EXAMENSARBETE INOM BIOTEKNIK,
AVANCERAD NIVÅ, 30 HP
STOCKHOLM, *SVERIGE 2021*

Investigating the autoimmunity profiles of Covid-19 patients

ALFRED KEDHAMMAR

KTH
SKOLAN FÖR KEMI, BIOTEKNOLOGI OCH HÄLSA

Abstract

The clinical severity of Covid-19 varies greatly between individuals, and all underlying risk factors are not yet well understood. Previous studies have shown Covid patients to be enriched with autoantibodies against type I interferons, suggesting autoimmunity may be an underlying factor of susceptibility to severe disease. In this project, the interplay between severe Covid-19 and autoimmunity was investigated in 114 Swedish patients, sampled in April-May 2020 as well as longitudinal re-samplings 4 and 8 months later, using the infrastructure of the Human Protein Atlas and the SciLife lab autoimmunity and serology profiling unit. First, 16 patients with few comorbidities were analyzed for autoantibodies at a near proteome-wide scale using planar microarrays, after which a custom antigen panel was assembled based on observed reactivities and literature studies. The antigen panel was implemented in a 384-plex suspension bead array which was run for all patient samples and a control group. Among the Covid patients, 23 antigens were called as differentially reactive and 8 of them were proposed as relevant to immunoregulation or Covid pathogenesis. The results partially replicated previous findings of autoimmunity directed to type I interferons and offer a list of candidate autoantigens for further inquiries.

Sammanfattning

Allvarlighetsgraden av sjukdomen Covid-19 varierar kraftigt mellan individer och alla underliggande riskfaktorer är ännu inte förstådda. Tidigare studier har påvisat Covidpatienter som överrepresenterade med autoantikroppar mot typ I interferoner, vilket förespråkar autoimmunitet som en möjlig underliggande riskfaktor till att utveckla allvarlig Covid. I detta projekt användes infrastrukturen av det mänskliga proteinatlasprojektet och enheten för autoimmunitets- och serologiprofilering på SciLife lab för att undersöka samspelet mellan allvarlig Covid-19 och autoimmunitet i 114 st svenska patienter inlagda under april-maj 2020, samt från uppföljningsprover 4 resp. 8 månader senare. Till en början undersöktes 16 patienter med låg grad av samsjukdom för förekomst av autoantikroppar mot proteomet i stort med hjälp av mikroarrayer. En panel av antigen sammanställdes därefter baserat på resultaten och litteraturstudier. Panelen implementerades som en 384-plex kulsuspensionsarray vilken kördes för alla patientprover samt en kontrollgrupp. Ibland Covidpatienterna klassades 23 st antigen som överrepresenterade, varav 8 st avsågs relevanta för immunoreglering eller sjukdomsförlopp. Resultaten visades delvis återskapa tidigare fynd av autoimmunitet riktad mot typ I interferoner och erbjuda en lista av potentiella autoantigen för vidare efterforskningar.

Keywords

Covid-19, SARS-CoV-2, Autoimmunity, Autoantibodies, Cytokines, Interferons

Contents

1 Introduction

1.1 Covid-19 and autoimmunity

Covid-19, is an infectious disease caused by the coronavirus SARS-CoV-2 which since the end of 2019 has spread into a pandemic with over 166,860,000 reported cases and 3,460,000 deaths [1], of which over 1,058,000 cases and 14,300 deaths are in Sweden [2].

Despite the relatively low case fatality rate, the severity and clinical outcome of the disease varies greatly between individuals. Even though several important risk factors such as age, sex and various underlying health conditions have been identified, many people still appear to be disproportionately affected (or unaffected) by the disease – implying all underlying risk factors are not yet well understood [3]. Furthermore, it has been established that the most severe damage caused by Covid- 19 is done by a dysregulated immune response, rather than the virus itself [4].

A proposed contributor to this clinical variability is autoimmunity which can be hypothesised as an underlying factor of susceptibility to severe disease as well as a tendency to develop serious complications post infection, so called "Long Covid". In autoimmunity, the immune system recognizes a native substance as foreign; such "autoreactive" substances are referred to as "autoantigens" and may lead to the production of specifically targeted "autoantibodies" (aABs) as well as a large variety of health complications and disease states. Most people exhibit autoreactivity to at least some native substances [5], but in some cases these reactivities may cause or predispose illness.

In Covid patients specifically, there have been demonstrations of autoimmunity directed towards cytokines – the molecules responsible for cross-talk and communication within the immune system. Interferons are a type of cytokine which specifically prevent and combat viral infections [6]. Bastard *et al* have shown that patients with severe Covid-19 are significantly enriched with autoantibodies against type I interferons and present evidence that these autoantibodies drive the severity of the disease [7]. Similarly, Zhang et al have shown that patients with non-functioning type I interferon responses, due to inborn errors, are more severely afflicted [8]. Both articles indicate that the genetic or autoimmune phenocopies of deficient or non-functional type I interferons are a driver of morbidity and mortality in Covid-19.

1.2 The Human Protein Atlas

The Human Protein Atlas (HPA) [9] is an extensive, Swedish-based program aiming to map the localization of all human proteins on a cell, tissue and organ level by integrating various omics technologies. This endeavour uses polyclonal antibodies, targeting different human protein epitopes, at a near proteome-wide scale. The HPA antibodies are produced by immunizing rabbits with synthetic peptides called protein epitope signature tags (PrESTs).

The PrESTs are expressed in *Escherichia coli* and are designed based on the following bioinformatical criteria [10]:

- Represent a predicted human protein, based on ensembl [11]
- Less than 60% sequence homology to other human proteins over a 50 aa sliding window
- Less than 8 aa sequence homology to other human proteins over a 10 aa sliding window
- No predicted transmembrane regions of signal peptides
- Between 25-150 amino acids long (average 80 aa)

Since 2003, the production of HPA antibodies has yielded over 42,000 PrESTs representing over 19,000 human proteins. Because the PrESTs are designed to represent a human protein epitope they can, besides the application of antibody production via immunization, also be coupled to planar or bead arrays to be utilized in autoimmunity assays.

1.3 Autoimmunity profiling technologies

In an autoimmunity assay, a panel of putative autoantigens is put in contact with a sample which is surveyed for the presence of autoantibodies directed towards their respective autoantigen in the panel. Once the autoantibodies are bound the total reactivity of each antigen can be evaluated, commonly by use of a secondary, anti-human detection antibody coupled to a fluorophore which is then measured optically.

1.3.1 Planar microarray

Autoimmunity assays can be performed on planar microarrays by coupling different antigens to ordered spots on a microarray slide. The broadest planar microarray within the HPA infrastructure, colloquially referred to as the "42k-array", contains the entire HPA PrEST collection and consists of two glass slides containing 21,000 PrESTs each [12]. While the antigens included in this array is a comprehensive representation of the complete set of the gene-derived proteome, there is no inherent possibility of sample multiplexing, so the throughput is limited to one sample or one sample pool per assay.

1.3.2 Suspension Bead Array

In a suspension bead array (SBA), antigens are instead coupled to magnetic beads in solution. Unlike the planar microarray where the identities of the antigens are retained by a spatial position in the array, the identity of each bead is spectrally discerned through a combination of different wavelengths and corresponding intensities of internal bead dyes [13]. Because all different beads can be pooled, allowing for antigen multiplexing, it is possible to scale the assay to run on hundreds or thousands of samples. The Luminex MagPlex beads and FlexMap 3D instrument allows for multiplexing of 384 antigens prepared in a four-plate protocol.

1.4 The Community study

The Community (short for Covid Immunity) study [14, 15] follows a cohort of over one hundred Covid patients and over two thousand healthcare workers from Danderyd university hospital. The purpose of the study is primarily to evaluate the longitudinal prevalence of anti-SARS-CoV-2 antibodies in the healthcare workers in connection with exposure and symptoms.

The 118 patients were sampled at multiple instances during hospital admission in April-May 2020 and invited for longitudinal re-sampling in September 2020 and January 2021. The initial sampling and two follow-up samplings are referred to as phases 1, 2 and 3 respectively.

At the time this project was initiated, the ethical framework only allowed for broad, prognostic biomarker discovery within the patients in the Community cohort, while the personnel were limited to investigations of SARS-CoV-2 antigens, specifically. Because this project aims at investigating the interplay between Covid and autoimmunity, only the patient samples were used.

1.5 Project outline

The intent of this project is to explore the autoimmunity profiles of the Covid patients of the Community study and compare them to a control group, leveraging the available technologies and infrastructure of the SciLife lab autoimmunity profiling facility and the Human Protein Atlas.

First, a subset of representative patients from phase 1 will be selected and analyzed via the broad, planar microarray. Secondly, based on the reactive PrESTs of the microarray and genes described in literature as relevant to the interplay between autoimmunity and Covid pathogenesis, an antigen panel will be constructed and implemented as a custom, 384-plex suspension bead array. Lastly, this SBA, which is more targeted than the planar microarray but highly scalable, will be utilized for all patients of all phases of the study as well as a control group.

The goal of the project is to generate a list of candidate autoantigen which are differentially reactive in Covid patients compared to the control group. These findings will also be put in relation to those of Bastard *et al*, by investigating in particular the observed autoreactivity towards type I interferons. The data from phases 2-3 will enable longitudinal analysis in a subset of patients, which can be used to investigate how reactivity towards antigens may change over time.

2 Materials and methods

2.1 Planar microarray

The limited throughput of the planar microarray protocol was sought to be increased in order to include more subjects and hence capture more autoreactivities. Previous endeavours have shown that the samples of up to four patients can be pooled for the analysis while still yielding interpretable results. For this reason, it was decided to run the analysis on four pools containing four samples each.

Because the exploratory analysis of the autoreactivities of these 16 patients was expected to broadly affect the composition of the custom antigen panel, it was deemed important to minimize the number of comorbidities. We reasoned that it would be more interesting to investigate the patients that had become inexplicably ill, rather than the ones whose disease severity could be attributed to any of a vast assortment of comorbidities.

2.1.1 Sample selection

In consultation with a collaborating clinician, the clinical data of 115 hospitalized Covid patients was manually surveyed. Variables of interest were age, sex, previously known health conditions (comorbidities) and disease severity. The care unit submission type was deemed an appropriate categorical marker of disease severity: submission to intensive care (IVA) meant the patients had at some point required invasive assisted respiration, while submission to intermediary care (IMA) meant the patients had at some point required non-invasive assisted respiration.

The presence and variation of comorbidities was very high in the cohort, and isolating patients with no previously known comorbidities and IVA/IMA submission yielded only 12 patients which could be stratified by sex and clinical severity into three pools. The first pool consisted of four men submitted to IVA, the second of four men submitted to IMA and the third of four women submitted to IVA or IMA. In this third pool, one woman had to be excluded and was replaced with a woman with documented asthma. In contrast, the fourth and final pool was filled with "wildcard" patients, selected manually by the collaborating clinician. Of these, two had multiple sclerosis (MS) and were undergoing treatment with the drug Mabthera, one had Guillain-Barre syndrome (GBS) contracted in connection to Covid and one was immunosuppressed.

Table 1: Four pools containing four patient samples each were selected based on known comorbidities, sex and disease severity.

Pool	Comorbidities	Sex	Severity
1	11 w. none	Male	IVA (critical)
2			IMA (severe)
3	1 w. asthma	Female	Mixed
4	Autoimmune disease or immunosuppression	Mixed	

2.1.2 Experimental procedure

The samples, consisting of heat-treated plasma were moved from -80°C storage, thawed overnight in -20°C storage and thawed prior to handling in 4°C. From each sample, 15 µL was transferred to its respective pool for a total volume of 60 µL per pool.

An assay buffer was prepared, consisting of 10 ml phosphate buffer solution (09-9400-100, Medicago) + 0.1% Tween20 (BP337-500, Fisher) (PBST), 0.3 g 3% Bovine Serum Albumin (B2000-500, Saveen Werner) (BSA) and 0.5 g 5% milk (70166-500G, Sigma).

Because each planar microarray consists of 2 glass slides containing 21,000 PrESTs each, the eight required glass slides (in-house batch 5) were retrieved from 8°C storage, mounted to a slide rack and dipped in milliQ several times. They were then dried in a slide centrifuge and transferred to horizontal slide mounts for handling. Microarray coverslips (100 µl, LifterSlip) were cleaned with ethanol, distilled water and blow dried.

Each sample pool was diluted 1:25 (to obtain 1:100 dilution for each sample) with 160 µg His_6ABP/ml in assay buffer. The diluted pools were then incubated for 15 min in an Elmi Intelli-Mixer RM-2L, program UU at 30 rpm. To each slide, 85 µL of its respective pool was applied in a vertical line after which the chamber cover was placed on top to allow the liquid to distribute uniformly across the entire slide. All slides were then incubated in RT for 1 h.

The slides were then put into 50 ml Falcon tubes filled with 0.1% PBST to loosen the chamber covers which were removed using a pincer. The slides were the transferred to a slide rack which was washed by submerging in 0.1% PBST for 5 min on a shake table at 80 rpm.

The slides were then incubated with 1:40 000 Hen anti-His6ABP IgY (25 mg/ml, Immunotech HPA), submerged in PBST on a shake table at 80 rpm for 1 h, after which they were washed twice.

The slides were then incubated with 1:15 000 goat anti chicken Alexa 555 (2 mg/ml, A-21437, Invitrogen) and goat anti-human IgG (H+L) Alexa 647 (2 mg/ml, A21445, Life Technology), submerged in PBST on a shake table at

80 rpm for 1 h. Finally, they were washed twice with PBST, once with PBS, dipped in MilliQ and dried in a slide centrifuge.

2.1.3 Image analysis

The slides were scanned using an InnoScan 1100AL (Innopsys) with the parameters width: 22,000, length: 72,000, red channel: PMT 950; Power 100 and Green channel: PMT 750; power 75. The green channel (532 nm) corresponds to the detection of the His_6ABP tag, i.e. PrESTs, while the red channel (635 nm) corresponds to the detection of human IgG.

The scanned images were then interpreted using the software GenePix 5.1 (Molecular Devices) and a grid detailing the spatial coordinates of each feature was manually fitted to the image. The settings were the default 2-colour mode settings for the mode "find irregular features" with the parameters composite pixel intensity: 300 MFI, resize features during alignment: feature diameter min 33% - max 150%.

For each slide, the result of the automated feature alignment tool was briefly manually inspected and any discoveries of calls which were deemed erroneous were flagged as such. The highest ranked intensities (approximately a dozen per slide) were also systematically inspected and curated the same way.

Once inspected and manually curated, each feature alignment was converted to a data structure in .gpr-format detailing the identity and metrics of each spot therein.

2.2 Antigen panel assembly

2.2.1 Control beads

The 384-plex SBA would include four control beads, leaving room for 380 antigens in the customized panel. The control beads were:

- An "empty" bead with no coupled antigen, included as a negative control for the array.
- A bead coupled with anti-human IgG Fc antibodies, included as a positive control for all samples containing human IgG.
- A bead coupled to Epstein–Barr virus nuclear antigen 1 (EBNA1) which is expected to exhibit a variable reactivity across human samples.
- A bead coupled to the His_6ABP tag present on all PrESTs, included as a negative control to demonstrate that the samples have been depleted of anti-His_6ABP antibodies and that any activity exhibited by a PrEST is not due to the tag, but the PrEST itself.

2.2.2 Full length type I interferons

Two full length type I interferons (IFNα2A and IFNω1) were purchased and included in the panel (SRP4594-100UG, SRP3061-100UG, Sigma).

This was done partially in order to replicate the findings of Bastard *et al* [7] but also to cross-reference the reactivities of the full length interferons to those of the PrESTs. If reactivity is seen for the former but not the latter, it would indicate the PrESTs do not fully represent the same, reactive epitopes as the full length proteins. Reactivity among the full length proteins and differential reactivity among the PrESTs might also indicate which epitopes are targeted, specifically.

2.2.3 PrEST antigen

The remaining 378 antigen in the panel would consist of PrESTs.

A literature study was conducted and a list was compiled containing genes relevant to the interplay between autoimmunity and Covid or other infectious diseases. These 24 genes were 13 subtypes of interferon α, interferons ω, β, κ, ε [7], γ [16], interleukins 6 [17], 17A, 17F , 22 [18], 23 and Granulocyte-macrophage colony-stimulating factor (CSF2) [19].

A list of all genes which had been represented by reactive PrESTs in the planar microarray analysis as well as all genes from the literature study was compiled and input to LIMS to yield a list of corresponding PrESTs IDs (HPRR) and antigen IDs (HPRA). HPRR identifies the PrEST as a bioinformatical construct while HPRA identifies the batch of the PrEST as a physically expressed antigen. Rows with missing HPRA were removed and the remaining

rows were deduplicated with respect to HPRR. This list was, in turn, run on an in-house script to generate a data frame containing the physical locations of the antigens which was then supplemented with information from the microarray results.

The goal was to include all PrESTs which had been reactive in the planar microarray analysis as well as all PrESTs representing genes from the literature list. Because all PrESTs were not expected to be available, a "backup" list of PrESTs was constructed for filling up whichever space remained in the panel. These were PrESTs that represented proteins with an immunological function and extracellular localization, and that had been represented in the planar microarrary analysis by some other reactive PrEST, sorted by descending reactivity.

The code for generating the location list can be found in appendix F.2.

2.2.4 Final composition

Available PrESTs were collected from in-house storage as well as the facilities of Atlas Antibodies in Bromma. The collection and final antigen panel is summarized in table 2.

Table 2: Composition of the final antigen panel in terms of antigen type, motivation for inclusion and storage type. Physically, the panel consisted of 328 tubes and 52 plate wells from 37 different plates.

#	Type	#	Comment	#	Storage
378	PrEST	325	Reactive >4 SD in microarray analysis	281	Tube
				44	Plate
		25	Represents proteins mentioned in literature	21	Tube
				4	Plate
		28	Represents proteins that have other reactive PrESTs, immunological function and extracellular localization	24	Tube
				4	Plate
2	Full length protein	2	IFNα2A and IFNω	2	Tube
4	Control beads	4	Empty, anti-human IgG Fc, EBNA1, His_6ABP	4	Tube
384		**384**		**384**	

2.3 Suspension Bead Array

The experimental procedure was carried out based on the in-house standard operating procedure (SOP) for SBA autoimmunity profiling (version 2.2), variations of which have been used for similar projects [5, 20].

2.3.1 Antigen dilution

The 380 antigens and four controls are diluted and transferred to four 96-well plates.

The finished antigen panel was formatted and fed into an in-house script (v11) which generated three work list files (.gwl) compatible with the liquid handling system used (Tecan Freedom EVO) as well as an updated antigen position list and worktable layouts. The updated list detailed the expected source and destiny positions of all tubes and wells and the volumes required to achieve their target concentrations. Some wells would be diluted manually.

First, 4x50 ml MES buffer (M2933-100G, Sigma) was thawed overnight at 4°C. The antigens and control reagents were also thawed at 4°C before handling. The control reagents consisted of rabbit anti-human IgG (309-005-082, Jackson), EBNA1 (ab138345, abcam), His6ABP (Human Protein Atlas, batch conc. 5.78 mg/ml).

The liquid handling system was inspected, loaded with tips and pumped with system liquid (dH_2O) twice to remove air bubbles. The worktable was loaded with four 10 ml tubes of MES buffer and the appropriate amount was distributed to each well of the four 96-well plates.

For the antigen dilution carried out by the liquid handling system, 322 antigen tubes were transferred to 48-tube holders in the order specified by the updated list, vortexed, spun down (2000 RPM, 1 min) and loaded onto the worktable. Two different worktable layouts were used to include all tubes in up to four tube holders at a time.

Similarly, 34 antigen plates containing 49 antigens of interest were vortexed, spun down (stacked two at a time, 1000 RPM, 1 min) and loaded in three different layouts running on 12, 12 and 10 plates at a time.

Three antigen plate wells (IDs 375-377) and three tubes (IDs 14, 70, 71) had to be transferred manually due to low well volume and incompatible tube format. An additional antigen was manually transferred to it's tube because it had not been included in the original script run (ID 380). The two full length proteins were hydrolyzed with 100 µl milliQ according to manufacturers' instructions, vortexed and transferred manually to destiny plates along with the three control reagents. The fourth control well was simply filled with buffer, totaling at 384 wells.

For all antigens, the full-length proteins and His6ABP, the final concentrations were 0.4 mg/ml. For the rabbit anti-human IgG and EBNA1 the final concentrations were 1.8 and 1.35 mg/ml, respectively. The destiny plates were sealed, vortexed, spun down (2000 RPM, 1 min) and transferred to overnight storage in 4°C.

2.3.2 Bead coupling

The beads are chemically activated and the antigens are coupled to the beads.

Four 96-well plates containing 40 µl Luminex MagPlex bead stock solution (500 000 beads) per well were prepared and stored in 4°C wrapped in foil. The MES buffer was used, as well as 3x50 ml AB buffer (S3139-250G, Sigma) thawed in 4°C overnight and 5 ml 10xBRE (11 112 589 001, Roche) thawed in 4°C the same day.

Four aliquots of 65 mg EDC (Proteochem) as well as stock NHS (24510, Thermo Scientific) were transferred from -20°C to RT for 30 min. Aliquots of NHS were made and weighed at 66.7, 66.7, 66.1 and 69.9 mg before being put on silica beads along with the EDC aliquots for 30 min.

The diluted antigen plates were vortexed and spun down (2000 RPM, 1 min) and the bead plates were moved to RT and spun down (2000 RPM, 1 min).

Due to the nature of the reagents used, the following steps are time sensitive and were performed twice, for two plates at a time.

First, the plates were washed by adding 80 µl AB buffer via automated multipipette and draining on magnet using a BioTek EL406 washer dispenser. A further 50 µl AB buffer was then added via automated multipipette.

For each plate, one aliquot of EDC and NHS were diluted separately in AB buffer to 50 mg/ml after which 1200 µl of both solutions were combined with a further 3600 µl AB buffer. The resulting 6000 µl activation solution was transferred to a trough and dispensed at 50 µl (or 0.5 mg EDC/NHS) per well into the corresponding bead plate using an automated multipipette.

The plates were then sealed, vortexed and incubated at RT in a hooded plate shaker (650 RPM, 20 min).

After incubation, the plates were pulsed at 1000 RPM and washed twice on magnet in 2x100 µl MES buffer using an automated multipipette for dispensation and BioTek for aspiration.

The diluted antigens (100 µl per well) were then added to the bead plates. The plates were then sealed, vortexed and incubated at RT in a hooded plate shaker (650 RPM, 2x1h with vortexing between).

Storage buffer was prepared by adding 45 ml milliQ and 50 µl ProClin (48912-U, Sigma) to the thawed 5 ml 10xBRE buffer.

The following steps were also performed twice, two plates at a time due to the time lag introduced earlier.

After incubation, the plates were washed twice in 100 µl 0.05 PBS-T on magnet using BioTek. Using an automated multipipette, 50 µl storage buffer was added to each well off-magnet. The plates were then sealed and vortexed.

Following protocol completion for all four plates, they were vortexed, pulsed at 1000 RPM and stored at 4°C until bead pooling 6 days later.

The complete list of bead couplings can be found in appendix E.

2.3.3 Bead pooling

The beads are pooled into one 96-plex tube per plate.

The plates were moved from storage and pulsed at 1000 RPM. The beads were then re-suspended in the plate wells by repeated pipetting using a manual multi-channel pipette.

For each plate:

A flat-bottomed, slide washer tray was placed, at a tilt, on a plate magnet after which the whole bead plate volume was transferred to the side of the tray in contact with the magnet. The pipette tips were reused for the whole transfer to limit bead loss. The wells were washed with an additional 50 µl storage buffer each and were then transferred to the tray in the same manner.

The tray was re-positioned horizontally on the magnet and 7 ml storage buffer was carefully aspirated from the opposite side of the beads and transferred to a 15 ml tube on magnet. The beads were re-suspended in remaining 3 ml storage buffer in the tray by repeated aspiration and transferred to a 5 mL LoBind tube on magnet where the beads were allowed to settle.

The entire tray was washed thrice using 700 µl supernatant from the 5 ml LoBind tube. The volume of the tube was then visually adjusted to 1.25 ml, resulting in a 96-plex SBA pool.

From each 96-plex pool, 300 µl was transferred after vortexing to a 1.5 ml LoBind tube, corresponding to a 1200 µl 384-plex SBA pool.

2.3.4 Pre-run assay tests

The coupling of the beads is evaluated.

In order to evaluate the quality of the SBA, a control assay run was performed on 21 wells which included three different types of tests:

- A coupling test, in which a hen anti-His_6ABP antibody is added in PBST to bind the His_6ABP tag present on all PrESTs (n=378) and the His_6ABP control bead. The purpose of the test is to verify that the antigens have successfully coupled to the beads. The test is performed in triplicate and with one PBST control.
- A sample test, in which commercial plasma is added in assay buffer. The reactivity profile may vary depending on the plasma, but reactivity is expected for the anti-human IgG control bead and likely for the EBNA1 control bead as well. The purpose of the test is to verify that running on plasma yields acceptable results. The test is performed in triplicate and with one assay buffer control.
- A number of specific antigen tests, in which HPA antibodies directed towards specific PrESTs are added in PBST to verify that they bind their specific targets on the beads. In this project, six HPA antibodies are used. Three of them target IFNα2A, IFNω and both respectively. They were included to evaluate binding to the full length interferons present in plate 4 as well as the IFNα2A/ω PrESTs in pool 1. The remaining three target arbitrarily chosen PrESTs in pool 2, 3 and 4. Each specific antigen test is performed in duplicate and with one PBST control. A summary of the chosen antibodies and their targets can be seen in table 3.

Table 3: Summary of the antibodies used for the specific test. All antibodies targeting beads in plate 1 were chosen for also evaluating reactivity to the full length interferons in plate 4, while the antibodies targeting plates 2-4 were chosen arbitrarily.

HPA antibody	Target PrEST	Target bead	Target genes	Target plate
HPA045659	HPRR3360041	32	IFNA various, IFNW1	1
HPA047557	HPRR3360040	33	IFNA various	1
HPA053957	HPRR3780122	89	IFNW1	1
HPA031823	HPRR2800004	147	SNAP29	2
HPA058595	HPRR3720498	288	ELF4	3
HPA059220	HPRR3560056	294	FJX1	4

The test solutions were prepared: For the coupling test, 0.5 µl hen anti-His_6ABP (produced in-house) was added to 5 ml PBST for a 1:10 000 dilution. For the sample test, 300 µl assay buffer w/o His_6ABP, consisting of phosphate buffer solution (09-9400-100, Medicago) + 0.1% Tween20 (BP337-500, Fisher) (PBST), 3% Bovine Serum Albumin

(B2000-500, Saveen Werner AB) (BSA) and 5% milk (70166-500G, Sigma) was combined with 8.54 µl 5.78 mg/ml His_6ABP solution (in-house) to yield assay buffer including 160 µg/ml His_6ABP. Of this assay buffer, 250 µl was combined with 1 µl commercial plasma (K2EDTA mixGender, cat#HUMANPLK2PNN, lot#HMN192452, BioIVT) for a 1:250 dilution. For each specific antigen test, 0.5 µl of the corresponding HPA antibody was added to 250 µl PBST for a 1:500 dilution.

To each of the 21 wells, 5 µl vortexed 384-plex SBA solution was added. To each of the 18 testing wells, 45 µl of the respective test solutions were added and to the 3 control wells, the same solutions without primary antibodies were added. The plate was sealed, vortexed and incubated on shake (650 RPM, 1h). The plate was then pulsed at 1000 RPM and washed thrice in 100 µl PBST using BioTek.

The detection antibodies were diluted: For the coupling test, 0.5 µl 0.5 mg/ml R-PE conjugated donkey F(ab')2 fragment anti-hen IgY (703 116 155, Jackson) was diluted in 500 µl PBST. For the sample test, 1 µl 0.5 mg/ml R-PE conjugated goat F(ab')2 fragment anti-human IgG (eBioScience) was diluted in 625 µl PBST. For the specific antigen test, 1 µl 1 mg/ml R-PE-conjugated anti rabbit IgG (111-116-144, Jackson) was diluted in 2000 µl PBST.

