to bioinformatically associate experiment-derived peptide-epitopes with potential antigens and calculate epitope-specific antibody response enrichment (Haynes, Kamath, Waitz, et al., 2021). Contemporary scientific methods frequently turn to machine learning models that incorporate epitope sequence and structural information to determine potential antigens (Villanueva-Flores et al., 2025).

# 3 Epitope predictions for precision medicine

Precise epitope mapping has potential implementation in numerous medical applications, including understanding autoimmune or infectious disease aetiology (Schwarz et al., 2021; Tariq et al., 2026), vaccine and immunotherapy development for chronic diseases or cancer (Parvizpour et al., 2020) and discovery of new biomarkers for precision medicine (Hörauf et al., 2024). Knowledge of epitope-specific immune response mechanisms help researchers understand the underlying processes of immune mediated diseases and can be incorporated into development of new vaccines, immunodiagnostics and -therapies.

## 3.1 Processes affecting immune memory

Past vaccinations and infections, even sickness experienced during childhood, can influence an individual's immune system throughout life. Specifically, primary immune response to one pathogen can induce and/or modify the elicitation of secondary immune responses to an unrelated pathogen, a phenomenon known as heterologous immunity (Nor Isamuddin et al., 2025). Researchers have shown that contracting a specific subtype of influenza in childhood "imprints" the pattern of the major hemagglutinin (HA) antigen to immune memory, which in turn can potentially confer lifelong protection from the avian-version of the influenza virus (Gostic et al., 2016). Immune imprinting (original antigenic sin) is a mechanism wherein the immune system prefers memory B cells raised to previously targeted antigens and recalls antibodies against already known epitopes instead of generating a new and specific response to the newly encountered antigen (Francis, 1960). This process can be used to design cross-reactive vaccines with broad coverage of protection against similar or even unrelated antigens. However, immune imprinting can hinder the development of new vaccines for pathogens with high mutation frequency that require yearly boosters (including influenza and SARS-CoV-2) particularly when the generation of non-neutralising cross-reactive antibodies from vaccinations of related viruses interferes with protection against seasonal influenza viruses (Khurana et al., 2013). SARS-CoV-2 infection boosts antibodies generated against other common seasonal human coronaviruses, but these antibodies are non-neutralising and do not confer protection from SARS-CoV-2 (Anderson et al., 2021; E. Shrock et al., 2020). In a study looking at the effects of pre-existing immunity on new antibody responses to influenza infection, Dugan et al. showed that both natural infection and vaccination induced antibody responses to neutralising epitopes of HA from past responses, but that natural infection also induced more responses to non-neutralising targets than vaccine, ultimately conferring less protection than vaccines (Dugan et al., 2020). mRNA vaccines towards new SARS-CoV-2 variants effectively elicit neutralising antibody responses, but these are generated by the recalling of the memory B cells generated during infection with the original Wuhan strain (Tortorici et al., 2024). In addition, B cell immunity, including antibody responses and immune memory, may be disrupted by naïve infections. Mina et al. showed that primary measles infection can erase pre-existing antibody repertoires, effectively eliminating protective immune memory against previously generated pathogens (Mina et al., 2019). In conclusion, immune memory shaped by previous infections, vaccinations and environmental exposures, has measurable effects on both adaptive and innate immune responses that facilitate long-term susceptibility to chronic inflammatory, autoimmune, neurodegenerative and metabolic diseases, and influence the individual response to

vaccines and treatment (including immunotherapies). Understanding how previous immune responses (immune history) might influence new responses can help predict biomarkers for disease and guide the design of new pharmaceuticals.

## 3.2 Disease-associated immune repertoire as potential biomarkers

Biomarkers are characteristics measured as indicators of normal biological processes (health), pathogenic processes (disease) or responses to exposure (infection, environmental factors) and intervention (including therapies) (FDA-NIH Biomarker Working Group, 2016). Biomarkers can be used to screen, diagnose, prognose or monitor disease and to predict therapy response (Aronson & Ferner, 2017). Diagnostic biomarkers are used to detect or confirm presence of a disease or its subtype, whereas predictive biomarkers identify individuals most likely to respond to a medical intervention or environmental agent in a positive way (improved survival, symptomatic benefit) or by developing an adverse effect (FDA-NIH Biomarker Working Group, 2016).

Immune repertoire is highly dynamic, responsive to change in homeostasis and therefore reflective of potential disease and healing processes. Many studies have reported associations between common pathogens and onset of chronic diseases. For example, neurotropic human herpes simplex virus 1 (HSV-1) infection has been associated with the onset of peripheral neuropathy (H. Lu et al., 2023), and immunisations against influenza (Jeong et al., 2025) and naïve influenza infection (Zhang, Gool, et al., 2021) have been associated with the onset of narcolepsy type 1. These studies indicate that upon infection or environmental exposure immune repertoire changes in a way that inadvertently or directly causes disease. Delineating specific immune responses like antibody response profiles and their mechanistic consequences could therefore provide potential diagnostic and/or predictive biomarkers. For example, specific antibody response patterns to common human pathogens, including human herpesviruses, *Helicobacter pylori* and bacteria found in oral microbiome have been associated with the development of atherosclerotic cardiovascular disease (Hansson, 2005; Jaago et al., 2022). Furthermore, antibodies generated against group A *Streptococcus* bacteria M-protein can cross-react with the cardiac myosin and in this way, streptococcal infection of the throat or skin can trigger myocarditis or kidney-disease (Adderson et al., 1998; Cunningham et al., 1997). In a study looking at whole-proteome autoantibody profiles of 250 multiple sclerosis patients, Zamecnik et al. described stable antibody signatures that were detected in 10% of patients both 5 years before and a year after disease onset (Zamecnik et al., 2024). The antibody-targeted sequences were found to share sequence similarity with both autoantigens and common human pathogens, and this signature has been proposed as a potential predictive biomarker for multiple sclerosis (Zamecnik et al., 2024). Furthermore, longitudinal studies have shown that antibody response to EBNA-1 protein of Epstein-Barr virus (EBV), a common herpesvirus infecting more than 90% of adult population (Kuri et al., 2020), is associated with increased risk in development of multiple sclerosis (Bjornevik et al., 2022; Sadam et al., 2021).

Fluctuations in immune repertoire that drive susceptibility and/or progression of chronic inflammatory diseases are candidates for diagnostic biomarkers. These measurable changes in immune repertoire can be due to past infections or environmental exposures and detected with single or multiplex antibody panels. For example, tick bites can induce so called alpha-gal syndrome, which is a food allergy to red meat mediated through the elicitation of IgE molecules targeting the carbohydrate

galactose-alpha-1,3-galactose (Commins et al., 2011). Common herpesvirus infections can also elicit immune responses that associate with autoimmunity-like phenotypes. Reactivation or infection with human herpesviruses can trigger autoimmune encephalitis by elicitation of autoantibodies that target autoantigens of the central nervous system (Maltsev & Fedirko, 2022). Furthermore, antibody signatures must be considered in the context of each disease, where the antibody patterns or antibodies targeting different epitopes might have distinct functional roles. For example, autoantibodies against type I interferon have been shown to serve as biomarkers for fatal outcomes in COVID-19 (Bastard et al., 2021), whereas the occurrence of anti-nuclear antibodies in systemic lupus erythematosus are associated with an ameliorated disease severity (Bradford et al., 2023). Understanding how immune repertoires are linked to disease development and progression can help predict diagnostic or predictive biomarkers and inform epitope-specific strategies for the design of new pharmaceuticals.

## 3.3 Epitope-specific strategies in the development of vaccines and therapeutics

Therapeutic modulation of maladaptive immune response-specific immune cascades can reduce risk and progression of disease (P. Kong et al., 2022). Therefore, determining antibody response at epitope precision can act as a guide in therapy design (Montero-Calle et al., 2024). From a therapeutic point of view, the immune response can both be induced and stimulated, also modulated or suppressed (Strzelec et al., 2023). Depending on the condition and disease of the patient, one individual may require immune activation while another modulation or inhibition. Therefore, understanding the specificity of immune response is essential for the development of effective therapeutic strategies. Antibodies are useful tools in clinical screening and diagnosis steps, since their collection is minimally invasive, and their use in biomarker and therapeutic discovery has been prevailing in recent years (Montero-Calle et al., 2024). Of note, antibodies comprised ~20% of United States of America Food and Drug Administration approved therapeutics according to 2018 data (Castelli et al., 2019).

Systems vaccinology claims that a person's baseline "immune setpoint", a combination of age, sex, genetic and epigenetic variation, prior infections or vaccinations, microbiome composition and metabolic state, can predict vaccine responsiveness (Raeven et al., 2019). Pre-existing antibody patterns have been demonstrated as a potential tool to predict individual immune response to future infections and vaccinations. In a study comprising more than 4000 people, including both immunosuppressed and healthy individuals, Song et al. showed that levels of pre-existing antibodies targeting common human pathogens can predict blunted immune response to SARS-CoV-2 vaccine and be used to predict antibody levels upon vaccination or infection (Song et al., 2026). Furthermore, monitoring and analysing the broad spectrum of antibody responses elicited by vaccinations can guide researchers in designing new more effective vaccines. Identifying the most antigenic and neutralising-antibody response eliciting regions of the pathogen antigens can be used to design vaccines with a lasting and protective immune response (Haynes, Kamath, Bozekowski, et al., 2021). Neutralising antibodies targeting epitopes in the receptor binding domain of SARS-CoV-2 Spike protein have been defined (Ju et al., 2020; Pinto et al., 2020; Yuan et al., 2020). In addition, computational approaches have designed vaccines containing multiple highly

immunogenic SARS-CoV-2 epitopes within a single multiplex-vaccine (Deepthi et al., 2025).

Mechanisms of innate immunity have been leveraged to create tumour-associated antigen and neoepitope-based dendritic cell vaccines, which help boost cancer patient's own immune response, create lasting immune memory and protect against relapses (Heras-Murillo et al., 2025; Q. Li et al., 2024). In cancer patients the presence of autoreactive antibodies is frequently reported and correlates with activated immune surveillance and better prognosis. For example, interferon alfa-2b treatment in melanoma patients elicited autoimmune responses in 25% of the study cohort, which were associated with fewer relapses and longer median survival compared to patients who did not develop autoantibodies (Gogas et al., 2006). Profiling antibody response at epitope precision may reveal cancer-specific neoepitopes that correlate with clinical features and treatment outcomes. Epitope-specific antibody boosting has been shown to correlate with therapeutic efficacy. In one study a breast cancer patient showed increased antibody response to a specific accessible epitope of a known cancer antigen called EPS8 in response to immune checkpoint blockade immunotherapy (Routh et al., 2023). Monitoring antibody responses can therefore be used to make informed decisions regarding the use of aggressive treatment regimen, allowing assessment of whether and individual patient is likely to get meaningful benefit. In addition, elucidation of public or private antigenic neoepitopes or cancer-associated epitopes can be useful for developing effective cancer vaccines that circumvent central tolerance mechanisms to elicit a strong immune response and prolong overall survival (Weber et al., 2024; Zaidi et al., 2025). For example, even a limited set of 15 HLA alleles and 50 shared cancer antigens can theoretically generate more than 28 000 neoepitope-HLA pairs (Gurung et al., 2024). Researchers have begun to describe public and private neoepitopes from mRNAs together with machine learning approaches to predict the MHC class restriction and epitope sequences showing promising tumour inhibition and cellular immune response in a murine colorectal cancer models (Ji et al., 2025). Random peptide libraries with epitope motif discovery (SERA platform together with IMUNE) and epitope mapping (PIWAS) have been used to delineate cancer-patient-specific antibody responses to tumour antigen NY-ESO-1 (W. S. Chen et al., 2020).

Immunotherapies have revolutionised the management of chronic diseases like cancers, autoimmune diseases and allergies. When designing immune system activating therapeutics, exact target epitopes must be thoroughly considered, since broad immunisation strategies might cause unintended or even potentially fatal immune related adverse events (Sharma et al., 2023). For example, in a murine model of breast cancer, monoclonal antibodies (mAbs) generated against different epitopes of the oncoprotein ErbB-2, had divergent effects dependent on their exact targets – one mAb inhibited tumour growth, whereas others stimulated active cancer proliferation (Yip et al., 2001). Convalescent plasma therapy derived from SARS-CoV-2 recovered patients has also been evaluated using Fc array binding assays to determine which antibody properties are linked to the effectiveness of the treatment and to potentially reverse engineer new therapeutics (Natarajan et al., 2021). In addition, the efficacy of immunotherapies may decline in individuals who develop anti-drug antibodies (ADA). Clinically significant ADAs are antibodies that affect the pharmacokinetics, pharmacodynamics, efficacy and safety of a given drug (Swanson, 2024). Neutralising ADAs interfere with the drug's therapeutic mode of action by binding the specific region engineered to bind to its therapeutic target (Shankar et al., 2014). ADAs can also elicit

adverse effects, like anaphylaxis or antibody-mediated immunodeficiency diseases (Gunn et al., 2016). Immune checkpoint inhibitor (ICI) therapies including anti-PD-1, anti-PD-L1, can elicit ADA responses – a systematic review covering 141 clinical trials reported that anti-PD-L1 treatment induced ADAs in 29.6% of liver cancer patients and that combination ICI therapies had higher incidence of ADAs than monotherapies (Galle et al., 2024). Therefore, monitoring antibody responses before, during and after treatment may help predict treatment response and possible side-effects.

In conclusion, antibody patterns in single-epitope resolution in the context of different chronic and infectious diseases could be candidates for potential biomarkers to either predict disease state, progression or therapy response and help inform the design of new vaccines and therapeutics (**Figure 4**).

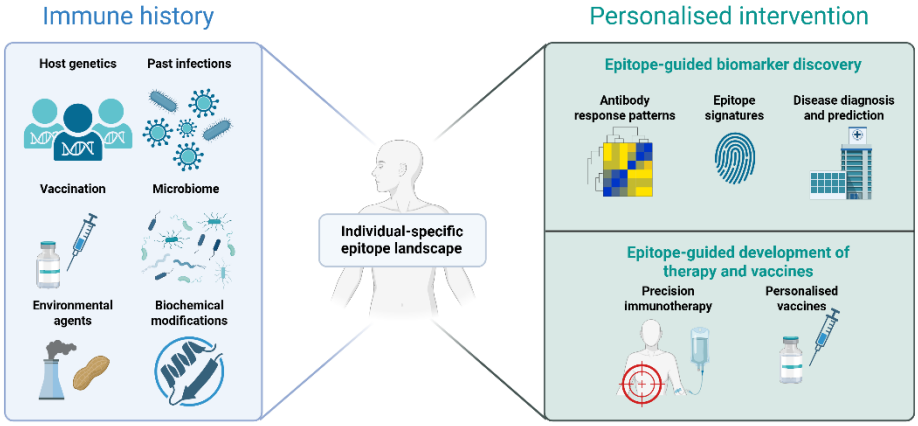



***Figure 4. Personalised approach to disease prevention, diagnosis and therapy.*** *An individual's immune history comprises of their genetics (HLA haplotypes, immune gene polymorphisms), past infections (molecular mimicry, immune imprinting), vaccination history, individual microbiome, contacted environmental agent (environmental and food allergens) and specific biochemical modifications that can occur on their proteins, including post-translational modifications like glycosylation (neoepitope generation, altered antigen processing). Immune history lays the ground for an individual-specific epitope landscape, that will serve as a basis for personalised medicine. Personalised approaches can include specific epitope patterns as biomarkers for disease diagnosis, prevention and prediction and epitope-specific responses can guide development of personalised therapies, including vaccines and immunotherapy. Figure created with BioRender.com.*

# 4 Aims of the study

Immune history can influence an individual's risk of developing disease and response to therapies. Insight into individual epitope-specific immune responses can therefore help researchers determine underlying processes of immune mediated disease and be useful in developing new diagnostics, vaccines and therapies.

Herein, we aimed to explore B cell epitope-specific antibody responses across diverse clinical cohorts, including in health, chronic disease, infections and immunotherapy, predict the original antigens associated with these epitope patterns, and evaluate the potential of predicted epitopes as candidate biomarkers for health and disease.

In detail, we aimed to:

1. Describe health and disease-associated antibody signatures and predict target epitopes using next-generation phage display in diverse clinical cohorts including malignant cancers, immunotherapy, cardiovascular disease and viral exposure.
2. Characterise epitope-specific antibody responses in cancer patients, chronically ill individuals and healthy controls to describe how immune history shapes cross-reactive or protective antibody signatures.
3. Evaluate the antibody signatures of disease- or health-associated epitopes as predictive tools for clinical outcomes, including cardiovascular disease risk, immunotherapy response and post-treatment complications after cancer surgery.

# 5 Materials and methods

The methodological approaches applied in this thesis are outlined below and described in greater detail in the corresponding publications:

- Mimotope Variation Analysis, MVA (Publications I, II, III and IV) *
- MVA with competing agent (Publication I) *
- Data pre-processing and quality control (Publications I, II, III and IV)
- Peptide grouping with SPEXS2 (Publications I, II, III and IV)
- Immunoprofile similarity testing (Publications I, II and IV)
- Group-specific feature selection with statistical testing (Publications I, II, III and IV)
- Sequence alignment analyses
    - With tool BLAST+ and reference database Swiss-Prot/UniProt (Publication II and III)
    - With regular expression-based tools and reference databases (Publications I, II, III and IV)
- Data visualisation (Publications I, II, III and IV)
- Commercial ELISA (Publications I, II, III and IV) *
- Phage display-based ELISA (Publication II and III) *
- Peptide display-based ELISA (Publication IV) *
- SARS-CoV-2 S native protein dot-ELISA (Publication II)

* Performed by others

# 6 Results and discussion

## 6.1 Characterisation of epitope-specific antibody response in individuals of diverse immune background

An individual's immune repertoire of both innate and adaptive immune response patterns provides an in depth look into past exposures to infections, immunisations, environmental agents and past or chronic disease (Zaslavsky et al., 2025). We employed experimental and bioinformatic B cell epitope prediction and validation methods to describe antibody-epitope response in 1) cancer patients receiving immunotherapy, in 2) individuals of wide clinical diagnoses to describe effects of heterologous immunity on new, previously un-encountered pathogens, chronic disease and therapeutics, and 3) described disease-specific immune patterns across multiple diagnoses for the development of new biomarkers (**Figure 5**).

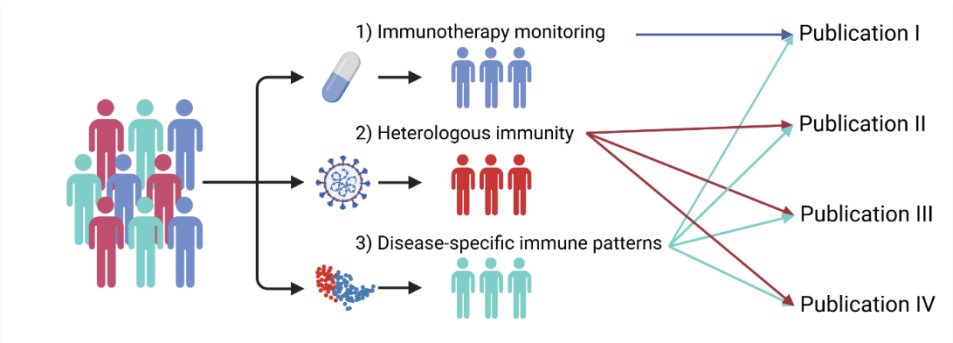



***Figure 5. B cell epitope determination strategies.*** *In publications underlying this thesis we employed mimotope variation analysis (MVA) to describe immune patterns in 1) cancer patients undergoing immunotherapy treatments, 2) the context of heterologous immunity pertaining to potential molecular mimicry of SARS-CoV-2 and associated coronaviruses together with other immune history in a diverse cohort of both chronically diseased and healthy individuals and 3) described disease-specific antibody responses as potential diagnostic and predictive biomarkers. Figure created with BioRender.com.*

In **publications I-IV** we analysed the antibody epitope patterns in different chronic, inflammatory and infectious diseases (**Table 2**). We described antibody signatures i.e. immuno-dominant epitope profiles in different clinical diagnosis groups, including cohorts of malignant cancer (non-small cell lung cancer, melanoma and breast cancer) (**Publication I - IV**), cardiovascular disease (coronary artery disease, acute coronary syndrome and myocardial infarction) (**Publication II, III**), neuropsychiatric disorder (first episode psychosis and schizophrenia) (**Publication II, III**), autoimmune disease (myasthenia gravis, multiple sclerosis, psoriasis) (**Publication II, III**), neurological disorder (narcolepsy) (**Publication III**), metabolic disease (type 2 diabetes) (**Publication II**), inflammatory disease (sarcoidosis) (**Publication III**), infectious disease (COVID-19) (**Publication II**), human papilloma virus (HPV)-associated cervical dysplasia (**Publication III**) and cancer immunotherapy (pembrolizumab and MelCancerVac®) (**Publication I**) (**Table 2**). In **publication IV** we described the antibody response patterns of breast cancer patients already included in **publications II** and **III** from the point of view of developing

post-surgical neuropathic pain (CPSNP) after intercostobrachial nerve resection. Healthy controls samples were included in discovery and validation cohorts of **publication I-III** and in the validation cohort of **publication IV**. Timepoint samples of immunotherapy receiving NSCLC patients, acute COVID-19 patients and breast cancer patients with CPSNP were included in **publications I**, **II** and **IV**. Altogether, we analysed 1068 unique samples from 938 individuals with disease- or health-associated immune backgrounds (**Publications I-IV**) (**Table 2**).

***Table 2**. **Characterisation of clinical study cohorts.** Disease sub-cohorts (1 to 13), including groups of individuals with clinical diagnoses (CASE) and along with ethnicity/sex/age matched controls (CTRL) or healthy control (HC) samples are described where applicable. The number of samples belonging to each group is described, asterisks (*) denote sub-cohorts, where timepoint samples were also included in analyses, but not described in the table. Mean age with standard deviation and number of men (M), women (F) or unknown (NA) is depicted. Reference column shows in which publication (I-IV) the cohort was included or if it has been published elsewhere. Of note, publication III included 2 ethnically different validation cohorts (III_2 and III_3). Description of disease sub-cohorts: 1 – Allergy (undefined), 2 – Autoimmune diseases, including myasthenia gravis (MG), multiple sclerosis (MS) and psoriasis. 3 – Cancer, patients with cancer diagnoses, including non-small cell lung cancer (NSCLC) and breast cancer (BC), 4 – Cancer/Immunotherapy, including NSCLC patients who received MelCancerVac® dendritic cell treatment (MelVac/MelVac-CTRL), melanoma patients who received Pembrolizumab treatment (PEM-Mel) and control samples for NSCLC group (CTRL-NSCLC), 5 – Cancer/Neuropathic pain, breast cancer (BC) patients who underwent surgical resection of the tumour and intercostobrachial nerve and who later developed post-surgical neuropathic pain (CPSNP) or did not develop neuropathic pain (non-CPSNP), 6 – Cardiovascular diseases, including stable coronary artery disease and acute coronary syndrome (CVD), and myocardial infarction (MI), 7 – Gynaecological disease, samples from patients with cervical intraepithelial pathologies and positive human papillomavirus (HPV) status, 8 – Infectious disease, patients with acute SARS-CoV-2 infection disease (COVID-19), 9 – Inflammatory disease, patients with sarcoidosis diagnosis, 10 – Metabolic disease, samples from type 2 diabetes (T2D) patients, who also had foot ulcers, 11 – Neurological disease, samples from narcolepsy (NRC) patients, 12 – Neuropsychiatric disorders, including first episode psychosis (FEP) and schizophrenia (SZ), 13 – Healthy control samples from North Estonian Medical Centre Blood Bank (CTRL-Mel and HC) and from individuals without any specified primary diagnosis (unspecified HC).*

| # | Group | Name | Country of origin | Samples (n) | Age (mean ± SD), years | Gender (M/F/NA) | Publica-tion(s) |
|---|---|---|---|---|---|---|---|
| 1 | CASE | Allergy | Estonia | 4 | 5.2 ± 2.5 | 1 M; 3 F | III |
| 2 | CASE | MG | Finland | 129 | 58.8 ± 15.5 | 57 M; 72 F | III |
| | CTRL | HC for MG | Finland | 50 | 41.8 ± 12.7 | 14 M; 36 F | III |
| | CASE | MS | Finland | 20 | 32.3 ± 7.8 | 4 M; 16 F | (Sadam et al., 2021), II and III |
| | CASE | Psoriasis | Estonia | 31 | 54.9 ± 15.4 | 7 M; 24 F | III |
| 3 | CASE | NSCLC | Denmark | 18* | 59.2 ± 7.9 | 7 M; 8 F; 3 NA | I |
| | CTRL | BC | Finland | 6* | 53.3±7.0 | 6 F | III_3, IV |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 4 | CASE | MelVac/ MelVac-CTRL | Denmark | 6* | 55.7 ± 8.4 | 4 M; 2 F | (Engell-Noerregaard et al., 2013), I |
| | CASE | PEM-Mel | Estonia | 5 | 67.6 ± 9.2 | 2 M; 3 F | I |
| | CTRL | CTRL-NSCLC | Belgium | 10 | 65.3 ± 8.4 | 5 M; 5 F | I |
| 5 | CASE | BC | Finland | 57 | 55.7±7.2 | 57 F | (Kaunisto et al., 2013), II and III_3 |
| | CASE | CPSNP | Finland | 27* | 53.9±6.1 | 27 F | (Kaunisto et al., 2013), IV |
| | CTRL | non-CPSNP | Finland | 30* | 57.4±7.8 | 30 F | |
| 6 | CASE | CVD | Finland | 64 | 62.8 ± 8.4 | 50 M; 14 F | (Jaago et al., 2022), II and III_3 |
| | CTRL | HC for CVD | Finland | 32 | 60.2±8.3 | 17 M; 15 F | |
| | CASE | MI | Estonia | 50 | 66.8 ± 12.6 | 29 M; 13 F; 8 NA | II and III_1 |
| | CASE | MI | Estonia | 40 | 66.8 ± 12.6 | 28 M; 12 F | III_2 |
| | CTRL | HC for MI | Estonia | 61 | 46.2 ± 15.5 | 25 M; 36 F | II |
| | CTRL | HC for MI | Estonia | 50 | 40.8 ± 9.9 | 20 M; 30 F | III_1 |
| | CTRL | HC for MI | Estonia | 49 | 40.4 ± 9.7 | 20 M; 29 F | III_2 |
| 7 | CASE | HPV | Estonia | 42 | 32.8 ± 5.8 | 42 F | III_2 |
| 8 | CASE | COVID-19 | Estonia | 6* | 57.2±7.8 | 4 M; 2 F | II |
| 9 | CASE | Sarcoidosis | Estonia | 47 | 41.1 ± 13 | 17 M; 30 F | III_2 |
| 10 | CASE | T2D | Belgium | 25 | 69.0 ± 8.7 | 21 M; 4 F | II |
| 11 | CASE | NRC | Finland | 16 | 17.6 ± 7.8 | 6 M; 10 F | (Sadam et al., 2018), III_3 |
| | CTRL | HC for NRC | Finland | 32 | 25.6 ± 8.3 | 18 M; 12 F; 2 NA | |
| 12 | CASE | FEP | Estonia | 30 | 25.6±4.9 | 16 M; 14 F | (Leppik et al., 2020; Parksepp et al., 2020), II |
| | CTRL | HC for FEP | Estonia | 30 | 24.0±6.1 | 15 M; 15 F | |
| | CASE | SZ | Finland | 30 | 41.9±18.4 | 12 M; 18 F | II and III_3 |
| | CTRL | HC for SZ | Finland | 30 | 42.1±18.2 | 12 M; 18 F | II and III_3 |
| 13 | CTRL | CTRL-Mel | Estonia | 80 | 38.5 ± 10.7 | 42 M; 38 F | I |
| | CTRL | HC | Estonia | 109 | 40.6±11.6 | 64 M; 45 F | II |
| | CTRL | HC | Estonia | 91 | 38.6 ± 11 | 63 M; 28 F | III_2 |
| | CTRL | Unspecified HC | Estonia | 24 | 49.5 ± 20.6 | 24 F | III_2 |

We used mimotope variation analysis (MVA), a high throughput M13 bacteriophage display technology that leverages a library of random 12-mer peptides with the complexity of $10^9$, SPEXS2 exhaustive pattern search algorithm (Elbre, 2017) and downstream statistical data analysis and annotation methods to describe the IgG

immune responses and associate them with candidate antigenic epitopes (**Figure 6**) (**Publications I-IV**). In high-yield sequencing data analysis, pre-analysis, quality assurances and normalisation steps are essential to minimise technical variations. During MVA data quality check erroneous sequencing reads or reads without the engineered peptide-coding insert are removed and quality sequences with confirmed peptide-coding inserts are bioinformatically translated into 12-mer peptides (**Figure 6a-b**). MVA generates, on average, 3 million reads that translate up to 0.6 million unique peptides per sample (**Publication I-IV**), which creates a curse of dimensionality (CoD) problem affecting downstream data analysis processes and possibly hindering the detection of biologically relevant relationships and patterns (Altman & Krzywinski, 2018). To mitigate CoD, while preserving biologically meaningful amino acid sequence information MVA incorporates an algorithm called SPEXS2, that reveals statistically enriched recognition patterns from peptide datasets (Elbre, 2017) (**Figure 6c**). A randomised bacterial peptide display platform called SERA facilitates a similar clustering algorithm called IMUNE, which is also used to identify disease biomarkers (Haynes, Kamath, Waitz, et al., 2021; Pantazes et al., 2016) and define pathogen-specific antibody epitopes (Haynes, Kamath, Bozekowski, et al., 2021). Pantazes et al. showed that IMUNE out-performs common motif discovery algorithm MEME (Bailey & Elkan, 1994) by enabling the simultaneous analysis of tens of millions of peptides in a fourth of the time (9 hours vs 40 hours) that it would take MEME (Pantazes et al., 2016). In MVA immunoprofiles are generated as a combination of individual- or group-specific motifs, where each motif gets a value that reflects the normalised count of all peptides containing the motif sequence in a given sample (**Figure 6d-e**).

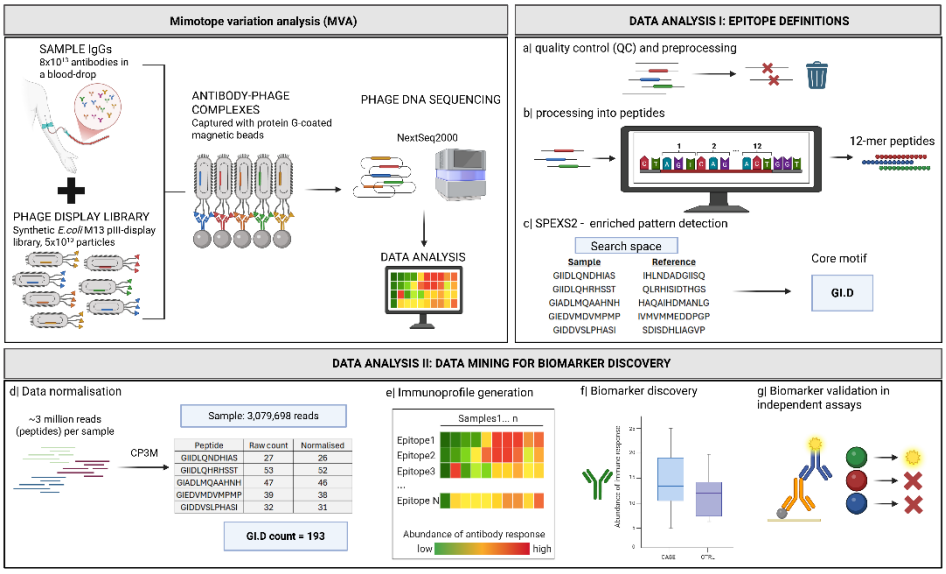



***Figure 6. Mimotope variation analysis and downstream data analysis steps. Mimotope variation analysis (MVA).*** *In MVA, human serum or plasma is incubated with E.coli M13 bacteriophage display library that displays up to $10^9$ unique random 12-mer peptides as part of the pIII coat protein. Each reaction receives $5x10^{10}$ phage particles to ensure phage diversity. After biopanning, wherein antibodies from the sample bind target epitope mimicking phage-presented peptides via Fab region, the phage-antibody complexes are captured via binding antibody Fc region with protein G-coated magnetic beads. The bound phage DNA is extracted, amplified with PCR, while each sample is also barcoded enabling multiplexed experiment setups, and sequenced with a short single read sequencing run.* ***DNA ANALYSIS I: EPITOPE DEFINITIONS.*** *a| Sequenced reads are quality checked and additionally mapped against the E.coli M13-library phage genome to remove reads coming from wild-type E.coli M13 or mutant wild-type E.coli M13 phages, that do not contain the engineered displayed peptide. b| Sequences with confirmed peptide-coding inserts are bioinformatically translated to 12-mer peptide sequences. c| Statistically enriched motif sequences are derived from sample peptides with SPEXS2 algorithm. SPEXS2 evaluates the occurrence of a pattern in samples against a reference peptide set and determines recurring motif sequences with length from 3 amino acids to 11 amino acids.* ***DATA ANALYSIS II: DATA MINING FOR BIOMARKER DISCOVERY.*** *d-e| Peptide counts are influenced by sequencing runs, different platforms and coverage and to analyse samples run on different experiments a normalisation step is required. On average MVA generates about 3 million reads per sample, therefore Counts Per 3 Million (CP3M) is used to normalise read count to library depth. Immunoprofiles can be generated from peptides and their normalised counts or based on representative epitope motifs, where the motif count is the sum of all peptides containing the sequence of that motif in each sample. This way, motif generation acts as one data dimensionality reduction method where datapoints from one sample can decrease from millions to thousands of features. f-g| Based on immunoprofile count data, group-specific (both disease and health) features can be delineated using different statistical methods, including additional dimensionality reduction, group comparisons, regression analyses and machine learning algorithms. Delineated biomarker candidates are validated using independent methods, including single peptide-based ELISA assays, Western blots, or dot blots. MVA and data analysis steps have also been detailed in previous publications (Elbre, 2017; Jaago et al., 2022; Sadam et al., 2018, 2021, 2026). Figure created with BioRender.com.*

## 6.2 Variability and dynamics of epitope-specific antibody response

The number of generated representative epitope motifs (with clustering algorithms like SPEXS2, IMUNE or MEME) depends on the number of peptides (and samples) used as input and can vary significantly. In our work we show that the number of epitopes defined as motif consensus sequences varies across different sub-cohorts (**Publication I-IV**) and depends on the specific approach and methods used to select peptides (**Publication II**). For example, in **publication II** we delineated epitope motif sequences for 8 clinically different subgroups with the combined total of ~23,000 epitope consensus sequences. The approach taken to determine consensus sequences was determined by the specific clinical question of each subgroup. In some groups, matched control cohorts were included to limit effect of covariable clinical features, whereas for others matched controls were not included and the parameters of SPEXS2 were optimised to maximise the number of epitope sequences generated for each group. Therefore, the final number of generated epitope sequences varied from a few hundred to thousands per study group (**Publication II**). The sheer number of generated epitope sequences reflects the differential composition of an individual's immunoprofile. Each person's immune repertoire is unique due to highly polymorphic human leukocyte antigen (HLA) genes, which introduce genetic heterogeneity that diversifies the capacity for epitope recognition across individuals (Khan et al., 2022). MVA analysis has demonstrated that different individuals with the same disease and immunotherapy background can have divergent antibody reactivities (in both composition and magnitude) against dominantly detected peptide antigens (**Publication I-IV, Figure 7**). Longitudinal samples collected from the same individuals show stable antibody response to the same dominant peptides and antibody profiles show significant similarities within individuals (Wilcoxon Rank Sum test, $p < 0.0001$) than compared to other individuals with the same or different disease background (**Figure 7**). We showed that dominant epitope-specific antibody response, in addition to being uniquely individual, is consistently stable in both short- (days to weeks) (**Publications I and II**) and long-term (years) intervals (**Publication I and IV**) irrespective of disease profile or therapeutic intervention in these specific studies (**Figure 7**). In **publication I** we show that the antibody response to dominant epitopes in an immunotherapy receiving NSCLC patient, who received 35 doses of a dendritic cell-derived autologous vaccine, was highly similar (cosine similarity index > 0.5) in all 4 time-series samples collected over a 2-year period and that 5 of the same immunotherapy-receiving NSCLC patients showed on average more similar antibody response patterns to their 2 week follow-up samples compared with randomly paired immunoprofiles (Wilcoxon Rank Sum test, $p < 0.0001$). In an independent cohort study of over 200 people, the antibody response patterns at 5-year follow-up showed higher average correlation between paired samples than unpaired samples (Pearson R values 0.78 versus 0.27) (Vogl et al., 2021). In another study involving 500 individuals, PhipSeq-derived antibody signatures unambiguously associated paired samples, that were collected 6 years apart, with ~96% accuracy (Zamecnik et al., 2024). Furthermore, in **publication II** we show, that antibody response to some SARS-CoV-2 Spike (S) resembling epitopes remain stable in COVID-19 patients in the acute and recovery phase of COVID-19 disease, although the timeframe to definitively determine stability of antibody response to these epitopes was short (i.e. weeks). However, sustained SARS-CoV-2-specific adaptive immune responses after COVID-19 infection, including anti-S neutralising IgGs with half-lives longer than 200 days have been documented elsewhere (K. W. Cohen et al., 2021). In **publication IV**, we showed that major antibody epitope

seroreactivities remained highly correlated between timepoint samples of breast cancer patients (average cosine similarity index 0.79 and Wilcoxon Rank Sum test, $p < 0.0001$) collected 4 to 9 years after tumour resection. These stable and sustained antibody response patterns act as an immune fingerprint, wherein the same individuals can be identified based on their serum antibody epitope repertoires even after years of exposure to different environmental and pathogenic agents (Dupic et al., 2021). In a case-study conducted over an 11-month time-period to study healthy immune repertoire fluctuations, B cell repertoires remained stable even after 2 episodes of naturally contracted rhinovirus infection (Mitsunaga & Snyder, 2020).

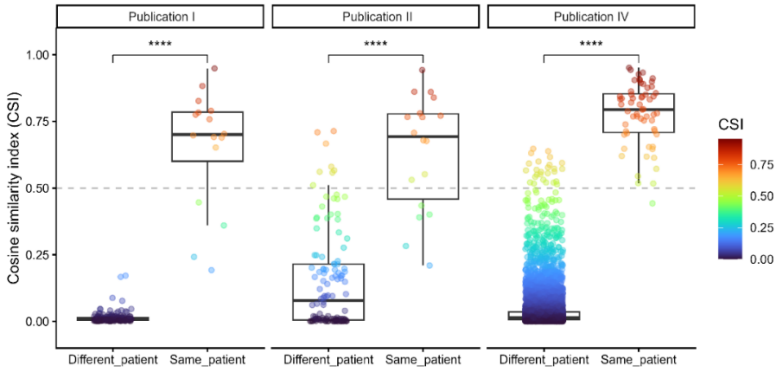



***Figure 7. Immunodominant antibody patterns remain stable across short-term (days to weeks) and long-term (years) timeframes.*** *Cosine similarity index (CSI) values were calculated based on the similarity of top 1000 most frequently detected peptides in each sample from patients in cohorts from publication I, II and IV who had timepoint samples available (n=170 samples). Publication I - 9 patients (7 patients had 2 timepoints, 1 patient has 3 timepoints and another 4 timepoint samples), Publication II - 6 patients (3 patients had 2 timepoints, 1 patient had 3 timepoints and 2 patients had 4 timepoints), Publication IV - 63 patients (2 timepoints). CSI values are calculated in pairs between samples that come from distinctive patients (Different_patient) or from the same individual (Same_patient). Mean CSI values are statistically significantly lower for samples that come from distinct patients compared to samples that come from the same individual: Publication I – 0.013 vs 0.652 (p-value $3.17*10^{-11}$), Publication II – 0.151 vs 0.636 (p-value $1.43*10^{-9}$), Publication IV – 0.037 vs 0.755 (p-value $5.12*10^{-39}$). Wilcoxon Rank Sum test, FDR adjusted p-values.*

In NSCLC patients receiving autologous dendritic cell immunotherapy, MelCancerVac® boosted antibody response to 15 specific epitopes of melanoma-associated antigens (**Publication I**). Although antibody response to majority of the 15 epitopes was patient-specific ("private"), antibody response to 2 epitopes was similarly boosted in all patients ("public") (**Publication I**). It has been documented that antibody response can be shaped by specific GRAB motifs in V gene segments, that predispose antibodies to recognise specific structural features, which creates "public" or shared immune responses between individuals (E. L. Shrock et al., 2023). Similarly, in SARS-CoV-2 infection a spectrum of public and private epitopes can be defined, wherein antibodies target common and patient-specific epitopes (E. Shrock et al., 2020). Analysis of most dominant peptide epitopes from longitudinal samples of the study cohort patients indicate that some shifting in dominant antibody responses occurred during different timepoints and that some patients shared more similar epitope-specific antibody response than others

(**Figure 7**). The composition of dominant epitopes in some patients shifted during disease progression, so that the longitudinal samples from the same individuals became dissimilar over time (CSI < 0.5, Same_patient, **Figure 7**), whereas some patients exhibited shared similar antibody reactivity to the same epitope mimicking peptides (CSI > 0.5, Different_patient, **Figure 7**). These analyses demonstrate that epitope-specific immune responses can be highly individual and heterogenous but also target same immunodominant sequences or structures. Delineation of these individual-specific patterns across different diseases, could help drive development of new pharmaceuticals and enable personalisation of treatment for each individual (Ng et al., 2025).

---

*Key insight:*

*An individual's antibody response pattern is stable across short (days to weeks) and long (years) timeframes and can be described as an immune fingerprint by which the same individuals can be identified even after years of repeated exposure to environmental and pathogenic antigens.*

---

## 6.3 Antibody response associated with autoantigen epitopes in immunotherapy

Immunotherapy is a treatment option taking advantage of the body's own immune system to combat disease, including cancer (Siddiqui et al., 2025). Cancer vaccines and immune checkpoint inhibitors (ICI) have shown promising results in the management of many cancers, however a large portion of patients show treatment resistance or immune-related adverse effects (Sharma et al., 2023) and therefore predicting treatment response could be beneficial to curb unwanted side-effects and treatment-resistance. Pre-existing autoreactive antibodies have been shown to be predictive of therapy response in cancer patients receiving ICI-s. For example, autoantibodies to interferon type I family proteins can predict ICI treatment responsive patients with odds ratios of 2.1 to 6.7 (Dai et al., 2025). In **publication I** we examined melanoma cell-derived antibody response upon autologous dendritic cell vaccine administration in non-small cell lung cancer (NSCLC) patients (MelVac) and upon pembrolizumab (anti-PD-1) treatments in melanoma patients (PEM-Mel) to predict common antigenic targets. Through sequence similarity searches using IEDB reported data and protein primary sequences from Swiss-Prot/UniProt we predicted 15 melanoma antigen-specific epitopes common to patients receiving immunotherapy (**Publication I**). In **publication I** we also show that top antibody response at peptide recognition level across immunotherapy-receiving melanoma patients was divergent. Upon vaccination, some individuals retained antibody response to specific dominant epitope motifs, in others antibody response to new epitopes prevailed (**Publication I**). This suggests that immunotherapy may have induced changes in dominant immune response, helped break immune tolerance and induce epitope spreading (Kellermann et al., 2024). Although, patients receiving different immunotherapy treatments had high reactivity to different epitopes of the same predicted target proteins, they still showed similarly increased response to melanoma-

associated antigens when compared to controls (**Publication I**). For example, both PEM-Mel and MelVac had significantly (Wilcoxon Rank Sum test, $p<0.001$) more antibodies detected against peptides mimicking known melanoma antigens CSPG4, PMEL and TGO1 when compared to controls (**Publication I**). This may be explainable by the diversity of HLA genotypes, which generates heterogenous immunopeptidomes between individuals and influences the recognition of tumour antigens (Chowell et al., 2019). In **publication I** we also validated that the enhanced epitope-specific immune response was specific to melanoma derived-antigens and not unspecific cross-reactive proteins. This was supported by the observation that competition with melanoma cell lysate, wherein melanoma-specific proteins bind serum antibodies and therefore reduce antibody availability, led to reduced epitope-specific antibody responses (Wilcoxon Rank Sum test, $p<0.05$), but did not affect immune response targeting unrelated common herpesvirus antigens, such as those of Epstein-Barr virus (EBV) and cytomegalovirus (CMV) (**Publication I**). Given that the 15 predicted epitopes bioinformatically mapped to widely expressed antigens (including MAGE family proteins), the humoral response towards these antigens could mark excessive immune attack and damage to self-tissues as concluded by others from studies of ICI-based therapies of melanoma and NSCLC (de Moel et al., 2019; Gowen et al., 2018; Toi et al., 2019). Recruitment of cellular adaptive immunity mechanisms were confirmed by data from MelCancerVac® trial, where INFγ analysis demonstrated T cell-specific response correlating with vaccine-specific immunity and sustained stable disease (Engell-Noerregaard et al., 2013). These results suggest that MVA-predicted epitopes of intracellular antigens could indicate immunotherapy or cancer associated tumour cell death and could be useful to monitor immunotherapy treatment response.

---

*Key insight:*

*Autologous cell therapy (MelCancerVac®) administered to non-small cell lung cancer patients elicits patient-specific antibody responses that are linked to melanoma-associated tumour antigens.*

---

## 6.4 Immune memory and heterologous immunity of pathogens associate with disease

In **publications II-IV** we studied antibody-mediated immune memory and heterologous immunity through specific pathogen-derived epitope-antibody relationships in cohorts of different chronic disorders, including cardiovascular disease and cancer.

In **publication II** we explored the potential role of heterologous B cell immunity in SARS-CoV-2 infection using MVA-derived immunoprofiles of SARS-CoV-2 naïve individuals. Previously it has been shown that there are cross-reactive memory T cells (Mateus et al., 2020), and cross-reactive, but non-neutralising SARS-CoV-2 antibodies in pre-pandemic COVID-19 naïve individuals (Selva et al., 2021). Our study cohort consisted of 538 individuals with a diverse clinical background, including chronic cardiovascular, neuropsychiatric and metabolic disease (**Table 2**). Therefore, immune features derived

from this cohort were also associated with different comorbidities (**Publication II**). MVA derives immune features from *E. coli* M13-bacteriophage random 12-mer peptide library, where the target peptide epitopes are not associated with specific parent molecules and therefore, requires sequence similarity analyses to determine true original and potentially cross-reactive antigens. To predict SARS-CoV-2 Spike (S) epitopes with a potential to cross-react with antibodies in naïve individuals we used a bioinformatical approach where we scanned the primary sequence of S for matching amino acid consensus sequences with features from our disease-associated immunoprofiles with the criteria of at least 4 matching amino acid residues and evaluated the cross-reactive potential of each amino acid residue with a rolling window method (**Publication II**). We also generated a random reference profile where the features were shuffled to qualitatively assess alignment within local protein regions to evaluate statistical significance of alignment using simulated random alignments (**Publication II**). In the end, we predicted 15 epitopes on the S protein that are potentially cross-reactive with the different components of immune memory (**Publication II**). Similar to our findings, groups utilising random bacterial display (SERA), have characterised SARS-CoV-2 whole proteome epitopes in both SARS-CoV-2 naïve and COVID-19 patients using two complimentary methods – IMUNE and PIWAS (Haynes, Kamath, Bozekowski, et al., 2021). Whereas our and IMUNE approach include reduction of 12-mer peptides to representative linear motif sequences and sequence similarity searches against a set of potential target proteins, PIWAS identifies candidate antigens by fragmenting 12-mer peptides into 5-6 amino acid k-mers and tiling them across protein sequences in a reference proteome (Haynes, Kamath, Waitz, et al., 2021). Regardless of approach, both MVA and SERA/IMUNE/PIWAS described SARS-CoV-2 S epitopes targeted by the naïve and COVID-19 patient populations, finding 3 partially overlapping IgG epitopes (S1.9, S2.2 and S2.5) (**Publication II**) (Haynes, Kamath, Bozekowski, et al., 2021). In addition, epitope S2.5 overlaps with sequences reported as highly reactive in individuals with severe, mild and asymptomatic SARS-CoV-2 infection (Y. Li et al., 2020; Mishra et al., 2021). Furthermore, out of the 15 MVA predicted S epitopes, 7 partially overlapped with epitopes previously described for COVID-19 naïve individuals and some overlapped with epitopes where antibody response has been described in patients with different stages of COVID-19 disease (**Publication II**). We hypothesised that these immune patterns could influence future SARS-CoV-2 immune responses though immune imprinting or heterologous immunity mechanisms, potentially interfering with the generation of an effective SARS-CoV-2 neutralising response or providing some protection against the virus. It has been documented that immune imprinting on earlier SARS-CoV-2 strains hinders the generation of potent neutralising antibody response to new vaccines in people who have received prior vaccinations or have been infected with earlier strains of the virus (Tortorici et al., 2024; Wrynla et al., 2025). Conversely, pre-existing antibody response, possibly due to previous common human coronavirus (HCoV) infections, did confer some protection against earlier strains of SARS-CoV-2 infection (Nguyen et al., 2024). We used BLAST+ (Basic Local Alignment Search Tool with command line application) (Camacho et al., 2009) and sequences from Swiss-Prot/UniProt database to show that this pre-existing and potentially SARS-CoV-2 Spike cross-reacting antibody response binds peptides that share sequence homology to common human viruses, including HCoVs, but also herpes-, papilloma- and respiratory viruses (**Publication II**). Sequence alignment with BLAST+ across human viral antigen reference database showed that MVA described peptides (i.e. epitope mimics) in sera samples of SARS-CoV-2 naïve

individuals frequently aligned to other HCoV proteins (**Publication II**). It has been documented that SARS-CoV-2 infection induces antibodies that are non-neutralising to SARS-CoV-2 and are derived from pre-existing memory B cells generated against other HcoVs (Anderson et al., 2021; E. Shrock et al., 2020). In addition, out of 15 MVA identified possibly cross-reactive epitopes on the Spike protein of SARS-CoV-2 virus, two (S1.8 and S2.2) showed sequence similarity to proteins of CMV and EBV (**Publication II**). Out of 199 of the tested cohort samples, 83.4% were CMV seropositive compared to 95.9% (out of 241 samples) that tested positive for anti-EBV antibodies (**Publication II**), suggesting that these epitopes may come from prior exposure to common herpesviruses. Notably, despite low sequence homology between CMV and SARS-CoV-2 it has been documented that CMV-specific T cells have been shown to cross-react with SARS-CoV-2 Spike protein (Pothast et al., 2022).

In **publication III** we explored the potential role of heterologous B cell immunity in the development of cardiovascular diseases (CVD) by using MVA-derived immunoprofiles of myocardial infarction (MI) patients and healthy controls. MI is a serious cardiovascular event that is usually preceded by development of atherosclerotic lesions in the arteries (Hansson, 2005). It has been shown that B cells can have conflicting roles in atherosclerotic processes. In detail, innate-like B1 cells produce natural IgM antibodies that protect against atherosclerosis, whereas conventional antigen driven B2 cells can generate adaptive antibodies that may promote progression of atherosclerosis (Que et al., 2018; Sage et al., 2019). In addition, pathogenic burden and elevated antibodies against bacterial and viral pathogens in MI have been documented (Jung & Lee, 2022; Reunanen et al., 2002). Our discovery cohort consisted of 50 MI patients and 50 healthy controls (**Table 2**). Therefore, immune features derived from this cohort were associated with both health and disease (**Publication III**). Antibody response to one of the epitopes from this set of immunoprofile features presented as significantly more enriched in healthy individuals when compared to MI patients (Mann-Whitney U test, p-value $1.5e^{-10}$) and the N-terminal epitope motif GI.D (glycine – isoleucine - any amino acid - aspartic acid) was predicted to be highly immunogenic (**Publication III**). To investigate the origin of this strong antibody response, we leveraged IEDB, Swiss-Prot/UniProt and BLAST+ and aligned GI.D-motif containing MVA-derived peptide epitopes to tiles of picornavirus antigens (**Publication III**). We determined that GI.D is a common motif found in viral protein 1 (VP1) sequences of picornaviruses, including enteroviruses like polio-, coxsackie- and vaccine-derived polioviruses (**Publication III**). VP1 protein and enterovirus RNA has been detected in heart tissue of CVD patients (Y. Li et al., 2000). Interestingly, increased serum antibody titres against coxsackie B1-5 viruses in patients following myocardial infarction has been documented before (Nikoskelainen et al., 1983). We concluded that the predicted cardiovascular health protective antibody response is likely derived from prior infections with enteroviruses or immunisations with poliovirus vaccines. In this context, immune memory and heterologous immunity may exert beneficial effects for health.

In **publication IV** we explored the potential role of heterologous B cell immunity in the development of post-surgical neuropathic pain (CPSNP) in breast cancer patients by using MVA-derived immunoprofiles and public databases. We determined 9344 epitope sequences either enriched in CPSNP or non-CPSNP-specific immunoprofiles (**Publication IV**). We referred to the IEDB database and showed that 1882 of the 9344 CPSNP and non-CPSNP-specific antibody responses target epitopes that map to specific antigens of common human pathogens including major herpesviruses like herpes simplex virus 1

(HSV-1) and 2 (HSV-2), EBV, CMV, enterovirus CVB3, human rhinovirus (HRV) and human papillomavirus 16 (HPV-16) (**Publication IV**). Serology analysis revealed that pathogen burden i.e. seropositivity of HSV-1, HSV-2, CMV and EBV, was proportionally higher in CPSNP group compared to non-CPSNP and that there were significantly more HSV2 seropositive individuals in the group who developed neuropathic pain (48% in CPSNP versus 20% in non-CPSNP, Fisher's exact test p value 0.027) (**Publication IV**). Infections can lead to chronic pain by causing direct nerve damage and by enhancing overall pain sensitivity and central sensitisation (S. P. Cohen et al., 2022). Previously it has been documented that HSV-1 can escape antiviral immunity and cause neuropathic pain through upregulating pro-inflammatory signals (E. Kong et al., 2024) and a rare unilateral pain syndrome with neuropathic pain features has been reported in association with recurrent HSV infections (Kallio-Laine et al., 2008). We showed that HSV-2 glycoprotein D (gD) epitope-mimicking peptide reactive antibodies were significantly more abundant in seropositive CPSNP patients (Wilcoxon Rank Sum test, FDR-adjusted p-value < 0.01) and experimental validation analyses confirmed that seropositivity to HSV-2 gD protein-derived peptides was significantly higher in CPSNP compared to controls (Wilcoxon Rank Sum test, FDR-adjusted p-value < 0.05) (**Publication IV**). We concluded that the bioinformatically defined epitope of HSV-2 glycoprotein D represents a genuine antibody target in HSV-2 seropositive samples (**Publication IV**). Case studies, describing HSV-2-associated pain, tingling, numbness and other abnormal sensations in the legs, feet and lower torso have been published (Ooi & Zawar, 2011; Shields & Alsorogi, 2019), however the mechanisms by which HSV-2 may contribute to the development of CPSNP remain unclear. Further research is required to clarify the role of human herpesviruses in the pathophysiology of neuropathic pain.

---

*Key insight:*

*SARS-CoV-2 naïve individuals harbour potentially cross-reactive antibody signatures derived from pre-existing immunity against common human viruses, that could interfere with the generation of a broad and neutralising SARS-CoV-2 immune response.*

---

## 6.5 Epitope-specific antibody response as biomarkers of disease and guidelines for therapy

Epitope-specific antibody response patterns have been proposed as biomarkers in human disease and numerous studies have been conducted to discover candidates. For example, SERA identified epitopes preferentially targeted by antibodies in celiac disease (Pantazes et al., 2016) and MVA has delineated antibody response patterns predicting optic neuritis progression to multiple sclerosis (Sadam et al., 2021). In **publication III**, we showed that antibody response to an epitope on enteroviral (EV) antigens may be used to predict risk of developing cardiovascular diseases (CVD) like myocardial infarction (MI), coronary artery disease and acute coronary syndrome. Antibody response to GI.D motif-encompassing peptides was 36% and 24% lower in men and women with a CVD diagnosis (**Publication III**). Development of atherosclerosis has been shown to be

promoted by enterovirus coxsackievirus B (CV-B) infections in animal models (Ilbäck et al., 1990) and the presence of EVs in atherosclerotic plaque has previously been documented (Musher et al., 2019). Of note, antibody response to this potentially protective GI.D epitope was dependent on vaccine type, with oral poliovirus-vaccine (OPV) showing increased antibody reactivity in both men and women (34% and 29 % respectively) compared to an intravenous (IPV) vaccine (**Publication III**). The differences between these two vaccines, including binding agents and antigen diversity, amount to differences in antibody response, that may serve as a guideline to the development of biomarkers, the design of vaccines and therapies targeting cardiovascular disease and MI.

Combining multiple epitope-specific antibody responses into a multivariate model may outperform the single epitope biomarkers by providing broader biological context (Rosado et al., 2021) and could increase test generalisability by accommodating heterogenous antibody responses across individuals (Jia et al., 2022). Here, we aimed to describe multiplex-epitope-specific immune patterns in different disease backgrounds and in immunotherapy. In **publication I** we demonstrated that a combination of antibody response to 3 epitopes with sequence similarity to melanoma-associated antigens can differentiate and could act as a biomarker of melanoma-associated immunity elicited by immunotherapies with 91% sensitivity and 100% specificity. Similarly, logistic regression model including autoantibody signatures against 6 tumour-associated antigens has shown good performance (AUC 0.841, sensitivity 76.6% and specificity 72.34%) in predicting diagnosis and differentiating different stages of gastric cancer from healthy controls (S. Wang et al., 2018). In addition, an autoantibody panel of 7 cancer-associated proteins showed good performance in detecting squamous cell carcinomas, where at least 76% of studied patients had autoantibody reactivity to at least 1 antigen in the panel (Chapman et al., 2008). Furthermore, in **publication II** we explored the association of predicted SARS-CoV-2 Spike antibody response patterns as biomarkers of ill health. Clinical characteristics like obesity, older age, diabetes, levels of C-reactive protein (CRP) and others each confer additional risk to developing severe COVID-19 upon SARS-CoV-2 infection (Laires et al., 2021; Rosa et al., 2025). Machine learning models combining the most impactful clinical characteristics have been made to transcend the description of individual pathologic conditions and to predict overall risk of severe COVID-19 with high accuracy (AUC 0.9) (Ozawa et al., 2025). We aimed to add description of individual antibody response to evaluate risk conferred by immune history. We showed that individuals with an acute or chronic disease background had significantly higher prevalence of predicted SARS-CoV-2 S pre-existing immunity than controls ($\chi 2$ test, $p < 0.00019$) (**Publication II**). A combined logistic regression model of antibody responses to 3 epitopes with sequence similarity to SARS-CoV-2 Spike protein differentiated patients with acute and chronic conditions (including cardiovascular disease, breast cancer, multiple sclerosis, type 2 diabetes and neuropsychiatric disorders) with 71% sensitivity and 68% specificity. Some of these epitopes were associated with specific clinical phenotypes, i.e. antibody response to peptides-mimicking Spike epitope “S18” was associated with hypertension (**Publication II**). We concluded that this model could be used to predict pre-existing risk of developing aggravated immunopathology upon exposure to SARS-CoV-2 and that it supplements other clinical biomarkers described for severe COVID-19.

In **publication IV** we explored the role of pre-existing antibody response in development of post-surgical neuropathic pain (CPSNP) in breast cancer patients.

Previously it has been shown that inflammatory markers, like high sensitivity CRP and orosomucoid protein are found in higher abundance in CPSNP patients relative to controls (Mustonen et al., 2019). High levels of pro-inflammatory cytokines in the cerebrospinal fluid of neuropathic pain patients has also been reported (Bäckryd et al., 2017). The development of neuropathic pain has also been associated with common human virus infections – neuropathic pain affects about 50% of people aged over 70 following herpesvirus infection (Watson, 2010) and 25% of people following COVID-19 (Herrero-Montes et al., 2022). We explored the pathogen-derived antibody load of breast cancer patients in both pre- and post-surgical timepoints and found that a combination test measuring antibody response to peptides mimicking specific epitopes of CVB-3, EBV, CMV, HPV-16 and HSV-2 predict the probability of CPSNP occurrence in breast cancer patients with 84.6% accuracy (**Publication IV**). The test was validated with an independent control cohort of Finnish women, confirming that the observed ROC performance was not cohort-specific, but generalisable to others as well (AUC 0.806) (**Publication IV**). Together, these findings establish epitope-resolved antibody profiling as a powerful framework for identifying predictive and diagnostic biomarkers spanning infection, chronic disease, and immune-mediated conditions.

