Plasmodium falciparum serology: A comparison of two protein production methods for analysis of antibody responses by protein microarray

Tate Oulton[1], Joshua Obiero[2], Isabel Rodriguez[3], Isaac Ssewanyana[4,5], Rebecca A. Dabbs[1], Bryan Greenhouse[3], Chris Drakeley[1], Phil L. Felgner[2], Will Stone[1], Kevin K.A. Tetteh[1]

[1] Department of Infection Biology, London school of Hygiene and Tropical Medicine, London, UK

[2] Department of Physiology and Biophysics, University of California, Irvine, Irvine, CA, USA

[3] University of California San Francisco, California, USA

[4] Infectious Diseases Research Collaboration, Kampala, Uganda

[5] Makerere University Collage of Health Sciences, Kampala, Uganda

Corresponding author: Kevin K.A. Tetteh**; kevin.tetteh@lshtm.ac.uk**

Keywords: protein microarray, reverse phase protein array, RPPA, malaria, Plasmodium

Abstract

The evaluation of protein antigens as putative serologic biomarkers of infection has increasingly shifted to high-throughput, multiplex approaches such as the protein microarray. *In vitro* Transcription/Translation (IVTT) systems – a similarly high-throughput protein expression method – are already widely utilised in the production of protein microarrays, though purified recombinant proteins derived from more traditional whole cell based expression systems also play an important role in biomarker characterisation. Here we have performed a side-by-side comparison of antigen-matched protein targets from an IVTT and purified recombinant system, on a protein microarray. The magnitude and range of antibody responses to purified recombinants was found to be greater than that of IVTT proteins, and responses between targets from opposing expression systems did not clearly correlate. However, responses between amino acid sequence-matched targets from each expression system were more closely correlated. Despite the lack of a clearly defined relationship between antigen-matched targets produced in each expression system, our data indicate that protein microarrays produced using either method can be used confidently, in a context

dependent manner, though care should be taken when comparing data derived from contrasting approaches.

Introduction

To date, the majority of malaria serologic studies have focussed on antibody responses to a small number of well-characterised, highly immunogenic *Plasmodium falciparum* antigens that have proven to be reliable markers of exposure to infection [1–8]. However, *P. falciparum* expresses more than 5000 proteins, each a potential antibody target [9,10]. Advances in technology have led to the development of new assay platforms that allow proteome scale investigation of antibody responses, such as the protein microarray [11,12] - boasting significantly greater experimental throughput than more classical monoplex methods (*e.g.* ELISA) [13,14]. The ability to simultaneously interrogate large numbers of putative targets, using low volumes of sample, significantly increases the rate at which antibody responses to antigens can be characterised. As such, protein microarray based approaches to biomarker identification and humoral response profiling in malaria, and other infectious diseases, have been increasingly adopted [15–24].

One widely utilised form of the protein microarray is based on an in vitro transcription/ translation (IVTT) system [25] – where protein products are produced through a PCR, *in vivo* recombination cloning and an *in vitro* expression pipeline, before being printed onto arrays [15]. The benefits of this methodology are considerable. In principle, whole organism proteome microarrays can be fabricated simply and quickly, enabling analysis of all potential protein driven immune responses to a pathogen. In terms of limitations, although reaction times are shorter, the rates of IVTT protein expression lag behind that of *in vivo* expression systems [26]. Therefore, the inherent heterogeneity between products is not assessed for every target making it difficult to normalise for reactivity between protein spots, which represent an impure mix of *E. coli* and target protein.

In contrast to the IVTT array methodology, the printing of purified proteins is cheaper and typically more quantifiable. Uniform amounts of product can therefore be incorporated into arrays, increasing confidence when comparing quantitative antibody responses between antigenic targets [27] and assessing relative immunogenicity. The process can be modified to support the scale up of recombinant proteins, and furthermore, affinity purification of protein targets reduces the risk of undesired background reactivity due to expression system components. However, the time required to produce panels of purified proteins is far in

excess of the IVTT system. For both systems, although the production of complex conformational proteins is possible it can sometimes be a challenge [28,29].

Here we present a comparison between an IVTT based and purified protein microarray. Matched malarial protein targets from each methodology were assessed for comparative reactivity in serum from Ugandan participant samples (n=899) [30] to determine the suitability of each approach in the context of high-throughput profiling of serological responses to protein antigens.

Results

Table 1 summarises the purified and IVTT protein targets for each antigen, with further detail in **Supplementary Table 1**. In brief, we assessed IgG antibody responses to 35 protein targets, derived from 11 well-characterised *P. falciparum* antigens (distinct gene products). Each antigen was represented by at least one IVTT and one purified protein target.

Magnitude of responses between expression systems

The magnitude of response to all protein targets was compared by antigen to evaluate differences in seroreactivity between IVTT derived and purified protein targets. As expected, responses varied significantly between antigens and between the protein targets mapping to each antigen.

Mean responses to all targets were compared by expression system (**Figure 1**) revealing a greater range of response to purified proteins (IQR Log2MFI = 3.88 – 6.40) than IVTT proteins (IQR Log2MFI 0.46 – 1.68), and a greater magnitude of response to purified than IVTT targets ($p = <0.001$). Similarly, the range and median intensity of individual antibody responses was found to be greater for purified proteins than their IVTT counterparts (*e.g.* AMA1 - IVTT_1, median [IQR] Log2MFI = 1.66 [0.80 – 2.53]; Pure_1, median [IQR] Log2MFI = 7.92 [6.16 – 8.52]) for all targets ($p = <0.001$) except MSP1 Pure_2, which more closely reflected the level of reactivity to the two MSP1 IVTT targets (**Figure 2**).

Correlation of responses between antigen matched targets

Considering all sequence matched IVTT and purified protein targets (*i.e.* excluding pairwise comparisons where purified protein sequence were non-overlapping with IVTT sequence for the same antigen) there was no evidence for a general correlation in mean response between expression platforms (Spearman's rho (r_s) = 0.279, $p = 0.23$). Antibody responses

to all protein targets for each antigen were therefore compared individually (representative example in **Figure 3** and all antigens in **Supplementary Figure 1**). This allowed for comparison between sequence matching IVTT and purified protein targets (*e.g.* EBA181 IVTT 2 vs. EBA181 Pure 1), non-matching IVTT and purified protein targets (*e.g.* EBA181 IVTT 1 vs. EBA181 Pure 1), and matching or non-matching targets produced in the same system (*e.g.* EBA181 IVTT 1 vs. EBA181 IVTT 2). Correlations were highly variable (r_s = 1.00 to -0.045) though all but one (GAMA; r_s = -0.045, p = 0.17) demonstrated a degree of positive, if not always statistically significant, association.