To each well, 50 µl of the corresponding detection antibody solution was added. The plate was then sealed, vortexed and incubated on shake (650 RPM, 30 min). The plate was then pulsed at 1000 RPM, washed thrice in 100 µl PBST and finally dispensed with 100 µl PBST using BioTek.

The plate was lastly transferred to a Luminex Flexmap 3D instrument. It was run using a 96-well protocol with a 60 second timeout per well.

2.3.5 Sample selection

The Covid-19 patient cohort consisted of samples from phases 1-3 of the Community study [15] i.e. 118 patient samples from April-May 2020, 60 follow-up samples from September 2020 and 51 additional follow-up samples from January 2021. In phase 1 many patients had been sampled multiple times and for these cases the last sample in the series was used. Three replicates consisting of commercial plasma and one empty control well were to be included in each of the four 96-well assay plates.

The 229 Covid-19 patient samples and 16 plate controls left space for 139 control samples. The control samples originated from a large in-house study called EIMS, consisting of multiple sclerosis (MS) patients and healthy controls (cohort ID MSC_6). To make the sample transfer as simple as possible it was decided that the 139 controls would be transferred plate-wise from approximately one and a half of the twenty-four 96-well plates in storage.

It was thus of interest to find the control plates which would be most appropriate to use. The cohort sample information was downloaded from the laboratory information system (LIMS) and processed in R [21] using package tidyverse [22]. The code can be found in appendix F.4. No single plate had more than 35% men or 21% healthy men so it was deemed relevant to find the one and a half plates which would maximize these parameters to strive for an even gender ratio and a large healthy control group. It was decided to use plate 14 and the second half of plate 16, together corresponding to 34% men and 60% healthy controls. The composition of all samples used in the assay can be seen in table 4.

Table 4: Samples used for the SBA run. For the Covid patients, statistics of sex and age were obtained from published data [15].

#	Type	#	Group	% men	mean age
229	Case	118	Hospital admission	69	56
		60	4 month follow-up		
		51	8 month follow-up		
139	Control	84	Healthy	34	40
		54	MS		
		1	Unknown		
16	Plate control	12	Commercial plasma	-	-
		4	Empty		
384		**384**		**56**	**50**

A graphical layout was constructed for performing plate-wide transfers from the six crude sample plates to the four assay plates using a semi-automatic pipettor and control wells were arbitrarily but evenly distributed across the plates (see appendix B).

2.3.6 Sample dilution

First, two sets of assay buffer were prepared: one with His_6ABP (set 1) and one without (set 2). Set 1 consisted of 50 ml phosphate buffer solution (09-9400-100, Medicago) + 0.1% Tween20 (BP337-500, Fisher) (PBST), 1.5 g 3% Bovine Serum Albumin (B2000-500, Saveen Werner) (BSA) and 2.5 g 5% milk (70166-500G, Sigma). Set 2 was prepared by adding 1.19 ml 4.4 mg/ml His_6ABP (in-house) to 15 ml of set 1 solution. Set 1 and set 2 were transferred (49 µl and 60 µl per well) to two sets of four skirted 96-well PCR-plates using an automated multipipette.

The two crude control sample plates were thawed from -80°C to -20°C overnight and then put on ice with the remaining four crude sample plates. Once thawed, they were vortexed and spun down (3000 RPM, 1 min).

In accordance with the layouts in appendix B, 1 µl of each of the selected samples was transferred from the six crude plates to the four set 1 plates. Plate-wide transfers were performed using a CyBio SELMA semi-automatic pipettor while some wells were transferred manually to allow a seamless merge of the source layouts as well as an even distribution of replicates and control wells within the destination plates.

The set 1 plates and crude plates were sealed, vortexed and centrifuged (3000 RPM, 1 min) and the crude plates were then returned to -20°C storage overnight to be returned to -80°C later. The CyBio SELMA semi-automatic pipettor was then used to directly transfer 15 µl per well from the set 1 plates to the set 2 plates, which were then sealed, vortexed, centrifuged (3000 RPM, 1 min) and incubated on a shake table (80 RPM, 1h).

2.3.7 Assay run

Of the 384-plex SBA pool, ∼1095 µl remained from the bead pooling (section 2.3.3). An additional 280 µl was added from each vortexed 96-plex pool to yield ∼2215 µl 384-plex pool. An automatic stepper pipette was used to transfer 5 µl per well of the 384-plex SBA pool to a 384-well Greiner assay plate, 16 wells at a time with vortexing between.

The set 2 plates were centrifuged (3000 RPM, 1 min) and transferred, one plate per quadrant, to the assay plate using a Tecan Freedom Evo liquid handling system. The assay plate was then sealed, vortexed and incubated in a hooded plate shaker (650 RPM, 2x1h with vortexing between).

To crosslink any bound autoantibodies to the beads, 28.5 ml PBST and 1.5 ml paraformaldehyde (43368, Alfa Caesar) (PFA) were mixed to a 0.2% solution. Using a BioTek EL406 washer dispenser, the assay plate was washed thrice in PBST, after which 50 µl of the 0.2% PFA was added to each well. The assay plate was then sealed, vortexed and incubated in a hooded shake table (650 RPM, 10 min).

Detection antibody solution was prepared by diluting 40 µl 0.5 mg/ml goat anti-human IgG Fc (12-4998-82, eBioscience) in 25 ml PBST for a final concentration of 0.8 µl/ml. The assay plate was pulsed (1000 RPM) and again washed thrice in PBST using the BioTek, after which 50 µl of the diluted detection antibody was added using an automated multipipette. The plate was then sealed, vortexed and incubated in a hooded plate shaker (650 RPM, 30 min). Finally, the plate was pulsed (1000 RPM), washed thrice in PBST and dispensed with 60 µl PBST per well using the BioTek.

The plate was ran on a Luminex Flexmap 3D instrument with 45 second timeout and a double wash every 96^{th} sample.

3 Data analysis

The data analysis was carried out in R [21] using the packages tidyverse [22], limma [23], ggbeeswarm [24] and ggpubr [25].

3.1 Planar microarray

The code is available in appendix F.1. Prior to this analysis, results were run on an in-house script to generate quality reports and metrics of e.g. database similarity, i.e. how frequently a certain reactivity had been observed in previous runs.

3.1.1 Filtering

The image analysis data for each slide was imported in .gpr-format and merged into a data frame. The database similarity was incorporated from the in-house script result. The data was filtered to remove spots which were either flagged in the image analysis, non-PrESTs, smaller than 30 pixels or whose PrEST detection signal was lower than 10 standard deviations above that of the spot background. For replicate spots (i.e. spots presenting the same PrEST), only the highest signal spot was kept.

3.1.2 Normalization

The signal was defined as the pixel median fluorescence intensity (MFI) of the spot subtracted with the MFI of its local background (see equation 1). The main, normalized, signal measure used in this report is referred to as standard deviations above baseline (SDs), defined as the difference between the spot MFI and the mean MFI of the slide, divided by the MFI standard deviation of the slide (see equation 2). The statistics are used on a per-slide basis to compensate for differing plasma concentrations across the pools as well as technical variation between the slides.

for spot i on slide j:

$$MFI_{i,j} = MFI_{i,j}^{foreground} - MFI_{i,j}^{background} \tag{1}$$

$$SDs_{i,j} = \frac{MFI_{i,j} - mean(MFI_j)}{SD(MFI_j)} \tag{2}$$

3.1.3 Annotation

Gene name, gene descriptions and Uniprot ID for each PrEST was incorporated from an in-house database. Gene Annotation (GO) [26] information was then downloaded via Uniprot [27] and incorporated as well. After manual inspection of the GO-annotated results, two lists of keywords were constructed to flag PrESTs which represented proteins with an immunological function or which were extracellularly accessible, see appendix A.

3.1.4 Reactivity calling

A threshold of 4 SDs were applied to determine which PrESTs were reactive. The final results included 352 reactive spots corresponding to 332 unique PrESTs.

3.2 Suspension Bead Array

The code for the pre-run assay tests can be found in appendix F.3 and the code for the analysis of the main run can be found in appendix F.5.

3.2.1 Pre-run assay tests

The Luminex bead count and MFI for each well and bead ID were merged with the test metadata for easier handling.

The bead count across all beads was surveyed. For the coupling test and sample test, the mean MFIs of the test triplicates were calculated for each bead ID and compared to the negative controls. For the specific antigen tests, the mean MFIs of the HPA antibody duplicate samples were calculated for each bead ID and the signal profiles were surveyed.

3.2.2 Data structuring

The Luminex data consisted of the measured bead count and median bead fluorescence intensity (MFI) for each sample and antigen. These two 384x384 matrices were pivoted into a long table containing 147,456 observations. All available patient information was downloaded from LIMS and merged into a comprehensive data structure. Similarly, the same methodology as section 3.1 was employed to collect information about all beads and their contents. The observations table was merged with both the sample information and the antigen information to yield a complete data structure containing all relevant information for each observation.

In order to assess to quality of the SBA, all observations of the control beads were inspected.

3.2.3 Rationale of statistical methods

When making group comparisons of autoimmunity profiles, especially in a discovery-oriented endeavour, the majority of observations across samples and antigens are expected to be non-reactive. For this reason classical quantitative approaches, such as the Wilcoxon Rank Sum Test or t-test, tend to measure small but consistent group differences in the non-reactive bulk of observations rather than the relatively few, reactive outliers which actually represent the property of interest.

A more suitable approach is to set a threshold, separating observations to reactive and non-reactive respectively. For the observations of each antigen, separating by group and reactivity results in a contingency table were the statistical significance can be computed with Fisher's exact test, indicating whether the antigen is differentially reactive in a particular group.

Setting a threshold to determine which observations are reactive and not is far from obvious: samples may differ in plasma concentration and quality, resulting in different per-sample background levels. Likewise, the different antigens may be subject to vastly different background levels due to differences in concentration or binding availability.

In order to separate all differetially reactive observations from their respective backgrounds, one approach is to make the observations directly comparable by normalizing each observation by it's own background and then applying a single threshold across all observations. In order to quantitatively compare the observations of a specific antigen across multiple samples, we can normalize by sample background. Likewise, in order to compare the observations of a specific sample across multiple antigens we can normalize by the antigen background. This is based on the assumption that the majority of observations across samples and antigens are nonreactive. This assumption is expected to hold true for all observations except those which measure the positive control beads.

One metric which can represent the background level across a set of values is the MAD, defined as the median absolute deviation of the values from the median of the set.

$$For\ \{d_1, d_2, d_3...\} \in D$$

$$MAD(D) = median(\{|d_i - median(D)|\}) \tag{3}$$

We can normalize the observations per sample, per antigen or both - defining the signal as the number of MADs from the median.

$$For\ observation\ MFI_{i,j}\ of\ sample\ s_i \in S\ and\ antigen\ ag_j \in Ag$$

$$MADs^{S}_{i,j} = \frac{MFI_{i,j} - median(s_i)}{MAD(s_i)} \tag{4}$$

$$MADs^{Ag}_{i,j} = \frac{MFI_{i,j} - median(ag_j)}{MAD(ag_j)} \tag{5}$$

$$MADs_{i,j}^{S,Ag} = \frac{MFI_{i,j} - median(s_i) - median(ag_j)}{MAD(s_i) * MAD(ag_j)} \tag{6}$$

3.2.4 Normal- and binarization

First, all observations associated with a bead count < 30 were removed (n = 771). Then, the median and MAD was calculated with respect to each sample and antigen (excluding the positive control beads). The normalized signals for each observation were then calculated as described in equations 4-6.

The threshold for reactivity was set to 1.5 $MADs^{S,Ag}$ based on the profile of the negative control beads. The subsequent number of reactivities per sample and antigen were surveyed.

3.2.5 Group comparison

Four patients were excluded from the Covid cohort (affecting four samples from phase 1 and one sample from phase 2) due to having tested negative for SARS-CoV-2 during hospital admission and one sample was excluded from the EIMS cohort due to missing data. The remaining samples of the EIMS cohort, consisting of MS patients and healthy controls, were treated as a single, non-Covid group for the sake of simplicity and statistical power.

For each phase of Covid patients and each antigen, the numbers of reactive and nonreactive samples were compared to those in the control group using Fisher's Exact Test, set to evaluate enriched reactivity in the Covid group.

The significance threshold was set to $p < 0.05$.

4 Results

4.1 Planar microarray

The four arrays represented a total of 235,008 observations, of which 117,730 (50.1%) remained after filtering. The effects of the per-glass normalization are visualized in figure 1. The normalized distributions of significant signals across the glasses are shown in figure 2. The distributions were not found to significantly differ between the glasses.

Of the 235,008 filtered spots, 352 spots (0.30%) representing 332 unique PrESTs showed a reactivity above 4 SDs. The full list of significantly reactive spots is available in appendix D.

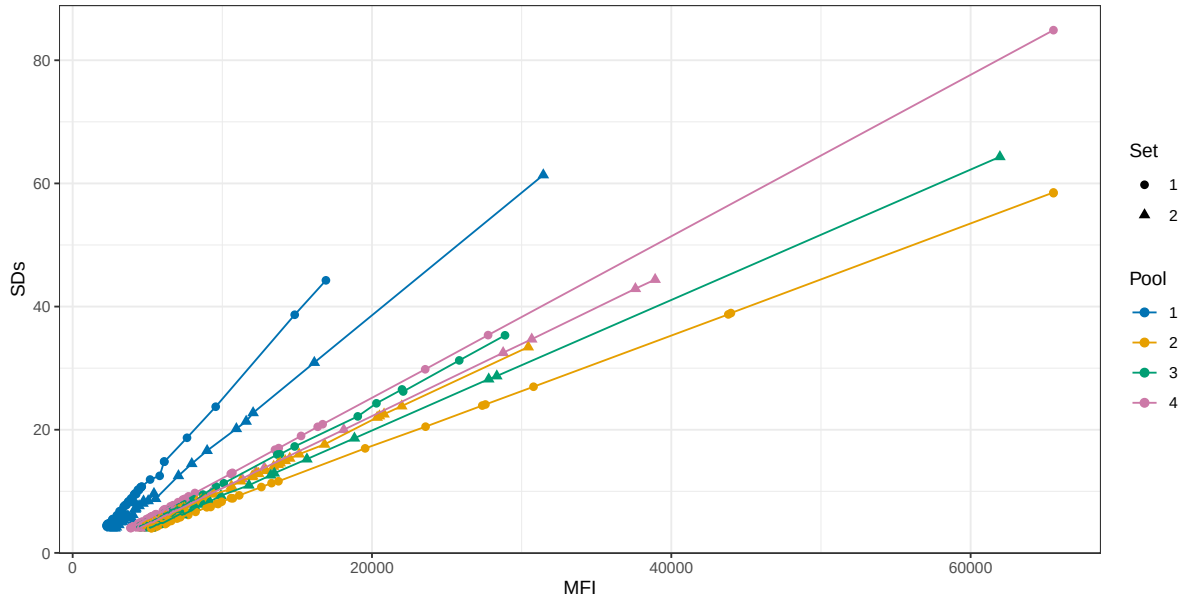


Figure 1: The x-axis shows the foreground pixel median signal intensity (MFI) of the reactive spots and the y-axis shows the corresponding normalized signal (SDs) as explained in equation 2. The points are colored by pool, shaped by slide set and shown connected by both to highlight the different correlations.

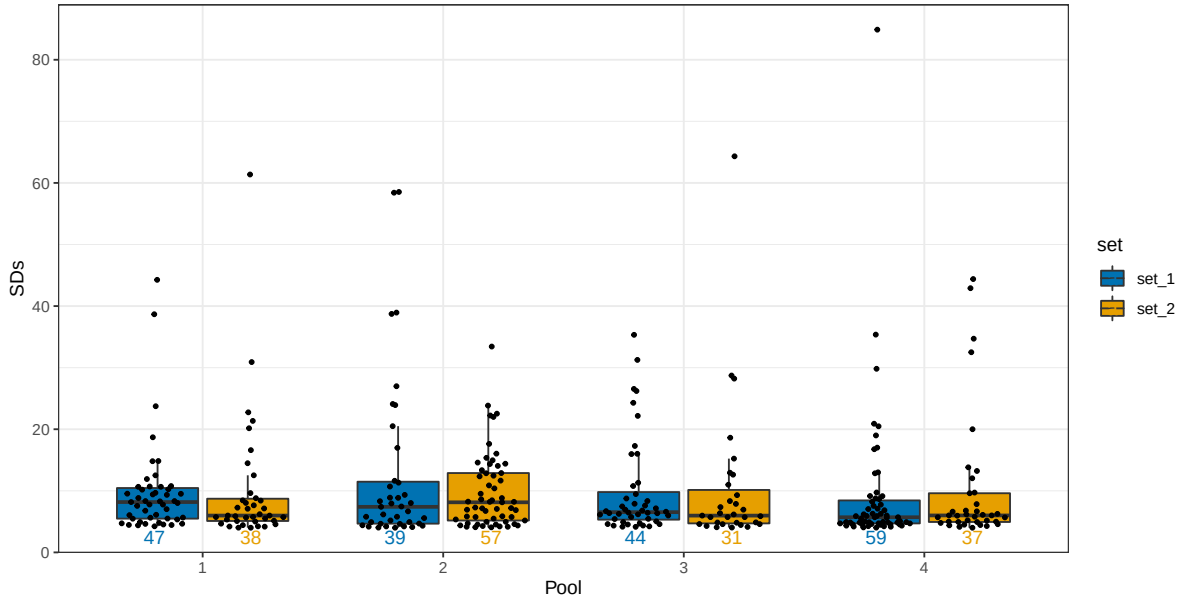


Figure 2: Boxplots and swarm plots showing the normalized signal of the significant observations for each pool and set. The number of observations is shown below each distribution.

4.2 Suspension Bead Array

4.2.1 Pre-run assay tests

The bead count across all bead IDs can be seen in figure 3. Only five beads (IDs 51, 73, 88, 221, 257) had observations with counts below 30 and only one bead (ID 221) had a count median below 30, at 29. The median bead counts per plate were 134, 112, 124 and 114 for plates 1-4 respectively. The plate-wide differences were not estimated to necessitate any volume adjustment to the 384-plex pooling.

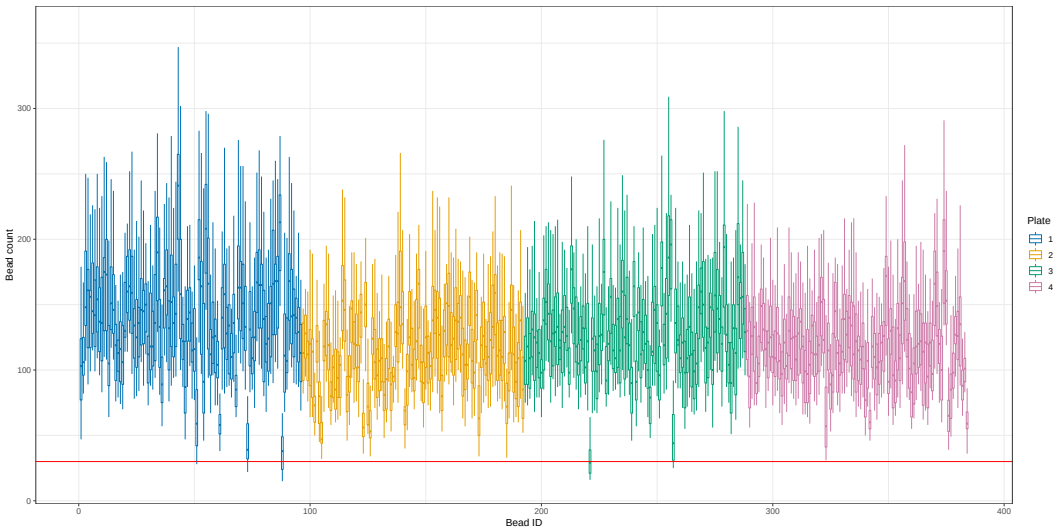


Figure 3: Boxplots consisting of the 21 observations of bead count across all bead IDs, color-separated by plate. The red horizontal line thresholds a bead count of 30.

The results of the coupling test are shown in figure 4. As expected, the His_6ABP beads exhibits a high signal while the remaining control beads and non-PrEST antigen exhibit a low signal. The high variation in signal strength across the PrEST antigen can be due to variations in antigen coupling efficacy, but also due to varying tag accessibility across the PrESTs.

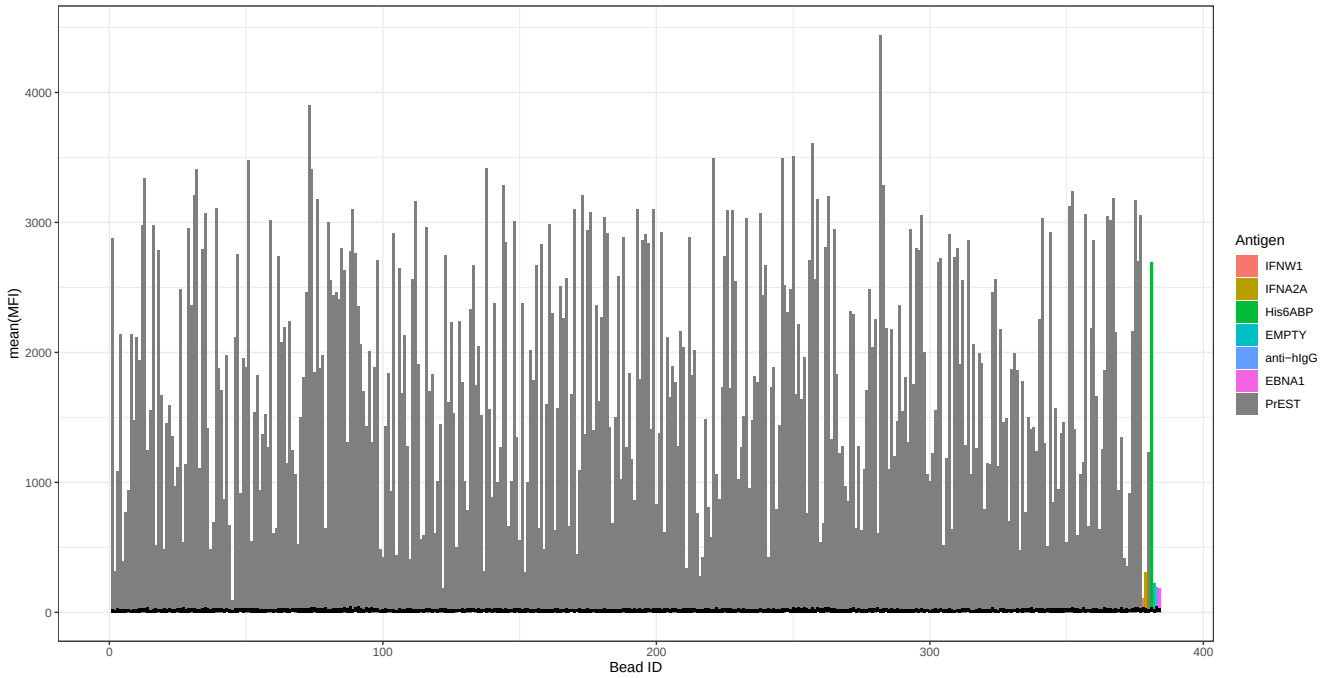


Figure 4: The mean MFI of all bead IDs in the coupling test. The profile of the negative control is plotted in black in the foreground. All control beads as well as antigen without a His_6ABP tag are shown in bright colors.

The results of the sample test are shown in figure 5. The triplicate of commercial plasma has a limited number of reactivities across the antigen and a high reactivity for the anti-hIgG and EBNA1 beads as expected. The negative control is very low across all beads except anti-hIgG, were it amounts to around a third of the signal of the plasma triplicate.

The results of the specific antigen test are shown in figure 6. All HPA antibodies appear to have found their intended target PrESTs. Additionally, the full length protein IFNω appears strongly reactive for the antibody targeting bead 89 (IFNω) and weakly reactive for the antibody targeting bead 32 (various IFNα and IFNω). The full length protein IFNα2A exhibits a very slight reactivity for the antibody targeting beads 33 (various IFNα) and is near indiscernible to the background for bead 32.

An unwanted, relatively low but consistent signal for anti-hIgG is seen across the negative controls of the sample test and specific antigen test which may be due to unspecific binding of the detection antibodies or contamination. All-in-all, the test results were seen as promising and no alterations were made to the array.

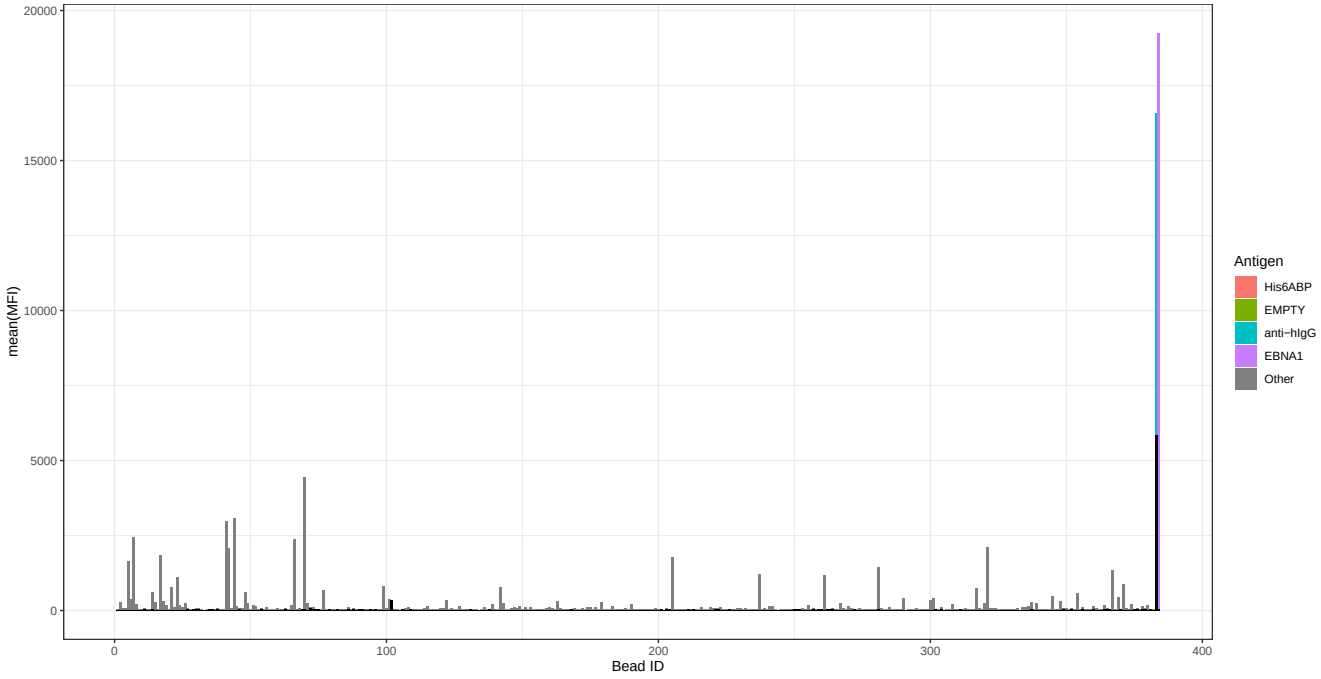


Figure 5: The mean MFI of all bead IDs in the sample test. The profile of the negative control is plotted in black in the foreground. All control beads are shown in bright colors.