---

*Key insights:*

*1. Pre-existing epitope-specific antibody response to poliovirus VP1 antigen, potentially derived from naïve infections or previous vaccinations, is associated with lower risk of developing cardiovascular disease, including myocardial infarction.*

*2. Epitope-specific antibody response generated against antigens of common human viral pathogens (including herpesviruses, rhinovirus and human papillomavirus) can predict the development of post-surgical neuropathic pain in breast cancer patients.*

---

# 7 Conclusion

In this thesis, next-generation phage display mimotope variation analysis was applied to explore the antibody response patterns elicited across health, chronic disease, infections, and immunotherapy. By integrating high-throughput experimental profiling with computational antigen annotation, we mapped disease-associated antibody epitopes, characterised differences between health and disease, evaluated their clinical predictive potential, and demonstrated the translational value of epitope-resolved immune profiling.

Based on analysis of 938 individuals and 1068 samples representing diverse disease backgrounds, the following conclusions were drawn:

1. MVA next-generation phage display profiling method enables prediction and systematic mapping of disease-associated epitope-specific antibody signatures across cancer, viral exposure, cardiovascular risk, and neuropathic pain cohorts, supporting large-scale immune biomarker discovery.
2. Epitope-resolved antibody repertoires differ between healthy and diseased populations but remain individually stable over time, reflecting the influence of immune history and heterologous immunity.
3. Epitope-specific antibody responses correlate with immunotherapy engagement, cardiovascular disease risk modulation, and susceptibility to post-surgical neuropathic pain, supporting their potential as predictive biomarkers of clinical outcome.
4. Integration of MVA method with computational antigen profiling enables scalable epitope annotation and identification of candidate antigens, supporting personalised diagnostics and therapies.

# References


Abbott, W. M., Damschroder, M. M., & Lowe, D. C. (2014). Current approaches to fine mapping of antigen–antibody interactions. *Immunology*, *142*(4), 526–535. https://doi.org/10.1111/imm.12284

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w

Adderson, E. E., Shikhman, A. R., Ward, K. E., & Cunningham, M. W. (1998). Molecular analysis of polyreactive monoclonal antibodies from rheumatic carditis: Human anti-N-acetylglucosamine/anti-myosin antibody V region genes. *Journal of Immunology (Baltimore, Md.: 1950)*, *161*(4), 2020–2031.

Akkaya, M., Kwak, K., & Pierce, S. K. (2020). B cell memory: Building two walls of protection against pathogens. *Nature Reviews Immunology*, *20*(4), 229–238. https://doi.org/10.1038/s41577-019-0244-2

Altman, N., & Krzywinski, M. (2018). The curse(s) of dimensionality. *Nature Methods*, *15*(6), 399–400. https://doi.org/10.1038/s41592-018-0019-x

Altschul, S. F. (1991). Amino acid substitution matrices from an information theoretic perspective. *Journal of Molecular Biology*, *219*(3), 555–565. https://doi.org/10.1016/0022-2836(91)90193-A

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. A., & Yu, Y.-K. (2005). Protein Database Searches Using Compositionally Adjusted Substitution Matrices. *The FEBS Journal*, *272*(20), 5101–5109. https://doi.org/10.1111/j.1742-4658.2005.04945.x

Alves, I., Santos-Pereira, B., Dalebout, H., Santos, S., Vicente, M. M., Campar, A., Thepaut, M., Fieschi, F., Strahl, S., Boyaval, F., Vizcaíno, R., Silva, R., Holst-Bernal, S., Vasconcelos, C., Santos, L., Wuhrer, M., Marinho, A., Heijs, B., & Pinho, S. S. (2021). Protein Mannosylation as a Diagnostic and Prognostic Biomarker of Lupus Nephritis: An Unusual Glycan Neoepitope in Systemic Lupus Erythematosus. *Arthritis & Rheumatology*, *73*(11), 2069–2077. https://doi.org/10.1002/art.41768

Anderson, E. M., Goodwin, E. C., Verma, A., Arevalo, C. P., Bolton, M. J., Weirick, M. E., Gouma, S., McAllister, C. M., Christensen, S. R., Weaver, J., Hicks, P., Manzoni, T. B., Oniyide, O., Ramage, H., Mathew, D., Baxter, A. E., Oldridge, D. A., Greenplate, A. R., Wu, J. E., … Hensley, S. E. (2021). Seasonal human coronavirus antibodies are boosted upon SARS-CoV-2 infection but not associated with protection. *Cell*, *184*(7), 1858-1864.e10. https://doi.org/10.1016/j.cell.2021.02.010

Arafat Hossain, Md. (2024). A comprehensive review of immune checkpoint inhibitors for cancer treatment. *International Immunopharmacology*, *143*, 113365. https://doi.org/10.1016/j.intimp.2024.113365

Aronson, J. K., & Ferner, R. E. (2017). Biomarkers—A General Review. *Current Protocols in Pharmacology*, *76*(1), 9.23.1-9.23.17. https://doi.org/10.1002/cpph.19

Ashkenazy, H., Avram, O., Ryvkin, A., Roitburd-Berman, A., Weiss-Ottolenghi, Y., Hada-Neeman, S., Gershoni, J. M., & Pupko, T. (2021). Motifier: An IgOme profiler based on peptide-motifs using machine learning. *Journal of Molecular Biology*, *433*(15), 167071. https://doi.org/10.1016/j.jmb.2021.167071

Bäckryd, E., Lind, A.-L., Thulin, M., Larsson, A., Gerdle, B., & Gordh, T. (2017). High levels of cerebrospinal fluid chemokines point to the presence of neuroinflammation in peripheral neuropathic pain: A cross-sectional study of 2 cohorts of patients compared with healthy controls. *PAIN*, *158*(12), 2487. https://doi.org/10.1097/j.pain.0000000000001061

Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, *2*, 28–36.

Bailey, T. L., & Gribskov, M. (1998). Combining evidence using p-values: Application to sequence homology searches. *Bioinformatics*, *14*(1), 48–54. https://doi.org/10.1093/bioinformatics/14.1.48

Barlow, D. J., Edwards, M. S., & Thornton, J. M. (1986). Continuous and discontinuous protein antigenic determinants. *Nature*, *322*(6081), 747–748. https://doi.org/10.1038/322747a0

Bartlett, A., Padfield, D., Lear, L., Bendall, R., & Vos, M. (2022). A comprehensive list of bacterial pathogens infecting humans. *Microbiology*, *168*(12), 001269. https://doi.org/10.1099/mic.0.001269

Bastard, P., Gervais, A., Le Voyer, T., Rosain, J., Philippot, Q., Manry, J., Michailidis, E., Hoffmann, H.-H., Eto, S., Garcia-Prat, M., Bizien, L., Parra-Martínez, A., Yang, R., Haljasmägi, L., Migaud, M., Särekannu, K., Maslovskaja, J., de Prost, N., Tandjaoui-Lambiotte, Y., … Casanova, J.-L. (2021). Autoantibodies neutralizing type I IFNs are present in ~4% of uninfected individuals over 70 years old and account for ~20% of COVID-19 deaths. *Science Immunology*, *6*(62), eabl4340. https://doi.org/10.1126/sciimmunol.abl4340

Bastas, G., Sompuram, S. R., Pierce, B., Vani, K., & Bogen, S. A. (2008). Bioinformatic Requirements for Protein Database Searching Using Predicted Epitopes from Disease-associated Antibodies*. *Molecular & Cellular Proteomics*, *7*(2), 247–256. https://doi.org/10.1074/mcp.M700107-MCP200

Berglund, L., Andrade, J., Odeberg, J., & Uhlén, M. (2008). The epitope space of the human proteome. *Protein Science*, *17*(4), 606–613. https://doi.org/10.1110/ps.073347208

Bian, S., Guo, X., Yang, X., Wei, Y., Yang, Z., Cheng, S., Yan, J., Chen, Y., Chen, G.-B., Du, X., Francis, S. S., Shu, Y., & Liu, S. (2024). Genetic determinants of IgG antibody response to COVID-19 vaccination. *The American Journal of Human Genetics*, *111*(1), 181–199. https://doi.org/10.1016/j.ajhg.2023.12.005

Bjornevik, K., Cortese, M., Healy, B. C., Kuhle, J., Mina, M. J., Leng, Y., Elledge, S. J., Niebuhr, D. W., Scher, A. I., Munger, K. L., & Ascherio, A. (2022). Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science*, *375*(6578), 296–301. https://doi.org/10.1126/science.abj8222

Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., & Madden, T. L. (2012). Domain enhanced lookup time accelerated BLAST. *Biology Direct*, *7*(1), 12. https://doi.org/10.1186/1745-6150-7-12

Bortnick, A., & Allman, D. (2013). What Is and What Should Always Have Been: Long-Lived Plasma Cells Induced by T Cell–Independent Antigens. *The Journal of Immunology*, *190*(12), 5913–5918. https://doi.org/10.4049/jimmunol.1300161

Bradford, H. F., Haljasmägi, L., Menon, M., McDonnell, T. C. R., Särekannu, K., Vanker, M., Peterson, P., Wincup, C., Abida, R., Gonzalez, R. F., Bondet, V., Duffy, D., Isenberg, D. A., Kisand, K., & Mauri, C. (2023). Inactive disease in patients with lupus is linked to autoantibodies to type I interferons that normalize blood IFNα and B cell subsets. *Cell Reports. Medicine*, *4*(1), 100894. https://doi.org/10.1016/j.xcrm.2022.100894

Burley, S. K., Bhatt, R., Bhikadiya, C., Bi, C., Biester, A., Biswas, P., Bittrich, S., Blaumann, S., Brown, R., Chao, H., Chithari, V. R., Craig, P. A., Crichlow, G. V., Duarte, J. M., Dutta, S., Feng, Z., Flatt, J. W., Ghosh, S., Goodsell, D. S., … Zardecki, C. (2025). Updated resources for exploring experimentally-determined PDB structures and Computed Structure Models at the RCSB Protein Data Bank. *Nucleic Acids Research*, *53*(D1), D564–D574. https://doi.org/10.1093/nar/gkae1091

Buus, S., Rockberg, J., Forsström, B., Nilsson, P., Uhlen, M., & Schafer-Nielsen, C. (2012). High-resolution Mapping of Linear Antibody Epitopes Using Ultrahigh-density Peptide Microarrays*. *Molecular & Cellular Proteomics*, *11*(12), 1790–1800. https://doi.org/10.1074/mcp.M112.020800

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Carpenter, E. P., Beis, K., Cameron, A. D., & Iwata, S. (2008). Overcoming the challenges of membrane protein crystallography. *Current Opinion in Structural Biology, Carbohydrates and Glycoconjugates / Biophysical Methods*, *18*(5), 581–586. https://doi.org/10.1016/j.sbi.2008.07.001

Carpenter, S., & O'Neill, L. A. J. (2024). From periphery to center stage: 50 years of advancements in innate immunity. *Cell*, *187*(9), 2030–2051. https://doi.org/10.1016/j.cell.2024.03.036

Carroll, M., Rosenbaum, E., & Viswanathan, R. (2024). Computational Methods to Predict Conformational B-Cell Epitopes. *Biomolecules*, *14*(8), 983. https://doi.org/10.3390/biom14080983

Castelli, M. S., McGonigle, P., & Hornby, P. J. (2019). The pharmacology and therapeutic applications of monoclonal antibodies. *Pharmacology Research & Perspectives*, *7*(6), e00535. https://doi.org/10.1002/prp2.535

Chandler, T. L., Yang, A., Otero, C. E., Permar, S. R., & Caddy, S. L. (2023). Protective mechanisms of nonneutralizing antiviral antibodies. *PLOS Pathogens*, *19*(10), e1011670. https://doi.org/10.1371/journal.ppat.1011670

Chapman, C. J., Murray, A., McElveen, J. E., Sahin, U., Luxemburger, U., Türeci, Ö., Wiewrodt, R., Barnes, A. C., & Robertson, J. F. (2008). *Autoantibodies in lung cancer: Possibilities for early detection and subsequent cure*. https://doi.org/10.1136/thx.2007.083592

Chen, W. S., Haynes, W. A., Waitz, R., Kamath, K., Vega-Crespo, A., Shrestha, R., Zhang, M., Foye, A., Carretero, I. B., Perez, I. G., Zhang, M., Zhao, S. G., Sjöström, M., Quigley, D. A., Chou, J., Beer, T. M., Rettig, M., Gleave, M., Evans, C. P., … Feng,

F. Y. (2020). Autoantibody landscape in patients with advanced prostate cancer. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *26*(23), 6204–6214. https://doi.org/10.1158/1078-0432.CCR-20-1966

Chen, Y.-C., Malfertheiner, P., Yu, H.-T., Kuo, C.-L., Chang, Y.-Y., Meng, F.-T., Wu, Y.-X., Hsiao, J.-L., Chen, M.-J., Lin, K.-P., Wu, C.-Y., Lin, J.-T., O'Morain, C., Megraud, F., Lee, W.-C., El-Omar, E. M., Wu, M.-S., & Liou, J.-M. (2024). Global Prevalence of Helicobacter pylori Infection and Incidence of Gastric Cancer Between 1980 and 2022. *Gastroenterology*, *166*(4), 605–619. https://doi.org/10.1053/j.gastro.2023.12.022

Chou, P. Y., & Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology and Related Areas of Molecular Biology*, *47*, 45–148. https://doi.org/10.1002/9780470122921.ch2

Chowell, D., Krishna, C., Pierini, F., Makarov, V., Rizvi, N. A., Kuo, F., Morris, L. G. T., Riaz, N., Lenz, T. L., & Chan, T. A. (2019). Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. *Nature Medicine*, *25*(11), 1715–1720. https://doi.org/10.1038/s41591-019-0639-4

Cia, G., Pucci, F., & Rooman, M. (2023). Critical review of conformational B-cell epitope prediction methods. *Briefings in Bioinformatics*, *24*(1), bbac567. https://doi.org/10.1093/bib/bbac567

Cohen, K. W., Linderman, S. L., Moodie, Z., Czartoski, J., Lai, L., Mantus, G., Norwood, C., Nyhoff, L. E., Edara, V. V., Floyd, K., Rosa, S. C. D., Ahmed, H., Whaley, R., Patel, S. N., Prigmore, B., Lemos, M. P., Davis, C. W., Furth, S., O'Keefe, J. B., … McElrath, M. J. (2021). Longitudinal analysis shows durable and broad immune memory after SARS-CoV-2 infection with persisting antibody responses and memory B and T cells. *Cell Reports Medicine*, *2*(7). https://doi.org/10.1016/j.xcrm.2021.100354

Cohen, S. P., Wang, E. J., Doshi, T. L., Vase, L., Cawcutt, K. A., & Tontisirin, N. (2022). Chronic pain and infection: Mechanisms, causes, conditions, treatments, and controversies. *BMJ Medicine*, *1*(1). https://doi.org/10.1136/bmjmed-2021-000108

Commins, S. P., James, H. R., Kelly, L. A., Pochan, S. L., Workman, L. J., Perzanowski, M. S., Kocan, K. M., Fahy, J. V., Nganga, L. W., Ronmark, E., Cooper, P. J., & Platts-Mills, T. A. E. (2011). The relevance of tick bites to the production of IgE antibodies to the mammalian oligosaccharide galactose-α-1,3-galactose. *The Journal of Allergy and Clinical Immunology*, *127*(5), 1286-1293.e6. https://doi.org/10.1016/j.jaci.2011.02.019

Crooke, S. N., Ovsyannikova, I. G., Kennedy, R. B., & Poland, G. A. (2020). Immunoinformatic identification of B cell and T cell epitopes in the SARS-CoV-2 proteome. *Scientific Reports*, *10*(1), 14179. https://doi.org/10.1038/s41598-020-70864-8

Cunningham, M. W., Antone, S. M., Smart, M., Liu, R., & Kosanke, S. (1997). Molecular analysis of human cardiac myosin-cross-reactive B- and T-cell epitopes of the group A streptococcal M5 protein. *Infection and Immunity*. (world). https://journals.asm.org/doi/10.1128/iai.65.9.3913-3923.1997

Dai, Y., Aizenbud, L., Qin, K., Austin, M., Jaycox, J. R., Cunningham, J., Wang, E. Y., Zhang, L., Fischer, S., Carroll, S. M., van Aggelen, H., Kluger, Y., Herold, K. C., Furchtgott, L., Kluger, H. M., & Ring, A. M. (2025). Humoral determinants of checkpoint

immunotherapy. *Nature*, *644*(8076), 527–536. https://doi.org/10.1038/s41586-025-09188-4

De Leon, A. J., Tjiam, M. C., & Yu, Y. (2024). B cell epitope mapping: The journey to better vaccines and therapeutic antibodies. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1868*(10), 130674. https://doi.org/10.1016/j.bbagen.2024.130674

de Moel, E. C., Rozeman, E. A., Kapiteijn, E. H., Verdegaal, E. M. E., Grummels, A., Bakker, J. A., Huizinga, T. W. J., Haanen, J. B., Toes, R. E. M., & van der Woude, D. (2019). Autoantibody Development under Treatment with Immune-Checkpoint Inhibitors. *Cancer Immunology Research*, *7*(1), 6–11. https://doi.org/10.1158/2326-6066.CIR-18-0245

de Wit, A. S., Bianchi, F., & van den Bogaart, G. (2025). Antigen presentation of post-translationally modified peptides in major histocompatibility complexes. *Immunology & Cell Biology*, *103*(2), 161–177. https://doi.org/10.1111/imcb.12839

Deepthi, V., Sasikumar, A., Mohanakumar, K. P., & Rajamma, U. (2025). Computationally designed multi-epitope vaccine construct targeting the SARS-CoV-2 spike protein elicits robust immune responses in silico. *Scientific Reports*, *15*(1), 9562. https://doi.org/10.1038/s41598-025-92956-z

Dowds, C. M., Kornell, S.-C., Blumberg, R. S., & Zeissig, S. (2014). Lipid antigens in immunity. *Biological Chemistry*, *395*(1), 61–81. https://doi.org/10.1515/hsz-2013-0220

Dugan, H. L., Guthmiller, J. J., Arevalo, P., Huang, M., Chen, Y.-Q., Neu, K. E., Henry, C., Zheng, N.-Y., Lan, L. Y.-L., Tepora, M. E., Stovicek, O., Bitar, D., Palm, A.-K. E., Stamper, C. T., Changrob, S., Utset, H. A., Coughlan, L., Krammer, F., Cobey, S., & Wilson, P. C. (2020). Preexisting immunity shapes distinct antibody landscapes after influenza virus infection and vaccination in humans. *Science Translational Medicine*, *12*(573), eabd3601. https://doi.org/10.1126/scitranslmed.abd3601

Dupic, T., Koraichi, M. B., Minervina, A. A., Pogorelyy, M. V., Mora, T., & Walczak, A. M. (2021). Immune fingerprinting through repertoire similarity. *PLOS Genetics*, *17*(1), e1009301. https://doi.org/10.1371/journal.pgen.1009301

Eisenberg, D., Weiss, R. M., & Terwilliger, T. C. (1984). The hydrophobic moment detects periodicity in protein hydrophobicity. *Proceedings of the National Academy of Sciences*, *81*(1), 140–144. https://doi.org/10.1073/pnas.81.1.140

Elbre, E. (2017). *SPEXS2—An exhaustive sequence pattern search tool* (Version v1.0.3) [Computer software]. https://github.com/egonelbre/spexs2

EL-Manzalawy, Y., & Honavar, V. (2010). Recent advances in B-cell epitope prediction methods. *Immunome Research*, *6*(2), S2. https://doi.org/10.1186/1745-7580-6-S2-S2

Elsen, V. D., & J, P. (2011). Expression Regulation of Major Histocompatibility Complex Class I and Class II Encoding Genes. *Frontiers in Immunology*, *2*. https://doi.org/10.3389/fimmu.2011.00048

Emini, E. A., Hughes, J. V., Perlow, D. S., & Boger, J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of Virology*, *55*(3), 836–839. https://doi.org/10.1128/JVI.55.3.836-839.1985

Engell-Noerregaard, L., Kvistborg, P., Zocca, M.-B., Pedersen, A. W., Claesson, M. H., & Mellemgaard, A. (2013). Clinical and Immunological Effects in Patients with Advanced Non-Small Cell Lung-Cancer after Vaccination with Dendritic Cells

Exposed to an Allogeneic Tumor Cell Lysate. *World Journal of Vaccines*, *3*(2), 68–76. https://doi.org/10.4236/wjv.2013.32011

Farriol-Duran, R., López-Aladid, R., Porta-Pardo, E., Torres, A., & Fernández-Barat, L. (2023). Brewpitopes: A pipeline to refine B-cell epitope predictions during public health emergencies. *Frontiers in Immunology*, *14*, 1278534. https://doi.org/10.3389/fimmu.2023.1278534

FDA-NIH Biomarker Working Group. (2016). *BEST (Biomarkers, EndpointS, and other Tools) Resource*. Food and Drug Administration (US). http://www.ncbi.nlm.nih.gov/books/NBK326791/

Ferguson, J. A., Raghavan, S. S. R., Alzua, G. P., Bhavsar, D., Huang, J., Rodriguez, A. J., Torres, J. L., Bottermann, M., Han, J., Krammer, F., Batista, F. D., & Ward, A. B. (2025). Functional and epitope specific monoclonal antibody discovery directly from immune sera using cryo-EM. *Science Advances*, *11*(33), eadv8257. https://doi.org/10.1126/sciadv.adv8257

Francis, T. (1960). On the Doctrine of Original Antigenic Sin. *Proceedings of the American Philosophical Society*, *104*(6), 572–578. JSTOR.

Fratta, E., Coral, S., Covre, A., Parisi, G., Colizzi, F., Danielli, R., Marie Nicolay, H. J., Sigalotti, L., & Maio, M. (2011). The biology of cancer testis antigens: Putative function, regulation and therapeutic potential. *Molecular Oncology*, *5*(2), 164–182. https://doi.org/10.1016/j.molonc.2011.02.001

Friedberg, E. C. (2008). Sydney Brenner. *Nature Reviews Molecular Cell Biology*, *9*(1), 8–9. https://doi.org/10.1038/nrm2320

Galle, P., Finn, R. S., Mitchell, C. R., Ndirangu, K., Ramji, Z., Redhead, G. S., & Pinato, D. J. (2024). Treatment-emergent antidrug antibodies related to PD-1, PD-L1, or CTLA-4 inhibitors across tumor types: A systematic review. *Journal for ImmunoTherapy of Cancer*, *12*(1). https://doi.org/10.1136/jitc-2023-008266

García-Nafría, J., & Tate, C. G. (2021). Structure determination of GPCRs: Cryo-EM compared with X-ray crystallography. *Biochemical Society Transactions*, *49*(5), 2345–2355. https://doi.org/10.1042/BST20210431

Geffen, Y., Anand, S., Akiyama, Y., Yaron, T. M., Song, Y., Johnson, J. L., Govindan, A., Babur, Ö., Li, Y., Huntsman, E., Wang, L.-B., Birger, C., Heiman, D. I., Zhang, Q., Miller, M., Maruvka, Y. E., Haradhvala, N. J., Calinawan, A., Belkin, S., … Zhou, D. C. (2023). Pan-cancer analysis of post-translational modifications reveals shared patterns of protein regulation. *Cell*, *186*(18), 3945-3967.e26. https://doi.org/10.1016/j.cell.2023.07.013

Gogas, H., Ioannovich, J., Dafni, U., Stavropoulou-Giokas, C., Frangia, K., Tsoutsos, D., Panagiotou, P., Polyzos, A., Papadopoulos, O., Stratigos, A., Markopoulos, C., Bafaloukos, D., Pectasides, D., Fountzilas, G., & Kirkwood, J. M. (2006). Prognostic significance of autoimmunity during treatment of melanoma with interferon. *The New England Journal of Medicine*, *354*(7), 709–718. https://doi.org/10.1056/NEJMoa053007

Gostic, K. M., Ambrose, M., Worobey, M., & Lloyd-Smith, J. O. (2016). Potent Protection against H5N1 and H7N9 Influenza via Childhood Hemagglutinin Imprinting. *Science (New York, N.Y.)*, *354*(6313), 722–726. https://doi.org/10.1126/science.aag1322

Gowen, M. F., Giles, K. M., Simpson, D., Tchack, J., Zhou, H., Moran, U., Dawood, Z., Pavlick, A. C., Hu, S., Wilson, M. A., Zhong, H., Krogsgaard, M., Kirchhoff, T., & Osman, I. (2018). Baseline antibody profiles predict toxicity in melanoma

patients treated with immune checkpoint inhibitors. *Journal of Translational Medicine*, *16*(1), 82. https://doi.org/10.1186/s12967-018-1452-4

Gunn, G. R., Sealey, D. C. F., Jamali, F., Meibohm, B., Ghosh, S., & Shankar, G. (2016). From the bench to clinical practice: Understanding the challenges and uncertainties in immunogenicity testing for biopharmaceuticals. *Clinical and Experimental Immunology*, *184*(2), 137–146. https://doi.org/10.1111/cei.12742

Gurung, H. R., Heidersbach, A. J., Darwish, M., Chan, P. P. F., Li, J., Beresini, M., Zill, O. A., Wallace, A., Tong, A.-J., Hascall, D., Torres, E., Chang, A., Lou, K. 'Hei-W., Abdolazimi, Y., Hammer, C., Xavier-Magalhães, A., Marcu, A., Vaidya, S., Le, D. D., … Rose, C. M. (2024). Systematic discovery of neoepitope–HLA pairs for neoantigens shared among patients and tumor types. *Nature Biotechnology*, *42*(7), 1107–1117. https://doi.org/10.1038/s41587-023-01945-y

Guthmiller, J. J., Han, J., Utset, H. A., Li, L., Lan, L. Y.-L., Henry, C., Stamper, C. T., McMahon, M., O'Dell, G., Fernández-Quintero, M. L., Freyn, A. W., Amanat, F., Stovicek, O., Gentles, L., Richey, S. T., de la Peña, A. T., Rosado, V., Dugan, H. L., Zheng, N.-Y., … Wilson, P. C. (2022). Broadly neutralizing antibodies target a haemagglutinin anchor epitope. *Nature*, *602*(7896), 314–320. https://doi.org/10.1038/s41586-021-04356-8

Hansson, G. K. (2005). Inflammation, Atherosclerosis, and Coronary Artery Disease. *New England Journal of Medicine*, *352*(16), 1685–1695. https://doi.org/10.1056/NEJMra043430

Haynes, W. A., Kamath, K., Bozekowski, J., Baum-Jones, E., Campbell, M., Casanovas-Massana, A., Daugherty, P. S., Dela Cruz, C. S., Dhal, A., Farhadian, S. F., Fitzgibbons, L., Fournier, J., Jhatro, M., Jordan, G., Klein, J., Lucas, C., Kessler, D., Luchsinger, L. L., Martinez, B., … Shon, J. C. (2021). High-resolution epitope mapping and characterization of SARS-CoV-2 antibodies in large cohorts of subjects with COVID-19. *Communications Biology*, *4*(1), 1317. https://doi.org/10.1038/s42003-021-02835-2

Haynes, W. A., Kamath, K., Waitz, R., Daugherty, P. S., & Shon, J. C. (2021). Protein-Based Immunome Wide Association Studies (PIWAS) for the Discovery of Significant Disease-Associated Antigens. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.625311

Heras-Murillo, I., Mañanes, D., Munné, P., Núñez, V., Herrera, J., Catalá-Montoro, M., Alvarez, M., del Pozo, M. A., Melero, I., Wculek, S. K., & Sancho, D. (2025). Immunotherapy with conventional type-1 dendritic cells induces immune memory and limits tumor relapse. *Nature Communications*, *16*(1), 3369. https://doi.org/10.1038/s41467-025-58289-1

Herrero-Montes, M., Fernández-de-las-Peñas, C., Ferrer-Pargada, D., Tello-Mena, S., Cancela-Cilleruelo, I., Rodríguez-Jiménez, J., Palacios-Ceña, D., & Parás-Bravo, P. (2022). Prevalence of Neuropathic Component in Post-COVID Pain Symptoms in Previously Hospitalized COVID-19 Survivors. *International Journal of Clinical Practice*, *2022*(1), 3532917. https://doi.org/10.1155/2022/3532917

Høie, M. H., Gade, F. S., Johansen, J. M., Würtzen, C., Winther, O., Nielsen, M., & Marcatili, P. (2024). DiscoTope-3.0: Improved B-cell epitope prediction using inverse folding latent representations. *Frontiers in Immunology*, *15*. https://doi.org/10.3389/fimmu.2024.1322712

Høie, M. H., Kiehl, E. N., Petersen, B., Nielsen, M., Winther, O., Nielsen, H., Hallgren, J., & Marcatili, P. (2022). NetSurfP-3.0: Accurate and fast prediction of protein

structural features by protein language models and deep learning. *Nucleic Acids Research*, *50*(W1), W510–W515. https://doi.org/10.1093/nar/gkac439

Hopp, T. P., & Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences*, *78*(6), 3824–3828. https://doi.org/10.1073/pnas.78.6.3824

Hörauf, J.-A., Schindler, C. R., Schaible, I., Wang, M., Weber, B., El Saman, A., Pallas, C., Widera, M., Marzi, I., Henrich, D., & Leppik, L. (2024). Extracellular vesicles epitopes as potential biomarker candidates in patients with traumatic spinal cord injury. *Frontiers in Immunology*, *15*, 1478786. https://doi.org/10.3389/fimmu.2024.1478786

Hu, D., & Irving, A. T. (2023). Massively-multiplexed epitope mapping techniques for viral antigen discovery. *Frontiers in Immunology*, *14*. https://doi.org/10.3389/fimmu.2023.1192385

Hu, Z., Zhao, J., Shi, L., Hu, J., Hu, S., & Liu, X. (2021). Identification of the dominant non-neutralizing epitope in the haemagglutinin of H7N9 avian influenza virus. *Virus Research*, *298*, 198409. https://doi.org/10.1016/j.virusres.2021.198409

Huang, C.-C., Cheng, K.-W., Hsieh, Y.-C., Lin, W.-W., Cheng, C.-M., Yuan, S.-S. F., Chen, I.-J., Cheng, Y.-A., Lu, Y.-C., Huang, B.-C., Tung, Y.-C., & Cheng, T.-L. (2019). Use of syngeneic cells expressing membrane-bound GM-CSF as an adjuvant to induce antibodies against native multi-pass transmembrane protein. *Scientific Reports*, *9*(1), 9931. https://doi.org/10.1038/s41598-019-45160-9

Hwang, J.-R., Byeon, Y., Kim, D., & Park, S.-G. (2020). Recent insights of T cell receptor-mediated signaling pathways for T cell activation and development. *Experimental & Molecular Medicine*, *52*(5), 750–761. https://doi.org/10.1038/s12276-020-0435-8

Ilbäck, N. G., Mohammed, A., Fohlman, J., & Friman, G. (1990). Cardiovascular lipid accumulation with Coxsackie B virus infection in mice. *The American Journal of Pathology*, *136*(1), 159–167.

Israeli, S., & Louzoun, Y. (2024). Single-residue linear and conformational B cell epitopes prediction using random and ESM-2 based projections. *Briefings in Bioinformatics*, *25*(2), bbae084. https://doi.org/10.1093/bib/bbae084

Jaago, M., Pupina, N., Rähni, A., Pihlak, A., Sadam, H., Vrana, N. E., Sinisalo, J., Pussinen, P., & Palm, K. (2022). Antibody response to oral biofilm is a biomarker for acute coronary syndrome in periodontal disease. *Communications Biology*, *5*(1), 205. https://doi.org/10.1038/s42003-022-03122-4

Jeong, Y. D., Jo, H., Yim, Y., Lee, S., Park, J., Lee, J., Kang, J., Jacob, L., Smith, L., Rahmati, M., López Sánchez, G. F., Lee, H., & Yon, D. K. (2025). Global estimates of vaccine-associated narcolepsy from 1967 to 2023. *Scientific Reports*, *15*(1), 21331. https://doi.org/10.1038/s41598-025-04049-6

Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*, *45*(W1), W24–W29. https://doi.org/10.1093/nar/gkx346

Jhunjhunwala, S., Hammer, C., & Delamarre, L. (2021). Antigen presentation in cancer: Insights into tumour immunogenicity and immune evasion. *Nature Reviews Cancer*, *21*(5), 298–312. https://doi.org/10.1038/s41568-021-00339-z

Ji, S., Wang, F., Wu, Y., Hu, H., Xing, Z., Zhu, J., Xu, S., Han, T., Liu, G., Wu, Z., Fei, C., Kong, L., Chen, J., Ding, Z., Huang, Z., & Zhang, J. (2025). Large-scale transcript variants

dictate neoepitopes for cancer immunotherapy. *Science Advances*, *11*(5), eado5600. https://doi.org/10.1126/sciadv.ado5600

Jia, Q., Wang, A., Yuan, Y., Zhu, B., & Long, H. (2022). Heterogeneity of the tumor immune microenvironment and its clinical relevance. *Experimental Hematology & Oncology*, *11*(1), 24. https://doi.org/10.1186/s40164-022-00277-y

Jonaitis, P., Petkeviciene, J., Salteniene, V., Ciupkeviciene, E., Jonaitis, L., Kriukas, M., Luksiene, D., Lesauskaite, V., Kupcinskas, J., & Kupcinskas, L. (2025). Helicobacter pylori Seroprevalence and Its Associations with Sociodemographic Characteristics, Environmental Factors, and Gastrointestinal Complaints: A Cross-Sectional Study in the Adult Population of Kaunas City, Lithuania. *Medicina*, *61*(6), 1049. https://doi.org/10.3390/medicina61061049

Ju, B., Zhang, Q., Ge, J., Wang, R., Sun, J., Ge, X., Yu, J., Shan, S., Zhou, B., Song, S., Tang, X., Yu, J., Lan, J., Yuan, J., Wang, H., Zhao, J., Zhang, S., Wang, Y., Shi, X., … Zhang, L. (2020). Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature*, *584*(7819), 115–119. https://doi.org/10.1038/s41586-020-2380-z

Jung, S.-H., & Lee, K.-T. (2022). Atherosclerosis by Virus Infection—A Short Review. *Biomedicines*, *10*(10), 2634. https://doi.org/10.3390/biomedicines10102634

Kallio-Laine, K., Seppänen, M., Lokki, M.-L., Lappalainen, M., Notkola, I.-L., Seppälä, I., Koskinen, M., Valtonen, V., & Kalso, E. (2008). Widespread Unilateral Pain Associated With Herpes Simplex Virus Infections. *The Journal of Pain*, *9*(7), 658–665. https://doi.org/10.1016/j.jpain.2008.02.003

Kamath, K., Reifert, J., Johnston, T., Gable, C., Pantazes, R. J., Rivera, H. N., McAuliffe, I., Handali, S., & Daugherty, P. S. (2020). Antibody epitope repertoire analysis enables rapid antigen discovery and multiplex serology. *Scientific Reports*, *10*(1), 5294. https://doi.org/10.1038/s41598-020-62256-9

Karplus, P. A., & Schulz, G. E. (1985). Prediction of chain flexibility in proteins. *Naturwissenschaften*, *72*(4), 212–213. https://doi.org/10.1007/BF01195768

Kaunisto, M. A., Jokela, R., Tallgren, M., Kambur, O., Tikkanen, E., Tasmuth, T., Sipilä, R., Palotie, A., Estlander, A.-M., Leidenius, M., Ripatti, S., & Kalso, E. A. (2013). Pain in 1,000 women treated for breast cancer: A prospective study of pain sensitivity and postoperative pain. *Anesthesiology*, *119*(6), 1410–1421. https://doi.org/10.1097/ALN.0000000000000012

Kellermann, G., Leulliot, N., Cherfils-Vicini, J., Blaud, M., & Brest, P. (2024). Activated B-Cells enhance epitope spreading to support successful cancer immunotherapy. *Frontiers in Immunology*, *15*. https://doi.org/10.3389/fimmu.2024.1382236

Kenison, J. E., Stevens, N. A., & Quintana, F. J. (2024). Therapeutic induction of antigen-specific immune tolerance. *Nature Reviews Immunology*, *24*(5), 338–357. https://doi.org/10.1038/s41577-023-00970-x

Khan, T., Rahman, M., Ahmed, I., Al Ali, F., Jithesh, P. V., & Marr, N. (2022). Human leukocyte antigen class II gene diversity tunes antibody repertoires to common pathogens. *Frontiers in Immunology*, *13*. https://doi.org/10.3389/fimmu.2022.856497

Khurana, S., Loving, C. L., Manischewitz, J., King, L. R., Gauger, P. C., Henningson, J., Vincent, A. L., & Golding, H. (2013). Vaccine-Induced Anti-HA2 Antibodies Promote Virus Fusion and Enhance Influenza Virus Respiratory Disease. *Science Translational Medicine*, *5*(200), 200ra114-200ra114. https://doi.org/10.1126/scitranslmed.3006366

Kidera, A., Konishi, Y., Oka, M., Ooi, T., & Scheraga, H. A. (1985). Statistical analysis of the physical properties of the 20 naturally occurring amino acids. *Journal of Protein Chemistry*, *4*(1), 23–55. https://doi.org/10.1007/BF01025492

Klebanoff, C. A., & Wolchok, J. D. (2017). Shared cancer neoantigens: Making private matters public. *Journal of Experimental Medicine*, *215*(1), 5–7. https://doi.org/10.1084/jem.20172188

Kolakowski, L. F., Leunissen, J. A., & Smith, J. E. (1992). ProSearch: Fast searching of protein sequences with regular expression patterns related to protein structure and function. *BioTechniques*, *13*(6), 919–921.

Kolaskar, A. s., & Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters*, *276*(1–2), 172–174. https://doi.org/10.1016/0014-5793(90)80535-Q

Kong, E., Hua, T., Li, J., Li, Y., Yang, M., Ding, R., Wang, H., Wei, H., Feng, X., Han, C., & Yuan, H. (2024). HSV-1 reactivation results in post-herpetic neuralgia by upregulating Prmt6 and inhibiting cGAS-STING. *Brain*, *147*(7), 2552–2565. https://doi.org/10.1093/brain/awae053

Kong, P., Cui, Z.-Y., Huang, X.-F., Zhang, D.-D., Guo, R.-J., & Han, M. (2022). Inflammation and atherosclerosis: Signaling pathways and therapeutic intervention. *Signal Transduction and Targeted Therapy*, *7*(1), 131. https://doi.org/10.1038/s41392-022-00955-7

Kringelum, J. V., Nielsen, M., Padkjær, S. B., & Lund, O. (2013). Structural analysis of B-cell epitopes in antibody:protein complexes. *Molecular Immunology*, *53*(1), 24–34. https://doi.org/10.1016/j.molimm.2012.06.001

Kuri, A., Jacobs, B. M., Vickaryous, N., Pakpoor, J., Middeldorp, J., Giovannoni, G., & Dobson, R. (2020). Epidemiology of Epstein-Barr virus infection and infectious mononucleosis in the United Kingdom. *BMC Public Health*, *20*(1), 912. https://doi.org/10.1186/s12889-020-09049-x

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*(1), 105–132. https://doi.org/10.1016/0022-2836(82)90515-0

Laires, P. A., Dias, S., Gama, A., Moniz, M., Pedro, A. R., Soares, P., Aguiar, P., & Nunes, C. (2021). The Association Between Chronic Disease and Serious COVID-19 Outcomes and Its Influence on Risk Perception: Survey Study and Database Analysis. *JMIR Public Health and Surveillance*, *7*(1), e22794. https://doi.org/10.2196/22794

Larman, H. B., Laserson, U., Querol, L., Verhaeghen, K., Solimini, N. L., Xu, G. J., Klarenbeek, P. L., Church, G. M., Hafler, D. A., Plenge, R. M., Nigrovic, P. A., De Jager, P. L., Weets, I., Martens, G. A., O'Connor, K. C., & Elledge, S. J. (2013). PhIP-Seq characterization of autoantibodies from patients with multiple sclerosis, type 1 diabetes and rheumatoid arthritis. *Journal of Autoimmunity*, *43*, 1–9. https://doi.org/10.1016/j.jaut.2013.01.013

Larman, H. B., Zhao, Z., Laserson, U., Li, M. Z., Ciccia, A., Gakidis, M. A. M., Church, G. M., Kesari, S., LeProust, E. M., Solimini, N. L., & Elledge, S. J. (2011). Application of a synthetic human proteome to autoantigen discovery through PhIP-Seq. *Nature Biotechnology*, *29*(6), 535–541. https://doi.org/10.1038/nbt.1856

Larsen, J. E. P., Lund, O., & Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome Research*, *2*, 2. https://doi.org/10.1186/1745-7580-2-2

Legutki, J. B., Zhao, Z.-G., Greving, M., Woodbury, N., Johnston, S. A., & Stafford, P. (2014). Scalable high-density peptide arrays for comprehensive health monitoring. *Nature Communications*, *5*(1), 4785. https://doi.org/10.1038/ncomms5785

Leja, M., Cine, E., Rudzite, D., Vilkoite, I., Huttunen, T., Daugule, I., Rumba-Rozenfelde, I., Pimanov, S., Liepniece-Karele, I., Pahomova, J., Purmalis, K., Eglitis, J., Pirags, V., Dzerve, V., & Erglis, A. (2012). Prevalence of Helicobacter pylori infection and atrophic gastritis in Latvia. *European Journal of Gastroenterology & Hepatology*, *24*(12), 1410. https://doi.org/10.1097/MEG.0b013e3283583ca5

Leppik, L., Parksepp, M., Janno, S., Koido, K., Haring, L., Vasar, E., & Zilmer, M. (2020). Profiling of lipidomics before and after antipsychotic treatment in first-episode psychosis. *European Archives of Psychiatry and Clinical Neuroscience*, *270*(1), 59–70. https://doi.org/10.1007/s00406-018-0971-6

Levine, M. H., Haberman, A. M., Sant'Angelo, D. B., Hannum, L. G., Cancro, M. P., Janeway, C. A., & Shlomchik, M. J. (2000). A B-cell receptor-specific selection step governs immature to mature B cell differentiation. *Proceedings of the National Academy of Sciences*, *97*(6), 2743–2748. https://doi.org/10.1073/pnas.050552997

Li, J., Ju, Y., Jiang, M., Li, S., & Yang, X.-Y. (2025). Epitope-Based Vaccines: The Next Generation of Promising Vaccines Against Bacterial Infection. *Vaccines*, *13*(3), 248. https://doi.org/10.3390/vaccines13030248

Li, J.-N., Wang, H., Han, Y.-X., Zhao, Y.-T., Zhou, H.-H., Xu, J., & Li, L. (2019). Novel peptides screened by phage display peptide library can mimic epitopes of the FnBPA-A protein and induce protective immunity against Staphylococcus aureus in mice. *MicrobiologyOpen*, *8*(10), e910. https://doi.org/10.1002/mbo3.910

Li, Q., Zeng, H., Liu, T., Wang, P., Zhang, R., Zhao, B., Feng, T., Yang, Y., Wu, J., Zheng, Y., Zhou, B., Shu, Y., Xu, H., Yang, L., & Ding, Z. (2024). A dendritic cell vaccine for both vaccination and neoantigen-reactive T cell preparation for cancer immunotherapy in mice. *Nature Communications*, *15*(1), 10419. https://doi.org/10.1038/s41467-024-54650-y

Li, Y., Bourlet, T., Andreoletti, L., Mosnier, J.-F., Peng, T., Yang, Y., Archard, L. C., Pozzetto, B., & Zhang, H. (2000). Enteroviral Capsid Protein VP1 Is Present in Myocardial Tissues From Some Patients With Myocarditis or Dilated Cardiomyopathy. *Circulation*, *101*(3), 231–234. https://doi.org/10.1161/01.CIR.101.3.231

Li, Y., Lai, D., Zhang, H., Jiang, H., Tian, X., Ma, M., Qi, H., Meng, Q., Guo, S., Wu, Y., Wang, W., Yang, X., Shi, D., Dai, J., Ying, T., Zhou, J., & Tao, S. (2020). Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. *Cellular & Molecular Immunology*, *17*(10), 1095–1097. https://doi.org/10.1038/s41423-020-00523-5

Li, Y., Liu, X., Zhang, X., Pan, W., Li, N., & Tang, B. (2021). Immune Cycle-Based Strategies for Cancer Immunotherapy. *Advanced Functional Materials*, *31*(50), 2107540. https://doi.org/10.1002/adfm.202107540

Liebhoff, A.-M., Venkataraman, T., Morgenlander, W. R., Na, M., Kula, T., Waugh, K., Morrison, C., Rewers, M., Longman, R., Round, J., Elledge, S., Ruczinski, I., Langmead, B., & Larman, H. B. (2024). Efficient encoding of large antigenic spaces by epitope prioritization with Dolphyn. *Nature Communications*, *15*(1), 1577. https://doi.org/10.1038/s41467-024-45601-8

Lima, N. S., Musayev, M., Johnston, T. S., Wagner, D. A., Henry, A. R., Wang, L., Yang, E. S., Zhang, Y., Birungi, K., Black, W. P., O'Dell, S., Schmidt, S. D., Moon, D., Lorang, C. G., Zhao, B., Chen, M., Boswell, K. L., Roberts-Torres, J., Davis, R. L., … Douek, D. C. (2022). Primary exposure to SARS-CoV-2 variants elicits convergent epitope specificities, immunoglobulin V gene usage and public B cell clones. *Nature Communications*, *13*(1), 7733. https://doi.org/10.1038/s41467-022-35456-2

Liu, F., Yuan, C., Chen, H., & Yang, F. (2024). Prediction of linear B-cell epitopes based on protein sequence features and BERT embeddings. *Scientific Reports*, *14*(1), 2464. https://doi.org/10.1038/s41598-024-53028-w

Liu, J., Wang, Y., Xiong, E., Hong, R., Lu, Q., Ohno, H., & Wang, J.-Y. (2019). Role of the IgM Fc Receptor in Immunity and Tolerance. *Frontiers in Immunology*, *10*. https://doi.org/10.3389/fimmu.2019.00529

Liu, J.-C., Zhang, K., Zhang, X., Guan, F., Zeng, H., Kubo, M., Lee, P., Candotti, F., James, L. K., Camara, N. O. S., Benlagha, K., Lei, J.-H., Forsman, H., Yang, L., Xiao, W., Liu, Z., & Liu, C.-H. (2024). Immunoglobulin class-switch recombination: Mechanism, regulation, and related diseases. *MedComm*, *5*(8), e662. https://doi.org/10.1002/mco2.662

Liu, X., & Wu, J. (2018). History, applications, and challenges of immune repertoire research. *Cell Biology and Toxicology*, *34*(6), 441–457. https://doi.org/10.1007/s10565-018-9426-0

Liu, X.-Y., Chen, Y.-L., Liu, G.-J., Deng, X.-N., Cui, Y., Tan, J., Dong, X.-C., Li, H.-Y., Chen, G.-J., Ou, Z.-M., & Wang, C.-H. (2022). Development of a variant of dinutuximab with enhanced antitumor efficacy and reduced induction of neuropathic pain. *FEBS Open Bio*, *12*(9), 1644–1656. https://doi.org/10.1002/2211-5463.13464

Loh, T. J., Lim, J. J., Jones, C. M., Dao, H. T., Tran, M. T., Baker, D. G., La Gruta, N. L., Reid, H. H., & Rossjohn, J. (2024). The molecular basis underlying T cell specificity towards citrullinated epitopes presented by HLA-DR4. *Nature Communications*, *15*(1), 6201. https://doi.org/10.1038/s41467-024-50511-w

Lossius, A., Johansen, J. N., Vartdal, F., & Holmøy, T. (2016). High-throughput sequencing of immune repertoires in multiple sclerosis. *Annals of Clinical and Translational Neurology*, *3*(4), 295–306. https://doi.org/10.1002/acn3.295

Lozier, J. N., Tayebi, N., & Zhang, P. (2005). Mapping of genes that control the antibody response to human factor IX in mice. *Blood*, *105*(3), 1029–1035. https://doi.org/10.1182/blood-2004-03-1126

Lu, H., Liao, Y., Zhang, C., Wen, W., Du, Y., Zhao, M., & Wang, L. (2023). A case of herpes simplex virus induced peripheral neuropathy and encephalitis with positive GM3 and CASPR2 antibody. *BMC Neurology*, *23*(1), 199. https://doi.org/10.1186/s12883-023-03238-y

Lu, L. L., Suscovich, T. J., Fortune, S. M., & Alter, G. (2018). Beyond binding: Antibody effector functions in infectious diseases. *Nature Reviews Immunology*, *18*(1), 46–61. https://doi.org/10.1038/nri.2017.106

Maltsev, D., & Fedirko, V. (2022). Refractory atypical trigeminal neuralgia associated with reactivated herpesvirus infection: Pathogenetic link and efficacy of combination antiviral therapy. *VirusDisease*, *33*(2), 155–165. https://doi.org/10.1007/s13337-022-00769-9

Manjula, B. N., Trus, B. L., & Fischetti, V. A. (1985). Presence of two distinct regions in the coiled-coil structure of the streptococcal Pep M5 protein: Relationship to mammalian coiled-coil proteins and implications to its biological properties.

*Proceedings of the National Academy of Sciences of the United States of America*, *82*(4), 1064–1068. https://doi.org/10.1073/pnas.82.4.1064

Márquez, E. J., Chung, C., Marches, R., Rossi, R. J., Nehar-Belaid, D., Eroglu, A., Mellert, D. J., Kuchel, G. A., Banchereau, J., & Ucar, D. (2020). Sexual-dimorphism in human immune system aging. *Nature Communications*, *11*(1), 751. https://doi.org/10.1038/s41467-020-14396-9

Mateus, J., Grifoni, A., Tarke, A., Sidney, J., Ramirez, S. I., Dan, J. M., Burger, Z. C., Rawlings, S. A., Smith, D. M., Phillips, E., Mallal, S., Lammers, M., Rubiro, P., Quiambao, L., Sutherland, A., Yu, E. D., da Silva Antunes, R., Greenbaum, J., Frazier, A., … Weiskopf, D. (2020). Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science*, *370*(6512), 89–94. https://doi.org/10.1126/science.abd3871

Matsumoto, K., Harada, S. Y., Yoshida, S. Y., Narumi, R., Mitani, T. T., Yada, S., Sato, A., Morii, E., Shimizu, Y., & Ueda, H. R. (2025). DECODE enables high-throughput mapping of antibody epitopes at single amino acid resolution. *PLOS Biology*, *23*(1), e3002707. https://doi.org/10.1371/journal.pbio.3002707

McClain, S. (2017). Bioinformatic screening and detection of allergen cross-reactive IgE-binding epitopes. *Molecular Nutrition & Food Research*, *61*(8), 1600676. https://doi.org/10.1002/mnfr.201600676

Mei, S., Li, F., Leier, A., Marquez-Lago, T. T., Giam, K., Croft, N. P., Akutsu, T., Smith, A. I., Li, J., Rossjohn, J., Purcell, A. W., & Song, J. (2020). A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction. *Briefings in Bioinformatics*, *21*(4), 1119–1135. https://doi.org/10.1093/bib/bbz051

Mentzer, A. J., Brenner, N., Allen, N., Littlejohns, T. J., Chong, A. Y., Cortes, A., Almond, R., Hill, M., Sheard, S., McVean, G., Collins, R., Hill, A. V. S., & Waterboer, T. (2022). Identification of host–pathogen-disease relationships using a scalable multiplex serology platform in UK Biobank. *Nature Communications*, *13*(1), 1818. https://doi.org/10.1038/s41467-022-29307-3

Merino, F., & Raunser, S. (2017). Electron Cryo-microscopy as a Tool for Structure-Based Drug Development. *Angewandte Chemie International Edition*, *56*(11), 2846–2860. https://doi.org/10.1002/anie.201608432

Merkenschlager, J., Pyo, A. G. T., Silva Santos, G. S., Schaefer-Babajew, D., Cipolla, M., Hartweger, H., Gitlin, A. D., Wingreen, N. S., & Nussenzweig, M. C. (2025). Regulated somatic hypermutation enhances antibody affinity maturation. *Nature*, *641*(8062), 495–502. https://doi.org/10.1038/s41586-025-08728-2

Mina, M. J., Kula, T., Leng, Y., Li, M., de Vries, R. D., Knip, M., Siljander, H., Rewers, M., Choy, D. F., Wilson, M. S., Larman, H. B., Nelson, A. N., Griffin, D. E., de Swart, R. L., & Elledge, S. J. (2019). Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. *Science*, *366*(6465), 599–606. https://doi.org/10.1126/science.aay6485

Mishra, N., Huang, X., Joshi, S., Guo, C., Ng, J., Thakkar, R., Wu, Y., Dong, X., Li, Q., Pinapati, R. S., Sullivan, E., Caciula, A., Tokarz, R., Briese, T., Lu, J., & Lipkin, W. I. (2021). Immunoreactive peptide maps of SARS-CoV-2. *Communications Biology*, *4*, 225. https://doi.org/10.1038/s42003-021-01743-9

Mitsunaga, E. M., & Snyder, M. P. (2020). Deep Characterization of the Human Antibody Response to Natural Infection Using Longitudinal Immune Repertoire

Sequencing*. *Molecular & Cellular Proteomics*, *19*(2), 278–293. https://doi.org/10.1074/mcp.RA119.001633

Mohan, D., Wansley, D. L., Sie, B. M., Noon, M. S., Baer, A. N., Laserson, U., & Larman, H. B. (2018). PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. *Nature Protocols*, *13*(9), 1958–1978. https://doi.org/10.1038/s41596-018-0025-6

Monroy-Iglesias, M. J., Crescioli, S., Beckmann, K., Le, N., Karagiannis, S. N., Van Hemelrijck, M., & Santaolalla, A. (2022). Antibodies as biomarkers for cancer risk: A systematic review. *Clinical and Experimental Immunology*, *209*(1), 46–63. https://doi.org/10.1093/cei/uxac030

Montero-Calle, A., Garranzo-Asensio, M., Moreno-Casbas, M. T., Campuzano, S., & Barderas, R. (2024). Autoantibodies in cancer: A systematic review of their clinical role in the most prevalent cancers. *Frontiers in Immunology*, *15*. https://doi.org/10.3389/fimmu.2024.1455602

Muenst, S., Soysal, S. D., Gao, F., Obermann, E. C., Oertli, D., & Gillanders, W. E. (2013). The presence of programmed death 1 (PD-1)-positive tumor-infiltrating lymphocytes is associated with poor prognosis in human breast cancer. *Breast Cancer Research and Treatment*, *139*(3), 667–676. https://doi.org/10.1007/s10549-013-2581-3

Musher, D. M., Abers, M. S., & Corrales-Medina, V. F. (2019). Acute Infection and Myocardial Infarction. *New England Journal of Medicine*, *380*(2), 171–176. https://doi.org/10.1056/NEJMra1808137

Mustonen, L., Aho, T., Harno, H., Sipilä, R., Meretoja, T., & Kalso, E. (2019). What makes surgical nerve injury painful? A 4-year to 9-year follow-up of patients with intercostobrachial nerve resection in women treated for breast cancer. *Pain*, *160*(1), 246–256. https://doi.org/10.1097/j.pain.0000000000001398

Najar, T. A., Khare, S., Pandey, R., Gupta, S. K., & Varadarajan, R. (2017). Mapping Protein Binding Sites and Conformational Epitopes Using Cysteine Labeling and Yeast Surface Display. *Structure*, *25*(3), 395–406. https://doi.org/10.1016/j.str.2016.12.016

Natarajan, H., Crowley, A. R., Butler, S. E., Xu, S., Weiner, J. A., Bloch, E. M., Littlefield, K., Wieland-Alter, W., Connor, R. I., Wright, P. F., Benner, S. E., Bonny, T. S., Laeyendecker, O., Sullivan, D., Shoham, S., Quinn, T. C., Larman, H. B., Casadevall, A., Pekosz, A., … Ackerman, M. E. (2021). Markers of Polyfunctional SARS-CoV-2 Antibodies in Convalescent Plasma. *mBio*, *12*(2), e00765-21. https://doi.org/10.1128/mBio.00765-21

Ng, E. S., Milotay, G., Tong, O., A. Taylor, C., Sun, S., Niu, G., Watson, R., Sun, B., MacKay, S., Gilchrist, J. J., Little, M., Fairfax, B. P., & Luo, Y. (2025). Defining the genetic determinants of CD8+ T cell receptor repertoire in the context of immune checkpoint blockade. *Science Advances*, *11*(30), eadu3461. https://doi.org/10.1126/sciadv.adu3461

Nguyen, T. T. N., Choo, E. M., Nakamura, Y., Suzuki, R., Shiina, T., Shin-I, T., Fukuta, M., Nguyen, C. T., Nguyen, T. T. T., Nguyen, L. K. H., Hoang, V. M. P., Morita, K., Dang, D. A., Hasebe, F., Le, T. Q. M., & Moi, M. L. (2024). Pre-existing cross-reactive neutralizing activity against SARS-CoV-2 and seasonal coronaviruses prior to the COVID-19 pandemic (2014-2019) with limited immunity against recent emerging SARS-CoV-2 variants, Vietnam. *International Journal of Infectious Diseases*, *139*, 109–117. https://doi.org/10.1016/j.ijid.2023.11.008

Nikoskelainen, J., Kalliomáki, J. L., Lapinleimu, K., Stenvik, M., & Halonen, P. E. (1983). Coxsackie B Virus Antibodies in Myocardial Infarction. *Acta Medica Scandinavica*, *214*(1), 29–32. https://doi.org/10.1111/j.0954-6820.1983.tb08566.x

Nor Isamuddin, N. H., AbuBakar, S., Chin, K.-L., & Zainal, N. (2025). Heterologous immunity and antibody-dependent enhancement in respiratory virus infections. *Molecular Biology Reports*, *53*(1), 27. https://doi.org/10.1007/s11033-025-11189-5

Ooi, C., & Zawar, V. (2011). Hyperaesthesia Following Genital Herpes: A Case Report. *Dermatology Research and Practice*, *2011*, 903595. https://doi.org/10.1155/2011/903595

Ozawa, T., Chubachi, S., Namkoong, H., Nemoto, S., Ikegami, R., Asakura, T., Tanaka, H., Lee, H., Fukushima, T., Azekawa, S., Otake, S., Nakagawara, K., Watase, M., Masaki, K., Kamata, H., Harada, N., Ueda, T., Ueda, S., Ishiguro, T., … Imoto, S. (2025). Predicting coronavirus disease 2019 severity using explainable artificial intelligence techniques. *Scientific Reports*, *15*(1), 9459. https://doi.org/10.1038/s41598-025-85733-5

Palmer, W. H., & Norman, P. J. (2023). The impact of HLA polymorphism on herpesvirus infection and disease. *Immunogenetics*, *75*(3), 231–247. https://doi.org/10.1007/s00251-022-01288-z

Pantazes, R. J., Reifert, J., Bozekowski, J., Ibsen, K. N., Murray, J. A., & Daugherty, P. S. (2016). Identification of disease-specific motifs in the antibody specificity repertoire via next-generation sequencing. *Scientific Reports*, *6*(1), 30312. https://doi.org/10.1038/srep30312

Parker, J. M., Guo, D., & Hodges, R. S. (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*, *25*(19), 5425–5432. https://doi.org/10.1021/bi00367a013

Parksepp, M., Leppik, L., Koch, K., Uppin, K., Kangro, R., Haring, L., Vasar, E., & Zilmer, M. (2020). Metabolomics approach revealed robust changes in amino acid and biogenic amine signatures in patients with schizophrenia in the early course of the disease. *Scientific Reports*, *10*, 13983. https://doi.org/10.1038/s41598-020-71014-w

Parvizpour, S., Pourseif, M. M., Razmara, J., Rafi, M. A., & Omidi, Y. (2020). Epitope-based vaccine design: A comprehensive overview of bioinformatics approaches. *Drug Discovery Today*, *25*(6), 1034–1042. https://doi.org/10.1016/j.drudis.2020.03.006

Paull, M. L., Johnston, T., Ibsen, K. N., Bozekowski, J. D., & Daugherty, P. S. (2019). A general approach for predicting protein epitopes targeted by antibody repertoires using whole proteomes. *PLOS ONE*, *14*(9), e0217668. https://doi.org/10.1371/journal.pone.0217668

Perera, J., Meng, L., Meng, F., & Huang, H. (2013). Autoreactive thymic B cells are efficient antigen-presenting cells of cognate self-antigens for T cell negative selection. *Proceedings of the National Academy of Sciences*, *110*(42), 17011–17016. https://doi.org/10.1073/pnas.1313001110

Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M., & Lundegaard, C. (2009). A generic method for assignment of reliability scores applied to solvent

accessibility predictions. *BMC Structural Biology*, *9*(1), 51. https://doi.org/10.1186/1472-6807-9-51

Pinho, S. S., & Reis, C. A. (2015). Glycosylation in cancer: Mechanisms and clinical implications. *Nature Reviews. Cancer*, *15*(9), 540–555. https://doi.org/10.1038/nrc3982

Pinto, D., Park, Y.-J., Beltramello, M., Walls, A. C., Tortorici, M. A., Bianchi, S., Jaconi, S., Culap, K., Zatta, F., De Marco, A., Peter, A., Guarino, B., Spreafico, R., Cameroni, E., Case, J. B., Chen, R. E., Havenar-Daughton, C., Snell, G., Telenti, A., … Corti, D. (2020). Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature*, *583*(7815), 290–295. https://doi.org/10.1038/s41586-020-2349-y

Pon, R., Marcil, A., Chen, W., Gadoury, C., Williams, D., Chan, K., Zhou, H., Ponce, A., Paquet, E., Gurnani, K., Chattopadhyay, A., & Zou, W. (2020). Masking terminal neo-epitopes of linear peptides through glycosylation favours immune responses towards core epitopes producing parental protein bound antibodies. *Scientific Reports*, *10*(1), 18497. https://doi.org/10.1038/s41598-020-75754-7

Ponomarenko, J., Bui, H.-H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, *9*(1), 514. https://doi.org/10.1186/1471-2105-9-514

Pothast, C. R., Dijkland, R. C., Thaler, M., Hagedoorn, R. S., Kester, M. G., Wouters, A. K., Hiemstra, P. S., van Hemert, M. J., Gras, S., Falkenburg, J. F., & Heemskerk, M. H. (2022). SARS-CoV-2-specific CD4+ and CD8+ T cell responses can originate from cross-reactive CMV-specific T cells. *eLife*, *11*, e82050. https://doi.org/10.7554/eLife.82050

Pupina, N., Avarlaid, A., Sadam, H., Pihlak, A., Jaago, M., Tuvikene, J., Rähni, A., Planken, A., Planken, M., Kalso, E., Tienari, P. J., Nieminen, J. K., Seppänen, M. R. J., Vaheri, A., Lindholm, D., Sinisalo, J., Pussinen, P., Timmusk, T., & Palm, K. (2022). Immune response to a conserved enteroviral epitope of the major capsid VP1 protein is associated with lower risk of cardiovascular disease. *EBioMedicine*, *76*, 103835. https://doi.org/10.1016/j.ebiom.2022.103835

Que, X., Hung, M.-Y., Yeang, C., Gonen, A., Prohaska, T. A., Sun, X., Diehl, C., Määttä, A., Gaddis, D. E., Bowden, K., Pattison, J., MacDonald, J. G., Ylä-Herttuala, S., Mellon, P. L., Hedrick, C. C., Ley, K., Miller, Y. I., Glass, C. K., Peterson, K. L., … Witztum, J. L. (2018). Oxidized Phospholipids are Proinflammatory and Proatherogenic in Hypercholesterolemic Mice. *Nature*, *558*(7709), 301–306. https://doi.org/10.1038/s41586-018-0198-8

R Taylor, W., M Thornton, J., & G Turnell, W. (1983). An ellipsoidal approximation of protein shape. *Journal of Molecular Graphics*, *1*(2), 30–38. https://doi.org/10.1016/0263-7855(83)80001-0

Radzicka, A., & Wolfenden, R. (1988). Comparing the polarities of the amino acids: Side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry.*, *27*(5), 1664–1670.