Multiple IVTT targets were produced for EBA181, HSP40, MSP1, MSP4 and MSP5. For all other than MSP5, non-sequence matching IVTTs were produced; correlation co-efficients for these targets were between 0.37 and 0.73 (**Supplementary Figure 1**). For EBA181 and MSP1, IVTT targets overlap by 17 amino acids – equivalent to a small peptide in terminal regions unlikely to cover immunogenic epitopes. As such, these targets are considered non-overlapping. For MSP4 and MSP5, duplicate IVTT products (identical but for 1 terminal amino acid) resulted in near perfect correlation of antibody responses (MSP4 r_s = 1.00, p = <0.001; MSP5 r_s = 0.94, p = <0.001). Multiple purified protein targets were produced for ACS5, ETRAMP4, ETRAMP5, HSP40 and MSP1 - none of which overlap. Correlation between these purified protein targets in each antigen varied between 0.31 and 0.59 (**Supplementary Figure 1**).

For the 8 antigens with >1 IVTT or purified protein target, the greatest level of correlation was found between an IVTT and purified target in 4/8 instances; between two IVTT targets (IVTT-IVTT) in 3/8 instances; and between two purified targets (purified-purified) in 1/8 instances (**Supplementary Table 2**). Comparing correlations between antigen-matched IVTT and purified proteins only, overlapping targets correlate more highly than non-overlapping targets. Sample sizes were too low to test the significance of this trend within antigens (**Figure 4**).

Discussion

Protein microarrays are a practical approach to the serological screening of large numbers of putative malaria antigen biomarkers. The throughput and flexibility of the microarray platform presents an opportunity to interrogate malarial antibody responses at a scale far exceeding traditional mono- or multiplex approaches, agnostic of predicted immunological targets. Here we have evaluated matched antigenic targets produced using two *E. coli* based expression

128 techniques – In Vitro Transcription/Translation (IVTT), and purified, whole-cell recombinants
129 – in the context of a protein microarray. We found that the magnitude of antibody responses
130 to purified protein targets was generally higher than for their IVTT counterparts, and that
131 correlation between protein target pairs at the individual serum sample level was variable
132 and related to degree of sequence homogeneity between targets. Our findings warn against
133 direct comparisons of microarray data from proteins produced in different expression
134 platforms without careful cross-validation of sequences and allelic types, but provide support
135 for the use of both IVTT and purified protein microarray platforms in the context of antigen
136 biomarker discovery and development.

137 In building this study, we hypothesised that the magnitude of responses to IVTT products -
138 which tended to be longer, often representing single exon sequences and therefore
139 potentially containing more epitopes - would be greater than purified targets truncated based
140 on species-specificity or domain boundaries which potentially represented fewer epitopes.
141 Contrary to this prediction, we found that purified proteins captured a greater range and
142 magnitude of responses (Purified, IQR Log2MFI = 3.88 – 6.40; IVTT, IQR Log2MFI 0.46 –
143 1.68; p = <0.001). The greater level of reactivity to purified targets may relate to differences
144 in the amount of protein deposited on the array, where consistent and defined amounts of
145 purified protein are spotted in contrast to the unquantified, and likely variable IVTT products.
146 These findings recommend a degree of caution in interpretation of array data from two
147 different platforms, for example: MSP5 showed the second highest mean MFI for any
148 purified protein, but showed among the lowest mean MFI of any IVTT protein.

149 In addition to differences in the magnitude of mean responses to targets stratified by
150 expression system, we observed a greater range of individual sample responses, stratified
151 by antigen, to purified proteins than in sequence matched IVTT-expressed targets (*e.g.*
152 AMA1 - IVTT_1, median [IQR] Log2MFI = 1.66 [0.80 – 2.53]; *Pichia pastoris* produced
153 Pure_1, median [IQR] Log2MFI = 7.92 [6.16 – 8.52]; p = <0.001). The *P. pastoris* AMA1 was
154 included as a control for the evaluation of the production of a conformational protein. AMA1
155 is a complex structure comprised of three domains defined by three disulphide bonds.
156 Production of AMA1 in *P. pastoris* has been fully characterised in terms of correct folding of
157 the purified protein [31,32] and this observation is likely a reflection of antibody reactivity to
158 correctly folded (*P. pastoris*) and incorrectly folded AMA1 (IVTT).

159 Considering all antigenic targets together, we found no evidence of correlation in mean
160 reactivity to sequence matched targets between expression systems (r_s = 0.28, p = 0.23).
161 This was not unexpected taking into account the differences in magnitude of response

between IVTT and purified proteins, and that the length of native protein sequence coverage between IVTT and purified targets was highly variable. In relation, targets grouped by antigen displayed highly variable correlations of response, but sequence matched proteins generally displayed stronger correlations of response than non-sequence matched targets, indicating the importance of capturing specific epitopes within expression sequences when producing antigens by either expression method.

Despite the lack of a clearly defined relationship between antigen-matched targets from opposing expression systems, we remain confident that microarrays utilising IVTT or purified recombinant proteins are able to produce compelling and biologically relevant data. Indeed, our data show age-dependent trends in antibody responses (typical of highly endemic populations [1–3]) irrespective of expression system (**Supplementary figure 2**), lending weight to the applicability of either methodology in serological assays [33–46].

The IVTT system lends itself to microarray applications, as vast numbers of proteins, or even entire proteomes, may be produced at scale relatively quickly. However, for application to serology there is concern that expressed proteins are not quantified before printing, and that expression levels of product may vary considerably; product yield in bacterial-based IVTT systems is generally considered to be lower (typically ~1 mg mL^{-1} or less) though higher protein yields have been reported [47,48]. This may lead to artificial discrepancies in reactivity between targets. Including expression tags on the protein products, such as glutathione-S-transferase (GST) or poly-histidine (His), allows detection of specific protein to confirm expression but not direct quantification. Similarly, it is important to acknowledge that the un-purified nature of printed reaction mixtures may mask, or otherwise adversely affect, the detection of antibody reactivity in a sample; Davies et al. report IVTT reaction compositions of 99% *E. coli* lysate to 1% target protein [15], though this will vary considerably, at scale, in practice.