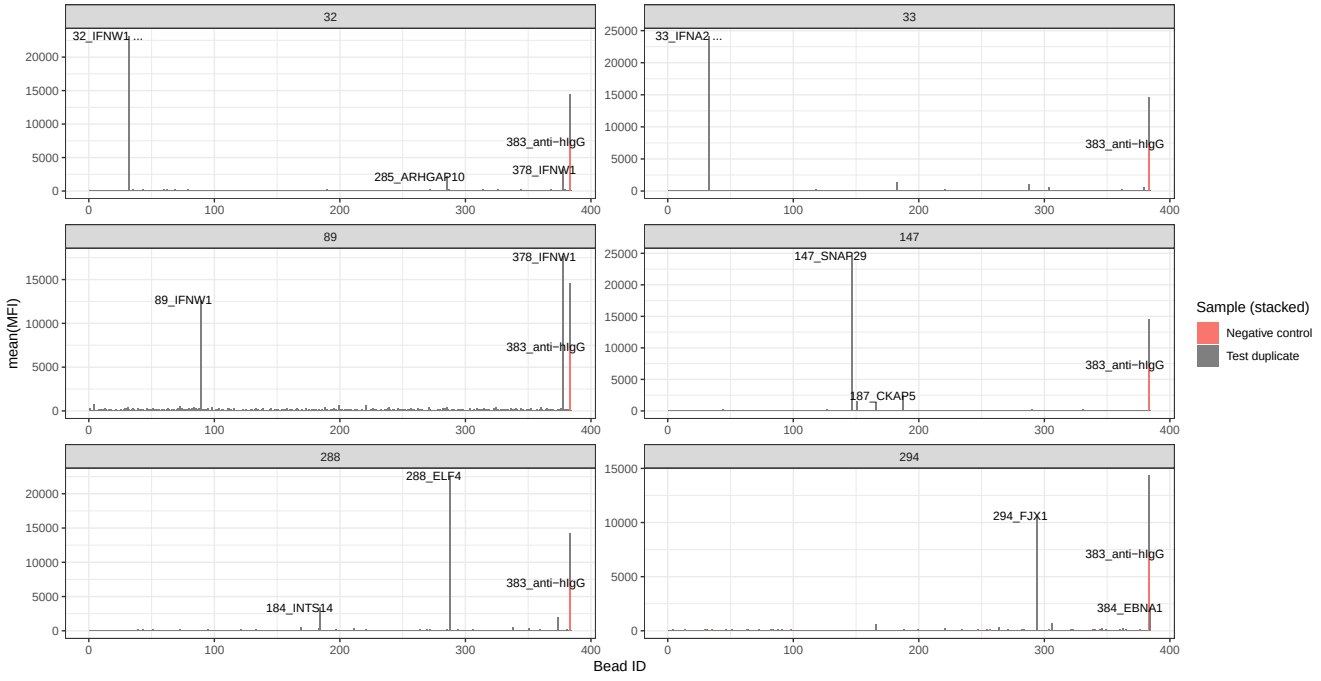


Figure 6: The mean MFI of all bead IDs, plotted for each of the six specific antigen tests, denoted by target bead ID. Outlier signals are captioned by bead ID and gene name. The columns of the negative control and antibody sample duplicates are stacked and shown in red and grey respectively.

4.2.2 Quality control

As can be seen in figure 7, the anti-hIgG bead indicates the binding of human IgG at uniform intensity across all non-empty wells. The EBNA1 bead exhibits a highly varying range of binding across all non-empty wells, but is more uniform for the commercial plasma and slightly lower in range for the Covid samples of phase 2 and 3. Both the empty bead and the His_6ABP appear inert across all wells.

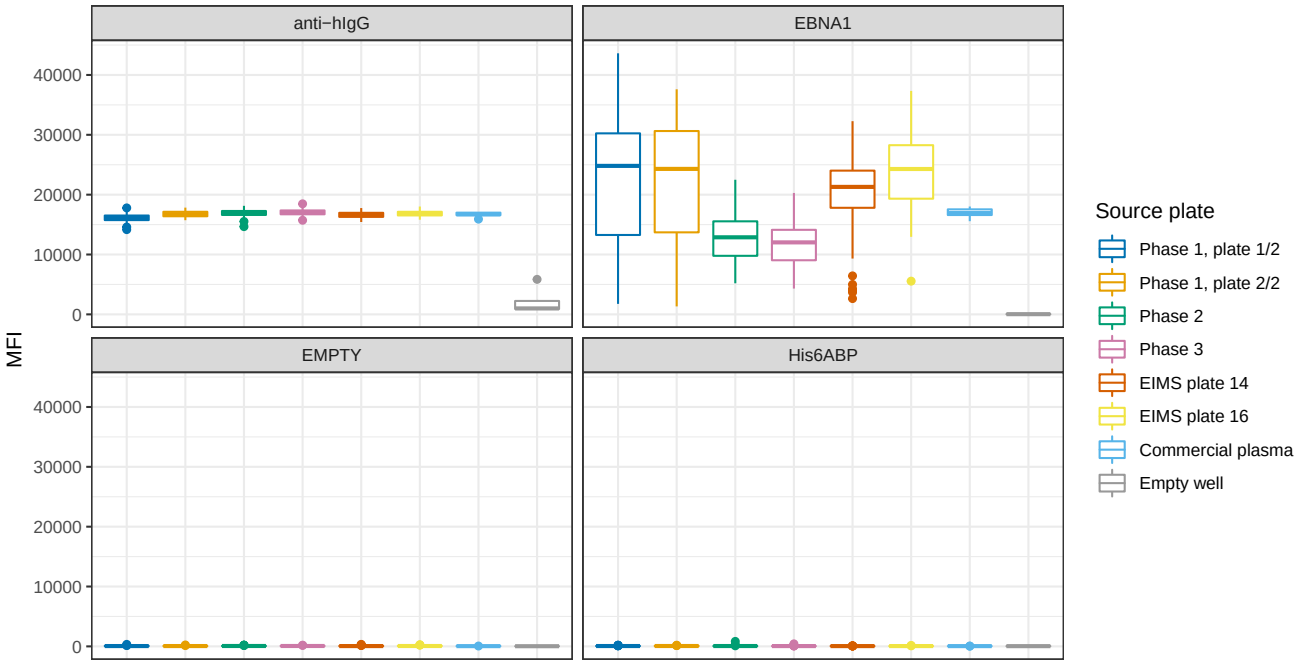


Figure 7: For all control beads (described in section 2.2.1), the MFI is shown as a boxplot for each source plate as well as the commersial plasma controls and empty wells.

4.2.3 Normal- and binarization

In figure 8, histograms of the number of reactivities per sample sample and antigen are plotted to provide an overview of the strictness of the set threshold. With medians of 12 reactive antigens per sample and 11 reactive samples per antigen, the chosen threshold does not contradict the assumption of the normalization (see 3.2.3).

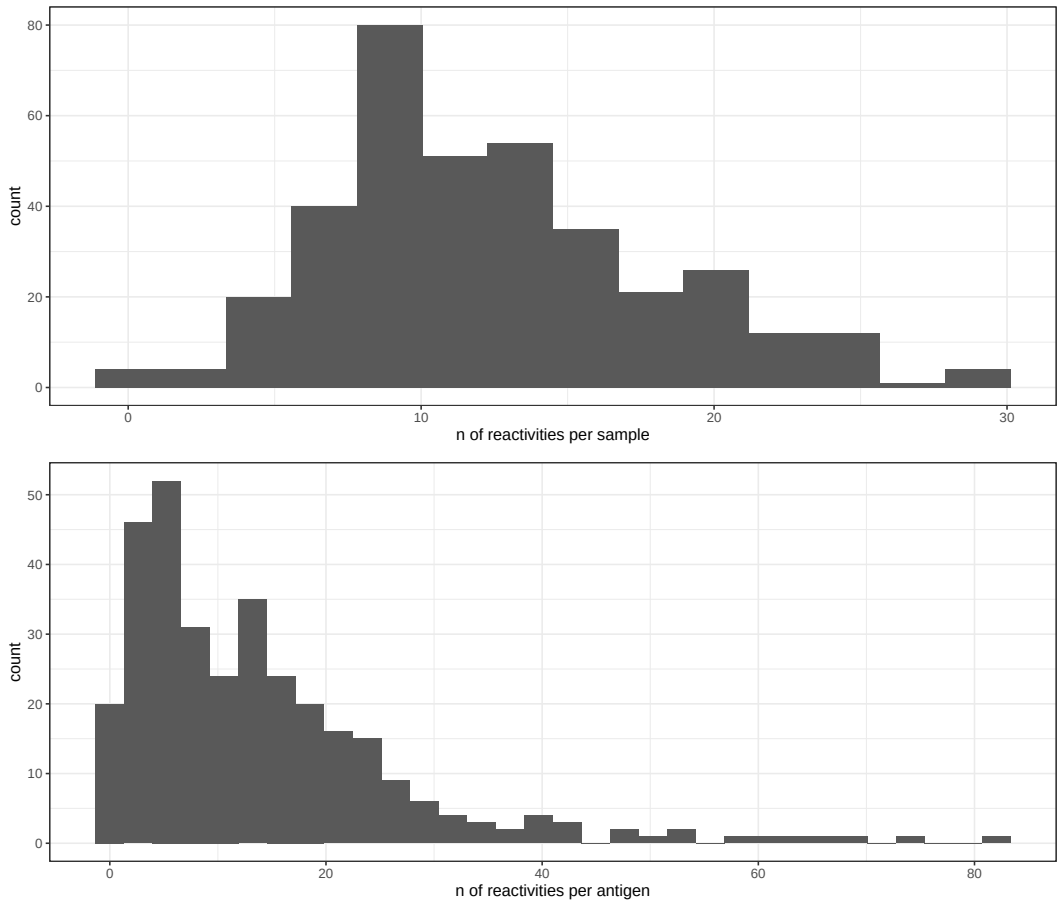


Figure 8: Histograms showing the number of called reactivities per sample (top) and per antigen (bottom).

Figure 9 shows the impact of normalizing each observation by the background of both it's sample and antigen. It is apparent that the transformation favors observations with relatively low backgrounds. For example, the PrEST with bead ID 5 representing the gene product RIN3 has a very high background compared to other antigens, but after transformation only a single observation is called as reactive. In contrast, the PrEST with bead ID 3 representing the gene product IL6 is appears to have it's signal range amplified in relation to the other antigens, due to low sample- and antigen background. It is worth noting that observations called as "reactive" are not claimed to possess any particular affinity or chemical property, merely that they are outliers in the normalized data. The threshold of 1.5 $MADs^{S,Ag}$ is above any observation in the negative control beads.

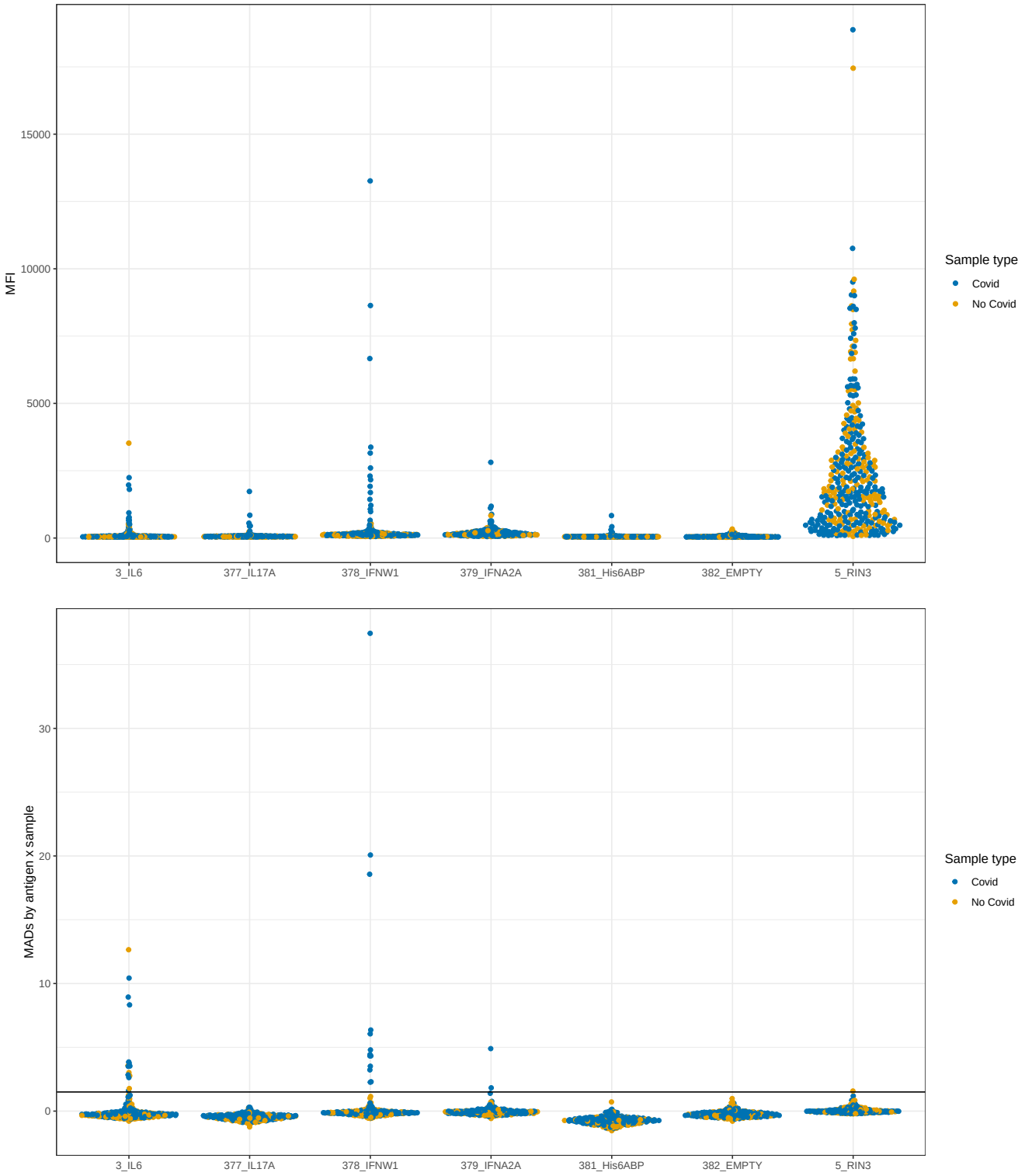


Figure 9: Normalization of some example antigens. The bead ID and antigen name are shown sequentially on the x-axis. On the y-axis, the signal is shown in MFI (top) and normalized against sample and antigen backgrounds as described in equation 6 (bottom). In the bottom plot, the horizontal line represents a cut-off of 1.5 $MADs^{S,Ag}$.

4.2.4 Group comparison

The complete results including p-values and contingency table values for each antigen and phase can be found in appendix C.

In total, 34 antigen were called as differentially reactive in at least one phase. In the comparison between the Covid patients of phase 1 (n = 114) and the EIMS controls (n = 138), 21 antigens were found to be differentially reactive ($p < 0.05$). From phase 2 (n = 59) and 3 (n = 51), a total of 13 antigens were found to be differentially reactive but only 2 of these had seen a significant increase in the number of reactive patients.

To provide some structure and prioritization among the results, the antigens were divided into three categories:

- **Excluded (n = 11)** Because the Covid patient samples of phase 2 and 3 are sub-cohorts of phase 1, their sample sizes are smaller. This means significance may be lost in phase 2-3 because the reactive patients did not participate in the follow up. It also means significance may be gained in phase 2 or 3 simply because the sample size has decreased, affecting the proportion of reactive patients. For this reason, antigens which reached significance in phase 2 or 3 but had not increased considerably in number of reactive patients since phase 1 were excluded from the results and are not discussed further.
- **Relevant (n = 8)** Antigens with immunoregulatory properties and/or particular perceived relevance to Covid pathogenesis, based on literature and gene ontology.
- **Various (n = 15)** Remaining antigens, which may still be of importance but will not be prioritized in the results or discussion to limit the scope of this project.

In figure 10, all observations of the antigens classed as relevant are visualized by antigen, phase, signal, reactivity and subject. The same plots can be found for the remaining various antigens in appendix C.2. Note that the signal ranges differ between antigens, but that they are scaled individually to better visualize their distributions.

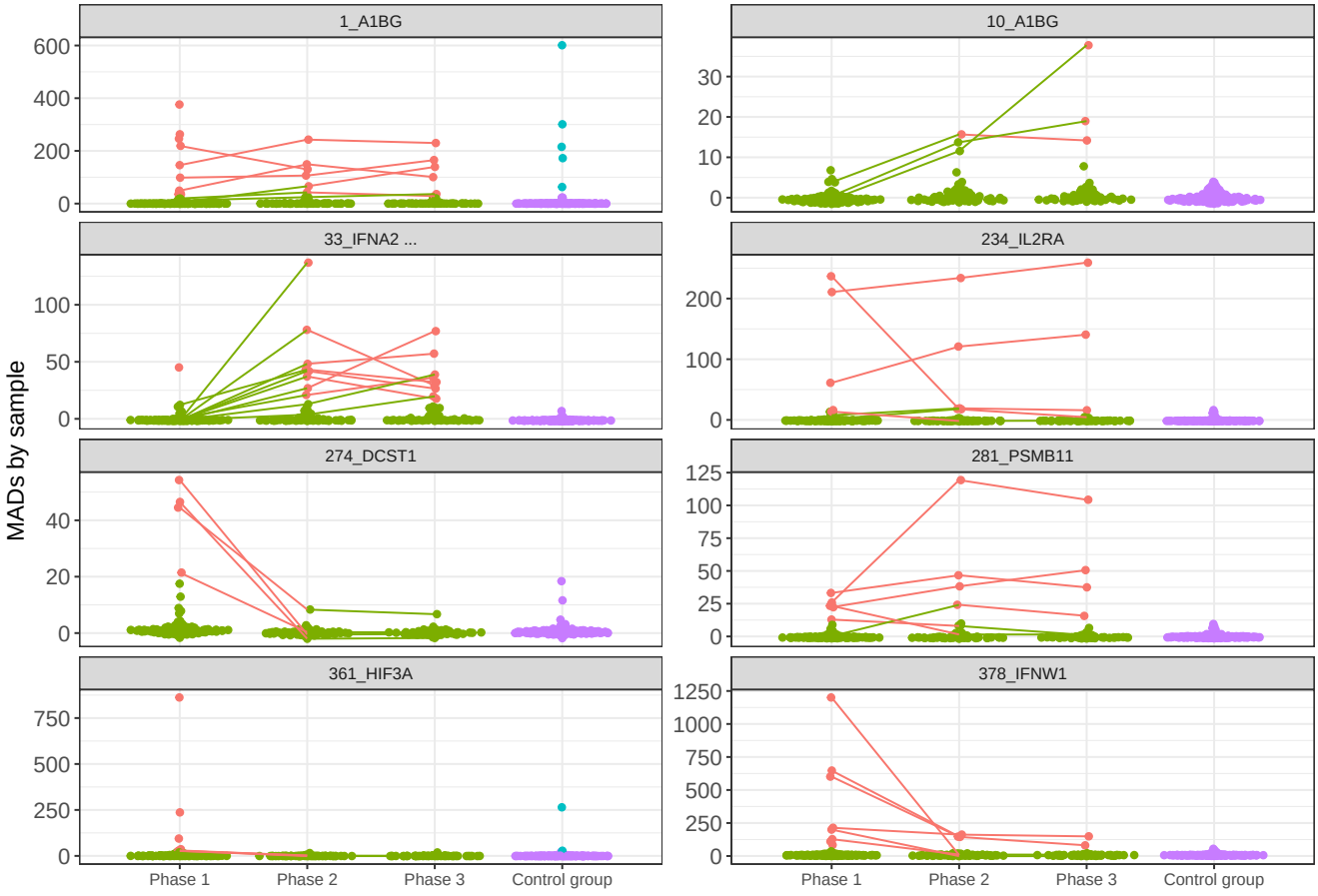


Figure 10: Swarm plots of all differentially reactive antigens classed as relevant. The x-axis is divided into all three phases of the Covid cohort as well as the control group. The y-axis shows the sample-normalized signal ($MADs^S$) scaled per antigen. Reactive and non-reactive observations are colored for the Covid patients (red, green) and control samples (blue, purple) respectively. Observations of patients who were reactive in at least one phase are interconnected, to visualize which patients did not participate in follow-up and how the reactivities of the same individuals change longitudinally.

The results of these antigens are briefly summarized below:

- Beads 1 and 10 both represent different epitopes of Alpha-1B-glycoprotein. Immunologically, the protein is involved in neutrophil degranulation which plays part in Covid pathogenesis and lung tissue damage [28]. In phase 1, bead 1 is reactive in 13/114 Covid patients compared to 5/138 controls and the number of reactivities decreases in phase 2-3 due to patient fall-off. Bead 10 on the other hand increases longitudinally in 3 patients compared to no controls. The PrEST of bead 1 is "sticky" with a high background and found reactive in 31% of in-house runs while that of bead 10 is not.
- For most antigens, the number of reactive patients generally decreases in phases 2-3. The most notable exception is bead 33 representing 12 subtypes of IFNα which increases from 1 to 9 reactive patients longitudinally (1 to 11 when including reactive patient fall-off). This corresponds to 17.6% of phase 3 or 7.9% of phase 1 patients exhibiting reactive levels of anti-type I IFN aAbs 8 months after infection, compared to none in the control group.
- Bead 234 represents the Interleukin-2 receptor subunit alpha which is reactive in the same 5 Covid patients across phases 1-2, after which 2 drop below the reactivity threshold in phase 3. The interleukin-2 receptor enables T-cell differentiation into regulatory T-cells, as well as effector T-cells and memory T-cells, upon antigen binding, to combat infections [29].
- Bead 274 representing E3 ubiquitin-protein ligase is reactive in 4 Covid patients in phase 1 of which all drop below the reactivity threshold in phase 2-3. The protein is a negative regulator of type I interferon-mediated signaling and plays part in the innate immune response [26].
- Bead 281 representing Proteasome subunit beta type-11 is reactive in 5 Covid patients in phase 1 after which two drop and another is gained in phase 2-3. The protein is a part of the proteasome which is responsible for antigen processing and presentation via MHC class I during viral infection [26].
- Bead 361 representing Hypoxia-inducible factor 3-alpha is called reactive in 8 Covid patients (although only 3 are visible above the bulk) in phase 1 compared to 2 controls, after which one patient goes below the reactivity threshold and the remaining seven fall off in phase 2-3. Some called reactivities here are admittedly close to the background, but one Covid patient exhibits a very high signal in phase 1. The protein is a transcription factor which regulates the transcriptional response to low oxygen tension [30], a hallmark symptom and pathological element in Covid-19.
- The full length protein IFNω has the 5^{th} highest signal range of any differentially reactive antigen and the highest among the antigens classed as relevant. In phase 1, 8/114 patients (7%) were classed as reactive compared to none in the control group, after which 3 and 2 remain reactive in phase 2 and 3 respectively.

All of the relevant antigens except HIF3A have an immunological or immunoregulatory function and all except PSMB11 are extracellularly accessible. Of the 15 various antigens, 2 are annotated to have minor immunological functions and 7 are extracellularly accessible (see appendix C).

Among the Covid patients who had participated in both phase 1 and 2 (n = 59), 194 out of 687 (28.2%) observed reactivities decreased past the reactivity threshold in phase 2. In comparison, only 334 out of 21,662 (1.5%) of non-reactive observations did the opposite transition of increasing past the reactivity threshold.

4.2.5 Type I Interferons

In the interest of comparing the results to the findings of of Bastard *et al*, it is worth mentioning the outcome of the beads representing type I IFNs. There were 11 such beads included in the panel: 2 were called as differentially reactive (mentioned above) and another 4 had at least one Covid patient classed as reactive in some phase. They are shown in figure 11. For each of these antigen, including the full length IFNα2A, only a few patients were reactive, albeit often with clear and consistent signal across the phases. The remaining 5 beads (ID 31,73,89,141,239) were not classed as reactive in any observation.

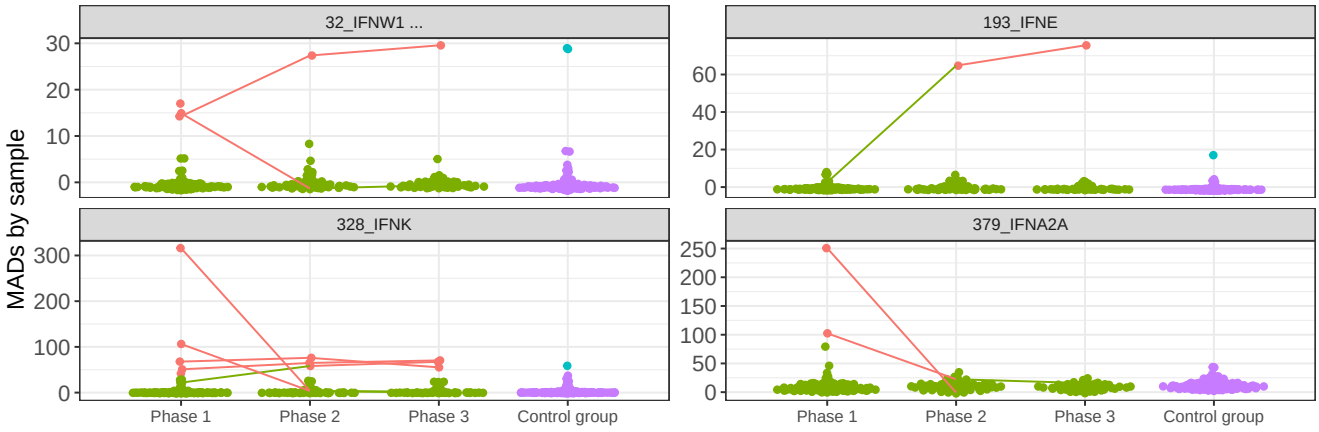


Figure 11: Swarm plots of beads representing type I IFNs which were reactive in at least one phase but not classed as differentially reactive. The x-axis is divided into all three phases of the Covid cohort as well as the control group. The y-axis shows the sample-normalized signal ($MADs^S$) scaled per antigen. Reactive and non-reactive observations are colored for the Covid patients (red, green) and control samples (blue, purple) respectively. Observations of patients who were reactive in at least one phase are interconnected, to visualize which patients did not participate in follow-up and how the reactivities of the same individuals change longitudinally.

5 Discussion

5.1 Findings

5.1.1 Candidate autoantigens

In total, 23 candidate autoantigens were called as differentially reactive in Covid patients, of which 8 were identified as biologically relevant to immunoregulation and/or Covid pathogenesis.

Due to the low sample size and exploratory nature of the project, no definite conclusions can be drawn solely based on these results. However, the results can be put in relation to previous findings and may be useful for future investigations of the interplay between Covid-19 and autoimmunity phenomena.

5.1.2 Type I interferons and previous findings

Based on the findings of Bastard *et al*, autoantibodies directed towards IFNα2 and/or IFNω were expected in 10% of hospitalized patients (n = 101/987). In this project, 8/114 (7%) of hospitalized patients were reactive to full length IFNω which is arguably well in line with the findings. Surprisingly however, only 2/114 (1.8%) were reactive to the full length IFNα which is described as the more common reactivity of the two.

Out of the four PrESTs representing epitopes of IFNα/ω, two (beads 31 and 89) were not called as reactive at any point and one (bead 32) was called as reactive in 3 Covid patients compared to 1 control. Lastly, bead 33 representing a common type I interferon epitope exhibited the single highest significance of any antigen at 9/51 (17.6%) reactive patients in phase 3 compared to no controls. As shown among the Covid patients, reactivities are much more commonly lost than gained after phase 1, which may indicate an elevated general presence of antibodies during hospitalization which later rescinds to baseline. This trend is contradicted by bead 33: out of all differentially reactive antigens, it is the only one to consistently increase in number of reactive patients throughout the 8 month period.

Whether this increase of autoantibodies is novel and contracted due to the Covid infection is not known as there is no data regarding the autoantibody levels prior to infection. It can be speculated that the autoantibody response of these patients were present before and during phase 1, but that the circulating autoantibodies were depleted by high levels of antigen *in vivo* during the course of the disease.