Raeven, R. H. M., van Riet, E., Meiring, H. D., Metz, B., & Kersten, G. F. A. (2019). Systems vaccinology and big data in the vaccine development chain. *Immunology*, *156*(1), 33–46. https://doi.org/10.1111/imm.13012

Rammensee, H., Bachmann, J., Emmerich, N. P., Bachor, O. A., & Stevanović, S. (1999). SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenetics*, *50*(3–4), 213–219. https://doi.org/10.1007/s002510050595

Ras-Carmona, A., Lehmann, A. A., Lehmann, P. V., & Reche, P. A. (2022). Prediction of B cell epitopes in proteins using a novel sequence similarity-based method. *Scientific Reports*, *12*(1), 13739. https://doi.org/10.1038/s41598-022-18021-1

Reunanen, A., Roivainen, M., Kleemola, M., Saikku, P., Leinonen, M., Hovi, T., Knekt, P., Leino, A., & Aromaa, A. (2002). Enterovirus, mycoplasma and other infections as predictors for myocardial infarction. *Journal of Internal Medicine*, *252*(5), 421–429. https://doi.org/10.1046/j.1365-2796.2002.01052.x

Richer, J., Johnston, S. A., & Stafford, P. (2015). Epitope Identification from Fixed-complexity Random-sequence Peptide Microarrays. *Molecular & Cellular Proteomics*, *14*(1), 136–147. https://doi.org/10.1074/mcp.M114.043513

Robinson, J., Guethlein, L. A., Cereb, N., Yang, S. Y., Norman, P. J., Marsh, S. G. E., & Parham, P. (2017). Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles. *PLoS Genetics*, *13*(6), e1006862. https://doi.org/10.1371/journal.pgen.1006862

Rosa, R. J., Andrade, R. L. de P., Perticarrara Ferezin, L., de Campos, M. C. T., Moura, H. S. D., Berra, T. Z., Ribeiro, N. M., Teibo, T. K. A., Vinci, A. L. T., Mendes Delpino, F., Torres, M. Á. F., & Arcêncio, R. A. (2025). Risk perception of severity or death from COVID-19: A systematic review of the factors associated. *Frontiers in Public Health*, *13*. https://doi.org/10.3389/fpubh.2025.1543629

Rosado, J., Pelleau, S., Cockram, C., Merkling, S. H., Nekkab, N., Demeret, C., Meola, A., Kerneis, S., Terrier, B., Fafi-Kremer, S., de Seze, J., Bruel, T., Dejardin, F., Petres, S., Longley, R., Fontanet, A., Backovic, M., Mueller, I., & White, M. T. (2021). Multiplex assays for the identification of serological signatures of SARS-CoV-2 infection: An antibody-based diagnostic and machine learning study. *The Lancet. Microbe*, *2*(2), e60–e69. https://doi.org/10.1016/S2666-5247(20)30197-X

Routh, E. D., Woodcock, M. G., Beckabir, W., Vensko, S. P., Serody, J. S., & Vincent, B. G. (2023). Evaluation of tumor antigen-specific antibody responses in patients with metastatic triple negative breast cancer treated with cyclophosphamide and pembrolizumab. *Journal for Immunotherapy of Cancer*, *11*(3), e005848. https://doi.org/10.1136/jitc-2022-005848

Sadam, H., Maruste, R., Rähni, A., Toots, M., Piirsalu, M., Pihlas, L., Ausman, H., Sirp, A., Pihlak, A., Laasfeld, T., Rull, K., Salumets, A., Antson, A., Tillmann, V., Aleksejeva, A., Rööp, T., Štšepetova, J., Sepp, E., Mändar, R., & Palm, K. (2026). Maternal antibodies shape infant immune response development in an epitope-specific manner. *eBioMedicine*, *123*, 106056. https://doi.org/10.1016/j.ebiom.2025.106056

Sadam, H., Pihlak, A., Jaago, M., Pupina, N., Rähni, A., Toots, M., Vaheri, A., Nieminen, J. K., Siuko, M., Tienari, P. J., & Palm, K. (2021). Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients. *EBioMedicine*, *64*, 103211. https://doi.org/10.1016/j.ebiom.2021.103211

Sadam, H., Pihlak, A., Kivil, A., Pihelgas, S., Jaago, M., Adler, P., Vilo, J., Vapalahti, O., Neuman, T., Lindholm, D., Partinen, M., Vaheri, A., & Palm, K. (2018). Prostaglandin D2 Receptor DP1 Antibodies Predict Vaccine-induced and Spontaneous Narcolepsy Type 1: Large-scale Study of Antibody Profiling. *EBioMedicine*, *29*, 47–59. https://doi.org/10.1016/j.ebiom.2018.01.043

Sage, A. P., Tsiantoulas, D., Binder, C. J., & Mallat, Z. (2019). The role of B cells in atherosclerosis. *Nature Reviews Cardiology*, *16*(3), 180–196. https://doi.org/10.1038/s41569-018-0106-9

Saha, S., Bhasin, M., & Raghava, G. P. (2005). Bcipep: A database of B-cell epitopes. *BMC Genomics*, *6*(1), 79. https://doi.org/10.1186/1471-2164-6-79

Sainath Rao, S., Mohan, K. V. K., Nguyen, N., Abraham, B., Abdouleva, G., Zhang, P., & Atreya, C. D. (2010). Peptides panned from a phage-displayed random peptide library are useful for the detection of *Bacillus anthracis* surrogates *B. cereus* 4342 and *B. anthracis* Sterne. *Biochemical and Biophysical Research Communications*, *395*(1), 93–98. https://doi.org/10.1016/j.bbrc.2010.03.145

Sanchez-Trincado, J. L., Gomez-Perosanz, M., & Reche, P. A. (2017). Fundamentals and Methods for T- and B-Cell Epitope Prediction. *Journal of Immunology Research*, *2017*(1), 2680160. https://doi.org/10.1155/2017/2680160

Schwarz, T., Heiss, K., Mahendran, Y., Casilag, F., Kurth, F., Sander, L. E., Wendtner, C.-M., Hoechstetter, M. A., Müller, M. A., Sekul, R., Drosten, C., Stadler, V., & Corman, V. M. (2021). SARS-CoV-2 Proteome-Wide Analysis Revealed Significant Epitope Signatures in COVID-19 Patients. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.629185

Selva, K. J., van de Sandt, C. E., Lemke, M. M., Lee, C. Y., Shoffner, S. K., Chua, B. Y., Davis, S. K., Nguyen, T. H. O., Rowntree, L. C., Hensen, L., Koutsakos, M., Wong, C. Y., Mordant, F., Jackson, D. C., Flanagan, K. L., Crowe, J., Tosif, S., Neeland, M. R., Sutton, P., … Chung, A. W. (2021). Systems serology detects functionally distinct coronavirus antibody features in children and elderly. *Nature Communications*, *12*(1), 2037. https://doi.org/10.1038/s41467-021-22236-7

Sender, R., Weiss, Y., Navon, Y., Milo, I., Azulay, N., Keren, L., Fuchs, S., Ben-Zvi, D., Noor, E., & Milo, R. (2023). The total mass, number, and distribution of immune cells in the human body. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(44), e2308511120. https://doi.org/10.1073/pnas.2308511120

Seo, K., & Choi, J. K. (2025). Comprehensive Analysis of TCR and BCR Repertoires: Insights into Methodologies, Challenges, and Applications. *Genomics & Informatics*, *23*(1), 6. https://doi.org/10.1186/s44342-024-00034-z

Shankar, G., Arkin, S., Cocea, L., Devanarayan, V., Kirshner, S., Kromminga, A., Quarmby, V., Richards, S., Schneider, C. K., Subramanyam, M., Swanson, S., Verthelyi, D., & Yim, S. (2014). Assessment and Reporting of the Clinical Immunogenicity of Therapeutic Proteins and Peptides—Harmonized Terminology and Tactical Recommendations. *The AAPS Journal*, *16*(4), 658–673. https://doi.org/10.1208/s12248-014-9599-2

Sharma, P., Goswami, S., Raychaudhuri, D., Siddiqui, B. A., Singh, P., Nagarajan, A., Liu, J., Subudhi, S. K., Poon, C., Gant, K. L., Herbrich, S. M., Anandhan, S., Islam, S., Amit, M., Anandappa, G., & Allison, J. P. (2023). Immune checkpoint therapy—Current perspectives and future directions. *Cell*, *186*(8), 1652–1669. https://doi.org/10.1016/j.cell.2023.03.006

Shields, L. B. E., & Alsorogi, M. S. (2019). Herpes Simplex Virus Type 2 Radiculomyelitis Disguised as Conversion Disorder. *Case Reports in Neurology*, *11*(1), 117–123. https://doi.org/10.1159/000499701

Shrock, E., Fujimura, E., Kula, T., Timms, R. T., Lee, I.-H., Leng, Y., Robinson, M. L., Sie, B. M., Li, M. Z., Chen, Y., Logue, J., Zuiani, A., McCulloch, D., Lelis, F. J. N., Henson,

S., Monaco, D. R., Travers, M., Habibi, S., Clarke, W. A., … Elledge, S. J. (2020). Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science*, *370*(6520), eabd4250. https://doi.org/10.1126/science.abd4250

Shrock, E. L., Timms, R. T., Kula, T., Mena, E. L., West, A. P., Guo, R., Lee, I.-H., Cohen, A. A., McKay, L. G. A., Bi, C., Keerti, Leng, Y., Fujimura, E., Horns, F., Li, M., Wesemann, D. R., Griffiths, A., Gewurz, B. E., Bjorkman, P. J., & Elledge, S. J. (2023). Germline-encoded amino acid–binding motifs drive immunodominant public antibody responses. *Science*, *380*(6640), eadc9498. https://doi.org/10.1126/science.adc9498

Siddiqui, M. A., Usmani, A., Ansari, M. N., & Almoselhy, R. I. M. (2025). Immune-related adverse events in immunotherapy: Challenges in diagnosis, monitoring, and management. *Toxicology Reports*, *14*, 102036. https://doi.org/10.1016/j.toxrep.2025.102036

Sikora, M., Bülow, S. von, Blanc, F. E. C., Gecht, M., Covino, R., & Hummer, G. (2021). Computational epitope map of SARS-CoV-2 spike protein. *PLOS Computational Biology*, *17*(4), e1008790. https://doi.org/10.1371/journal.pcbi.1008790

Song, L., Park, J. G., Qiu, J., Murugan, V., Mousavi, S., Hou, C.-W., Magee, D. M., Chen, Z., Chung, Y., Williams, S., Adam, D., Guo, B., McDaniel, C., Trendler, T., Rice, G., Connard, K., Bobbett, B., Ritchie, M., Garcia, I., … LaBaer, J. (2026). Pre-vaccine sentinel antibodies predict blunted vaccine responses. *Cell Press Blue*, 100088. https://doi.org/10.1016/j.cpblue.2026.100088

Stern, L. J., Clement, C., Galluzzi, L., & Santambrogio, L. (2024). Non-mutational neoantigens in disease. *Nature Immunology*, *25*(1), 29–40. https://doi.org/10.1038/s41590-023-01664-1

Strzelec, M., Detka, J., Mieszczak, P., Sobocińska, M. K., & Majka, M. (2023). Immunomodulation—A general review of the current state-of-the-art and new therapeutic strategies for targeting the immune system. *Frontiers in Immunology*, *14*, 1127704. https://doi.org/10.3389/fimmu.2023.1127704

Swanson, S. J. (2024). What are clinically significant anti-drug antibodies and why is it important to identify them. *Frontiers in Immunology*, *15*. https://doi.org/10.3389/fimmu.2024.1401178

Sweet, R. M., & Eisenberg, D. (1983). Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. *Journal of Molecular Biology*, *171*(4), 479–488. https://doi.org/10.1016/0022-2836(83)90041-4

Taams, L. S., & Wauben, M. H. M. (2000). Anergic T cells as active regulators of the immune response. *Human Immunology*, *61*(7), 633–639. https://doi.org/10.1016/S0198-8859(00)00127-0

Tang, X. L., Tregear, G. W., White, D. O., & Jackson, D. C. (1988). Minimum requirements for immunogenic and antigenic activities of homologs of a synthetic peptide of influenza virus hemagglutinin. *Journal of Virology*, *62*(12), 4745–4751. https://doi.org/10.1128/JVI.62.12.4745-4751.1988

Tariq, F., Charavanmuttu, Y., Labiba, K., & Wincup, C. (2026). Aberrant B cell responses as drivers of autoantibody generation and epitope diversification in SLE pathogenesis. *Frontiers in Immunology*, *16*. https://doi.org/10.3389/fimmu.2025.1731285

Teraguchi, S., Saputri, D. S., Llamas-Covarrubias, M. A., Davila, A., Diez, D., Nazlica, S. A., Rozewicki, J., Ismanto, H. S., Wilamowski, J., Xie, J., Xu, Z., Loza-Lopez, M. de J.,

van Eerden, F. J., Li, S., & Standley, D. M. (2020). Methods for sequence and structural analysis of B and T cell receptor repertoires. *Computational and Structural Biotechnology Journal*, *18*, 2000–2011. https://doi.org/10.1016/j.csbj.2020.07.008

Thalén, N. B., Karlander, M., Lundqvist, M., Persson, H., Hofström, C., Turunen, S. P., Godzwon, M., Volk, A.-L., Malm, M., Ohlin, M., & Rockberg, J. (2024). Mammalian cell display with automated oligo design and library assembly allows for rapid residue level conformational epitope mapping. *Communications Biology*, *7*(1), 805. https://doi.org/10.1038/s42003-024-06508-8

The UniProt Consortium. (2025). UniProt: The Universal Protein Knowledgebase in 2025. *Nucleic Acids Research*, *53*(D1), D609–D617. https://doi.org/10.1093/nar/gkae1010

Thjodleifsson, B., Asbjörnsdottir, H., Björg Sigurjonsdottir, R., Gíslason, D., Olafsson, Í., Cook, E., Gíslason, T., Jogi, R., & Janson, C. (2007). Seroprevalence of Helicobacter pylori and cagA antibodies in Iceland, Estonia and Sweden. *Scandinavian Journal of Infectious Diseases*, *39*(8), 683–689. https://doi.org/10.1080/00365540701225736

Thomas, M., Calamito, M., Srivastava, B., Maillard, I., Pear, W. S., & Allman, D. (2006). Notch activity synergizes with B-cell–receptor and CD40 signaling to enhance B-cell activation. *Blood*, *109*(8), 3342–3350. https://doi.org/10.1182/blood-2006-09-046698

Thornton, J. M., Edwards, M. S., Taylor, W. R., & Barlow, D. J. (1986). Location of 'continuous' antigenic determinants in the protruding regions of proteins. *The EMBO Journal*, *5*(2), 409–413. https://doi.org/10.1002/j.1460-2075.1986.tb04226.x

Toi, Y., Sugawara, S., Sugisaka, J., Ono, H., Kawashima, Y., Aiba, T., Kawana, S., Saito, R., Aso, M., Tsurumi, K., Suzuki, K., Shimizu, H., Domeki, Y., Terayama, K., Nakamura, A., Yamanda, S., Kimura, Y., & Honda, Y. (2019). Profiling Preexisting Antibodies in Patients Treated With Anti–PD-1 Therapy for Advanced Non–Small Cell Lung Cancer. *JAMA Oncology*, *5*(3), 376–383. https://doi.org/10.1001/jamaoncol.2018.5860

Tokarz, R., Mishra, N., Tagliafierro, T., Sameroff, S., Caciula, A., Chauhan, L., Patel, J., Sullivan, E., Gucwa, A., Fallon, B., Golightly, M., Molins, C., Schriefer, M., Marques, A., Briese, T., & Lipkin, W. I. (2018). A multiplex serologic platform for diagnosis of tick-borne diseases. *Scientific Reports*, *8*(1), 3158. https://doi.org/10.1038/s41598-018-21349-2

Tortorici, M. A., Addetia, A., Seo, A. J., Brown, J., Sprouse, K., Logue, J., Clark, E., Franko, N., Chu, H., & Veesler, D. (2024). Persistent immune imprinting occurs after vaccination with the COVID-19 XBB.1.5 mRNA booster in humans. *Immunity*, *57*(4), 904-911.e4. https://doi.org/10.1016/j.immuni.2024.02.016

van den Elsen, P. J., Holling, T. M., Kuipers, H. F., & van der Stoep, N. (2004). Transcriptional regulation of antigen presentation. *Current Opinion in Immunology*, *16*(1), 67–75. https://doi.org/10.1016/j.coi.2003.11.015

Van Regenmortel, M. H. V. (1996). Mapping Epitope Structure and Activity: From One-Dimensional Prediction to Four-Dimensional Description of Antigenic Specificity. *Methods*, *9*(3), 465–472. https://doi.org/10.1006/meth.1996.0054

Van Regenmortel, M. H. V. (2009). What Is a B-Cell Epitope? In M. Schutkowski & U. Reineke (Eds), *Epitope Mapping Protocols: Second Edition* (pp. 3–20). Humana Press. https://doi.org/10.1007/978-1-59745-450-6_1

Vidarsson, G., Dekkers, G., & Rispens, T. (2014). IgG Subclasses and Allotypes: From Structure to Effector Functions. *Frontiers in Immunology*, *5*, 520. https://doi.org/10.3389/fimmu.2014.00520

Villanueva-Flores, F., Sanchez-Villamil, J. I., & Garcia-Atutxa, I. (2025). AI-driven epitope prediction: A systematic review, comparative analysis, and practical guide for vaccine development. *Npj Vaccines*, *10*(1), 207. https://doi.org/10.1038/s41541-025-01258-y

Vita, R., Blazeska, N., Marrama, D., Duesing, S., Bennett, J., Greenbaum, J., De Almeida Mendes, M., Mahita, J., Wheeler, D. K., Cantrell, J. R., Overton, J. A., Natale, D. A., Sette, A., & Peters, B. (2024). The Immune Epitope Database (IEDB): 2024 update. *Nucleic Acids Research*, *53*(D1), D436–D443. https://doi.org/10.1093/nar/gkae1092

Vogl, T., Klompus, S., Leviatan, S., Kalka, I. N., Weinberger, A., Wijmenga, C., Fu, J., Zhernakova, A., Weersma, R. K., & Segal, E. (2021). Population-wide diversity and stability of serum antibody epitope repertoires against human microbiota. *Nature Medicine*, *27*(8), 1442–1450. https://doi.org/10.1038/s41591-021-01409-3

Wang, R., Lan, C., Benlagha, K., Camara, N. O. S., Miller, H., Kubo, M., Heegaard, S., Lee, P., Yang, L., Forsman, H., Li, X., Zhai, Z., & Liu, C. (2024). The interaction of innate immune and adaptive immune system. *MedComm*, *5*(10), e714. https://doi.org/10.1002/mco2.714

Wang, S., Qin, J., Ye, H., Wang, K., Shi, J., Ma, Y., Duan, Y., Song, C., Wang, X., Dai, L., Wang, K., Wang, P., & Zhang, J. (2018). Using a panel of multiple tumor-associated antigens to enhance autoantibody detection for immunodiagnosis of gastric cancer. *OncoImmunology*, *7*(8), e1452582. https://doi.org/10.1080/2162402X.2018.1452582

Watson, C. P. N. (2010). Herpes zoster and postherpetic neuralgia. *CMAJ*, *182*(16), 1713–1714. https://doi.org/10.1503/cmaj.101409

Weber, J. S., Carlino, M. S., Khattak, A., Meniawy, T., Ansstas, G., Taylor, M. H., Kim, K. B., McKean, M., Long, G. V., Sullivan, R. J., Faries, M., Tran, T. T., Cowey, C. L., Pecora, A., Shaheen, M., Segar, J., Medina, T., Atkinson, V., Gibney, G. T., … Zaks, T. (2024). Individualised neoantigen therapy mRNA-4157 (V940) plus pembrolizumab versus pembrolizumab monotherapy in resected melanoma (KEYNOTE-942): A randomised, phase 2b study. *The Lancet*, *403*(10427), 632–644. https://doi.org/10.1016/S0140-6736(23)02268-7

Wintjens, R., Bifani, A. M., & Bifani, P. (2020). Impact of glycan cloud on the B-cell epitope prediction of SARS-CoV-2 Spike protein. *Npj Vaccines*, *5*(1), 81. https://doi.org/10.1038/s41541-020-00237-9

Wrynla, X. H., Bates, T. A., Trank-Greene, M., Wahedi, M., Hinchliff, A., Curlin, M. E., & Tafesse, F. G. (2025). Immune imprinting and vaccine interval determine antibody responses to monovalent XBB.1.5 COVID-19 vaccination. *Communications Medicine*, *5*(1), 182. https://doi.org/10.1038/s43856-025-00898-4

Wu, F.-L., Lai, D.-Y., Ding, H.-H., Tang, Y.-J., Xu, Z.-W., Ma, M.-L., Guo, S.-J., Wang, J.-F., Shen, N., Zhao, X.-D., Qi, H., Li, H., & Tao, S.-C. (2019). Identification of Serum

Biomarkers for Systemic Lupus Erythematosus Using a Library of Phage Displayed Random Peptides and Deep Sequencing*[S]. *Molecular & Cellular Proteomics*, *18*(9), 1851–1863. https://doi.org/10.1074/mcp.RA119.001582

Xu, G. J., Kula, T., Xu, Q., Li, M. Z., Vernon, S. D., Ndung'u, T., Ruxrungtham, K., Sanchez, J., Brander, C., Chung, R. T., O'Connor, K. C., Walker, B., Larman, H. B., & Elledge, S. J. (2015). Comprehensive serological profiling of human populations using a synthetic human virome. *Science*, *348*(6239), aaa0698. https://doi.org/10.1126/science.aaa0698

Xu, H., Palpant, T., Wang, Q., & Shaw, D. E. (2025). Design of immunogens to present a tumor-specific cryptic epitope. *Scientific Reports*, *15*(1), 11322. https://doi.org/10.1038/s41598-025-94295-5

Ye, S., Lu, C., Qiu, Y., Zheng, H., Ge, X., Wu, A., Xia, Z., Jiang, T., Zhu, H., & Peng, Y. (2022). An atlas of human viruses provides new insights into diversity and tissue tropism of human viruses. *Bioinformatics*, *38*(11), 3087–3093. https://doi.org/10.1093/bioinformatics/btac275

Yip, Y. L., Smith, G., Koch, J., Dübel, S., & Ward, R. L. (2001). Identification of epitope regions recognized by tumor inhibitory and stimulatory anti-ErbB-2 monoclonal antibodies: Implications for vaccine design. *Journal of Immunology (Baltimore, Md.: 1950)*, *166*(8), 5271–5278. https://doi.org/10.4049/jimmunol.166.8.5271

Yuan, M., Wu, N. C., Zhu, X., Lee, C.-C. D., So, R. T. Y., Lv, H., Mok, C. K. P., & Wilson, I. A. (2020). A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science (New York, N.Y.)*, *368*(6491), 630–633. https://doi.org/10.1126/science.abb7269

Zaidi, N., Jaffee, E. M., & Yarchoan, M. (2025). Recent advances in therapeutic cancer vaccines. *Nature Reviews Cancer*, *25*(7), 517–533. https://doi.org/10.1038/s41568-025-00820-z

Zamecnik, C. R., Sowa, G. M., Abdelhak, A., Dandekar, R., Bair, R. D., Wade, K. J., Bartley, C. M., Kizer, K., Augusto, D. G., Tubati, A., Gomez, R., Fouassier, C., Gerungan, C., Caspar, C. M., Alexander, J., Wapniarski, A. E., Loudermilk, R. P., Eggers, E. L., Zorn, K. C., … Wilson, M. R. (2024). An autoantibody signature predictive for multiple sclerosis. *Nature Medicine*, *30*(5), 1300–1308. https://doi.org/10.1038/s41591-024-02938-3

Zandian, A., Forsström, B., Häggmark-Månberg, A., Schwenk, J. M., Uhlén, M., Nilsson, P., & Ayoglu, B. (2017). Whole-Proteome Peptide Microarrays for Profiling Autoantibody Repertoires within Multiple Sclerosis and Narcolepsy. *Journal of Proteome Research*, *16*(3), 1300–1314. https://doi.org/10.1021/acs.jproteome.6b00916

Zaslavsky, M. E., Craig, E., Michuda, J. K., Sehgal, N., Ram-Mohan, N., Lee, J.-Y., Nguyen, K. D., Hoh, R. A., Pham, T. D., Röltgen, K., Lam, B., Parsons, E. S., Macwana, S. R., DeJager, W., Drapeau, E. M., Roskin, K. M., Cunningham-Rundles, C., Moody, M. A., Haynes, B. F., … Boyd, S. D. (2025). Disease diagnostics using machine learning of B cell and T cell receptor sequences. *Science*, *387*(6736), eadp2407. https://doi.org/10.1126/science.adp2407

Zefferino, R., Di Gioia, S., & Conese, M. (2020). Molecular links between endocrine, nervous and immune system during chronic stress. *Brain and Behavior*, *11*(2), e01960. https://doi.org/10.1002/brb3.1960

Zhang, Z., Gool, J. K., Fronczek, R., Dauvilliers, Y., Bassetti, C. L. A., Mayer, G., Plazzi, G., Pizza, F., Santamaria, J., Partinen, M., Overeem, S., Peraita-Adrados, R., da Silva,

A. M., Sonka, K., del Rio-Villegas, R., Heinzer, R., Wierzbicka, A., Young, P., Högl, B., … Khatami, R. (2021). New 2013 incidence peak in childhood narcolepsy: More than vaccination? *Sleep*, *44*(2), zsaa172. https://doi.org/10.1093/sleep/zsaa172

Zhang, Z., Lu, M., Qin, Y., Gao, W., Tao, L., Su, W., & Zhong, J. (2021). Neoantigen: A New Breakthrough in Tumor Immunotherapy. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.672356

## Acknowledgements

First and foremost, I would like to thank my thesis supervisors Dr. Kaia Palm and Dr. Tõnis Timmusk. I am grateful for the opportunity to work with data from a wide range of human diseases, expand my bioinformatics skill set, and see the world during my studies – all made possible by the care, support and guidance of Kaia. I am also thankful to Tõnis for initiating my research journey at the Neurolab all those years ago, and for where that journey has led me today.

This thesis outlined many years of work that has been conducted by the amazing Protobios team. It would not have been possible without the contributions of each special and wonderful team member who are all deserving of a big and heartfelt "Thank you!". A round of applause for: Maarja Toots, Arno Pihlak, Regina Maruste, Maria Piirsalu, Alex Sirp, Loviisa Pihlas, Epp Väli, Helen Ausman, Tõnis Laasfeld and Luiza Tolokonnikova. A special thank you to past team members who all had an impact and left a mark on the research journey of this author – thank you Helle Sadam, Mariliis Uusväli, Nadežda Pupina and Jürgen Tuvikene!

Last, but certainly not least – I am grateful to my family and friends for never-ending support and endless hours of dog-sitting, so that this thesis could see the light of day.

---

The research presented in this thesis was conducted at Protobios LLC and supported by funding resources of European Union projects DTRIP4H (No. 101188432), SZTEST (No. 734791), ERA-NET-NEURON (No. 964215), LongCovid (No. 101057553), OPADE (No. 101095436), by Estonian Research Council grants PRG573 and PRG1953, by the Estonian Ministry of Economic Affairs and Communications (8-1/79-1) and the Stipend of Tamkivi Foundation of Natural Sciences (*anno* 2022) issued by the Estonian National Culture Foundation.

## Abstract


## Epitope-level Antibody Profiling for the Development of Diagnostics and Therapeutics Using Next Generation Phage Display

Immune system protects the organism from pathogens, environmental allergens and even pathological cellular processes that might lead to cancer. Previous antigenic exposure can shape the immune response to unrelated antigens of microbial or self-protein origin. Cross-reactive immune memory both decodes disease-associated immune patterns and confers protection from new pathogens and diseases. In this thesis mimotope variation analysis, a high throughput phage display technology, was used to investigate the antibody response patterns elicited in response to personalised immunotherapy treatments, the role of heterologous immunity via antibody reactivities in different disease backgrounds and described antibody response patterns characteristic of different chronic and infectious diseases that could potentially be used as predictive or diagnostic biomarkers for health and disease. Based on a combined study cohort of 938 individuals, we showed that epitope-specific antibody patterns differ between healthy and diseased populations but remain individually stable over time and reflect the influence of immune history and heterologous immunity. Furthermore, epitope-specific and individualistic antibody responses harboured markers reflecting immunotherapy responses, cardiovascular disease risk and susceptibility to post-surgical neuropathic pain. Overall, integration of high-throughput epitope-specific antibody response profiling with computational antigen-mapping methods across diverse disease cohorts as large-scale immune biomarker discovery, supports advances in the development of personalised diagnostics and therapies.

# Lühikokkuvõte

## Antikehaepitoopide kaardistamine diagnostika ja teraapia arenduseks järgmise põlvkonna faagikuvamise meetodiga

Inimese immuunsüsteemil on oluline roll organismi kaitsmisel seda potentsiaalselt kahjustavate tegurite eest, mille hulka kuuluvad muuhulgas erinevad haigustekitajad (sh viirused, bakterid), keskkonna allergeenid ja isegi halvaloomulised rakulised protsessid, mis võivad viia vähkkasvajate tekkeni. Iga kokkupuude eelmainitud teguritega vermib immuunreaktsiooni, mis võib hiljem rist-reaktiivsuse protsessi käigus ära tunda sarnase molekulaarse mustriga mikroobset või inimvalgulist päritolu antigeeni. Rist-reaktiivsed immuunmälu komponendid peidavad endas haigusseoselisi immuun-mustreid ja aitavad kaitsta uute haigustekitajate ja haiguste tekke eest.

Käesolevas töös kasutati suure läbilaskvusega faagi-kuvamis-tehnoloogiat nimega mimotoop variatsioon analüüsi (MVA) kirjeldamaks spetsiifilisi antikehaepitoopide mustreid personaliseeritud immuunteraapiat saanud vähipatsientides, uurimaks heteroloogse immuunsuse rolli erinevates nakkus- ja kroonilistes haigustes ning kirjeldamaks haigusspetsiifilisi antikehaepitoopide immuunreaktsioone, mis oleksid kasutatavad potentsiaalsete biomarkeritena. Selle töö ühendatud uuringukohorti kuulus 938 erinevat inimest (kokku 1068 prooviga) ning töö peamised tulemused on järgnevad:

1. MVA järgmise põlvkonna profileerimismeetod võimaldab süstemaatiliselt kaardistada haigusseoselisi antikehaepitoopide immuunmustreid ning toetab sellega immuunbiomarkerite avastamist.
2. Epitoobi-spetsiifilised antikehamustrid erinevad tervete ja haigete populatsioonide vahel, kuid püsivad üksikisiku tasandil ajas stabiilsed ning peegeldavad sellega immuunmälu ja heteroloogse immuunsuse mõju.
3. Spetsiifilised antikehaepitoop-mustrid on seotud immuunteraapia eduka toimimisega, südame- ja veresoonkonna haiguste ning operatsioonijärgse neuropaatilise valu tekkeriskiga, ning esindavad seega võimalikke biomarkereid.
4. MVA koos arvutuslike antigeeni profileerimiskatsetega võimaldab suures mahus epitoopide annoteerimist ja kandidaat-antigeenide identifitseerimist, mis omakorda toetab personaliseeritud diagnostikameetodite ja teraapiate väljatöötamist.

# Appendix 1

**Publication I**

**Annika Rähni**, Jaago, M., Sadam, H., Pupina, N., Pihlak, A., Tuvikene, J., Annuk, M., Mägi, A., Timmusk, T., Ghaemmaghami, A. M., & Palm, K. (2022). **Melanoma-specific antigen-associated antitumor antibody reactivity as an immune-related biomarker for targeted immunotherapies.** *Communications Medicine*, 2, 48. https://doi.org/10.1038/s43856-022-00114-7

communications medicine

ARTICLE

https://doi.org/10.1038/s43856-022-00114-7 OPEN

# Melanoma-specific antigen-associated antitumor antibody reactivity as an immune-related biomarker for targeted immunotherapies

Annika Rähni[1,2], Mariliis Jaago[1,2], Helle Sadam[1,2], Nadežda Pupina[1], Arno Pihlak[1], Jürgen Tuvikene[1,2,3], Margus Annuk[4], Andrus Mägi[5], Tõnis Timmusk[1,2], Amir M. Ghaemmaghami[6] & Kaia Palm[1,2]✉

## Abstract

**Background:** Immunotherapies, including cancer vaccines and immune checkpoint inhibitors have transformed the management of many cancers. However, a large number of patients show resistance to these immunotherapies and current research has provided limited findings for predicting response to precision immunotherapy treatments.

**Methods:** Here, we applied the next generation phage display mimotope variation analysis (MVA) to profile antibody response and dissect the role of humoral immunity in targeted cancer therapies, namely anti-tumor dendritic cell vaccine (MelCancerVac®) and immunotherapy with anti-PD-1 monoclonal antibodies (pembrolizumab).

**Results:** Analysis of the antibody immune response led to the characterization of epitopes that were linked to melanoma-associated and cancer-testis antigens (CTA) whose antibody response was induced upon MelCancerVac® treatments of lung cancer. Several of these epitopes aligned to antigens with strong immune response in patients with unresectable metastatic melanoma receiving anti-PD-1 therapy.

**Conclusions:** This study provides insights into the differences and similarities in tumor-specific immunogenicity related to targeted immune treatments. The antibody epitopes as biomarkers reflect melanoma-associated features of immune response, and also provide insights into the molecular pathways contributing to the pathogenesis of cancer. Concluding, antibody epitope response can be useful in predicting anti-cancer immunity elicited by immunotherapy.

## Plain language summary

Immunotherapy treatments, which utilize the patient's own immune system to fight cancer, have become a standard treatment of cancer. However, for many patients' immunotherapy does not work. During the immune response the body produces proteins called antibodies. This study characterized the antibodies produced following treatment with two different types of immunotherapies that treat skin cancer, to gain insights into how the immune system responds in different individuals. Our results demonstrate that multiple proteins that are present in patients with skin cancer are specifically targeted by the immune system during skin cancer specific immunotherapy. Our results should help further anti-cancer drug development.

[1] Protobios Llc, Tallinn, Estonia. [2] Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia. [3] dxlabs Llc, Tallinn, Estonia. [4] EGeen International Inc., Mountain View, CA, USA. [5] Tartu University Hospital, Tartu, Estonia. [6] Immunology and Immuno-Bioengineering Group, School of Life Science, Faculty of Medicine and Health Sciences, University of Nottingham, Nottingham, UK. ✉email: kaia@protobios.com

Knowledge of the immunosuppressive tumor microenvironment has markedly improved within the last decade (reviewed in ref. [1]). To achieve immunogenicity, tumor cells must express antigens capable of eliciting immune activation. The identification of applicable tumor antigens is indispensable for the development of effective cancer immunotherapy. Most known tumor antigens are considered canonical if derived from protein-coding regions in contrast to noncanonical antigens that include sequences outside protein-coding regions or that are generated by antigen-processing[2]. Melanoma cells are considered highly immunogenic with well-described tumor-associated antigens (TAAs)[3], including cancer-testis antigens (CTAs)[4] and neoantigens carrying novel epitopes of self-antigens[5]. Some well-known examples include carcinoembryonic antigen (CEA), B melanoma antigen 1 (BAGE), G antigens (GAGEs), cancer/testis antigen 1 (CTAG1; also known as NY-ESO1), and melanoma-associated antigens (MAGEs) (Rev in ref. [6]). The antigenic repertoire is a critical factor for immunosurveillance and cancer progression[7]. However, most studies have focused on the role of T cells in these battles[8], while considerably less is known about B cell response[9]. Humoral response against cross-reactive autoantigens has been detected in different cancers[10]. A burst of recent publications is pointing to the role of antibodies contributing to tumor control[11] as cancer-associated autoimmunity targeting non-malignant tissues may reflect favorable disease outcome[12]. On the other hand, the reasons underlying the immunogenicity of the tumor, or the lack of it, are not well understood[13]. The antitumor immunity can result from many factors including MHC genetic variation, tumor mutational load, tissue microenvironment[13], but also by cell stress, reactivation of embryonic or gonadal transcription, epigenetic instability, aberrant RNA splicing, and others[14,15]. For example, it is argued that the capture of either apoptotic or necrotic cancer cells by macrophages and dendritic cells in the tumor microenvironment may lead to immune suppression or stimulate inflammatory pathways contributing to antitumor cytotoxicity[16].

Discoveries in cancer biology have led to new strategies in awakening tumor immunogenicity, including checkpoint blockade, adoptive cellular therapy, and cancer vaccines, underscoring the role of the immune system in waging the war on cancer tissue. Among these are monoclonal antibodies that target cancer immune checkpoint inhibitors (ICIs) including anti-CTLA-4, anti-PD-1, and anti-PD-L1/2 antibodies that are able to restore anticancer immunity and are widely used for the management of various cancers, including melanoma[17]. Immunogenicity of CTAs has led to the use of melanoma-associated antigens as promising candidates for novel cancer treatments[18,19]. In addition to monoclonal antibodies, cancer vaccines, in particular those based on dendritic cells (DCs) as vectors for antigen delivery, are a major focus of current developments[20]. To date, personalized neoantigen-based DC vaccines are evolving and have shown clinical success in melanoma and other solid tumors[21].

Biomarkers associated with clinical prognosis of the cancer and/or severe immune-related adverse effects (irAEs) of the drugs are areas of active investigation. Different biomarkers have been tackled with variable success, such as levels of PD-L1[22], genetic mutations[23], inflammatory cytokines[24], and the presence of tumor-infiltrating lymphocytes (reviewed in ref. [25]). Tumor infiltrating B lymphocytes contribute to anti-tumor immunity by promoting antibody response to tumor antigens[26,27]. High titer antibodies against melanoma differentiation antigens (TRP1/TYRP1, TRP2/TYRP2, gp100, MelanA/MART1) were observed in responder group of melanoma patients treated with ICI mAbs (monotherapies with Nivolumab, Pembrolizumab or Ipilimumab, or the combination of Nivolumab and Ipilimumab)[28,29]. However, pre-treatment autoantibody profiles in melanoma patients were reported to predict ICI treatment-associated toxicity[30]. Connectedly, DC vaccines also stimulate robust antibody response[31–33] and in some cases, this is associated with prolonged recurrence-free survival[32]. Despite big hopes, clinical benefit of immunotherapies has remained limited only to a subset of patients[34,35] and it is currently undetermined whether increase or decrease in immune response to specific tumor antigens is beneficial to the patient[36,37].

Here, we explore the use of a high precision approach called mimotope variation analysis (MVA), a next generation random peptide phage display method to delineate cancer therapy-associated antibody immune response at epitope resolution. We hypothesize that the pre-existing and treatment-induced antibodies against specific antigen targets could reflect the response elicited by anti-tumor drug and that this response could be predictive of cancer immunogenicity and thus, sensitivity to immune therapy. We generate data to test this hypothesis by immunoprofiling analysis of the anti-melanoma antibody response in the sera samples from the phase II clinical trial of patients with non-small cell lung cancer (NSCLC) receiving autologous DC therapy based on allogenic melanoma cell lysate (MelCancerVac®)[38,39]. We correlate the findings on melanoma-specific antigen profiles with those from a group of patients with unresectable metastatic melanoma receiving anti-PD1 (pembrolizumab) treatment as a part of their standard-of-care. We verify the melanoma-antigen specificity using MVA-based competition, and further determine a three-epitope biomarker signature of melanoma-specific antibody response elicited by both immunotherapies. Our results demonstrate the relevance of antibody epitope profiling to better understand the fine line separating beneficial immunosurveillance from harmful autoimmunity in the anticancer immune response elicited by different types of therapy.

## Methods

**Study population**. The present study analyzed samples from a total of 119 individuals from 2 different clinical cohorts of NSCLC and melanoma patients and their appropriate controls, whose clinical characteristics are shown in Table 1 and Supplementary Table 1. The study was conducted in accordance with the guiding principles of the Declaration of Helsinki and the study participants gave informed consent before enrollment.

The NSCLC patient cohort ($n = 24$) included longitudinal study of patients diagnosed with advanced NSCLC, who participated in the phase II clinical trial evaluating the effectiveness of MelCancerVac® vaccine[38,39] (Supplementary Table 2). The clinical trial, completed at the time of this study, was designed and carried out by Dandrit Biotech A/S and approved by European Medicines Agency (https://www.clinicaltrialsregister.eu/ctr-search/trial/2006-002202-54/DK). Out of the 24 study participants, 6 NSCLC patients donated blood samples before vaccination (group: *MelVac-CTRL*) and after receiving MelCancerVac® (group: *MelVac)*, while 18 NSCLC patients had not received any doses of the vaccine at the time of sample donation (group: *NSCLC*).

The melanoma group comprised of patients with unresectable and metastatic melanoma ($n = 5$, ICD-10: C43; group: *PEM-Mel*), who received KEYTRUDA® (anti-PD-1 monoclonal antibody pembrolizumab, Schering-Plough Labo NV) immunotherapy as a part of standard-of-care. Serum samples of melanoma patients were collected 3 weeks after the first immunotherapy treatment, when patients came to receive the second dose (European Medicines Agency guidelines for KEYTRUDA therapy) and were provided by EGeen International (Mountain View CA, USA; ethical permit: 236/T-5).

Control groups included subjects with no history of cancer ($n = 10$, group: *CTRL-NSCLC*), with approvals for recruitment to

**Table 1 Description of clinical cohorts.**

| Cohort | Cohort ($n = 119$ individuals) | | | | |
|---|---|---|---|---|---|
| Sub-cohort | NSCLC sub-cohort | | | Melanoma sub-cohort | |
| | ($n = 34$) | | | ($n = 85$) | |
| Group | CTRL-NSCLC | NSCLC | MelVac-CTRL /MelVac | CTRL-Mel | PEM-Mel |
| Individuals | Controls without cancer ($n = 10$)[a] | NSCLC without MelCancerVac® therapy ($n = 18$)[b] | NSCLC with MelCancerVac® therapy ($n = 6$)[c] | Healthy controls ($n = 80$)[a] | Melanoma patients with pembrolizumab therapy ($n = 5$)[a] |
| Age (mean ± SD) | 65.3 ± 8.4 | 59.2 ± 7.9 | 55.7 ± 8.4 | 38.5 ± 10.7 | 67.6 ± 9.2 |
| Gender (M/F/NA) | 5/5/0 | 7/8/3 | 4/2/0 | 42/38/0 | 2/3/0 |
| Samples | $\sum$ samples = 130 | | | | |

*NSCLC* – non-small cell lung cancer patients; *CTRL-NSCLC* – non-cancer controls for NSCLC group; *MelVac* – NSCLC patients who received MelCancerVac® vaccine; *MelVac-CTRL* – paired samples of MelVac group taken before vaccination; *CTRL-Mel* – healthy controls for melanoma group; *PEM-Mel* – melanoma patients receiving pembrolizumab treatment; [a] – 1 sample per person available to researchers; [b] – 1 sample per person available to researchers, except for 3 patients (NSCLC1, NSCLC2, and NSCLC7) who had 2 samples available; [c] – 1 pre- and 1 post-vaccination sample of the patient available to researchers, except for one patient with 1 pre- and 3 post-vaccination samples. *F* – female, *M* – male, *n* – number of individuals; *NA* – not available.

the study from the Ethics Committee of the University Hospital of Liège (permit: 2018/77), and healthy blood donors ($n = 80$, ICD-10: Z52.0; group: *CTRL-Mel*) from the Blood Center of North Estonia Medical Center with the approval of the Ethics Review Committee on Human Research of the National Institute for Health Development, Estonia (permit: 1045).

**Mimotope variation analysis (MVA)**. MVA, the next generation phage display method was used to determine individual immunoprofiles reflecting antibody repertoires for the study cohort[40,41]. Two µl of serum or plasma, previously precleared to plastic and E. coli/wt M13 phage particles, was incubated with 5 µl of phage library (~$5 \times 10^{11}$ phage particles, derivative of Ph.D.-12, NEB, UK) overnight at +4 °C. The human immunoglobulin G (IgG)-captured phages were pulled down by protein G-coated magnetic beads (NEB, S1506S). IgG-bound phage DNA was extracted and samples were barcoded and sequences amplified by PCR. Pooled samples were analyzed by Illumina sequencing (50 bp single end read, Brigham Young University DNA Sequencing Center, Utah, USA).

**MVA with DDM-1.7 cell line lysate competition**. MelCancerVac® (DanDrit Biotech, Denmark/Enochian Biosciences, USA) is a therapeutic cellular vaccine based on autologous dendritic cells pulsed with the lysate of allogeneic melanoma cells (DDM-1.7) expressing several tumor antigens, including melanoma-associated antigens[42]. In MVA competition assay, freshly produced lysate from DDM-1.7 melanoma cells (Cellin Technologies, Estonia) was used to pre-block the study samples before MVA assay. Briefly, 30 µl of cell lysate (3 mg/ml) was incubated with 2 µl of serum or plasma before overnight incubation with the phage library and MVA was conducted as described.

**Data analysis and peptide antigen clustering**. Data were processed with peptide data sets cleaned of sequencing errors and known artefacts, and counts normalized to 3 million reads[40,41]. Final dataset of 12-mer peptides consisted on average of $3.26 \times 10^6$ peptide sequences ($5.8 \times 10^5$ unique) per sample, with a combined total of ~$4.2 \times 10^8$ peptide sequences. SPEXS2 exhaustive pattern search algorithm[40,41] was used to group similar peptides and reveal enriched recognition patterns (epitopes) in the studied peptide sets (Supplementary Fig. 1a). Each sample was analyzed separately for identification of sample-specific epitopes that had ≥4 fixed amino acid positions. For data analysis of *MelVac* samples, the identification of epitopes was performed in a discriminative manner, where peptide sets from *MelVac-CTRL* and *MelVac* samples of the same patient were compared to each other. Epitopes that represented peptides that were at least 2-fold more enriched in the query sample (*MelVac*) as compared to paired sample peptide set (*MelVac-CTRL*) and with a hypergeometric $p$-value < $1 \times 10^{-8}$ were selected for further analysis. For melanoma cohort ($n = 5$, *PEM-Mel*) the identification of epitopes was performed as non-discriminatory, where patient-specific epitopes were identified in comparison to a random-generated peptide set. Epitopes that represented peptides that were 10-fold more enriched in the query (*PEM-Mel*) than randomly generated reference peptide set and had a hypergeometric $p$-value < $1 \times 10^{-8}$, were selected for further analysis. Altogether 54,055 core epitopes for melanoma and 18,021 epitopes for *MelVac* groups were selected, representing a dataset of melanoma-specific antibody immune response. In addition, pairwise comparison of *MelVac-CTRL* and *MelVac* sample datasets generated 17,690 pre-treatment-specific core epitopes.

**Sequence alignment**. The set of melanoma-associated antigens used in sequence alignment were chosen from Weinert et al., 2009 data describing genes expressed in the DDM-1.7 melanoma cells[42] (Supplementary Fig. 1b). Sequences of the epitopes of the antigens were downloaded from Immune Epitope Database (IEDB[43], date accessed: 24.09.2020, www.iedb.org). Altogether, the IEDB database contained 2234 epitopes of 102 proteins expressed in the melanoma cell lysate DDM-1.7[42]. All antigen alignments were conducted using custom Excel VBA scripts.

For sequence similarity analysis, 2234 linear IEDB epitopes were exactly aligned with 54,055 melanoma and 18,021 vaccination-specific epitopes generated with SPEXS2. Thirty-five database entries (altogether 34 unique proteins) with sequence identity to at least 1 epitope from both melanoma and vaccination-specific epitope sets were recruited for further antigen-specific analysis. Primary protein sequences were downloaded from UniProtKB database[44] using accession codes matching IEDB epitope entry names (date accessed: 09.10.2020, www.uniprot.org). These 35 protein sequences were aligned with 54,055 melanoma, 18,021 vaccination-specific, and 17,690 pre-vaccination-specific epitopes, with the criteria that every fixed amino acid from SPEXS2-determined epitopes was to match with the protein sequence. Out of these, altogether 8562 epitopes aligned to sequences of 35 melanoma-associated antigens.

**ELISA**. Human cytomegalovirus (CMV) and Epstein-Barr virus (EBV) serostatuses were measured from blood samples with ISO-17025 accredited methods. In brief, serological analyses were performed with anti-CMV ELISA (IgG) method (EUROIMMUN EI 2570–9601G) and with anti-EBV-CA ELISA (IgG) method (EUROIMMUN EI 2791–9601G) according to the manufacturer's specifications. Absorbance was measured at 450 nm with SpectraMax Paradigm (Molecular Devices). For CMV serology, 41 samples tested positive, 13 negative and 2 samples were borderline and therefore excluded from further correlation analyses. For EBV serology, all measured samples were conclusive: 35 tested positive, 3 samples were negative.

**Statistics and reproducibility**. The study included 119 independent study subjects. Samples donated at different time points were considered as paired samples of the individual ($n = 130$). Technical replicates are defined as the same sample profiled in independent MVA experiments. No randomization or blinding to sample characteristics was conducted, samples were divided into groups based on clinically relevant diagnoses. Group-wise comparisons of median values were visualized using violin- or boxplots with individual data points, and statistical significance is shown where applicable. To evaluate the reproducibility of MVA data, the values of peptide abundance in two technical replicates were compared using Pearson's correlation coefficient analysis (R package "ggpubr") and the correlation value between replicates was established as $R = 0.95$ ($P < 0.0001$). Other samples were not measured repeatedly.

**Statistical analysis**. Statistical analyses were conducted with R statistical programming language v.4.0.4 and RStudio environment v.1.4.1106[45,46]. Data were analyzed, graphs were produced and visualized using R packages "reshape2", "tidyverse", "precrec", "ggpubr", "ggsci", "scales", "patchwork", "egg", "ggalt" 2021 versions[45–58].

Cosine similarity indices (CSIs) for sample comparisons based on top 2500 peptide abundance values and composition were calculated with the cosine function in R package "lsa"[59].

Top 50 immunodominant characteristics were defined from group-specific epitopes generated in SPEXS2 analysis. For post- (*Vac*, $n = 6$) or pre- (*Pre*, $n = 6$) vaccination samples the abundance of group-specific epitopes (18,021 for *Vac* and 17,690 for *Pre*, respectively) were calculated as the number of IgG-bound peptides containing the epitope sequence in the sample. The 50 epitopes with the highest abundance values were selected for analysis. Z-scores for the comparison of antibody response to top 50 immunodominant characteristics were calculated individually for each patient. First, the mean of top epitope abundance values across both *Pre* and *Vac* samples was calculated, then the mean was subtracted from the value of each epitope (mean centered) and the result divided by the standard deviation (autoscaled). For graphical presentation the values are capped off at the 97.5th percentile value of each patient.

Boxplots were generated using the style of Tukey with R packages "ggpubr" or "ggplot2"[47,48]. In figures the upper, middle and lower boxplot lines represent the 75th, 50th, and 25th percentiles, while whiskers represent the largest or smallest value within 1.5 times interquartile range above the 75th percentile or below the 25th percentile, respectively. The *p*-values of two-sided Wilcoxon Rank Sum test were visualized with "ggpubr" or "ggplot2" packages[47,48].

Wilcoxon Rank Sum test (with continuity correction, base R "stats" package[46]) was used to assess the group-differentiating features of 8562 unique epitopes aligning to melanoma-associated antigens, while custom Excel VBA script was used to determine the sensitivity and specificity while maximizing Youden's index for each biomarker. MedCalc® Statistical Software (v.19.7.2, www.medcalc.org; 2021) was used to conduct logistic regression and ROC analysis of 15 epitopes as a combinational test.

**Reporting summary**. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Results

**Highly individual patterns of top antibody response are elicited by immune therapy**. To characterize immunotherapy-specific antibody repertoires and dissect the role of the immune system in biological therapies, our analysis included two different immunotherapy cohorts and controls comprising 119 individuals (Table 1 and Supplementary Table 1). First, sera samples of NSCLC patients from the phase II clinical trial receiving autologous DC vaccine MelCancerVac® (Supplementary Table 2) and second, sera samples of melanoma patients receiving monoclonal anti-PD-1 antibody pembrolizumab as part of their standard care. To characterize the most prevalent antibody immune response in our study cohort, we used mimotope variation analysis (Supplementary Fig. 1)[40,41]. Briefly, individual blood samples were incubated with M13 phage-displayed 12-mer peptide library to capture individual-specific IgG antibody repertoires and high-throughput sequencing was used to uncover captured peptides. Cosine similarity index (CSI), a measure of similarity between the samples, was calculated to compare the top antibody response by analysis of seroresponse to 2500 peptides with the highest antibody reactivity in cohort samples ($n = 130$, Table 1). Analysis showed that different individuals with the same disease and immunotherapy background presented remarkable differences in the composition and magnitude of the dominant antibody response to different peptide antigens (CSI < 0.3, Supplementary Fig. 2a, Supplementary Data 1). The dominant antibody response to peptide antigens within groups of MelVac and MelVac-CTRL, PEM-Mel and CTRL-Mel was highly dissimilar (CSI < 0.4, Fig. 1a). However, longitudinal samples from the same individual showed similar IgG response to the top antigens (Supplementary Fig. 2b). For example, clear similarity in the top immunoprofile features of a patient denoted as MelVac1, who received 35 different

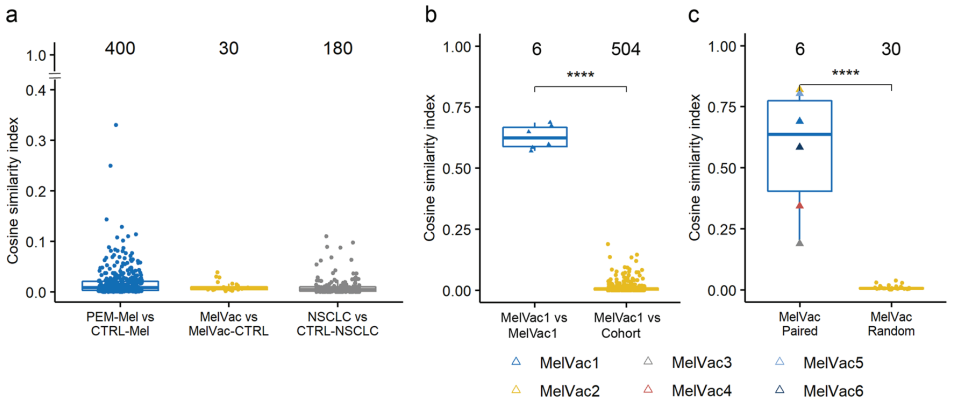



**Fig. 1 Top antibody response is individual-specific. a–c** The top antibody response was analyzed using cosine similarity indices (CSI) by comparing the composition and abundance values of the 2500 most IgG-bound peptides in each sample to that of the rest of the cohort in pairs (Supplementary Data 1). CSI values (range from 0 to 1, *y*-axis) between samples belonging to the indicated groups (*x*-axis) are depicted. Numbers above boxplots indicate the number of comparison pairs shown as dots. Comparisons of samples to themselves (CSI = 1) are not depicted. Comparisons between different individuals are indicated with circles while, comparison of the samples of the same patient are indicated with triangles. **a** Pairwise comparison between study groups and their matched controls. *PEM-Mel* – melanoma patients receiving pembrolizumab treatment (n = 5); *CTRL-Mel* – healthy controls for melanoma group ($n = 80$); *MelVac* – NSCLC patients who received MelCancerVac® vaccine ($n = 6$); *MelVac-CTRL* – paired samples of MelVac group taken before vaccination ($n = 6$); *NSCLC* – non-small cell lung cancer patients ($n = 18$); *CTRL-NSCLC* – non-cancer controls for NSCLC group ($n = 10$). **b** Pairwise comparisons of the 4 longitudinal samples of one NSCLC patient, who received 35 doses of MelCancerVac® and remained with stable disease (Supplementary Table 2), to the 4 samples themselves (*MelVac1 vs MelVac1*) and to the rest of the study cohort ($n = 126$ samples, *MelVac1 vs Cohort*). **c** Pairwise comparisons of pre- and post-vaccination immunoprofiles of vaccinated NSCLC patients ($n = 6$). *MelVac Paired* – comparison of pre- and post-vaccination samples of the same patient; *MelVac Random* - comparison of the pre-vaccination sample of one patient to the post-vaccination samples of all 5 other patients. Two-sided Wilcoxon Rank Sum test, **** $p < 0.0001$, *p*-values not adjusted for multiple comparisons.