In contrast to IVTT-based microarrays, printing purified protein allows a highly quantifiable approach to be taken. Affinity purification and dialysis of expression products substantially reduces the risk of background reactivity to bacterial components, and the simple determination of target protein concentrations allows defined quantities of product to be spotted, providing much greater confidence when comparing reactivity between targets. However, these advantages come at a substantial cost; the need for *in silico* analysis to design vectors, transfection procedures, expression and purification drastically slows the rate at which putative targets can be produced and screened. As such, proteins chosen for expression are most often down-selected from larger lists of potential markers of interest,

196 and important targets may be missed that would otherwise be revealed by proteome level
197 screening. Furthermore, longer proteins and those including transmembrane regions or
198 signal peptides, often fail to express satisfactorily; the precipitation of product or the
199 formation of inclusion bodies commonly occurs in these instances. Purified protein antigens
200 are therefore often much smaller fragments of full-length, native protein, selected based on
201 the outputs of predictive tools aiming to identify the most immunoreactive and expressible
202 epitopes. Ultimately, mapping precise antibody binding sequences in this way provides
203 greater resolution in the dissection of specific epitope reactivity in screened sera, which may
204 be of immunological relevance. Shorter, epitope specific sequences may in theory be
205 transposed to IVTT systems with a view to generating more granular serological screens,
206 though we accept that truncated protein targets will in some cases favour linear B cell
207 epitopes, while missing conformational epitopes. However, for measuring exposure to
208 infection there is less importance on the targeting of confirmation epitopes than would be
209 required for protective epitopes [33].

210 For a number of samples in all antigens, data points (Log2 MFI) are negative – a result of
211 background subtraction and normalisation processes. Protein spots showing lower than
212 background signal most often occur due to technical assay effects or artefacts in imaging,
213 and hence are not considered to be of biological relevance. An additional yet important
214 consideration is that data for IVTT and purified protein spots were not identically corrected
215 and normalised, owing to differences in the composition of respective background control
216 spots and the GST background reactivity subtracted from tagged purified proteins. This
217 makes direct comparisons between the two less certain, though the normalisation processes
218 employed go some way to overcoming this by transforming data to a ratio above expression
219 system-specific background reactivity. Furthermore, in consideration of these differences,
220 correlations were calculated for both methods after normalisation to Log2 ratios, and are
221 based on rank to allow for the identification of non-linear relationships. Nonetheless, these
222 methods are recognised as imperfect and a degree of uncertainty remains around
223 comparisons between IVTT and purified proteins.

224 The primary benefit of the microarray platform is the ability to screen orders of magnitude
225 more targets simultaneously than more standard serological assays. Our analysis shows
226 that both IVTT and purified proteins can be successfully used to capture malarial protein-
227 antigen specific antibody responses on a reverse phase protein microarray. Although
228 correlations of response between expression systems are not as strong as may have been
229 expected, a number of acknowledged technical differences in the methods of protein

230 production may account for this finding. Considering the data more broadly, observed trends
231 lend support to the utilisation of both IVTT and purified arrays depending on the objectives
232 and context of hypotheses to be investigated. The strengths and weaknesses of each
233 expression system should dictate the chosen approach on a case-by-case basis. For
234 example, very high density proteome level screening to identify 'shortlists' of candidate
235 markers based on binary categorisation of seropositivity may be best achieved using IVTT
236 systems. In contrast, smaller numbers of 'shortlisted' targets expressed as purified proteins
237 may allow for more nuanced characterisation of antibody responses on a more continuous
238 scale. As already described, the key limitation in the production of purified recombinants in
239 our current expression pipeline is throughput. The adaption of our methods to increase the
240 capacity of protein production would improve our ability to more widely mine the biomarker
241 information derived from the IVTT platform. As such, we are currently exploring a number of
242 existing approaches to address this methodological bottleneck [49,50].

243 In summary, the IVTT protein microarray approach has proven to be a powerful, high-
244 throughput, biomarker discovery platform with applicability across a range of infectious
245 diseases. When combined with a cheap, scalable and flexible protein expression platform
246 such as the *E.coli in vivo* expression platform we have the ability to mine potential diagnostic
247 and vaccine related targets.

248

249 **Methods**

250 ***Ethics***

251 All serum samples were collected after written informed consent from the participant or their
252 parent/guardian. The protocol for sample collection was reviewed and approved by the
253 Makerere University School of Medicine Research and Ethics Committee (#2011–149 and
254 #2011–167), the London School of Hygiene and Tropical Medicine Ethics Committee (#5943
255 and #5944), the Durham University School of Biological and Biomedical Sciences Ethics
256 Committee, the University of California, San Francisco, Committee on Human Research
257 (#11–05539 and #11–05995) and the Uganda National Council for Science and Technology
258 (#HS-978 and #HS-1019).

259

260 ***Samples***

261 Sera were originally collected as part of a comprehensive longitudinal surveillance study
262 conducted in three sub-counties in Uganda (Walukuba, Jinja District; Kihihi, Kanungu

263 District, and Nagongera, Tororo). The study design and methods have been previously
264 reported and are described in detail elsewhere [30]. A sub-selection of samples (n= was
265 made from individuals with a breadth of recorded clinical episodes of malaria to ensure a
266 range of sero-reactivity.

267 ***Protein targets***

268 *Purified protein expression*

269 Recombinant proteins were generated and expressed in *Escherichia coli* as glutathione S-
270 transferase (GST)-tagged fusion proteins using previously described methods [51–53]. The
271 exception to this was PfAMA1, which was expressed as a histidine tagged protein in *Pichia*
272 *pastoris* [31]. Purification of the expressed proteins was performed using affinity
273 chromatography (Glutathione Sepharose 4B (GE Healthcare Life Sciences) or HisPur Ni-
274 NTA (Invitrogen) resins for GST and His tagged proteins, respectively). Protein
275 concentration was assessed using the Bradford protein assay, with quality, and purity
276 assessed by 4-20% SDS-PAGE.

277 *IVTT protein expression*

278 An *in vitro* transcription and translation (IVTT) system was used to express proteins of
279 interest as previously described [15]. Briefly, *Plasmodium falciparum* DNA (3D7 isolate)
280 coding sequences were PCR-amplified and cloned into T7 expression vectors via
281 homologous recombination. Target sequences were expressed at 21°C for 16h in *E. coli*-
282 based, cell-free transcription/translation reactions, and products were printed onto arrays as
283 un-purified, whole reaction mixtures.