Some reactivities were observed in the Covid patients for PrESTs representing other type I interferons such as IFNε and IFNκ, yet the significance of the findings are limited primarily by sample size.

5.1.3 Differences between PrESTs and full length proteins

The inclusion of the full length proteins in the SBA offers the opportunity to compare their reactivity to that of the PrESTs. Several beads representing IFNα/ω were called as completely inert, and most were less reactive than the full length IFNω. This indicates that the PrESTs on these beads do not fully present the antigenic epitopes of the full length protein, possibly due to the epitopes in question being conformational in nature or simply situated in a sequence-part of the protein which is not represented. Intuitively, it makes sense that the reactivities of the PrESTs would constitute a subset of the reactivities towards the full length proteins, because they exhibit a subset of the epitopes.

On the other hand, bead 33 demonstrated both a higher number of reactivities and a completely inverted longitudinal profile compared to both full length proteins. In this case, it could be speculated that the PrEST does present a reactive epitope as well as being more accessible to binding due to a smaller size and an increased conformational freedom.

As was apparent in the specific antigen test and figure 6, antibodies which were specifically derived from PrESTs may not bind the full length protein strongly. The HPA antibody of bead 89 representing IFNω did bind to both the PrEST and the full length protein, yet the bead was called as completely inert to the antibodies of the samples, that instead bound other representations of IFNω. This suggests that even epitopes that are accessible on the full length proteins are not necessarily immunogenic.

All in all, the results of this project indicate that PrESTs and full length proteins can be expected to differ in reactivity. The PrESTs may be less reactive than their mimicked proteins due to not being fully representative. On the other hand, using many PrESTs representing the same protein may provide a multi-faceted approach in which binding is evaluated for different epitope representations of different conformational freedoms which may partially compensate for the artifacts arising in an *in vitro* approach. A combination approach of including both full-length proteins and their PrESTs may be beneficial in combining the advantages of both.

5.2 Limitations

5.2.1 The challenges of autoimmunity profiling

Autoimmunity profiles are inherently diverse and scattered, and with a broad enough antigen panel even clinically healthy samples will exhibit some activity [5]. Because even significant autoreactivities may only be present in a relatively small proportion of patients, it may be difficult to determine overlaps and significant enrichments even in a sizeable cohort.

A further complication is that free antibody levels in the plasma may fluctuate over time depending on the state of the immune system and the availability of target antigen. In the results there are several examples of reactivities decreasing or increasing, as well as remaining stable, over time. Another challenging aspect is that these assays measure free antibodies, entailing that antibodies may be "depleted" to a varying extent depending on the level of target antigen in the plasma.

5.2.2 Sample quality and quantity

The most limiting aspect of this project is arguably the sample size. When measuring autoimmunity present in less than 10% of hospitalized Covid patients, even hundreds of samples may not provide statistically significant results. The relative rarity of autoimmune features also makes it difficult to make statistically sound investigations within the feature group, e.g. looking at the impact of sex and age. These parameters were also only available for the EIMS controls and the Covid patients of phase 2 and were not largely utilized in the analysis.

When looking for autoimmunity features impacting the severity of the disease in hospitalized patients, the direct comparison to mild Covid cases would have been more powerful than using a control group consisting of a combination of MS-patients and healthy controls with no known relation to the disease. The usage of the EIMS control group was primarily motivated by limited access to other samples. However, the frequency of Covid cases requiring hospitalization in the population at large is low and therefore the majority of the controls should arguably represent mild cases.

5.2.3 Panel composition bias

Out of the 380 antigen included in the panel, 353 are based on the planar microarray analysis of 16 patients and 27 are based on literature. Basing such a large fraction of the panel on such a small fraction of patients is optimistic,

especially for autoimmunity profiling. This is because not all antigens which may be differentially reactive at group level can be expected to be captured in such a small sample size. It is important to clarify however, that the "pilot" group of 16 patients never was expected to fully and comprehensibly represent all Covid patients, but rather provide a list of possible autoantigens to further investigate at a group level. The pool-wise stratification by disease severity, sex and comorbidity of the pilot group was not largely utilized in the rest of the project, but the justification of minimizing comorbidities (see section 2.1) still applies.

The differentially reactive, non-excluded antigens (n = 23) consisted of 21 (91.3%) microarray antigens and 2 (8.7%) literature antigens compared to the panel at large (n = 380) which consisted of 353 (92.9%) microarray antigens and 27/380 (7.1%) literature antigens. In summary, antigens which were specifically motivated as relevant in literature were slightly enriched in the results.

5.2.4 Cut-offs

There are two important cut-offs set in the data analysis of this project and both largely affect the results.

The first is the cut-off for binarizing reactivity, set at 1.5 $MADs^{S,Ag}$. Normalizing all observations by both sample and antigen background in order to set a universal cut-off value is applicable for an assay containing a large number of samples and antigen, where the majority of both are nonreactive. Another, clinical approach is to use a negative control group to define the range at which a signal is deemed non-reactive for a particular antigen. This presumes certainty that the control group is wholly nonreactive towards the antigen in question, which is not applicable in this broad and discovery-based project. A more complex, data-driven approach is to analyze each antigen distribution separately with respect to the impact of different cut-offs on comparisons across the groups. The tools for such an approach were available, but were not applied in order to prioritize self reliant work and learning under a time constraint.

The method chosen for this project is justified in that it takes both sample and antigen backgrounds into consideration while still performing a relatively simple transformation. The threshold is fairly generous, set slightly above the distribution of the empty beads, but this generosity can be justified in that the end purpose is to compare the reactivity between the groups, rather than the reactivity within the groups. E.g. as long as the cut-off is equally generous for both groups in the comparison, the impact is small.

The second cut-off is for the p-value of Fisher's Exact Test, under which a finding is called significant. Statistically speaking, the SBA assay investigates 380 antigens in conjunction, corresponding to an equal number of alternative hypotheses to be tested. It can be argued that this warrants multiple hypothesis correction to compensate for the increased chance of erroneous significance calls. However, it can be argued that multiple hypothesis approaches, such as Bonferroni correction and False Discovery Rate (FDR), are more appropriate for quantitative comparisons in which the p-values fall in a mostly uniform distribution. Due to the contingency basis of Fisher's Exact Test (which in itself is a direct result of the binarization cut-off) and the relatively low sample size, the p-value distribution does not follow the assumption of the mostly uniform distribution on which many correction methods are based. Furthermore, the purpose of the significance cut-off in this exploratory project is not to draw statistically ensured conclusions but rather to arrange the results in order of interest.

6 Conclusion and future perspectives

This project proposes several candidate autoantigens for which autoantibodies were found to be enriched in a subset of patients afflicted with severe Covid in a Swedish cohort. Out of the 23 differentially reactive antigens, 8 are motivated as biologically relevant to immunoregulation or Covid pathogenesis. Previous findings of cytokine-directed autoimmunity in Covid patients [7] were replicated for full length IFNω as well as a PrEST representing a common type I interferon epitope, with the former peaking in reactivity during the hospital admission and the latter increasing in reactivity over an 8 month period.

The results of this project further advance the notion of autoimmunity as an underlying risk factor of severe Covid and invite further investigation. As a suggestion, similar inquires may, in the future, benefit from a larger sample size and a qualitatively similar control group (discussed in section 5.2.2). It is also worth noting that the scope of this project has been limited to detecting autoimmunity directed to proteins and protein fragments, while autoantibodies can be directed to other classes of biomolecules as well, such as phospholipids.

The variation of reactivity seen across the full length proteins and their respective PrESTs (discussed in section 5.1.3) indicate that neither approach captures all reactivities seen in the other, suggesting that a combination of full-length proteins and PrESTs may be a good approach for future queries relating to cytokine directed autoimmunity profiling. The differential reactivity across the PrESTs may also be of use in an endeavor to map the reactive epitopes of the proteins.

The data generated during the course of this project is far from exhausted and may be utilized further. For instance, no work has been done to investigate the co-occurrence and longitudinal changes of multiple reactivities in the same patient; neither has patient data of age and sex been incorporated in the analysis, primarily due to low numbers of reactive samples for any antigen. Studying in detail the longitudinal autoimmunity profile of a single patient may be useful to, for example, a treating physician.

An as of yet not peer reviewed study indicate the presence of a large variety of possibly pathogenic autoantibodies in a considerable fraction (>50%) of Covid patients [31]. There is also ongoing research pertaining to explaining the phenomenon of "Long Covid" as an acquired autoimmune disease [32, 33].

As of now, the pandemic is hopefully soon at an end due to the ongoing mass vaccinations. However, the autoimmunity phenomena described herein may be relevant to post-infectious complications or the autoimmunity aspects of other infectious diseases and contribute to a broader understanding of the factors governing disease susceptibility.

Acknowledgements

I wish to thank the following people for their contributions and support

Anna Månberg,
for her work as supervisor and words of encouragement

Ronald Sjöberg,
for stepping in as co-supervisor at short notice and his continuous support

Jennie Olofsson,
for lab supervision and various sound effects

Cecilia Hellström,
for discussions of protocols and data analysis approaches

August Jernbom Falk,
for early project discussion and data analysis tips

Harry Soar,
for proofreading and nitpicking

Sebastian Havervall,
for help in surveying the clinical data

Charlotte Thålin,
for access to the samples and data of the Community study

Peter Nilsson,
for allocating time and resources to the project and to my learning

All of PAPP,
for the warm welcome

References

[1] Weekly Operational Update on COVID-19, Issue No. 56. Word Health Organization; 2021. Available from: https://www.who.int/publications/m/item/weekly-operational-update-on-covid-19---24-may-2021.

[2] Folkhälsomyndigheten. Bekräftade fall i Sverige – daglig uppdatering; 2021. Available from: https://www.folkhalsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19/statistik-och-analyser/bekraftade-fall-i-sverige/.

[3] Beck DB, Aksentijevich I. Susceptibility to severe COVID-19. Science. 2020 Oct;370(6515):404–405. Publisher: American Association for the Advancement of Science Section: Perspective. Available from: https://science.sciencemag.org/content/370/6515/404.

[4] Mohamed Khosroshahi L, Rokni M, Mokhtari T, Noorbakhsh F. Immunology, immunopathogenesis and immunotherapeutics of COVID-19; an overview. International Immunopharmacology. 2021 Apr;93:107364. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7784533/.

[5] Neiman M, Hellström C, Just D, Mattsson C, Fagerberg L, Schuppe-Koistinen I, et al. Individual and stable autoantibody repertoires in healthy individuals. Autoimmunity. 2019 Jan;52(1):1–11. Available from: https://www.tandfonline.com/doi/full/10.1080/08916934.2019.1581774.

[6] De Andrea M, Ravera R, Gioia D, Gariglio M, Landolfo S. The interferon system: an overview. European Journal of Paediatric Neurology. 2002 May;6:A41–A46. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1090379802905738.

[7] Bastard P, Rosen LB, Zhang Q, Michailidis E, Hoffmann HH, Zhang Y, et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. Science. 2020 Oct;370(6515). Publisher: American Association for the Advancement of Science Section: Research Article. Available from: https://science.sciencemag.org/content/370/6515/eabd4585.

[8] Zhang Q, Bastard P, Liu Z, Pen JL, Moncada-Velez M, Chen J, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science. 2020 Oct;370(6515). Publisher: American Association for the Advancement of Science Section: Research Article. Available from: https://science.sciencemag.org/content/370/6515/eabd4570.

[9] Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015 Jan;347(6220):1260419–1260419. Available from: https://www.sciencemag.org/lookup/doi/10.1126/science.1260419.

[10] Berglund L, Björling E, Jonasson K, Rockberg J, Fagerberg L, Szigyarto CAK, et al. A whole-genome bioinformatics approach to selection of antigens for systematic antibody generation. PROTEOMICS. 2008;8(14):2832–2839. _eprint: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/pdf/10.1002/pmic.200800203. Available from: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/pmic.200800203.

[11] Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, et al. Ensembl 2021. Nucleic Acids Research. 2021 Jan;49(D1):D884–D891. Available from: https://academic.oup.com/nar/article/49/D1/D884/5952199.

[12] Sjöberg R, Mattsson C, Andersson E, Hellström C, Uhlen M, Schwenk JM, et al. Exploration of high-density protein microarrays for antibody validation and autoimmunity profiling. New Biotechnology. 2016 Sep;33(5):582–592. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1871678415001545.

[13] Stephen Angeloni,, Robert Cordes, Sherry Dunbar,, Carlos Garcia,, Grant Gibson,, Charles Martin,, Valerie Stone. xMAP Cookbook, A collection of methods and protocols for developing multiplex assays with xMAP technology. Luminex; 2013. BR_574.01_0613. Available from: https://cdn2.hubspot.net/hub/128032/file-213083097-pdf/Luminex-xMap_Cookbook.pdf.

[14] Rudberg AS, Havervall S, Månberg A, Jernbom Falk A, Aguilera K, Ng H, et al. SARS-CoV-2 exposure, symptoms and seroprevalence in healthcare workers in Sweden. Nature Communications. 2020 Dec;11(1):5064. Available from: http://www.nature.com/articles/s41467-020-18848-0.

[15] Havervall S, Falk AJ, Klingström J, Ng H, Greilert-Norin N, Gabrielsson L, et al. SARS-CoV-2 induces a durable and antigen specific humoral immunity after asymptomatic to mild COVID-19 infection. Infectious Diseases (except HIV/AIDS); 2021. Available from: http://medrxiv.org/lookup/doi/10.1101/2021.01.03.21249162.

[16] Döffinger R, Helbert MR, Barcenas-Morales G, Yang K, Dupuis S, Ceron-Gutierrez L, et al. Autoantibodies to Interferon-γ in a Patient with Selective Susceptibility to Mycobacterial Infection and Organ-Specific Autoimmunity. Clinical Infectious Diseases. 2004 01;38(1):e10–e14. Available from: https://doi.org/10.1086/380453.

[17] Mojtabavi H, Saghazadeh A, Rezaei N. Interleukin-6 and severe COVID-19: a systematic review and meta-analysis. European Cytokine Network. 2020 Jun;31(2):44–49. Available from: http://www.john-libbey-eurotext.fr/medline.md?doi=10.1684/ecn.2020.0448.

[18] Puel A, Döffinger R, Natividad A, Chrabieh M, Barcenas-Morales G, Picard C, et al. Autoantibodies against IL-17A, IL-17F, and IL-22 in patients with chronic mucocutaneous candidiasis and autoimmune polyendocrine syndrome type I. Journal of Experimental Medicine. 2010 Feb;207(2):291–297. Available from: https://rupress.org/jem/article/207/2/291/40841/Autoantibodies-against-IL17A-IL17F-and-IL22-in.

[19] Ku CL, Chi CY, von Bernuth H, Doffinger R. Autoantibodies against cytokines: phenocopies of primary immunodeficiencies? Human Genetics. 2020 Jun;139(6-7):783–794. Available from: https://link.springer.com/10.1007/s00439-020-02180-0.

[20] Pin E, Sjöberg R, Andersson E, Hellström C, Olofsson J, Jernbom Falk A, et al. Array-Based Profiling of Proteins and Autoantibody Repertoires in CSF. In: Santamaría E, Fernández-Irigoyen J, editors. Cerebrospinal Fluid (CSF) Proteomics. vol. 2044. New York, NY: Springer New York; 2019. p. 303–318. Series Title: Methods in Molecular Biology. Available from: http://link.springer.com/10.1007/978-1-4939-9706-0_19.

[21] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2017. Available from: https://www.R-project.org/.

[22] Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. Journal of Open Source Software. 2019;4(43):1686.

[23] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015;43(7):e47.

[24] Clarke E, Sherrill-Mix S. ggbeeswarm: Categorical Scatter (Violin Point) Plots; 2017. R package version 0.6.0. Available from: https://CRAN.R-project.org/package=ggbeeswarm.

[25] Kassambara A. ggpubr: 'ggplot2' Based Publication Ready Plots; 2020. R package version 0.4.0. Available from: https://CRAN.R-project.org/package=ggpubr.

[26] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nature genetics. 2000 May;25(1):25–29. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3037419/.

[27] The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research. 2021 Jan;49(D1):D480–D489. Available from: https://doi.org/10.1093/nar/gkaa1100.

[28] Wang J, Li Q, Yin Y, Zhang Y, Cao Y, Lin X, et al. Excessive Neutrophils and Neutrophil Extracellular Traps in COVID-19. Frontiers in Immunology. 2020;11:2063. Available from: https://www.frontiersin.org/article/10.3389/fimmu.2020.02063.

[29] Liao W, Lin JX, Leonard WJ. IL-2 family cytokines: new insights into the complex roles of IL-2 as a broad regulator of T helper cell differentiation. Current Opinion in Immunology. 2011;23(5):598–604. Special section: Cytokines/Immunogenetics and transplantation. Available from: https://www.sciencedirect.com/science/article/pii/S0952791511001063.

[30] Semenza G. Hypoxia-Inducible Factors in Physiology and Medicine. Cell. 2012;148(3):399–408. Available from: https://www.sciencedirect.com/science/article/pii/S0092867412000876.

[31] Chang SE, Feng A, Meng W, Apostolidis SA, Mack E, Artandi M, et al. New-Onset IgG Autoantibodies in Hospitalized Patients with COVID-19. Allergy and Immunology; 2021. Available from: http://medrxiv.org/lookup/doi/10.1101/2021.01.27.21250559.

[32] The Long Covid Kids study group, Ortona E, Buonsenso D, Carfi A, Malorni W. Long COVID: an estrogen-associated autoimmune disease? Cell Death Discovery. 2021 Jun;7(1):77. Available from: http://www.nature.com/articles/s41420-021-00464-6.

[33] Khamsi R. Rogue antibodies could be driving severe COVID-19. Nature. 2021 Feb;590(7844):29–31. Available from: http://www.nature.com/articles/d41586-021-00149-1.

A Key words used for flagging gene ontology

Biological process

cytokine
interleukin
interferon
tumor necrosis factor
virus
viral
immun
hypoxia
humoral
Fc
T cell
natural killer cell
NK T cell
phagocyte
leukocyte
lymphocyte
neutrophil
complement activation

Cellular component

plasma membrane
blood microparticle
cell surface
extracellular
extrinsic
secretory
membrane attack complex
immune
immuno

B Sample layout

The top six plates represent the samples from the Community study phases 1-3 and the MSC_6 samples used as control. The lower four plates represent the destination assay plates were the colour and numbering is retained from the source plates. Wells which have been switched out to allow even distribution of replicates are shown with white numbering.

DSPATI MIX 1

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42	50	58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12	20	28	36	44	52	60	68	76	84	92
E	5	13	21	29	37	45	53	61	69	77	85	93
F	6	14	22	30	38	46	54	62	70	78	86	94
G	7	15	23	31	39	47	55	63	71	79	87	95
H	8	16	24	32	40	48	56	64	72	80	88	96

DSPATI MIX 2

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42	50	58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12	20	28	36	44	52	60	68	76	84	92
E	5	13	21	29	37	45	53	61	69	77	85	93
F	6	14	22	30	38	46	54	62	70	78	86	94
G	7	15	23	31	39	47	55	63	71	79	87	95
H	8	16	24	32	40	48	56	64	72	80	88	96

Ej använt
Flyttas manuellt

DSPATI 09

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42	50	58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12	20	28	36	44	52	60	68	76	84	92
E	5	13	21	29	37	45	53	61	69	77	85	93
F	6	14	22	30	38	46	54	62	70	78	86	94
G	7	15	23	31	39	47	55	63	71	79	87	95
H	8	16	24	32	40	48	56	64	72	80	88	96

DSPATI 10

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42	50	58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12	20	28	36	44	52	60	68	76	84	92
E	5	13	21	29	37	45	53	61	69	77	85	93
F	6	14	22	30	38	46	54	62	70	78	86	94
G	7	15	23	31	39	47	55	63	71	79	87	95
H	8	16	24	32	40	48	56	64	72	80	88	96

EIMS 14

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42	50	58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12	20	28	36	44	52	60	68	76	84	92
E	5	13	21	29	37	45	53	61	69	77	85	93
F	6	14	22	30	38	46	54	62	70	78	86	94
G	7	15	23	31	39	47	55	63	71	79	87	95
H	8	16	24	32	40	48	56	64	72	80	88	96

EIMS 16

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42	50	58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12	20	28	36	44	52	60	68	76	84	92
E	5	13	21	29	37	45	53	61	69	77	85	93
F	6	14	22	30	38	46	54	62	70	78	86	94
G	7	15	23	31	39	47	55	63	71	79	87	95
H	8	16	24	32	40	48	56	64	72	80	88	96

Platta 1 **DSPATI MIX 1**

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42	50	58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12		28	36	44	52		68	76	84	92
E	5	13	21	29	37	45	53	61	69	77	85	
F	6	14	22	30	38	46	54	62	70	78	86	20
G	7	15	23	31	39	47	55	63	71	79	87	40
H	8	16	24	32		48	56	64	72	80	88	60

Platta 2 **DSPATI MIX 2 + EIMS 16 + DSPATI 09**

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	52	56	48	40	32	24	16	8
B	2	10	18	26	53	55	47	39	31	23	15	7
C	3	11	19		54	54	46	38	30		14	6
D	4	12	20		55	53	45	37	29	21	13	5
E	5	13	21	43	60	52	44	36	28	20	12	4
F	6	14	22	22	59	51		35	27	19	11	3
G	7	15	23	50	58	50	42	34	26	18	10	2
H	8	16	24	51	57	49	41	33	25	17	9	1

DSPATI fas 1
DSPATI fas 2
DSPATI fas 3
EIMS

Replikat
Neg. kontroll

Platta 3 **EIMS 14**

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	9	17	25	33	41	49	57	65	73	81	89
B	2	10	18	26	34	42		58	66	74	82	90
C	3	11	19	27	35	43	51	59	67	75	83	91
D	4	12	20	28	36	44	52	60	68	76	84	92
E	5	13		29	37	45	53	61	69	77	85	21
F	6	14	22	30	38	46	54	62	70	78	86	50
G	7	15	23	31	39	47	55	63	71	79	87	
H	8	16	24	32	40	48	56	64	72	80	88	

Platta 4 **EIMS 16 + DSPATI 10**

	1	2	3	4	5	6	7	8	9	10	11	12
A	96	88	80	72	64	56	48	40	32	24	16	8
B	95	87	79	71	63		47	39	31		15	7
C	94	86	78	70	62	76	46	38	30	22	14	6
D	93	85	77	69	61	23	45	37	29	21	13	5
E	92	84		68	60		44	36	28	20	12	4
F	91	83	75	67	59	51	43	35	27	19	11	3
G	90	82	74	66	58	50	42	34	26	18	10	2
H	89	81	73	65	57	49	41	33	25	17	9	1

C Extended group comparison results

C.1 p-values and contingency values

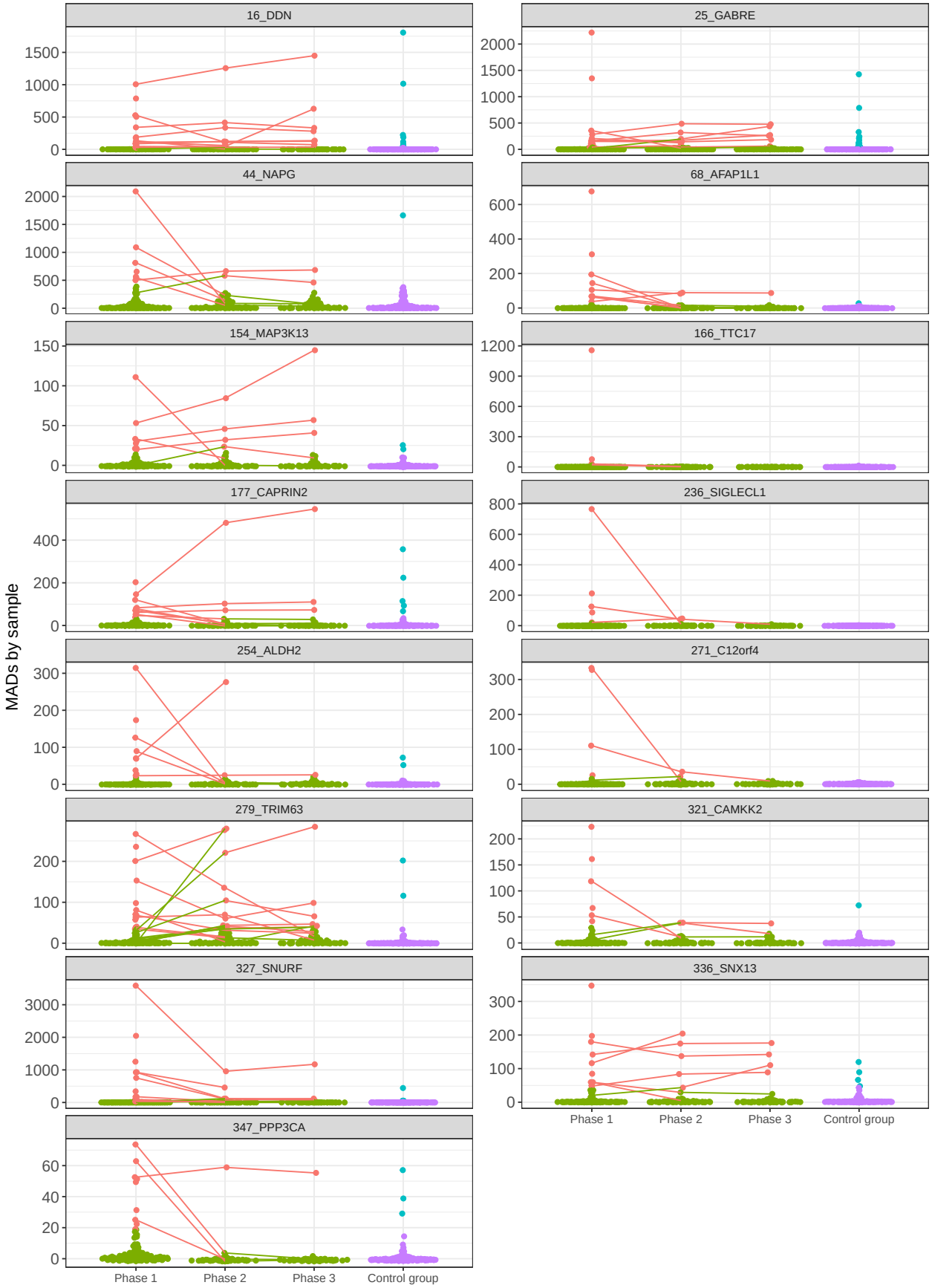
The following table shows all antigen called as significant for the comparison detailed in section 3.2.5. Column p is the p-value obtained from Fisher's Exact Test when comparing the current phase of Covid patients to the EIMS controls, where green indicates a value below 0.0025 and yellow indicates a value below 0.05. The column n/N n/N denotes the number or reactive samples (n) and non-reactive samples (N) among the Covid patients and EIMS controls respectively. The columns kw1 and kw2 represent flagging of immunological function and extracellular localization, respectively, based on the Gene Ontology keywords detailed in appendix A.

The entries are divided in three groups which are explained in section 3.2.5 and explored in section 4.2.4.