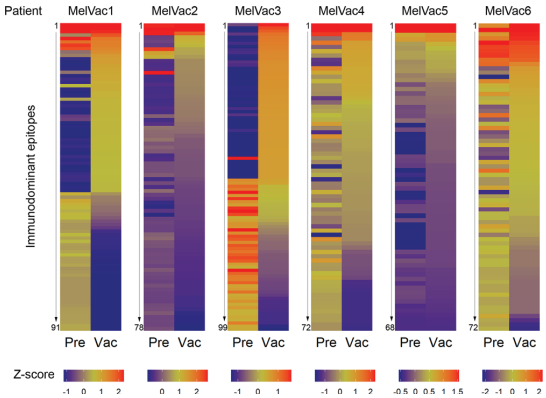



**Fig. 2 The heterogeneity of antibody response converges on immunodominant epitopes.** Heatmaps depict differential antibody response to the 50 most immunodominant epitopes detected in pre- (*Pre*, $n = 6$) and post- (*Vac*, $n = 6$) vaccination samples of patients who received MelCancerVac® treatment (*MelVac1-MelVac6*). Rows depict immunodominant epitopes with numbers on the left of each panel referring to the specific epitope sequences provided in Supplementary Data 2. The number of epitopes differs for each patient as some epitopes were in the top 50 for both *Pre* and *Vac* samples, while some were detected in only one sample of the patient. Z-scores depict the abundance of IgG-binding peptides containing the immunodominant epitopes in each sample and are calculated separately for every patient by mean centering and autoscaling the abundance values across both *Pre* and *Vac* samples. Epitopes are ranked by highest-to-lowest abundance values as observed in the *Vac* sample. Z-score scale is cut-off at 97.5th percentile for better visualization of each panel.

vaccinations of MelCancerVac® and showed stable disease over a 2-year period, was evident irrespective of the vaccination stage (CSI > 0.5, Fig. 1b). Overall, the antibody response profiles from the paired MelVac and MelVac-CTRL samples shared more similar features with each other than with unpaired NSCLC samples (Fig. 1c, Wilcoxon Rank Sum test, $p < 0.0001$, Supplementary Fig. 2b). Of note, the similarity to melanoma patients in the top 2500 peptide composition of NSCLC patients did not increase significantly upon vaccination with MelCancerVac® (Wilcoxon Rank Sum test, $p > 0.05$, Supplementary Fig. 2c). Next, to delineate epitopes characteristic to MelVac group we used SPEXS2 exhaustive pattern search algorithm (Supplementary Fig. 1a). Comparison of seroresponse values to the top 50 most targeted epitopes in both pre- and post-vaccination samples of the same patient revealed common antigenic features present in both conditions (Fig. 2 and Supplementary Data 2). However, we also observed that seroresponse to the top 50 antigens changed upon MelCancerVac® vaccination, although these changes were largely individual-specific (Fig. 2). Specifically, some patients maintained antibody reactivity to majority of epitopes upon vaccination (MelVac2, MelVac4–6), whereas the top response epitopes in others changed (MelVac1, MelVac3, Fig. 2). Overall, our data showed that even though intra-individual cancer immunoprofiles were more similar than inter-individual ones, the antibody immune response in NSCLC cohort was dynamic upon MelCancerVac® vaccination with changes both in the composition and abundance to immunodominant antigens.

**Immune reactivity targets epitopes of melanoma antigens**. We hypothesized that dendritic cell vaccine therapy based on melanoma cell lysate could elicit melanoma-antigen-specific antibody

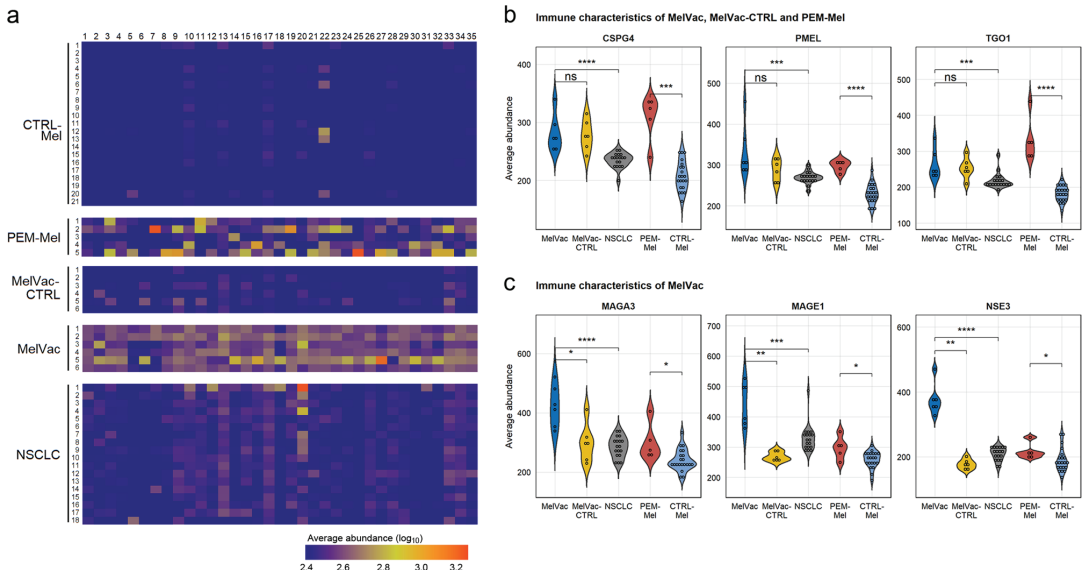



**Fig. 3 Individual-specific immunoprofiles of antibody response to melanoma-associated antigens. a** Heatmap showing the antibody response to melanoma-antigens in patients with cancer and in the controls. $Log_{10}$-transformation of the average abundance of the IgG-bound peptides containing epitopes with 100% identity to the indicated proteins are depicted. Value range ≤ 2.4 and ≥ 3.3 is provided for better visual representation. Group names of samples are depicted on the *y*-axis, individuals shown in numbers. *x*-axis indicates different melanoma-associated antigens (a total of 35 proteins) shown in Supplementary Table 3 and Supplementary Data 3. **b**, **c** Violin plots showing the average abundance of IgG-bound peptides containing epitopes with 100% identity to the specific proteins (shown above each graph, Supplementary Data 4) from panel **a** across study sub-cohorts. Two-sided Wilcoxon Rank Sum test, ns $p > 0.5$, *$p < 0.05$, **$p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, *p*-values not adjusted for multiple comparisons. *CTRL-Mel* – healthy controls for melanoma group ($n = 21$, all individuals older than 45 years); *PEM-Mel* – melanoma patients receiving pembrolizumab treatment ($n = 5$); *MelVac-CTRL* – paired samples of MelVac group taken before vaccination ($n = 6$); *MelVac* – NSCLC patients who received MelCancerVac® vaccine ($n = 6$); *NSCLC* – non-small cell lung cancer patients ($n = 18$).

response in NSCLC patients. In particular, considering that out of all protein antigens known to be expressed by the DDM-1.7 melanoma cells[42], 102 proteins were reported to have epitopes showing serologically positive findings in Immune Epitope Database (IEDB). To characterize protein-specific immune responses in immunotherapy patients, we used SPEXS2 exhaustive pattern search algorithm to group individual peptides of MelVac, MelVac-CTRL, and PEM-Mel groups into representative epitopes and compared the delineated epitopes with known antigenic sequences. Altogether, antibody response to IEDB epitopes with sequence similarity to 35 proteins was detected in both MelVac and PEM-Mel groups (Supplementary Table 3). To characterize potential new antibody targets, we aligned the cancer-group-specific epitopes with the primary sequence of these 35 proteins. With an average of 340 epitope alignments per protein, 8562 unique epitopes from either MelVac, MelVac-CTRL or PEM-Mel group matched exactly with these antigens (Fig. 3a and Supplementary Data 3). Although, immunotherapy groups showed high reactivity to different sets of epitopes, the overall antibody response converged on the same antigenic proteins (Fig. 3a). Significant difference in antibody response to epitopes of known melanoma antigens CSPG4 (#4), PMEL (#6) and TGO1 (#31) was noticeable between PEM-Mel and MelVac and their respective control groups (Wilcoxon Rank Sum test, $p < 0.001$, Fig. 3b, Supplementary Data 4). Antibody response to a subset of melanoma-associated antigens (MAGA3 (#9), MAGE1 (#17), and NSE3/MAGEG1 (#25)) was pronounced in MelVac group and high in PEM-Mel group (Fig. 3c, Wilcoxon Rank Sum test, $p < 0.05$). Overall, we observed melanoma-associated antigen-specific IgG signatures elicited by MelCancerVac® and anti-PD-1 immunotherapies.

**Fifteen group-discriminating epitopes converge on antigens associated with modulation of extracellular matrix and tumor cell survival pathways**. To identify specific changes in immune response upon immunotherapy, we divided the study cohort into two subsets: controls ($n = 90$) and cancer samples with melanoma-associated attributes ($n = 11$, comprising of both MelVac and PEM-Mel groups). ROC analysis of 8562 antigen-associated epitopes resulted in 15 most group-discriminating antigenic determinants (markers M1 to M15) with sensitivity > 0.72 and specificity > 0.67 (Supplementary Fig. 3 and Supplementary Table 4). Notably, several of the resolved 15 biomarkers (M1-M15) mapped to the same antigens, but to different epitopes (M2 and M7 to MAGD2; M2 and M5 to MAGE1; M2 and M6 to PMEL; M3 and M11 to MORC4; M4 and M10 to MAGEMG50; M8 and M14 to CSPG4; M9 and M13 to CRGB1, Supplementary Table 5). Analysis of protein structure and biological relevance data from UniProtKB database indicated that majority of these 15 epitopes aligned to antigenic regions that were enriched in polar amino acids and located preferentially in regions with no known structural domains or in disordered segments. For example, M2 and M7 to MAGD2, M2, and M5 to MAGE1, M7 to MAGD1, M10 to MAGEMG50; M9 and M13 to CRGB1 and M12 to MAGE6. Some epitopes encompassed well-conformed domains like coiled-coil repeat of G3V599 for M1, leucine rich repeat of PRA22 for M3, cadherin-like CSPG repeats for M8 and M14. Biologically, these antigens are associated with extracellular-matrix formation (MAGEMG50) and modulation collagen protein turnover pathways (G3V599, CSPG4, and TGO1), with P53-associated apoptosis (MAGEMG50) and/or via ubiquitin ligase activity (MAGE1, MAGD1, MAGA3), but also with melanosome

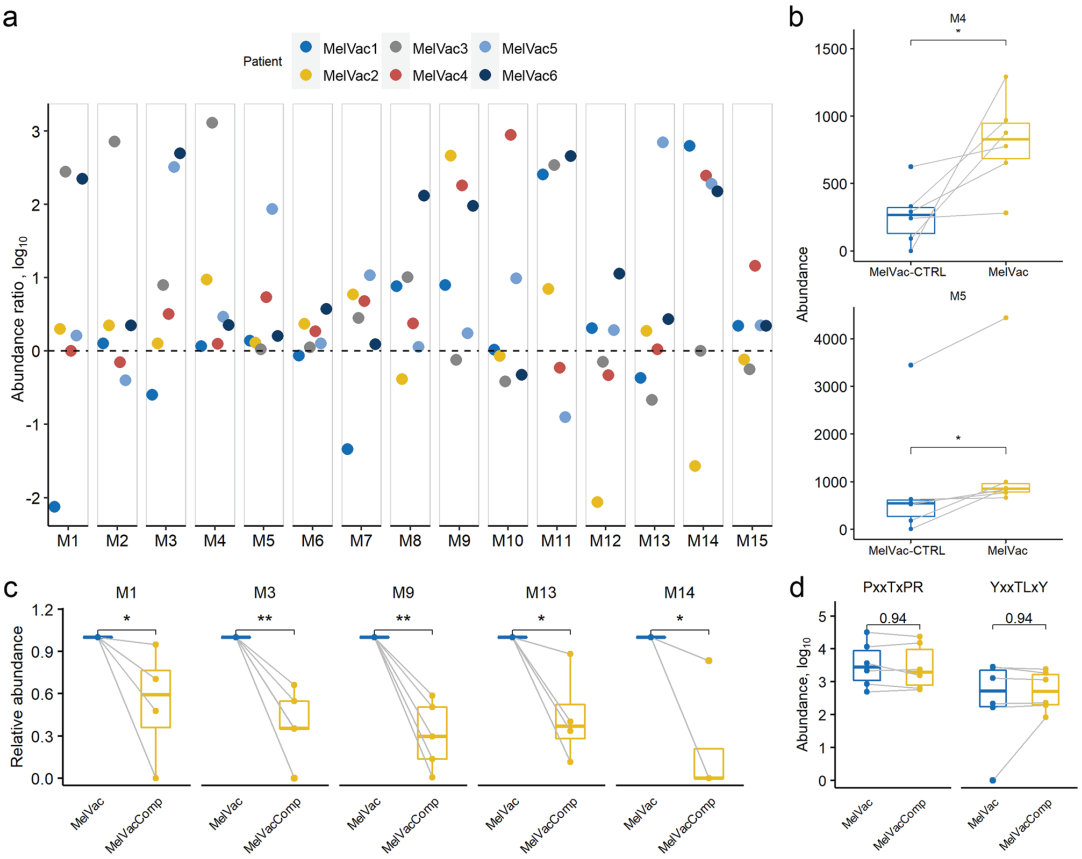


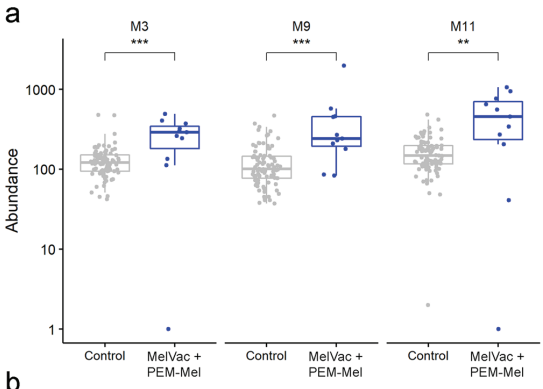

**Fig. 4 Antibody response to 15 melanoma-specific epitopes is pre-existing before MelCancerVac® vaccination and boosted upon vaccine stimulation.** **a** Comparison of antibody response to 15 epitopes in samples taken before ($n = 6$, *MelVac-CTRL*) and after vaccination ($n = 6$, *MelVac*) in six MelCancerVac® receiving patients (*MelVac1-MelVac6*). *x*-axis denotes 15 epitopes as biomarkers (*M1-M15*), *y*-axis (*Abundance ratio*) shows the ratio of abundance values of IgG-bound peptides between paired MelVac and MelVac-CTRL samples of the patient ($\text{MelVac}_i[M_{\text{abundance}} + 1]/\text{MelVac-CTRL}_i[M_{\text{abundance}} + 1]$, *i* – number of patient, *M* – biomarker) in base 10 logarithmic scale. Dashed line indicates ratio value 0 (1 in linear scale), i.e., where antibody reactivity to peptides containing the specific epitopes remained unchanged in MelCancerVac® post-vaccination cohort. Values > 0 indicate rise in seroreactivity after vaccination while <0 indicates decrease. Source abundance values for each epitope are presented in Supplementary Data 5. **b** Vaccine-dependent antibody response enhancement to the resolved epitopes was common. Data are shown for epitopes M4 and M5 by comparing abundances of IgG-bound peptides from the vaccinated patients, before (*MelVac-CTRL*) and after vaccination (*MelVac*). *Abundance* – number of IgG-bound peptides containing the specified epitope sequence detected in the sample. Two-tailed paired Wilcoxon Rank Sum test, * $p < 0.05$, *p*-values not adjusted for multiple comparisons. **c** Box plots show the abundance of IgG-bound peptides containing the specified epitopes (M1, M3, M9, M13, and M14) upon MVA competition analysis. *MelVacComp* – data from competition with DDM-1.7 melanoma cell lysate is shown. *Relative abundance* – the abundance of IgG-bound peptides containing the specified epitopes normalized to values of the paired vaccination-specific sample (*MelVac*) for each patient. **d** Box plots show the abundance of IgG-bound peptides containing sequences of the viral capsid antigen p18 (EBV VCA p18 epitope ($_{161}$ GGQPHDTAPRGARKK $_{175}$) and the epitope of glycoprotein B (CMV gB; $_{70}$ ETIYNTTLKY $_{80}$)[40] from MVA competition analysis. *MelVacComp* – data from competition with DDM-1.7 melanoma cell lysate is shown. *Abundance* – the abundance of IgG-bound peptides containing the specified epitopes in base 10 logarithmic scale.

biogenesis (PMEL), ciliary signaling (ARMC) and lipoprotein signaling (G3V599). We also analyzed whether antibody response to common human herpesviruses, including cytomegalovirus (CMV) and Epstein-Barr virus (EBV) contributed to the treatment-elicited anti-cancer immunity given that EBV and CMV are the most prevalent infection types in tumors[60] and can act as independent biomarkers for cancer immunotherapy[61]. For that, we analyzed EBV and CMV serology and showed that seropositivity to these common herpesviruses (Supplementary Fig. 4) was not correlated with treatment-elicited antibody response to the resolved top 15 melanoma-associated epitopes (Supplementary Data 5, 6). Based on both clinical serology and MVA data, we concluded that melanoma-associated antibody response linked to immunotherapy pointed onto apoptotic signaling and extracellular matrix-remodeling pathways conveyed by tumor-antigens but was not correlated with the common herpesviral antigens.

**MelCancerVac® boosts prior antibody response against a subset of melanoma-associated antigens**. Analysis of longitudinal samples revealed that antibody response to the majority

of the 15 melanoma-differentiating epitopes was detectable in MelVac-CTRL samples with levels boosted by MelCancerVac® administration in paired MelVac samples (Fig. 4a and Supplementary Fig. 5). Although the seroresponse changes to the majority of the 15 epitopes were patient-specific, reactivity to the epitope markers M4 (MAGEMG50) and M5 (MAGE1) was similarly boosted by vaccine in all patients (Wilcoxon Rank Sum test, $p < 0.05$, Fig. 4b). Furthermore, MVA with melanoma cell lysate competition confirmed that the high antibody reactivity observed in MelVac samples was specific to melanoma proteins as blocking with cell lysate interfered with IgG binding to most of the aforementioned epitopes, with significant effects for M1, M3, M9, M13, and M14 (Fig. 4c and Supplementary Fig. 6). The specificity of blocking with anti-melanoma-lysate was further confirmed by analyzing independent EBV and CMV-associated epitopes. Namely, the antibody response to epitopes of the viral capsid antigen p18 (EBV VCA p18; $_{161}$ GGQ**P**HD**T**A**PR**GARKK $_{175}$) and glycoprotein B (CMV gB; $_{70}$ ETI**Y**NT**TLKY** $_{80}$)[40] did not change upon competition (Fig. 4d). Therefore, we conclude that antibody response boosted upon MelCancerVac® vaccination is specific to epitopes related to melanoma-associated antigens and is frequently correlated with pre-existing immune response to these antigens.

**Three-epitope signature as a biomarker of immunotherapy-elicited melanoma-specific response**. As high melanoma-specific response appears to be a feature of tumor immunogenicity and its sensitivity to targeted treatments, we were interested to examine the prognostic utility of the resolved epitopes in therapy-elicited response. We used logistic regression and ROC analysis and found that 3 epitope biomarkers: M3 (ARMC9/PRA22/MORC4) (73% sensitivity, 97% specificity), M9 (CRBG1) (Sens 82% sensitivity, 81% specificity) and M11 (MORC4) (73% sensitivity, 90% specificity) in combination differentiated cancer patients based on melanoma-specific response elicited by treatments from controls with area under curve (AUC) of 0.991, ~91% sensitivity and 100% specificity (Fig. 5a, b and Supplementary Fig. 7). Here we show that antibody reactivity to a small subset of epitopes of known melanoma-associated antigenic determinants serves as a biomarker associated with melanoma-specific immunity elicited by cancer immunotherapies.




b

| Study cohort | AUC (p>0.0001) | Sensitivity | Specificity | 95% CI |
|---|---|---|---|---|
| MelVac+ PEM-Mel (n=11) vs Control (n=90) | 0.911 (SE 0.00953) | 90.91 | 100 | 0.951-1.000 |

**Fig. 5 Antibody reactivity to epitopes of melanoma-antigens is associated with immunotherapy. a** Box plots showing the IgG response to peptides containing the 3 epitopes that were most differentiating for melanoma-specific immunotherapy as deemed by logistic regression model analysis. *Abundance* – number of IgG-bound peptides containing the epitope biomarker sequence detected in a sample. *Control* – healthy controls for melanoma group (CTRL-Mel, $n = 80$) and non-cancer controls of NSCLC group (CTRL-NSCLC, $n = 10$); *MelVac* – NSCLC patients who received MelCancerVac® vaccine ($n = 6$); *PEM-Mel* – melanoma patients receiving pembrolizumab treatment ($n = 5$). Two-sided Wilcoxon Rank Sum test $^{**}p < 0.01$, $^{***}p < 0.001$, *p*-values not adjusted for multiple comparisons. **b** Logistic regression model of biomarkers M3, M9, and M11. AUC – area under curve; SE – standard error; CI – confidence interval.

## Discussion

Here, we note unique and divergent changes in melanoma-antigen immune profiles in cancer patients receiving immunotherapy treatments that implicate distinct humoral immune functions connected to the therapy. The IgG response to specific epitopes of a subset of melanoma-antigens was associated with dendritic cell vaccine treatments in lung cancer. Patients receiving MelCancerVac® showed prior response to some epitopes which was enhanced upon treatments, concluding these as potential biomarkers associated with anti-melanoma immunity already at the pre-treatment stage. The same epitopes were targeted by antibodies in melanoma patients receiving anti-PD-1 therapy. The resolved antigenic determinants were in proteins involved in the formation and modulation of extracellular matrix, and also tumor cell survival. Resolved antibody response targeting melanoma or melanoma-like features in cancer elicited by different types of immunotherapies sets the stage for future investigations of the epitopes and their clinical relevance as biomarkers for predicting therapy efficiency in larger studies.

Herein, we used MVA to discover epitope biomarkers associated with anti-cancer antibody response. Our data establish that individuals with cancer show highly heterogenous immune response to peptide antigens that is neither clinical group, cancer type nor immunotherapy specific. This finding was expected and could be related to the genetic variation and history of previously encountered pathogens[62]. However, we found clear similarities in antibody response to melanoma-associated proteins, including CTAs, a group of antigens now exceeding more than 200 proteins[63] in patients upon different immunotherapy treatments. The resolved epitopes were mapped to extra- and intracellular proteins associated with the formation and modulation of extracellular matrix, but also with tumor cell survival, ciliary functions and lipoprotein signaling. Despite the fact that CTAs are mostly internal tumor antigens, the restricted expression of CTAs in tissues and their antigenicity has promoted utilizing them as targets for immunotherapies[64]. Interestingly, among other epitopes targeted by treatment-elicited antibody response in melanoma and NSCLC cases was of MAGE-A3 protein. MAGE-A3 has been detected in up to 76% of melanomas and in 30–50% of NSCLCs and is thus currently trialed as a target for immunotherapy[65,66]. Antibody response to epitopes coalescing on other MAGE group of proteins could be related to the poor prognostic features of metastases and melanoma progression[37,67]. On the other hand, given that the resolved group of antigens included widely expressed proteins (including MAGE family of antigens like MAGE-D), the humoral response towards these could mark excessive immune-attack and damage to self-tissues as concluded by others from studies of ICI-based therapies of melanoma and NSCLC[30,68,69]. This suggests that MVA-defined epitopes from intracellular antigens could be indicators of immunotherapy-associated tumor cell death. Our data on heterogenic immune response to melanoma-antigens are in good harmony with the findings on the heterogeneous expression of

CTAs[70]. Therefore, characterization of the antibody response towards to canonical tumor antigens, including CTAs at pre- and post-treatment stage at epitope resolution could provide new strategies to detect and tackle cancer.

Most of the anti-melanoma-associated antigen immune response that we described for the NSCLC MelCancerVac® cohort was also detected in patients with unresectable metastatic melanoma who received anti-PD-1 antibody (pembrolizumab) immunotherapy. We found that the resolved epitopes of MAGE-A antigens were similarly targeted by antibodies in both cases, suggesting the redundant cellular functions of the underlying antigens in lung cancer and melanoma. We expected to detect the anti-melanoma immunogenicity mainly in post-vaccination cases, but interestingly we determined antitumor immunity already at the pre-vaccine stage. Data from IFNγ analysis of MelCancerVac® trial demonstrated T cell-specific response correlating with vaccine-specific immunity and sustained stable disease[38]. These findings on the ICI-associated restoration of T cell activity by MelCancerVac® are in good agreement with the observed anti-melanoma-specific humoral response patterns detected by our analysis. This once again highlights the importance to determine the elicited antibody response to specific tumor antigens as a measure of the anti-tumor activity associated with immune treatments for assessing the clinical utility of the treatment. However, to confirm the relationship between the melanoma-antigen associated epitopes and clinical efficacy, the number of patients needs to be substantially increased in future studies.

Autoantibodies have the potential to provide unique fingerprints that reflect the nature of the malignant process in the affected organ. Studies of B cell immunity on melanoma have demonstrated its important role in anti-tumor response, but also in irAEs associated with ICIs[71]. Relatedly, Toi et al. reported that although the presence of preexisting antibodies was associated with clinical benefits, it also led to the development of irAEs in NSCLC patients treated with anti-PD-1 monotherapy[69]. Therefore, the antibody epitopes we determined might be useful to assess immunopathological effects in order to minimize the probability of deleterious autoimmunity.

We validated the findings of epitopes mapped to melanoma-associated antigens from MVA competition analysis using melanoma cell lysate-specific antibody depletion approach from MelCancerVac® samples. Interestingly, antibody response against three melanoma antigens observed in the pre-vaccination group, that were further confirmed to be therapy-connected by melanoma-lysate competition assay suggests that these could act as positive biomarkers of anti-tumor response to immune therapy. This is interesting in the light of the recent studies showing that preexisting antibody response in peripheral blood to tumor antigens, amongst these to MAGE1 carrying the resolved M5 epitope, is predictive of good clinical response to anti-PD-1 monotherapy of NSCLC[72–74]. Currently, high mutational burden in a tumor has been shown to predict sensitivity to anti-PD-1 therapies[75]. The expression of certain CTAs (for example MAGE-A) has been linked to poor disease prognosis due to reduced treatment response[76] or is used to predict resistance to immunotherapy[77]. Our research on mapping the anticancer immunity to specific epitopes of tumor antigens prior to therapy could provide valuable prognostic knowledge for making the best therapeutic decisions, minimizing the probability of deleterious autoimmunity, and also for identification of novel targets for immunotherapy.

A big obstacle for effective cancer immunotherapy is the scarcity of immunotherapeutic biomarkers with specific clinical correlation to treatment efficacy and in relation to possible side effects. Here, we connected antibody profiles as defined by MVA to anti-tumor immunity elicited by immunotherapy at epitope precision. Our data provide support for the use of epitopes of tumor antigens as biomarkers for patient stratification and in immune monitoring. Future studies will elaborate on the connection between antibody profiles and ICI treatment outcomes.

There are several important caveats to this study. Although the MVA data covers a broad range of peptide antigens, extensive knowledge of the immune proteome of cancer is quite limited. Therefore, the mechanisms of how antibody response to the examined melanoma-specific antigens precisely impacts tumor immunogenicity, effective therapeutic targeting, or therapy efficacy remain to be solved. In addition, despite the use of multiple statistical methods and efforts to provide authentic targets to our findings, it is certainly possible for an epitope to play an important role in cancer and may mimic tumor-antigens but be derived from other antigens. For example, MAGE-A6 (amino acids 172–187) and *Mycoplasma penetrans* HF-2 permease (amino acids 216–229) share structural homology and elicit complex immunologic cross-reactivity[78]. We were thus careful with our conclusions: we can implicate an epitope by similarity to a self-protein but it is premature to exclude neoepitopes, cryptic epitopes, and metagenome-associated epitopes for which we do not have data. Finally, although this study harnessed the use of samples from a phase II clinical trial for the discovery of blood biomarkers, due to the limited number of samples further clinical studies are warranted.

## Data availability

Source data underlying the main figures are provided as Supplementary Data files 1–6. The whole sequencing datasets generated and/or analyzed in this study are not publicly available due to containing sensitive clinical information but are available from the corresponding author upon reasonable request via a material transfer agreement.

## Code availability

The R packages used to generate and analyze data presented in this study are described in the Methods section and available to all. The SPEXS2 algorithm for delineating epitopes from peptide sets can be found at https://github.com/egonelbre/spexs2. The custom Excel VBA used to find sequence similarities between generated epitopes and protein sequences are available from the corresponding author upon request.

Received: 20 July 2021; Accepted: 25 April 2022;
Published online: 11 May 2022

## References

1. Cerezo-Wallis, D. & Soengas, M. S. Understanding tumor-antigen presentation in the new era of cancer immunotherapy. *Curr. Pharm Des.* **22**, 6234–6250 (2016).
2. Chong, C., Coukos, G. & Bassani-Sternberg, M. Identification of tumor antigens with immunopeptidomics. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-021-01038-8 (2021).
3. Michels, J. et al. Multiplex bead-based measurement of humoral immune responses against tumor-associated antigens in stage II melanoma patients of the EORTC18961 trial. *Oncoimmunology* **7**, e1428157 (2018).
4. Gordeeva, O. Cancer-testis antigens: Unique cancer stem cell biomarkers and targets for cancer therapy. *Semin. Cancer Biol.* **53**, 75–89 (2018).
5. Verdegaal, E. M. et al. Neoantigen landscape dynamics during human melanoma-T cell interactions. *Nature* **536**, 91–95 (2016).
6. Finn, O. J. The dawn of vaccines for cancer prevention. *Nat. Rev. Immunol.* **18**, 183–194 (2018).
7. Dunn, G. P., Old, L. J. & Schreiber, R. D. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity* **21**, 137–148 (2004).
8. Waldman, A. D., Fritz, J. M. & Lenardo, M. J. A guide to cancer immunotherapy: From T cell basic science to clinical practice. *Nat. Rev. Immunol.* **20**, 651–668 (2020).
9. Shalapour, S. & Karin, M. The neglected brothers come of age: B cells and cancer. *Semin. Immunol.* **52**, 101479 (2021).
10. Mirandola, L. et al. Novel antigens in non-small cell lung cancer: SP17, AKAP4, and PTTG1 are potential immunotherapeutic targets. *Oncotarget* **6**, 2812–2826 (2015).
11. Fridman, W. H. et al. B cells and cancer: To B or not to B? *J. Exp. Med.* https://doi.org/10.1084/jem.20200851 (2021).

12. Zitvogel, L., Perreault, C., Finn, O. J. & Kroemer, G. Beneficial autoimmunity improves cancer prognosis. *Nat. Rev. Clin. Oncol.* **18**, 591–602 (2021).
13. Blankenstein, T., Coulie, P. G., Gilboa, E. & Jaffee, E. M. The determinants of tumour immunogenicity. *Nat. Rev. Cancer* **12**, 307–313 (2012).
14. Dersh, D., Holly, J. & Yewdell, J. W. A few good peptides: MHC class I-based cancer immunosurveillance and immunoevasion. *Nat. Rev. Immunol.* **21**, 116–128 (2021).
15. Leko, V. & Rosenberg, S. A. Identifying and targeting human tumor antigens for T cell-based immunotherapy of solid tumors. *Cancer Cell* **38**, 454–472 (2020).
16. Fonseca, C. & Dranoff, G. Capitalizing on the immunogenicity of dying tumor cells. *Clin. Cancer Res.* **14**, 1603–1608 (2008).
17. Passarelli, A., Mannavola, F., Stucci, L. S., Tucci, M. & Silvestris, F. Immune system and melanoma biology: A balance between immunosurveillance and immune escape. *Oncotarget* **8**, 106132–106142 (2017).
18. Kruit, W. H. et al. Selection of immunostimulant AS15 for active immunization with MAGE-A3 protein: results of a randomized phase II study of the European Organisation for Research and Treatment of Cancer Melanoma Group in Metastatic Melanoma. *J. Clin. Oncol.* **31**, 2413–2420 (2013).
19. Vansteenkiste, J. et al. Adjuvant MAGE-A3 immunotherapy in resected non-small-cell lung cancer: phase II randomized study results. *J. Clin. Oncol.* **31**, 2396–2403 (2013).
20. Wculek, S. K. et al. Dendritic cells in cancer immunology and immunotherapy. *Nat. Rev. Immunol.* **20**, 7–24 (2020).
21. Carreno, B. M. et al. Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348**, 803–808 (2015).
22. Shen, X. & Zhao, B. Efficacy of PD-1 or PD-L1 inhibitors and PD-L1 expression status in cancer: Meta-analysis. *BMJ* **362**, k3529 (2018).
23. Rizvi, N. A. et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
24. Trebeschi, S. et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann. Oncol.* **30**, 998–1004 (2019).
25. Linette, G. P. & Carreno, B. M. Tumor-infiltrating lymphocytes in the checkpoint inhibitor era. *Curr. Hematol. Malig. Rep.* **14**, 286–291 (2019).
26. Gilbert, A. E. et al. Monitoring the systemic human memory B cell compartment of melanoma patients for anti-tumor IgG antibodies. *PLoS One* **6**, e19330 (2011).
27. Ladanyi, A. Prognostic and predictive significance of immune cells infiltrating cutaneous melanoma. *Pigment Cell Melanoma Res.* **28**, 490–500 (2015).
28. Diem, S. et al. Immunoglobulin G and subclasses as potential biomarkers in metastatic melanoma patients starting checkpoint inhibitor treatment. *J. Immunother.* **42**, 89–93 (2019).
29. Fassler, M. et al. Antibodies as biomarker candidates for response and survival to checkpoint inhibitors in melanoma patients. *J. Immunother. Cancer* **7**, 50 (2019).
30. Gowen, M. F. et al. Baseline antibody profiles predict toxicity in melanoma patients treated with immune checkpoint inhibitors. *J. Transl. Med.* **16**, 82 (2018).
31. Battaini, F. et al. Antibody response after vaccination with antigen-pulsed dendritic cells. *Int. J. Biol. Markers* **19**, 213–220 (2004).
32. Block, M. S. et al. Th17-inducing autologous dendritic cell vaccination promotes antigen-specific cellular and humoral immunity in ovarian cancer patients. *Nat. Commun.* **11**, 5173 (2020).
33. Santin, A. D. et al. Human papillomavirus type 16 and 18 E7-pulsed dendritic cell vaccination of stage IB or IIA cervical cancer patients: A phase I escalating-dose trial. *J. Virol.* **82**, 1968–1979 (2008).
34. Lu, J., Lee-Gabel, L., Nadeau, M. C., Ferencz, T. M. & Soefje, S. A. Clinical evaluation of compounds targeting PD-1/PD-L1 pathway for cancer immunotherapy. *J. Oncol. Pharm. Pract.* **21**, 451–467 (2015).
35. Yun, S., Vincelette, N. D., Green, M. R., Wahner Hendrickson, A. E. & Abraham, I. Targeting immune checkpoints in unresectable metastatic cutaneous melanoma: A systematic review and meta-analysis of anti-CTLA-4 and anti-PD-1 agents trials. *Cancer Med.* **5**, 1481–1491 (2016).
36. Stamell, E. F., Wolchok, J. D., Gnjatic, S., Lee, N. Y. & Brownell, I. The abscopal effect associated with a systemic anti-melanoma immune response. *Int. J. Radiat. Oncol. Biol. Phys.* **85**, 293–295 (2013).
37. Barrow, C. et al. Tumor antigen expression in melanoma varies according to antigen and stage. *Clin. Cancer Res.* **12**, 764–771 (2006).
38. Engell-Noerregaard, L. et al. Clinical and immunological effects in patients with advanced non-small cell lung-cancer after vaccination with dendritic cells exposed to an allogeneic tumor cell lysate. *World J. Vaccines* **03**, 68–76 (2013).
39. Kvistborg, P. et al. Comparison of monocyte-derived dendritic cells from colorectal cancer patients, non-small-cell-lung-cancer patients and healthy donors. *Vaccine* **28**, 542–547 (2009).
40. Sadam, H. et al. Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients. *EBioMedicine* **64**, 103211 (2021).
41. Sadam, H. et al. Prostaglandin D2 receptor DP1 antibodies predict vaccine-induced and spontaneous narcolepsy type 1: Large-scale study of antibody profiling. *EBioMedicine* **29**, 47–59 (2018).
42. Weinert, B. T. et al. Real-time PCR analysis of genes encoding tumor antigens in esophageal tumors and a cancer vaccine. *Cancer Immun.* **9**, 9 (2009).
43. Vita, R. et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2019).
44. UniProt, C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
45. RStudio Team. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/ (2021).
46. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (2020).
47. Kassambra, A. ggpubr: 'ggplot2' Based Publication Ready Plots. *R package version 0.4.0.* (2020).
48. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York* (2016).
49. Wickham, H., Averick, M., Bryan, J., Chang, W. & McGowan, L. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 43 (2019).
50. Sito, T. & Rehmsmeier, M. Precrec: Fast and accurate precision-recall and ROC curve calculations in R. *Bioinformatics* **33**, 145–147 (2017).
51. Wickham, H. & Seidel, D. scales: Scale Functions for Visualization. *R package version 1.1.1.* (2020).
52. Henry, L. & Wickham, H. purrr: Functional Programming Tools. *R package version 0.3.4.* (2020).
53. Pedersen, T. L. patchwork: The Composer of Plots. *R package version 1.1.1.*(2020).
54. Auguie, B. egg: Extensions for 'ggplot2': Custom Geom, Custom Themes, Plot Alignment, Labelled Panels, Symmetric Scales, and Fixed Panel Size. *R packageversion 0.4.5.* (2019).
55. Rudis, B., Bolker, B. & Schulz, J. ggalt: Extra Coordinate Systems, 'Geoms', Statistical Transformations, Scales and Fonts for 'ggplot2' *R packageversion 0.4.0.* (2017).
56. Wickham, H. Reshaping Data with the reshape Package. *Journal of Statistical Software*, **21**, 1–20 (2007).
57. Saito, T. & Rehmsmeier, M. Precrec: fast and accurate precision-recall and ROC curve calculations in *Bioinformatics*, **33**, 145-147 (2017).
58. Xiao, N. ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for 'ggplot2'.*R package version 2.9.* (2018).
59. Wild, F. lsa: Latent Semantic Analysis. *R package version 0.73.2.* (2020).
60. Chen, S. et al. The viral expression and immune status in human cancers and insights into novel biomarkers of immunotherapy. *BMC Cancer* **21**, 1183 (2021).
61. Varn, F. S., Schaafsma, E., Wang, Y. & Cheng, C. Genomic characterization of six virus-associated cancers identifies changes in the tumor immune microenvironment and altered genetic programs. *Cancer Res.* **78**, 6413–6423 (2018).
62. Watson, C. T., Glanville, J. & Marasco, W. A. The individual and population genetics of antibody immunity. *Trends Immunol.* **38**, 459–470 (2017).
63. Gjerstorff, M. F., Andersen, M. H. & Ditzel, H. J. Oncogenic cancer/testis antigens: Prime candidates for immunotherapy. *Oncotarget* **6**, 15772–15787 (2015).
64. Li, X. F., Ren, P., Shen, W. Z., Jin, X. & Zhang, J. The expression, modulation, and use of cancer-testis antigens as potential biomarkers for cancer immunotherapy. *Am. J. Transl. Res.* **12**, 7002–7019 (2020).
65. Vansteenkiste, J. F. et al. Efficacy of the MAGE-A3 cancer immunotherapeutic as adjuvant therapy in patients with resected MAGE-A3-positive non-small-cell lung cancer (MAGRIT): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* **17**, 822–835 (2016).
66. Dreno, B. et al. MAGE-A3 immunotherapeutic as adjuvant therapy for patients with resected, MAGE-A3-positive, stage III melanoma (DERMA): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Oncol.* **19**, 916–929 (2018).
67. Roeder, C. et al. MAGE-A3 is a frequent tumor antigen of metastasized melanoma. *Arch. Dermatol. Res.* **296**, 314–319 (2005).
68. de Moel, E. C. et al. Autoantibody development under treatment with immune-checkpoint inhibitors. *Cancer Immunol. Res.* **7**, 6–11 (2019).
69. Toi, Y. et al. Profiling preexisting antibodies in patients treated with anti-PD-1 therapy for advanced non-small cell lung cancer. *JAMA Oncol.* **5**, 376–383 (2019).
70. Jakobsen, M. K. et al. The cancer/testis antigen gene VCX2 is rarely expressed in malignancies but can be epigenetically activated using DNA methyltransferase and histone deacetylase inhibitors. *Front. Oncol.* **10**, 584024 (2020).
71. Willsmore, Z. N. et al. B cells in patients with melanoma: Implications for treatment with checkpoint inhibitor antibodies. *Front. Immunol.* **11**, 622442 (2020).

72. Ohue, Y. et al. Serum antibody against NY-ESO-1 and XAGE1 antigens potentially predicts clinical responses to anti-programmed cell death-1 therapy in NSCLC. *J Thorac. Oncol.* **14**, 2071–2083 (2019).
73. Tan, Q. et al. Autoantibody profiling identifies predictive biomarkers of response to anti-PD1 therapy in cancer patients. *Theranostics* **10**, 6399–6410 (2020).
74. Zhou, J. et al. Peripheral blood autoantibodies against to tumor-associated antigen predict clinical outcome to immune checkpoint inhibitor-based treatment in advanced non-small cell lung cancer. *Front. Oncol.* **11**, 625578 (2021).
75. McGranahan, N. et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351**, 1463–1469 (2016).
76. Wong, P. P. et al. Identification of MAGEA antigens as causal players in the development of tamoxifen-resistant breast cancer. *Oncogene* **33**, 4579–4588 (2014).
77. Shukla, S. A. et al. Cancer-germline antigen expression discriminates clinical outcome to CTLA-4 blockade. *Cell* **173**, 624–633 e628 (2018).
78. Vujanovic, L., Shi, J., Kirkwood, J. M., Storkus, W. J. & Butterfield, L. H. Molecular mimicry of MAGE-A6 and Mycoplasma penetrans HF-2 epitopes in the induction of antitumor CD8(+) T-cell responses. *Oncoimmunology* **3**, e954501 (2014).

## Acknowledgements

Previous members of DanDrit Biotech A/S team Eric Leire (Genflow Biosciences, Immunethep), Mai-Britt Zocca (IO Biotech ApS) are thanked for assistance on study execution. We thank A.-H. Pool from the Department of Neuroscience, University of Texas Southwestern Medical Center, for the expertize that supported this work. The members of Protobios team are thanked for their excellent technical support. This study was supported by research funding grants of Protobios (5.1–4/20/170, and PRG573) from the Estonian Ministry of Education and Estonian Research Council, respectively, and H2020-MSCA-RISE-2016 (EU734791) and H2020 PANBioRA (EU760921) projects from the European Union. T.T. and J.T. were partially supported by Estonian Research Council (grant PRG805) and European Union through the European Regional Development Fund (Project No. 2014–2020.4.01.15–0012). J.T. was partially supported by the grant from Estonian Ministry of Education and Research (2014–2020.4.01.21–0315).

## Author contributions

A.R., M.J., H.S., N.P., A.P., M.A., A.M. and K.P. conceived and designed the study. M.A. and A.M. consulted the clinical data. A.R., M.J., H.S., N.P., A.P., J.T., and K.P. analyzed and interpreted the data. M.A., T.T., and K.P. supervised the study. A.R., A.G., and K.P. drafted the manuscript. All authors revised and approved the final manuscript for submission.

## Competing interests

A.P. and K.P. are inventors of the patent application (PCT Application No. US/14079626) filed by Protobios that covers the use of phage display method for manipulating and monitoring humoral immunity. All other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s43856-022-00114-7.

**Correspondence** and requests for materials should be addressed to Kaia Palm.

**Peer review information** *Communications Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022

# Appendix 2

**Publication II**

Jaago, M.*, **Annika Rähni***, Pupina, N., Pihlak, A., Sadam, H., Tuvikene, J., Avarlaid, A., Planken, A., Planken, M., Haring, L., Vasar, E., Baćević, M., Lambert, F., Kalso, E., Pussinen, P., Tienari, P. J., Vaheri, A., Lindholm, D., Timmusk, T., … Palm, K. (2022). **Differential patterns of cross-reactive antibody response against SARS-CoV-2 spike protein detected for chronically ill and healthy COVID-19 naïve individuals.** Scientific Reports, 12(1), 16817. https://doi.org/10.1038/s41598-022-20849-6

scientific reports

Check for updates

OPEN

# Differential patterns of cross-reactive antibody response against SARS-CoV-2 spike protein detected for chronically ill and healthy COVID-19 naïve individuals

Mariliis Jaago[1,2,18], Annika Rähni[1,2,18], Nadežda Pupina[1], Arno Pihlak[1], Helle Sadam[1,2], Jürgen Tuvikene[1,2,17], Annela Avarlaid[2], Anu Planken[3], Margus Planken[3], Liina Haring[4], Eero Vasar[5,6], Miljana Baćević[7], France Lambert[8], Eija Kalso[9,10], Pirkko Pussinen[11], Pentti J. Tienari[12], Antti Vaheri[13], Dan Lindholm[14,15], Tõnis Timmusk[1,2], Amir M. Ghaemmaghami[16] & Kaia Palm[1]✉

**Immunity to previously encountered viruses can alter response to unrelated pathogens. We reasoned that similar mechanism may also involve SARS-CoV-2 and thereby affect the specificity and the quality of the immune response against the virus. Here, we employed high-throughput next generation phage display method to explore the link between antibody immune response to previously encountered antigens and spike (S) glycoprotein. By profiling the antibody response in COVID-19 naïve individuals with a diverse clinical history (including cardiovascular, neurological, or oncological diseases), we identified 15 highly antigenic epitopes on spike protein that showed cross-reactivity with antigens of seasonal, persistent, latent or chronic infections from common human viruses. We observed varying degrees of cross-reactivity of different viral antigens with S in an epitope-specific manner. The data show that pre-existing SARS-CoV-2 S1 and S2 cross-reactive serum antibody is readily detectable in pre-pandemic cohort. In the severe COVID-19 cases, we found differential antibody response to the 15 defined antigenic and cross-reactive epitopes on spike. We also noted that despite the high mutation rates of Omicron (B.1.1.529) variants of SARS-CoV-2, some of the epitopes overlapped with the described mutations. Finally, we propose that the resolved epitopes on spike if targeted by re-called antibody response from SARS-CoV-2 infections or vaccinations can function in chronically ill**

[1]Protobios LLC, Tallinn, Estonia. [2]Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia. [3]North Estonia Medical Centre Foundation, Tallinn, Estonia. [4]Institute of Clinical Medicine, Psychiatry Clinic of Tartu University Hospital, University of Tartu, Tartu, Estonia. [5]Department of Physiology, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia. [6]Center of Excellence for Genomics and Translational Medicine, University of Tartu, Tartu, Estonia. [7]Dental Biomaterial Research Unit (d-BRU), Faculty of Medicine, University of Liege, Liege, Belgium. [8]Department of Periodontology and Oral Surgery, Faculty of Medicine, University of Liege, Liege, Belgium. [9]Department of Anaesthesiology, Intensive Care and Pain Medicine, Helsinki University Hospital, Helsinki, Finland. [10]SleepWell Research Programme, Department of Pharmacology, University of Helsinki, Helsinki, Finland. [11]Oral and Maxillofacial Diseases, Helsinki University Hospital, University of Helsinki, Helsinki, Finland. [12]Translational Immunology Research Program, Department of Neurology, Neurocenter, Helsinki University Hospital, University of Helsinki, Helsinki, Finland. [13]Department of Virology, Medicum, University of Helsinki, Helsinki, Finland. [14]Department of Biochemistry and Developmental Biology, Faculty of Medicine, University of Helsinki, Helsinki, Finland. [15]Minerva Foundation Institute for Medical Research, Helsinki, Finland. [16]Immunology and Immuno-Bioengineering Group, School of Life Science, Faculty of Medicine and Health Sciences, University of Nottingham, Nottingham, United Kingdom. [17]Present address: DXLabs LLC, Tallinn, Estonia. [18]These authors contributed equally: Mariliis Jaago and Annika Rähni. ✉email: kaia@protobios.com

**COVID-19 naïve/unvaccinated individuals as immunogenic targets to boost antibodies augmenting the chronic conditions. Understanding the relationships between prior antigen exposure at the antibody epitope level and the immune response to subsequent infections with viruses from a different strain is paramount to guiding strategies to exit the COVID-19 pandemic.**

**Abbreviations**

| | |
|---|---|
| ACE2 | Angiotensin-converting enzyme 2 |
| ADE | Antibody-dependent enhancement |
| AUROC | Area under the receiver operating characteristic |
| BC | Breast cancer |
| CAD | Coronary artery disease |
| CMV | Human cytomegalovirus |
| CoV | Coronavirus |
| CVD | Cardiovascular disease |
| Dot-ELISA | Dot enzyme-linked immunosorbent assay |
| EBV | Epstein-Barr virus |
| FEP | First episode psychosis |
| HCoV | Human coronavirus |
| HLA | Human leukocyte antigen |
| HSV-1 | Herpes simplex virus-1 |
| HT | Hypertension |
| mAb | Monoclonal antibody |
| MI | Myocardial infarction |
| MS | Multiple sclerosis |
| MVA | Mimotope-variation analysis |
| ND | Neuropsychiatric disorder |
| NP | Neuropathic pain |
| RBD | Receptor binding domain |
| RU | Relative unit |
| S | SARS-CoV-2 spike glycoprotein |
| S1 | S1 subunit of S |
| S2 | S2 subunit of S |
| SZ | Schizophrenia |
| T2D | Type 2 diabetes |

The coronavirus disease 2019 (COVID-19) pandemic has unveiled the pathogenicity of SARS-CoV-2 with surges by the currently prevailing SARS-CoV-2 variant B.1.1.529 (described first in 2021)[1] designated as Omicron displaying unusually large number of mutations and fast-spreading sublineages[2,3]. In general, clinical manifestations of SARS-CoV-2 infection range from asymptomatic and relatively milder, flu-like symptoms[4–6] to long-lasting complications known as the post-COVID syndrome or long COVID[7,8]. This complex clinical picture along with confirmative studies that vaccination protects against severe forms of disease[9] points to the immune system as a key factor in the control of SARS-CoV-2[10].

The relative manifestation of symptoms in infected is likely attributable to the partial protection conferred by the pre-existing immune memory. The preference of the immune system to recall existing memory cells, rather than stimulate a de novo response when encountering a novel but closely related antigen is referred to as immune imprinting, historically known as original antigenic sin[11]. Overall, immune imprinting would lead to enhanced immunity, whereas established pre-immunity may also increase cross-reactive antibody response towards epitopes that are shared between the current and the previously encountered antigen[12,13]. Studies have observed cross-reactivity between endemic common cold human coronaviruses (HCoVs) and SARS-CoV-2[14–16],whereas whether this cross-reactivity[17,18] is beneficial or detrimental to COVID-19 disease is not clear[14,19–25]. Cross-reactivity through heterologous immunity may arise through recognition of identical antigenic epitopes shared by different pathogens, or through recognition of unrelated epitopes owing to cross-reactivity of individual T and B cell receptors[26]. Shifts in antibody response to respiratory syncytial virus, cytomegalovirus (CMV) and herpes simplex virus-1 (HSV-1) were noted in patients with severe COVID-19[27]. Cross-protective effects of non-COVID-19 vaccines against SARS-CoV-2 are currently tested in clinical trials for polio, measles-mumps-rubella, influenza, and Bacillus Calmette–Guérin vaccines (rev in[28]) with promising pre-publication findings[29]. Collectively these data demonstrate that people at the stage of the current pandemic carry heterogeneous, immune-imprinted repertoires derived from their distinctive histories of infection and vaccination.

Given that the infections with SARS-CoV-2, in particular with its Omicron variants have become so common, it is likely that these confer boosting to the prior immune repertoire. Reports are emerging on findings of IgG autoantibodies in COVID-19 patients with a significant subset of patients developing new-onset autoantibodies[30–33] that could place them at risk for progression to autoimmunity. Studies reporting on myocarditis after receiving mRNA vaccines against COVID-19[34–38] suggest molecular mimicry between the vaccine product and self-antigens as an underlying mechanism[39].

Given that the relationships between prior antigen exposure, through infection or vaccination with SARS-CoV-2, and the immune responses to subsequent infections with emerging viruses is still incompletely understood, but is of paramount importance to exit the COVID-19 pandemic, we employed an unbiased approach

of next generation peptide phage display mimotope variation analysis (MVA)[40,41], to delineate cross-reactive immunity hallmarks on SARS-CoV-2 S glycoprotein in COVID-19 naïve subjects. Using samples from both COVID-19 naïve individuals and patients with a COVID-19 diagnosis, we identified pre-existing antibody response to multiple S protein sites by using recombinant S protein subunits of SARS-CoV-2 that was increased in patients with severe COVID-19 disease. Among these, three epitopes with cross-reactivity to SARS-CoV-2 S were features of underlying acute and/or chronic clinical conditions. Thus, highlighting the risk for chronic condition exacerbation following SARS-CoV-2 S protein exposure due to infection or vaccination.

## Star★methods

### Key resources table.

| Reagent or resource | Source | Identifier |
|---|---|---|
| **Antibodies** | | |
| Human IgG reference pool | Sigma-Aldrich | Cat#i4506 |
| Rabbit anti-human IgG (H&L) (HRP) | Abcam | Cat#ab6759 |
| **Biological samples (serum, plasma)** | | |
| COVID-19 | Tartu University Hospital | COVID-19 |
| Coronary artery disease | Helsinki University Hospital | CVD (CAD) |
| Myocardial infarction | North Estonia Regional Hospital | CVD (MI) |
| T2D and T2D with foot ulcers (DFU) | University of Liege | T2D |
| Multiple sclerosis | Helsinki University Hospital | MS |
| Breast cancer | Helsinki University Hospital | BC |
| First episode psychosis, schizophrenia | Psychiatry Clinic of Tartu University Hospital, Helsinki University Hospital | ND (FEP; SZ) |
| Healthy donor | North Estonia Medical Blood Centre | HC |
| **Chemicals, peptides, and recombinant proteins** | | |
| SARS-CoV-2 spike protein S1 subunit | Icosagen | Cat#P-305-100 |
| SARS-CoV-2 spike protein S2 subunit | Icosagen | Cat#P-306-100 |
| Ph.D.™-12 phage display peptide library (modified from original library) | New England Biolabs | Cat#E8110S |
| **Critical commercial assays** | | |
| Catalysed signal amplification (CSA) system II, biotin-free, HRP, DAB+ | Dako (Agilent) | Cat#K1497 |
| Anti-CMV ELISA (IgG) | EUROIMMUN | Cat# EI 2570-9601 G |
| Anti-EBV-CA ELISA (IgG) | EUROIMMUN | Cat# EI 2791-9601 G |
| **Software and algorithms** | | |
| SPEXS2 algorithm | Courtesy of Egon Elbre | https://github.com/egonelbre/spexs2 |
| ImageJ v. 1.53a | Schneider et al., 2012 | https://imagej.nih.gov/ij/ |
| RStudio v. 1.3.959 | RStudio Team, 2020 | https://www.rstudio.com |
| R "tidyverse" packages | Wickham et al., 2019 | https://doi.org/10.21105/joss.01686 |
| R "HPAanalyze" package | Tran et al., 2019 | https://doi.org/10.1186/s12859-019-3059-z |
| R "ggpubr" package | Courtesy of Kassambara | https://CRAN.R-project.org/package=ggpubr |
| R "ggbeeswarm" package | Courtesy of Clarke and Sherrill-Mix | https://CRAN.R-project.org/package=ggbeeswarm |
| MS Office Excel 2016 | Microsoft Corporation | https://www.microsoft.com |
| Adobe Photoshop CS4 version 11.0 | Adobe Systems Inc | https://www.adobe.com |
| **Other** | | |
| Protein G magnetic beads | New England Biolabs | Cat#S1430S |
| Immune epitope database | Immune Epitope Database | https://www.iedb.org |
| Human protein atlas v.20.0 | Uhlen et al., 2015[42] | https://www.proteinatlas.org |
| UniProtKB human reference proteome | [43] | https://www.uniprot.org/proteomes/UP000005640; ID: UP000005640 |
| SARS-CoV-2 proteome sequences | [44] | https://www.viralzone.expasy.org/8996 |

## Experimental model and subject details

**Ethics declarations.** The study was conducted in accordance with the guiding principles of the Declaration of 182 Helsinki and the study participants gave written informed consent before enrolment. The sample data on myocardial infarction (MI), breast cancer (BC) and schizophrenia (SZ) samples are described in detail in Pupina et al. 2022[45]. The use of multiple sclerosis (MS) and coronary artery disease (CAD) cohort samples was approved by the regional ethics committees (Dno 83/13/03/01/2013[40] and licence no 106/2007[46], respectively). Use of samples from type II diabetes (T2D) patients was approved by the Ethics Committee of University Hospital of Liege (permit no 2018/78). First-episode psychosis (FEP) cohort and samples from COVID-19 patients

| No | Cohort (abbreviation) | Study group | Group size (n) | Gender | Age (mean ± standard deviation (SD)), years | Origin | Case or ctrl | Previous studies |
|---|---|---|---|---|---|---|---|---|
| **Cardiovascular diseases (CVD, with comorbidities including obesity, hypertension, diabetes)** | | | | | | | | |
| 1 | CAD | No-CAD | 32 | 17M/15 F | 60.2 ± 8.3 | Finland | Ctrl | [46] |
| | | Stable-CAD | 32 | 26M/6F | 64.3 ± 8.4 | Finland | Case | [46] |
| | | Acute coronary syndrome | 32 | 24M/8F | 61.3 ± 8.2 | Finland | Case | [46] |
| 2 | MI | HC | 61 | 25M/36F | 46.2 ± 15.5 | Estonia | Ctrl | [45] |
| | | MI | 50 | 29M/13F/10 not available (NA) | 66.8 ± 12.6 | Estonia | Case | [45] |
| 3 | T2D and foot ulcer | T2D | 25 | 21M/4F | 69.0 ± 8.7 | Belgium | Case | None |
| **Autoimmune disease** | | | | | | | | |
| 4 | MS | MS | 20 | 4M/16F | 32.3 ± 7.8 | Finland | Case | [40,45] |
| **Cancer** | | | | | | | | |
| 5 | BC | BC | 57 | 0M/57F | 55.7 ± 7.2 | Finland | Case | [45] |
| **Neuropsychiatric disorders (ND)** | | | | | | | | |
| 6 | FEP | HC | 30 | 15M/15F | 24.0 ± 6.1 | Estonia | Ctrl | [47,48] |
| | | FEP | 30 | 16M/14F | 25.6 ± 4.9 | Estonia | Case | [40,47,48] |
| 7 | SZ | HC | 30 | 12M/18F | 42.1 ± 18.2 | Finland | Ctrl | [45] |
| | | SZ | 30 | 12M/18F | 41.9 ± 18.4 | Finland | Case | [45] |
| **Healthy blood donors** | | | | | | | | |
| 8 | Donors | HC | 109 | 64M/45F | 40.6 ± 11.6 | Estonia | Ctrl | [40,45] |

**Table 1.** Descriptions of clinical cohorts of COVID-19 naïve individuals.

were used with the approval of the Ethics Committee of the University of Tartu (licenses no. 362/T-3 and 312T-2, respectively) and have previously been studied in[47,48]. Healthy blood donor (HC) samples were procured from the North Estonia Medical Blood Centre (Tallinn, Estonia) and approved with study licence TAIEK no 1045.

**Clinical cohorts.** The antibody immune profiles were generated by MVA from SARS-CoV-2 naïve individuals (n = 538, Table 1), comprising patients with different clinical diagnoses (n = 276) and healthy controls (n = 262). COVID-19 naïve blood samples were collected prior to SARS-CoV-2 emergence (between April 2010 to Dec 2018) and chosen to represent a cross-section of the general uninfected, unexposed, unvaccinated population. Cardiovascular disease (CAD[46] and MI[45]) and neuropsychiatric disease (FEP and SZ[45,47,48]) samples are matched with control samples (Table 1) by the design of the original studies. Type II diabetes (T2D), multiple sclerosis (MS) and breast cancer (BC) sub-cohort patient samples have been analysed against selected controls samples (Table 1). In addition, time-series samples from six patients with COVID-19 diagnoses were analysed by MVA, and samples from two of these patients were analysed by dot-ELISA methods (Table S1). These samples were initially collected in the hospital emergency room (ER) and consecutive time-series samples were collected in the stationary care unit during disease progression or recovery process.

## Method details

**Mimotope-variation analysis.** For qualitative and quantitative characterisation of humoral immune response, we used an in-house developed mimotope-variation analysis (MVA) method as described previously[40,41]. In brief, modified random 12-mer peptide phage library (Ph.D.-12, New England Biolabs) was used. Two µl of serum/plasma samples was precleared to plastic and *E. coli*/wild type M13 phage particles. After preclearing, serum/plasma samples were incubated with 2.5 µl library (~ $5 \times 10^{11}$ phage particles) and phage-immunoglobulin G (IgG) antibody complexes were recovered using protein G-coated magnetic beads (S1430S, New England Biolabs). Captured phage DNA was analysed by next generation sequencing (Illumina HiSeq, 50-bp single end reads) with barcoded primers for sample multiplexing. To evaluate the reproducibility of the data, we compared peptide abundance in two replicates using Pearson correlation coefficient, which was 0.985 ($p < 0.0001$) (Fig. S1). Obtained sequences were bioinformatically cleaned of sequencing errors and known artefacts, yielding peptide profiles for further analysis.