284 *Overview of compared IVTT and purified protein antigens*

285 We assessed antibody responses to protein targets mapping to eleven antigens (*i.e.* distinct
286 gene products), each represented on the array by at least one IVTT and one purified protein
287 target. Full details are in **Table 1** and **Supplementary Table 1**. The number of purified
288 protein targets varied according to availability, while the number of IVTT targets was
289 dependent on the exon composition of each the gene sequence; multiple exon sequences
290 were expressed as multiple protein targets based on exon delineation. Similarly, single exon
291 gene sequences were generally expressed as a single protein. As a result, of the 11
292 antigens investigated, 8 were represented by >1 IVTT or purified protein target; 5 had >1
293 IVTT protein target (EBA181, HSP40, MSP1, MSP4 and MSP5) and 5 had >1 purified
294 protein target (ACS5, ETRAMP4, ETRAMP5, HSP40 and MSP1). Near identical IVTT

295 proteins (1 aa difference) were produced independently and printed in parallel for two
296 antigens: MSP4 and MSP5. Purified protein targets were smaller than their IVTT
297 counterparts, which generally cover the entire reference sequence; purified protein targets
298 are produced as fragments of the full-length protein, with the aim of capturing antibodies
299 specific to predicted epitopes based on *in silico* analysis.

300 ***Protein microarray***

301 Prior to printing, Tween 20 was added to purified proteins to yield a final concentration of
302 0.001% Tween 20. Arrays were printed onto nitrocellulose-coated slides (AVID, Grace Bio-
303 Labs, Inc., Bend, OR, USA) using an Omni Grid Accent microarray printer (Digilabs, Inc.,
304 Marlborough, MA, USA). Alongside proteins of interest, buffer (PBS) and no-DNA (empty T7
305 vector reactions) were included as controls to allow for background normalisation of purified
306 and IVTT proteins respectively.

307 *Sample probing*

308 For analysis of antibody reactivity on the protein microarray, serum samples were diluted
309 1:200 in a 3 mg mL^{-1} *E. coli* lysate solution in protein arraying buffer (Maine Manufacturing,
310 Sanford, ME, USA) and incubated at room temperature for 30 min. Arrays were rehydrated
311 in blocking buffer for 30 min. Blocking buffer was removed, and arrays were probed with pre-
312 incubated serum samples using sealed, fitted slide chambers to ensure no cross-
313 contamination of sample between pads. Chips were incubated overnight at 4°C with
314 agitation. Arrays were washed five times with TBS-0.05% Tween 20, followed by incubation
315 with biotin-conjugated goat anti-human IgG (Jackson ImmunoResearch, West Grove, PA,
316 USA) diluted 1:200 in blocking buffer at room temperature. Arrays were washed three times
317 with TBS-0.05% Tween 20, followed by incubation with streptavidin-conjugated SureLight P-
318 3 (Columbia Biosciences, Frederick, MD, USA) at room temperature protected from light.
319 Arrays were washed three times with TBS-0.05% Tween 20, three times with TBS, and once
320 with water. Arrays were air dried by centrifugation at 500 x *g* for 5 min and scanned on a
321 GenePix 4300A High-Resolution Microarray Scanner (Molecular Devices, Sunnyvale, CA,
322 USA). Target and background intensities were measured using an annotated grid file (.GAL).

323 *Data normalisation*

324 Microarray spot foreground and local background fluorescence data were imported into R
325 (Foundation for Statistical Computing, Vienna, Austria) for correction, normalisation and
326 analysis. Local background intensities were subtracted from foreground using the
327 backgroundCorrect function of the limma package [54]. The backgroundCorrect function was

328 then further applied to GST-tagged purified proteins, whereby background-corrected GST
329 fluorescence was subtracted from background-corrected target fluorescence to account for
330 any GST-specific reactivity in samples. All data were then Log2 transformed and the mean
331 signal intensity of buffer and no-DNA control spots were subtracted from purified and IVTT
332 proteins respectively to give a relative measure of reactivity to targets over background
333 (**Supplementary Figure 3**) [20].

334 **Acknowledgements**
335 Many thanks to Tony Holder for the kind provision of purified recombinant proteins (GAMA
336 and MSP7).
337
338
339 **References**

340 1. Corran P, Coleman P, Riley E, Drakeley C. Serology: a robust indicator of malaria transmission
341 intensity? Trends in Parasitology. 2007;23: 575–582.
342 doi:https://doi.org/10.1016/j.pt.2007.08.023

343 2. Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SLR, Carneiro I, et al. Estimating
344 medium- and long-term trends in malaria transmission by using serological markers of malaria
345 exposure. Proceedings of the National Academy of Sciences of the United States of America.
346 2005;102: 5108–5113. doi:10.1073/pnas.0408725102

347 3. Stewart L, Gosling R, Griffin J, Gesase S, Campo J, Hashim R, et al. Rapid assessment of malaria
348 transmission using age-specific sero-conversion rates. PloS one. 2009;4: e6083–e6083.
349 doi:10.1371/journal.pone.0006083

350 4. Bousema T, Youssef RM, Cook J, Cox J, Alegana VA, Amran J, et al. Serologic markers for
351 detecting malaria in areas of low endemicity, Somalia, 2008. Emerging infectious diseases.
352 2010;16: 392–399. doi:10.3201/eid1603.090732

353 5. Theisen M, Vuust J, Gottschau A, Jepsen S, Høgh B. Antigenicity and immunogenicity of
354 recombinant glutamate-rich protein of Plasmodium falciparum expressed in Escherichia coli.
355 Clin Diagn Lab Immunol. 1995;2: 30–34.