Bead	Gene	Phase 1		Phase 2		Phase 3		kw1	kw2	Uniprot
		p	n/N n/n	p	n/N n/n	p	n/N n/n			
Relevant										
378	IFNW1	0,00153	8/106 0/138	0,02591	3/56 0/138	0,07177	2/49 0/138	TRUE	TRUE	P05000
1	A1BG	0,01586	13/101 5/133	0,07180	6/53 5/133	0,04376	6/45 5/133	TRUE	TRUE	P04217
234	IL2RA	0,01804	5/109 0/138	0,00213	5/54 0/138	0,01881	3/48 0/138	TRUE	TRUE	P01589
281	PSMB11	0,01804	5/109 0/138	0,00748	4/55 0/138	0,00485	4/47 0/138	TRUE	FALSE	A5LHX3
361	HIF3A	0,02596	8/106 2/136	1,00000	0/59 2/136	1,00000	0/51 2/136	TRUE	TRUE	Q9Y2N7
274	DCST1	0,04067	4/110 0/138	1,00000	0/59 0/138	1,00000	0/51 0/138	TRUE	TRUE	Q5T197
33	IFNA(various)	0,45238	1/113 0/138	0,00005	8/51 0/138	0,00000	9/42 0/138	TRUE	TRUE	
10	A1BG	1,00000	0/114 0/138	0,29949	1/58 0/138	0,01881	3/48 0/138	TRUE	TRUE	P04217
Various										
279	TRIM63	0,00020	15/99 2/136	0,00001	12/47 2/136	0,00170	7/44 2/136	TRUE	FALSE	Q969Q1
327	SNURF	0,00042	19/95 5/133	0,07180	6/53 5/133	0,09754	5/46 5/133		FALSE	Q9Y675
68	AFAP1L1	0,00405	9/105 1/137	0,21406	2/57 1/137	0,46792	1/50 1/137		FALSE	Q8TED9
254	ALDH2	0,01365	9/105 2/136	0,34694	2/57 2/136	0,61303	1/50 2/136	FALSE	TRUE	P05091
44	NAPG	0,01728	7/107 1/137	0,21406	2/57 1/137	0,17769	2/49 1/137	FALSE	TRUE	Q99747
236	SIGLECL1	0,01804	5/109 0/138	0,08863	2/57 0/138	1,00000	0/51 0/138	FALSE	TRUE	Q8N7X8
154	MAP3K13	0,02596	8/106 2/136	0,06702	4/55 2/136	0,12265	3/48 2/136	FALSE	FALSE	O43283
25	GABRE	0,02910	21/93 13/125	0,17177	9/50 13/125	0,27161	7/44 13/125	FALSE	TRUE	P78334
347	PPP3CA	0,03350	9/105 3/135	0,76236	1/58 3/135	0,71913	1/50 3/135	TRUE	TRUE	Q08209
321	CAMKK2	0,03476	6/108 1/137	0,21406	2/57 1/137	0,46792	1/50 1/137	FALSE	FALSE	Q96RR4
16	DDN	0,03814	16/98 9/129	0,02556	10/49 9/129	0,05219	8/43 9/129		FALSE	O94850
336	SNX13	0,03987	10/104 4/134	0,09311	5/54 4/134	0,13835	4/47 4/134	FALSE	FALSE	Q9Y5W8
271	C12orf4	0,04067	4/110 0/138	0,08863	2/57 0/138	1,00000	0/51 0/138	FALSE	FALSE	Q9NQ89
166	TTC17	0,04067	4/110 0/138	1,00000	0/59 0/138	1,00000	0/51 0/138	FALSE	TRUE	Q96AE7
177	CAPRIN2	0,04517	11/103 5/133	0,44756	3/56 5/133	0,37063	3/48 5/133	FALSE	TRUE	Q6IMN6
Excluded										
369	IQSEC2	0,06016	16/98 10/128	0,01965	11/48 10/128	0,37826	5/46 10/128	FALSE	FALSE	Q5JU85
72	CFP	0,09124	3/111 0/138	0,02591	3/56 0/138	0,01881	3/48 0/138	TRUE	TRUE	P27918
240	LIN37	0,09124	3/111 0/138	0,08863	2/57 0/138	0,01881	3/48 0/138	FALSE	FALSE	Q96GY3
61	LCN15	0,13302	4/110 1/136	0,02942	4/55 1/136	0,17476	2/48 1/136		TRUE	Q6UWW0
192	ACSS1	0,15264	5/109 2/136	0,02627	5/54 2/136	0,01624	5/46 2/136	FALSE	FALSE	Q9NUB1
131	SLC20A2	0,18051	10/104 7/131	0,04290	8/51 7/131	0,10182	6/45 7/131	FALSE	TRUE	Q08357
342	SUSD6	0,20829	2/111 0/134	0,00735	4/53 0/134	0,00498	4/46 0/134	FALSE	FALSE	Q92537
304	DCST1	0,24294	3/111 1/137	0,08119	3/56 1/137	0,01933	4/47 1/137	TRUE	TRUE	Q5T197
137	CCDC181	0,25674	4/110 2/136	0,15968	3/56 2/136	0,04629	4/47 2/136		FALSE	Q5TID7
189	CELSR3	0,26259	6/108 4/134	0,04267	6/53 4/134	0,00966	7/44 4/134	FALSE	TRUE	Q9NYQ7
101	MED13	0,74355	4/110 6/132	0,02634	8/51 6/132	0,14236	5/46 6/132	FALSE	FALSE	Q9UHV7

C.2 Additional plots

The same plots as in figure 10 but for the various antigens.


16_DDN
25_GABRE
44_NAPG
68_AFAP1L1
154_MAP3K13
166_TTC17
177_CAPRIN2
236_SIGLECL1
254_ALDH2
271_C12orf4
279_TRIM63
321_CAMKK2
327_SNURF
336_SNX13
347_PPP3CA
MADs by sample
Phase 1
Phase 2
Phase 3
Control group

D Planar microarray results

The following are the significant results of the planar microarray analysis, containing 352 observations with a reactivity above 4 SDs.

The columns indicate the following:

1. HPRR: The identifier of the PrEST as a bioinformatical construct.
2. db: The percentage of the previous in-house runs in which the spot is deemed reactive.
3. SDs: The normalized signal, as described in equation 2.
4. MFI: The pixel median fluorescence intensity of the spot foreground.
5. p_rank: Descending rank of the SDs of the current pool.
6. Pool: Current pool.
7. set: States which of the two glasses making up the microarray the observation occurred on.
8. kw1: Gene ontology biological process was flagged as containing keywords in A.
9. kw2: Gene ontology cellular compartment was flagged as containing keywords in A.
10. Gene: HGNC gene identifier.
11. Uniprot: Uniprot identifier

HPRR	db	SDs	MFI	p_rank	Pool	set	kw1	kw2	Gene	Uniprot
HPRR1450131	0	84,9	65364	1	pool_4	1	FALSE	FALSE	TAF1C	Q15572
HPRR3280420	0	64,3	61408	1	pool_3	2	FALSE	FALSE	LIN37	Q96GY3
HPRR3460022	6	61,4	31222	1	pool_1	2	FALSE	FALSE	FIG4	Q92562
HPRR3050035	0	58,5	65318	1	pool_2	1	FALSE	TRUE	TTC17	Q96AE7
HPRR3140817	0	58,4	65188	2	pool_2	1	TRUE	TRUE	HIF3A	Q9Y2N7
HPRR3100022	10	44,4	38706	2	pool_4	2	FALSE	FALSE	AP3B2	Q13367
HPRR3140513	31	44,3	16721	3	pool_1	1	TRUE	TRUE	A1BG	P04217
HPRR2500042	0	42,9	37438	3	pool_4	2	FALSE	FALSE	PRR11	Q96HE9
HPRR4280292	2	38,9	43638	3	pool_2	1	NA	FALSE	SNURF	Q9Y675
HPRR3420640	2	38,7	43399	4	pool_2	1	FALSE	FALSE	CEND1	Q8N111
HPRR3880320	0	38,7	14661	4	pool_1	1	FALSE	FALSE	KCTD20	Q7Z5Y7
HPRR3100091	0	35,4	27595	4	pool_4	1	FALSE	TRUE	TGM5	O43548
HPRR2930017	0	35,3	28688	3	pool_3	1	FALSE	FALSE	METTL5	Q9NRN9
HPRR2630023	0	34,7	30444	5	pool_4	2	NA	NA	SPATS1	Q496A3
HPRR1820080	0	33,4	30075	5	pool_2	2	NA	FALSE	AFAP1L1	Q8TED9
HPRR2280131	0	32,5	28567	6	pool_4	2	FALSE	FALSE	RIPPLY3	P57055
HPRR3050002	0	31,3	25471	4	pool_3	1	FALSE	FALSE	PHLDB1	Q86UU1
HPRR3420667	10	30,9	15913	5	pool_1	2	NA	TRUE	S100A3	P33764
HPRR4390331	0	29,8	23362	7	pool_4	1	NA	FALSE	BOLA1	Q9Y3E2
HPRR3300084	0	28,7	27922	5	pool_3	2	FALSE	FALSE	MTG2	Q9H4K7
HPRR1400018	2	28,2	27449	6	pool_3	2	TRUE	TRUE	TNFRSF14	Q92956
HPRR3420383	0	27,0	30400	6	pool_2	1	FALSE	TRUE	SIGLECL1	Q8N7X8
HPRR4220339	31	26,5	21752	7	pool_3	1	TRUE	FALSE	CCDC88B	A6NC98
HPRR4150088	2	26,2	21486	8	pool_3	1	FALSE	FALSE	MNX1	P50219
HPRR2470528	0	24,3	19971	9	pool_3	1	NA	NA	FAM219A	Q8IW50
HPRR3540053	0	24,1	27207	7	pool_2	1	FALSE	TRUE	MMP23B	O75900
HPRR3310202	0	23,9	27014	8	pool_2	1	FALSE	FALSE	ACSS1	Q9NUB1
HPRR3420747	0	23,8	21683	9	pool_2	2	FALSE	TRUE	ADGRA1	Q86SQ6
HPRR4190823	47	23,7	9156	6	pool_1	1	FALSE	FALSE	ATF3	P18847
HPRR2810079	20	22,7	11810	7	pool_1	2	FALSE	TRUE	GABRE	P78334
HPRR3760566	2	22,5	20533	10	pool_2	2	FALSE	TRUE	ROBO4	Q8WZ75
HPRR3760797	0	22,2	20271	11	pool_2	2	TRUE	TRUE	DCST1	Q5T197
HPRR4200020	0	22,2	18296	10	pool_3	1	FALSE	FALSE	XAB2	Q9HCS7
HPRR4320454	2	22,0	20070	12	pool_2	2	FALSE	NA	FAM83E	Q2M2I3
HPRR3090005	0	21,4	11111	8	pool_1	2	FALSE	TRUE	MYO16	Q9Y6X6
HPRR3860339	0	20,9	16558	8	pool_4	1	TRUE	TRUE	STK10	O94804
HPRR2090170	4	20,5	23221	13	pool_2	1	FALSE	TRUE	CHRM1	P11229
HPRR3020598	0	20,5	16239	9	pool_4	1	FALSE	FALSE	DYDC1	Q8WWB3
HPRR4290533	0	20,2	10511	9	pool_1	2	FALSE	TRUE	ISLR2	Q6UXK2
HPRR3480071	0	20,0	17903	10	pool_4	2	FALSE	FALSE	CA5A	P35218
HPRR3340440	18	19,0	15107	11	pool_4	1	NA	NA	PRR20A;PRR20C;PRR20B;PRR20D;PRR20E	P86496;P86479;P86481;P86480;P86478
HPRR3141049	0	18,7	7297	10	pool_1	1	FALSE	TRUE	BSND	Q8WZ55
HPRR3830208	2	18,6	18422	11	pool_3	2	FALSE	FALSE	ZRANB2	O95218
HPRR2930265	2	17,6	16239	14	pool_2	2	FALSE	FALSE	HAAO	P46952
HPRR3140513	31	17,3	14441	12	pool_3	1	TRUE	TRUE	A1BG	P04217
HPRR3140513	31	17,0	13597	13	pool_4	1	TRUE	TRUE	A1BG	P04217
HPRR3860253	2	17,0	19321	15	pool_2	1	FALSE	FALSE	MRPS24	Q96EL2

HPRR	db	SDs	MFI	p_rank	Pool	set	kw1	kw2	Gene	Uniprot
HPRR3480041	0	16,8	13401	14	pool_4	1	FALSE	FALSE	RRN3	Q9NYV6
HPRR3051136	8	16,6	8724	11	pool_1	2	FALSE	NA	ACCSL	Q4AC99
HPRR3730635	0	16,0	14849	16	pool_2	2	FALSE	FALSE	TFCP2	Q12800
HPRR3210127	0	16,0	13434	14	pool_3	1	FALSE	TRUE	ABCA7	Q8IZY2
HPRR2950103	0	16,0	13402	15	pool_3	1	FALSE	TRUE	SPRED2	Q7Z698
HPRR3730263	0	15,4	14254	17	pool_2	2	FALSE	FALSE	SMNDC1	O75940
HPRR2552012	2	15,2	15230	16	pool_3	2	FALSE	FALSE	PERM1	Q5SV97
HPRR4110114	0	15,0	13907	18	pool_2	2	TRUE	FALSE	PIK3R2	O00459
HPRR4010009	0	14,8	5876	12	pool_1	1	FALSE	FALSE	NRBF2	Q96F24
HPRR3420383	0	14,8	5865	13	pool_1	1	FALSE	TRUE	SIGLECL1	Q8N7X8
HPRR3141012	0	14,6	13581	19	pool_2	2	TRUE	TRUE	ADAMTS13	Q76LX8
HPRR3500398	0	14,5	7657	14	pool_1	2	FALSE	NA	PNMA5	Q96PV4
HPRR3310245	18	14,4	13420	20	pool_2	2	FALSE	FALSE	SPTLC3	Q9NUV7
HPRR4090035	10	14,4	13375	21	pool_2	2	FALSE	FALSE	MYBPH	Q13203
HPRR1370153	16	14,0	13107	22	pool_2	2	TRUE	TRUE	LPCAT1	Q8NF37
HPRR4030492	0	13,8	12626	15	pool_4	2	TRUE	TRUE	DAG1	Q14118
HPRR3090254	4	13,4	12500	23	pool_2	2	FALSE	FALSE	RBM26	Q5T8P6
HPRR3500251	0	13,2	12120	16	pool_4	2	FALSE	TRUE	SLC9A7	Q96T83
HPRR4040612	27	13,0	10535	17	pool_4	1	FALSE	FALSE	MEX3C	Q5U5Q3
HPRR3280403	0	12,9	13075	17	pool_3	2	FALSE	FALSE	DACT3	Q96B18
HPRR4110319	0	12,9	12091	24	pool_2	2	NA	NA	PLEKHG3	A1L390
HPRR3071040	0	12,8	12061	25	pool_2	2	FALSE	TRUE	SCN8A	Q9UQD0
HPRR4160920	0	12,8	10418	18	pool_4	1	FALSE	FALSE	TSC22D1	Q15714
HPRR4130094	0	12,6	12777	18	pool_3	2	FALSE	FALSE	DBNDD2	Q9BQY9
HPRR3070024	0	12,5	6671	15	pool_1	2	FALSE	FALSE	C12orf4	Q9NQ89
HPRR3700378	39	12,5	5019	16	pool_1	1	FALSE	FALSE	RNF185	Q96GF1
HPRR3670067	0	12,5	11717	26	pool_2	2	FALSE	FALSE	ZKSCAN7	Q9P0L1
HPRR4030134	0	12,3	11621	27	pool_2	2	TRUE	TRUE	CFP	P27918
HPRR4040106	0	12,0	11107	19	pool_4	2	NA	TRUE	OR2K2	Q8NGT1
HPRR3280067	29	11,9	4798	17	pool_1	1	FALSE	FALSE	TIMM29	Q9BSF4
HPRR3870215	0	11,7	11018	28	pool_2	2	TRUE	TRUE	SNAP29	O95721
HPRR3070111	4	11,6	13407	29	pool_2	1	TRUE	FALSE	NUP37	Q8NFH4
HPRR4050488	0	11,3	13072	30	pool_2	1	FALSE	FALSE	POU3F4	P49335
HPRR3460376	2	11,3	9723	19	pool_3	1	FALSE	FALSE	WDR70	Q9NW82
HPRR4320252	0	11,0	11240	20	pool_3	2	FALSE	FALSE	PIR	O00625
HPRR3120222	0	10,9	10334	31	pool_2	2	FALSE	TRUE	EMILIN2	Q9BXX0
HPRR4190832	12	10,8	9305	21	pool_3	1	FALSE	FALSE	ANKLE1	Q8NAG6
HPRR3070036	18	10,8	4377	18	pool_1	1	FALSE	TRUE	ANO2	Q9NQ90
HPRR3720599	0	10,7	4350	19	pool_1	1	FALSE	FALSE	DLX4	Q92988
HPRR3310300	0	10,7	12371	32	pool_2	1	FALSE	TRUE	RALGAPA2	Q2PPJ7
HPRR3070385	0	10,7	4333	20	pool_1	1	NA	TRUE	PRH1;AC006518,7;PRH2	P02810
HPRR2250411	0	10,6	4332	21	pool_1	1	FALSE	FALSE	MIER1	Q8N108
HPRR3720498	0	10,4	9913	35	pool_2	2	TRUE	FALSE	ELF4	Q99607
HPRR4280358	0	10,3	4187	22	pool_1	1	FALSE	FALSE	STMND1	H3BQB6
HPRR3730056	0	10,2	4162	23	pool_1	1	FALSE	TRUE	ALDH2	P05091
HPRR3070013	0	9,7	8042	20	pool_4	1	FALSE	FALSE	PLEKHA5	Q9HAU0
HPRR2551962	2	9,7	9128	21	pool_4	2	NA	NA	CCDC30	Q5VVM6
HPRR4340196	2	9,7	3973	24	pool_1	1	FALSE	TRUE	HCN4	Q9Y3Q4

HPRR	db	SDs	MFI	p_rank	Pool	set	kw1	kw2	Gene	Uniprot
HPRR4090121	4	9,6	5219	25	pool_1	2	NA	FALSE	C3orf30	Q96M34
HPRR3140639	0	9,6	9034	22	pool_4	2	TRUE	TRUE	IL2RA	P01589
HPRR3420661	16	9,5	3923	26	pool_1	1	FALSE	FALSE	ZNF669	Q96BR6
HPRR2551553	0	9,5	9152	36	pool_2	2	FALSE	FALSE	IGFN1	Q86VF2
HPRR3140512	0	9,5	3914	27	pool_1	1	TRUE	TRUE	A1BG	P04217
HPRR3140485	0	9,5	8272	23	pool_3	1	FALSE	TRUE	GLE1	Q53GS7
HPRR3760711	0	9,4	3876	28	pool_1	1	TRUE	FALSE	TRIM63	Q969Q1
HPRR1450564	0	9,3	3851	29	pool_1	1	TRUE	FALSE	NPM1	P06748
HPRR3280504	18	9,3	10870	37	pool_2	1	FALSE	FALSE	AKT1S1	Q96B36
HPRR2660035	12	9,3	9650	24	pool_3	2	FALSE	NA	C6orf141	Q5SZD1
HPRR3050186	2	9,2	7603	25	pool_4	1	FALSE	FALSE	PRPF19	Q9UMS4
HPRR3970190	0	9,1	7588	26	pool_4	1	FALSE	FALSE	ZDHHC3	Q9NYG2
HPRR2920354	8	8,9	10374	38	pool_2	1	FALSE	FALSE	AFTPH	Q6ULP2
HPRR3010321	0	8,9	10342	39	pool_2	1	FALSE	TRUE	LECT2	O14960
HPRR3120104	0	8,8	3663	30	pool_1	1	NA	FALSE	ZCCHC2	Q9C0B9
HPRR3070193	0	8,8	8537	40	pool_2	2	FALSE	TRUE	CAPRIN2	Q6IMN6
HPRR3780077	0	8,8	4805	31	pool_1	2	FALSE	TRUE	FUT1	P19526
HPRR2540556	0	8,8	7318	27	pool_4	1	FALSE	TRUE	SLC20A2	Q08357
HPRR4290077	0	8,8	7707	25	pool_3	1	FALSE	TRUE	FRMPD1	Q5SYB0
HPRR3790916	2	8,6	7217	28	pool_4	1	TRUE	FALSE	PSMB11	A5LHX3
HPRR4160573	2	8,6	8350	41	pool_2	2	TRUE	TRUE	FYB1	O15117
HPRR3050138	2	8,5	8219	42	pool_2	2	FALSE	FALSE	CEP126	Q9P2H0
HPRR2960810	2	8,4	4626	32	pool_1	2	FALSE	TRUE	SCN10A	Q9Y5Y9
HPRR2810104	0	8,4	4606	33	pool_1	2	FALSE	FALSE	KDM5C	P41229
HPRR3000060	0	8,3	7384	26	pool_3	1	FALSE	FALSE	NOA1	Q8NC60
HPRR2620027	6	8,3	3479	34	pool_1	1	FALSE	FALSE	DNAH8	Q96JB1
HPRR4290309	0	8,3	3479	35	pool_1	1	FALSE	TRUE	GCOM1;MYZAP	P0CAP1
HPRR4370219	14	8,3	8741	27	pool_3	2	FALSE	FALSE	PFKFB1	P16118
HPRR3610292	57	8,3	9734	43	pool_2	1	FALSE	FALSE	ZNF688	P0C7X2
HPRR3730550	0	8,3	3460	36	pool_1	1	FALSE	TRUE	NEDD8	Q15843
HPRR3540042	4	8,2	8025	44	pool_2	2	NA	NA	SLCO1B3;AC011604,2;SLCO1B7;SLCO1B1	Q9NPD5;G3V0H7;Q9Y6L6
HPRR2540601	0	8,2	6899	29	pool_4	1	NA	NA	C8orf74	Q6P047
HPRR3050851	0	8,2	8001	45	pool_2	2	FALSE	TRUE	CKAP5	Q14008
HPRR3070447	0	8,2	8000	46	pool_2	2	FALSE	FALSE	ZC3H10	Q96K80
HPRR4090049	0	8,2	3420	37	pool_1	1	FALSE	FALSE	MBD3	O95983
HPRR3100436	2	8,1	7934	47	pool_2	2	NA	FALSE	C15orf39	Q6ZRI6
HPRR2540604	2	8,0	4419	38	pool_1	2	FALSE	FALSE	ESCO2	Q56NI9
HPRR4200014	2	8,0	3363	39	pool_1	1	FALSE	TRUE	FAT2	Q9NYQ8
HPRR4190138	2	8,0	9409	48	pool_2	1	TRUE	TRUE	PPP3CA	Q08209
HPRR1920067	10	7,9	9325	49	pool_2	1	FALSE	TRUE	NAPG	Q99747
HPRR4110218	4	7,9	7033	29	pool_3	1	FALSE	FALSE	CAMKK2	Q96RR4
HPRR3790073	2	7,9	8314	30	pool_3	2	NA	TRUE	LCN15	Q6UWW0
HPRR3300056	0	7,8	7527	30	pool_4	2	FALSE	TRUE	CPXM1	Q96SM3
HPRR2550582	0	7,8	3274	40	pool_1	1	FALSE	TRUE	NENF	Q9UMX5
HPRR1140021	4	7,8	3270	41	pool_1	1	FALSE	TRUE	FMN1	Q68DA7
HPRR3000234	0	7,8	6533	31	pool_4	1	FALSE	FALSE	COPS4	Q9BT78

HPRR	db	SDs	MFI	p_rank	Pool	set	kw1	kw2	Gene	Uniprot
HPRR3420319	0	7,6	4220	42	pool_1	2	TRUE	TRUE	C8G	P07360
HPRR231711	0	7,6	6799	31	pool_3	1	FALSE	FALSE	ATXN3L	Q9H3M9
HPRR2540601	0	7,6	3191	43	pool_1	1	NA	NA	C8orf74	Q6P047
HPRR3920097	12	7,5	6378	32	pool_4	1	TRUE	FALSE	KLF2	Q9Y5W3
HPRR2920484	0	7,5	6686	32	pool_3	1	TRUE	FALSE	TANK	Q92844
HPRR3020645	0	7,5	8783	50	pool_2	1	TRUE	TRUE	CALML5	Q9NZT1
HPRR3420406	8	7,4	8724	51	pool_2	1	FALSE	FALSE	ARID3B	Q8IVW6
HPRR3400060	0	7,4	7823	33	pool_3	2	TRUE	TRUE	YPEL5	P62699
HPRR3010670	0	7,3	7190	53	pool_2	2	NA	TRUE	ARSI	Q5FYB1
HPRR3460850	0	7,3	4032	45	pool_1	2	NA	FALSE	ANKRD34B	A5PLL1
HPRR2550430	0	7,2	7109	54	pool_2	2	FALSE	NA	SIPA1L2	Q9P2F8
HPRR3790027	0	7,2	7075	55	pool_2	2	FALSE	FALSE	PDXDC1	Q6P996
HPRR3890790	0	7,2	6443	34	pool_3	1	FALSE	FALSE	MSI1	O43347
HPRR3890580	0	7,1	3965	46	pool_1	2	FALSE	FALSE	CLIP1	P30622
HPRR2620003	2	7,1	6036	33	pool_4	1	FALSE	FALSE	MOCS1	Q9NZB8
HPRR2810103	2	7,1	3942	47	pool_1	2	TRUE	FALSE	CITED1	Q99966
HPRR3140495	4	7,1	6991	56	pool_2	2	TRUE	FALSE	NR4A3	Q92570
HPRR4080085	0	7,0	5990	34	pool_4	1	FALSE	FALSE	CRIP2	P52943
HPRR2290159	0	7,0	2993	48	pool_1	1	TRUE	FALSE	IPO7	O95373
HPRR3470279	10	6,9	7433	35	pool_3	2	FALSE	TRUE	PRCD	Q00LT1
HPRR4170045	0	6,9	6271	36	pool_3	1	FALSE	FALSE	OAZ1	P54368
HPRR3400251	8	6,9	6881	57	pool_2	2	FALSE	FALSE	DRC1	Q96MC2
HPRR4320117	0	6,9	6824	58	pool_2	2	FALSE	TRUE	GUCA2B	Q16661
HPRR4160972	0	6,8	5842	35	pool_4	1	FALSE	FALSE	SUSD6	Q92537
HPRR4190823	47	6,8	6162	37	pool_3	1	FALSE	FALSE	ATF3	P18847
HPRR3890111	0	6,8	6631	36	pool_4	2	FALSE	TRUE	KCNC4	Q03721
HPRR4050555	0	6,8	2901	49	pool_1	1	FALSE	FALSE	ZNF560	Q96MR9
HPRR3720315	2	6,7	6705	60	pool_2	2	FALSE	FALSE	SLC35C2	Q9NQQ7
HPRR1440073	0	6,7	6082	38	pool_3	1	TRUE	TRUE	HPN	P05981
HPRR3760566	2	6,7	6524	38	pool_4	2	FALSE	TRUE	ROBO4	Q8WZ75
HPRR4110051	2	6,7	7908	61	pool_2	1	FALSE	FALSE	PCGF2	P35227
HPRR4130031	0	6,6	6484	40	pool_4	2	NA	FALSE	C10orf105	Q8TEF2
HPRR3020832	0	6,6	6008	39	pool_3	1	FALSE	FALSE	RBM20	Q5T481
HPRR3100096	0	6,5	5896	40	pool_3	1	FALSE	FALSE	RPAP1	Q9BWH6
HPRR3730387	0	6,4	5876	41	pool_3	1	FALSE	FALSE	TRERF1	Q96PN7
HPRR230959	0	6,4	5831	42	pool_3	1	FALSE	FALSE	ARMCX3	Q9UH62
HPRR4040718	6	6,4	6895	43	pool_3	2	FALSE	FALSE	SIAH2	O43255
HPRR2360049	0	6,4	5815	44	pool_3	1	FALSE	FALSE	GAB1	Q13480
HPRR3250208	0	6,3	5415	41	pool_4	1	TRUE	TRUE	HCST	Q9UBK5
HPRR3860253	2	6,3	5399	42	pool_4	1	FALSE	FALSE	MRPS24	Q96EL2
HPRR3250026	0	6,3	5733	45	pool_3	1	FALSE	FALSE	LIG1	P18858
HPRR3790142	0	6,3	6175	43	pool_4	2	FALSE	TRUE	MRGPRG	Q86SM5
HPRR3100218	0	6,2	3496	50	pool_1	2	FALSE	FALSE	INTS14	Q96SY0
HPRR3860359	6	6,2	5338	45	pool_4	1	TRUE	FALSE	IGF2BP2	Q9Y6M1
HPRR2440663	0	6,2	5677	46	pool_3	1	FALSE	FALSE	COG5	Q9UP83
HPRR2660035	12	6,2	6100	46	pool_4	2	FALSE	NA	C6orf141	Q5SZD1
HPRR3140620	0	6,2	7354	62	pool_2	1	FALSE	TRUE	ANGPTL6	Q8NI99
HPRR2850699	2	6,2	6700	47	pool_3	2	FALSE	FALSE	AGMO	Q6ZNB7
HPRR4160709	0	6,1	5639	48	pool_3	1	FALSE	FALSE	USP33	Q8TEY7