**Clustering of immunodominant peptide antigens.** Peptides in a sample dataset cleaned of sequencing errors and known artefacts were normalised to 3 million reads (RPM units). SPEXS2 exhaustive pattern search algorithm[41] was used to cluster similar peptides and reveal recognition patters (epitope consensuses) that were enriched in studied peptide sets. The identification of core consensus sequences was either performed discriminatorily (see below for details), where selected peptide sets were studied for enrichment compared to control group, or non-discriminatorily (see below for details), where enrichment was identified compared to a random-generated peptide set (hypothesis-free). The decision of whether to compare sample peptides to random-generated reference peptide sets or to peptides from matched controls samples was determined by the clinical question of the study[40,45,46]with parameters chosen to maximise the number of epitope consensuses in the output. In brief, for each clinical group, 100,000s of distinct peptide antigens were used as input, from which

1000s of distinct core epitopes were identified with SPEXS2 algorithm. The output of distinct core epitopes generated for each of the sub-cohorts (CAD, MI, T2D, BC, FEP, SZ, MS) were combined together and further analysis was carried out across the whole study cohort. Altogether a total set of core epitopes (n = 22,949) was generated, for the entire cohort including data from samples of different age groups, genders and clinical background. All the 22,949 consensus epitope sequences were aligned to primary sequence of the SARS-CoV-2 spike glycoprotein (S) (UniProtKB code: P0DTC2).

**Non-discriminatory clustering.** The generation of distinct consensus epitopes (n = 8088) for CAD cohort is described in detail in[46]. The generation of distinct consensus epitopes (n = 4019) for MI cohort is in[45].

For analysis of T2D with foot ulcer condition, most abundant and shared peptides from MVA immunoprofiles were selected with criteria: peptide must be present in ≥ 10 repeats in one sample; and must be present in ≥ 2 samples. The resulting peptide set was compared with random-generated peptide set of same length and enriched consensuses were identified using SPEXS2 algorithm, using hypergeometric p-value < $10^{-7}$ and motif to be present in ≥ 4 distinct peptides. As a result, epitopes (n = 1169) with ≥ 4 fixed amino acid positions were identified.

For cancer-related immunoprofiles, patients with surgically removed BC (denoted as BC), top 900 most abundant peptides from each sample were screened separately for epitope consensus identification (≥ 5 fixed amino acid positions). These identified from ≥ 2 samples were extracted for further analysis and motifs more detected in BC were selected. Additionally, peptides from a large non-unique dataset (shared in ≥ 2 samples) that did not contain any consensuses identified by the above-mentioned approach, were examined separately. Peptides present in ≥ 4 samples (> 10%) in BC group were extracted and common consensus sequences were identified with SPEXS2 tool (with hypergeometric p-value < $10^{-3}$, ≥ 5 fixed amino acid positions, epitope consensus to be present in ≥ 4 unique peptides). Altogether, distinct consensus epitopes (n = 1,014) were selected for further analysis.

**Discriminatory clustering.** Within FEP and SZ cohorts, most abundant and shared group-specific peptide antigens were extracted for each of the control and case groups, based on criteria: the chosen peptide must be present in ≥ 10 repeats in one sample; and must be present in ≥ 10% of the samples within the given group. Core consensus epitopes were defined for each group independently with SPEXS2 as motifs that were more enriched when compared to at least one of the 3 different reference sets: (1) random-generated reference set of same length, where amino acids were scrambled within-peptides; (2) random-generated reference set of same length, where amino acids were scrambled within amino acid positions- and across peptides; and (3) the top abundant peptide set from the control group. Distinct consensus epitopes identified from SPEXS2 analyses were selected, based on criteria: hypergeometric p-value < $10^{-6}$ (FEP), < $10^{-5}$ (HC of FEP), < $10^{-8}$ (SZ), or < $10^{-6}$ (HC of SZ); consensus epitope present in ≥ 4 distinct peptides; ≥ 4 fixed amino acid positions. Epitope consensuses matching these criteria were selected for FEP (n = 228), SZ (n = 1785), HC of FEP (n = 1935), HC of SZ (n = 760).

Within MS group, subgroups with or without initial optic neuritis diagnosis were analysed separately. The topmost abundant peptides from both subgroups were extracted with criteria: peptide count ≥ 5 in ≥ 1 sample. Using SPEXS2 the peptide sets were compared to top peptide set of age- and sex-matched controls with no MS diagnosis[40] using criteria: hypergeometric p-value < $10^{-7}$; epitope consensus to be present in ≥ 4 distinct peptides and have ≥ 4 fixed amino acid positions. As a result, distinct epitopes (n = 3,500) were identified for MS group.

**Epitope prediction of SARS-CoV-2 S.** The primary sequence of SARS-CoV-2 S (P0DTC2) was obtained from www.viralzone.expasy.org/8996 (date accessed 25.03.2020). To predict immunogenic regions on S, we followed two complimentary alignment approaches. First, primary sequence of SARS-CoV-2 S protein was scanned with distinct epitope consensus (n = 22,949) with the criterion of ≥ 4 exact-matching amino acid positions. Random reference profile was generated by scanning with shuffled-sequence motifs (≥ 4 exact positions) in order to qualitatively assess alignment enrichment on local protein regions. An additional random alignment distribution was simulated with 3 independent alignments with scrambled motifs in order to assess statistical significance of specific alignment results. Specific alignments with values above 97.5th percentile (two-tailed) of simulated distribution were considered statistically significant (*p < 0.025, **p < 0.005, ***p < 0.0005). Two regions (amino acid positions 623–633 and 708–713) were excluded based on lower count of unique peptide epitopes (< 200) or insufficient specific/random ratio (< 3 positions with ratio ≥ 2) values. Secondly, to identify regions with less prevalent yet sufficiently high alignment enrichment results, an additional approach was taken. For the epitope consensuses, the number of fixed amino acids for alignments was 4. Therefore, on a protein region of 20 amino acids, the probability of exact match with this region was $17/20^4 = 1.1 \times 10^{-4}$. With an input of all epitope consensuses (n = 22,949), the number of theoretical exact alignments to any 20-amino acid region would be 2.4. Therefore, to identify potential immunogenic regions, the threshold for significance was set to ≥ 3 of exact-aligned epitope consensuses. Based on this criterion, epitopes S1.4 and S2.1 were included with the other described immunogenic regions. The composition of top 10 most abundant peptides representing the 15 defined S epitopes across the cohort of 538 samples is shown in Table S2.

Alignment profiles on the primary sequence of SARS-CoV-2 S protein were generated using custom Excel VBA scripts and MS Office Excel, and visualised using R "tidyverse" packages[49]. The data was visualised using centred weighted moving averages across 9 amino acid positions. The displayed value (*value*) was calculated per each amino acid position ($n$), taking into account the raw values ($a$) of given and four adjacent positions in both directions and multiplying those with weights (from 1 to 5):

$$value = \frac{1 * (a_{n-4} + a_{n+4}) + 2 * (a_{n-3} + a_{n+3}) + 3 * (a_{n-2} + a_{n+2}) + 4 * (a_{n-1} + a_{n+1}) + 5 * a_n}{25}$$

**Sequence alignment.** *Individual sample peptide alignment.* For the analysis of individual peptide alignments of samples that were used in dot-ELISA (n = 9), all peptide sequences underlying the 15 MVA-predicted epitopes were aligned to SARS-CoV-2 S protein primary sequence, taking into account the different abundance values of peptides in individual immunoprofiles. Alignment was performed with ≥ 4 matching amino acid positions, ratio of specific/random alignment (with shuffled peptide sequences) was calculated for each amino acid position on SARS-CoV-2 S protein primary sequence, and data was visualised as heatmaps using centred weighted moving averages of ratios across nine amino acid positions.

*Human pathogen and human proteome alignment.* Reference proteins of common human viruses were obtained from UniProtKB with keywords "taxonomy: "Viruses [10239]"" + "host: human" + "reviewed: yes" (date accessed: 10.11.2020, altogether 6114 viral proteins). Human reference proteome was obtained from UniProt Proteome database (Proteome ID: UP000005640, date accessed: 16.11.2020, altogether 75,069 sequences).

Constructed viral protein and human protein databases were first scanned with 15 SARS-CoV-2 S protein epitope representative motifs. Using custom Excel VBA scripts a reference set of viral and human proteins that contained at least 1 predicted S protein epitope was generated. 7.5% (458) of human-associated virus proteins and 6.3% (4,968) of human proteome sequences matched ≥ 1 MVA predicted S protein epitope(s). Next, the primary sequences of reference set proteins from Swiss-Prot (excluding TrEMBL) were aligned with MVA captured peptides containing these 15 S epitopes (900 most abundant peptides per each defined epitope across 538 samples) by using standalone BLAST + *blastp* function (v. 2.11.0) with the following criteria: (i) word size: 2, (ii) gap open penalty: 9, (iii) gap extend penalty: 1, (iv) substitution matrix: PAM30, (v) threshold: 11, and by disallowing composition-based statistics correction. E-value threshold for human virus proteins was set to 200,000, while E-value threshold for human proteome alignment was set to 2000. Output was further curated to select for the highest scoring target alignments of the 15 epitopes of SARS-CoV-2 S based on E-value (< 0.05, except for human coronavirus proteins, where E-value > 0.05 was allowed) and identity (> 48%, i.e. at least 5 matching amino acid positions in the 12-mer peptide for viral proteins and > 52%, i.e. at least 5 matching amino acid positions from the 12-mer peptide for human proteins) (Tables S3 and S4).

The alignment data were visualised using R statistical programming (v. 4.0.2, https://www.R-project.org/) and RStudio environment (http://www.rstudio.com/). Violin plots were produced and visualised using R "tidyverse" packages[49].

Human Protein Atlas (HPA, v. 20.0, http://www.proteinatlas.org)[42] gene expression data was used to visualise the expression of human proteins that aligned with SARS-CoV-2 S protein epitope peptides. "HPAanalyze" R package[50] was used to access gene expression data of normal tissue from HPA repository. Expression data deemed "Uncertain" by the "Reliability" parameter were excluded and data was visualised with "ggplot2"[51].

**Modelling immune response patterns in COVID-19 naïve subjects.** IgG-bound peptide abundance values of 15 epitopes were calculated for each sample in the COVID-19 naïve cohort (n = 538) and normalised per epitope with the 97.5% percentile value (to allow for comparison of epitopes). Subjects with a relatively higher response (normalised value > 0.5) to at least two epitopes on S protein (of the 15) were grouped into the "high" group, whereas others to the "low" group.

**Visualisation of data from published data sets.** B cell epitopes for the SARS-CoV-2 S protein for naïve individuals were derived from Grifoni et al. study[52], for COVID-19 from several published studies[27,53–57]. Amino acids of SARS-CoV-2 S protein important for binding of human ACE2 have been reported by different groups[58] and anti-spike RBD neutralising antibody CR3022 discontinuous epitope were reported by ter Meulen[59] and Rogers and colleagues[53]. Domains of SARS-CoV-2 S protein were accessed from[54,60]. Raw data of PEPperCHIP® SARS-CoV-2 Proteome Microarray is from Schwarz et al., 2021[61]. Reported epitopes in Immune Epitope Database (IEDB) were accessed on 11.07.2022 with the following criteria: *Organism*—SARS-CoV2 (ID:2697049); *Antigen*—Spike glycoprotein (PODTC2); *B cell Assay*—Outcome "Positive" and Neutralization" (Table S5).

**Dot-ELISA (CMV and EBV epitopes).** Independent validation experiments for IgG immunoreactivity measured by MVA were performed with dot-ELISA experiments. Peptides containing CMV- and EBV-specific epitopes previously described in[40] were printed as follows. Recombinant phages displaying in the N-terminus of the pIII of M13 peptides of interest, specifically (TLPMDTSPRAHW containing epitope of viral capsid antigen (VCA) p18 of Epstein-Barr virus (EBV)) (*specific*) and TLPMDASPRAHW (control). In addition, peptides containing epitope of glycoprotein B (gB) of human cytomegalovirus (CMV) NETIYNTTLKYGGGGDYKDDD(LYS(BIOTIN)) were synthesised by Genescript (US). For dot-ELISA, peptides or peptide-displaying phages printed onto nitrocellulose filter pads (Amersham Bioscience) as duplicates by SpotBot® 4 (Arrayit) were exposed to human precleared sera/plasma (dilution 1:100 for MS sub-group samples and 1:50 for CAD sub-group samples) for 1 h at room temperature and then incubated with rabbit anti-human IgG (HRP) (Abcam, ab6759; dilution 1:1000) as a secondary antibody. Images were scanned using EttanTM DIGEImager (GE Healthcare Life Sciences). All printed dot intensities were calculated and analysed further as averages of duplicates. With CAD samples, n = 28 samples were assayed for seroreactivity to epitope on gB of CMV and n = 54 samples for VCA p18 of EBV. Seroreactivity to epitope on gB of CMV was calculated by subtracting the background intensity of the dot and displayed in arbitrary units (AU). Seroreactivity to epitope on VCA p18 of EBV was calculated as the difference between signals of specific vs control phage (signal difference, in AU) in the MS cohort, whereas in the CAD cohort as the ratio between signals of specific vs control displayed epitopes (signal ratio, in AU).

**Dot-ELISA (spike subunits).** Spike protein fragments of S1 subunit (amino acids 14–681, cat no. P-305–100), S2 subunit (aa 693–1218, cat no. P-306–100) and RBD (aa 319–541, cat no. P-307–100) (Icosagen AS) were diluted in 1× PBS (pH 7.4), blotted onto nitrocellulose membranes (50 ng of protein per dot) (Amersham Biosciences, cat no. RPN 1520D) and blocked with 5% non-fat dried milk powder (DMP) (Applichem, cat. no. A0830) in 1xPBS-0.05%-Tween20 for 1 h at room temperature. For linear epitope analysis, spike protein subunits S1, S2 and RBD were denatured in 4 M Urea (Applichem, cat. no. A8113) for 1 h at room temperature prior to blotting. Sera/plasma samples (1:50 dilutions) were pre-treated overnight with 1:2 solution of *E. coli* phage lysate in 2.5% DMP in 1xPBS-0.05%-Tween20 at + 4 °C to reduce non-specific signals. Blocked nitrocellulose membranes (5% DMP in 1xPBS-0.05%-Tween20, 1 h at RT) were washed with 1xPBS-0.05%-Tween20 and incubated for 4 h at room temperature with human sera/plasma sample dilutions to form SARS-CoV-2 S protein-specific immunoglobulin and S subunit complexes. After washes, membranes were incubated with 1:1000 rabbit anti-human IgG HRP-conjugated secondary antibody (Abcam, cat no. ab6759) in 2.5% DMP in 1× PBS-0.05%-Tween20 for 1 h at room temperature. Signal amplification system (Dako, CSA II System kit, cat no. K1497) was used for sensitivity enhancement with reagents diluted to either 1:10 for amplification reagents or to 1:100 for DAB chromogen in 1× PBS-0.05%-Tween20 buffer. Membranes were digitally scanned and signals quantified using ImageJ software (version 1.53a)[62]. The averages from two technical replicate dots were calculated per each experiment (n = 3) and intensity values were represented as proportional to a COVID-19 positive case (P1#1, Table S1). Averages ± SEM of three independent experiments are represented on figures.

**ELISA.** Anti-CMV and anti-EBV serostatus was measured from serum/plasma samples with ISO/IEC-17025:2017 accredited methods. In brief, serological analyses were performed with anti-CMV IgG ELISA method (EUROIMMUN EI 2570-9601 G) and with anti-EBV CA IgG ELISA method (EUROIMMUN EI 2791-9601 G) according to manufacturer's specifications. Absorbance was measured at 450 nm with SpectraMax Paradigm (Molecular Devices). Altogether serum or plasma samples of 199 subjects from the clinical cohorts of healthy donors (n = 83), MS (n = 20), or CAD (n = 96) were analysed for anti-CMV IgG antibodies and 241 subjects from the clinical cohorts of healthy donors (n = 19), MS (n = 20), FEP (n = 60), SZ (n = 46), or CAD (n = 96) were analysed for anti-EBV CA IgG antibodies.

## Quantification and statistical analysis

**Statistical analysis.** All statistical analyses (Pearson correlation calculation, Mann–Whitney U test, multiple linear regression analysis) were done using R statistical programming and RStudio environment (https://www.R-project.org/; URL: http://www.rstudio.com/). Boxplot graphs were produced and visualised using R "tidyverse" packages[49], "ggbeeswarm", and "ggpubr"[63,64]. On boxplots, the upper, middle and lower boxplot lines represent the 75th, 50th and 25th percentiles, while whiskers represent the largest or smallest value within 1.5 times interquartile range above the 75th percentile or below the 25th percentile, respectively and *outer dots* indicate outliers. Reported p-values were not adjusted for multiple comparisons, as this study was viewed as a hypothesis-free approach with an emphasis on discovering new relationships in a retrospective cohort.

## Results

**Cross-reactive immune response to SARS-CoV-2 spike protein in COVID-19 naïve people.** We used a high throughput random peptide phage display method (MVA)[33,34] to investigate potential cross-reactive antibody epitopes on SARS-CoV-2 S antigen in a cohort of SARS-CoV-2 unexposed (= COVID-19 naïve, also unvaccinated) individuals (n = 538, Table 1). Our discovery cohort of COVID-19 naïve included sera samples collected before 2017 from both, healthy individuals (Ctrl) and people diagnosed with various acute illnesses and chronic conditions (Case) to reflect the overall diversity of general population (Table 1). The mean age in case sub-cohorts of adults varied from 24 to 69 years, and the proportion of men varied between 20 and 84% for most sub-cohorts (Table 1).

Using MVA we defined 15 highly antigenic epitopes on the S protein, of which ten were on subunit 1 (S1) and five on subunit 2 (S2) (Fig. 1 and Table 2). The majority of these epitopes (epitopes S1.1 to S2.4) were exposed on the exterior surface of SARS-CoV-2 S trimer (Fig. S2A). Seven of the 15 identified epitopes were partially overlapping with epitopes previously reported for COVID-19 unexposed individuals[52] with an average overlap of 60% per epitope (Table 2). Epitopes S2.2 and S2.5 colocalised precisely with antigenic determinants reported by others, whereas epitope S1.8 extended to a more C-terminal region (amino acids 570–582) (Table 2). Furthermore, almost half of resolved epitopes (S1.8, S1.9, S1.10, S2.1, S2.2, S2.3, and S2.5) mapped to antigenic regions of S against which immune response has been detected in asymptomatic, mild, and severe COVID-19 cases (Table 2 and[27,55–57]). In good agreement with published data from SARS-CoV-2 proteome-based peptide arrays[61], we found that linear peptides from these studies that contained our resolved epitopes on S protein showed differential seroreactivity in naïve, mild and severe COVID-19 samples (Fig. S3).

Epitopes S1.1 to S7 locating to the N-terminal region of S1, encompassed the antigenic sites against which antibody response was detected by others in COVID-19 patients[27,55–57,61,65] and also its neutralising effects (Table 2 and Table S5). Interestingly, three epitopes (S1.1 to S1.3) were encompassing the anti-NTD supersite of S with neutralisation activity[66–68]. Four epitopes (S1.4 to S1.7) were identified in the receptor binding domain (RBD, amino acids 319 to 542) of SARS-CoV-2 S protein, spanning amino acid residues (319 to 542) that are involved in angiotensin-converting enzyme 2 (ACE2) binding[53], and targeted by neutralising antibodies[69,70] (Fig. 1, Fig. S2B). The epitope S1.5 overlapped at the I468 residue with the binding site of highly neutralising antibodies which showed good breadth against SARS-CoV-2 variants but not against BA.2.12.1 and BA.4/BA.5 Omicron sublineages with L452 mutations[71]. Additionally, epitopes S1.1, S1.3-S1.8, S1.10 and S2.2 encompassed antigenic regions with SARS-CoV-2 S neutralising activity (Tables 2 and S5) with S1.6 with overlapping in the E484 residue

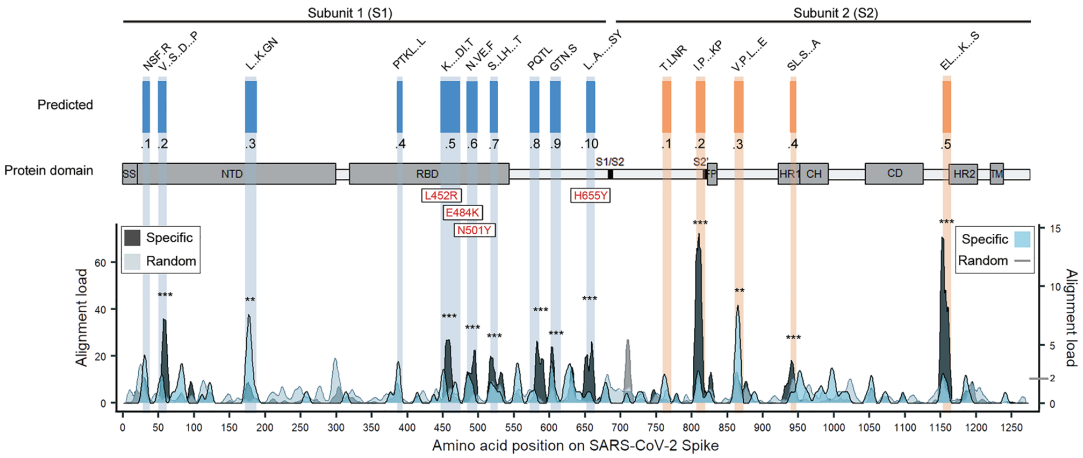



**Figure 1.** MVA-defined epitopes on SARS-CoV-2 spike protein. Alignment profiles of the most abundant and common immune response features from MVA immunoprofiling data of COVID-19 naïve subjects (n = 538), including altogether 22,949 unique epitopes characterising core motif sequences (hypergeometric p-value < $10^{-3}$ for core epitope recognition) on the primary sequence of SARS-CoV-2 S protein (UniProtKB code: P0DTC2). The abundance of aligned core epitope sequences (black Specific, primary y-axis) in peaks was significantly higher (**p < 0.01, ***p < 0.001) compared to random alignment (light gray random, primary y-axis). Of aligned core epitopes, 111 were with exact matches in all amino acid positions (blue specific, secondary y-axis). Regions of primary sequence with alignment load of > 2 motifs (the calculated theoretical random) (gray random line, secondary y-axis) were considered as potentially immunogenic and included in further analysis. Alignment profiles were visualised as centred moving averages across 9 amino acids. 15 epitopes (designated S1.1….10 on subunit 1 of S (S1) and S2.1….5 on subunit 2 of S (S2)) from MVA data analysis (predicted) were defined on SARS-CoV-2 S protein (x-axis) with representative consensus sequences shown. Common mutations observed in emerging SARS-CoV-2 variants were mapped to MVA-identified S epitopes and highlighted with red labels. Most common variants with enhanced transmissibility of SARS-CoV-2 variants with L452R, E484K/Q, and H655Y mutations[99,109,123] encompass epitopes in S1.5, S1.6, or S1.10 respectively. Primary y-axis—abundance of specific- vs random-aligned sequences with ≥ 4 matching positions (includes also exactly-matching sequences). Secondary y-axis—abundance of aligned motif sequences with all positions matching. SARS-CoV-2 S protein domains are adapted from Wrapp et al.[54] with additional information on RBD from Yuan et al.[53]. *SS* signal peptide [1–12]; *NTD* N-terminal domain (13–303); *RBD* receptor binding domain (319–542); *S1/S2* S1 subunit end and S2 start site (683–686); *S2'* S2' protease cleavage site (815–816); *FP* fusion peptide (816–833); *HR1* heptad repeat 1 (908–985); *CH* central helix (986–1035); *CD* connector domain (1076–1141); *HR2* heptad repeat 2 (1163–1202); *TM* transmembrane domain (1214–1234).

of the Omicron BA.1 escape mutant (E484A)[2]. A few of the resolved epitopes (S1.9, S1.10 and S2.5) included N-glycosylation sites ([72] and Table 2). However, some data demonstrate that glycosylation is not essential for serorecognition of linear epitopes in spike upon COVID-19 infection[73]. Of note, some of the epitopes, including with neutralising/protective activity, embedded the common mutations of the emerging Omicron variants (Fig. 1, Tables 2 and S5).

Collectively, these data suggest that IgG antibody responses to distinct epitopes of SARS-CoV-2 S protein is common across the naïve population and reactivity to the same antigenic regions is detected by serostudies of COVID-19 patients (Table 2 and ref in the Table 2).

**Epitopes on SARS-CoV-2 S protein identified in COVID-19 naïve sera are linked to heterologous pathogens.** Next, we wanted to know whether cross-strain or cross-species immunity could be behind the observed epitope-specific pre-existing anti-SARS-CoV-2 S immunoreactivity. Sequence alignment analysis across human viral antigens resulted in frequent detection of other human coronaviruses (HCoVs, including SARS-CoV, OC43 and HKU) (Fig. 2A, Table S3). In addition, significant homology of the resolved S epitopes was also observed with common herpes-, papilloma-, and respiratory (including influenza) viral antigens (Table S3). For example, antigens of human cytomegalovirus (CMV) and of Epstein-Barr virus (EBV), shared significant similarity with peptides containing epitopes S1.8 and S2.2 (Fig. 2A, Table S3) and these epitopes similar to CMV and EBV seroprevalence were also associated with age (Fig. S4A). Diagnostic serology measurements confirmed CMV and EBV seropositivity in analysed samples (Fig. S5A). However, in samples with CMV and EBV serology findings differential epitope-specific anti-S antibody response was evident (Figs. S6, S7), suggesting that herpesviral antigens can be direct molecular mimics of S antigenic determinants or indirectly associated with the heterologous immunity towards SARS-CoV-2 S. Epitopes S1.10 and S2.5 showed higher antibody responses in CMV + samples of both Ctrl and Case groups when compared to CMV- samples (Fig. S6), while seroresponse

| Epitope identification (a) | Amino acid position (b) | Sequence (c) | Representative epitope (d) | Data on antigenic regions from other studies (e) | | | | S mutants of variants of SARS-CoV-2 (f) |
|---|---|---|---|---|---|---|---|---|
| | | | | Bio-informatic predictions in COVID-19 naïve | Serostudies | | IEBD: B cell neutralising antibody response | |
| | | | | | COVID-19 naïve | COVID-19 diseased | | |
| **S1 subunit** | | | | | | | | |
| S1.1 | 26–34 | PAYTNSFTR | NSF.R | | | | P26 | P26S[124] |
| S1.2 | 47–58 | VLHSTQDL-FLPF | V..S..D…P | | | | – | Q52R[124] |
| S1.3 | 170–185 | YVSQPFLM-DLEGKQGN | L..K.GN | | | | 178–185 | |
| S1.4 | 384–390 | PTKLNDL | PTKL..L | | | | 384–390 | |
| S1.5 | 445–471 | VGGNYNY-LYRLFRKSNL-KPFERDISTE | K….DI.T | 469–483[52] | | | 445–471 | G446S[O:A1,A3][2,71] L452M[OA2][71] L452Q[O:A2][71] L452R[O:A4,5][71] L452R/Q*[127] G446S[O][2] |
| S1.6 | 481–495 | NGVEGFNCYF-PLQSY | N.VE.F | 469–483[52] | | | 481–495 | E484K/Q*[123,127] E484A[O][2] F486V[O:A5][71] F490S[124] Q493/K[O][2] |
| S1.7 | 514–523 | SFELLHAPAT | S…LH…T | | | | 514–523 | |
| S1.8 | 570–582 | ADTTDAVRD-PQTL | PQTL | | 550–570[27] | 553–570[55] N** 550–570[27] 532–588[27] 560–616[27] | A570 | A570D[124] |
| S1.9 | 599–612 | TPGT**N**TSNQ-VAVLY | GTN.S | 592–620[52] | | 588–644[27] 560–616[27] | – | |
| S1.10 | 650–660 | LIGAEHV**N**NSY | L..A…..SY | 652–661[52] | | 655–672[56] N** 616–672[27] | H655 | H655Y[128*O][124] N658S[O:A5][71] |
| **S2 subunit** | | | | | | | | |
| S2.1 | 757–768 | GSFCTQLN-RALT | T.LNR | 757–769[52] | | 766–785[27] | – | N764K[O][125] |
| S2.2 | 804–815 | QILPDPSKPSKR | I.P…KP | | 810–816[57] 785–805[27] 810–830[27] | 809–826[55] N** 787–822[56] 810–816[57] 785–805[27] 810–830[27] 812–868[27] | – | |
| S2.3 | 858–869 | LTVLPPLLT-DEM | V.P.L…E | 867–880[52] | | 812–868[27] | – | T859N[126] |
| S2.4 | 937–944 | SLSSTASA | SL.S…A | 915–946[52] | | | – | |
| S2.5 | 1151–1161 | ELDKYFK**N**HTS | EL….K…S | 1157–1164[52] | 1146–1166[27] | 1147–1158[56] 1146–1166[27] | 1148–1158 | |

**Table 2.** The epitopes identified by MVA on SARS-CoV-2 spike glycoprotein are validated by using external data showing partial overlap with antigenic domains reported for COVID-19 naïve and diseased and with neutralising epitopes. The following information is given in columns: unique identification (a), amino acid position (b), amino acid sequence with glycosylation patterns bolded and underlined (c), representative epitope consensus sequence (d), immunogenic regions with amino acid positions on S indicated from other informatic and seroreactivity studies and from IEBD data on B cell neutralising antibody response (see Table S5) (e), frequent mutations described in new variants of being monitored (VBM) and of concern (VOC) of SARS-CoV-2 S[122], where the highly mutated Omicron (B.1.1.529) sublineages (designated with "O") that have enhanced transmissibility and mutations that show differential (often escape from) neutralising antibody response (marked with "*") are shown in (f) as *A1, A2, A3, A4, A5*—Omicron sublineages, BA1, BA2, BA3, BA4 and BA4 respectively.[2,71,99,109,122–126]. *N* neutralising/protective antibodies shown, **, high IgG titre was associated with poor clinical outcome (development of pneumonia, needing care in the intensive unit or needing assisted pulmonary ventilation).

to epitopes S1.6, S1.8, S1.9 and S2.1 was significantly higher (S1.8, S1.9, S2.5) or lower (S1.6) in EBV + samples of Case group when compared to EBV + Ctrl group (Fig. S7). High sequence similarity was also found between epitopes of SARS-CoV-2 S protein and antigens of influenza A H1N1 (FLUA), respiratory syncytial virus type B (HRSV-B), rhinoviruses 2/16 (HRV-2/16), adenovirus A type 12 (HAdV-A) and most frequent papillomaviruses (HPV6 and HPV11, Fig. 2A, Table S3). By using dot-ELISA, we independently validated the IgG antibody response detected by MVA at a peptide level to common epitopes of CMV glycoprotein B and EBV VCA p18 proteins[40] (Fig. S5B). Collectively, our data conclude that heterologous immunity between epitopes of various common human viruses and SARS-CoV-2 S can be common.

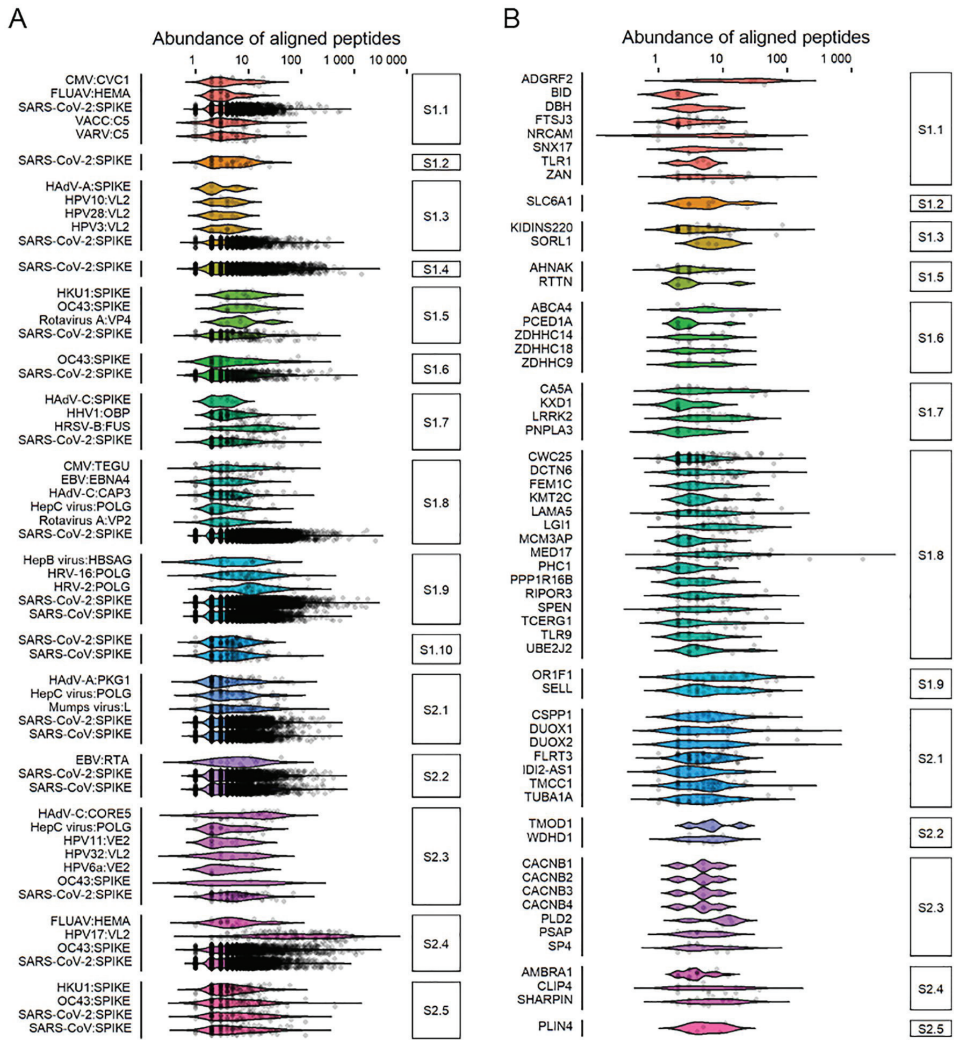



**Figure 2.** 15 SARS-CoV-2 S protein epitopes mimic common viral protein antigens and self-proteins implicated in normal development and disease. Pathogen proteins and human proteome were accessed from UniProtKB and aligned with 13,500 most abundant IgG-bound 12-mer peptides containing one representative SARS-CoV-2 S protein epitope (from S1.1 to S2.5) using standalone BLAST with alignment criteria customised to short sequences. (**A,B**) Viral and human proteins mimicked by SARS-CoV-2 S protein epitopes are depicted on violin plots and were identified using *blastp* alignment analysis at E-value ≤ 0.05 (except for visualising SARS-CoV, SARS-CoV-2, HKU1, and OC43 alignments where E-value was not restricted). Each dot represents the relative abundance of IgG response to a peptide in one sample from the cohort of SARS-CoV-2 naïve subjects (n = 538). See Tables S3 and S4 for detailed information on pathogen and human protein alignments, full protein names and accession codes.

Recent evidence also suggests that the existence of pre-COVID-19 autoimmunity plays a role in disease outcome[30,74]. Therefore, we focused on the human proteome and identified at least 63 human proteins with highly similar antigenic determinants to resolved epitopes of SARS-CoV-2 S (Fig. 2B, Table S4). Furthermore, analysis of the Human Protein Atlas expression data[42] revealed that these proteins were differentially expressed with several of them displaying central nervous system and immune system specificity (Fig. S8). For example, epitope S1.3 mapped to kinase D-interacting substrate of 220 kDa (KIDINS220, Fig. 2B, Table S4), which has a crucial role in neuronal and cardiovascular development[75]. Relatedly, immune responses against epitope S1.3 were stronger in COVID-19 naïve men with heart disease (Fig. S4B). Epitope S1.7 showed mimicry to leucine-rich repeat serine/threonine protein kinase 2 (LRRK2; Fig. 2B, Table S4), which is associated with Parkinson's and inflammatory

bowel diseases[76,77]. High sequence similarity was found between epitope S2.2 and tropomodulin 1 (TMOD1) that is linked to synaptogenesis, chronic pulmonary disease and cancer[78]. Interestingly, epitope S2.3 that showed differential immunoreactivity in COVID-19 naïve smokers and women with hypertension (Fig. S4C,D) shared high sequence similarity with proteins associated with hypertension[79], acute respiratory distress syndrome, periodontitis in smokers[80,81], but also with proteins involved in the development of the nervous system[81], and psychiatric disorders[82] (Fig. 2B, Table S4). Collectively, our data show that distinct seroresponse to epitopes of SARS-CoV-2 S protein in COVID-19 naïve could accommodate cross-reactive targets of B cell response against both, viral pathogens and self-proteins.

**Epitopes on spike protein identified in COVID-19 naïve are differentially targeted by antibodies in COVID-19 diseased.** Next, we randomly picked samples (n = 8) from the COVID-19 naïve cohort with differential epitope profiles (Fig. 3A) along with the pooled IgGs of healthy subjects (n = 2700) (Fig. 3B,C, Table S1) to test for the presence of anti-S seroreactivity by dot-ELISA. Samples from patients from intensive care unit with severe COVID-19 disease (n = 2, COVID-19 +) were included for reference (Fig. 3B,C, Table S1).

The dot-ELISA signals detecting IgG response to recombinant spike subunits in both COVID-19 naïve and COVID-19 + samples were at similar value range (Fig. 3B, Fig. S9). Notably, anti-SARS-CoV-2 S protein seroreactivity was detected in all studied COVID-19 naïve samples (n = 9) (Fig. 3B) with more cross-reactivity to S2 subunit as compared to S1 (S2/S1 signal ratio > 1), although one naïve individual (#1) showed a reversed pattern (S2/S1 signal ratio < 1, Fig. 3C).

Next, we analysed samples from patients who had been admitted to the hospital with the COVID-19 diagnosis (Table S1). Detailed, epitope-specific analysis of time-series samples from SARS-CoV-2 infected patients (n = 6) and matched controls (n = 6) was performed using MVA. While reactivity to epitopes S1.1 and S1.3 increased during COVID-19 progression, reactivity to epitopes S1.8, S2.1 and S2.4 remained high across the studied time-window (Fig. 3D, Table S1). High reactivity to epitopes S1.1, S1.3, S1.4 (located in the RBD) and S1.8 was detected in three COVID-19 patients and in some COVID-19 naïve individuals, however the high antibody reactivity to S1.4 decreased in time of the COVID-19 patients' stay in the hospital. Low immune reactivity to S1.10 and S2.5 was observed in COVID-19 diseased, wherein response to these epitopes in naïve was high (Fig. 3D). Overall, we show that epitopes on spike protein identified in COVID-19 naïve are differentially targeted by antibodies elicited in COVID-19 patients.

**Pathogenic epitopes on the S protein of SARS-CoV-2.** We investigated whether pre-existing antibody response to SARS-CoV-2 S epitopes in COVID-19 naïve people could be landmarks of ill-health. For that, COVID-19 naïve cohort was divided into two: a Case group (n = 276) of subjects with diagnosis of different acute and chronic conditions (cardiovascular disease (CVD), breast cancer (BC), multiple sclerosis (MS), type 2 diabetes (T2D), or neuropsychiatric disorders (ND)), and a Control group (n = 262) (Table 1). First, we observed that pre-existing high immune response to epitopes of the S protein was significantly prevalent in the Case group ($\chi^2$ test, ****$p < 0.0001$) (Fig. 4A and Fig. S10B), whereas the classification into low- or high- seroresponse groups was neither associated with age nor gender ($\chi^2$ test, ns $p > 0.05$, Fig. S10A). Correlation analysis showed no correlation between age of subjects and abundance of immune response as detected by MVA (Pearson $R < 0.3$, Table S6).

As Case group included a different proportion of "low" and "high" sub-groups compared to Control group (Fig. 4A), we developed multivariable logistic regression models by using 80% of the subset as a training data set to identify which pre-existing antibody epitopes on S were pathogenic and associated with serious health problems. We found that antibody response to epitopes S1.6, S1.8 and S2.1 was effective for identifying Case subjects with diagnosis of serious acute and chronic conditions (95% confidence interval (CI) for the area under the receiver operating characteristic (AUROC) was 0.697…0.790), sensitivity 71.0%, specificity 68.6% and the balanced accuracy 69.8%) (Fig. 4B). Separately, prior high antibody response to S1.6 was more frequent in Control group, whereas pre-existing response to S1.8 and S2.1 was more prevalent in Case group (Fig. 4C). Cross-validation of the model against independent testing dataset (20% subset of data) was equally accurate within both "low" and "high" groups, with balanced accuracy values of 62% and 65%, respectively (Fig. S10C). Further, we observed that pre-existing antibody response to some of these S-specific epitopes was associated with certain disease groups. For example, pre-existing response to epitope S1.8 was higher in subjects with hypertension (Fig. S4B). Overall, we suggest that prior seroresponse to a combination of three epitopes of S (S1.6, S1.8 and S2.1) may be used for predicting the underlying risk of aggravated immunopathology of acute/chronic condition due to or associated with exposure to SARS-CoV-2 S antigen.

## Discussion

**Seroreactive immunodominant epitopes on spike of SARS-CoV-2 in COVID-19 naïve.** Here, we employed a high-resolution antibody response profiling technology (MVA) on a heterogenous cohort of COVID-19 naïve subjects including healthy subjects and people with one or more chronic or acute condition(s). Highly varied antibody response to SARS-CoV-2 and its emerging variants, in respect to antibody titres and neutralising activity, in the infected, vaccinated or people with hybrid immunity has been reported by several studies[83–86]. However, few of these studies have characterized the individual differences in the pre-existing antibody epitope repertoire. We identified 15 highly immunogenic epitopes characteristic to naïve populations in SARS-CoV-2 S protein localising in NTD, RBD, FP, HR1 and HR2 domains with notable heterogeneity among individuals (Figs. 1 and 4A). Our 3D analysis showed that majority of these were surface-exposed epitopes (Fig. S2) and encompassed antigenic regions determined by other studies underscoring the clinical relevance of the resolved 15 epitopes (Table 2). However, a notable difference between our findings and others (see Table 2

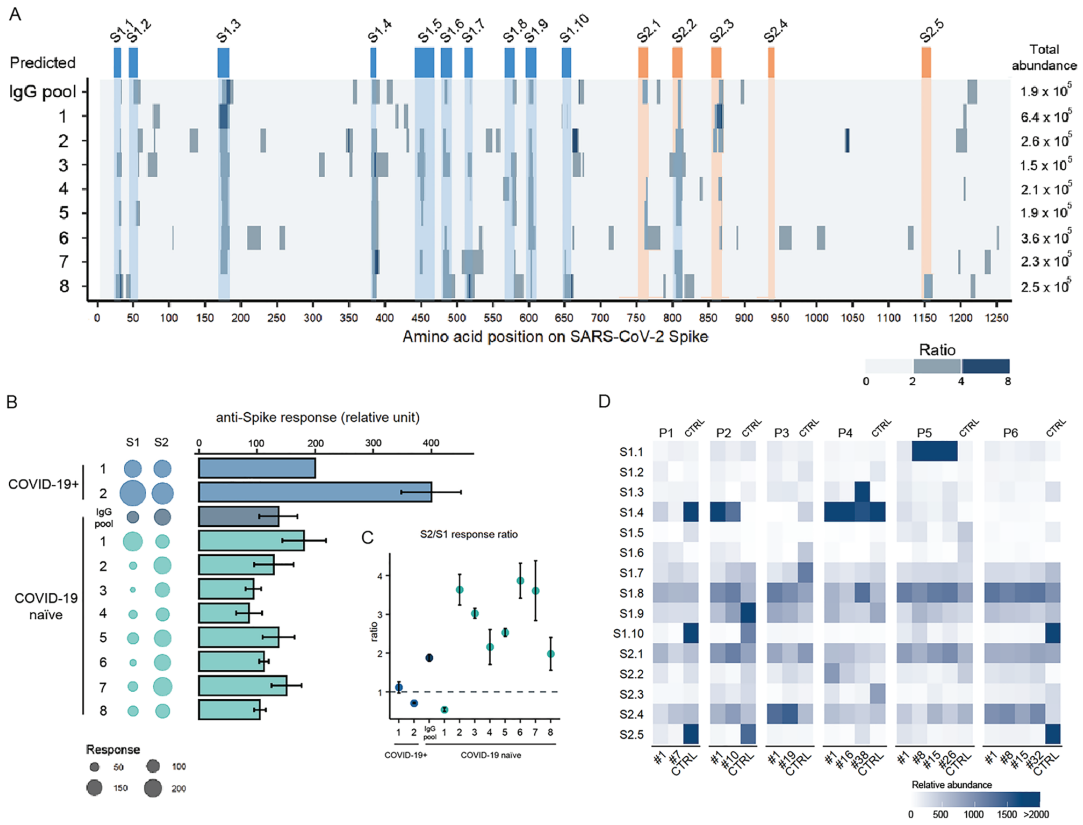



**Figure 3.** Differential antibody response to SARS-CoV-2 spike protein in COVID-19 diseased and naïve individuals. **(A)** Individual MVA immunoprofiles of antibody response to 15 epitopes of SARS-CoV-2 S in COVID-19 naïve samples shown as a ratio of specific vs random alignments. Ratios are visualised as centred moving averages across 9 amino acids. Y-axis—identifiers of samples, IgG pool refers to human pooled IgG sample, numbers 1–8 refer to COVID-19 naïve subjects (same as in **B,C**), Predicted—MVA-delineated epitopes on SARS-CoV-2 S; S1.1 to S1.10—epitopes on subunit 1 of S (dark blue); S2.1 to S2.5—epitopes on subunit 2 of S (orange colour); total abundance—calculated total abundance of IgG-bound peptides aligned to S per individual sample; colour bar (ratio)—ratio of alignment loads of specific vs random (scrambled) IgG-bound peptide sets. **(B)** Seroreactivity of COVID-19 naive (n = 8) and COVID-19 patients (n = 2) to spike protein subunits S1 and S2. 50 ng of SARS-CoV-2 S protein recombinant subunits S1 and S2 were immobilised on nitrocellulose slides and incubated with human serum/plasma samples to measure immunoreactivity to SARS-CoV-2 spike protein. Samples (n = 8) of selected subjects from COVID-19 naïve cohort (COVID-19 naïve 1–8) were used, alongside with the pooled human IgG sample (IgG pool, n = 2700 individual healthy donors, Sigma-Aldrich, # I4506). Samples (n = 2) taken at hospital intensive care admittance from patients (P1#1, P2#1) diagnosed with COVID-19 were included as positive controls for anti-spike immunoreactivity (COVID-19+, 1–2). Dots represent dot-ELISA data normalised to the sample from patient 1 (P1#1) (COVID-19+) separately for S1 and S2 subunits (100 represents 100%). Bar plot shows the sum of average of dot-ELISA results for S1 and S2 from independent experiments (n = 3). Error bars represent summarised SEM from independent S1 and S2 results. **(C)** Scatter plot depicts average ratio of SARS-CoV-2 S2 and S1 signals in dot-ELISA experiments (n = 3) from (**B**). Error bars represent SEM. **(D)** Heatmap shows the abundance of IgG-bound peptides containing the 15 epitopes of SARS-CoV-2 S protein in COVID-19 patients and controls. MVA immunoprofiles of serial samples from COVID-19 patients (n = 6) at different time points, where #1 is the sample taken at time of hospital admittance and "#n" where n designates the number of days from the first sample withdrawal. CTRL (n = 6) are age- and gender-matched healthy subjects selected from the cohort of 538 people. Relative abundance depicts the abundance of IgG-bound peptides containing the corresponding S epitopes, where values above 2000 are capped to 2000.

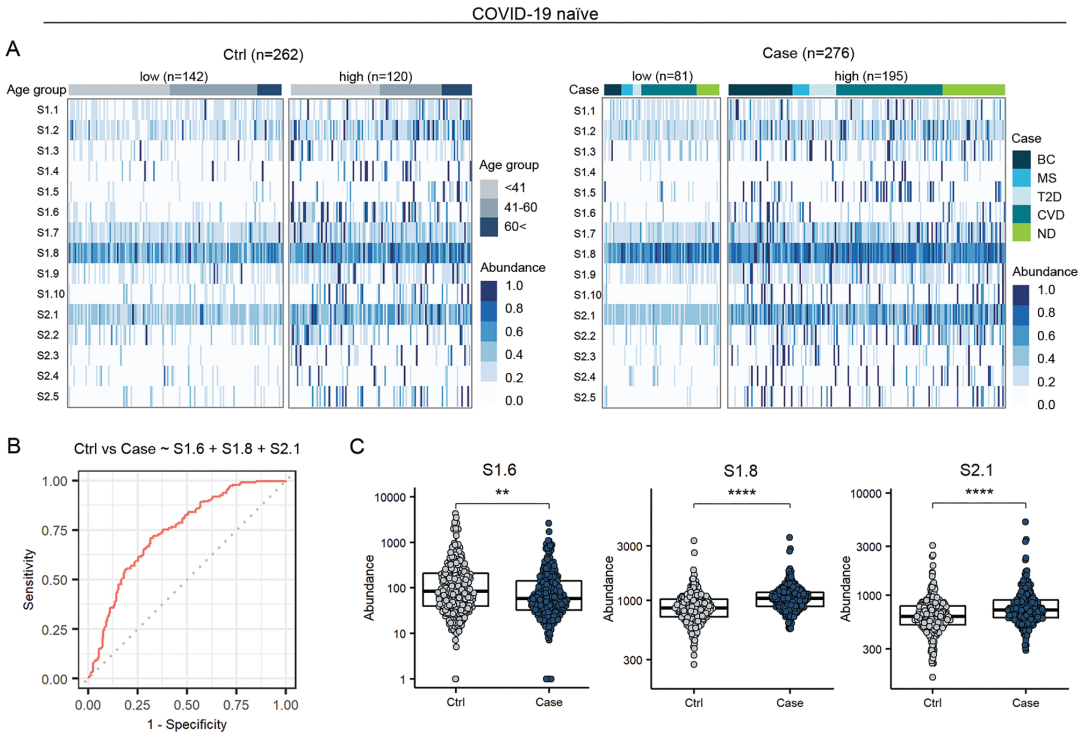



**Figure 4.** Antibody response to immunodominant epitopes on SARS-CoV-2 spike protein as predictors of ill health. **(A)** Individual immunoprofiles of SARS-CoV-2 S protein epitopes in COVID-19 naïve subjects, grouped into Ctrl (n = 262) or Case (with presented chronic diseases) (n = 276). "low" or "high"—subjects with relatively lower or higher overall response to S, calculated based on abundance of IgG-bound peptides containing the epitopes (S1.1 to S2.5) on SARS-CoV-2 spike protein (see "Methods"). x-axis—individual samples sub-grouped by Abundance (blue colour bar) where colour intensity shows individual abundance of IgG-bound peptides containing epitopes of spike normalised with 97.5th percentile value for visualisation. Age (grey colour bar) in Ctrl; y-axis—epitopes on Spike; *S1* epitopes on subunit 1 of S; *S2* on subunit 2 of S; COVID-19 naïve Case sub-groups: *BC* breast cancer (n = 57), *MS* multiple sclerosis (n = 20), *T2D* type 2 diabetes (n = 25), *CVD*—cardiovascular disease (n = 114), *ND*—neuropsychiatric disorders (n = 60). **(B)** Multivariable logistic regression analysis was used to describe the associations of epitope seroresponse predictors with the acute and/or chronic disease outcomes. Figure shows receiver operating characteristic (ROC) curve of using response to epitopes S1.6, S1.8, and S2.1 on training data (80% subset, i.e. 431/538 samples) for classifying Case (n = 276) vs Ctrl (n = 262). Area under curve (AUC) = 0.74 with 95% CI = (0.70…0.79). On validation with test set of 20% (107/538) samples, the select model classified samples into Case vs Ctrl with balanced accuracy of 62.0% for "low" group and 65.2% for "high" group (Fig. S10C). **(C)** Separately, antibody response to epitope S1.6 was identified as prevalent among Ctrl group subjects, whereas immune responses to epitopes S1.8 and S2.1 were prevalent among Case group in COVID-19 naïve cohort. Mann–Whitney U test, $^{**}p < 0.01$; $^{****}p < 0.0001$. Group sizes: Ctrl (n = 262), Case (n = 276); abundance—abundance of IgG-bound peptides containing respective epitopes.

for references) was that we observed cross-reacting IgG reactivity in COVID-19 naïve individuals to epitopes (S1.4, S1.5, S1.6) in RBD that included residues of the binding sites of neutralising antibodies (Table S5). We observed that predominant seroreactivity in naïve sera was against antigenic determinants outside the RBD of SARS-CoV-2 S (Fig. 1), consistent with the increasing recognition of non-RBD neutralising antibody response[87]. Several of these epitopes concurred with the dominant linear epitopes that were identified by the study of mRNA vaccines as shared by vaccine naïve individuals and COVID-19 infected subjects[88]. Further, NTD-targeting antibodies with neutralising activities are of high interest (ref in[67]). Our epitope S1.3 locates to the N4 loop of the S protein, which is an antigenic site of monoclonal antibodies with potent neutralising activity directed to NTD[66] thus suggesting that anti-S1.3 antibodies might interfere with anti-NTD supersite neutralising potential[89]. Several of the resolved epitopes within the S2 subunit overlapped with neutralisation determinants of SARS-CoV-2 S protein reported by others[21,43,45,52] suggesting that there is quite substantial polyclonal neutralising potency towards SARS-CoV-2 in COVID-19 naïve sera. Of note, immunisation with S2 in mice has shown more potent

neutralising antibody response than booster vaccines[90]. Given that recent molecular studies show the abundance of neutralising antibody targets on the SARS-CoV-2 S protein[91], our data on pre-existing immunity highlights the need to correlate these findings on polyclonal neutralisation potency and its breadth with histories of infections and vaccinations[92–94]. Countries are likely to have distinct immune profiles because their histories of COVID-19 waves and vaccination rates differ, suggesting that these differences manifest in differential cross-recognition of SARS CoV-2 infections/vaccines at the level of binding and neutralising antibody-based immunity.

Many of the SARS-CoV-2 Omicron sublineages have evolved that carry distinct spike mutations and represent an antigenic shift resulting in escape from antibodies induced by previous infection or vaccination[95]. Our data show that some of these mutations (specifically, L452Q/R, E484K, F486V, N658S), frequently present in novel variants of concern[68,71]) are located within epitopes S1.5, S1.6, and S1.10 respectively (Fig. 1 and Tables 2 and S5). E484K has been shown to decrease the neutralisation ability of anti-spike antibodies tenfold[96–99], L452Q/R and F486V are escape mutations in the RBD of Omicron sublineages from the cross-reactivity of neutralising antibodies, and N658S contributes to interference on hACE2 binding[71]. Cross-reactive antibodies targeting the dominant linear epitopes on SARS-CoV-2 S may contribute to neutralisation. Foremost, neutralising antibodies targeting the linear epitope 440–449[2], such as REGN-10987 (Imdevimab)[100], COV2-2130 (Cilgavimab, component of Evusheld)[101] and LY-CoV1404 (Bebtelovimab[102]) were reported to neutralise BA.2 subvariants and BA.4/5, with LY-CoV14044 notably demonstrating high potency against all Omicron subvariants[71]. On the other hand, cross-reactive activity from the binding of antibodies to SARS-CoV-2 could contribute to the control of infection by antibody-dependent mechanisms[83] and potentially amplify the damage that the virus causes to the body. The related mechanism involves antibody-dependent enhancement (ADE), a phenomenon in which non-neutralising/binding or sub-neutralising antibodies promote virus infection (rev in[103]). Most recently, early evidence of ADE in SARS-CoV-2 has begun emerging[104]. All these data further underscore the necessity for precise molecular characterisation of the effects of the pre-existing antibody on shaping the immunity to emerging SARS-CoV-2 infections and vaccines.

**Heterologous immunity landscape on SARS-CoV-2 S.** Most recently, the question of whether immune history affects SARS-CoV-2 infection outcome has been replaced by the question to what extent pre-existing immunity plays a role. Similar to studies on influenza whereby the antibody response to older virus strains had profound and negative impacts on subsequent immunity[13] cross-reactive antibody reactivity conferred by prior seasonal coronaviruses has widely been reported (ref in[20,105]). Higher titres of IgG against the HCoV-OC43 S protein were observed in patients with severe COVID-19[106], concluding that such immunological imprinting by previous seasonal coronavirus infections negatively impacted on the antibody response against SARS-CoV-2 infection[107]. Similarly, the exposure to the antigenically shifted Omicron primarily leads to a recall of existing memory B cells specific for epitopes shared by multiple SARS-CoV-2 variants rather than by priming naïve B cells recognising Omicron-specific epitopes showing that previous SARS-CoV-2 infection history can imprint a profound negative impact on the subsequent protective immunity[86].

In addition to cross-reactive epitopes with Omicron sublineages and also with endemic coronaviruses evidence of heterologous immunity between SARS-CoV-2 and pathogenic bacteria was reported[108]. Our study advances the concept showing that the cross-reactivity to epitopes of SARS-CoV-2 S protein with potential functional impacts could stem from the molecular mimicry with antigens of previously encountered other pathogens, including herpes-, papilloma-, adeno-, rhino-, influenza and other viruses (Fig. 2A). In good agreement with this, CMV seropositivity and age-related reduction in antibody titres against certain CMV antigens associated with the severity of SARS-CoV-2 infection[109]. Conversely, several other studies suggest cross-protection against COVID-19 incidence and severity from vaccines of influenza[28,110–112] and of other pathogens (polio, HIB, MMR, Varicella, PCV13, and HepA–HepB)[28,110–113]. These mechanisms may include generation of cross-protective antibodies through molecular mimicry. Cross-reacting antibodies with SARS-CoV-2 proteins elicited by poliovirus[114] and pneumococcal bacteria[115] have been identified, whereas for the mumps virus (via the MMR vaccine), the cross-reactivity of the vaccine antigen (measles fusion glycoprotein) with RBD of SARS-CoV-2 spike was suggested[113]. Additionally, monoclonal antibodies against SARS-CoV-2 S RBD have been shown to cross-react with the Ebola glycoprotein and HIV-1 gp140[116]. Our data predict 15 cross-reactive SARS-CoV-2 spike-like epitopes in common pathogens (Fig. 2A). Although it is not clear how this pre-existing anti-pathogen humoral immunity impacts SARS-CoV-2 infections or vaccine efficacy, Heterologous Vaccine Intervention as a strategy against COVID-19 is advocated by governmental bodies as advantageous for populations with suboptimal response to vaccination (e.g., patients with altered immunocompetence)[117]. Altogether these findings along with ours illustrate how immunological imprinting from prior exposure, i.e., 'original antigenic sin', can strongly affect the response to novel antigens. Whether the 15 SARS-CoV-2 S epitopes presented here elicit cross-reactive antibody response with antigens of pathogens and/or of human origin needs further investigation by functional studies.

**Pathogenic epitopes on SARS-CoV-2 S.** To our knowledge, this is a sole study to fine-map pre-existing antibody immunity to epitopes on SARS-CoV-2 S that associate with acute and/or chronic conditions like cardiovascular, neurological and oncological disease. Vastly different serological signatures to SARS-CoV-2 S that we detected in subjects with COVID-19 were also observed in healthy and people with chronic disease diagnosis suggesting a model where (recurrent) exposure to SARS-CoV-2 S-specific immune stimuli would progressively induce antibodies against certain epitopes landmarking chronic disease. We show the value of three epitopes on S as biomarkers to discriminate within COVID-19 naïve subjects between healthy and chronically diseased (with 95% CI, AUROC 0.69…0.79) (Fig. 4B and Fig. S10C). Ever since the first COVID-19 cases, immune-mediated manifestations have been reported[118]. A total of 55 long-term effects are associated with COVID-19[119]. Among these are myriad neurologic complications—including confusion, stroke, and neuromuscular disorders—which

manifest during acute COVID-19 and related maladies lasting for months, and implicate immune dysfunction, including anti-neural autoimmune dysregulation[120]. Autoimmunity has become the hallmark of post-COVID syndrome and latent autoimmunity correlates with humoral response to SARS-CoV-2[74]. Here, we elaborate this aspect of findings further by our data showing that the resolved 15 epitopes on SARS-CoV-2 S protein share similarity with many human proteins of immune and nervous system origins (Fig. S8). Specifically, epitope S1.6 of RBD shares similarity with regions in highly expressed CNS proteins (Fig. 2B). Relatedly, H1N1 infection and the Pandemrix vaccine were found to be associated with narcolepsy[121] by stimulating immune response targeting many cross-reactive autoantigens[41]. This suggests that conservation of antigenic sites across pathogen and human proteomes that we observed as epitopes on SARS-CoV-2 S protein in COVID-19 naïve individuals may have resulted (at least partially) from limited immune pressure for compatible immune fitness. However, further studies on epitopes of SARS-CoV-2 S are warranted, in particular in the settings of emerging strains and growing number of vaccines.