356 6. Carvalho LH, Fontes CJ, Fernandes AA, Marinuzzi HC, Krettli AU. Cross-reactive cellular immune
357 response to circumsporozoite proteins of Plasmodium vivax and P. falciparum in malaria-
358 exposed individuals. Parasite Immunol. 1997;19: 47–59. doi:10.1046/j.1365-3024.1997.d01-
359 182.x

360 7. Milek RL, DeVries AA, Roeffen WF, Stunnenberg H, Rottier PJ, Konings RN. Plasmodium
361 falciparum: heterologous synthesis of the transmission-blocking vaccine candidate Pfs48/45 in
362 recombinant vaccinia virus-infected cells. Exp Parasitol. 1998;90: 165–174.
363 doi:10.1006/expr.1998.4315

364 8. Kanunfre KA, Leoratti FMS, Hoffmann EHE, Durlacher RR, Ferreira AW, Moraes-Avila SL, et al.
365 Differential recognition of Plasmodium falciparum merozoite surface protein 2 variants by
366 antibodies from malaria patients in Brazil. Clin Diagn Lab Immunol. 2003;10: 973–976.
367 doi:10.1128/cdli.10.5.973-976.2003

368 9. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the
369 human malaria parasite Plasmodium falciparum. Nature. 2002;419: 498–511.
370 doi:10.1038/nature01097

371 10. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, et al. A proteomic
372 view of the Plasmodium falciparum life cycle. Nature. 2002;419: 520–526.
373 doi:10.1038/nature01107

374 11. Burbelo PD, Ching KH, Bush ER, Han BL, Iadarola MJ. Antibody-profiling technologies for
375 studying humoral responses to infectious agents. Expert Rev Vaccines. 2010;9: 567–578.
376 doi:10.1586/erv.10.50

377 12. Manzano-Román R, Dasilva N, Díez P, Díaz-Martín V, Pérez-Sánchez R, Orfao A, et al. Protein
378 arrays as tool for studies at the host-pathogen interface. J Proteomics. 2013;94: 387–400.
379 doi:10.1016/j.jprot.2013.10.010

380 13. Adu B, Issahaque Q-A, Sarkodie-Addo T, Kumordjie S, Kyei-Baafour E, Sinclear CK, et al.
381 Microscopic and Submicroscopic Asymptomatic Plasmodium falciparum Infections in Ghanaian
382 Children and Protection against Febrile Malaria. Infect Immun. 2020;88.
383 doi:10.1128/IAI.00125-20

384 14. Ondigo BN, Hamre KES, Frosch AEP, Ayodo G, White MT, John CC. Antibody Profiles to P.
385 falciparum Antigens Over Time Characterize Acute and Long-Term Malaria Exposure in an Area
386 of Low and Unstable Transmission. Am J Trop Med Hyg. 2020. doi:10.4269/ajtmh.19-0480

387 15. Davies DH, Liang XW, Hernandez JE, Randall A, Hirst S, Mu YX, et al. Profiling the humoral
388 immune response to infection by using proteome microarrays: High-throughput vaccine and
389 diagnostic antigen discovery. Proceedings of the National Academy of Sciences of the United
390 States of America. 2005;102: 547–552. doi:DOI 10.1073/pnas.0408782102

391 16. Arevalo-Herrera M, Lopez-Perez M, Dotsey E, Jain A, Rubiano K, Felgner PL, et al. Antibody
392 Profiling in Naive and Semi-immune Individuals Experimentally Challenged with Plasmodium
393 vivax Sporozoites. PLoS Negl Trop Dis. 2016;10: e0004563. doi:10.1371/journal.pntd.0004563

394 17. Crompton PD, Kayala MA, Traore B, Kayentao K, Ongoiba A, Weiss GE, et al. A prospective
395 analysis of the Ab response to Plasmodium falciparum before and after a malaria season by
396 protein microarray. Proceedings of the National Academy of Sciences of the United States of
397 America. 2010;107: 6958–6963. doi:10.1073/pnas.1001323107

398 18. Sundaresh S, Randall A, Unal B, Petersen JM, Belisle JT, Hartley MG, et al. From protein
399 microarrays to diagnostic antigen discovery: a study of the pathogen Francisella tularensis.
400 Bioinformatics. 2007;23: I508–I518. doi:10.1093/bioinformatics/btm207

401 19. Helb DA, Tetteh KKA, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, et al. Novel serologic
402 biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for
403 individuals and communities. Proceedings of the National Academy of Sciences. 2015;112:
404 E4438–E4447. doi:10.1073/pnas.1501705112

405 20. Stone WJR, Campo JJ, Ouédraogo AL, Meerstein-Kessel L, Morlais I, Da D, et al. Unravelling the
406 immune signature of Plasmodium falciparum transmission-reducing immunity. Nature
407 Communications. 2018;9: 558. doi:10.1038/s41467-017-02646-2

408 21. van den Hoogen LL, Walk J, Oulton T, Reuling IJ, Reiling L, Beeson JG, et al. Antibody Responses
409 to Antigenic Targets of Recent Exposure Are Associated With Low-Density Parasitemia in
410 Controlled Human Plasmodium falciparum Infections. Frontiers in Microbiology. 2019;9.
411 doi:10.3389/fmicb.2018.03300

412 22. Sen S, Sanders EC, Gabriel KN, Miller BM, Isoda HM, Salcedo GS, et al. Antibody Response to a
413 Nucleocapsid Epitope as a Marker for COVID-19 Disease Severity. bioRxiv. 2020.
414 doi:10.1101/2020.10.15.341743

415 23. Khan S, Nakajima R, Jain A, de Assis RR, Jasinskas A, Obiero JM, et al. Analysis of Serologic
416 Cross-Reactivity Between Common Human Coronaviruses and SARS-CoV-2 Using Coronavirus
417 Antigen Microarray. bioRxiv. 2020. doi:10.1101/2020.03.24.006544

418 24. Chakraborty S, Randall A, Vickers TJ, Molina D, Harro CD, DeNearing B, et al. Interrogation of a
419 live-attenuated enterotoxigenic Escherichia coli vaccine highlights features unique to wild-type
420 infection. NPJ Vaccines. 2019;4: 37. doi:10.1038/s41541-019-0131-7

421 25. Murray CJ, Baliga R. Cell-free translation of peptides and proteins: from high throughput
422 screening to clinical production. Curr Opin Chem Biol. 2013;17: 420–426.
423 doi:10.1016/j.cbpa.2013.02.014

424 26. Failmezger J. Understanding limitations to increased potential of cell-free protein synthesis.
425 2018 [cited 2 Dec 2020]. doi:10.18419/OPUS-9701

426 27. Kamuyu G, Tuju J, Kimathi R, Mwai K, Mburu J, Kibinge N, et al. KILchip v1.0: A Novel
427 Plasmodium falciparum Merozoite Protein Microarray to Facilitate Malaria Vaccine Candidate
428 Prioritization. Front Immunol. 2018;9: 2866. doi:10.3389/fimmu.2018.02866

429 28. Baneyx F, Mujacic M. Recombinant protein folding and misfolding in Escherichia coli. Nat
430 Biotechnol. 2004;22: 1399–1408. doi:10.1038/nbt1029