HPRR	db	SDs	MFI	p_rank	Pool	set	kw1	kw2	Gene	Uniprot
HPRR3340279	0	6,1	2658	51	pool_1	1	FALSE	FALSE	MEIS3	Q99687
HPRR3730007	0	6,1	6039	47	pool_4	2	TRUE	TRUE	CRTAM	O95727
HPRR2960328	18	6,1	2638	53	pool_1	1	FALSE	FALSE	IQSEC1	Q6DN90
HPRR3320021	2	6,0	5969	48	pool_4	2	TRUE	TRUE	NPEPPS	P55786
HPRR2920005	0	6,0	5961	49	pool_4	2	FALSE	FALSE	FARP2	O94887
HPRR3310022	0	6,0	5526	49	pool_3	1	FALSE	FALSE	MGME1	Q9BQP7
HPRR3250127	0	6,0	5187	50	pool_4	1	FALSE	FALSE	CD3EAP	O15446
HPRR4030048	0	6,0	3387	54	pool_1	2	TRUE	FALSE	RUNX3	Q13761
HPRR2930265	2	6,0	3385	55	pool_1	2	FALSE	FALSE	HAAO	P46952
HPRR3500103	4	6,0	6525	50	pool_3	2	FALSE	TRUE	EMP1	P54849
HPRR3870276	0	6,0	5923	51	pool_4	2	FALSE	FALSE	ZNF516	Q92618
HPRR3760818	0	5,9	5137	52	pool_4	1	FALSE	TRUE	DDR2	Q16832
HPRR3770042	22	5,9	5124	53	pool_4	1	TRUE	TRUE	PTPRS	Q13332
HPRR4180611	2	5,9	5112	54	pool_4	1	FALSE	FALSE	IQSEC2	Q5JU85
HPRR4060033	0	5,9	6439	51	pool_3	2	FALSE	FALSE	RCN1	Q15293
HPRR1450401	0	5,9	3336	56	pool_1	2	FALSE	FALSE	ZNF18	P17022
HPRR4280338	0	5,9	5949	63	pool_2	2	FALSE	TRUE	DIO1	P49895
HPRR3390107	4	5,8	5917	64	pool_2	2	NA	NA	PRR3	P79522
HPRR3640047	4	5,8	5781	55	pool_4	2	FALSE	FALSE	SRRM1	Q8IYB3
HPRR3030026	0	5,8	6943	65	pool_2	1	FALSE	FALSE	MAP3K13	O43283
HPRR3410133	0	5,8	6941	66	pool_2	1	TRUE	TRUE	SUSD4	Q5VX71
HPRR2930229	0	5,8	3276	57	pool_1	2	FALSE	TRUE	FAM126B	Q8IXS8
HPRR1370007	0	5,7	5844	67	pool_2	2	TRUE	FALSE	ANLN	Q9NQW6
HPRR3010665	0	5,7	5843	68	pool_2	2	FALSE	FALSE	KCTD16	Q68DU8
HPRR220396	0	5,7	5332	52	pool_3	1	TRUE	FALSE	MORC2	Q9Y6X9
HPRR4160359	0	5,7	3265	58	pool_1	2	FALSE	FALSE	SNX13	Q9Y5W8
HPRR3790690	0	5,7	6296	53	pool_3	2	FALSE	FALSE	PAPSS2	O95340
HPRR2970072	4	5,7	3258	59	pool_1	2	TRUE	TRUE	GRB7	Q14451
HPRR3010234	2	5,7	6292	54	pool_3	2	FALSE	FALSE	PAPD7	Q5XG87
HPRR3100022	10	5,7	6291	55	pool_3	2	FALSE	FALSE	AP3B2	Q13367
HPRR2880060	0	5,7	4976	56	pool_4	1	TRUE	TRUE	STAMBP	O95630
HPRR3760334	0	5,7	4976	57	pool_4	1	FALSE	FALSE	NR1I3	Q14994
HPRR2550478	0	5,7	5706	58	pool_4	2	NA	FALSE	CCDC181	Q5TID7
HPRR1350006	0	5,7	5686	59	pool_4	2	TRUE	TRUE	PTPRN	Q16849
HPRR4160381	0	5,7	3224	60	pool_1	2	TRUE	FALSE	KDM4A	O75164
HPRR2551567	0	5,7	2491	61	pool_1	1	FALSE	TRUE	SRGAP2	O75044
HPRR4180186	0	5,6	4927	60	pool_4	1	FALSE	TRUE	COL12A1	Q99715
HPRR3060026	8	5,6	2483	62	pool_1	1	FALSE	FALSE	TEX15	Q9BXT5
HPRR2501355	8	5,6	6702	69	pool_2	1	FALSE	FALSE	SHISA6	Q6ZSJ9
HPRR3460552	0	5,5	3152	63	pool_1	2	FALSE	FALSE	TOPAZ1	Q8N9V7
HPRR3730670	0	5,5	2432	64	pool_1	1	FALSE	FALSE	SBF2	Q86WG5
HPRR3830074	0	5,5	4806	61	pool_4	1	FALSE	FALSE	SARS2;AC011455,2	Q9NP81
HPRR3010047	2	5,5	5614	70	pool_2	2	FALSE	TRUE	FSTL4	Q6MZW2
HPRR2501355	8	5,5	2424	65	pool_1	1	FALSE	FALSE	SHISA6	Q6ZSJ9
HPRR3610292	57	5,4	5097	56	pool_3	1	FALSE	FALSE	ZNF688	P0C7X2
HPRR3500150	2	5,4	5069	57	pool_3	1	FALSE	FALSE	YME1L1	Q96TA2
HPRR4180818	2	5,3	2378	66	pool_1	1	FALSE	FALSE	RLIM	Q9NVW2
HPRR4220131	0	5,3	5492	71	pool_2	2	FALSE	FALSE	JCAD	Q9P266

HPRR	db	SDs	MFI	p_rank	Pool	set	kw1	kw2	Gene	Uniprot
HPRR4220450	0	5,3	3058	67	pool_1	2	FALSE	TRUE	MMRN2	Q9H8L6
HPRR3580152	0	5,2	5282	63	pool_4	2	NA	TRUE	OR8S1	Q8NH09
HPRR1950384	0	5,2	6278	72	pool_2	1	TRUE	TRUE	IL21R	Q9HBE5
HPRR3140706	0	5,2	5346	73	pool_2	2	FALSE	TRUE	KCNS1	Q96KK3
HPRR4050663	33	5,2	5252	64	pool_4	2	FALSE	FALSE	CCDC85C	A6NKD9
HPRR400117	4	5,1	2964	68	pool_1	2	TRUE	TRUE	TSC1	Q92574
HPRR2850169	2	5,1	2936	69	pool_1	2	FALSE	FALSE	TMEM33	P57088
HPRR1450124	0	5,1	4490	67	pool_4	1	FALSE	FALSE	ADNP2	Q6IQ32
HPRR4030190	0	5,0	4775	58	pool_3	1	FALSE	TRUE	GDF11	O95390
HPRR2570136	0	5,0	4446	68	pool_4	1	FALSE	FALSE	SESN1	Q9Y6P5
HPRR2551992	2	5,0	5107	69	pool_4	2	FALSE	FALSE	RNF220	Q5VTB9
HPRR3450023	0	5,0	2893	72	pool_1	2	FALSE	FALSE	LUZP4	Q9P127
HPRR3010367	27	5,0	5095	70	pool_4	2	FALSE	FALSE	POC5	Q8NA72
HPRR3680222	0	5,0	5160	74	pool_2	2	FALSE	TRUE	HMCN2	Q8NDA2
HPRR510010	0	5,0	5157	75	pool_2	2	FALSE	FALSE	RUBCNL	Q9H714
HPRR3100294	0	5,0	4407	71	pool_4	1	FALSE	FALSE	SH3GL3	Q99963
HPRR2470425	0	4,9	5054	72	pool_4	2	FALSE	FALSE	SURF1	Q15526
HPRR3390003	2	4,9	5047	73	pool_4	2	FALSE	FALSE	ESPN	B1AK53
HPRR2360051	0	4,9	5991	76	pool_2	1	FALSE	FALSE	GAB1	Q13480
HPRR3890790	0	4,9	2223	73	pool_1	1	FALSE	FALSE	MSI1	O43347
HPRR2440808	0	4,9	4379	74	pool_4	1	NA	FALSE	TNRC18	O15417
HPRR2550975	10	4,9	5984	77	pool_2	1	FALSE	FALSE	DCAF6	Q58WW2
HPRR680317	0	4,9	4681	59	pool_3	1	TRUE	FALSE	BACH2	Q9BYV9
HPRR3420256	0	4,9	4370	75	pool_4	1	FALSE	FALSE	IRX1	P78414
HPRR3100112	12	4,9	4358	76	pool_4	1	FALSE	TRUE	MESD	Q14696
HPRR2810079	20	4,9	5489	60	pool_3	2	FALSE	TRUE	GABRE	P78334
HPRR3890398	0	4,9	4328	77	pool_4	1	FALSE	TRUE	NRSN2	Q9GZP1
HPRR3000044	2	4,9	4325	78	pool_4	1	FALSE	TRUE	ARHGAP10	A1A4S6
HPRR3050415	0	4,9	5472	61	pool_3	2	NA	NA	NKAPD1	Q6ZUT1
HPRR3890362	0	4,8	4319	79	pool_4	1	TRUE	TRUE	NPY	P01303
HPRR2950103	0	4,8	4316	80	pool_4	1	FALSE	TRUE	SPRED2	Q7Z698
HPRR3250136	0	4,8	5440	63	pool_3	2	FALSE	TRUE	SIPA1L3	O60292
HPRR3210226	0	4,8	2172	75	pool_1	1	FALSE	NA	PPP1R37	O75864
HPRR3790027	0	4,8	4923	81	pool_4	2	FALSE	FALSE	PDXDC1	Q6P996
HPRR2501355	8	4,7	4534	64	pool_3	1	FALSE	FALSE	SHISA6	Q6ZSJ9
HPRR3700378	39	4,7	4224	82	pool_4	1	FALSE	FALSE	RNF185	Q96GF1
HPRR3470014	12	4,7	2146	76	pool_1	1	NA	NA	C6orf163	Q5TEZ5
HPRR3460813	2	4,7	4936	78	pool_2	2	NA	FALSE	RABL6	Q3YEC7
HPRR2050129	2	4,7	4212	83	pool_4	1	FALSE	FALSE	NSRP1	Q9H0G5
HPRR2870030	0	4,7	4212	84	pool_4	1	FALSE	FALSE	OCA2	Q04671
HPRR3280213	2	4,7	5718	79	pool_2	1	TRUE	TRUE	LILRA4	P59901
HPRR3210135	2	4,7	2725	77	pool_1	2	FALSE	FALSE	RNF126	Q9BV68
HPRR2570042	2	4,7	2122	78	pool_1	1	TRUE	FALSE	CUL7	Q14999
HPRR3790353	0	4,7	5684	80	pool_2	1	FALSE	TRUE	L3MBTL1	Q9Y468
HPRR2180111	0	4,6	2114	79	pool_1	1	FALSE	TRUE	KCNK5	O95279
HPRR3500227	6	4,6	5660	81	pool_2	1	NA	FALSE	CAPSL	Q8WWF8
HPRR2552223	0	4,6	4137	85	pool_4	1	TRUE	FALSE	GPATCH3	Q96I76
HPRR4340274	22	4,6	4429	65	pool_3	1	FALSE	TRUE	EPHA1	P21709
HPRR1450175	0	4,6	5233	66	pool_3	2	FALSE	FALSE	MED13	Q9UHV7

HPRR	db	SDs	MFI	p_rank	Pool	set	kw1	kw2	Gene	Uniprot
HPRR3020218	4	4,6	4832	82	pool_2	2	TRUE	TRUE	GPAM	Q9HCL2
HPRR3340204	6	4,6	5223	67	pool_3	2	NA	NA	WASHC2C;WASHC2A	Q9Y4E1;Q641Q2
HPRR3460092	2	4,6	4416	68	pool_3	1	FALSE	FALSE	TBCCD1	Q9NVR7
HPRR3890406	2	4,6	5218	69	pool_3	2	FALSE	FALSE	GLIS2	Q9BZE0
HPRR3010005	0	4,6	4743	86	pool_4	2	FALSE	FALSE	FAM13B	Q9NYF5
HPRR1770002	6	4,6	4103	87	pool_4	1	FALSE	TRUE	CELSR3	Q9NYQ7
HPRR3300234	10	4,6	2671	80	pool_1	2	FALSE	TRUE	NCOA5	Q9HCD5
HPRR3460017	0	4,5	4090	88	pool_4	1	FALSE	FALSE	SPDL1	Q96EA4
HPRR3000498	0	4,5	4084	89	pool_4	1	FALSE	TRUE	PARM1	Q6UWI2
HPRR2760282	0	4,5	4373	70	pool_3	1	TRUE	TRUE	IL15	P40933
HPRR3070101	0	4,5	4779	83	pool_2	2	FALSE	FALSE	EIF4B	P23588
HPRR3070828	0	4,5	2077	81	pool_1	1	FALSE	FALSE	RHNO1	Q9BSD3
HPRR3070082	29	4,5	4777	84	pool_2	2	FALSE	TRUE	DIP2B	Q9P265
HPRR3020680	0	4,5	4052	90	pool_4	1	FALSE	FALSE	TUBAL3	A6NHL2
HPRR2640051	27	4,5	4676	91	pool_4	2	TRUE	TRUE	PTPRS	Q13332
HPRR3160158	2	4,5	4038	92	pool_4	1	FALSE	FALSE	RBBP6	Q7Z6E9
HPRR1450707	0	4,5	2051	82	pool_1	1	FALSE	FALSE	ZSCAN31	Q96LW9
HPRR3860253	2	4,4	2047	83	pool_1	1	FALSE	FALSE	MRPS24	Q96EL2
HPRR3790157	2	4,4	5087	71	pool_3	2	NA	FALSE	DDN	O94850
HPRR3970188	0	4,4	4630	93	pool_4	2	NA	NA	KIAA1143	Q96AT1
HPRR3120082	0	4,4	2604	84	pool_1	2	FALSE	TRUE	ATP5F1A	P25705
HPRR2770023	4	4,4	5407	85	pool_2	1	TRUE	TRUE	TACR1	P25103
HPRR2930132	0	4,4	4588	94	pool_4	2	FALSE	TRUE	SPAG16	Q8N0X2
HPRR3420566	0	4,4	2026	85	pool_1	1	TRUE	TRUE	SPNS2	Q8IVW8
HPRR3940247	0	4,4	4655	86	pool_2	2	FALSE	FALSE	BTBD11	A6QL63
HPRR3320150	0	4,4	4638	87	pool_2	2	FALSE	TRUE	FKBPL	Q9UIM3
HPRR2970029	57	4,4	4241	72	pool_3	1	FALSE	FALSE	RIN3	Q8TB24
HPRR3050221	0	4,4	4230	73	pool_3	1	FALSE	FALSE	PEX16	Q9Y5Y5
HPRR1470106	0	4,3	3937	95	pool_4	1	TRUE	TRUE	DDB1	Q16531
HPRR3760638	0	4,3	3935	96	pool_4	1	FALSE	FALSE	TEDC2	Q7L2K0
HPRR231399	0	4,3	5321	88	pool_2	1	FALSE	FALSE	POU3F4	P49335
HPRR4120351	2	4,3	4937	74	pool_3	2	FALSE	FALSE	SERTAD1	Q9UHV2
HPRR2990212	4	4,3	4561	90	pool_2	2	NA	FALSE	TMEM139	Q8IV31
HPRR2551214	20	4,3	4493	97	pool_4	2	FALSE	FALSE	CCSAP	Q6IQ19
HPRR4110248	0	4,3	3868	98	pool_4	1	FALSE	FALSE	DCAF4	Q8WV16
HPRR4340082	0	4,2	1969	87	pool_1	1	TRUE	TRUE	LGALS3	P17931
HPRR3940206	0	4,2	2507	88	pool_1	2	FALSE	TRUE	GABRA2	P47869
HPRR1140021	4	4,2	3850	99	pool_4	1	FALSE	TRUE	FMN1	Q68DA7
HPRR2550183	0	4,2	4130	75	pool_3	1	TRUE	FALSE	KLHL20	Q9Y2M5
HPRR3000368	0	4,2	5202	91	pool_2	1	FALSE	FALSE	TRMT44	Q8IYL2
HPRR3300256	0	4,2	5199	92	pool_2	1	TRUE	TRUE	SEMG1	P04279
HPRR2850737	4	4,2	4501	93	pool_2	2	FALSE	FALSE	SELENOT	P62341
HPRR3460530	10	4,2	2491	89	pool_1	2	NA	FALSE	GON7	Q9BXV9
HPRR2440481	0	4,2	3822	100	pool_4	1	FALSE	FALSE	CCM2	Q9BSQ5
HPRR4090131	0	4,2	2479	90	pool_1	2	FALSE	FALSE	MAIP1	Q8WWC4
HPRR3010516	0	4,2	5157	94	pool_2	1	FALSE	FALSE	DMXL1	Q9Y485
HPRR2551687	0	4,2	4087	76	pool_3	1	FALSE	FALSE	ANGEL2	Q5VTE6
HPRR3470199	0	4,2	3799	101	pool_4	1	FALSE	FALSE	ANKRD28	O15084

HPRR	db	SDs	MFI	p_rank	Pool	set	kw1	kw2	Gene	Uniprot
HPRR3700376	0	4,2	4456	95	pool_2	2	FALSE	TRUE	KIF21A	Q7Z4S6
HPRR4040434	0	4,2	4455	96	pool_2	2	FALSE	TRUE	NPPA	P01160
HPRR3730443	0	4,2	4073	77	pool_3	1	FALSE	FALSE	SLC25A23	Q9BV35
HPRR3140155	0	4,1	4379	102	pool_4	2	FALSE	FALSE	HES2	Q9Y543
HPRR3560056	0	4,1	4798	78	pool_3	2	FALSE	TRUE	FJX1	Q86VR8
HPRR2930282	0	4,1	5082	97	pool_2	1	FALSE	NA	WDCP	Q9H6R7
HPRR3970185	0	4,1	4391	98	pool_2	2	FALSE	TRUE	CLASP2	O75122
HPRR3430094	0	4,1	4743	79	pool_3	2	NA	NA	FCGR1B;FCGR1A	Q92637;P12314
HPRR3020304	0	4,1	4375	100	pool_2	2	FALSE	TRUE	LOXL4	Q96JB6
HPRR3050344	2	4,1	3718	103	pool_4	1	NA	FALSE	CCDC82	Q8N4S0
HPRR3470137	2	4,0	4699	80	pool_3	2	NA	FALSE	C4orf46	Q504U0
HPRR3720239	0	4,0	4993	101	pool_2	1	TRUE	TRUE	ST6GAL1	P15907
HPRR3140525	0	4,0	3689	104	pool_4	1	FALSE	TRUE	KCNC1	P48547
HPRR4080101	0	4,0	2399	91	pool_1	2	FALSE	FALSE	ZNF699	Q32M78
HPRR1950034	0	4,0	2394	92	pool_1	2	TRUE	TRUE	CD6	P30203
HPRR1840163	6	4,0	4966	102	pool_2	1	FALSE	TRUE	ROR2	Q01974
HPRR1950241	0	4,0	4254	105	pool_4	2	FALSE	FALSE	NAT14	Q8WUY8

E Suspension bead array

ID	Plate	Pos	HPRA	HPRR
1	1	1	27604	HPRR3140513
2	1	2	38491	HPRR4090035
3	1	3	1023	HPRR330007
4	1	4	39089	HPRR3970260
5	1	5	19376	HPRR2970029
6	1	6	31477	HPRR3640047
7	1	7	39907	HPRR4040612
8	1	8	23321	HPRR2640051
9	1	9	32913	HPRR3760566
10	1	10	25727	HPRR3140512
11	1	11	36257	HPRR3920097
12	1	12	15637	HPRR2760215
13	1	13	20564	HPRR2760112
14	1	14	31163	HPRR3610292
15	1	15	37285	HPRR4040718
16	1	16	37989	HPRR3790157
17	1	17	14939	HPRR2550975
18	1	18	36332	HPRR3860359
19	1	19	38657	HPRR3090254
20	1	20	5742	HPRR1450564
21	1	21	22506	HPRR2960328
22	1	22	31617	HPRR3540042
23	1	23	34899	HPRR3770042
24	1	24	21067	HPRR3070036
25	1	25	22994	HPRR2810079
26	1	26	22270	HPRR3070082
27	1	27	25765	HPRR3160158
28	1	28	23822	HPRR2760285
29	1	29	40000	HPRR4040434
30	1	30	13161	HPRR1060056
31	1	31	27910	HPRR3360011
32	1	32	29015	HPRR3360041
33	1	33	30010	HPRR3360040
34	1	34	32051	HPRR3730122
35	1	35	38367	HPRR4110234
36	1	36	4214	HPRR1350006
37	1	37	25413	HPRR3140525
38	1	38	26446	HPRR3420640
39	1	39	27768	HPRR3280067
40	1	40	34842	HPRR3420667
41	1	41	38050	HPRR4050663
42	1	42	21821	HPRR3010367
43	1	43	20380	HPRR2660035
44	1	44	5366	HPRR1920067
45	1	45	11418	HPRR1840163
46	1	46	16320	HPRR2620027
47	1	47	27748	HPRR3280504
48	1	48	29419	HPRR3470279
49	1	49	37401	HPRR3340440
50	1	50	36165	HPRR3830074
51	1	51	36527	HPRR3940206
52	1	52	25204	HPRR3140817
53	1	53	23768	HPRR2970072
54	1	54	27120	HPRR3140495
55	1	55	29273	HPRR3420661
56	1	56	31750	HPRR3300234
57	1	57	31870	HPRR3420406
58	1	58	33088	HPRR3450023
59	1	59	21696	HPRR3020304
60	1	60	29393	HPRR3500103
61	1	61	32833	HPRR3790073
62	1	62	12145	HPRR2501355
63	1	63	6182	HPRR1920077
64	1	64	23070	HPRR3010047
65	1	65	27204	HPRR2551214
66	1	66	30414	HPRR3310245
67	1	67	39049	HPRR4030492
68	1	68	5809	HPRR1820080
69	1	69	36541	HPRR3700376
70	1	70	30317	HPRR3700378
71	1	71	44228	HPRR4190823
72	1	72	37397	HPRR4030133
73	1	73	37662	HPRR3830230
74	1	74	38088	HPRR4030134
75	1	75	1045	HPRR230370
76	1	76	34343	HPRR3730070
77	1	77	36005	HPRR3890362
78	1	78	26060	HPRR3141009
79	1	79	26782	HPRR3141011
80	1	80	33311	HPRR3141012
81	1	81	31277	HPRR3720239
82	1	82	31397	HPRR3730069
83	1	83	21721	HPRR2640093
84	1	84	21972	HPRR2760282
85	1	85	24978	HPRR3420318
86	1	86	25121	HPRR3010321
87	1	87	25235	HPRR3140620
88	1	88	30576	HPRR3420319
89	1	89	33240	HPRR3780122
90	1	90	15970	HPRR2570223
91	1	91	458	HPRR231711
92	1	92	499	HPRR231399
93	1	93	513	HPRR220396
94	1	94	1616	HPRR400117
95	1	95	1797	HPRR370205
96	1	96	2248	HPRR370206
97	2	1	3203	HPRR520018
98	2	2	3390	HPRR330039

ID	Plate	Pos	HPRA	HPRR
99	2	3	3709	HPRR1370007
100	2	4	3796	HPRR1450401
101	2	5	3820	HPRR1450175
102	2	6	3837	HPRR1400018
103	2	7	3838	HPRR1400019
104	2	8	4131	HPRR1440073
105	2	9	4185	HPRR1450707
106	2	10	4187	HPRR1450124
107	2	11	4368	HPRR1470106
108	2	12	5344	HPRR1950241
109	2	13	5397	HPRR1950384
110	2	14	5677	HPRR1950416
111	2	15	5768	HPRR1440074
112	2	16	5862	HPRR1370153
113	2	17	7032	HPRR2180111
114	2	18	7163	HPRR2210075
115	2	19	7383	HPRR2050129
116	2	20	7592	HPRR1450131
117	2	21	7915	HPRR680317
118	2	22	8326	HPRR2090170
119	2	23	8670	HPRR2250411
120	2	24	8949	HPRR2360049
121	2	25	8956	HPRR2360051
122	2	26	9128	HPRR2290159
123	2	27	9505	HPRR2280131
124	2	28	10360	HPRR2470528
125	2	29	10584	HPRR2470425
126	2	30	10688	HPRR2440481
127	2	31	11193	HPRR2440808
128	2	32	11473	HPRR2500042
129	2	33	11567	HPRR2210076
130	2	34	11593	HPRR1370154
131	2	35	13107	HPRR2540556
132	2	36	13607	HPRR2540604
133	2	37	14069	HPRR2550183
134	2	38	14254	HPRR510010
135	2	39	14377	HPRR2550582
136	2	40	14553	HPRR2550430
137	2	41	14589	HPRR2550478
138	2	42	14870	HPRR2620003
139	2	43	15007	HPRR2551567
140	2	44	15586	HPRR2551992
141	2	45	16109	HPRR1951141
142	2	46	16306	HPRR2570042
143	2	47	16432	HPRR2770023
144	2	48	17503	HPRR2920005
145	2	49	18501	HPRR2551687
146	2	50	18636	HPRR2552012
147	2	51	18749	HPRR2800004
148	2	52	18975	HPRR2630023
149	2	53	20347	HPRR2930229
150	2	54	20627	HPRR2880060
151	2	55	20665	HPRR3000060
152	2	56	20752	HPRR3070013
153	2	57	20825	HPRR3010516
154	2	58	20980	HPRR3030026
155	2	59	21181	HPRR2551962
156	2	60	21248	HPRR2930265
157	2	61	21341	HPRR3060026
158	2	62	21470	HPRR3010665
159	2	63	21564	HPRR2990212
160	2	64	21740	HPRR3010670
161	2	65	21859	HPRR2930132
162	2	66	21935	HPRR3020832
163	2	67	22025	HPRR3020598
164	2	68	22166	HPRR3020680
165	2	69	22211	HPRR3050002
166	2	70	22369	HPRR3050035
167	2	71	22449	HPRR3000368
168	2	72	22666	HPRR2810104
169	2	73	22680	HPRR2950103
170	2	74	22766	HPRR3070828
171	2	75	22822	HPRR3050344
172	2	76	22993	HPRR2810103
173	2	77	23233	HPRR3051136
174	2	78	23235	HPRR3070101
175	2	79	23402	HPRR3070447
176	2	80	23422	HPRR3050415
177	2	81	23680	HPRR3070193
178	2	82	23715	HPRR2850737
179	2	83	23916	HPRR3100294
180	2	84	23999	HPRR3140155
181	2	85	24102	HPRR3100436
182	2	86	24243	HPRR2850169
183	2	87	24261	HPRR3090005
184	2	88	24409	HPRR3100218
185	2	89	24436	HPRR2850699
186	2	90	24441	HPRR2960810
187	2	91	24481	HPRR3050851
188	2	92	24554	HPRR1950034
189	2	93	24766	HPRR1770002
190	2	94	24976	HPRR3310022
191	2	95	25013	HPRR3120104
192	2	96	25072	HPRR3310202
193	3	1	25087	HPRR2850680
194	3	2	25116	HPRR2930282
195	3	3	25213	HPRR3070111
196	3	4	25339	HPRR3000498