Collectively, our study provides evidence on the pre-existing immunity in COVID-19 naïve/unvaccinated individuals targeting 15 dominant epitopes on S protein. We show that this cross-reactive antibody response is boosted during SARS-CoV-2 infection in epitope-specific manner (Fig. 3D). Our findings are consistent with similar reports on pre-existing anti-SARS-CoV-2 S antibody immunity from others on cohorts with different genetic and demographic backgrounds[27,55–57,61,86]. We also show that pre-existing immunity on SARS-CoV-2 S shares epitope mimics associated with ill health. Overall, these data support the role of personal immune history with functional consequences to the diversity of antibodies elicited due to the phenomenon of heterologous immunity, i.e. back-boosting, i.e. immune imprinting, i.e. "original antigenic sin"[86,118]. The study lends further credence to MVA generated immunoprofiles as robust and generalisable.

**Study limitations.** A weakness of the study is the scarcity of information on the study participants, with no available data on the immune history of their infections and vaccinations. Another caveat is the limited sample size of COVID-19 infected individuals which we used for validating the pre-existing seroreactivity patterns observed in naïve samples. Further studies will be necessary to determine what roles the resolved epitopes play in anti-SARS-CoV-2 immunity. Although the resolved epitopes were located on the exterior surface of the spike trimer by and spike-specific seroreactivity in naïve sera was confirmed independently by dot-ELISA using globular and denatured S subunits, it needs further analysis of the function of these antigenic sites of whether these target neutralising or binding antibodies. Although our data analysis results show differential anti-SARS-CoV-2 S epitope reactivity in herpesvirus positive/negative individuals, further studies are warranted to prove direct cross-reactivity between SARS-CoV-2 S and other pathogens. The unexpected finding of the SARS-CoV-2 S epitopes to embed potential pathogenic antigenic determinants will require further investigation in the future. Deciphering the biological role of the conservation of the heterologous immunity hot-spots on SARS-CoV-2 S may contribute to the design of future vaccines.

## Data availability

The data are not publicly available due to containing information that could compromise research participant privacy/consent. Any materials that can be shared will be released via a material transfer agreement.

## Material availability

No new reagents or materials were generated as part of this study.

Received: 24 January 2022; Accepted: 20 September 2022
Published online: 07 October 2022

## References

1. Karim, S. S. A. & Karim, Q. A. Omicron SARS-CoV-2 variant: A new chapter in the COVID-19 pandemic. *Lancet* **398**(10317), 2126–2128 (2021).
2. Cao, Y. *et al.* Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **602**(7898), 657–663 (2022).
3. Elliott, P. *et al.* Rapid increase in Omicron infections in England during December 2021: REACT-1 study. *Science* **375**(6587), 1406–1411 (2022).
4. Collie, S., Champion, J., Moultrie, H., Bekker, L. G. & Gray, G. Effectiveness of BNT162b2 vaccine against Omicron variant in South Africa. *N. Engl. J. Med.* **386**(5), 494–496 (2022).
5. Madhi, S. A. *et al.* Population immunity and Covid-19 severity with Omicron variant in South Africa. *N. Engl. J. Med.* **386**(14), 1314–1326 (2022).
6. Sigal, A. Milder disease with Omicron: Is it the virus or the pre-existing immunity?. *Nat. Rev. Immunol.* **22**(2), 69–71 (2022).
7. Osuchowski, M. F. *et al.* The COVID-19 puzzle: Deciphering pathophysiology and phenotypes of a new disease entity. *Lancet Respir. Med.* **9**(6), 622–642 (2021).
8. Crook, H., Raza, S., Nowell, J., Young, M. & Edison, P. Long covid-mechanisms, risk factors, and management. *BMJ* **374**, n1648 (2021).
9. Tseng, H. F. *et al.* Effectiveness of mRNA-1273 against SARS-CoV-2 Omicron and Delta variants. *Nat Med.* **28**(5), 1063–1071 (2022).
10. Hope, J. L. & Bradley, L. M. Lessons in antiviral immunity. *Science* **371**(6528), 464–465 (2021).
11. Kelvin, A. A. & Zambon, M. Influenza imprinting in childhood and the influence on vaccine response later in life. *Euro Surveill.* **24**, 48 (2019).
12. Dugan, H. L. *et al.* Preexisting immunity shapes distinct antibody landscapes after influenza virus infection and vaccination in humans. *Sci. Transl. Med.* **12**(573), eabd3601 (2020).
13. Fonville, J. M. *et al.* Antibody landscapes after influenza virus infection or vaccination. *Science* **346**(6212), 996–1000 (2014).
14. Lipsitch, M., Grad, Y. H., Sette, A. & Crotty, S. Cross-reactive memory T cells and herd immunity to SARS-CoV-2. *Nat. Rev. Immunol.* **20**(11), 709–713 (2020).

15. Aran, D., Beachler, D. C., Lanes, S. & Overhage, J. M. Prior presumed coronavirus infection reduces COVID-19 risk: A cohort study. *J. Infect.* **81**(6), 923–930 (2020).
16. Sagar, M. *et al.* Recent endemic coronavirus infection is associated with less severe COVID-19. *J. Clin. Invest.* **13**, 1 (2020).
17. Amanat, F. *et al.* SARS-CoV-2 mRNA vaccination induces functionally diverse antibodies to NTD, RBD, and S2. *Cell* **184**(15), 3936-3948e10 (2021).
18. Dangi, T. *et al.* Cross-protective immunity following coronavirus vaccination and coronavirus infection. *J. Clin. Invest.* **131**, 24 (2021).
19. Aguilar-Bretones, M. *et al.* Seasonal coronavirus-specific B cells with limited SARS-CoV-2 cross-reactivity dominate the IgG response in severe COVID-19. *J. Clin. Invest.* **131**, 21 (2021).
20. Crowley, A. R. *et al.* Boosting of cross-reactive antibodies to endemic coronaviruses by SARS-CoV-2 infection but not vaccination with stabilized spike. *Elife* **2022**, 11 (2022).
21. Focosi, D. *et al.* Previous humoral immunity to the endemic seasonal alphacoronaviruses NL63 and 229E is associated with worse clinical outcome in COVID-19 and suggests original antigenic sin. *Life (Basel).* **11**(4), 298 (2021).
22. Kundu, R. *et al.* Cross-reactive memory T cells associate with protection against SARS-CoV-2 infection in COVID-19 contacts. *Nat. Commun.* **13**(1), 80 (2022).
23. Selva, K. J. *et al.* Systems serology detects functionally distinct coronavirus antibody features in children and elderly. *Nat. Commun.* **12**(1), 2037 (2021).
24. Sokal, A. *et al.* Maturation and persistence of the anti-SARS-CoV-2 memory B cell response. *Cell* **184**, 1201 (2021).
25. Thevarajan, I. *et al.* Breadth of concomitant immune responses prior to patient recovery: A case report of non-severe COVID-19. *Nat. Med.* **26**(4), 453–455 (2020).
26. Cohn, M. Degeneracy, mimicry and crossreactivity in immune recognition. *Mol. Immunol.* **42**(5), 651–655 (2005).
27. Shrock, E. *et al.* Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* **370**, 6520 (2020).
28. Pawlowski, C. *et al.* Exploratory analysis of immunization records highlights decreased SARS-CoV-2 rates in individuals with recent non-COVID-19 vaccinations. *Sci. Rep.* **11**(1), 4741 (2021).
29. Hupert, N., Marin-Hernandez, D., Gao, B., Aguas, R. & Nixon, D. F. Heterologous vaccination interventions to reduce pandemic morbidity and mortality: Modeling the US winter 2020 COVID-19 wave. *Proc. Natl. Acad. Sci. U S A.* **119**(3), 25448119 (2022).
30. Bastard, P. *et al.* Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* **370**(6515), eabd4585 (2020).
31. Zhang, A., Stacey, H. D., Mullarkey, C. E. & Miller, M. S. Original antigenic sin: How first exposure shapes lifelong anti-influenza virus immune responses. *J. Immunol.* **202**(2), 335–340 (2019).
32. Chang, S. E. *et al.* New-onset IgG autoantibodies in hospitalized patients with COVID-19. *Nat. Commun.* **12**(1), 5417 (2021).
33. Lingel, H. *et al.* Unique autoantibody prevalence in long-term recovered SARS-CoV-2-infected individuals. *J. Autoimmun.* **122**, 102682 (2021).
34. Barda, N. *et al.* Safety of the BNT162b2 mRNA Covid-19 vaccine in a nationwide setting. *N. Engl. J. Med.* **385**(12), 1078–1090 (2021).
35. Mevorach, D. *et al.* Myocarditis after BNT162b2 mRNA vaccine against Covid-19 in Israel. *N. Engl. J. Med.* **385**, 2140 (2021).
36. Verma, A. K., Lavine, K. J. & Lin, C. Y. Myocarditis after Covid-19 mRNA vaccination. *N. Engl. J. Med.* **385**(14), 1332–1334 (2021).
37. Witberg, G., Barda, N., Hoss, S., Richter, I., Wiessman, M., Aviv, Y., *et al.* Myocarditis after Covid-19 vaccination in a large health care organization. *N. Engl. J. Med.* (2021).
38. Oster, M. E., Shay, D. K. & Shimabukuro, T. T. Myocarditis cases after mRNA-based COVID-19 vaccination in the US-reply. *JAMA* **327**(20), 2020–2021 (2022).
39. Vojdani, A. & Kharrazian, D. Potential antigenic cross-reactivity between SARS-CoV-2 and human tissue with a possible link to an increase in autoimmune diseases. *Clin. Immunol.* **217**, 108480 (2020).
40. Sadam, H. *et al.* Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients. *EBioMedicine* **64**, 103211 (2021).
41. Sadam, H. *et al.* Prostaglandin D2 receptor DP1 antibodies predict vaccine-induced and spontaneous narcolepsy type 1: Large-scale study of antibody profiling. *EBioMedicine* **29**, 47–59 (2018).
42. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**(6220), 1260419 (2015).
43. UniProt, C. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**(D1), D480–D489 (2021).
44. Hulo, C. *et al.* ViralZone: A knowledge resource to understand virus diversity. *Nucleic Acids Res.* **39**(Database issue), D576–D582 (2011).
45. Pupina, N. *et al.* Immune response to a conserved enteroviral epitope of the major capsid VP1 protein is associated with lower risk of cardiovascular disease. *EBioMedicine* **76**, 103835 (2022).
46. Jaago, M. *et al.* Antibody response to oral biofilm is a biomarker for acute coronary syndrome in periodontal disease. *Commun. Biol.* **5**(1), 205 (2022).
47. Leppik, L. *et al.* Profiling of lipidomics before and after antipsychotic treatment in first-episode psychosis. *Eur. Arch. Psychiatry Clin. Neurosci.* **270**(1), 59–70 (2020).
48. Parksepp, M. *et al.* Metabolomics approach revealed robust changes in amino acid and biogenic amine signatures in patients with schizophrenia in the early course of the disease. *Sci. Rep.* **10**(1), 13983 (2020).
49. Wickham, H., Averick, M., Bryan, J., Chang, W. & McGowan, L. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 43 (2019).
50. Tran, A. N., Dussaq, A. M., Kennell, T. Jr., Willey, C. D. & Hjelmeland, A. B. HPAanalyze: an R package that facilitates the retrieval and analysis of the Human Protein Atlas data. *BMC Bioinform.* **20**(1), 463 (2019).
51. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2021).
52. Grifoni, A. *et al.* A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe.* **27**(4), 671-680e2 (2020).
53. Rogers, T. F. *et al.* Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* **369**(6506), 956–963 (2020).
54. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**(6483), 1260–1263 (2020).
55. Amrun, S. N. *et al.* Linear B-cell epitopes in the spike and nucleocapsid proteins as markers of SARS-CoV-2 exposure and disease severity. *EBioMedicine* **58**, 102911 (2020).
56. Farrera-Soler, L. *et al.* Identification of immunodominant linear epitopes from SARS-CoV-2 patient plasma. *PLoS ONE* **15**(9), e0238089 (2020).
57. Ng, K. W. *et al.* Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science* **370**(6522), 1339–1343 (2020).
58. Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**(7807), 215–220 (2020).
59. ter Meulen, J. *et al.* Human monoclonal antibody combination against SARS coronavirus: Synergy and coverage of escape mutants. *PLoS Med.* **3**(7), e237 (2006).

60. Yuan, M. *et al.* A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* **368**(6491), 630–633 (2020).
61. Schwarz, T. *et al.* SARS-CoV-2 proteome-wide analysis revealed significant epitope signatures in COVID-19 patients. *Front. Immunol.* **12**, 629185 (2021).
62. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH image to ImageJ: 25 years of image analysis. *Nat. Methods.* **9**(7), 671–675 (2012).
63. Clarke, E., & Sherrill-Mix, S. *ggbeeswarm: Categorical Scatter (Violin Point) Plots. R Package Version 060*. (2019).
64. Kassambra A. *ggpubr: 'ggplot2' Based Publication Ready Plots. R Package Version 040*. (2020).
65. Poh, C. M. *et al.* Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat. Commun.* **11**(1), 2806 (2020).
66. Chi, X. *et al.* A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **369**(6504), 650–655 (2020).
67. Haslwanter, D. *et al.* A combination of receptor-binding domain and N-terminal domain neutralizing antibodies limits the generation of SARS-CoV-2 spike neutralization-escape mutants. *MBio* **12**(5), e0247321 (2021).
68. McCallum, M. *et al.* Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants. *Science* **374**, eabl8506 (2021).
69. Barnes, C. O. *et al.* SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**(7839), 682–687 (2020).
70. Yuan, M. *et al.* Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. *Science* **373**(6556), 818–823 (2021).
71. Cao, Y., Yisimayi, A., Jian, F., Song, W., Xiao, T., Wang, L., *et al.* BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature*. (2022).
72. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**(6501), 330–333 (2020).
73. Li, Y. *et al.* Linear epitope landscape of the SARS-CoV-2 Spike protein constructed from 1,051 COVID-19 patients. *Cell Rep.* **34**(13), 108915 (2021).
74. Rojas, M. *et al.* Autoimmunity is a hallmark of post-COVID syndrome. *J. Transl. Med.* **20**(1), 129 (2022).
75. El-Dessouky, S. H. *et al.* Prenatal delineation of a distinct lethal fetal syndrome caused by a homozygous truncating KIDINS220 variant. *Am. J. Med. Genet. A.* **182**(12), 2867–2876 (2020).
76. DiMaio, R. *et al.* LRRK2 activation in idiopathic Parkinson's disease. *Sci. Transl. Med.* **10**(451), 5429 (2018).
77. Wallings, R. L. & Tansey, M. G. LRRK2 regulation of immune-pathways and inflammatory disease. *Biochem. Soc. Trans.* **47**(6), 1581–1595 (2019).
78. Duan, Y. *et al.* Prediction of key genes and miRNAs responsible for loss of muscle force in patients during an acute exacerbation of chronic obstructive pulmonary disease. *Int. J. Mol. Med.* **38**(5), 1450–1462 (2016).
79. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* **41**(6), 677–687 (2009).
80. Mendez-Gomez, H. R. *et al.* The lipase activity of phospholipase D2 is responsible for nigral neurodegeneration in a rat model of Parkinson's disease. *Neuroscience* **377**, 174–183 (2018).
81. Meyer, R. C., Giddens, M. M., Coleman, B. M. & Hall, R. A. The protective role of prosaposin and its receptors in the nervous system. *Brain Res.* **1585**, 1–12 (2014).
82. Andrade, A. *et al.* Genetic associations between voltage-gated calcium channels and psychiatric disorders. *Int. J. Mol. Sci.* **20**(14), 3537 (2019).
83. Cho, A., Muecksch, F., Schaefer-Babajew, D., Wang, Z., Finkin, S., Gaebler, C. *et al.* Anti-SARS-CoV-2 receptor binding domain antibody evolution after mRNA vaccination. *Nature* (2021).
84. Gaebler, C. *et al.* Evolution of antibody immunity to SARS-CoV-2. *Nature* **591**(7851), 639–644 (2021).
85. Wang, Z. *et al.* Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. *Nature* **595**(7867), 426–431 (2021).
86. Reynolds, C. J. *et al.* Immune boosting by B11529 (Omicron) depends on previous SARS-CoV-2 exposure. *Science* **377**, eabq1841 (2022).
87. Voss, C. *et al.* Epitope-specific antibody responses differentiate COVID-19 outcomes and variants of concern. *JCI Insight.* **6**, 13 (2021).
88. Wisnewski, A. V. *et al.* Immunogenic amino acid motifs and linear epitopes of COVID-19 mRNA vaccines. *PLoS ONE* **16**(9), e0252849 (2021).
89. McCallum, M. *et al.* N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **184**(9), 2332-2347e16 (2021).
90. Ng, K. W. *et al.* SARS-CoV-2 S2-targeted vaccination elicits broadly neutralizing antibodies. *Sci. Transl. Med.* **14**(655), eabn3715 (2022).
91. Schmidt, F., Weisblum, Y., Rutkowska, M., Poston, D., DaSilva, J., Zhang, F. *et al.* High genetic barrier to SARS-CoV-2 polyclonal neutralizing antibody escape. *Nature*. (2021).
92. Forgacs, D. *et al.* SARS-CoV-2 mRNA vaccines elicit different responses in immunologically naive and pre-immune humans. *Front. Immunol.* **12**, 728021 (2021).
93. Collier, A.Y., Yu, J., McMahan, K., Liu, J., Chandrashekar, A., Maron, J.S., *et al.* Differential kinetics of immune responses elicited by covid-19 vaccines. N. Engl. J. Med. (2021).
94. Martinez-Flores, D. *et al.* SARS-CoV-2 vaccines based on the spike glycoprotein and implications of new viral variants. *Front. Immunol.* **12**, 701501 (2021).
95. Park, Y.J., Pinto, D., Walls, A.C., Liu, Z., De Marco, A., Benigni, F. *et al.* Imprinted antibody responses against SARS-CoV-2 Omicron sublineages. *bioRxiv* (2022).
96. Kupferschmidt, K. New mutations raise specter of "immune escape". *Science* **371**(6527), 329–330 (2021).
97. Xie, X., Liu, Y., Liu, J., Zhang, X., Zou, J., Fontes-Garfias, C.R. *et al.* Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat. Med.* (2021).
98. Ku, Z. *et al.* Molecular determinants and mechanism for antibody cocktail preventing SARS-CoV-2 escape. *Nat. Commun.* **12**(1), 469 (2021).
99. Greaney, A. J. *et al.* Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe.* **29**(3), 463-476e6 (2021).
100. Hansen, J. *et al.* Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* **369**(6506), 1010–1014 (2020).
101. Zost, S. J. *et al.* Potently neutralizing and protective human antibodies against SARS-CoV-2. *Nature* **584**(7821), 443–449 (2020).
102. Westendorf, K. *et al.* LY-CoV1404 (bebtelovimab) potently neutralizes SARS-CoV-2 variants. *Cell Rep.* **39**(7), 110812 (2022).
103. Bournazos, S., Gupta, A. & Ravetch, J. V. The role of IgG Fc receptors in antibody-dependent enhancement. *Nat. Rev. Immunol.* **20**(10), 633–643 (2020).
104. Qi, H., Liu, B., Wang, X., & Zhang, L. The humoral response and antibodies against SARS-CoV-2 infection. *Nat. Immunol.* (2022).

105. Ng, K. W., Faulkner, N., Wrobel, A. G., Gamblin, S. J. & Kassiotis, G. Heterologous humoral immunity to human and zoonotic coronaviruses: Aiming for the achilles heel. *Semin. Immunol.* **55**, 101507 (2021).
106. Guo, L. *et al.* Cross-reactive antibody against human coronavirus OC43 spike protein correlates with disease severity in COVID-19 patients: A retrospective study. *Emerg. Microbes Infect.* **10**(1), 664–676 (2021).
107. Aydillo, T. *et al.* Immunological imprinting of the antibody response in COVID-19 patients. *Nat. Commun.* **12**(1), 3781 (2021).
108. Eggenhuizen, P. J. *et al.* Heterologous immunity between SARS-CoV-2 and pathogenic bacteria. *Front. Immunol.* **13**, 821595 (2022).
109. Shang, J. *et al.* Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U S A* **117**(21), 11727–11734 (2020).
110. Amato, M. *et al.* Relationship between influenza vaccination coverage rate and COVID-19 outbreak: An Italian ecological study. *Vaccines (Basel).* **8**(3), 535 (2020).
111. Conlon, A., Ashur, C., Washer, L., Eagle, K. A. & Hofmann Bowman, M. A. Impact of the influenza vaccine on COVID-19 infection rates and severity. *Am. J. Infect. Control.* **49**(6), 694–700 (2021).
112. Fink, G., Orlova-Fink, N., Schindler, T., Grisi, S., Ferrer, A.P.S., Daubenberger, C. *et al.* Inactivated trivalent influenza vaccination is associated with lower mortality among patients with COVID-19 in Brazil. *BMJ Evid. Based Med.* (2020).
113. Marakasova, E. & Baranova, A. MMR vaccine and COVID-19: Measles protein homology may contribute to cross-reactivity or to complement activation protection. *MBio* **12**(1), 3447 (2021).
114. Comunale, B. A. *et al.* Poliovirus vaccination induces a humoral immune response that cross reacts with SARS-CoV-2. *Front. Med. (Lausanne).* **8**, 710010 (2021).
115. Root-Bernstein, R. Possible cross-reactivity between SARS-CoV-2 proteins, CRM197 and proteins in pneumococcal vaccines may protect against symptomatic SARS-CoV-2 disease and death. *Vaccines (Basel).* **8**(4), 559 (2020).
116. Kim, S. I. *et al.* Stereotypic neutralizing VH antibodies against SARS-CoV-2 spike protein receptor binding domain in patients with COVID-19 and healthy individuals. *Sci. Transl. Med.* **13**(578), 6990 (2021).
117. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *Lancet* **395**(10229), 1054–1062 (2020).
118. Yewdell, J. W. & Santos, J. J. S. Original antigenic sin: How original? How sinful?. *Cold Spring Harb. Perspect. Med.* **11**(5), 387816 (2021).
119. Lopez-Leon, S. *et al.* More than 50 long-term effects of COVID-19: A systematic review and meta-analysis. *Sci. Rep.* **11**(1), 16144 (2021).
120. Spudich, S. & Nath, A. Nervous system consequences of COVID-19. *Science* **375**(6578), 267–269 (2022).
121. Sarkanen, T., Alakuijala, A., Julkunen, I. & Partinen, M. Narcolepsy associated with pandemrix vaccine. *Curr. Neurol. Neurosci. Rep.* **18**(7), 43 (2018).
122. Mittal, A., Khattri, A. & Verma, V. Structural and antigenic variations in the spike protein of emerging SARS-CoV-2 variants. *PLoS Pathog.* **18**(2), e1010260 (2022).
123. Chen, R.E., Zhang, X., Case, J.B., Winkler, E.S., Liu, Y., Van Blargan, L.A. *et al.* Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nat. Med.* (2021).
124. Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **19**(7), 409–424 (2021).
125. Ou, J. *et al.* Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events. *Signal Transduct. Target Ther.* **7**(1), 138 (2022).
126. Tian, D., Sun, Y., Zhou, J. & Ye, Q. The global epidemic of SARS-CoV-2 variants and their mutational immune escape. *J. Med. Virol.* **94**(3), 847–857 (2022).
127. Starr, T. N., Greaney, A. J., Dingens, A. S. & Bloom, J. D. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Rep. Med.* **2**(4), 100255 (2021).
128. Baum, A. *et al.* Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* **369**(6506), 1014–1018 (2020).


## Acknowledgements

We thank A.-H. Pool (Caltech) for his expertise that supported this work. The members of Protobios team are thanked for their excellent technical assistance. Icosagen (Estonia) is highly acknowledged for the kind gift of SARS-CoV-2 spike subunit recombinant proteins.

## Author contributions

K.P., M.J., A.R., N.P., conceived and designed the study. A.P., M.P., L.H., E.V., M.B., F.L., E.K., P.P., P.T., A.V. and D.L. collected samples and acquired the clinical data. K.P., M.J., A.R., N.P., A.P., H.S., J.T., A.A. and A.M.G. analysed and interpreted the data. E.V., P.T., D.L., T.T. and K.P. supervised the study. K.P., M.J., and A.R. drafted the manuscript. All authors revised and approved the final manuscript for submission.

## Funding

This study was supported by research funding grants of Protobios (5.1-4/20/170, and PRG573) from the Estonian Ministry of Education and Research and Estonian Research Council, respectively, and H2020-MSCA-RISE-2016 (EU734791) and H2020 PANBioRA (EU760921) projects from the European Union, Helsinki University Hospital grants, Mary and Georg C. Ehrnrooth Foundation, Finnish Eye Foundation. AA and TT were partially supported by Estonian Research Council (institutional research funding IUT19-18 and grant PRG805) and European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012). JT was partially supported by the SekMO grant from Estonian Ministry of Education and Research (2014-2020.4.01.21-0315). DL was supported by Finska Läkaresällskapet, Liv och Hälsa Foundation and The Finnish Society of Sciences and Letters. AV was supported by Magnus Ehrnrooth Foundation and Sigrid Jusélius Foundation.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-20849-6.

**Correspondence** and requests for materials should be addressed to K.P.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022

# Appendix 3

**Publication III**

Pupina, N., Avarlaid, A., Sadam, H., Pihlak, A., Jaago, M., Tuvikene, J., **Annika Rähni**, Planken, A., Planken, M., Kalso, E., Tienari, P. J., Nieminen, J. K., Seppänen, M. R. J., Vaheri, A., Lindholm, D., Sinisalo, J., Pussinen, P., Timmusk, T., & Palm, K. (2022). **Immune response to a conserved enteroviral epitope of the major capsid VP1 protein is associated with lower risk of cardiovascular disease.** EBioMedicine, 76, 103835. https://doi.org/10.1016/j.ebiom.2022.103835

# Immune response to a conserved enteroviral epitope of the major capsid VP1 protein is associated with lower risk of cardiovascular disease

Nadežda Pupina,[a] Annela Avarlaid,[b] Helle Sadam,[a,b] Arno Pihlak,[a] Mariliis Jaago,[a,b] Jürgen Tuvikene,[a,b,c] Annika Rähni,[a,b] Anu Planken,[d] Margus Planken,[d] Eija Kalso,[e] Pentti J. Tienari,[f] Janne K. Nieminen,[f] Mikko R.J. Seppänen,[f] Antti Vaheri,[g] Dan Lindholm,[h,i] Juha Sinisalo,[j] Pirkko Pussinen,[k] Tõnis Timmusk,[a,b] and Kaia Palm[a,b]*

[a]Protobios LLC, Mäealuse 4, Tallinn 12618, Estonia
[b]Department of Chemistry and Biotechnology, Tallinn University of Technology, Estonia
[c]dxlabs LLC, Mäealuse 4, Tallinn 12618, Estonia
[d]The North Estonia Medical Center, Tallinn, Estonia
[e]Department of Anaesthesiology, Intensive Care and Pain Medicine, Helsinki University Hospital and Department of Pharmacology and SleepWell Research Programme, University of Helsinki, Finland
[f]Department of Neurology, Neurocenter, Helsinki University Hospital, and Translational Immunology Research Program, University of Helsinki, Finland
[g]Department of Virology, Medicum, University of Helsinki, Finland
[h]Department of Biochemistry and Developmental Biology, Faculty of Medicine, University of Helsinki, Finland
[i]Minerva Foundation Institute for Medical Research, Helsinki, Finland
[j]Heart and Lung Center, Helsinki University Hospital, University of Helsinki, Finland
[k]Oral and Maxillofacial Diseases, University of Helsinki and Helsinki University Hospital, Finland

eBioMedicine 2022;76: 103835
Published online 25 January 2022
https://doi.org/10.1016/j.ebiom.2022.103835

## Summary

**Background** Major cardiac events including myocardial infarction (MI) are associated with viral infections. However, how specific infections contribute to the cardiovascular insults has remained largely unclear.

**Methods** We employed next generation phage display mimotope-variation analysis (MVA) to explore the link between antibody-based immune response and severe cardiovascular conditions. Here, we used a case-control design, including the first-stage discovery cohort ($n$ = 100), along with cohorts for second-stage discovery ($n$ = 329) and validation ($n$ = 466).

**Findings** We observed strong antibody response to the peptide antigens with Gly-Ile-X-Asp (G-I-X-D) core structure in healthy individuals but not in patients with MI. Analysis of the origin of this epitope linked it with the N-terminus of the VP1 protein of poliovirus 3 (PV3), but also other species of picornaviruses. Consistently, we found low levels of antibody response to the G-I-X-D epitope in individuals with severe cardiac disease complications.

**Interpretation** Our findings imply that antibody response to the G-I-X-D epitope is associated with polio vaccinations and that high antibody levels to this epitope could discriminate healthy individuals from prospective MI patients as a blood-derived biomarker. Together, these findings highlight the importance of epitope-specific antibody response and suggest that protective immunity against the polio- and non-polio enteroviral infections support improved cardiovascular health.

**Funding** Estonian Ministry of Education (5.1-4/20/170), Estonian Research Council (PRG573, PRG805), H2020-MSCA-RISE-2016 (EU734791), H2020 PANBioRA (EU760921), European Union through the European Regional Development Fund (Project no. 2014-2020.4.01.15-0012), Helsinki University Hospital grants, Mary and Georg C. Ehrnrooth Foundation, Finnish Eye Foundation, Finska Läkaresällskapet, The Finnish Society of Sciences and Letters, Magnus Ehrnrooth Foundation and Sigrid Jusélius Foundation.

**Copyright** © 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

**Keywords:** Epitope; Antibodies; Myocardial infarction; Cardiovascular diseases; Poliovirus; Enteroviruses

*Corresponding author at: Protobios LLC, Mäealuse 4, Tallinn 12618, Estonia.
*E-mail address:* kaia@protobios.com (K. Palm).

## Research in context

### *Evidence before this study*

Cardiovascular diseases (CVDs) are the leading cause of death globally. High blood pressure, smoking, high cholesterol levels, obesity, and diabetes are among the major risk factors contributing to CVD development.

### *Added value of this study*

Here, we used high-throughput analysis of the antibody immune response in search of novel CVD-associated biomarkers and found value in measuring the immunoreactivity to a single viral epitope as a blood-based biomarker for the risk assessment of the CVD associated acute cardiac events.

### *Implications of all the available evidence*

The understanding of novel viral infection-associated mechanisms along with noninvasive cardiac-specific biomarkers may yield valuable insights into the CVD pathophysiology and prevention at an early stage of the disease. Given that we relate this epitope to cross-protective poliovirus antigens, the study data would contribute to further vaccine use and improvements.

## Introduction

Cardiovascular diseases (CVDs) are the leading global burden with coronary artery disease (CAD) as a major contributor to the fatalities, and myocardial infarction (MI) accounting for most of the mortality.[1,2]

External stress factors that have detrimental impact on the integrity of the cellular machinery promote the development of cardiac disorders.[3] CAD is the most common heart problem that involves the blood vessels and usually starts as a result of atherosclerosis and may culminate in MI, a major adverse cardiovascular event.[4] In the atherosclerotic disease process, the role of inflammation has become well established.[5,6] Considerable evidence has emerged indicating that infections contribute to chronic inflammation either through direct effects by hijacking cellular processes and inducing host inflammatory cascades, or indirectly via an autoimmune-mediated response, or a combination of both.[7]

Infectious diseases of the heart are heterogenous. They have variable clinical presentations as any type of microorganism, e.g. bacteria, fungi, parasites, and viruses, may affect more than one cardiac structure.[8] A number of acute and chronic infections including *Chlamydophila pneumoniae, Helicobacter pylori, Porphyromonas gingivalis,* influenza and coronaviruses, e.g. SARS-CoV-2[9,10] has been linked to atherosclerosis[11] and MI.[12] In both cross-sectional and cohort studies, elevated antibodies against *C. pneumonia, H.pylori*, periodontal pathogens, *Mycoplasma pneumoniae*, herpes viruses, and enteroviruses have been demonstrated.[13–16] Further, data suggest that the general infectious burden poses a greater risk for coronary heart disease (CHD) than any single pathogen.[17–19] A unique study by Pesonen and co-workers showed that the antibody titers against pathogens like enteroviruses cumulatively increased the risk for CHD, whereas the risk for acute coronary events decreased with increasing number of childhood contagious diseases.[20] On the other hand, childhood oral infections and infection-related hospitalisations were associated with subclinical carotid atherosclerosis and adverse atherosclerosis phenotype in adults.[21,22] Increased risk for acute coronary syndrome (ACS) was found for acute bacterial pneumonia, urinary tract infection and bacteremia with different bacterial strains.[23]

Excess mortality from cardiovascular disease during influenza epidemics was recognised early in the 20th century, and recent reports indicate that COVID-19 may also be associated with a similar increase.[23] There is no direct evidence that viral vaccines would have an influence on atherosclerosis in humans.[24] However, studies showed that influenza and pneumococcal vaccination reduce the risk for cardiovascular events in adults by 36%[25] and 17%,[26] respectively. The molecular mechanisms by which vaccinations exert these protective effects have not been defined.

The current study investigated the relationship between antibody immune response associated with MI by using an untargeted, high-throughput next generation peptide phage display method (MVA). We have correlated these results with the data from the case-control cohort comprising of patients with angiographically verified CVDs status, i.e. ACS and chronic CAD, and population-based controls to evaluate associations of antibody immune response as a predictive blood biomarker for CVD conditions.

## Methods

### Ethics statement

The study was conducted in accordance with the guiding principles of the Declaration of 182 Helsinki and the study participants gave written informed consent before enrolment. The ethical approvals were obtained from the local ethics review committees of Estonia (177/T-2, 211/M-22, 281/T-5; 351), and Finland (the Helsinki University Central Hospital ethics committee (106/2007, 303/2007); Helsinki University Hospital (Dno 83/13/03/01/2013); Helsinki Biobank (HBP2018006)). Healthy sera donor samples were procured from the North Estonia Medical Blood Centre (Tallinn, Estonia).

### Study populations

The summary of the study design is illustrated in Figure S1. Baseline characteristics for the cohorts used are shown in Tables 1 and S1, S3 and S4. In brief, study was

| | Healthy control (HC) | Myocardial infarction (MI) |
|---|---|---|
| **Group size (n)** | 50 | 50 |
| **Gender** | 20 M; 30F | 29 M; 13 F; 8 NA |
| **Mean age ±SD (min-max), years** | 40.76 ± 9.86 (21−58) | 66.78 ± 12.58 (38−88)*; 10 NA |

***Table 1*: Characterisation of the Estonia I cohort (Figure S1).**
The age of the individuals was fixed at the time of sample collection. *HC − healthy control; MI − myocardial infarction patients; M − male; F − female; NA − not available: patient agreeing to the study did not disclose the exact age or gender.*
* − data on 40 subjects.

designed as 2-stage case-control study including three cohorts, Estonia I and II, and Finland I. Using a case-control design, the first-stage discovery cohort (Estonia I, $n$ = 100) was built from the samples collected by North Estonian Regional Hospital (Tallinn, Estonia) including serum or plasma of MI ($n$ = 50) and HC ($n$ = 50). MI was diagnosed according to electrocardiogram and cardiac marker analysis (troponin, creatine kinase-MB). The case condition included ST-elevation (STEMI) and non-ST-elevation (NSTEMI) myocardial infarction patients. All MI samples were collected less than 12 h before or after the coronary angiography procedures and stored at −80 °C until further use. Heart conditions were excluded for the HC group. In the second stage of the study, we were aiming to associate the discovered epitope with different types of poliovirus vaccines. Another discovery cohort Estonia II ($n$ = 329) was created by combining samples of previous retrospective studies from (a) subjects born in years 1927−1958 ($n$ = 101, *No vaccine*), when no polio vaccine was available, also in Estonia, (b) subjects born in years of OPV vaccination according to the national immunisation history (years 1959−2007, $n$ = 224, *OPV*), and (c) individuals born in years of IPV vaccination programmes in Estonia (years 2008−2011, $n$ = 4, *IPV*) (Table S3). It should be mentioned that individuals from *No vaccine* group could have received OPV during mass vaccination programmes in 1959 and 1960 in their adulthood[27,28] (Figure S2). Regarding clinical background, the Estonia II cohort consisted of MI cases ($n$ = 40) and either healthy blood donors (ICD-10: Z52.0, $n$ = 135) or samples from individuals with different primary diagnoses with no history of MI ($n$ = 154) (including patients with cervical intraepithelial pathologies ($n$ = 42), prediabetes ($n$ = 6), psoriasis ($n$ = 31), sarcoidosis ($n$ = 47), allergies ($n$ = 4), or individuals without any specified primary diagnosis ($n$ = 24)) collected during physician appointments. The validation cohort Finland I ($n$ = 466) was built to replicate the findings from Estonia I (Table S4). This cohort included the plasma samples of the Corogene study[29] to represent subjects with CAD/ACS (CVD, $n$ = 64). The CVD group included samples from patients with chronic coronary artery disease (CAD, $n$ = 32) and with acute coronary syndrome (ACS, $n$ = 32) who were diagnosed by coronary angiography, cardiac enzyme and symptomatic findings[30] (Table S4). The control group (CTRL, $n$ = 402) included samples from individuals with no known history of CAD, ACS or MI and who were either healthy ($n$ = 125) or had different primary diagnoses ($n$ = 277) (including gum disease ($n$ = 25),[30] breast cancer ($n$ = 57),[31] myasthenia gravis ($n$ = 129), multiple sclerosis ($n$ = 20),[32] narcolepsy ($n$ = 16),[33] or schizophrenia ($n$ = 30)).

To examine confounding of age and gender, matched subgroups were generated from Estonia II and Finland I cohorts using R package "MatchIt". Estonia II sex and age matched subgroups for CTRL and MI included age and gender matched males from *No vaccine* group (*No vaccine-M-CTRL* ($n$ = 13), *No vaccine-M-MI* ($n$ = 18)), and females (*No vaccine-F-CTRL* ($n$ = 8), *No vaccine-F-MI* ($n$ = 12)) (Table S5). Also, age and gender matched males from OPV group were included (*OPV-M-CTRL* ($n$ = 18), *OPV-M-MI* ($n$ = 10) as there were no females in OPV-MI group (Table S5). For Finland I, *M-CTRL* ($n$ = 50), *M-CVD* ($n$ = 50), *F-CTRL* ($n$ = 56), and *F-CVD* ($n$ = 14) subgroups were generated (Table S6).

## Mimotope-variation analysis

All study samples were analysed using the mimotope-variation analysis method (MVA,[32,33]). In brief, modified random 12-mer peptide phage library (originally derived from Ph.D.-12, New England Biolabs) was amplified according to the manufacturer's protocol. 2 $\mu$l of plasma/serum samples, previously precleared to plastic and E. coli/wt M13 phage particles, were incubated with 2.5 $\mu$l library (~$5 \times 10^{11}$ phage particles) and immunoglobulin G (IgG) fraction was recovered using protein G-coated magnetic beads (S1430S, New England Biolabs). IgG-bound phage DNA was analysed by next generation sequencing (Illumina HiSeq, 50-bp single end reads). Sequence data were screened for sequencing errors and known artefacts, yielding specific peptide datasets for further analysis. A brief description of MVA method is provided in Figure S3.

## Data processing and motif clustering

Data processing was performed as described previously.[32,33] Final dataset consisted of $1{,}3 \times 10^6$ unique peptides across Estonia I discovery cohort

samples. For further analysis, we selected peptides by the criteria that a peptide was to be present in at least 10% of the samples in the study group and detected at least 10 times in one sample. For sequence-based peptide analysis and hypergeometric tests, SPEXS2 algorithm was used (http://egonelbre.github.io/spexs2/) with the following criteria: hypergeometric p-value $<10^{-5}$; motif present in ≥4 distinct peptides; ≥4 fixed amino acid positions. As a result, 3260 and 759 distinct epitopes were identified for HC and MI groups, respectively. As some of the epitopes observed in HC and MI groups were identical, 3711 unique epitope consensuses were found altogether.

### Data analysis and statistical methods

Fisher scores for unique 3711 epitope consensuses were calculated to find HC and MI group-differentiating epitopes (threshold >0.04). For further analysis, 202 the most characteristic group-differentiating epitopes were selected. Epitope-forming clusters were defined using Pearson correlation matrix analysis with hierarchical clustering of peptides based on log10 abundance values. Visualisation of Pearson correlation matrix was performed using R statistical programming in RStudio environment using "pheatmap" package.[34] To define characteristic sequence logos for each cluster, position weight matrices were generated and visualised as sequence logos using in-house "MotifTree" algorithm.[35]

Pairwise comparisons with Mann-Whitney U test were performed for discovering the difference in antibody response to the peptides of epitopes between analysed groups. All comparisons made are indicated in the figures with brackets between respective groups. In the cases of multiple comparisons, p.adjust function in R was used to adjust p-values using Holm-Bonferroni correction. Boxplots were used to display variation in samples for tested groups and statistical results. Each boxplot indicates the median (*bold line*), 25th and 75th percentiles (*lower* and *upper lines* of the box), the upper whisker of the boxplots extends from upper line of the box to the largest value no further than 1.5 * interquartile range (IQR, the distance between the 25th and 75th percentiles), the lower whisker extends from the lower line of the box to the smallest value at most 1.5 * IQR. Boxplots were created using "ggpubr"[36] package in RStudio environment. y-axis of boxplots represents total abundance of peptides meaning the sum of decimal logarithms of the peptide sequence counts detected in one individual sample.

For selecting the group-differentiating peptides contributing to the 51 epitopes with G-I-X-D consensus sequence, Fisher scores were calculated (threshold >0.03), resulting in 140 unique peptides (TOP140). TOP140 peptide abundance values were used in heatmap image analyses for the visualisation of differences in the antibody response between tested groups. Heatmap image analyses were visualised using "pheatmap"[34] package in RStudio environment.

Kendall correlation analysis was performed for studying correlations between MVA and ELISA or dot ELISA results and Spearman correlation analysis for studying correlations between age and the strength of antibody response to the G-I-X-D epitope. "ggpubr"[36] package in RStudio environment was used to visualise correlation analyses.

For the initial G-I-X-D epitope identification, IEDB database was used (v3.0, date accessed: 24.03.2021). Since IEDB search engine does not allow searches with undetermined amino acids in the query sequences the peptide sequence G-I-E-D-L was used. The undefined amino acid in G-I-X-D was assigned as glutamic acid (E) based on observational data of amino acids in the peptides contributing to G-I-X-D. Leucine (L) was added to the 4 amino acid motif-based on observational data of amino acids in the peptides contributing to G-I-X-D to lengthen the motif and garner more substantial and specific results.

To study the origin of G-I-X-D epitope more precisely, peptides characterising Cluster I, II, and III, were selected by Fisher score analysis (TOP258, threshold >0.03). Alignments of TOP258 peptides were performed against 100 amino acid fragments of picornaviruses (Table S2) using standalone BLAST (v. 2.8.1). Alignments were performed using "blastp-short" task[37] which is optimised for query sequences shorter than 30 residues (Scoring matrix: PAM30). The t-distributed Stochastic neighbor Embedding (t-SNE) analysis was performed for the visualisation of the alignment results as plots using the "Rtsne" package in R.[38]

To assess biomarker performance on predicting MI diagnosis, receiver operating characteristic (ROC) analysis was performed on Estonia I cohort ($n = 100$; Table 1). Generalised linear model-based statistical analysis was performed with "glm" function using gaussian distribution and identity link, and visualised with "jtools" package, statistical comparison of different models was done using "anova" function with F-test in R. Also, packages "ggplot2" and "tidyverse"[39] in RStudio environment were used for data manipulation and visualisation.

For age and sex-matched cohort design, the age and sex of study subjects were fixed at the time of sample collection considering that MI samples were collected < 12 h before or after angiography procedures. Finland I cohort CVD samples were collected as in.[29]

### Phage dot ELISA

M13 phage particles displaying the relevant peptides were printed onto nitrocellulose coated glass slides (10,485,323, Whatman) using SpotBot® 4 arrayer (Arrayit) in four dilutions (1:2, 1:10, 1:100, 1:1000). 1xPBS-20% glycerol and mQ were used as negative controls and human IgG serial dilutions (12−50 ng/µl) as

positive controls. Human serum or plasma (1:50), previously precleared to plastic and *E. coli*/wild type M13 phage antigens, was used as primary antibody while rabbit anti-human IgG H&L (HRP) (#ab6759, Abcam, RRID: AB_955,434) (1:1000) was used as a secondary antibody. Both antibody incubations were conducted at room temperature for 1 h with agitation. The presence of bound human IgG antibodies was detected using reaction with catalyzed Signal Amplification (CSA) System II, Biotin-Free, HRP, DAB+ (Dako, Agilent, Cat#: K1497), diluted in the substrate buffer (1:100). Results were visualised using Ettan DigeImager (GE Healthcare Life Sciences) and signals (image colour-intensities) quantified using ImageQuant software version 8.1 (GE Healthcare Life Sciences).

### Human coxsackievirus IgG ELISA

Anti-coxsackievirus B1, B3 and B5 VP1 epitopes serostatus was measured from 14 (7 HC and 7 MI patients) serum or plasma samples. In brief, serological analyses were performed with anti-coxsackievirus IgG ELISA method (ESR134G, SERION ELISA classic) in duplicates according to manufacturer's protocol. Absorbance was measured using SpectraMax Paradigm microplate reader (Molecular Devices). The results were interpreted according to manufacturer's specifications.

### Role of funding source

We confirm no role of funders in study design, data collection, analysis and interpretation, or writing the report. The corresponding author had full access to all the data of this study and submitted the manuscript for publication.

## Results

### Characterisation of the top antibody response in MI patients

In order to explore antibody epitope features associated with MI, we generated immunoprofiles of the myocardial infarction patients (MI; $n = 50$) and healthy donors (HC; $n = 50$) constituting the discovery cohort Estonia I (Table 1, Figure S1 and Table S1) using the next generation peptide phage display MVA technology[32,33] (Figure S3). In brief, individual plasma or serum samples were incubated with M13 phage-displayed 12-mer random peptide library to capture individual-specific antibody (IgG) repertoires. The DNA of IgG-bound phages was lysed and antigen coding-sequences were amplified using PCR. Then, amplicons were subjected to high-throughput DNA sequencing and the obtained data was translated to peptides underlying the subject- and group-specific immunoprofiles. For delineating antigen recognition patterns, peptides were clustered by using unsupervised hierarchical clustering analysis yielding ultimately in 3711 consensus epitopes. From these, 202 that discriminated the healthy donors from MI patients were selected with the cut-off Fisher score values >0.04. Further Pearson correlation analysis revealed that 75 of the healthy group-specific epitopes (TOP75) formed three distinct clusters: Cluster I, II and III (Figure 1a). Cluster I comprised of peptides with G-I-X-D consensus, Cluster II of peptides containing W-W-N, and Cluster III of peptides containing [AS]-X-Y-X-[YF]-X-X-K consensus patterns (Figure 1b). Notably, high-titer antibody response to peptides containing the G-I-X-D pattern was specific for healthy people ($p = 1.5e{-}10$; Figure 1b), regardless of gender (Figures 1c and S4). Together, we observed a strong and homogenous antibody response to a few epitopes that was characteristic to healthy and absent or weak in patients with MI ($p = 1e{-}06$ for *Cluster II*, $p = 6.6e{-}06$ for *Cluster III*; Figure 1b). By assessing the G-I-X-D performance on Estonia I cohort ($n = 100$) as a biomarker in predicting MI using receiver operating characteristic (ROC) analysis, we found that measuring antibody response to a single (G-I-X-D) epitope as a biomarker could discriminate MI patients from healthy controls with 64% sensitivity at 84% specificity (AUC 0.740, CI 0.643 to 0.823, $p < 0.0001$; Figure S5).

### Conserved epitope harbours the N-terminal G-I-X-D core structure

Since we observed a strong and rather homogenous antibody response to a few antigenic epitopes that was characteristic to healthy individuals and absent or weak in patients with MI, we analytically validated the MVA findings of the immunogenicity of the biggest G-I-X-D-defined cluster (*Cluster I*, Figure 1a,b) using phage dot ELISA method (Figures 2a and S6). All TOP140 G-I-X-D epitope containing peptides encompassed G-I-X-D motif N-terminally. Therefore, to examine the importance of G-I-X-D epitope location in the peptide sequence, we analysed phages displaying peptides where G-I-X-D was located either in the N-terminus (**GIED**VMDVMPMP; Figure 2a) in reference to peptides or the epitope sequence was located further downstream from the N-terminus (ADGPGGP**GIPD**G; Figure S6). Based on MVA data, samples with either high immunoreactivity to TOP140 peptides containing the G-I-X-D epitope (abundance values > 40; *HC, nr1–5 and MI, nr1,2*) or low immunoreactivity (abundance values < 25; *HC, nr6,7 and MI, nr3–7*) from Estonia I cohort (Table S1) were selected for analysis. Phage dot ELISA analysis data confirmed the MVA data findings on the antibody reactivity to the G-I-X-D epitope (Figure 2a), but also established that N-terminal location of the epitope was imperative to discriminate between samples with high and low immunoreactivity (Kendall correlation, $R = 0.6$, $p = 0.002$; Figure 2b). Altogether, these analyses confirm that N-terminal G-I-X-D is a highly immunogenic

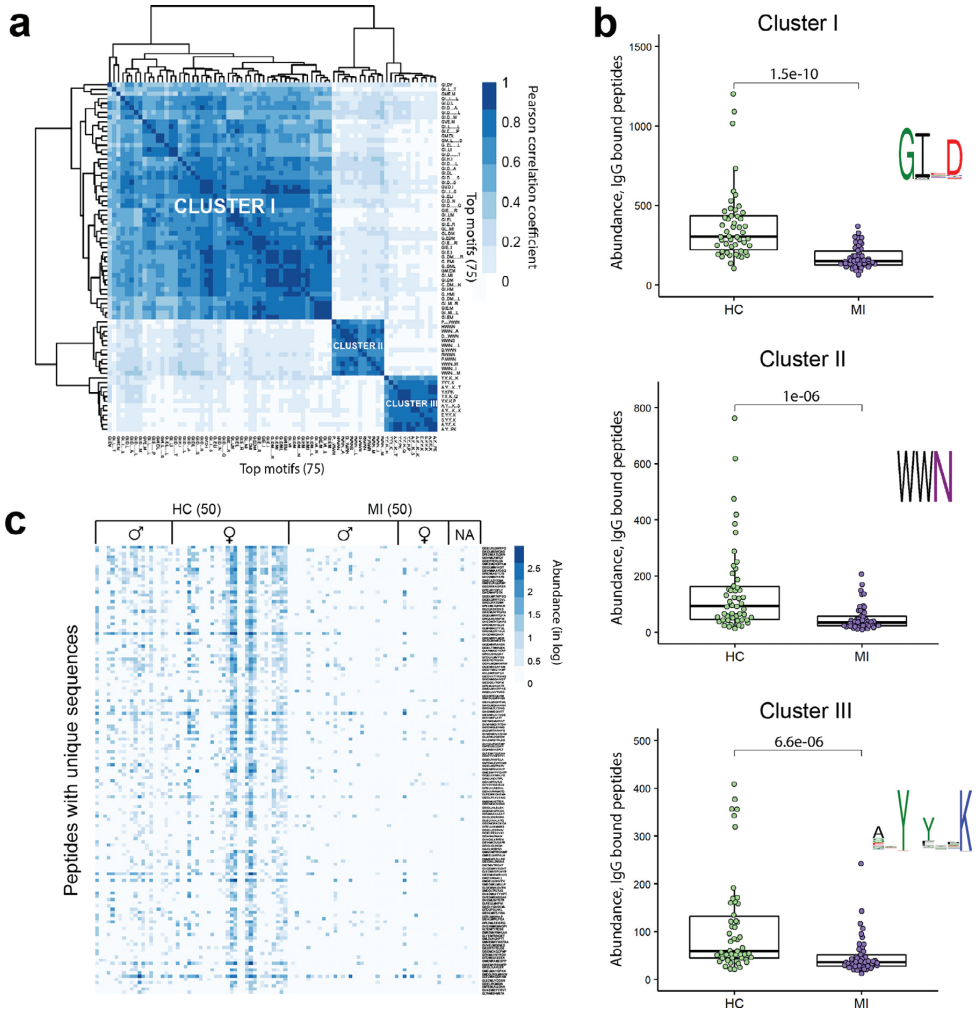



**Figure 1.** MVA analysis reveals three major epitopes associated with a differential antibody response in MI diseased compared to healthy controls. a. Clustering of peptides of the Estonia I cohort resulted altogether in 3711 unique epitope patterns. Fisher score (threshold >0.04) was used to select the most group-differentiating ones, resulting in 202 epitopes. Using Pearson correlation matrix analysis and hierarchical clustering based on log10 abundance, TOP75 epitope patterns clustered into 3 group-differentiating epitopes (Clusters I, II and III). Dark blue colour shows high correlation between TOP75 epitopes (>0.7), light blue - low correlation (<0.3). b. The immunoreactivity to the 3 group-differentiating epitopes discriminates between HC and subjects with MI. On boxplots, *bold line* indicates the median, *lower* and *upper lines* indicate 25th and 75th percentiles, and *whiskers* are shown in the style of Tukey. Pairwise comparison with Mann-Whitney U test, *p-values* reported above the brackets. Sequence logos (on the right) of G-I-X-D for Cluster I, W-W-N for Cluster II, and [AS]-X-Y-X-[YF]-X-X-K for Cluster III. *y axes - Abundance, IgG bound peptides*, anti-G-I-X-D seroresponse measured by MVA. c. Heatmap image analysis of TOP140 peptides from Cluster I shows that antibody response to this epitope is more common in healthy individuals. Each column represents the TOP140 peptide profile of the individual sample. Each line represents a peptide with unique sequence. Colour code (*Abundance in log10*) depicts the strength of immunoreactivity to unique peptides in rows containing the G-I-X-D epitope. *HC — healthy controls; MI — myocardial infarction patients.*

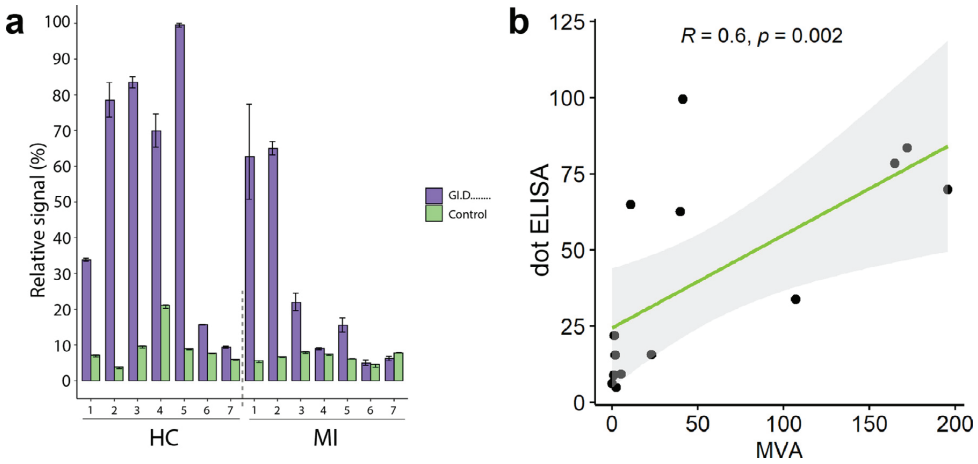



**Figure 2.** Phage dot ELISA data confirms MVA findings linking immune response to G-I-X-D core epitope with improved cardiac health. a. Seven samples from both, HC and MI groups were selected for phage dot ELISA assay based on high (HC1−5, MI1,2) or low seroresponse (HC6,7, MI3−7) to TOP140 peptides containing the G-I-X-D epitope. N-terminally G-I-X-D epitope displaying phages (**GIED**VMDVMPMP) along with FLAG epitope (DYKDDDDK) displaying phages as negative controls (*Control*) were incubated with selected serum or plasma samples. Results are shown as relative (%) to the highest detected signal across study samples (*y-axis*). Error bars indicate SEM ($n$ = 2 technical replicates). b. Phage dot ELISA and MVA data of selected samples are in strong positive correlation (Kendall correlation, $R$ = 0.6, $p$ = 0.002), whereas in moderate correlation with coxsackie viral antigens (Fig. S8a). *HC − healthy controls; MI − myocardial infarction patients*.

epitope detected in antibody immune response profiles of healthy subjects of Estonia I cohort.

### Conservation of the antibody G-I-X-D epitopes across the enteroviral species

We next used the immunoprofile data collected from the Estonia I cohort to look for possible associations of the defined motifs with pathogen-associated antigens. Substring searches of the Immune Epitope Database and Analysis Resources (IEDB, date accessed: 24.03.2021; www.iedb.org) using linear peptides with "GIEDL" sequence that contribute substantially to the harbored G-I-X-D epitope were performed. This resulted in two referenced epitopes with sequences of GIEDLISEVAQGALTLSL (ID 79,229,[40]) and GIEDLISEVAQGAL (ID 100,065,[41]), respectively, that both were derived of the genome polyprotein of EV-C (human poliovirus type 3). Next, to widen the pool of potential pathogens, substring searches with shorter "GI" sequence across Enteroviruses (ID 12,059, Enterovirus) resulted in 30 records. Data showed that PVs, coxsackieviruses A (CV-A) and coxsackieviruses B (CV-B) all share high sequence similarity encompassing amino acids 1 to 11 downstream of the VP3/VP1-specific protease 3C cleavage site (Figure 3a and Table S2). Further alignment analysis of peptides from Clusters I, II, and III differentiating the HC group of Estonia I (Figure 3b and Table S2) confirmed that Cluster I defined by the G-I-X-D consensus epitope was highly specific to the N-terminus of EV-C VP1 antigens (Figure 3c). Notably, the N-termini of VP1 proteins of the human poliovirus type 3 (PV3), coxsackieviruses A17 (CA17) and A20 (CA20) showed the highest homology with G-I-X-D-containing peptides, whereas peptides of Clusters II and III did not align to EV-C proteomes (Figure 3b). Next, we investigated whether the G-I-X-D epitope could originate from other picornaviral species. Data showed that in addition to EV-C, G-I-X-D core epitope-containing peptides aligned also to different species of EV-A, EV-B and parechovirus A types (Figure 3c and Table S2). In addition, high similarity was observed with VP1 N-termini of vaccine-derived polioviruses (VDPVs), namely VDPV3 (VDPV3/EST/Env2008 (GenBank: KC784372.1, Figure 3a). Of note, analysing our MVA data we found that seroreactivity to the recently described peptide sequences PALTAVETG and PALTAAETG, derived of the VP1 proteins of human PV-1/3 and CV-B1/B3 viruses and also shared by many other picornaviruses,[42] did not differentiate MI cases from controls (Figure S7), further confirming the functional relevance of the G-I-X-D findings. Then, we used conventional ELISA method to measure IgG response in Estonia I cohort sera ($n$ = 14; same samples as in Figure 2a) against VP1 of several enteroviral species, including coxsackieviruses B1, B3 and B5 and with reported cross-reactivity to poliovirus type 3.[43] Interestingly, differential readouts were obtained for all samples analysed and there was a

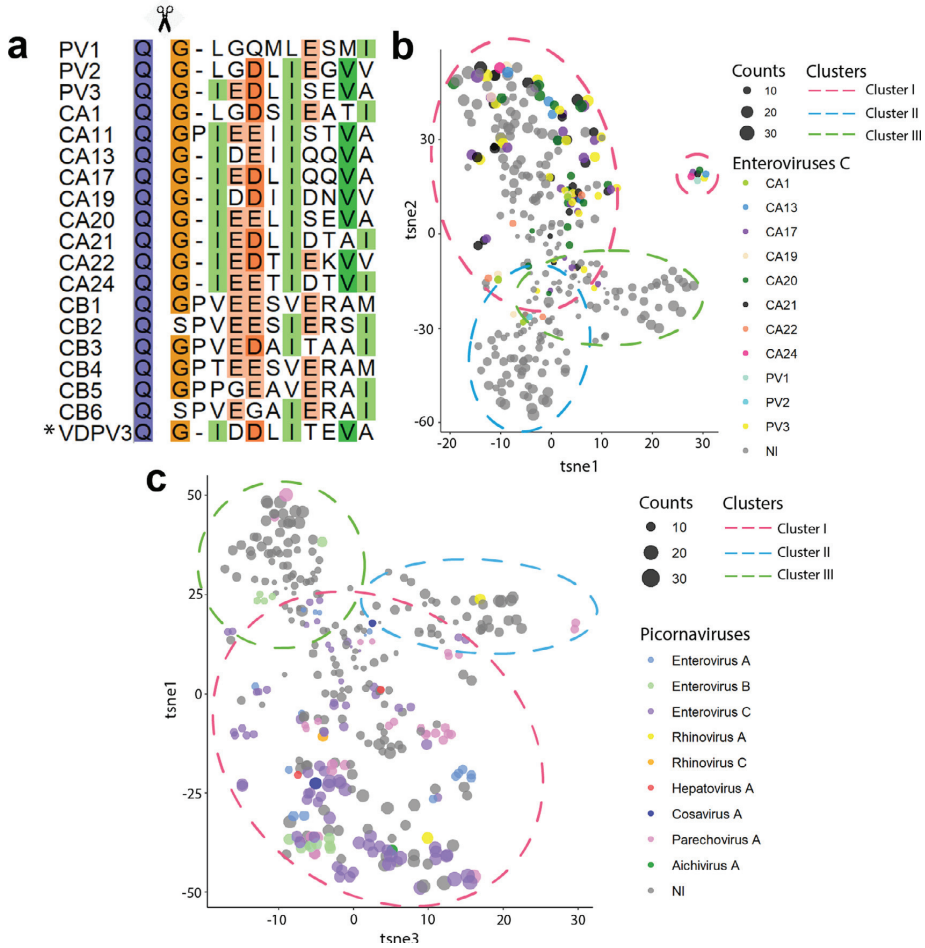



**Figure 3.** N-terminal G-I-X-D sequence is common in picornaviral VP1 proteins. a. Alignments of N-termini of VP1 sequences of poliovirus (PV) and non-PV species (Table S2) are shown. Scissors indicate PV3/PV1 protease 3C cleavage sites in PVs, VDPVs and coxsackieviruses A and B. Colours indicate >40% amino acid identities in select positions (different colours indicate different amino acids). b,c. TOP258 peptides with epitopes of Cluster I, II and III were aligned to 100 aa fragments of select enteroviral C (b) and other picornaviral proteomes (c) (Table S2) using standalone BLAST. Interrupted lines indicate peptide affiliation to the cluster *(pink-coloured line* – Cluster I, *blue-coloured line* – Cluster II, *green-coloured line* – Cluster III). The dot indicates one alignment with each of the TOP258 peptides. Peptides with no alignments or peptides aligned to one region only are shown as one dot, peptides aligned to many fragments are shown as repeated dots according to the number of alignments. The size of each dot corresponds to peptide abundance (*Counts*) across the HC cohort. Viruses are colour coded. *VDPV – vaccine-derived poliovirus; PV1, PV2, PV3 – poliovirus 1, 2, or 3, respectively; CA1, CA11, CA13, CA17, CA19, CA20, CA21, CA22, CA24, CB1, CB2, CB3, CB4, CB5, CB6 – coxsackievirus A1, A11, A13, A17, A19, A20, A21, A22, A24, B1, B2, B3, B4, B5, or B6, respectively; NI – peptides with no alignments to the picornaviruses (grey dots); y- and x-axes: tsne1, 2, or 3 – t-SNE dimension 1, 2, or 3, respectively.*

moderate linear relationship with MVA G-I-X-D findings (Kendall correlation, $R = 0.35$, $p = 0.079$; Figure S8a). However, only a few ($n = 3$) serum samples exceeded the thresholds for coxsackieviral IgG seropositivity (Figure S8b). Overall, from this data it can be concluded that the antibody response to G-I-X-D epitopes could stem from the PVs and/or PV vaccine antigens, also from vaccine-derived polioviruses (VDPVs) and/or from alike viral antigens, such of VP1 of EV-B types (coxsackieviruses B), but also other species of picornavirus family.