431 29. Ban B, Sharma M, Shetty J. Optimization of Methods for the Production and Refolding of
432 Biologically Active Disulfide Bond-Rich Antibody Fragments in Microbial Hosts. Antibodies
433 (Basel). 2020;9. doi:10.3390/antib9030039

434 30. Kamya MR, Arinaitwe E, Wanzira H, Katureebe A, Barusya C, Kigozi SP, et al. Malaria
435 transmission, infection, and disease at three sites with varied transmission intensity in Uganda:
436 implications for malaria control. Am J Trop Med Hyg. 2015;92: 903–912. doi:10.4269/ajtmh.14-
437 0312

438 31. Collins CR, Withers-Martinez C, Bentley GA, Batchelor AH, Thomas AW, Blackman MJ. Fine
439 mapping of an epitope recognized by an invasion-inhibitory monoclonal antibody on the
440 malaria vaccine candidate apical membrane antigen 1. J Biol Chem. 2007;282: 7431–7441.
441 doi:10.1074/jbc.M610562200

442 32. Collins CR, Withers-Martinez C, Hackett F, Blackman MJ. An inhibitory antibody blocks
443 interactions between components of the malarial invasion machinery. PLoS Pathog. 2009;5:
444 e1000273. doi:10.1371/journal.ppat.1000273

445 33. Bailey JA, Berry AA, Travassos MA, Ouattara A, Boudova S, Dotsey EY, et al. Microarray analyses
446 reveal strain-specific antibody responses to Plasmodium falciparum apical membrane antigen 1
447 variants following natural infection and vaccination. Sci Rep. 2020;10: 3952.
448 doi:10.1038/s41598-020-60551-z

449 34. Tessema SK, Nakajima R, Jasinskas A, Monk SL, Lekieffre L, Lin E, et al. Protective Immunity
450 against Severe Malaria in Children Is Associated with a Limited Repertoire of Antibodies to
451 Conserved PfEMP1 Variants. Cell Host Microbe. 2019;26: 579-590.e5.
452 doi:10.1016/j.chom.2019.10.012

453 35. Proietti C, Krause L, Trieu A, Dodoo D, Gyan B, Koram KA, et al. Immune Signature Against
454 Plasmodium falciparum Antigens Predicts Clinical Immunity in Distinct Malaria Endemic
455 Communities. Mol Cell Proteomics. 2020;19: 101–113. doi:10.1074/mcp.RA118.001256

456 36. Stucke EM, Niangaly A, Berry AA, Bailey JA, Coulibaly D, Ouattara A, et al. Serologic responses
457 to the PfEMP1 DBL-CIDR head structure may be a better indicator of malaria exposure than
458 those to the DBL-α tag. Malar J. 2019;18: 273. doi:10.1186/s12936-019-2905-9

459 37. Venkatesh A, Jain A, Davies H, Periera L, Maki JN, Gomes E, et al. Hospital-derived antibody
460 profiles of malaria patients in Southwest India. Malar J. 2019;18: 138. doi:10.1186/s12936-019-
461 2771-5

462 38. Friedman-Klabanoff DJ, Travassos MA, Ifeonu OO, Agrawal S, Ouattara A, Pike A, et al. Epitope-
463 specific antibody responses to a Plasmodium falciparum subunit vaccine target in a malaria-
464 endemic population. J Infect Dis. 2020. doi:10.1093/infdis/jiaa611

465 39. Wu L, Mwesigwa J, Affara M, Bah M, Correa S, Hall T, et al. Sero-epidemiological evaluation of
466 malaria transmission in The Gambia before and after mass drug administration. BMC Med.
467 2020;18: 331. doi:10.1186/s12916-020-01785-6

468 40. Labadie-Bracho MY, van Genderen FT, Adhin MR. Malaria serology data from the Guiana
469 shield: first insight in IgG antibody responses to Plasmodium falciparum, Plasmodium vivax and
470 Plasmodium malariae antigens in Suriname. Malar J. 2020;19: 360. doi:10.1186/s12936-020-
471 03434-y

472 41. Wu L, Mwesigwa J, Affara M, Bah M, Correa S, Hall T, et al. Antibody responses to a suite of
473 novel serological markers for malaria surveillance demonstrate strong correlation with clinical
474 and parasitological infection across seasons and transmission settings in The Gambia. BMC
475 Med. 2020;18: 304. doi:10.1186/s12916-020-01724-5

476 42. Steinhardt L, Ravaoarisoa E, Wiegand R, Harimanana A, Hedje J, Cotte AH, et al. School-based
477 sero-surveys to assess the validity of using routine health facility data to target malaria
478 interventions in the Central Highlands of Madagascar. J Infect Dis. 2020.
479 doi:10.1093/infdis/jiaa476

480 43. Lu A, Cote O, Dimitrova SD, Cooley G, Alamgir A, Uzzaman MS, et al. Screening for malaria
481 antigen and anti-malarial IgG antibody in forcibly-displaced Myanmar nationals: Cox's Bazar
482 district, Bangladesh, 2018. Malar J. 2020;19: 130. doi:10.1186/s12936-020-03199-4

483 44. Surendra H, Supargiyono, Ahmad RA, Kusumasari RA, Rahayujati TB, Damayanti SY, et al. Using
484 health facility-based serological surveillance to predict receptive areas at risk of malaria
485 outbreaks in elimination areas. BMC Med. 2020;18: 9. doi:10.1186/s12916-019-1482-7

486 45. Müller-Sienerth N, Shilts J, Kadir KA, Yman V, Homann MV, Asghar M, et al. A panel of
487 recombinant proteins from human-infective Plasmodium species for serological surveillance.
488 Malar J. 2020;19: 31. doi:10.1186/s12936-020-3111-5

46. Assefa A, Ali Ahmed A, Deressa W, Sime H, Mohammed H, Kebede A, et al. Multiplex serology demonstrate cumulative prevalence and spatial distribution of malaria in Ethiopia. Malar J. 2019;18: 246. doi:10.1186/s12936-019-2874-z

47. Iskakova MB, Szaflarski W, Dreyfus M, Remme J, Nierhaus KH. Troubleshooting coupled in vitro transcription–translation system derived from Escherichia coli cells: synthesis of high-yield fully active proteins. Nucleic Acids Research. 2006;34: e135–e135. doi:10.1093/nar/gkl462

48. Caschera F, Noireaux V. Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription–translation system. Biochimie. 2014;99: 162–168. doi:10.1016/j.biochi.2013.11.025