ID	Plate	Pos	HPRA	HPRR
197	3	5	25535	HPRR3250026
198	3	6	25550	HPRR3210226
199	3	7	25636	HPRR2440663
200	3	8	25659	HPRR2570136
201	3	9	25688	HPRR3100091
202	3	10	25702	HPRR3280213
203	3	11	25703	HPRR3250127
204	3	12	25729	HPRR3140485
205	3	13	25761	HPRR3100112
206	3	14	25769	HPRR3250208
207	3	15	25779	HPRR3100096
208	3	16	25861	HPRR3010005
209	3	17	26010	HPRR3140816
210	3	18	26395	HPRR3300256
211	3	19	26661	HPRR3070385
212	3	20	26720	HPRR3460092
213	3	21	26747	HPRR3400059
214	3	22	26766	HPRR3000234
215	3	23	26789	HPRR3210135
216	3	24	26904	HPRR3420256
217	3	25	26951	HPRR2552223
218	3	26	27040	HPRR3050221
219	3	27	27052	HPRR2540601
220	3	28	27111	HPRR3140706
221	3	29	27255	HPRR230418
222	3	30	27698	HPRR3210127
223	3	31	27780	HPRR3310300
224	3	32	28271	HPRR3500398
225	3	33	28459	HPRR3020218
226	3	34	28564	HPRR3390107
227	3	35	28645	HPRR3480071
228	3	36	28843	HPRR3010234
229	3	37	28951	HPRR3580152
230	3	38	29090	HPRR3540053
231	3	39	29105	HPRR3470199
232	3	40	29186	HPRR3340279
233	3	41	29188	HPRR3480041
234	3	42	29226	HPRR3140641
235	3	43	29274	HPRR1140021
236	3	44	29277	HPRR3420383
237	3	45	29300	HPRR3500227
238	3	46	29506	HPRR3430094
239	3	47	30126	HPRR2850679
240	3	48	30149	HPRR3280420
241	3	49	30509	HPRR3460017
242	3	50	30512	HPRR3460376
243	3	51	30605	HPRR3400060
244	3	52	31180	HPRR3500251
245	3	53	31321	HPRR3320021
246	3	54	31353	HPRR3460022
247	3	55	31680	HPRR3320150
248	3	56	31838	HPRR3730670
249	3	57	31967	HPRR2551553
250	3	58	32009	HPRR3300056
251	3	59	32204	HPRR3460813
252	3	60	32216	HPRR3390003
253	3	61	32286	HPRR3460552
254	3	62	32337	HPRR3730056
255	3	63	32382	HPRR3720599
256	3	64	32472	HPRR3730550
257	3	65	32491	HPRR3400251
258	3	66	32517	HPRR3730387
259	3	67	32615	HPRR3760334
260	3	68	32738	HPRR3790027
261	3	69	32851	HPRR3460530
262	3	70	32864	HPRR3780077
263	3	71	32952	HPRR3790142
264	3	72	33086	HPRR3300084
265	3	73	33170	HPRR3680222
266	3	74	33192	HPRR3790690
267	3	75	33396	HPRR3760818
268	3	76	33584	HPRR3470137
269	3	77	33907	HPRR3470014
270	3	78	33942	HPRR3141049
271	3	79	34253	HPRR3070024
272	3	80	34409	HPRR3730635
273	3	81	34473	HPRR3730263
274	3	82	34643	HPRR3760797
275	3	83	34705	HPRR3890580
276	3	84	34745	HPRR3730007
277	3	85	34772	HPRR3140639
278	3	86	35108	HPRR3760638
279	3	87	35285	HPRR3760711
280	3	88	35700	HPRR3870215
281	3	89	36180	HPRR3790916
282	3	90	36223	HPRR3730443
283	3	91	36226	HPRR3860253
284	3	92	36248	HPRR3410133
285	3	93	36270	HPRR3000044
286	3	94	36280	HPRR3890398
287	3	95	36383	HPRR3890111
288	3	96	36741	HPRR3720498
289	4	1	36809	HPRR3720315
290	4	2	36819	HPRR3870276
291	4	3	36977	HPRR4030048
292	4	4	37086	HPRR4030493
293	4	5	37133	HPRR4090049
294	4	6	37367	HPRR3560056

ID	Plate	Pos	HPRA	HPRR
295	4	7	37393	HPRR4010009
296	4	8	37485	HPRR4110248
297	4	9	37686	HPRR3420566
298	4	10	37906	HPRR3940247
299	4	11	38093	HPRR3830208
300	4	12	38125	HPRR3340204
301	4	13	38166	HPRR4030190
302	4	14	38203	HPRR4080101
303	4	15	38221	HPRR3970185
304	4	16	38258	HPRR3760798
305	4	17	38302	HPRR4090121
306	4	18	38347	HPRR3890406
307	4	19	38369	HPRR4130031
308	4	20	38404	HPRR4120351
309	4	21	38624	HPRR3050138
310	4	22	38710	HPRR3670067
311	4	23	38948	HPRR3460850
312	4	24	38954	HPRR4040106
313	4	25	38991	HPRR3280403
314	4	26	39022	HPRR3970188
315	4	27	39102	HPRR3120222
316	4	28	39151	HPRR3420747
317	4	29	39214	HPRR4060033
318	4	30	39434	HPRR4050488
319	4	31	39467	HPRR3890790
320	4	32	39547	HPRR4110051
321	4	33	39732	HPRR4110218
322	4	34	40113	HPRR4150088
323	4	35	40148	HPRR4200014
324	4	36	40149	HPRR4200020
325	4	37	21971	HPRR2760283
326	4	38	41174	HPRR4280338
327	4	39	41319	HPRR4280292
328	4	40	41403	HPRR4290455
329	4	41	41644	HPRR4290533
330	4	42	41739	HPRR4160573
331	4	43	41771	HPRR4130094
332	4	44	41958	HPRR4320117
333	4	45	41980	HPRR4160920
334	4	46	42022	HPRR4050555
335	4	47	42027	HPRR4290309
336	4	48	42067	HPRR4160359
337	4	49	42002	HPRR4280358
338	4	50	40364	HPRR3880320
339	4	51	42207	HPRR4340274
340	4	52	42491	HPRR4030495
341	4	53	42587	HPRR4290352
342	4	54	40464	HPRR4160972
343	4	55	40475	HPRR4160709
344	4	56	40538	HPRR3970190
345	4	57	43203	HPRR4180818
346	4	58	43333	HPRR4080085
347	4	59	43352	HPRR4190138
348	4	60	43428	HPRR4170045
349	4	61	43819	HPRR2870030
350	4	62	40626	HPRR4110319
351	4	63	44067	HPRR3500150
352	4	64	44073	HPRR4030136
353	4	65	44201	HPRR3860339
354	4	66	44246	HPRR4190832
355	4	67	44278	HPRR4340196
356	4	68	44396	HPRR2920354
357	4	69	44522	HPRR4420250
358	4	70	44667	HPRR4110114
359	4	71	44668	HPRR4090131
360	4	72	45141	HPRR4280017
361	4	73	45175	HPRR3140818
362	4	74	45201	HPRR4220450
363	4	75	45278	HPRR4220131
364	4	76	45376	HPRR4320252
365	4	77	45668	HPRR4160381
366	4	78	45794	HPRR4320454
367	4	79	45819	HPRR4370219
368	4	80	45910	HPRR4340082
369	4	81	45915	HPRR4180611
370	4	82	46422	HPRR4390331
371	4	83	46470	HPRR4220339
372	4	84	46563	HPRR4290077
373	4	85	40976	HPRR3100022
374	4	86	41026	HPRR3250136
375	4	87	43032	HPRR3790353
376	4	88	43072	HPRR4180186
377	4	89	45945	HPRR4290151
378	4	90	IFNW1	
379	4	91	IFNA2A	
380				HPRR4160191
381	4	93	His6ABP	
382	4	94	EMPTY	
383	4	95	anti-hIgG	
384	4	96	EBNA1	

F Code

F.1 Planar microarray data analysis

42k_analysis.R

```
library(tidyverse)
library(limma)
library(ggbeeswarm)
library(ggpubr)

file_names <- list.files('gpr/', pattern='.gpr')
file_paths <- paste('gpr/', file_names, sep="")


# All columns from STARTSCRIPT.R
gpr_columns <- c("Block", "Column", "Row", "Name", "ID", "X", "Y", "Dia.",
                 "F635 Median", "F635 Mean", "F635 SD", "F635 CV", "B635",
                 "B635 Median", "B635 Mean", "B635 SD", "B635 CV",
                 "% > B635+1SD", "% > B635+2SD", "F635 % Sat.", "F532 Median",
                 "F532 Mean", "F532 SD", "F532 CV", "B532", "B532 Median",
                 "B532 Mean", "B532 SD", "B532 CV", "% > B532+1SD",
                 "% > B532+2SD", "F532 % Sat.", "Ratio of Medians (635/532)",
                 "Ratio of Means (635/532)", "Median of Ratios (635/532)",
                 "Mean of Ratios (635/532)", "Ratios SD (635/532)",
                 "Rgn Ratio (635/532)", "Rgn R2 (635/532)", "F Pixels",
                 "B Pixels", "Circularity", "Sum of Medians (635/532)",
                 "Sum of Means (635/532)", "Log Ratio (635/532)",
                 "F635 Median - B635", "F532 Median - B532", "F635 Mean - B635",
                 "F532 Mean - B532", "F635 Total Intensity",
                 "F532 Total Intensity", "SNR 635", "SNR 532", "Flags",
                 "Normalize", "Autoflag")

# My subset of columns
gpr_subset <- c("Name", "ID",
                "F635 Median", "F635 Mean", "F635 SD",
                "B635 Median", "B635 Mean", "B635 SD",
                "F532 Median", "F532 Mean", "F532 SD",
                "B532 Median", "B532 Mean", "B532 SD",
                "F Pixels", "B Pixels", "Circularity",
                "F635 Median - B635", "F532 Median - B532",
                "Flags")

# Easy variable names for subset of columns (snake case)
gpr_colnames <- c("name", "id",
                  "f635_median", "f635_mean", "f635_sd",
                  "b635_median", "b635_mean", "b635_sd",
                  "f532_median", "f532_mean", "f532_sd",
                  "b532_median", "b532_mean", "b532_sd",
                  "f_pixels", "b_pixels", "circularity",
                  "fb635_median", "fb532_median",
                  "flags")

## DATA IMPORT
gpr = tibble()
pools = paste("pool", 1:4, sep="_")
sets = paste("set", 1:2, sep="_")

for(i in 1:8) {
  slide <- read.maimages(file_paths[i], source="genepix.median",
```

```
                          other.columns=gpr_subset)$other
  slide <- as.data.frame(slide)
  colnames(slide) = gpr_colnames

  slide <- as_tibble(slide) %>%
    mutate(pool = pools[ceiling(i/2)],
           set = sets[2-i%%2])

  gpr <- rbind(gpr, slide)
}

## FILTERING AND DB%
db <- read_tsv("data/fragment_rsd.txt") %>%
  select(PrEST, `%DB`) %>%
  rename(name = PrEST, db = `%DB`)

d_clean <- gpr %>%
  # Incorporate %DB
  left_join(db, by = "name") %>%
  # Filter rows that are flagged, non-HPPR, too small or too weak
  filter(flags == 0,
         str_starts(name, "HPRR"),
         f_pixels >= 30,
         f532_median - b532_median > 10 * b532_sd) %>%
  # If there are replicates, only keep the strongest one
  arrange(name, pool, desc(f532_median)) %>%
  distinct(name, pool, .keep_all = T)


## TRANSFORM AND DEFINE SIGNAL
calc_stats <- function(d){
  mfi <- d$fb635_median
  mn <- mean(mfi)
  sd <- sd(mfi)
  d_new <- d %>%
    mutate(mfi = mfi,
           sds = (mfi-mn) / sd)
  return(d_new)}

# Use per glass stats
d_rank = tibble()
for(p in pools){
  for(s in sets){
    glass <- calc_stats(filter(d_clean, pool == p, set == s))
    d_rank <- rbind(d_rank, glass)}}
d_rank <- d_rank %>%
  select(name, db, pool, set, sds, everything()) %>%
  arrange(desc(sds))

## Incorporate gene info
d_42k <- as_tibble(read.delim("data/42k_array.txt")) %>%
  rename(name = PrEST)

d_id <- d_rank %>%
  left_join(d_42k, by = "name")
d_id <- d_id %>%
  arrange(pool, desc(sds)) %>%
  mutate(prank = c(1:nrow(filter(d_id, pool == "pool_1")),
```

```
                    1:nrow(filter(d_id, pool == "pool_2")),
                    1:nrow(filter(d_id, pool == "pool_3")),
                    1:nrow(filter(d_id, pool == "pool_4")))) %>%
  # Remove rows with no Uniprot ID
  filter(Uniprot != "") %>%
  arrange(desc(sds)) %>%
  # Keep all except antigen names
  select(name, db, pool, set, sds, prank, Gene, Gene.desc, Uniprot,
         everything(), -Ag.name)

## Write Uniprot IDs
# to_write <- filter(d_id, sds > 4)
# write_csv(as.data.frame(to_write$Uniprot), col_names = F,
#           file = paste("data/uniprot_id_",
#                        format(Sys.time(),"%y%m%d_%H%M%S"),
#                        ".txt", sep = ""))
# -> then manually obtain GO-list from Uniprot and save as uniprot_GO.tsv

## Incorporate downloaded GO and flag keywords
go <- read_tsv("data/uniprot_GO.tsv") %>%
  rename(Uniprot = Entry)

# Create regexp filters based on keywords
kw_bioprocess <- paste(
  read_delim("keywords/keywords_bioprocess.txt", "\t", col_names=F)$X1,
  collapse = "|")
kw_cellcomp <- paste(
  read_delim("keywords/keywords_cellcomp.txt", "\t", col_names=F)$X1,
  collapse = "|")

d_go <- d_id %>%
  left_join(go, by = "Uniprot") %>%
  arrange(desc(sds)) %>%
  # kw1: immunological function
  # kw2: extracellular/membrane localization
  mutate(kw1 = str_detect(`Gene ontology (biological process)`,kw_bioprocess),
         kw2 = str_detect(`Gene ontology (cellular component)`,kw_cellcomp)) %>%
  select(name, db, sds, prank, pool, kw1, kw2, Gene,
         Uniprot, `Protein names`, names(go),
         everything(), -Gene.desc)

# Subset all >4 sd
d_sub <- filter(d_go, sds>4)

# Subset kw
d_kw_and <- filter(d_sub, kw1 == TRUE & kw2 == TRUE)
d_kw_or <- filter(d_sub,
                  !(kw1 == TRUE & kw2 == TRUE) & (kw1 == TRUE | kw2 == TRUE))

## Write results
# write_csv(d_sub, file = paste("results",format(Sys.time(),"%y%m%d_%H%M%S"),
#                               ".txt", sep = ""))

################################################################################

## Write de-duplicated results
d_dedup <- d_sub %>%
  add_count(name) %>%
```

```
  distinct(name, .keep_all = TRUE)
# write_tsv(d_dedup, file = paste("dedup_results",
#                                 format(Sys.time(),"%y%m%d_%H%M%S"),".txt",
#                                 sep = ""))

## Write gene list for antigen collection
# gene_list <- str_split(d_sub$Gene, ";")
# gene_vector <- unlist(gene_list)
# write(gene_vector, "42k_genes.txt")
```

42k_plot.R

```
## PLOTS
# Which data structure will be used for plotting?
###############
d_plot <- d_sub
###############

# Custom colorblind palette
mypal <- c("#0072B2", "#E69F00", "#009E73", "#CC79A7",
           "#D55E00", "#F0E442", "#56B4E9", "#999999")

# Index pools and count occurence
pool_index <- c(1:nrow(filter(d_plot, pool == "pool_1")),
                1:nrow(filter(d_plot, pool == "pool_2")),
                1:nrow(filter(d_plot, pool == "pool_3")),
                1:nrow(filter(d_plot, pool == "pool_4")))
(count_df <- count(d_plot, pool, set))

# Plot fall-off for each pool
ggplot(arrange(d_plot, pool, desc(sds))) +
  geom_point(aes(y = sds,
                 x = pool_index,
                 color = pool,
                 shape = set),
             size = 1.5) +
  geom_hline(yintercept = 4)

# Boxplot of pools and sets
ggplot(d_plot, aes(x = pool, y = sds)) +
  geom_boxplot(outlier.shape = NA, aes(fill = set)) +
  labs(y = "SDs", x = "Pool") +
  geom_quasirandom(dodge.width=0.8, size = 0.9, aes(group = set)) +
  geom_text(inherit.aes = F,
            stat="count",
            aes(x = pool,label=..count.., group = set, color = set),
            y = 2.5,
            position = position_dodge(width = 0.8)) +
  theme_bw() +
  scale_x_discrete(labels = factor(1:4)) +
  scale_color_manual(values=mypal) +
  scale_fill_manual(values=mypal)

# Wilcox difference between sets for every pool
ggplot(d_plot, aes(x = set, y = sds, fill = set)) +
  geom_boxplot() +
  facet_wrap(~ pool, ncol = 2) +
  geom_quasirandom(dodge.width=0.8) +
  stat_compare_means(label.y = 75)

```

```r
# Plot xy-correlation
ggplot(d_plot, aes(y=sds, x=f635_median, color = pool, shape = set)) +
  geom_point(size=2) +
  geom_line(aes(group = interaction(pool,set))) +
  labs(y = "SDs", x = "MFI", shape = "Set", color = "Pool") +
  theme_bw() +
  scale_shape_discrete(labels = factor(1:2)) +
  scale_color_manual(labels = factor(1:4),values=mypal)
```

F.2 Compiling antigen panel information for localization

panel_position_input.R

```r
library(tidyverse)

## Prerequisite step
# LIMS: 1) Input list of genes from 42k results and literature studies
#       2) Export df containing HPRR and HPRA, save as lims_genes.xls

# Here we merge those gene lists, remove rows without HPRA and de-duplicate HPRR
lims_42k_hpra <- read_tsv("lims_42k_genes.xls") %>%
  mutate(origin = "42k_genes")
lims_lit_genes <- read_tsv("lims_lit_genes.xls") %>%
  mutate(origin = "lit_genes")

hprr2collect <- rbind(lims_42k_hpra, lims_lit_genes) %>%
  select(-X4) %>%
  filter(`Ag name` != "") %>%
  distinct(PrEST, .keep_all = T)


# This goes to position-finding script
write_tsv(hprr2collect, "AK_covid.tsv")
```

panel_sort_output.R

```r
library(tidyverse)

# Import hprr2gene
hprr2gene <- as_tibble(read.delim("../42k/data/42k_array.txt")) %>%
  select(PrEST, Gene, Gene.desc)

# Make position df with gene info
plock <- read_tsv("AK_covid_pos.txt") %>%
  # Remove mostly empty columns occupying namespace
  select(-Gene, -PrEST.1) %>%
  left_join(hprr2gene, by = "PrEST") %>%
  select(PrEST, Ag.name, Gene, Gene.desc, everything())

# Make position df without gene info
hprr2pos <- plock %>%
  select(-Gene, -Gene.desc)

pos_info <- names(hprr2pos)[2:length(hprr2pos)]


# 1) 42k hpra
d_42k <- read_tsv("42k_dedup.txt") %>%
  # Rename "name" -> "PrEST
  rename(PrEST = name)
# This df has HPRA as unique key, meaning all other columns may be copied for some rows
```

```
d_42k_pos <- d_42k %>%
  inner_join(hprr2pos, by = "PrEST") %>%
  distinct(Ag.name, .keep_all = T) %>%
  mutate(prio = "1_42k") %>%
  select(prio, PrEST, db, sds, Gene, pos_info)

# 2) Literature genes
d_lit <- read_tsv("lims_lit_genes.xls") %>%
  filter(`Ag name` != "") %>%
  select(-X4, -`Ag name`)
d_lit_pos <- d_lit %>%
  inner_join(hprr2pos, by = "PrEST") %>%
  mutate(sds = NA, db = NA) %>%
  mutate(prio = "1_lit") %>%
  select(prio, PrEST, db, sds, Gene, pos_info)

# 3) Fill upp remainder of panel on HPRR that:
# - were not reactive themselves
# - represents genes that showed other reactive HPRR
# - represents genes that were immunologically relevant and extracellularly accessible
# - are sorted in descending signal strength
bonus_genes <- unique(filter(d_42k, kw1 == T & kw2 == T)$Gene)
bonus_genes_sds <- d_42k %>%
  filter(kw1 == T & kw2 == T) %>%
  select(Gene, sds) %>%
  distinct(Gene, .keep_all=T)
bl <- str_split(plock$Gene,";") %in% bonus_genes
d_bonus_pos <- plock[bl,] %>%
  left_join(bonus_genes_sds, by = "Gene") %>%
  mutate(db = NA) %>%
  mutate(prio = "2_bonus") %>%
  select(prio, PrEST, db, sds, Gene, pos_info) %>%
  arrange(desc(sds))

final <- rbind(d_42k_pos, d_lit_pos, d_bonus_pos) %>%
  arrange(prio,
          ProjBox.4,
          ProjBox.3,
          ProjBox.2,
          ProjBox.1)


## WRITE RESULTS
# write_tsv(d_lit_pos, file = paste("plock_lit_",
# format(Sys.time(),"%y%m%d_%H%M%S"),".xls", sep = ""), na = "")
```

F.3 Pre-run assay tests data analysis

test_analysis.R

```
library(tidyverse)

# Read well information
wells <- read_tsv("AK_wells.txt")

# Tidy count data
count <- read_tsv("bead_count.txt", col_names = F) %>%
  mutate(well = 1:21) %>%
  select(well, everything()) %>%
```

```
  pivot_longer(!well, names_to = "id", values_to = "count") %>%
  mutate(id = rep(1:384, times = 21))

# Tidy MFI data
mfi <- read_tsv("mfi.txt", col_names = F) %>%
  mutate(well = 1:21) %>%
  select(well, everything()) %>%
  pivot_longer(!well, names_to = "id", values_to = "mfi") %>%
  mutate(id = rep(1:384, times = 21))

# Join count, MFI and well data
d <- count %>%
  mutate(mfi = mfi$mfi) %>%
  left_join(wells, by = "well") %>%
  # Add plate info
  arrange(id) %>%
  mutate(plate = c(rep(1,8064/4), rep(2,8064/4),
                   rep(3,8064/4), rep(4,8064/4)))

bead_info <- read_tsv("bead_info.txt") %>%
  mutate(id = as.integer(id))

################################################################################


################################## Check bead count
# Plot bead count for all id
plotdata = d
ggplot(data = plotdata) +
  geom_boxplot(aes(x = id, y = count, group = id, color = factor(plate)),
               outlier.shape = NA) +
  scale_color_discrete(name = "Plate") +
  geom_hline(yintercept = 30, color = "red") +
  theme_bw() +
  labs(y = "Bead count", x = "Bead ID")


# Summarize bead count per plate
summarise(group_by(plotdata,plate), median = median(count), mean = mean(count),
          min = min(count), max = max(count))

# Summarize bead count per id
summarise(group_by(plotdata,id), median = median(count), mean = mean(count),
          min = min(count), max = max(count)) %>%
  arrange(median)

################################## Check coupling test

coupling_sum <- summarise(
  group_by(
    filter(d, test=="Coupling"),
    id, control),
  mean_MFI = mean(mfi))

coupling <- ungroup(coupling_sum) %>%
  add_column(plate = c(rep(1,192), rep(2,192), rep(3,192), rep(4,192))) %>%
  mutate(ctrl = if_else(id %in% c(378:379,381:384) & is.na(control),
                        id, as.integer(0)))
```

```
68
69  ggplot() +
70    geom_col(data = filter(coupling, is.na(control)),
71             aes(x = id, y = mean_MFI,
72                 fill = if_else(ctrl != 0, factor(ctrl), NULL)), width=1) +
73    geom_col(data = filter(coupling, control == TRUE),
74             aes(x = id, y = mean_MFI), fill = "black", width=1) +
75    theme_bw() +
76    labs(y = "mean(MFI)", x = "Bead ID", fill = "Antigen") +
77    scale_fill_discrete(labels = c("IFNW1", "IFNA2A", "His6ABP", "EMPTY",
78                                   "anti-hIgG", "EBNA1", "PrEST"))
79
80  ################################## Check sample test
81
82  samp_sum <- summarise(
83    group_by(
84      filter(d, test=="Sample"),
85      id, control),
86    mean_MFI = mean(mfi))
87
88  samp <- samp_sum %>%
89    mutate(ctrl = if_else(id %in% c(381:384) & is.na(control), id, as.integer(0)))
90
91  ggplot() +
92    geom_col(data = filter(samp, is.na(control)),
93             aes(x = id, y = mean_MFI,
94                 fill = if_else(ctrl != 0, factor(ctrl), NULL)), width=1) +
95    geom_col(data = filter(samp, control == TRUE),
96             aes(x = id, y = mean_MFI), fill = "black", width=1) +
97    theme_bw() +
98    labs(x = "Bead ID", y = "mean(MFI)", fill = "Antigen") +
99    scale_fill_discrete(labels =
100                          c("His6ABP", "EMPTY", "anti-hIgG", "EBNA1", "Other"))
101
102
103 ################################## Check antigen test
104
105 spec_control <- d %>%
106   filter(test == "Specific", control == TRUE) %>%
107   mutate(group = 1)
108 spec_control_rep <- spec_control
109 for(i in 2:6){
110   spec_control <- d %>%
111     filter(test == "Specific", control == TRUE) %>%
112     mutate(group = i)
113   spec_control_rep <- full_join(spec_control_rep, spec_control)
114 }
115
116 ag <- d %>%
117   filter(test == "Specific", is.na(control)) %>%
118   full_join(spec_control_rep) %>%
119   group_by(group, id, control) %>%
120   summarise(mean_MFI = mean(mfi)) %>%
121   ungroup() %>%
122   mutate(ctrl = if_else(id %in% c(381:384) & is.na(control),
123                         id, as.integer(0))) %>%
124   left_join(distinct(select(d, group, target_id)), by = "group") %>%
125   left_join(bead_info, by = "id")
```

```
126
127 ggplot(data = filter(ag),
128        aes(x = id, y = mean_MFI, fill = control)) +
129   geom_col(position = position_stack(reverse = T),width=1) +
130   # geom_col(data = filter(ag, control == TRUE),
131   #          aes(x = id, y = mean_MFI), color = "black") +
132   geom_text(aes(x = id, y = mean_MFI,
133                 label = if_else(mean_MFI > 2000 & is.na(control),
134                                 bead_info, "")),
135             size = 3, hjust = 0.8) +
136   facet_wrap(~target_id, ncol = 2, scales = "free") +
137   theme_bw() +
138   labs(x = "Bead ID", y = "mean(MFI)", fill = "Sample (stacked)") +
139   scale_fill_discrete(labels = c("Negative control", "Test duplicate"))
```