### Immunoreactivity to G-I-X-D epitope is associated with OPV/IPV use and is weak in patients with myocardial infarction

We hypothesised that the observed differences in seroreactivity patterns to G-I-X-D could be associated with PV immunisation (Figure S2). To investigate this hypothesis, we created the second stage of the study with the second Estonian discovery cohort (Estonia II, Figure S1) that included samples of individuals ($n$ = 329) that were grouped according to the year of birth. People born before the year 1959 when no vaccination with OPV against PV was available in Estonia comprised group *No vaccine* ($n$ = 101) (Figure 4 and Table S3). Individuals born in the years of OPV vaccination in Estonia, 1959–2007, comprised group *OPV* ($n$ = 224) and individuals born in the years of IPV vaccination, the year 2008 and later were included in *IPV* group ($n$ = 4) (Figures 4 and S2 and Table S3). Data analysis concluded that high seroreactivity to the G-I-X-D epitope was common to the *OPV* group and was low in *No vaccine* and *IPV* groups (Mann Whitney U test, $p < 0.0001$ or $p < 0.05$, respectively, Holm-Bonferroni adjustment for 3 comparisons) (Figure 4a,b). No gender differences in any vaccine groups were detected (Figures 4a and S9). Additionally, individuals without a history of MI from Estonia II cohort ($n$ = 309) in the *OPV* group showed higher seroreactivity to the G-I-X-D epitope as compared with individuals from *No vaccine* and *IPV* groups (Mann Whitney U test, $p < 0.001$ or $p < 0.05$, respectively, Holm-Bonferroni adjustment for 3 comparisons) (Figure S10). However, no correlation between the antibody response to the G-I-X-D epitope and the age of the subjects within the OPV group was found (Spearman correlation, $R = 0.071$, $p = 0.3$; Figure S11). We further used statistical modelling of data to determine whether vaccination status, age and gender affect antibody response to G-I-X-D epitope, and found that in contrast to OPV vaccination, gender and age had no significant effects on the seroresponse to G-I-X-D epitope (Table S7). Adding an interaction term between age and gender failed to improve the model (F-test, $p = 0.5092$). Therefore, we concluded that the antibody response to G-I-X-D epitope did not depend on gender or age, but only on OPV vaccination and MI diagnosis.

Next, we analysed the Estonia I cohort to evaluate association of the immunoreactivity to G-I-X-D with the risk of MI independent of the poliovirus vaccine exposure. Importantly, patients with MI exhibited a significant decrease in immune response to the G-I-X-D epitope compared to healthy individuals of *No vaccine* and *OPV* groups (Mann Whitney U test, $p < 0.05$, Holm-Bonferroni adjustment for 4 comparisons) (Figure 4c). Moreover, analysis of subgroups of the Estonia II cohorts matched by the vaccination status, age and gender (Table S5) further confirmed the findings that patients with MI showed significantly reduced response to G-I-X-D epitope compared to people without primary diagnosis of MI (Figure S12).

Next, to replicate findings of the G-I-X-D epitope association with cardiovascular disease and poliovirus vaccines, we comparatively analysed the data from the immunoprofiles of Finland I validation cohort ($n$ = 466) (Figure S1, Table S4). IPV has been the predominantly used vaccine by the Finnish national poliovirus immunisation programme, apart from years 1984–1985 with high rate of OPV use (https://www.cdc.gov/mmwr/preview/mmwrhtml/00000682.htm). Data analysis showed that antibodies against G-I-X-D epitopes were more prevalent in healthy subjects of Estonia I cohort than in the control subjects of Finland I cohort ($p < 0.0001$, Holm-Bonferroni adjustment for 4 comparisons, Figure 4d). However, the anti-G-I-X-D antibody response was significantly lower in CVD (MI or CAD/ACS) subgroup as compared to controls of both Estonia I and Finland I (Mann Whitney U test, $p < 0.0001$ and $p < 0.001$, respectively, Holm-Bonferroni adjustment for 4 comparisons, Figures 4d and S13 (CTRL vs ACS)). A weak negative correlation between age and the anti-G-I-X-D seroresponse (Spearman correlation, $R$=-0.14, $p = 0.0041$) in the Finland I control group ($n$ = 402) was observed (Figure S14). Therefore, we excluded the effects of age on the observed association between the increased risk of CVD and decreased antibody response to the G-I-X-D epitope. Analysing the age and sex-matched subgroup of the Finland I cohort (Table S6), we found that anti-G-I-X-D antibody response was significantly lower in men with CVD compared to controls (Mann Whitney U test, $p < 0.05$, Holm-Bonferroni adjustment for 2 comparisons), whereas no statistically significant effect was seen for women (Mann Whitney U test, $p = 0.83$, Holm-Bonferroni adjustment for 2 comparisons), possibly due to the small sample size for this gender group (Figure S15).

Finally, we used generalised linear model-based statistical modelling across our cohorts (Estonia II, and Finland I) to determine the relationship between the anti-G-I-X-D seroresponse and CVD, taking into account the vaccination status, age and gender (Figure 4e, Table S8). Our results indicated that vaccination with OPV was strongly associated with an increased anti-G-I-X-D seroreactivity in both men and women (34% and 29%, respectively), whereas vaccination with IPV and increasing age of patients showed slight decreasing but statistically unsignificant trends in the anti-G-I-X-D seroresponse. Importantly, the modelling across all data showed that CVD diagnosis was associated with 36% and 24% drop in the anti-G-I-X-D seroresponse in both men and women, respectively (Figure 4e, Table S8).

Altogether, our data conclude that a low titer antibody response to the G-I-X-D epitope is associated with cardiac complications. Our results additionally suggest beneficial effects of polio vaccinations with OPV even in countries that have eradicated poliovirus.

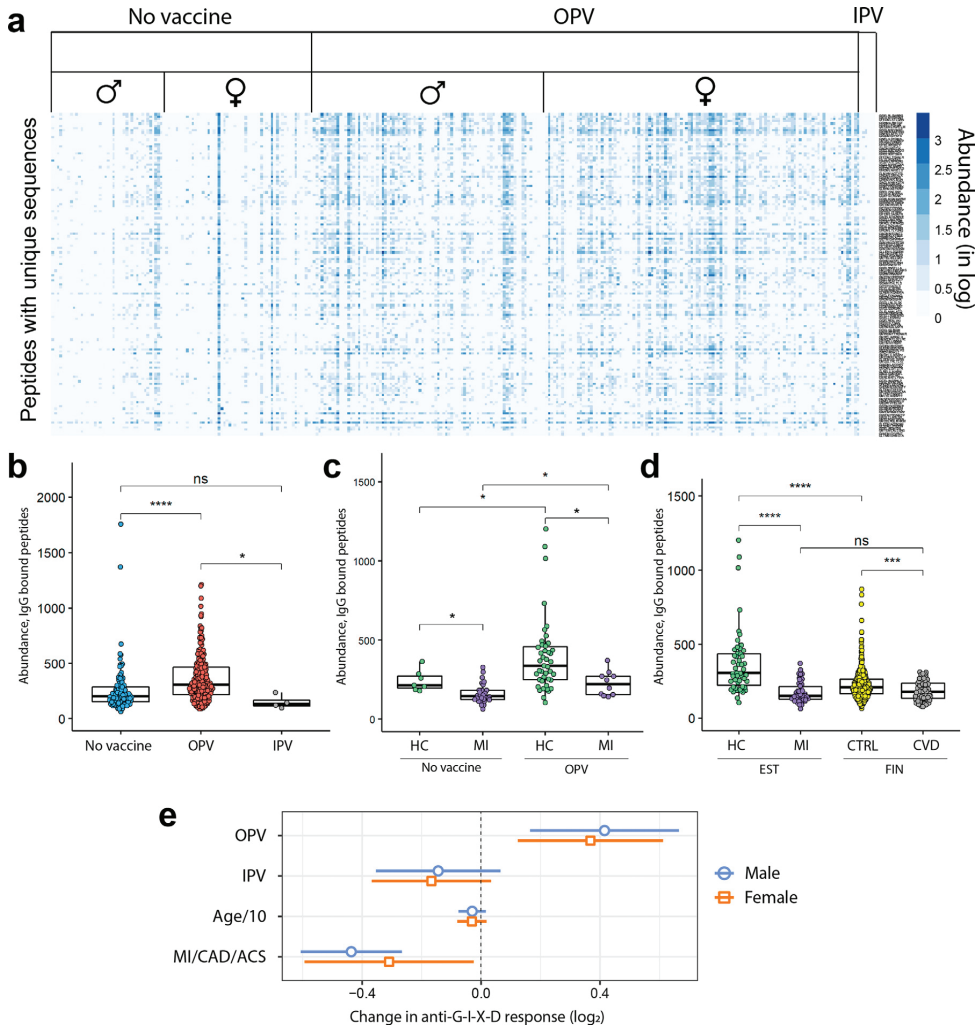



**Figure 4.** High immunoreactivity to the G-I-X-D epitope is associated with improved cardiac health. a. MVA data of subjects from Estonia II cohort with the known year of birth independent of their diagnosis were selected to analyse the association of the antibody reactivity to G-I-X-D epitope with polio vaccinations (Fig. S2). Regarding clinical background, the Estonia II cohort consisted of MI cases ($n$ = 40) and either healthy blood donors (ICD-10: Z52.0, $n$ = 135) or samples from individuals with different primary diagnoses with no history of MI ($n$ = 154). All individuals were grouped according to their year of birth, forming three major groups as follows: *No vaccine* ($n$ = 101; people born before 1959, years when no vaccination against poliomyelitis was done in Estonia), *OPV* ($n$ = 224; individuals born in years 1959−2007 when OPV vaccination in Estonia was in the national immunisation schedule), *IPV* ($n$ = 4; individuals born in 2008 and after, years of IPV vaccination in Estonia). Heatmap shows the antibody response to TOP140 peptides containing the G-I-X-D epitope. Each column represents the TOP140 peptide immunoprofiles in individual samples. Each line represents a peptide with unique sequence. Intensity of the colour correlates with the strength of immunoreactivity to the G-I-X-D peptides in rows (*Abundance*, in log10). b. People born in years of OPV vaccination in Estonia (1959−2007, *OPV*) have higher immunoreactivity to G-I-X-D peptides as compared with *No vaccine* or *IPV* groups. Groups: *No vaccine* ($n$ = 101); *OPV* ($n$ = 224) and *IPV* ($n$ = 4), the entire Estonia II cohort, including MI cases (Table S3). c. High immunoreactivity to peptides with G-I-X-D detected in healthy individuals born in Estonia in years of OPV vaccination (1959−2007, *HC-OPV*) as well as in those born before 1959 (*HC-No vaccine*) as compared to low antibody response observed in MI patients (*MI-No vaccine* and *MI-OPV*). Groups: *HC-No vaccine* ($n$ = 7); *HC-OPV* ($n$ = 43); *MI-No vaccine* ($n$ = 30); *MI-OPV* ($n$ = 10). d. High immune reactivity to G-I-X-D epitope observed in subjects of

## Discussion

Herein we identified a highly antigenic epitope located at the N-terminus of VP1 proteins of EV-B and -C type enteroviruses (EVs), including polioviruses (PVs), and observed that strong antibody immune response to this epitope was associated with improved cardiovascular health. Thus, this epitope has the potential as a site of broadly neutralising reactivity against EVs and possibly other species of the picornavirus family. Secondly, the conserved G-I-X-D epitope has potential as blood biomarker for risk assessments of major cardiac events, as poor response against the conserved G-I-X-D epitope was strongly associated with severe cardiac conditions of MI and CAD/ACS. Still, the mechanisms by which inadequate antibody response to the G-I-X-D epitope contributes to the progression of MI need further studies. Lastly, our data highlight the potential beneficial role of heterologous cross-reactive immunity of PV vaccines or natural EV infections against a spectrum of clinical heart conditions.

Exploratory data analysis of this study using MVA identified an epitope mimicking the N-terminus of VP1 protein of EV-C type and alike viral species that was strongly associated with improved cardiac health (Figure 1). Namely, poor antibody response to this epitope with G-I-X-D consensus stratified MI subjects from healthy with 64% sensitivity at 84% specificity (Figure S5). Further analytical validation and annotation analyses concluded that the major antigenic site (the N-terminal glycine of the processed VP1 viral antigen) is highly conserved across EVs, in particular among the EV-B and EV-C species, also in PVs and VDPVs, and additionally among other *Picornaviridae* (Figure 3c). Comparative conventional ELISA testing for enteroviral anti-VP1 seroresponse showed limited concordance with MVA findings, further supporting the wider viral origin of the G-I-X-D epitope. These data have global implications. The genus *Enterovirus* includes rhinoviruses and EVs (types A to D, with EV-C including PVs 1 to 3) that infect humans.[44] The climate and socio-economic factors contribute to the geographic distribution of different EV types.[44,45] In Europe and the United States, for example, EVs are found in 5–12% of the population.[46] Most of these belong to EV-B, whereas in Asia, EV-A types are common.[47–49] In contrast, non-polio type EV-Cs are relatively rare.[45] Notably, EV genotypes change over time primarily through recombination of both conventional and unconventional EV types that most frequently occurs at non-capsid regions, as reported for EV-As.[50] Viruses constituting the oral poliovirus vaccine (OPV) are inherently genetically instable.[51,52] Recombination with other EV-C types into highly divergent strains[46] leads to the OPV-derived VDPVs that achieve the level of transmissibility similar to natural PVs[53–55] and can cause acute flaccid paralysis.[56] It is well documented that following persistent EV infections, cardiac tissues become inflamed[57] and that EV types A and B are responsible for cardiac disease.[58,59] In this study, we show that the G-I-X-D epitope is well-conserved across EVs and even wider across other species of *Picornaviridae,* suggesting that heterologous immunity can be mediated by cross-reactive antibodies to shared antigens of this rapidly evolving virus family.

The immunoprofiling of people of Estonia II cohort born during the OPV vaccination (1959–2007) revealed that high titer antibody response against the G-I-X-D epitope was more common compared to those individuals born before 1959, when no vaccines were available (Figure 4b), and also compared to the individuals of Finland I cohort, who had mostly received IPV polio vaccine according to the Finnish national immunisation programmes (https://www.cdc.gov/mmwr/preview/mmwrhtml/00000682.htm) (Figure 4d,e). These differences in the observed anti-G-I-X-D response could be explained by the difference in the antigenic structure of poliovirus vaccines. Commercial IPVs are produced by incubations with formaldehyde that is known to result in some loss of immunogenicity due to the partial destruction of specific antigenic epitopes.[60] Notably, the antibody response to VP3/VP1 cleavage site of PVs has been detected in OPV but not in IPV recipients, thus leaving a gap in the PV-associated immune response.[61] Most recent studies have shown high sequence variability of the VP1 N-terminal domain,[62] that is externalised in the $\alpha$-helical conformations for interactions with host cell membranes.[63, 64] Despite the

Finland I with no CVD diagnostic history (*CTRL-FIN*) and in donors from Estonia I (*HC-EST*) as compared to individuals from groups of cardiovascular disease (CAD/ACS, *CVD-FIN* and MI, *MI-EST*, correspondingly). Groups: *HC-EST* ($n$ = 50) and *MI-EST* ($n$ = 50) – Estonia I discovery cohort (Table 1); *CTRL-FIN* ($n$ = 402) and *CVD-FIN* ($n$ = 64) – Finland I cohort (Table S4). b–d. Pairwise comparison with Mann-Whitney U test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, Holm-Bonferroni correction was used to adjust $p$-values for multiple comparisons. On boxplots, *bold line* indicates the median, *lower* and *upper lines* indicate 25th and 75th percentiles, and *whiskers* are shown in the style of Tukey. *y axes – Abundance, IgG bound peptides,* anti-G-I-X-D seroresponse measured by MVA. *HC – healthy controls; MI – myocardial infarction patients; CVD – cardiovascular disease (including acute coronary syndrome patients (CAD/ACS)); CTRL – control group* e. Modelling of anti-G-I-X-D response in Estonia II and Finland I cohorts with generalised linear model using the formula "$\log_2$(G-I-X-D abundance) ~ Vaccination status + Age/10 + CVD diagnosis". People younger than 18 years were excluded from the analysis, non-vaccinated people with no primary diagnosis of CVD were used as the reference level. Estimates of the indicated coefficients (*y-axis*) are shown with circles and squares, 95% confidence interval is shown with line. *OPV* and *IPV* indicate vaccination with the respective vaccine, *Age/10* indicates age divided by 10 (i.e., the effect of being 10 years older), *MI/CAD/ACS* designates any of the indicated primary diagnosis.

VP1 N-terminal hypervariability, the Gly residue of the Q/G dipeptide forming the VP3/VP1 cleavage site[65] that is part of the G-I-X-D epitope defined by our study, remains highly conserved.[66] Thus, the findings of the current study suggest that the data on the G-I-X-D epitope could be useful in advancing poliovirus vaccines or in testing vaccine efficacy against evolving PV strains. To the best our knowledge, there is neither monoclonal antibody that would correlate with protective properties in humans and be used to estimate polio vaccine potency nor a validated test for the analysis of epitope-specific poliovirus-induced antibody response. This is important, given that the same widely accepted structural rearrangement model of the VP1 capsid proteins is valid across picornaviruses.[64]

Here we define the role of an EV-associated common epitope as protective against acute CVD events (MI and ACS). The latter is not surprising as EVs have been detected in atherosclerotic plaques[67] and, as highlighted recently, in association with MI plaque disruptions.[68] In one study, antibody response to EV antigens was detected in 49% of patients with CAD and in 54.3% of those with MI.[69] Other studies observed that coxsackievirus B (CV-B) promoted atherosclerosis in animal models.[70] However, the mechanisms of how EVs contribute to CVDs are not clear. Based on our study data, inadequate antibody response against the N-terminus of PV1 presenting the G-I-X-D epitope could hypothetically allow direct infection and lytic effects or the poor clearance of the EV species and lead to chronic inflammation of the heart tissue. We observed that anti-G-I-X-D antibody response was relatively lower in subjects of Finland I in general as compared to Estonia I cohort and was determined as particularly low in the MI subgroup of Estonia I and CVD subgroup of Finland I (Figures. 4d and S13). Based on our modelling data (Figure 4e), we suggest that these differences could be explained by the differential use of PV vaccines (OPV vs IPV) between countries, also by individual heterologous immunity towards EV/picornaviral species, both of which are essential to support cardiovascular health. Potentially, changes in the gut microbiota composition linked with CVD[71] could impair gut immune regulation promoting EV replication.[72] Our data highlight the importance of the long-term, stable, and high-titer antibody response against the N-terminal epitope of VP1 antigen of EVs in protecting against severe heart damage (MI, CAD, ACS). However, further studies are needed to establish robust correlations.

The main strengths of this study are concordant results in two ethnic populations covering a wide age range that the antibody response to a highly antigenic epitope reflects improved cardiovascular health. Additionally, the study includes a comprehensive analysis of the antibody epitope repertoire associated with enteroviruses and cardiovascular health. The main weaknesses are associated with the retrospective study design that we complemented with a descriptive analysis of publicly available data on national immunisation programmes against polio in Estonia and in Finland (Figure S2). We describe a potential biomarker for the assessment of risks associated with severe CVD (MI and ACS) and raise important questions on the role of common viral infections in the heart pathophysiology. The exact mechanisms by which these may lead to heart failure remain to be established. Our results stress the importance of dissecting the role of heterologous immunity against EV infections, also primed by globally used PV vaccines.

### Contributors

**NP**, **AA**, **MJ**, **HS**, **AR**, **AP** (Arno Pihlak), and **KP** conceived and designed the study. **AP** (Anu Planken), **MP**, **EK**, **MRJS**, **PJT**, **JKN**, **AV**, **DL**, **JS**, **PP** and **TT** collected samples and acquired the clinical data. **NP**, **AA**, **JT**, **HS**, **AP** (Arno Pihlak), **MJ**, **AR**, and **KP** analysed and interpreted the data. **NP**, **AA**, **JT**, **AP**, **MJ**, and **AR** verified the data. **EK**, **PJT**, **DL**, **PP**, **TT**, and **KP** supervised the study. **NP** and **KP** drafted the manuscript. All authors had access to the data in the study. All authors revised and approved the final manuscript for submission.

### Data sharing statement

The data are not publicly available due to containing information that could compromise research participant privacy/consent. Any materials that can be shared will be released via a material transfer agreement.

### Declaration of interests

AP (Arno Pihlak) and KP are inventors of the patent application (PCT Application No. US/14,079,626) filed by Protobios that covers the use of phage display method for manipulating and monitoring humoral immunity. NP, HS, AP (Arno Pihlak), MJ, JT, AR, TT, and KP are affiliated with Protobios Llc. JS reported payment/honoraria from Astra- Zeneca, Amgen, Abbott and Bayer. All other authors declare no competing interests.

### Acknowledgments

We thank A.-H. Pool, Department of Neuroscience, University of Texas Southwestern Medical center, for his expertise that supported this work. The members of Protobios team are thanked for their excellent technical assistance. This study was supported by institutional research funding grants of Protobios (5.1-4/20/170, and PRG573) from the Estonian Ministry of Education and Estonian Research Council, respectively, and H2020-MSCA-RISE-2016 (EU734791) and H2020 PANBioRA (EU760921) projects from the European Union, Helsinki University Hospital grants, Mary and Georg C.

Ehrnrooth Foundation, and Finnish Eye Foundation. AA, MP and TT were partially supported by Estonian Research Council (grant PRG805) and European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012). DL was supported by Finska Läkaresällskapet and The Finnish Society of Sciences and Letters. AV was supported by Magnus Ehrnrooth Foundation and Sigrid Jusélius Foundation. JT was partially supported by the grant from Estonian Ministry of Education and Research (2014-2020.4.01.21-0315).

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2022.103835.

### References


1 Roth GA, Mensah GA, CO Johnson, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. *J Am Coll Cardiol*. 2020;76(25):2982–3021.
2 Timmis A. European heart journal - quality of care & clinical outcomes: becoming the go-to journal for cardiovascular outcomes research worldwide. *Eur Heart J*. 2019;40(46):3753.
3 Michalak M, Agellon LB. Stress coping strategies in the heart: an integrated view. *Front Cardiovasc Med*. 2018;5:168.
4 Fox KAA, Metra M, Morais J, Atar D. The myth of 'stable' coronary artery disease. *Nat Rev Cardiol*. 2020;17(1):9–21.
5 Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med*. 2005;352(16):1685–1695.
6 Ross R. Atherosclerosis is an inflammatory disease. *Am Heart J*. 1999;138(5 Pt 2):S419–S420.
7 Rosenfeld ME, Campbell LA. Pathogens and atherosclerosis: update on the potential contribution of multiple infectious organisms to the pathogenesis of atherosclerosis. *Thromb Haemost*. 2011;106(5):858–867.
8 Murillo H, Restrepo CS, Marmol-Velez JA, et al. Infectious diseases of the heart: pathophysiology, clinical and imaging overview. *Radiographics*. 2016;36(4):963–983.
9 Sato K, Sinclair JE, Sadeghirad H, Fraser JF, Short KR, Kulasinghe A. Cardiovascular disease in SARS-CoV-2 infection. *Clin Transl Immunol*. 2021;10(9):e1343.
10 Szwed P, Gasecka A, Zawadka M, et al. Infections as novel risk factors of atherosclerotic cardiovascular diseases: pathophysiological links and therapeutic implications. *J Clin Med*. 2021;10(12):2539.
11 Mitra S, Drautz-Moses DI, Alhede M, et al. In silico analyses of metagenomes from human atherosclerotic plaque samples. *Microbiome*. 2015;3:38.
12 Pothineni NVK, Subramany S, Kuriakose K, et al. Infections, atherosclerosis, and coronary heart disease. *Eur Heart J*. 2017;38(43):3195–3201.
13 Patel P, Mendall MA, Carrington D, et al. Association of helicobacter pylori and chlamydia pneumoniae infections with coronary heart disease and cardiovascular risk factors. *BMJ*. 1995;311(7007):711–714.
14 Pussinen PJ, Jousilahti P, Alfthan G, Palosuo T, Asikainen S, Salomaa V. Antibodies to periodontal pathogens are associated with coronary heart disease. *Arterioscler Thromb Vasc Biol*. 2003;23(7):1250–1254.
15 Reunanen A, Roivainen M, Kleemola M, et al. Enterovirus, mycoplasma and other infections as predictors for myocardial infarction. *J Intern Med*. 2002;252(5):421–429.
16 Roivainen M, Viik-Kajander M, Palosuo T, et al. Infections, inflammation, and the risk of coronary heart disease. *Circulation*. 2000;101(3):252–257.
17 Espinola-Klein C, Rupprecht HJ, Blankenberg S, et al. Impact of infectious burden on progression of carotid atherosclerosis. *Stroke*. 2002;33(11):2581–2586.
18 Rupprecht HJ, Blankenberg S, Bickel C, et al. Impact of viral and bacterial infectious burden on long-term prognosis in patients with coronary artery disease. *Circulation*. 2001;104(1):25–31.
19 Zhu J, Quyyumi AA, Norman JE, et al. Effects of total pathogen burden on coronary artery disease risk and C-reactive protein levels. *Am J Cardiol*. 2000;85(2):140–146.
20 Pesonen E, Andsberg E, Ohlin H, et al. Dual role of infections as risk factors for coronary heart disease. *Atherosclerosis*. 2007;192(2):370–375.
21 Burgner DP, Sabin MA, Magnussen CG, et al. Early childhood hospitalisation with infection and subclinical atherosclerosis in adulthood: the Cardiovascular Risk in Young Finns Study. *Atherosclerosis*. 2015;239(2):496–502.
22 Pussinen PJ, Paju S, Koponen J, et al. Association of childhood oral infections with cardiovascular risk factors and subclinical atherosclerosis in adulthood. *JAMA Netw Open*. 2019;2:(4) e192523.
23 Soehnlein O, Libby P. Targeting inflammation in atherosclerosis - from experimental insights to the clinic. *Nat Rev Drug Discov*. 2021;20(8):589–610.
24 Lawson JS. Multiple infectious agents and the origins of atherosclerotic coronary artery disease. *Front Cardiovasc Med*. 2016;3:30.
25 Musher DM, Abers MS, Corrales-Medina VF. Acute infection and myocardial infarction. *N Engl J Med*. 2019;380(2):171–176.
26 Warren-Gash C, Blackburn R, Whitaker H, McMenamin J, Hayward AC. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. *Eur Respir J*. 2018;51(3):1701794.
27 Chumakov MP. Some results of the work on mass immunization in the Soviet Union with live poliovirus vaccine prepared from Sabin strains. *Bull World Health Organ*. 1961;25:79–91.
28 Horstmann DM. The Sabin live poliovirus vaccination trials in the USSR, 1959. *Yale J Biol Med*. 1991;64(5):499–512.
29 Vaara S, Nieminen MS, Lokki ML, et al. Cohort profile: the Corogene study. *Int J Epidemiol*. 2012;41(5):1265–1271.
30 Liljestrand JM, Paju S, Pietiainen M, et al. Immunologic burden links periodontitis to acute coronary syndrome. *Atherosclerosis*. 2018;268:177–184.
31 Mustonen L, Aho T, Harno H, Sipila R, Meretoja T, Kalso E. What makes surgical nerve injury painful? A 4-year to 9-year follow-up of patients with intercostobrachial nerve resection in women treated for breast cancer. *Pain*. 2019;160(1):246–256.
32 Sadam H, Pihlak A, Jaago M, et al. Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients. *EBioMedicine*. 2021;64: 103211.
33 Sadam H, Pihlak A, Kivil A, et al. Prostaglandin D2 receptor DP1 antibodies predict vaccine-induced and spontaneous narcolepsy type 1: large-scale study of antibody profiling. *EBioMedicine*. 2018;29:47–59.
34 Kolde R. pheatmap: Pretty Heatmaps [Internet]. 2019 [cited 2021 Apr 21]. Available from: https://CRAN.R-project.org/package=pheatmap.
35 Kruup M. Finding motifs from short peptides [Internet]. 2013 [cited 2021 Jun 11]. Available from: https://comserv.cs.ut.ee/home/files/Mari-Liis+Kruup+-+Finding+motifs+from+short+peptides.pdf?study=ATILoputoo&reference=C2F24DE244D358C55E32D7A221FEE5CB523D4E9C.
36 Kassambara A. ggpubr: "ggplotz" Based Publication Ready Plots [Internet]. 2020 [cited 2021 Apr 21]. Available from: https://CRAN.R-project.org/package=ggpubr.
37 Madden T. BLAST+ features. BLAST® Command Line Applications User Manual. National Center for Biotechnology Information (US) [Internet]. 2020 [cited 2021 Oct 1]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK569839/.
38 Krithe JH. Rtsne: t-distributed stochastic neighbor embedding using a Barnes-Hut implementation [Internet]. 2015 [cited 2021 Apr 21]. Available from: https://github.com/jkrijthe/Rtsne.
39 Wickham H, Averick M, Bryan J, Chang W, McGowan L. Welcome to the tidyverse. *J Open Source Softw*. 2019;4:43.
40 Graham S, Wang EC, Jenkins O, Borysiewicz LK. Analysis of the human T-cell response to picornaviruses: identification of T-cell epitopes close to B-cell epitopes in poliovirus. *J Virol*. 1993;67(3):1627–1637.
41 Roivainen M, Piirainen L, Rysa T, Narvanen A, Hovi T. An immunodominant N-terminal region of VP1 protein of poliovirion that is buried in crystal structure can be exposed in solution. *Virology*. 1993;195(2):762–765.
42 Harb J, Mennesson N, Lepetit C, et al. Comparison of monoclonal gammopathies linked to poliovirus or coxsackievirus vs. other infectious pathogens. *Cells*. 2021;10(2):438.

43 Swanink CM, Veenstra L, Poort YA, Kaan JA, Galama JM. Coxsackievirus B1-based antibody-capture enzyme-linked immunosorbent assay for detection of immunoglobulin G (IgG), IgM, and IgA with broad specificity for enteroviruses. *J Clin Microbiol*. 1993;31 (12):3240–3246.
44 Lukashev AN, Vakulenko YA. Molecular evolution of types in non-polio enteroviruses. *J Gen Virol*. 2017;98(12):2968–2981.
45 Brown DM, Zhang Y, Scheuermann RH. Epidemiology and sequence-based evolutionary analysis of circulating non-polio enteroviruses. *Microorganisms*. 2020;8(12):E1856.
46 Brouwer L, van der Sanden SMG, Calis JCJ, et al. High frequency of polio-like enterovirus C strains with differential clustering of CVA-13 and EV-C99 subgenotypes in a cohort of Malawian children. *Arch Virol*. 2018;163(10):2645–2653.
47 Mohamadpoor T, Nabavinia M, Gholoobi A, Alavi M, Meshkat Z. Enteroviruses in acute myocardial infarction. *Iran J Public Health*. 2012;41(8):71–74.
48 Roger VL. Epidemiology of myocardial infarction. *Med Clin N Am*. 2007;91(4):537–552. ix.
49 Shin SY, Kim KS, Lee YS, et al. Identification of enteroviruses by using monoclonal antibodies against a putative common epitope. *J Clin Microbiol*. 2003;41(7):3028–3034.
50 Wang M, Zhu L, Fan J, et al. Rules governing genetic exchanges among viral types from different enterovirus a clusters. *J Gen Virol*. 2020;101(11):1145–1155.
51 Agol VI. Molecular mechanisms of poliovirus variation and evolution. *Curr Top Microbiol Immunol*. 2006;299:211–259.
52 Agol VI. Vaccine-derived polioviruses. *Biologicals*. 2006;34(2):103–108.
53 Duintjer Tebbens RJ, Pallansch MA, Chumakov KM, et al. Review and assessment of poliovirus immunity and transmission: synthesis of knowledge gaps and identification of research needs. *Risk Anal*. 2013;33(4):606–646.
54 Famulare M, Selinger C, McCarthy KA, Eckhoff PA, Chabot-Couture G. Assessing the stability of polio eradication after the withdrawal of oral polio vaccine. *PLOS Biol*. 2018;16(4):e2002468.
55 Jenkins HE, Aylward RB, Gasasira A, et al. Implications of a circulating vaccine-derived poliovirus in Nigeria. *N Engl J Med*. 2010;362(25):2360–2369.
56 Fernandez-Garcia MD, Kebe O, Fall AD, Ndiaye K. Identification and molecular characterization of non-polio enteroviruses from children with acute flaccid paralysis in West Africa, 2013-2014. *Sci Rep*. 2017;7(1):3808.
57 Bouin A, Gretteau PA, Wehbe M, et al. Enterovirus persistence in cardiac cells of patients with idiopathic dilated cardiomyopathy is linked to 5′ terminal genomic RNA-deleted viral populations with viral-encoded proteinase activities. *Circulation*. 2019;139(20):2326–2338.
58 Dennert R, Crijns HJ, Heymans S. Acute viral myocarditis. *Eur Heart J*. 2008;29(17):2073–2082.
59 Knowlton KU. Dilated cardiomyopathy. *Circulation*. 2019;139 (20):2339–2341.
60 Wilton T, Dunn G, Eastwood D, Minor PD, Martin J. Effect of formaldehyde inactivation on poliovirus. *J Virol*. 2014;88 (20):11955–11964.
61 Herremans T, Reimerink JH, Kimman TG, van Der Avoort HG, Koopmans MP. Antibody responses to antigenic sites 1 and 3 of serotype 3 poliovirus after vaccination with oral live attenuated or inactivated poliovirus vaccine and after natural exposure. *Clin Diagn Lab Immunol*. 2000;7(1):40–44.
62 Shah PNM, Filman DJ, Karunatilaka KS, et al. Cryo-EM structures reveal two distinct conformational states in a picornavirus cell entry intermediate. *PLOS Pathog*. 2020;16:(9) e1008920.
63 Tschope C, Ammirati E, Bozkurt B, et al. Myocarditis and inflammatory cardiomyopathy: current evidence and future directions. *Nat Rev Cardiol*. 2021;18(3):169–193.
64 Jimenez-Clavero MA, Escribano-Romero E, Douglas AJ, Ley V. The N-terminal region of the VP1 protein of swine vesicular disease virus contains a neutralization site that arises upon cell attachment and is involved in viral entry. *J Virol*. 2001;75(2):1044–1047.
65 Xia X, Xie Z. Protein structure, neighbor effect, and a new index of amino acid dissimilarities. *Mol Biol Evol*. 2002;19 (1):58–67.
66 Shaw J, Jorba J, Zhao K, et al. Dynamics of evolution of poliovirus neutralizing antigenic sites and other capsid functional domains during a large and prolonged outbreak. *J Virol*. 2018;92(9): e01949–17.
67 Campbell LA, Rosenfeld ME. Infection and atherosclerosis development. *Arch Med Res*. 2015;46(5):339–350.
68 Musher DM, Abers MS, Corrales-Medina VF. Acute infection and myocardial infarction. Reply *N Engl J Med*. 2019;380(15):e21.
69 Plotkin V, Voronel VL, Timoshina MA, Zaripova ZA, Murina EA, Khromov-Borisov NN. Enterovirus infection as a risk factor of acute coronary syndrome and its complications. *Klin Med (Mosk)*. 2011;89(2):25–29.
70 Ilback NG, Mohammed A, Fohlman J, Friman G. Cardiovascular lipid accumulation with Coxsackie B virus infection in mice. *Am J Pathol*. 1990;136(1):159–167.
71 Koren O, Spor A, Felin J, et al. Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc Natl Acad Sci USA*. 2011;108(Suppl 1):4592–4598.
72 Kuss SK, Best GT, Etheredge CA, et al. Intestinal microbiota promote enteric virus replication and systemic pathogenesis. *Science*. 2011;334(6053):249–252.

# Appendix 4

**Publication IV**

Sadam, H.*, Mustonen, L.*, **Annika Rähni***, Nieminen, J. K., Toots, M., Uusväli, M., Pihlak, A., Tienari, P. J., Harno, H., Palm, K.**, & Kalso, E.** (2026). **Comprehensive antigen profiling predicts post-surgical neuropathic pain in women treated for breast cancer.** Scientific Reports, 16(1), 12511. https://doi.org/10.1038/s41598-026-41637-6

scientific reports

Check for updates

OPEN

# Comprehensive antigen profiling predicts post-surgical neuropathic pain in women treated for breast cancer

Helle Sadam[1,6], Laura Mustonen[2,3,6]✉, Annika Rähni[1,4,6], Janne K. Nieminen[3,5], Maarja Toots[1], Mariliis Uusväli[1], Arno Pihlak[1], Pentti J. Tienari[3,5], Hanna Harno[2,3], Kaia Palm[1,4] & Eija Kalso[2]

**The importance of neuroimmune interactions in neuropathic pain (NP) has been established, but antibody-mediated mechanisms remain underexplored. In this explorative case-control study, we analyzed antibody profiles in patients with intercostobrachial nerve injury during breast cancer (BC) surgery. We compared 27 patients who developed chronic NP with 30 who remained NP-free, despite similar nerve injury. Plasma samples were collected before surgery and 4–9 years later. Mimotope variation analysis (MVA), a next generation random peptide phage display method revealed highly individual yet shared antigen profiles. We identified 1882 antibody epitopes differing between the study groups and that were associated with 79 common human pathogens. NP patients showed elevated pre-surgical antibody responses to viral epitopes of CMV (cytomegalovirus), EBV (Epstein-Barr virus), human papilloma virus-16 (HPV-16), human rhinovirus C3 (HRV C3), Herpes Simplex-1 (HSV-1), Herpes Simplex-2 (HSV-2), while antibody levels against Coxsackievirus B3 (CVB3) were lower. These findings persisted over time. The combination of responses to five viral epitopes (CVB3, EBV, CMV, HPV-16, HSV-2) predicted persistent NP (AUC 0.9, 95% CI 0.794–0.963). These findings implicate elevated antiviral immune responses in NP pathogenesis and encourage further clinical and basic research on the molecular mechanisms and novel treatment strategies for managing NP.**

**Keywords** Neuropathic pain, Post-surgical pain, Antibody response, Herpesvirus

Neuropathic pain (NP) is defined as pain caused by a lesion or disease of the somatosensory nervous system[1]. Surgical nerve injuries can lead to persistent peripheral NP (chronic post-surgical NP, CPSNP). CPSNP is especially prevalent (14–31%) in patients operated on for breast cancer (BC)[2,3], injury to the intercostobrachial nerve (ICBN) during axillary surgery being an important cause. However, not all ICBN injuries lead to CPSNP. The biological mechanisms explaining why similar nerve injuries become painful in some, but not in all, still need to be elucidated.

Development and maintenance of pain following the onset of peripheral nerve lesion is a complex process involving multiple peripheral and central mechanisms[4,5]. Extensive preclinical and growing clinical evidence shows that neuroimmune interactions play a key role in this process and involve both innate and adaptive immune systems[6–8]. However, antibodies are less studied in the context of NP, although it has been shown that antibodies and their complexes have the potential to induce nociceptor hyperexcitability through multiple mechanisms[9,10]. In the context of nerve injury, pre-existing antibodies derived from previous exposures to environmental pathogens[11], and autoantibodies, could enhance nociception due to trauma-related events[9,10]. Also, tissue injury could lead to exposure of hidden antigens and the generation of autoantibodies with pro-nociceptive properties[12,13]. However, the role of humoral immune responses in painful peripheral nerve injury has largely remained unexplored and clinical studies are scarce.

[1]Protobios Llc, Tallinn, Estonia. [2]Department of Anaesthesiology, Intensive Care and Pain Medicine, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. [3]Neurocenter, Neurology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. [4]Department of Chemistry and Biotechnology, Tallinn University of Technology, Tallinn, Estonia. [5]Translational Immunology Research Program, University of Helsinki, Helsinki, Finland. [6]These authors contributed equally: Helle Sadam, Laura Mustonen and Annika Rähni.These authors contributed equally: Kaia Palm and Eija Kalso. ✉email: laura.mustonen@helsinki.fi

In this explorative study, we used mimotope variation analysis (MVA)[14], a next generation random 12-mer peptide phage display method, for large-scale assessment of the antibody epitope repertoire in CPSNP after BC surgery, which had been complicated by nerve injury. The data generated by MVA immunoprofiling have successfully been used to identify novel antibody epitope biomarkers for vaccination-induced narcolepsy, multiple sclerosis, and COVID-19[14–17]. We compared the targeted peptide epitope profiles (immunoprofiles) in the clinical samples of patients with persistent painful versus painless surgical injuries to ICBN. In this, our aim was to assess whether specific immunodominant antibody epitopes associate with CPSNP after BC treatments. We analyzed samples from two timepoints: before surgery and at follow-up (4–9 years from surgery). In these settings, we addressed two questions: (1) Do patients with painful and painless surgical nerve injury present with a pre-existing distinct antibody epitope spectrum before nerve injury? (2) Are there CPSNP-associated changes in the immunodominant antibody response that emerges after surgical nerve injury and BC treatment? With this approach, our aim was to increase understanding of the immunological features in the pathophysiology of CPSNP and identify novel biomarkers for CPSNP.

## Results

### Patient characteristics

We used the same study group that was previously described by Lötsch et al.[18] and Mustonen et al.[19]. We included 57 patients, of whom 27 had CPSNP, with a matched control group of 30 non-CPSNP patients (Fig. 1 and Table S1). The patient characteristics, including the type of cancer and cancer treatments, have been previously reported elsewhere[19]. At the follow-up visit, patients in the CPSNP group were on average approximately four years younger (mean age 60,3 vs. 64.0 years) than patients in the non-CPSNP group (Fig. 2b).

The groups had no significant differences in surgical and cancer treatments or in cancer type. Mastectomy was performed on 16 (59%) patients in the CPSNP group and 20 (67%) patients in the non-CPSNP group. Most patients (96% in CPSNP group and 83% in non-CPSNP group) had undergone axillary lymph node clearance. Chemotherapy was administered to 23 (85%) patients in the CPSNP group and 25 (83%) patients in non-CPSNP group. Docetaxel was included in the chemotherapy regimen in all but one patient in CPSPN and two patients in non-CPSNP group. Presurgical samples were taken from both groups (Timepoint 1 [T1]); the mean follow-up time was 6.5 years, ranging from 4 to 9 years (Timepoint 2 [T2], Fig. 2a). In the CPSNP group, the median worst pain past week score was 5/10 NRS, with overall pain intensity ranging from 4 to 8/10 NRS. For CPSNP patients, the median score in the Douleur Neuropathique 4 (DN4) screening questionnaire was 6/10 ranging from 3 to 8/10. At the follow-up visit, four patients in the CPSNP group reported on-demand use of paracetamol or NSAID, one patient reported on-demand use of paracetamol-codeine combination drug, and one patient used regular neuropathic pain medication (gabapentinoid and amitriptyline). None of the patients in the non-CPSNP group reported any use of pain medications. None of the patients had received any neuromodulatory pain treatments.

### The most abundant antibody epitopes are uniquely individual and consistently stable

To identify the antibody epitope response repertoire, we performed MVA immunoprofiling of serum samples from CPSNP and non-CPSNP groups collected at two timepoints ($n = 126$; Fig. 2c, Fig. S1). Data analysis when

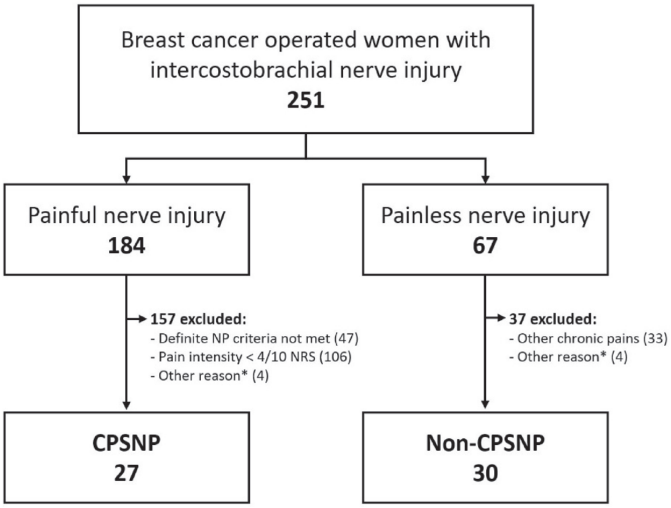



**Fig. 1**. Patient selection. *Other reasons include metastasized cancer and diagnosis of inflammatory or neurological disease. CPSNP: persistent post-surgical neuropathic pain; NP: neuropathic pain; NRS: numerical rating scale.

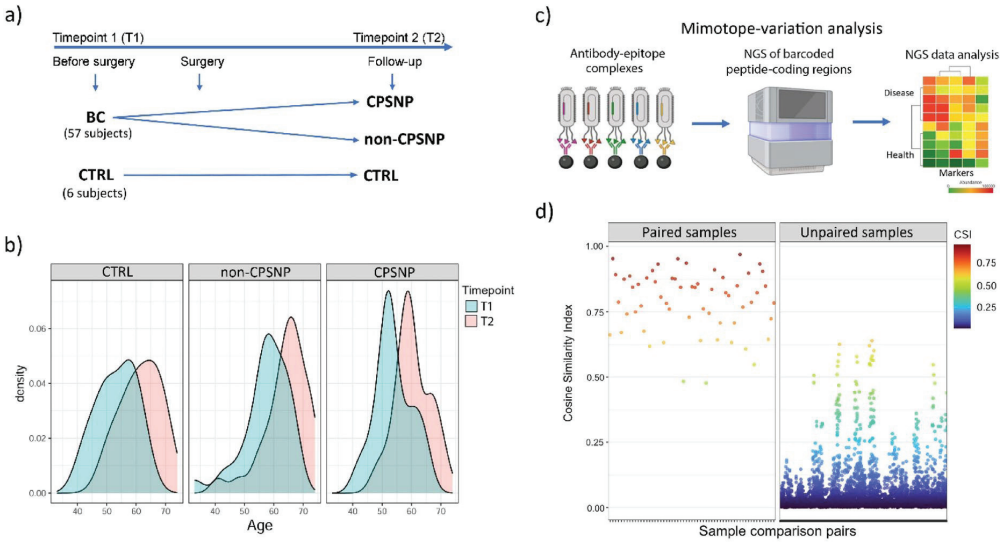
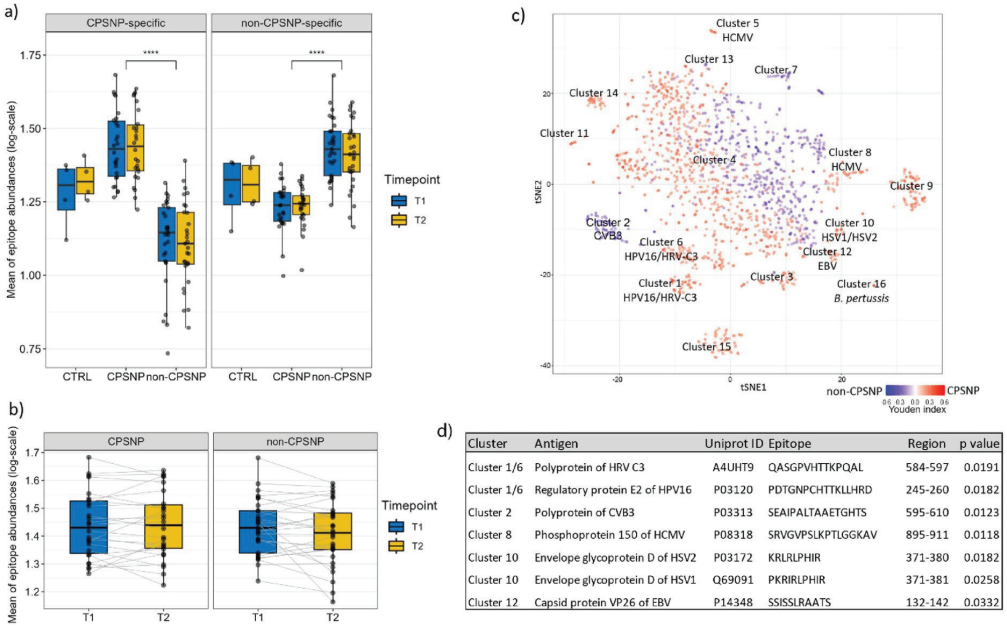



**Fig. 2**. Characterization of antibody epitope profiles from the serum of patients with CPSNP ($n = 27$) and non-CPSNP ($n = 30$) by MVA. (**a**) Clinical study design. Samples from 57 patients with breast cancer (BC) diagnosis were analyzed. Each patient had given two samples: one before (timepoint 1, T1) and another after (timepoint 2, T2) surgery. The period between T1 and T2 varied between 4 and 9 years. The control group (CTRL) included serum samples derived from T1 and T2 timepoints from six patients from the same cohort without ICBN injury and without pain. (**b**) Age distribution of the 63 patients at T1 and T2. (**c**) Antibody epitope profiling with MVA. MVA workflow comprised different sequential steps: (i) immunocapture of IgG-phage complexes from a sample using the phage library (random 12-mer peptide M13 phage library); (ii) high-throughput Illumina HiSeq2500 short DNA sequencing of bar-coded phage DNA; (iii) bioinformatical data analysis resulting in antibody epitope identification. The sequenced DNA reads obtained were quality-checked, translated to peptide sequences, and sample- and cohort-specific epitopes from peptide sets were developed by the SPEXS2 pattern search algorithm. Schematic presentation was created with biorender.com (**d**) Highly individual immunoreactive epitope profiles shared similar features in paired sera (T1 and T2) samples, as observed from MVA. The 5000 most IgG-bound (abundant) peptide values (read counts) from each sample were taken into immunoprofile similarity analysis. For all sample pairs, the normalized scalar products of peptide count vectors were calculated for the cosine similarity index (CSI, R package lsa). The CSI values depicted are 0.70 and above (color-coded scale on the right), showing highly correlated immune profiles. Based on CSI analysis, the data of one CPSNP, one non-CPSNP and two CTRL samples were removed from quantitative analyses (Table S1). Group sizes: Paired samples $n = 59$, Unpaired sample pairs = 6844. Sample pairs were combined from timepoint samples of cohort patients: CPSNP $n = 26$, non-CPSNP $n = 29$ and CTRL $n = 4$ (Table S1). BC: breast cancer; CPSNP: chronic post-surgical neuropathic pain; CSI: cosine similarity index; CTRL: control; MVA: mimotope variance analysis; NGS: next generation sequencing; NP: neuropathic pain; T1: timepoint 1; T2: timepoint 2.

comparing the individual pre-operative and follow-up samples resulted in defined sets of peptides that were highly individual-specific but similar at both timepoints. The cosine similarity index (CSI) was remarkably high on average (0.79) for paired samples ($n = 59$ comparisons), while CSI was significantly lower on average (0.04) in unpaired samples (Fig. 2d, Fig. S2). These results demonstrated that the seroreactivity against major antibody epitopes remained similar in time in the same subject but differed between subjects.

## BC patients with CPSNP have distinct pre-surgery IgG profiles

To define differences in antibody epitope profiles among study groups, we performed an unsupervised clustering analysis of the most abundant (immunodominant) peptides detected by MVA and determined group-specific epitopes (Fig. S1). To further investigate the origin of these immune features, we aligned the epitope sequences ($n = 9344$) to the peer-reviewed epitopes from the IEDB with exact matching. This annotation analysis identified 1882 epitopes showing high sequence similarity with the reported epitopes of 79 common human pathogens (Table S2). Table S2 shows annotated linear epitopes associated with pathogen antigens. Antibody immune response to these epitopes was detected in both pre-surgery and follow-up (Timepoints T1 and T2) samples from the patients with CPSNP (Fig. 3a). Pairwise data analysis confirmed the persistence of immune response to the selected epitopes over time (Spearman rho and p-values between timepoint samples showed high correlation: $\rho = 0.88$, $p = 1.7{*}10^{-6}$ (CPSNP) and $\rho = 0.80$, $p = 1.6{*}10^{-6}$ (non-CPSNP), Fig. 3b). Further clustering analysis resulted in 17 distinct pathogen-associated epitope clusters (Fig. 3c). Annotation analysis indicated that,




| Cluster | Antigen | Uniprot ID | Epitope | Region | p value |
|---|---|---|---|---|---|
| Cluster 1/6 | Polyprotein of HRV C3 | A4UHT9 | QASGPVHTTKPQAL | 584-597 | 0.0191 |
| Cluster 1/6 | Regulatory protein E2 of HPV16 | P03120 | PDTGNPCHTTKLLHRD | 245-260 | 0.0182 |
| Cluster 2 | Polyprotein of CVB3 | P03313 | SEAIPALTAAETGHTS | 595-610 | 0.0123 |
| Cluster 8 | Phosphoprotein 150 of HCMV | P08318 | SRVGVPSLKPTLGGKAV | 895-911 | 0.0118 |
| Cluster 10 | Envelope glycoprotein D of HSV2 | P03172 | KRLRLPHIR | 371-380 | 0.0182 |
| Cluster 10 | Envelope glycoprotein D of HSV1 | Q69091 | PKRIRLPHIR | 371-381 | 0.0258 |
| Cluster 12 | Capsid protein VP26 of EBV | P14348 | SSISSLRAATS | 132-142 | 0.0332 |

**Fig. 3**. Distinct antibody immune response for CPSNP relates to antigens of common pathogens. (**a**–**c**) Annotation analysis of group discriminative epitope mimics resulted in epitopes ($n = 1882$) showing sequence similarity to common pathogens ($n = 79$) from the Immune Epitope Database (IEDB) (Table S1). 4188 antigenic determinants of common pathogens were selected. Summarized log-transformed values of seroreactivity to these IgG bound epitopes in each sample of CPSNP ($n = 26$), non-CPSNP ($n = 29$) or CTRL ($n = 4$) groups are depicted in a boxplot. Wilcoxon Rank Sum test, p-values adjusted with FDR: **** $p \leq 0.0001$. (**b**) Pairwise analysis of pre-operative (T1) and follow-up (T2) samples of CPSNP ($n = 26$) and non-CPSNP ($n = 29$) patients. Summarized log-transformed values of seroreactivity to IgG bound epitopes in each sample are depicted in a boxplot. Spearman rho and p-values between timepoint samples showed high correlation: $\rho = 0.88$, $p = 1.7{*}10^{-6}$ (CPSNP) and $\rho = 0.80$, $p = 1.6{*}10^{-6}$ (non-CPSNP). (**c**) t-SNE analysis of 1882 antibody epitopes identified 17 different clusters (Table S2). For t-SNE and clustering analysis, the abundance of immune response values to 1882 epitopes in 63 cohort pre-operative samples (27 CPSNP, 30 non-CPSNP and 6 CTRL) were normalized to 3 million reads and log-transformed. Epitope clusters were detected using the R package Rtsne (dims 2, perplexity 15, max_iter 500) and R package dbscan (epsilon 2, minPts 10) (Table S2). (**d**) Seven epitopes from different clusters stood out as the most interesting group-discriminating epitopes that aligned to 79 different human pathogen epitopes presented in IEDB (Table S1). The sequences of these epitopes belonging to common human viruses are depicted with amino acid regions. A permutation-based test was performed to evaluate statistical significance of each aligned protein–motif pair (Table S2). For each pair, the protein sequence was held constant while the amino acid positions of the matched motif were randomly shuffled 10,000 times to generate permuted motif variants preserving amino acid composition. Each permuted motif was aligned to the target protein sequence, and the number of exact matches was recorded. The empirical p-value was calculated as: (number of permuted matches + 1)/(number of permutations + 1) and p-values were adjusted with FDR. The highest adjusted p-value is shown. AUC: area under the ROC curve; B. pertussis: *Bordetella pertussis*; CPSNP: chronic post-surgical neuropathic pain; CTRL: control; CVB3: Coxsackievirus B3; EBV: Epstein-Barr virus; CMV: human cytomegalovirus; HRV-C3: human rhinovirus C3; HSV1/2: Herpes simplex virus 1/2; T1: timepoint 1; T2: timepoint 2.

while most clusters were heterogenous and contained epitopes of different viral antigens (Cluster 4 and 15, for example, Table S2), some clusters were very clearly linked to a specific epitope of a specific pathogen (Fig. 3c). For example, Clusters 5 and 8 were strongly associated with the human cytomegalovirus (CMV) antigen of phosphoprotein 150 (pp150). Clusters 10 and 12 contained epitopes associated with the glycoprotein D (gpD) of human herpes simplex viruses (HSV1 and HSV2). Cluster 2 was more distinct for non-CPSNP than for CPSNP and was enriched in epitopes associated with enteroviruses (Fig. 3c and Table S2). We then quantified the antibody response against these major epitope clusters (Fig. 3c) that were specifically associated with CPSNP. Amongst these, antibody responses against the epitopes of seven viruses emerged as the most discriminating (Tables S2- S3), specifically against antigenic determinants of the polyprotein of human rhinovirus (HRV C3); regulatory protein E2 of human papilloma virus 16 (HPV16); polyprotein of coxsackievirus B3 (CVB3); pp150 of CMV; glycoprotein D of HSV1 and HSV2; capsid protein VP26 of EBV (Fig. 3d). We further hypothesized

that some of the identified epitopes related to these common pathogens could be linked to autoimmune mimicry (Table S4). We observed accumulations of immune response targeting extracellular domains of various human protein or secretory protein antigens, as well as those that could directly influence pain signaling. Among others, for example, the epitope of HRV-C3/HPV16 (Cluster 11/Cluster 6) showed mimicry of the epitope of the human collagen alpha-3(VI) chain, a beaded filament collagen shown to have a role in cancer and acute wound healing (UNIPROT ID P12111) (Table S4). These results indicated that a high pathogen load was a dominating feature differentiating patients with or without CPSNP.