49. Konczal J, Gray CH. Streamlining workflow and automation to accelerate laboratory scale protein production. Protein Expr Purif. 2017;133: 160–169. doi:10.1016/j.pep.2017.03.016

50. Nettleship JE, Rada H, Owens RJ. Overview of a High-Throughput Pipeline for Streamlining the Production of Recombinant Proteins. Methods Mol Biol. 2019;2025: 33–49. doi:10.1007/978-1-4939-9624-7_2

51. Herman LS, Fornace K, Phelan J, Grigg MJ, Anstey NM, William T, et al. Identification and validation of a novel panel of Plasmodium knowlesi biomarkers of serological exposure. PLoS Negl Trop Dis. 2018;12: e0006457. doi:10.1371/journal.pntd.0006457

52. Tetteh KKA, Osier FHA, Salanti A, Kamuyu G, Drought L, Failly M, et al. Analysis of antibodies to newly described Plasmodium falciparum merozoite antigens supports MSPDBL2 as a predicted target of naturally acquired immunity. Infect Immun. 2013;81: 3835–3842. doi:10.1128/IAI.00301-13

53. Polley SD, Tetteh KKA, Cavanagh DR, Pearce RJ, Lloyd JM, Bojang KA, et al. Repeat sequences in block 2 of Plasmodium falciparum merozoite surface protein 1 are targets of antibodies associated with protection from malaria. Infect Immun. 2003;71: 1833–1842. doi:10.1128/iai.71.4.1833-1842.2003

54. 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: e47–e47. doi:10.1093/nar/gkv007

Figure legends

Figure 1. Mean magnitude of antibody responses to targets. The mean magnitude of response of each protein target stratified by expression system, presented with median and interquartile range of all mean responses.

Figure 2. Magnitude and range of response to IVTT and purified proteins. All sample responses (n = 899) to all protein targets grouped by antigen, presented with median and interquartile range.

525 **Figure 3. Correlation of antibody responses and sequence mapping.** A representative
526 example correlogram of multiple antigen-matched targets (left). Spearman's rank correlation
527 reported (r_s) and increasing blue colour scale indicates relative strength of correlation based
528 on calculated correlations for all proteins included in this analysis. Protein schematic (right)
529 represents amino-acid aligned representation of IVTT (green) and purified (orange) proteins
530 to the full-length native protein (grey). Proteins in the correlogram and schematic are
531 correspondingly aligned.

532 **Figure 4. Spearman's rho correlation coefficients between antigen-matched IVTT and**
533 **purified proteins.** Targets with overlapping amino acid sequences are indicated by closed
534 circles, compared to non-overlapping sequences indicated by open circles

Tables

Table 1. Description of antigens and their corresponding IVTT and purified protein targets.

Protein	Description	Full length (AA)	Protein target/expression system	Size (Start AA - End AA)
ACS5	Acyl CoA synthase	811	IVTT_1	811 (1–811)
			Pure_1	117 (294–410)
			Pure_2	160 (414–573)
			Pure_3	150 (578–727)
AMA1	Apical membrane antigen 1	622	IVTT_1	622 (1–622)
			Pure_1	450 (97–546)
EBA181	Erythrocyte binding antigen 181	1567	IVTT_1	754 (1–754)
			IVTT_2	752 (737–1488)
			Pure_1	585 (755–1339)
ETRAMP4	Early transcribed membrane antigen 4	136	IVTT_1	136 (1–136)
			Pure_1	25 (28–52)
			Pure_2	61 (76–136)
ETRAMP5	Early transcribed membrane antigen 5	181	IVTT_1	181 (1–181)
			Pure_1	86 (26–111)
			Pure_2	47 (135–181)
GAMA	GPI-anchored membrane antigen	738	IVTT_1	738 (1–738)
			Pure_1	99 (68–166)
HSP40	Heat shock protein 40 type II	402	IVTT_1	134 (80-213)
			IVTT_2	171 (213-401)
			Pure_1	83 (71-153)
			Pure_2	189 (214-402)
MSP1	Merozoite surface protein 1	1720	IVTT_1	870 (1-870)
			IVTT_2	868 (853-1720)
			Pure_1	45 (64-108)
			Pure_2	35 (54-63;109-133)
			Pure_3	116 (1605-1720)
MSP4	Merozite surface protein 4	272	IVTT_1	162 (1-162)
			IVTT_2	161 (1-161)
			IVTT_3	110 (163-272)
			Pure_1	65 (43-107)
MSP5	Merozoite surface protein 5	272	IVTT_1	172 (1-172)
			IVTT_2	171 (1-171)
			Pure_1	61 (147-207)
MSP7	Merozoite surface protein 7	351	IVTT_1	351 (1-351)
			Pure_1	175 (177-351)

Figure 1

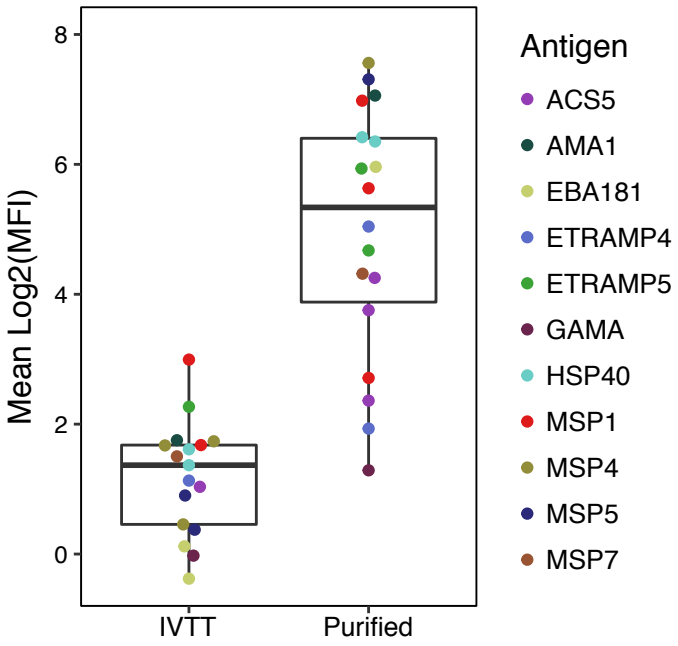

Mean Log2(MFI)
8
6
4
2
0
IVTT
Purified
Antigen
ACS5
AMA1
EBA181
ETRAMP4
ETRAMP5
GAMA
HSP40
MSP1
MSP4
MSP5
MSP7