F.4 Finding appropriate control plates

MSC_6_wrangle.R

```
1  library(tidyverse)
2
3  df <- read_tsv("MSC_6.tsv")
4
5  df_all <- df %>%
6    select("sample_name", "class", "gender", "Plate96", "Well96", "Pos96",
7           "Date of birth", "Colection Date") %>%
8    rename(date_sampled = `Colection Date`) %>%
9    mutate(date_born = as.Date(paste("19", `Date of birth`, sep=""),
10                              format = "%Y%m%d"),
11          sample_age = lubridate::time_length(difftime(date_sampled, date_born),
12                                              "years"),
13          half_plate = if_else(Pos96 <= 96/2,
14                               paste(Plate96,1,sep="_"),
15                               paste(Plate96,2,sep="_"))) %>%
16   arrange(Plate96, Pos96)
17
18 by_plate <- group_by(df_all, Plate96)
19 by_half_plate <- group_by(df_all, half_plate)
20
21 plate <- summarise(by_plate,
22                     n_HC = sum(class == "Control", na.rm=T),
23                     n_MS = sum(class == "Case", na.rm=T),
24                     prop_HC = n_HC/(n_HC+n_MS),
25                     n_M = sum(gender == "M", na.rm=T),
26                     n_F = sum(gender == "F", na.rm=T),
27                     prop_M = n_M/(n_M+n_F),
28                     n_HC_M = sum(gender == "M" & class == "Control", na.rm=T),
29                     n_HC_F = sum(gender == "F" & class == "Control", na.rm=T),
30                     prop_HC_M = n_HC_M/(n_M+n_F),
31                     med_age = median(sample_age, na.rm=T))
32 half_plate <- summarise(by_half_plate,
33                       n_HC = sum(class == "Control", na.rm=T),
34                       n_MS = sum(class == "Case", na.rm=T),
35                       prop_HC = n_HC/(n_HC+n_MS),
36                       n_M = sum(gender == "M", na.rm=T),
37                       n_F = sum(gender == "F", na.rm=T),
38                       prop_M = n_M/(n_M+n_F),
39                       n_HC_M = sum(gender == "M" & class == "Control", na.rm=T),
40                       n_HC_F = sum(gender == "F" & class == "Control", na.rm=T),
```

```
                      prop_HC_M = n_HC_M/(n_M+n_F),
                      med_age = median(sample_age, na.rm=T))

ggplot(plate, mapping = aes(n_HC, n_M)) +
  geom_point() +
  geom_text(aes(label=Plate96), hjust=-0.5, vjust=-0.5)

ggplot(half_plate, mapping = aes(n_HC, n_M)) +
  geom_point() +
  geom_text(aes(label=half_plate), hjust=-0.5, vjust=-0.5)

# Whole of plate 14 and second half of plate 16

d_14 <- df_all %>%
  filter(Plate96 == 14, Pos96 < 93) %>%
  rename(age = sample_age, source_pos = Pos96) %>%
  add_column(source_plate = "EIMS_14") %>%
  mutate(class = if_else(class == "Case", "MS", "HC")) %>%
  select(sample_name, class, gender, age, source_plate, source_pos) %>%
  arrange(source_pos)

d_16 <- df_all %>%
  filter(Plate96 == 16, Pos96 > 49) %>%
  rename(age = sample_age, source_pos = Pos96) %>%
  add_column(source_plate = "EIMS_14") %>%
  mutate(class = if_else(class == "Case", "MS", "HC")) %>%
  select(sample_name, class, gender, age, source_plate, source_pos) %>%
  arrange(source_pos)

# write_tsv(d_14, "p14_info.txt")
# write_tsv(d_16, "p16_info.txt")
```

F.5 SBA data analysis

sba_analysis.R

```
library(tidyverse)
library(ggbeeswarm)

########### Position information
## Convert 384 to 96 plate
# Greiner384 read up->down, left->right with plates 1-4 in quadrants 2,3,1,4
greiner2pcr_plate <- rep(c(rep(c(1,2),8), rep(c(3,4),8)),12)

greiner2pcr_well_seq <- rep(c(1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8),2)
greiner2pcr_well <- c()
for(i in 0:11){
  greiner2pcr_well <- c(greiner2pcr_well, greiner2pcr_well_seq+i*8)
}

greiner2pcr_id <- (greiner2pcr_plate-1)*96+greiner2pcr_well

## Source positions

source2sba <- read_tsv("source2sba.txt")

########### Sample information
## eims

```

```
eims_14 <- read_tsv("p14_info.txt")
eims_16 <- read_tsv("p16_info.txt") %>%
  mutate(source_plate = "EIMS_16")

eims <- eims_14 %>%
  add_row(eims_16) %>%
  add_column(phase = NA)

## covid

dspati_info <- read_tsv("dspati_info.txt") %>%
  rename("sample_name" = sample_id)

dspati <- read_tsv("dspati.txt") %>%
  left_join(dspati_info, by = "sample_name") %>%
  mutate(sample_name = str_pad(sample_name, 3, side = "left", pad = "0"))


# Merge covid, eims and positions
all_samples <- full_join(dspati,eims)

sba_id_info <- tibble("sba_id" = greiner2pcr_id,
                      "sba_plate" = greiner2pcr_plate,
                      "sba_pos" = greiner2pcr_well,
                      "greiner_pos" = 1:384) %>%
  arrange(sba_id) %>%
  left_join(source2sba, by = "sba_id") %>%
  left_join(all_samples, by = c("source_plate", "source_pos")) %>%
  mutate(class = if_else(source_plate == "control", "control",
                         if_else(source_plate == "empty", "empty", class)),
         sample_name = if_else(source_plate %in% c("control", "empty"),
                               source_plate, sample_name),
         simple_class = if_else(class == "covid",
                                "covid",
                                if_else(class == "MS" | class == "HC",
                                        "no_covid",
                                        class)),
         sample_id = paste(sample_name, phase, sep ="_"))


# Import Luminex data and transform layout

mfi <- read_tsv("mfi.txt", col_names = FALSE) %>%
  add_column(greiner_pos = 1:384) %>%
  pivot_longer(X1:X384, names_to = "bead_id", values_to = "mfi") %>%
  mutate(bead_id = rep(1:384, 384))

count <- read_tsv("count.txt", col_names = FALSE) %>%
  add_column(greiner_pos = 1:384) %>%
  pivot_longer(X1:X384, names_to = "bead_id", values_to = "count") %>%
  mutate(bead_id = rep(1:384, 384))

# Merge mfi and count
luminex <- count %>%
  add_column(mfi = mfi$mfi)


######### Get antigen info from different sources
```

```r

# from PrEST to uniprot
d_42k <- as_tibble(read.delim("42k_array.txt")) %>%
  rename("uniprot" = Uniprot) %>%
  select(PrEST, uniprot)
# export and get uniprot info

# from uniprot to GO
go <- read_tsv("go.tsv") %>%
  select(-1)
names(go) <- c("uniprot", "protein", "go_process", "go_cellcomp")

# keywords
kw_bioprocess <- paste(
  read_delim("keywords/keywords_bioprocess.txt", "\t", col_names=F)$X1,
  collapse = "|")
kw_cellcomp <- paste(
  read_delim("keywords/keywords_cellcomp.txt", "\t", col_names=F)$X1,
  collapse = "|")

antigen_info <- read_tsv("AK_SBA.txt") %>%
  left_join(d_42k, by = "PrEST") %>%
  left_join(go, by = "uniprot") %>%
  mutate(bead_info = if_else(is.na(gene),
                             paste(bead_id, ag_name, sep="_"),
                             paste(bead_id, gene, sep="_")),
         bead_info = if_else(str_length(bead_info) > 20,
                                      paste(str_sub(bead_info, 1, 8), "..."),
                                      bead_info),
         kw1 = str_detect(go_process, kw_bioprocess),
         kw2 = str_detect(go_cellcomp, kw_cellcomp))



# merge information
d_full <- luminex %>%
  left_join(sba_id_info, by = "greiner_pos") %>%
  left_join(antigen_info, by = "bead_id")

######### Normalize by sample
# MAD (Median Absolute Deviation) = Median of absolute deviations from median

# MAD assumption does not apply for positive control beads
d_calc_stats <- d_full %>%
  filter(bead_id %in% 1:382,
         count > 30)

sample_metrics <-  d_calc_stats%>%
  group_by(sba_id) %>%
  summarise(mfi_sample_median = median(mfi),
            mfi_sample_mad = mad(mfi, constant = 1))

ag_metrics <- d_calc_stats %>%
  group_by(bead_id) %>%
  summarise(mfi_ag_median = median(mfi),
            mfi_ag_mad = mad(mfi, constant = 1))

d_norm <- d_calc_stats %>%
```

```
  left_join(sample_metrics, by = "sba_id") %>%
  mutate(mfi_sub_sample = mfi - mfi_sample_median,
         mads_sample = mfi_sub_sample / mfi_sample_mad) %>%
  left_join(ag_metrics, by = "bead_id") %>%
  mutate(mfi_sub_ag = mfi - mfi_ag_median,
         mads_ag = mfi_sub_ag / mfi_ag_mad) %>%
  mutate(mfi_sub_both = mfi - mfi_sample_median - mfi_ag_median,
         mads_both = mfi_sub_both / (mfi_sample_mad * mfi_ag_mad))

######### Trimming
## Exclusion criteria
# bead count < 30 (facility standard is remove <16, flag <30)
# set aside PCR-negative patients
# one EIMS sample has NA class and is therefore excluded via simple_class

d_trimmed <- d_norm %>%
  filter(!(sample_name %in% c("009","059","063","085")),
         simple_class %in% c("covid", "no_covid"))

d_called <- d_trimmed %>%
  mutate(reactive = mads_both > 1.5)

n_reac_per_sample <- d_called %>% group_by(sba_id) %>% count(reactive) %>%
  pivot_wider(names_from = reactive, values_from = n) %>%
  rename(n = `TRUE`, N = `FALSE`) %>%
  select(sba_id, n, N)
sba_id_info_reac <- sba_id_info %>%
  left_join(n_reac_per_sample, by = "sba_id")


######### Comparison full phase to control ###############

# The data of each phase are compared separately to the eims controls
for(phase_counter in 1:3){
  d_comp_current_phase <- d_called %>%
    filter(phase == phase_counter | is.na(phase))

  cont <- count(d_comp_current_phase, bead_id, simple_class, reactive) %>%
    mutate(reactive = if_else(as.character(reactive) == "TRUE", "n", "N")) %>%
    pivot_wider(names_from = reactive, values_from = n, values_fill = 0) %>%
    select(bead_id, simple_class, n, N) %>%
    arrange(bead_id, simple_class)

  # Fisher
  f_val <- c()
  for(id in sort(unique(d_comp_current_phase$bead_id))){
    cont_table <- as.matrix(filter(cont, bead_id == id)[,3:4])
    dimnames(cont_table) = list(c("covid", "no_covid"), c("n", "N"))

    ftest <- fisher.test(x = cont_table, alternative = "greater")
    f_val <- c(f_val, ftest$p.value)
  }

  cont_wide <- cont %>%
    pivot_wider(names_from = simple_class,
                values_from = c(n, N),
                names_sep = "_") %>%
    mutate(larger_nN_in =
```

```
            if_else( (n_covid / (n_covid + N_covid)) >
                       (n_no_covid / (n_no_covid + N_no_covid)),
                     "covid", "no_covid"),
          f = f_val,
          #q = q_val,
          phase_comp = phase_counter) %>%
    mutate(n_shorthand =
             paste(n_covid, "/", N_covid, " ",
                   n_no_covid, "/", N_no_covid, sep ="")) %>%
    select(bead_id, phase_comp, f, #q,
           n_covid, N_covid, n_no_covid, N_no_covid, n_shorthand, larger_nN_in)

  if(phase_counter == 1){
    cont_all <- cont_wide
  }else{
    cont_all <- full_join(cont_all, cont_wide)
  }
}

antigen_info_f <- antigen_info %>%
  full_join(cont_all, by = "bead_id") %>%
  select(names(cont_all), db, bead_info, protein, kw1, kw2, go_process,
         everything()) %>%
  arrange(f)

d_comp <- d_called %>%
  left_join(cont_all, by = "bead_id") %>%
  select(sba_id, sample_id, mfi, mads_sample, mads_ag, mads_both,
         names(cont_all), db, bead_info, protein, kw1, kw2,
         go_process, everything())

sig_phase <- filter(antigen_info_f, f < 0.05) %>%
  select(bead_id, gene, kw1, kw2, protein, go_process,
         phase_comp, f, n_shorthand) %>%
  arrange(phase_comp) %>%
  pivot_wider(names_from = phase_comp,
              values_from = c(f,n_shorthand), names_prefix = "phase_") %>%
  select(bead_id, gene,
         f_phase_1, n_shorthand_phase_1,
         f_phase_2, n_shorthand_phase_2,
         f_phase_3, n_shorthand_phase_3,
         everything()) %>%
  arrange(is.na(f_phase_1), is.na(f_phase_2), is.na(f_phase_3),
          f_phase_1, f_phase_2, f_phase_3)

sig_beads <- unique(filter(antigen_info_f, f < 0.05)$bead_id)
sig_phase_full <- filter(antigen_info_f, bead_id %in% sig_beads) %>%
  select(bead_id, gene, kw1, kw2, protein, uniprot, go_process,
         phase_comp, f, n_shorthand) %>%
  arrange(phase_comp) %>%
  pivot_wider(names_from = phase_comp,
              values_from = c(f,n_shorthand), names_prefix = "phase_") %>%
  select(bead_id, gene,
         f_phase_1, n_shorthand_phase_1,
         f_phase_2, n_shorthand_phase_2,
         f_phase_3, n_shorthand_phase_3,
         everything()) %>%
  arrange(is.na(f_phase_1), is.na(f_phase_2), is.na(f_phase_3),
```

```
          f_phase_1, f_phase_2, f_phase_3)

# write_tsv(sig_phase_full,
#         paste("results_",format(Sys.time(),"%y%m%d_%H%M%S"),".txt",sep = ""))


########## Numbering

x <- d_called %>%
  filter(phase == 3,
         reactive == TRUE)

# 85080 obserations in covid patients
#   of which 2874 (3.4%) were reactive

# 43281 obs in phase 1
#   of which 1341 (3.1%) were reactive

# 22397 obs in phase 2
#   of which 828 (3.7%) were reactive

# 19402
#   of which 705 were reactive

######### Longitudinal macro

x <- d_called %>%
  filter(!is.na(phase)) %>%
  select(sample_name, phase, bead_id, reactive, mads_sample) %>%
  pivot_wider(names_from = phase, values_from = c(reactive,mads_sample))

x1 <- x %>%
  filter(!reactive_1, !is.na(reactive_2)) %>%
  mutate(increased = reactive_2)

count(x1, increased)
```

sba_plot_old.R

```
########## Look at control beads across all samples

# Points
ggplot(filter(d_full, bead_id %in% 381:384), aes(x=sba_id, y = mfi)) +
  geom_point(aes(color = source_plate)) +
  facet_wrap(~as.factor(bead_id), ncol=2)

# Boxplot
ggplot(filter(d, ag_name %in% c("anti-hIgG", "EBNA1", "EMPTY", "His6ABP")),
       aes(x=source_plate, y = mfi)) +
  geom_boxplot(aes(color = source_plate)) +
  facet_wrap(~ag_name, ncol=2) +
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank())


########## General
## Look at specific bead across all samples
ggplot(filter(d_comp, ag_name == "IFNA2A"), aes(x=sba_id, y = mfi)) +
  geom_point(aes(color = simple_class))
```

```
## Look at specific sample across all beads
ggplot(filter(d_norm, sba_id %in% 149:151), aes(x=bead_id, y = mfi)) +
  geom_col(aes(color = class)) +
  geom_hline(aes(yintercept = mfi_sample_median)) +
  geom_hline(aes(yintercept = mfi_sample_median + mfi_sample_mad),
             color = "red") +
  facet_wrap(~sba_id)


########## Normalization evaluation

ggplot(filter(d_norm, sba_id %in% 149:151), aes(x=bead_id, y = mfi)) +
  geom_col(aes(color = class)) +
  geom_hline(aes(yintercept = mfi_sample_median)) +
  geom_hline(aes(yintercept = mfi_sample_median + mfi_sample_mad),
             color = "red") +
  facet_wrap(~sba_id)

ggplot(filter(d,
              sample_name %in% 17:32,
              bead_id %in% 1:380),
       aes(x=bead_id, y = mfi)) +
  geom_col(aes(color = class)) +
  facet_wrap(~sample_name, ncol=4)

# The antigen-normalized median correlates to stickyness
ggplot(filter(by_ag, bead_id %in% 1:380),
       aes(x = antigen_info$db[1:380], y = mfi_ag_mad)) +
  geom_point() +
  geom_text(aes(label=ifelse(mfi_ag_mad > 100,as.character(bead_id),'')),
            hjust=0,vjust=0)

# Normalizing by antigen decreases this tendency
ggplot(filter(d, bead_id %in% 1:380),
       aes(x = db, y = mads)) +
  geom_jitter(position = "jitter")
ggplot(filter(d, bead_id %in% 1:380),
       aes(x = db, y = madsmads)) +
  geom_point(position = "jitter")


####### Big plot
d_matrix <- filter(d, bead_id %in% 1:380,
                   !(class %in% c("control", "empty"))) %>%
  mutate(sample_id = if_else(is.na(phase), sample_name,
                             paste(sample_name, phase, sep = "_"))) %>%
  select(sample_id, ag_name, madsmads) %>%
  pivot_wider(id_cols = c(sample_id, ag_name),
              names_from = ag_name,
              values_from = madsmads)
d_matrix_input <- as.matrix(select(d_matrix, -sample_id))
heatmap(d_matrix_input)

d_big <- d %>%
  filter(bead_id %in% 350:380) %>%
  mutate(reactive =
```

```
            mfi > 2000 |
            mads > 50)

ggplot(d_big %>%
         mutate(sample_id = if_else(is.na(phase),
                                    sample_name,
                                    paste(sample_name, phase, sep = "_"))),
       aes(x = sample_id, y = ag_name)) +
  geom_tile(aes(fill = reactive)) +
  theme(axis.text.x = element_text(size = 8, angle = 90, hjust = 0.95,
                                   vjust = 0.2),
        axis.text.y = element_text(size = 8))

########## Full length interferons
ggplot(filter(d, ag_name %in% c("IFNW1", "IFNA2A")), aes(x=sba_id, y = mads)) +
  geom_point(aes(color = class)) +
  facet_wrap(~ag_name, ncol=2)


########## Ordered fall-off of all reactivities
# mfi
ggplot(filter(d) %>% arrange(desc(mfi)), aes(x=1:nrow(d), y = mfi)) +
  geom_point(aes(color = bead_id == 383))


########## Antigen profile across samples + testing cut-off
d_plot <- d %>%
  filter(bead_id %in% 1:46| bead_id %in% 381:384) %>%
  mutate(reactive =
            mfi > 2000 |
            mads > 50)

# Variable grid Fall-off
ggplot(d_plot %>%
         pivot_longer(cols = c(mfi,mads,madsmads),
                      names_to = "var", values_to = "val") %>%
         mutate(var_f = factor(var, levels = c("mfi", "mads", "madsmads"))) %>%
         arrange(desc(val))) +
  geom_point(aes(x = 1:length(val), y = val, color = reactive)) +
  facet_wrap(bead_id ~ var_f, ncol = 6, scales = "free")

# Fall-off
ggplot(arrange(d_plot, desc(mfi))) +
  geom_violin(aes(x = 1:length(sba_id), y = madsmads, color = reactive)) +
  facet_wrap(~bead_id, ncol = 10, scales = "free_y")

# Boxplot
ggplot(d_plot) +
  geom_histogram(aes(y = madsmads_sd)) +
  geom_vline(xintercept = 2) +
  facet_wrap(~bead_id, ncol = 10)

# Samples
d_plot_samples <- d %>%
  filter(sba_id %in% unique(sba_id)[1:16]) %>%
  mutate(reactive = if_else(mfi > 2000, "mfi > 2000",
                            if_else(mads > 50, "mads > 50", "None")))
ggplot(d_plot_samples) +
```

```
  geom_point(aes(x = bead_id, y = mfi, color = reactive)) +
  facet_wrap(~sba_id, ncol = 8, scales = "free_y")


########## Various interferons
ggplot(d_ifn %>% filter(class %in% c("covid", "MS", "HC"))) +
  geom_quasirandom(dodge.width=0.8, size = 0.9,
                   aes(y = madsmads, x = class, color = bead_info)) +
  facet_wrap(~bead_info, ncol = 5, scales = "free_y")


######### Empty wells
ggplot(filter(d, class == "empty"), aes(x=bead_id, y = mfi)) +
  geom_col(aes(color = class)) +
  geom_text(aes(label=ifelse(mads > 50,as.character(ag_name),'')),
            hjust=0,vjust=0) +
  facet_wrap(~sba_plate, ncol=2)


########## Longitudinal
# Based on diff and baseline
ggplot(filter(d_follow, diff > 4, baseline < 5)) +
  geom_line(aes(y = mads, x = phase,
                color = interaction(ag_name, sample_name))) +
  theme(legend.position = "none")

# Based on genes
ggplot(filter(d_follow, diff < -50, str_detect(gene, "IFN")),
       aes(y = mads, x = phase)) +
  geom_line(aes(color = interaction(ag_name, sample_name))) +
  theme(legend.position = "none") #+
  geom_text(aes(label=ifelse(diff< -1 & phase == 2, as.character(ag_name),'')),
            hjust=0,vjust=0)


ggplot(filter(d_follow, diff < -500), aes(y = mads, x = phase)) +
  geom_line(aes(color = interaction(ag_name, sample_name))) +
  theme(legend.position = "none") #+
geom_text(aes(label=ifelse(diff< -1 & phase == 2, as.character(ag_name),'')),
          hjust=0,vjust=0)

################################################################################
## Findings

# HIF3A reactivity from 42k analysis pool 2 can be traced to patient
ggplot(filter(d, PrEST == "HPRR3140817"), aes(x=sba_id, y = sds)) +
  geom_point(aes(color = class)) +
  facet_wrap(~ag_name, ncol=2)
```

sba_plot_results.R

```
# Custom colorblind palette
mypal <- c("#0072B2", "#E69F00", "#009E73", "#CC79A7",
           "#D55E00", "#F0E442", "#56B4E9", "#999999")

# Show control beads
d_plot <- d_full %>%
  mutate(source_plate_f = factor(source_plate, levels = c("DSPATI_MIX_1",
                                                          "DSPATI_MIX_2",
                                                          "DSPATI_9",
```

```
                                                    "DSPATI_10",
                                                    "EIMS_14",
                                                    "EIMS_16",
                                                    "control",
                                                    "empty")))

ggplot(filter(d_plot, ag_name %in% c("anti-hIgG", "EBNA1", "EMPTY", "His6ABP")),
       aes(x=source_plate_f, y = mfi)) +
  geom_boxplot(aes(color = source_plate_f)) +
  facet_wrap(~ag_name, ncol=2) +
  theme_bw() +
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank()) +
  labs(y = "MFI", color = "Source plate") +
  scale_color_manual(values=mypal, labels = c("Phase 1, plate 1/2",
                                              "Phase 1, plate 2/2",
                                              "Phase 2",
                                              "Phase 3",
                                              "EIMS plate 14",
                                              "EIMS plate 16",
                                              "Commercial plasma",
                                              "Empty well"
                                              ))



## Per-sample normalization

# mfi
ggplot(filter(d_trimmed, bead_id %in% c(3, 5, 377, 378, 379, 381, 382)),
       aes(x = bead_info,
           y = mfi)) +
  geom_quasirandom(aes(color = simple_class)) +
  theme(axis.text.x = element_text(size = 8, angle = 45, hjust=0, vjust=0)) +
  xlab("") +
  ylab("MFI") +
  labs(color = "Sample type") +
  theme_bw() +
  scale_color_manual(values=mypal, labels = c("Covid", "No Covid"))

# mads
ggplot(filter(d_trimmed, bead_id %in% c(3, 5, 377, 378, 379, 381, 382)),
       aes(x = bead_info,
           y = mads_both)) +
  geom_quasirandom(aes(color = simple_class)) +
  geom_hline(yintercept = 1.5) +
  theme(axis.text.x = element_text(size = 8, angle = 45, hjust=0, vjust=0)) +
  xlab("") +
  ylab("MADs by antigen x sample") +
  labs(color = "Sample type") +
  theme_bw() +
  scale_color_manual(values=mypal, labels = c("Covid", "No Covid"))


## Called reactive
# How many reactive per sample?
reactive_per_sample <- d_called %>%
```

```r
  group_by(sba_id) %>%
  count(reactive) %>%
  filter(reactive == TRUE)

ggplot(reactive_per_sample) +
  geom_histogram(aes(x = n), bins = 14) +
  xlab("n of reactivities per sample") +
  theme_bw()

reactive_per_ag <- d_called %>%
  group_by(bead_id) %>%
  count(reactive) %>%
  filter(reactive == TRUE)

ggplot(reactive_per_ag) +
  geom_histogram(aes(x = n), bins = 32) +
  xlab("n of reactivities per antigen") +
  theme_bw()


########## Significant ag ############################
beads_relevant <- c(378,1,234,281,361,274,33,10)
beads_various <- c(279,254,44,236,154,25,347,321,336,271,166,177,327,68,16)
beads_excluded <- c(369,72,240,61,192,131,342,304,137,189,101)


# Facet
tag_sig_pati <- d_called %>%
  select(sample_name, phase, bead_id, reactive) %>%
  pivot_wider(names_from = phase, names_prefix = "phase_",
              values_from = reactive) %>%
  # If for the current antigen, the current patient is reactive in any phase
  mutate(sig_pat = if_else(phase_1|phase_2|phase_3, TRUE, FALSE)) %>%
  select(sample_name, bead_id, sig_pat)

plot_facet <- d_called %>%
  filter(bead_id %in% c(beads_relevant)) %>%
  left_join(tag_sig_pati, by = c("sample_name", "bead_id"))

ggplot(plot_facet,
       aes(x = factor(phase), y = mads_sample,
           color = factor(interaction(simple_class, reactive),
                          levels = c("covid.TRUE", "covid.FALSE",
                                     "no_covid.TRUE", "no_covid.FALSE")),
           group = simple_class)) +
  geom_quasirandom(dodge.width = 0.8) +
  geom_line(data = filter(plot_facet, sig_pat), alpha = 0.5, aes(group = sample_name)) +
  labs(x = "", y = "MADs by sample") +
  facet_wrap(~reorder(bead_info, bead_id), ncol = 2, scales = "free_y") +
  theme_bw() +
  theme(axis.text.y = element_text(size = 12),
        axis.title.y = element_text(size = 13)) +
  theme(legend.position = "none") +
  scale_x_discrete(labels =
                     c("Phase 1", "Phase 2", "Phase 3", "Control group")) +
  scale_color_manual(values=mypal)
```

```
######### IFN ##########################################################################

# Facet
tag_sig_pati <- d_called %>%
  select(sample_name, phase, bead_id, reactive) %>%
  pivot_wider(names_from = phase, names_prefix = "phase_",
              values_from = reactive) %>%
  # If for the current antigen, the current patient is reactive in any phase
  mutate(sig_pat = if_else(phase_1|phase_2|phase_3, TRUE, FALSE)) %>%
  select(sample_name, bead_id, sig_pat)

t1_ifn <- antigen_info %>%
  filter(str_detect(gene, "IFN"),
         !str_detect(gene, "IFNG"))

d_ifn <- d_called %>%
  filter(bead_id %in% c(32,193,328,379)) %>%
  left_join(tag_sig_pati, by = c("sample_name", "bead_id"))

ggplot(d_ifn,
       aes(x = factor(phase), y = mads_sample,
           color = factor(interaction(simple_class, reactive),
                          levels = c("covid.TRUE", "covid.FALSE",
                                     "no_covid.TRUE", "no_covid.FALSE")),
           group = simple_class)) +
  geom_quasirandom(dodge.width = 0.8) +
  geom_line(data = filter(d_ifn, sig_pat), alpha = 0.5, aes(group = sample_name)) +
  labs(x = "", y = "MADs by sample") +
  facet_wrap(~reorder(bead_info, bead_id), ncol = 2, scales = "free_y") +
  theme_bw() +
  theme(axis.text.y = element_text(size = 12),
        axis.title.y = element_text(size = 13)) +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Phase 1", "Phase 2", "Phase 3", "Control group"))  +
  scale_color_manual(values=mypal)
```