### Viral seropositivity by ELISA confirm the MVA data

We analyzed overall seropositivity to EBV, CMV, HSV1 and HSV2 and observed that those for EBV and CMV were prevalent across all study groups, whereas seropositivity for HSV1, and particularly HSV2, were less common (Fig. 4a, Table S1). The percentage of HSV2 seropositive individuals was higher in the CPSNP group than in the non-CPSNP or CTRL groups (HSV2: 48.2%/20.0%/33.3%, Fisher's Exact test p value 0.027) (Fig. 4a). Next, because the annotation analysis of the antibody profiles to IEDB-reviewed epitopes revealed clear links with common human pathogens, we used the overall seropositivity results to validate the findings of MVA. We assessed the potential of seroreactivity to herpesviral antigens (pp150 of CMV, gpD of HSV1, gpD of HSV2) for accurate classification (Fig. 4b). The antibody response to the examined epitopes was significantly higher in the seropositive samples (Wilcoxon Rank Sum test, FDR-adjusted $p < 0.01$, Fig. 4b). Additionally, the response to gpD antigen of HSV2 was greater in CPSNP samples than in non-CPSNP samples (Wilcoxon Rank Sum $p < 0.05$, Fig. S3). Finally, an epitope-specific ELISA using peptides corresponding to epitopes of pp150 CMV ($_{892}$GRGSRVGVPSLKPTLGGKAV$_{911}$) and gpD HSV2 ($_{368}$MAPKRLRLPHIRDDD$_{383}$) independently confirmed that the MVA findings were genuine targets of the IgG response associated with these pathogens and potentially linked to neuroinflammation (Fig. 4c, Fig. S4).

### Predictive value of immune load to pathogens for CPSNP in breast cancer patients

The marked difference in seroreactivity to herpesviruses between the CPSNP and non-CPSNP groups led us to explore the potential of using seroreactivity to antigenic epitopes of the seven common viruses, both against individual viruses and collectively for classification. Individually, antibody responses to single epitopes of CMV, EBV, HRV, HPV16, HSV1 and HSV2 were notably higher in pre-surgery samples of CPSNP than non-CPSNP samples (Fig. 5a). Anti-CVB3 was an exception, as the responses against its epitope were lower in CPSNP than in non-CPSNP samples. In post-operative samples collected 4–9 years after surgery, the seroresponse to most epitopes remained similarly associated with CPSNP (Fig. S5).

We then analyzed all seven features to develop a logistic regression model aimed at predicting CPSNP in BC patients. These data showed that the seroresponse to the defined set of viral epitopes of CVB3, EBV, CMV, HPV-16, and HSV-2, presented a novel predictive biomarker that distinguished CPSNP patients from non-CPSNP patients pre-surgery, achieving a balanced accuracy of 84.6% (Fig. 5b). The observed cross-validated AUC was 0.85 and was significantly greater than expected by chance based on a permutation test ($p = 0.00089$). Baseline clinical characteristics were comparable between the CPSNP and non-CPSNP groups[19]. The variables demonstrating minor difference between the two groups were age and BMI before (T1) and after surgery (T2). The reported biomarker associations remained stable after age and BMI adjustments (Fig. S6).

This model was validated on an independent control cohort[16,17] comprising age-, sex-, and country-matched individuals ($n = 28$, achieving AUC = 0.806, with a sensitivity of 0.73 and a specificity of 0.75, Fig. S7). Altogether, our findings suggested that immunity load to pathogens may play a significant role in modulating pain signaling and could provide new avenues for therapeutic interventions.

## Discussion

In this study, we present a hypothesis-free analysis of antibody epitopes in patients with painful (CPSNP) and painless (non-CPSNP) nerve injury after BC surgery, using MVA immunoprofiling. MVA uncovered highly individualized immunoprofiles in the paired samples across study groups (pre-operative and follow-up), while also identifying shared immunoreactive features. We identified 1882 epitopes that were differently targeted by antibody response in the two study groups, and these epitopes showed high sequence similarity to antigens of common human pathogens. These included epitope mimics of the polyprotein of HRV-C3, regulatory protein E2 of HPV16, phosphoprotein 150 of CMV, capsid protein VP26 of EBV and envelope glycoprotein D of HSV1 and HSV2, whereas epitopes of polyprotein of CVB3 were associated with painless nerve injury. Similar immune responses to the epitopes of these common human viruses were detected in the samples collected before surgery and after the follow-up showing that the findings persisted over time. Five of these specific epitopes (epitopes of CMV, EBV, HSV2, CVB3, and HPV-16) could be potentially clinically useful as preoperative predictors of CPSNP in patients undergoing BC surgery (AUC of 0.9, 95% CI, 0.794–0.963). This suggests that immune load linked to common viruses is associated with an increased risk of developing CPSNP.

The antibody repertoire reflects both the past and current activity of the adaptive immune system. Immune response against previously encountered environmental pathogens has been identified as a potential risk factor for various autoimmune disorders[20], as well as in other contexts, including psychiatric disorders[21]. Recently, serological evidence of multiple infections was associated with multisite chronic pain[22]. To the best of our knowledge, this is the first study to report an association between antibody responses against common and latent viruses and the risk of developing persistent pain after surgical nerve injury.

Neuropathic pain (NP) is caused by disease or a lesion in the somatosensory system. However, not all nerve lesions become chronically painful. Immune system has a major contribution in the resolution on pain after nerve injury but also in the maladaptive mechanisms that lead to persistent pain[23,24]. Moreover, cellular immune response to nerve injury is regarded as a possible therapeutic target[25]. The crosstalk between immune system and

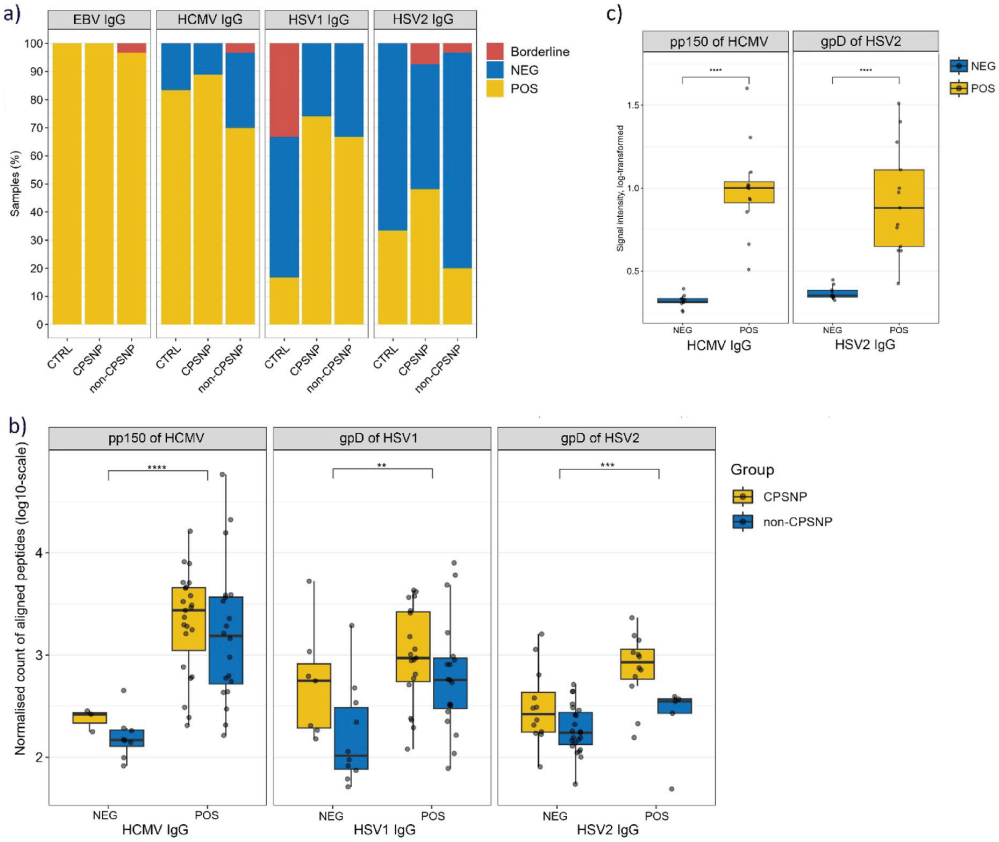



**Fig. 4**. Serology findings of high seropositivity confirm the chronic pathogen burden in CPSNP and non-CPSNP groups. (**a**) IgG response measured by ELISA for common viruses and/or their specific antigens. Serology was measured for EBV (Anti-EBV-CA IgG ELISA, 2791–9601 G), CMV (Anti-CMV, EI 2570–9601 G), HSV1 (Anti-HSV1-gC1, EI2531-9601-2G), HSV2 (Anti-HSV2-gG2, EI2532-9601-2G) for 59 clinical samples ((CPSNP ($n=26$), non-CPSNP ($n=29$), CTRL ($n=4$)). The percentage of HSV seropositivity was higher in the CPSNP group than in non-CPSNP or CTRL groups (Fisher's Exact Test $p \leq 0.05$) whereas statistically significant only for HSV2 (Fisher's Exact Test $p=0.027$). CPSNP group – 24/3/0 CMV, 27/0/0 EBV, 20/7/0 HSV1, 13/12/2; non-CPSNP – 21/8/1 CMV, 29/0/1 EBV, 20/10/0 HSV1, 6/23/1 HSV2; CTRL – 3/1/0 CMV, 5/1/0 EBV, 1/3/2 HSV1, 2/4/0 HSV2, (numbers referring to POS/NEG/B). (**b**) MVA-detected immunodominant epitopes classify seroresponse to herpesviral antigens pp150 of CMV, gpD of HSV1 and gpD of HSV2. Summarized seroreactivity values of IgG-bound peptides that contained epitopes with sequence similarity to the three epitopes of common herpesviruses are shown as cohort samples are divided by serology status (NEG, POS) and groups (CPSNP, non-CPSNP); Wilcoxon Rank Sum test, p-values adjusted with FDR, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$. (**c**) Epitopes of pp150 of CMV and gpD of HSV1/2 were validated using epitope-specific ELISA with CMV and HSV2 POS/NEG serum samples and peptides displaying epitope of pp150 of CMV (left: $n=23$) and epitope of gpD of HSV2 (right: $n=23$). Signal intensities were calculated as Signal/Background ratio (RLU). Wilcoxon Rank Sum test, FDR-adjusted p-values: **** $p \leq 0.0001$. CPSNP: chronic post-surgical neuropathic pain; CTRL: control; EBV: Epstein Barr virus (human gammaherpesvirus 4); CMV: human cytomegalovirus (human betaherpesvirus 5); HSV1/2: Herpes simplex virus 1/2 (human herpesvirus 1/2); NEG: seronegative; POS: seropositive; B: borderline RLU: relative light unit.

NP includes interactions with both innate and adaptive immunity. This process involves almost every immune system component and cell type, including macrophages, neutrophils, T cells, glial cells, cytokines, chemokines and inflammatory mediators[7,13,23–25]. However, the role of humoral immune mechanisms has been less studied in the context of painful nerve injury, although growing evidence implicates its potential contribution in certain persistent pain conditions[9,10,26,27].

Antibodies may exert pronociceptive effects through immune-complex formation, complement activation, and Fc-receptor–mediated sensitization of peripheral nociceptors, providing a potential mechanistic link

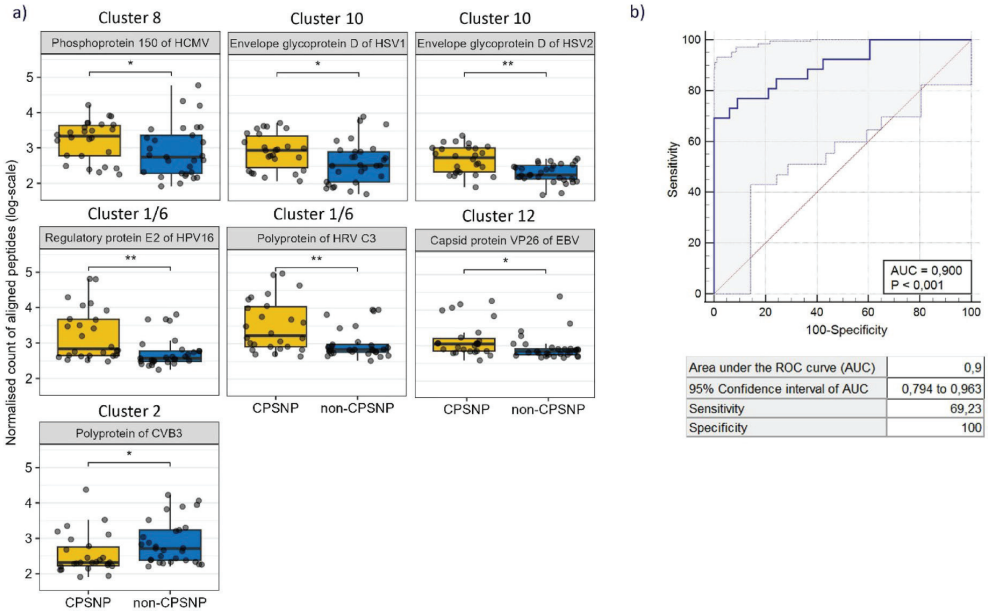



**Fig. 5**. Antibody response to seven highly antigenic epitopes of common viruses as candidate markers for predicting CPSNP pre-operatively. (**a**) Seven epitopes, including those of pp150 of CMV, gpD of HSV1 and HSV2, polyprotein of HRV C3, regulatory protein E2 of HPV16, capsid protein VP26 of EBV, and polyprotein of CVB3, were deemed most discriminating (Tables S2 and S3). Summarized values of IgG-bound peptides with sequence similarity to the seven epitopes from Fig. 3d are depicted in a boxplot for each pre-surgery sample of the CPSNP ($n = 26$) or non-CPSNP ($n = 29$) groups. Wilcoxon Rank Sum test, p-values FDR adjusted: * $p \leq 0.05$; ** $p \leq 0.01$. (**b**) Logistic regression model with values of antibody responses to five of these epitopes (pp150 of CMV, gpD of HSV2, E2 of HPV16, VP26 of EBV, polyprotein of CVB3) in pre-operative samples predicts CPSNP with high specificity (1) and sensitivity (0.69) (AUC 0.9). The observed cross-validated AUC was 0.85 and was significantly greater than expected by chance based on a permutation test ($p = 0.00089$). Groups compared: CPSNP ($n = 26$); non-CPSNP + CTRL ($n = 29 + 4$). AUC: area under the ROC curve; CPSNP: chronic post-surgical neuropathic pain; CTRL: control; CVB3: Coxsackievirus B3; EBV: Epstein-Barr virus (human gammaherpesvirus 4); CMV: human cytomegalovirus (human betaherpesvirus 5); HPV16: human papillomavirus 16; HRV-C3: human rhinovirus C3; HSV1/2: Herpes simplex virus 1/2 (human herpesvirus 1/2).

between adaptive immunity and pain processing[9,10]. For example, IgG from complex regional pain syndrome (CRPS) and fibromyalgia patients can transfer pain-like hypersensitivity to rodents[26,27]. Furthermore, in a subset of fibromyalgia patients, autoantibodies bind to satellite glia cells in the dorsal root ganglia (DGR) of the sensory nerves[28]. In the context of nerve injury, a mouse model showed accumulation of IgG in DRG and dorsal spinal cord with potential pro-nociceptive effect[13]. A clinical study of spinal cord injury patients showed an antibody-mediated autoimmune response that was associated with a higher need of NP medication[12]. However, there is a lack of clinical studies concerning the role of antibodies in painful peripheral nerve injury.

Our data suggest an elevated immune response to multiple human herpesviruses as a risk factor for developing persistent NP after breast cancer surgery. Human herpesvirus infections or their post-infection syndromes have been associated with persistent NP in different contexts, with post-herpetic neuralgia the most well-known. Apart from these, the association of latent herpesvirus infections and other NP conditions is less studied. A rare unilateral pain syndrome with NP features has been reported in association with recurrent HSV infections[29]. Indeed, HSV1 and −2 remain latent in the ganglia of sensory nerves with the potential of clinical or subclinical reactivation later in life. Although these viruses are known to reside primarily in trigeminal (HSV1) and sacral (HSV2) ganglia, HSV1 can sometimes reside in thoracic ganglia[30] suggesting a plausible anatomical link to the CPSNP-related findings in the present study.

EBV infects mainly B cells and epithelial cells, CMV infects monocytes, lymphocytes and epithelial cells. After primary infection the reactivations of CMV and EBV are usually asymptomatic. Previously, seropositivity to EBV has been associated with chronic multisite musculoskeletal pain and inflammatory joint pain[22,31]. Other chronic pains are common comorbidities in NP, and they associate with risk of developing CPSNP after surgery[32,33]. High antibody titers to both EBV and CMV have been associated with elevated levels of other inflammatory mediators such as CRP and IL-6 indicating that the activity of multiple herpesviruses can drive

general inflammation[34]. Furthermore, general inflammation and higher levels of pro-inflammatory cytokines have been associated with NP in different contexts in both pre-clinical and clinical studies[6,7,23]. This is in line with our data showing elevated antibody response to multiple herpesviruses in CPSNP. High antibody levels to latent herpesviruses are indirect indicators of viral reactivations and they also serve as an indicator of immune dysregulation, due to their associations with elevated levels of inflammation[34,35]. The compromised immune function which predisposes individuals to recurrent and prolonged herpesviral reactivation specifically could be the underlying factor for the elevated risk of CPSNP following nerve injury[36]. However, our study does not provide direct information about immune function, inflammation, or herpesvirus reactivations and further studies are needed to confirm this hypothesis.

Notably, compared with the CPSNP group, the non-CPSNP group patients exhibited an elevated IgG response targeting epitopes that mimic the polyprotein of CVB3, a common human enterovirus causing flu-like symptoms and sometimes myocarditis. It remains unclear how antibody response to CBV3 could associate with favorable outcome in terms of pain development after surgical nerve injury. The clinical relevance of this finding needs further evidence. Interestingly, in a recent MVA study, robust antibody response to a conserved enteroviral epitope was associated with better cardiovascular health[17].

Our modelling results suggest that antibody response to specific antigen epitopes of CMV, EBV, HSV2, CVB3, and HPV-16 prior to surgery could be useful predictors of CPSNP in patients undergoing BC surgery. There are several mechanisms that could be hypothesized. Firstly, antibodies could be nociceptive per se, like in CRPS and fibromyalgia. Secondly, antibodies may jointly induce general inflammation[34], which interferes with normal nerve regeneration and leads to pain sensitization[24,36]. Thirdly, antibodies may be markers of compromised immune functions, which would be the biological cause of lack of regeneration and pain. Lastly, reverse causality is also worthy of consideration, because chronic NP is associated with increased stress and psychological burden[32,37] and stress is known to induce reactivations of herpesviruses. All in all, our findings support the hypothesis that prior infections and viral re-activations potentially prime neuroimmune pathways that facilitate the development and persistence of NP following surgical injury. Since our study does not provide direct information about the underlying mechanisms, further research is necessary to clarify the role of antibody responses in NP development.

Strengths of this study include a prospective setting of thoroughly characterized study groups of painful and painless nerve injury patients after BC surgery. Our data is unique, because a similar setting is very difficult to obtain in most other NP etiologies where the onset of nerve lesion and NP are usually unpredictable. Another strength of this study is its comprehensive immune response analysis. The interaction between the immune system, inflammation, and pain is intricate. MVA allowed us to screen a vast set of peptides, accelerating the discovery process of antigens associated with chronic pain at the epitope level. However, our study does not provide direct information on active infection, pathogen load, or viral reactivation. The patient cohort is small, but the patient groups were carefully characterized with a detailed assessment of NP along with preoperative and follow-up samples with timespan up to nine years. These together add considerable depth to our study. Our results may not be directly generalizable to men and to other NP etiologies, because our study groups comprised of BC treated women with surgical nerve injury.

Our analysis demonstrated that the antibody response profiles in BC patients experiencing CPSNP showed reactivity to multiple viral antigens. These immune response patterns may be candidates as predictive markers assessing the risk of pain development during the preoperative stage. This finding could pave ways to new therapeutic opportunities, such as use of antiviral drugs, for the prevention and management of neuropathic pain. Since our results from clinical patients are preliminary, we encourage further studies to elucidate the multidisciplinary mechanisms and potential treatment avenues in persistent NP.

## Patients and methods

### Patient selection

We have previously described the diagnosis of NP in 251 women whose treatment for BC involved axillary surgery and ICBN resection (as reported by the surgeons). We examined the patients for the presence or absence of CPSNP of the ICBN 4–9 years after surgery[32]. All patients were recruited from a previous prospective cohort of 1000 women operated on for BC at the Helsinki University Hospital during 2006–2010[38]. Thus, we were able to use data and samples collected preoperatively from the patients in the current study. During surgery, the operating surgeon reported the handling of ICBN (totally resected, partially resected, spared or not identified). Patients with total or partial ICBN resection were invited for the follow-up research visits for the current study. Follow-up research visits took place during 2014–2016, 4–9 years after surgery[32]. The visits included thorough clinical sensory examination by a neurologist (HH) for diagnosis of NP according to stepwise grading criteria (unlikely, possible, probable, definite NP)[39], and collection of blood samples. The contents of sensory examination and NP grading have previously been reported in detail elsewhere[32]. To fulfill the criteria for definite NP, the pain and sensory alterations had to be located within the area of ICBN innervation and the ICBN resection had to be confirmed by the operating surgeon. Because of the surgeon's report of nerve lesion, no additional diagnostic tests are needed to fulfill the criteria of definite NP[39]. We used the "worst pain past week" in the Brief Pain Inventory (BPI) long form to measure pain intensity on a Numerical Rating Scale (NRS, 0–10). NRS ≥ 4/10 was considered to reflect at least moderate pain intensity and clinically meaningful pain[40]. All patients were asked for the presence of pain in other body locations (back, neck, joints, head, other) and use of current pain medications. In addition, Douleur Neuropathique 4 (DN4) questionnaire[41] was completed for each patient with pain.

For the current study assessing the role of antibody response profiles in CPSNP, we selected subgroups using the following criteria: (1) the CPSNP group included patients fulfilling the criteria of definite NP[39] and pain intensity ≥ 4/10 on the NRS and (2) the non-CPSNP group included patients with ICBN resection with no CPSNP and no other pains. Patients with ongoing cancer treatment, neurological or autoimmune diseases

were excluded from the present analysis. Patient selection is presented in Fig. 1. For technical control purposes, preoperative and follow-up samples from six BC-operated patients without ICBN injury and without pain were included in the MVA studies. These patients were from the same original cohort as the analyzed groups[32].

The Coordinating Ethics Board of the Helsinki and Uusimaa Hospital District approved the study (149/13/03/00/14) which was also registered in ClinicalTrials.gov (NCT02487524). This study was performed in accordance with the Declaration of Helsinki. All patients gave written informed consent.

## Breast cancer treatments

The cancer treatment data were accessible from the previous study[38]. All patients had been operated on for unilateral BC, and none had received neoadjuvant treatment. Breast surgery was either mastectomy or breast conserving surgery (BCS) accompanied with axillary surgery, which was either sentinel lymph node biopsy (SLNB) or axillary lymph node dissection (ALND). The resection of the ICBN (either total or partial) was verified by the operating surgeon. Oncological treatments, including radiotherapy, chemotherapy, and endocrine treatment, were administered according to national treatment protocols. The chemotherapy regimen comprised a combination of docetaxel and CEF (cyclophosphamide, epirubicine, and 5-fluorouracil). Trastuzumab was administered to the patients with HER2 (human epidermal growth factor 2) -positive BC. Tamoxifen (premenopausal women) or aromatase inhibitor (postmenopausal women) comprised the endocrine treatment.

## Sample collection

Plasma samples for the MVA were collected at two separate timepoints: at induction of anesthesia for the BC surgery (Timepoint 1 [T1]) and at the research visit 4–9 years after surgery (Timepoint 2 [T2]). Plasma from blood samples collected into ethylenediaminetetra-acetic acid (EDTA) tubes was extracted by centrifugation at 3000 revolutions per minute (RPM) for 10 min, and transferred to cryogenic vials, which were immediately frozen and stored at −80 °C. Both the preinjury and follow-up samples were drawn and prepared by the same research nurse. The follow-up research visit protocol included the analysis of basic laboratory parameters and inflammatory markers, including high-sensitivity C-reactive protein (hs-CRP). The collection of study samples has been reported previously[32].

## Mimotope variation analysis

Sample ($n = 126$) processing by MVA was performed according to standard MVA protocols[14]. In brief, plasma samples ($n = 126$, 2 µl of plasma per analysis) were incubated with *Escherichia coli* M13 phage library displaying random 12-mer peptides (complexity $1 \times 10^9$ peptides; NEB, E8111L). IgG–phage complexes were isolated with protein G-coated magnetic beads (NEB, S1430) and amplified by Phusion High Fidelity PCR (Thermo Scientific) for next-generation sequencing. An average of 3 million peptide-encoding DNA sequences per sample was achieved: 5000 most abundantly detected peptides by abundance values (read counts) from each sample were taken into immunoprofile-similarity analysis. For comparisons, the normalized scalar products of peptide count vectors were calculated for the cosine similarity index (R package lsa) that ultimately resulted in a $126 \times 126$ sample immunoprofile-similarity matrix. Due to not meeting technical criteria, four samples (two controls, one CPSNP and one non-CPSNP), were excluded from downstream group comparison and regression analyses (Table S1).

## Delineating cohort specific epitope responses

*Cohort-specific epitope selection*

To delineate cohort-specific antibody response, the data comparison of the most immunodominant epitopes between CPSNP and non-CPSNP groups was carried out (see schematic overview of the data analysis workflow, Fig. S1). The most immunodominant epitopes were determined by clustering analysis (defining 5–11 positions-long epitopes that all contained five fixed amino acids) and hypergeometric test analysis where SPEXS2 Software was used[14]. Samples of these two groups at either timepoint (T1, before surgery; T2, at the 4–9 years follow-up) were compared with each other using optimal cut-point analysis (R package cutpointr, maximizing the Youden metric) and Student's *t*-test (R Stats Package, log-transformed). Based on the group comparisons, 9344 epitopes were chosen for further analysis (*t*-test $p < 0.05$ and Youden $> 0.25$).

*Alignment analysis on infectious epitopes of IEDB and marker epitope selection*

To look for potential mimicry with human and pathogen antigens, group-specific ($n = 9344$) epitopes were exactly aligned with sequences of T and B cell peer-reviewed epitopes in a public database (Immune Epitope Database [IEDB], version: epitope_full_v3.tsv, 2,225,965 epitopes of which 2,214,634 were linear, date accessed: 15.04.2024) (see schematic overview of data analysis workflow, Fig. S1). Based on the relevance (determined by assessed probability of pathogen exposure) and alignment scores, the most relevant microorganisms were selected ($n = 79$, Table S2). To evaluate the statistical significance of specific epitope–motif alignments, we performed permutation-based tests for each aligned epitope–motif pair ($n = 6254$). For each pair, the epitope sequence was held constant while the amino acid positions of the matched motif were randomly shuffled 10,000 times to generate permuted motif variants preserving amino acid composition. Each permuted motif was aligned to the original epitope sequence, and the number of exact matches was recorded. The empirical p-value was calculated as: (number of permuted matches + 1)/(number of permutations + 1) and p-values were adjusted with FDR. 1882 epitopes that showed sequence similarity with the revised epitopes from IEDB, were visualized and clustered using R packages Rtsne (where t-SNE is stochastic neighbor embedding)[42–44] and dbscan[45]. For clustering analysis, the relative values of immune response to the 1882 epitopes in T1 samples ($n = 63$) were log-transformed. Epitope clusters were identified using the Rtsne function from R's Rtsne package (dims = 2, perplexity = 15, max_iter = 500) and the dbscan function from the dbscan package (epsilon = 2, minPts = 10).

Altogether, 17 different clusters were identified (Table S2). Further, the most relevant epitopes and antigens identified by the study cohort were described by alignment scores (Table S3) where the seven most interesting leads (with the highest alignment scores) were selected (see schematic overview of data analysis workflow, Fig. S1). The total immune response value for each epitope was calculated by summarizing the counts of all unique peptides containing the aligned motifs and log-transformed per sample (Wilcoxon Rank Sum test, p-values unadjusted). Following regression and receiver operating characteristic (ROC) analyses, corresponding visualizations were created using MedCalc Statistical Software (version 17.0.4, MedCalc Software Bvba, Belgium) (see schematic overview of data analysis workflow, Fig. S1). For independent validation (regression and ROC analyses), previously published immunoprofiles were used for country- (Finland), age- (52.9 ± 6.21 years) and sex- (female) matched clinical samples ($n = 26$) with no diagnoses of BC and NP pain[16].

*Alignment analysis on autoantigenic epitopes of IEDB*

Autoimmune-associated epitopes were filtered from the original IEDB database (version: epitope_full_v3.tsv, export on 15.04.2024) alignment with 1882 queried epitopes (see schematic overview of data analysis workflow, Fig. S1). All autoimmune-associated peer-reviewed B cell epitopes containing 82 infection-related epitope mimics were included in the study (Table S4). For the antigens identified, cellular locations were gained from the UNIPROT database (column intracellular/extracellular, column features – cytoplasm, endoplasmic reticulum, cell membrane, secreted, cell junction, cell projection, chromosome, Golgi apparatus, lysosome, mitochondrion, nucleus, recycling endosome, NA – data not available) (Table S4).

## Epitope validation

For epitope validation, in-house epitope-specific enzyme-linked immunosorbent assay (ELISA) was used to analyze randomly selected samples with proportional seronegative and -positive sample representation (human cytomegalovirus [CMV] pp150, $n = 23$; herpes simples virus 2 [HSV2] gpD, $n = 23$). Peptides containing target epitopes identified by MVA (for CMV pp150 $_{901}$SLKPTLGGK$_{910}$, for HSV2 gpD $_{370}$PKRLRLPHIRDD$_{383}$) and control epitope ($_{1}$AASAASAA$_{8}$) were synthesized with C-terminal addition (GGGDYKDDDKK(biotin), CASLO ApS). Nunc Immobilizer Streptavidin-coated 96-well plates (Thermo Scientific, 436015) were coated with 0.03 µg/ml (3 ng/well) peptides in 1xPBS for 1 h at room temperature and washed with 0.1% PBS-Tween. After blocking unspecific binding with 5% BSA for 1 h and washing with 0.1% PBS-Tween, diluted plasma samples (plasma dilution 1:1000) in 1% BSA 0.1% PBS-Tween were allowed to bind to peptides overnight at + 4 °C. Next day, the plate was washed with 0.1% PBS-Tween and incubated with secondary antibodies (polyclonal anti-human-HRP (1:1000; goat anti-human IgG (H + L) (Invitrogen, 31410) for 1 h at room temperature. Finally, the plate was washed with 1x phosphate-buffered saline (PBS), and 20x dilution of chemiluminescent substrate (SuperSignal™ ELISA Femto Substrate (Thermo Scientific, 37075)) was added. Results were measured within 5 min using a multimodal microplate reader (Hidex Oy). The signal intensities were calculated as Signal/Background ratio (relative light unit [RLU]) (Wilcoxon Rank Sum test p-values, Pearson correlation coefficients).

## EBV, CMV, HSV1 and HSV2 seropositivity

Commercial ELISA tests were used to determine the serology of 59 samples (CTRL [control, $n = 4$], CPSNP [$n = 26$], non-CPSNP [$n = 29$]) to herpesviruses Ebstein-Barr virus (EBV; Anti-EBV-CA IgG ELISA, 2791–9601 G), CMV (Anti-CMV, EI 2570–9601 G), HSV1 (Anti-HSV1-gG1, EI2531-9601-2G), HSV2 (Anti-HSV2-gG2, EI2532-9601-2G) by analyzing samples from T1 (Table S1). Analyses were carried out in accordance with the manufacturers' specifications. Absorbency was measured at 450 nm multimodal microplate reader (Hidex Oy).

## Statistics

Statistical analyses were conducted with R statistical programming language (v.4.2.3) and RStudio environment (v. 2024.09.1 Build 394). Data were analyzed, graphs were produced and visualized using R packages tidyverse[46], ggpubr[47], ggsci[48] and scales[49]. Cosine similarity indices (CSIs) for sample comparisons based on the top 5000 peptide abundance values and composition were calculated with the cosine function in R package lsa[50]. Boxplots were generated using the style of Tukey with R packages ggpubr and ggplot2[51]. In figures, the upper, middle and lower boxplot lines represent the 75th, 50th, and 25th percentiles, while whiskers represent the largest or smallest value within 1.5x interquartile range above the 75th percentile or below the 25th percentile, respectively. The p-values of Wilcoxon Rank Sum tests were visualized with R ggpubr or ggplot2 packages, and statistical significance is shown where applicable. Student's *t*-tests (base R Stats Package), along with each epitope's Youden index calculation analysis (R package cutpointr[52], maximizing the Youden metric), were used to identify 9344 unique group-differentiating features. MedCalc® Statistical Software (v.20.121.2, www.medcalc.org; 2022) was used to conduct logistic regression and ROC analyses of five pathogen epitopes were used as combinational tests. The proportion of seropositive and -negative samples was evaluated with Fisher's Exact Test from the R Stats Package[53]. Spearman's ρ statistic for epitope-specific ELISA and MVA data correlation were calculated with cor.stat from the R Stats Package. Age adjusted logistic regression, ROC/AUC and cross-validation calculations were performed in R with glm() function from "stats", "pROC"[54] and "caret"[55] packages. Model performance was assessed using 5-fold cross-validation. Statistical significance was evaluated using a permutation test in which outcome labels were randomly shuffled 10,000 times, generating a null distribution of cross-validated AUC values. The permutation p-value was calculated as the proportion of permuted AUCs greater than or equal to the observed AUC obtained from the unpermuted data.

## Data availability

All data are available upon reasonable request to the corresponding author (clinical data) and Kaia Palm (MVA-data, [kaia@protobios.com](mailto: kaia@protobios.com)).

Received: 13 November 2025; Accepted: 23 February 2026
Published online: 03 March 2026

## References


1. Treede, R. D. et al. Neuropathic pain: Redefinition and a grading system for clinical and research purposes. *Neurology* **70**, 1630–1635 (2008).
2. Haroutiunian, S., Nikolajsen, L., Finnerup, N. B. & Jensen, T. S. The neuropathic component in persistent postsurgical pain: A systematic literature review. *Pain* **154**, 95–102 (2013).
3. Ilhan, E., Chee, E., Hush, J. & Moloney, N. The prevalence of neuropathic pain is high after treatment for breast cancer: A systematic review. *Pain* **158**, 2082–2091 (2017).
4. Colloca, L. et al. Neuropathic pain. *Nat. Rev. Dis. Primers*. **3**, 17002 (2017).
5. Meacham, K., Shepherd, A., Mohapatra, D. P. & Haroutounian, S. Neuropathic Pain: Central vs. Peripheral Mechanisms. *Curr. Pain Headache Rep.* **21**, 28 (2017).
6. Calvo, M., Dawes, J. M. & Bennett, D. L. The role of the immune system in the generation of neuropathic pain. *Lancet Neurol.* **11**, 629–642 (2012).
7. Sommer, C., Leinders, M. & Üçeyler, N. Inflammation in the pathophysiology of neuropathic pain. *Pain* **159**, 595–602 (2018).
8. Jain, A., Hakim, S. & Woolf, C. J. Immune drivers of physiological and pathological pain. *J. Exp. Med.* **221**, e20221687 (2024).
9. Goebel, A. Autoantibody pain. *Autoimmun. Rev.* **15**, 552–557 (2016).
10. Lacagnina, M. J., Heijnen, C. J., Watkins, L. R. & Grace, P. M. Autoimmune regulation of chronic pain. *Pain Rep.* **6**, e905 (2021).
11. Sadam, H. et al. Maternal antibodies shape infant immune response development in an epitope-spesific manner. *EBioMedicine* **123**, 106056 (2026).
12. Schwab, J. M. et al. Lesional antibody synthesis and complement deposition associate with de novo antineuronal antibody synthesis after spinal cord injury. *Neurol. Neuroimmunol. Neuroinflamm.* **10**, e200099 (2023).
13. Lacagnina, M. J. et al. B cells drive neuropathic pain-related behaviors in mice through IgG-Fc gamma receptor signaling. *Sci. Transl Med.* **16**, eadj1277 (2024).
14. Sadam, H. et al. Prostaglandin D2 receptor DP1 antibodies predict vaccine-induced and spontaneous narcolepsy type 1: Large-scale study of antibody profiling. *EBioMedicine* **29**, 47–59 (2018).
15. Sadam, H. et al. Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients. *EBioMedicine* **64**, 103211 (2021).
16. Jaago, M. et al. Differential patterns of cross-reactive antibody response against SARS-CoV-2 spike protein detected for chronically ill and healthy COVID-19 naïve individuals. *Sci. Rep.* **12**, 16817 (2022).
17. Pupina, N. et al. Immune response to a conserved enteroviral epitope of the major capsid VP1 protein is associated with lower risk of cardiovascular disease. *EBioMedicine* **76**, 103835 (2022).
18. Lötsch, J., Mustonen, L., Harno, H. & Kalso, E. Machine-learning analysis of serum proteomics in neuropathic pain after nerve injury in breast cancer surgery points at chemokine signaling via SIRT2 regulation. *Int. J. Mol. Sci.* **23**, 3488 (2022).
19. Mustonen, L. et al. HLA-region genetic association analysis of breast cancer patients with and without persistent postsurgical neuropathic pain. *Eur. J. Pain*. **29**, e70009 (2025).
20. Sundaresan, B., Shirafkan, F., Ripperger, K. & Rattay, K. The role of viral infections in the onset of autoimmune diseases. *Viruses* **15**, 782 (2023).
21. Kotsiri, I. et al. Viral infections and schizophrenia: A comprehensive review. *Viruses* **15**, 1345 (2023).
22. Fang, J. et al. Pathogen-specific exposure is associated with multisite chronic pain: A prospective cohort study. *Brain, Behav., Immun.* **129**, 157–164 (2025).
23. Fiore, N. T., Debs, S. R., Hayes, J. P., Duffy, S. S. & Moalem-Taylor, G. Pain-resolving immune mechanisms in neuropathic pain. *Nat. Rev. Neurol.* **19**, 199–220 (2023).
24. Deng, L., Gillis, J. E., Chiu, I. M. & Kaplan, D. H. Sensory neurons: An integrated component of innate immunity. *Immunity* **57**, 815–831 (2024).
25. Kim, H. W., Wang, S., Davies, A. J. & Oh, S. B. The therapeutic potential of natural killer cells in neuropathic pain. *Trends Neurosci.* **146**, 617–627 (2023).
26. Tékus, V. et al. A CRPS-IgG-transfer-trauma model reproducing inflammatory and positive sensory signs associated with complex regional pain syndrome. *Pain* **155**, 299–308 (2014).
27. Goebel, A. et al. Passive transfer of fibromyalgia symptoms from patients to mice. *J. Clin. Invest.* https://doi.org/10.1172/JCI144201 (2021).
28. Fanton, S. et al. Anti-satellite glia cell IgG antibodies in fibromyalgia patients are related to symptom severity and to metabolite concentrations in thalamus and rostral anterior cinculate cortex. *Brain Behav. Immun.* **114**, 371–382 (2023).
29. Kallio-Laine, K. et al. Widespread unilateral pain associated with herpes simplex virus infections. *J. Pain* **9**, 658–665 (2008).
30. Gielh, K. A. et al. Identification and characterization of 20 immunocompentent patients with simultaneous varicella zoster and herpes simplex virus infection. *JEADV* **22**, 722–728 (2008).
31. Broadwall, E. M., Pedersen, M., Asprusten, T. T. & Wyller, V. B. B. Pain in adolescent chronic fatigue syndrome following Epstein-Barr virus infection. *Scand. J. Pain* **20**, 765–773 (2020).
32. Mustonen, L. et al. What makes surgical nerve injury painful? A 4-year to 9-year follow-up of patients with intercostobrachial nerve resection in women treated for breast cancer. *Pain* **160**, 246–256 (2019).
33. Rosenberger, D. C. & Pogatzki-Zahn, E. M. Chronic post-surgical pain – Update on incidence, risk factors and preventive treatment options. *BJA Educ.* **22**, 190–196 (2022).
34. Grinde, B. Herpesviruses: latency and reactivation – viral strategies and host response. *J. Oral Microbiol.* **5**, 22766 (2013).
35. Bennett, J. M. et al. Inflammation and reactivation of latent herpesviruses in older adults. *Brain Behav. Immun.* **26**, 739–746 (2012).
36. Cohen, S. P. et al. Chronic pain and infection: mechanisms, causes, conditions, treatments, and controversies. *BMJ Med.* **1**, e000108 (2022).
37. Vieira, W. F. et al. Neuropathic pain, mood, and stress-related disorders: A literature review of comorbidity and co-pathogenesis. *Neurosci. Biobehav Rev.* **161**, 105673 (2024).
38. Kaunisto, M. A. et al. Pain in 1,000 women treated for breast cancer: A prospective study of pain sensitivity and postoperative pain. *Anesthesiology* **119**, 1410–1421 (2013).
39. Finnerup, N. B. et al. Neuropathic pain: An updated grading system for research and clinical practice. *Pain* **157**, 1599–1606 (2016).
40. Gerbershagen, H. J., Rothaug, J., Kalkman, C. J. & Meissner, W. Determination of moderate-to-severe postoperative pain on the numeric rating scale: A cut-off point analysis applying four different methods. *Br. J. Anaesth.* **107**, 619–626 (2011).
41. Bouhassira, D. et al. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic quenstionnaire (DN4). *Pain* **114**, 29–36 (2005).
42. van der Maaten, L. & Hinton, G. Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
43. van der Maaten, L. Accelerating t-SNE using tree-based algorithms. *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).
44. Krijthe, J. H. & Rtsne T-distributed stochastic neighbor embedding using a Barnes-Hut implementation. *R package version.* **0**, 17 (2015).
45. Hahsler, M., Piekenbrock, M. & Doran, D. Dbscan: Fast density-based clustering with R. *J. Stat. Softw.* **91**, 1–30 (2019).

46. Wickham, H. et al. Welcome to the Tidyverse. *J. Open. Source Softw.* **4**, 1686 (2019).
47. Kassambra, A. ggpubr: 'ggplot2' based publication ready plots. *R package version 0.6.0* (2023).
48. Xiao, N. & ggsci Scientific journal and sci-fi themed color palettes for 'ggplot2'. *R package version 3.0.0* (2023).
49. Wickham, H., Pedersen, T. & Seidel, D. scales: Scale functions for visualization. *R package version 1.3.0* (2023).
50. Wild, F. & lsa Latent semantic analysis. *R package version 0.73.3* (2022).
51. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
52. Thiele, C. & Hirschfeld, G. Cutpointr: Improved estimation and validation of optimal cutpoints in R. *J. Stat. Softw.* **98**, 1–27 (2021).
53. R Core Team. R: *A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org (2003).
54. Robin, X. et al. pROC: An open-source package for R and S+to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
55. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Soft* **28**, 1–26 (2008).



## Acknowledgements
The authors would like to thank the research nurse Eija Ruoppa for her excellent work. We thank all patients who participated in this study. Protobios acknowledges the excellent technical support from its team members Alex Sirp, Maria Piirsalu, Regina Maruste, Loviisa Pihlas, Epp Väli, and Helen Ausman. The authors thank Les Hearn, MSc, for reviewing the English language of the manuscript.


## Author contributions
EK, KP, HH, HS, AR and PJT contributed to the study design. EK and HH contributed to the data collection. HH conducted the clinical examination of the patients. HS, AR, MU, MT and AP performed immunoproliferation and validation experiments. HS, AR, MU, AP, MT, JN, LM and KP contributed to the data analysis and making of the figures and tables. HS, KP, AR and LM wrote the first draft of the manuscript. All authors were involved in the data interpretation, review and approval of the final version of the manuscript.


## Funding
Open access funded by Helsinki University Library. This study was supported by a grant to Protobios from the EU's HE research and innovation program under grant agreement No 101095436, with the clause that "*Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them*". In addition, the research by Protobios was supported by funding from the Estonian Ministry of Education and Research (5.1-4/20/170) and Estonian Research Council (PRG1953, KP and PSG691, MT). The clinical part of the study at the Helsinki University Hospital was supported by European Union FP7 (#Health_F2_2013-602891) grant NeuroPain and the Helsinki University Hospital Research Funds to the Department of Anaesthesiology, Intensive Care and Pain Medicine (Y102011092).


## Declarations

## Competing interests
The authors declare no competing interests.

## Additional information
**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-41637-6.


**Correspondence** and requests for materials should be addressed to L.M.



**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

# Curriculum Vitae


**Personal data**

| | |
|---|---|
| Name | Annika Rähni |
| Date of birth | 19.07.1995 |
| Location | Tallinn, Estonia |
| E-mail | annika@protobios.com |
| ORCID | 0000-0002-2826-4636 |


**Education**

| | |
|---|---|
| 2020 – | Tallinn University of Technology, doctoral studies |
| 2017 – 2019 | Tallinn University of Technology, MSc (Gene technology), diploma *cum laude* |
| 2014 – 2017 | Tallinn University of Technology, BSc (Gene technology) |

**Employment**

| | |
|---|---|
| 2019 – …. | Protobios LLC, technician-specialist |
| 2021 | Tallinn University of Technology, School of Science, Department of Chemistry and Biotechnology, course instructor |
| 2020 | Ludwig Maximilian's University of Munich, Germany, visiting scientist at the lab of Molecular and Behavioural Neurobiology |
| 2019 | University of Tartu, Institute of Genomics, assistant |

**Research training and dissemination**

| | |
|---|---|
| 2025 | PEGS Europe: Protein and Antibody Engineering Summit, Lisbon, Portugal, poster presentation "Decoding the antibody repertoire for diagnostics and therapeutics with mimotope variation analysis (MVA)" |
| 2025 | Estonian Society of Human Genetics, Annual Conference, Haapsalu, Estonia, poster presentation "Antibody repertoires indicate herpesvirus-associated immune responses in Long COVID" |
| 2025 | Winter Seminar in Gene Sciences, Jäneda, Estonia, presentation "Antibody Response Patterns in Psychosis Spectrum Disorders: Investigating Antibody-Epitope Signatures Using Phage Display and Next Generation Sequencing" |
| 2024 | TalTech, School of Science, Department of Chemistry and Biotechnology Christmas Conference, PhD student flash presentation "From Sniffles to Psychosis: Investigating Antibody-Epitope Signatures Using Phage Display and Next Generation Sequencing" |
| 2024 | Estonian Society of Human Genetics, Annual Conference, Viljandi, Estonia, poster presentation "Immunoprofiling: towards a new understanding of disease development and response to treatments" |
| 2024 | EMBL-EBI Bioinformatics Summer School, Wellcome Genome Campus, Hinxton, United Kingdom, poster presentation "Next generation phage display mimotope-variation analysis (MVA) |

defines antibody response patterns linking different phases of schizophrenia”

2023 Estonian Society of Human Genetics, Annual Conference, oral presentation “Uue põlvkonna faagidisplei rakendamine skisofreenia uuringutes”

2023 Estonian Society of Human Genetics, Annual Conference, Rakvere, Estonia, poster presentation “The antibody response patterns linking breast cancer and *Helicobacter pylori* infection – a next-generation phage display analysis”

2023 The 22nd International Gene Forum, Tartu, Estonia, poster presentation “Next generation phage display mimotope-variation analysis (MVA) defines antibody response patterns linking different phases of schizophrenia”

2023 Keystone Conference "Neuroimmune Interactions: From Basic Mechanisms to Novel Therapeutic Directions", Whistler, British Columbia, Canada, poster presentation “Next generation phage display mimotope-variation analysis (MVA) defines antibody response patterns linking different phases of schizophrenia”

2022 Course “Molecular diagnostics and precision medicine” (5 ECTS), Technical University of Denmark, online

2021 Conference Rudolf Buchheim 200: „New Essays on the Doctrine of Drugs“, Tartu, Estonia

2021 University of Tartu Spin-off programme pitching event, Tartu, Estonia, 3 minute pitch titled „From immune profiling to personalised medicine“

2021 FAIR Principle Workshop, University of Tartu and ELIXIR Estonia, Tartu, Estonia

**Publications**

Sadam, H.*, Mustonen, L.*, **Rähni, A.***, Nieminen, J. K., Toots, M., Uusväli, M., Pihlak, A., Tienari, P. J., Harno, H., Palm, K.**, & Kalso, E.** (2026). Comprehensive antigen profiling predicts post-surgical neuropathic pain in women treated for breast cancer. *Scientific Reports*, *16*(1), 12511. https://doi.org/10.1038/s41598-026-41637-6

Sadam, H., Maruste, R., **Rähni, A.**, Toots, M., Piirsalu, M., Pihlas, L., Ausman, H., Sirp, A., Pihlak, A., Laasfeld, T., Rull, K., Salumets, A., Antson, A., Tillmann, V., Aleksejeva, A., Rööp, T., Štšepetova, J., Sepp, E., Mändar, R., & Palm, K. (2026). Maternal antibodies shape infant immune response development in an epitope-specific manner. eBioMedicine, 123, 106056. https://doi.org/10.1016/j.ebiom.2025.106056

Lekk, I.*, Cabrera-Cabrera, F.*, Turconi, G.*, Tuvikene, J., Esvald, E.-E., **Rähni, A.**, Casserly, L., Garton, D. R., Andressoo, J.-O., Timmusk, T., & Koppel, I. (2023). Untranslated regions of brain-derived neurotrophic factor mRNA control its translatability and subcellular localization. *The Journal of Biological Chemistry*, *299*(2), 102897. https://doi.org/10.1016/j.jbc.2023.102897

**Rähni, A.**, Jaago, M., Sadam, H., Pupina, N., Pihlak, A., Tuvikene, J., Annuk, M., Mägi, A., Timmusk, T., Ghaemmaghami, A. M., & Palm, K. (2022). Melanoma-

specific antigen-associated antitumor antibody reactivity as an immune-related biomarker for targeted immunotherapies. *Communications Medicine*, *2*, 48. https://doi.org/10.1038/s43856-022-00114-7

Jaago, M.*, **Rähni, A.***, Pupina, N., Pihlak, A., Sadam, H., Tuvikene, J., Avarlaid, A., Planken, A., Planken, M., Haring, L., Vasar, E., Bićević, M., Lambert, F., Kalso, E., Pussinen, P., Tienari, P. J., Vaheri, A., Lindholm, D., Timmusk, T., … Palm, K. (2022). Differential patterns of cross-reactive antibody response against SARS-CoV-2 spike protein detected for chronically ill and healthy COVID-19 naïve individuals. *Scientific Reports*, *12*(1), 16817. https://doi.org/10.1038/s41598-022-20849-6

Jaago, M., Pupina, N., **Rähni, A.**, Pihlak, A., Sadam, H., Vrana, N. E., Sinisalo, J., Pussinen, P., & Palm, K. (2022). Antibody response to oral biofilm is a biomarker for acute coronary syndrome in periodontal disease. *Communications Biology*, *5*(1), 205. https://doi.org/10.1038/s42003-022-03122-4

Pupina, N., Avarlaid, A., Sadam, H., Pihlak, A., Jaago, M., Tuvikene, J., **Rähni, A.**, Planken, A., Planken, M., Kalso, E., Tienari, P. J., Nieminen, J. K., Seppänen, M. R. J., Vaheri, A., Lindholm, D., Sinisalo, J., Pussinen, P., Timmusk, T., & Palm, K. (2022). Immune response to a conserved enteroviral epitope of the major capsid VP1 protein is associated with lower risk of cardiovascular disease. *EBioMedicine*, *76*, 103835. https://doi.org/10.1016/j.ebiom.2022.103835

Sadam, H., Pihlak, A., Jaago, M., Pupina, N., **Rähni, A.**, Toots, M., Vaheri, A., Nieminen, J. K., Siuko, M., Tienari, P. J., & Palm, K. (2021). Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients. *EBioMedicine*, *64*, 103211. https://doi.org/10.1016/j.ebiom.2021.103211

Tuvikene, J., Esvald, E.-E.*, **Rähni, A.***, Uustalu, K.*, Zhuravskaya, A., Avarlaid, A., Makeyev, E. V., & Timmusk, T. (2021). Intronic enhancer region governs transcript-specific Bdnf expression in rodent neurons. *eLife*, *10*, e65161. https://doi.org/10.7554/eLife.65161

*, ** Equal contribution

# Elulookirjeldus

**Isikuandmed**

| | |
|---|---|
| Nimi | Annika Rähni |
| Sünniaeg | 19.07.1995 |
| Elukoht | Tallinn, Eesti |
| E-mail | annika@protobios.com |
| ORCID | 0000-0002-2826-4636 |

**Haridustee**

| | |
|---|---|
| 2020 – | Tallinna Tehnikaülikool, doktoriõpe |
| 2017 – 2019 | Tallinna Tehnikaülikool, magistrikraad (geenitehnoloogia), *cum laude* |
| 2014 – 2017 | Tallinna Tehnikaülikool, bakalaureusekraad (geenitehnoloogia) |

**Töökogemus**

| | |
|---|---|
| 2019 – …. | OÜ Protobios, tehnik-spetsialist |
| 2021 | Tallinna Tehnikaülikool, Loodusteaduskond, Keemia ja biotehnoloogia instituut, praktikumi juhendaja |
| 2020 | Münheni Ludwig Maximilian Ülikool, Saksamaa, lähetus Molekulaarse neurobioloogia laboris |
| 2019 | Tartu Ülikool, Genoomika instituut, assistent |

**Erialased kursused, konverentsid ja ettekanded**

| | |
|---|---|
| 2025 | PEGS Europe: Protein and Antibody Engineering Summit Lissabonis Portugalis, poster-ettekanne “Decoding the antibody repertoire for diagnostics and therapeutics with mimotope variation analysis (MVA)” |
| 2025 | Eesti Inimesegeneetika Ühingu aastakonverents Haapsalus, poster-ettekanne “Antibody repertoires indicate herpesvirus-associated immune responses in Long COVID” |
| 2025 | Geenitehnoloogide talvekool Jänedal, ettekanne “Antibody Response Patterns in Psychosis Spectrum Disorders: Investigating Antibody-Epitope Signatures Using Phage Display and Next Generation Sequencing” |
| 2024 | TalTech Loodusteaduskonna Keemia ja biotehnoloogia instituudi Jõulukonverents, suuline ettekanne “From Sniffles to Psychosis: Investigating Antibody-Epitope Signatures Using Phage Display and Next Generation Sequencing” |
| 2024 | Eesti Inimesegeneetika Ühingu aastakonverents Viljandis, poster-ettekanne “Immunoprofiling: towards a new understanding of disease development and response to treatments” |
| 2024 | EMBL-EBI Bioinformaatika suvekool, Wellcome Genome Campus Hinxtonis Suurbritannias, poster-ettekanne “Next generation phage display mimotope-variation analysis (MVA) defines antibody response patterns linking different phases of schizophrenia” |

| | |
|---|---|
| 2023 | Eesti Inimesegeneetika Ühingu aastakonverents Rakveres, suuline ettekanne “Uue põlvkonna faagidisplei rakendamine skisofreenia uuringutes” |
| 2023 | Eesti Inimesegeneetika Ühingu aastakonverents Rakveres, poster presentation “The antibody response patterns linking breast cancer and *Helicobacter pylori* infection – a next-generation phage display analysis” |
| 2023 | XXII Rahvusvaheline Geenifoorum Tartus, poster-ettekanne “Next generation phage display mimotope-variation analysis (MVA) defines antibody response patterns linking different phases of schizophrenia” |
| 2023 | Keystone-i konverents "Neuroimmune Interactions: From Basic Mechanisms to Novel Therapeutic Directions", Whistler-is (British Columbia) Kanadas, poster-ettekanne “Next generation phage display mimotope-variation analysis (MVA) defines antibody response patterns linking different phases of schizophrenia” |
| 2022 | E-kursus “Molecular diagnostics and precision medicine” (5 ECTS), Taani Tehnikaülikool |
| 2021 | Konverents Rudolf Buchheim 200: „New Essays on the Doctrine of Drugs“ Tartus |
| 2021 | Tartu Ülikooli Spin-off programmi idee-esitluse üritus Tartus, suuline ettekanne „From immune profiling to personalised medicine“ |
| 2021 | FAIR printsiipide töötuba, Tartu Ülikool ja ELIXIR Estonia, Tartus |

**Publikatsioonid**


Sadam, H.*, Mustonen, L.*, **Rähni, A.***, Nieminen, J. K., Toots, M., Uusväli, M., Pihlak, A., Tienari, P. J., Harno, H., Palm, K.**, & Kalso, E.** (2026). Comprehensive antigen profiling predicts post-surgical neuropathic pain in women treated for breast cancer. *Scientific Reports*, *16*(1), 12511. https://doi.org/10.1038/s41598-026-41637-6

Sadam, H., Maruste, R., **Rähni, A.**, Toots, M., Piirsalu, M., Pihlas, L., Ausman, H., Sirp, A., Pihlak, A., Laasfeld, T., Rull, K., Salumets, A., Antson, A., Tillmann, V., Aleksejeva, A., Rööp, T., Štšepetova, J., Sepp, E., Mändar, R., & Palm, K. (2026). Maternal antibodies shape infant immune response development in an epitope-specific manner. eBioMedicine, 123, 106056. https://doi.org/10.1016/j.ebiom.2025.106056

Lekk, I.*, Cabrera-Cabrera, F.*, Turconi, G.*, Tuvikene, J., Esvald, E.-E., **Rähni, A.**, Casserly, L., Garton, D. R., Andressoo, J.-O., Timmusk, T., & Koppel, I. (2023). Untranslated regions of brain-derived neurotrophic factor mRNA control its translatability and subcellular localization. *The Journal of Biological Chemistry*, *299*(2), 102897. https://doi.org/10.1016/j.jbc.2023.102897

**Rähni, A.**, Jaago, M., Sadam, H., Pupina, N., Pihlak, A., Tuvikene, J., Annuk, M., Mägi, A., Timmusk, T., Ghaemmaghami, A. M., & Palm, K. (2022). Melanoma-specific antigen-associated antitumor antibody reactivity as an immune-related biomarker for targeted immunotherapies. *Communications Medicine*, *2*, 48. https://doi.org/10.1038/s43856-022-00114-7

Jaago, M.*, **Rähni, A.***, Pupina, N., Pihlak, A., Sadam, H., Tuvikene, J., Avarlaid, A., Planken, A., Planken, M., Haring, L., Vasar, E., Baćević, M., Lambert, F., Kalso, E., Pussinen, P., Tienari, P. J., Vaheri, A., Lindholm, D., Timmusk, T., … Palm, K. (2022). Differential patterns of cross-reactive antibody response against SARS-CoV-2 spike protein detected for chronically ill and healthy COVID-19 naïve individuals. *Scientific Reports*, *12*(1), 16817. https://doi.org/10.1038/s41598-022-20849-6

Jaago, M., Pupina, N., **Rähni, A.**, Pihlak, A., Sadam, H., Vrana, N. E., Sinisalo, J., Pussinen, P., & Palm, K. (2022). Antibody response to oral biofilm is a biomarker for acute coronary syndrome in periodontal disease. *Communications Biology*, *5*(1), 205. https://doi.org/10.1038/s42003-022-03122-4

Pupina, N., Avarlaid, A., Sadam, H., Pihlak, A., Jaago, M., Tuvikene, J., **Rähni, A.**, Planken, A., Planken, M., Kalso, E., Tienari, P. J., Nieminen, J. K., Seppänen, M. R. J., Vaheri, A., Lindholm, D., Sinisalo, J., Pussinen, P., Timmusk, T., & Palm, K. (2022). Immune response to a conserved enteroviral epitope of the major capsid VP1 protein is associated with lower risk of cardiovascular disease. *EBioMedicine*, *76*, 103835. https://doi.org/10.1016/j.ebiom.2022.103835

Sadam, H., Pihlak, A., Jaago, M., Pupina, N., **Rähni, A.**, Toots, M., Vaheri, A., Nieminen, J. K., Siuko, M., Tienari, P. J., & Palm, K. (2021). Identification of two highly antigenic epitope markers predicting multiple sclerosis in optic neuritis patients. *EBioMedicine*, *64*, 103211. https://doi.org/10.1016/j.ebiom.2021.103211

Tuvikene, J., Esvald, E.-E.*, **Rähni, A.***, Uustalu, K.*, Zhuravskaya, A., Avarlaid, A., Makeyev, E. V., & Timmusk, T. (2021). Intronic enhancer region governs transcript-specific Bdnf expression in rodent neurons. *eLife*, *10*, e65161. https://doi.org/10.7554/eLife.65161

*, ** Võrdne panus

ISSN 2585-6901
ISBN 978-9916-80-570-1