Figure 2

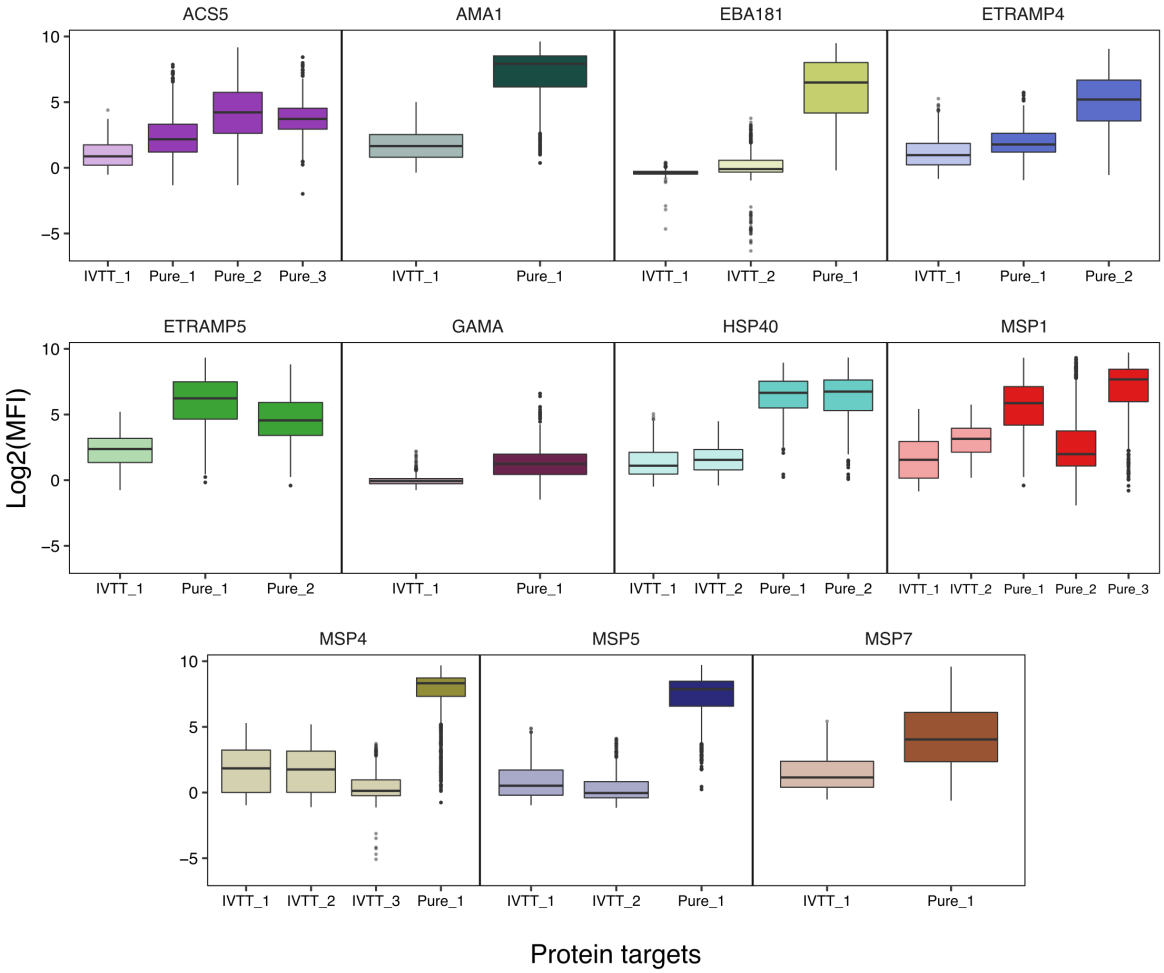

ACS5
AMA1
EBA181
ETRAMP4
ETRAMP5
GAMA
HSP40
MSP1
MSP4
MSP5
MSP7
Log2(MFI)
10
5
0
−5
IVTT_1
IVTT_2
IVTT_3
Pure_1
Pure_2
Pure_3
Protein targets

Figure 3

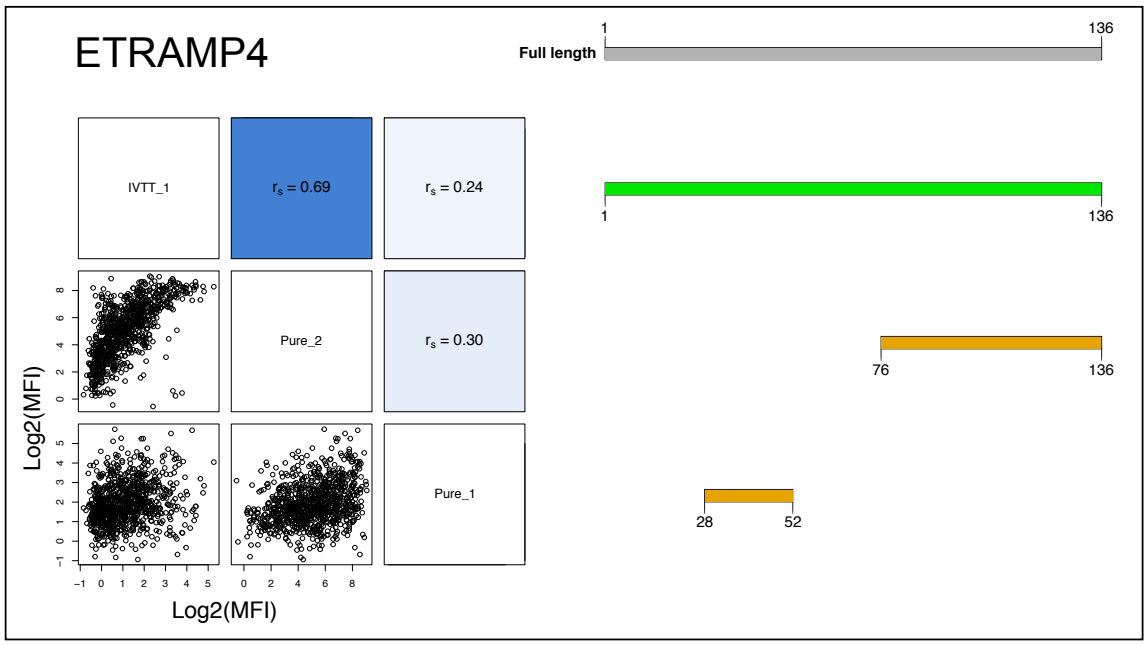
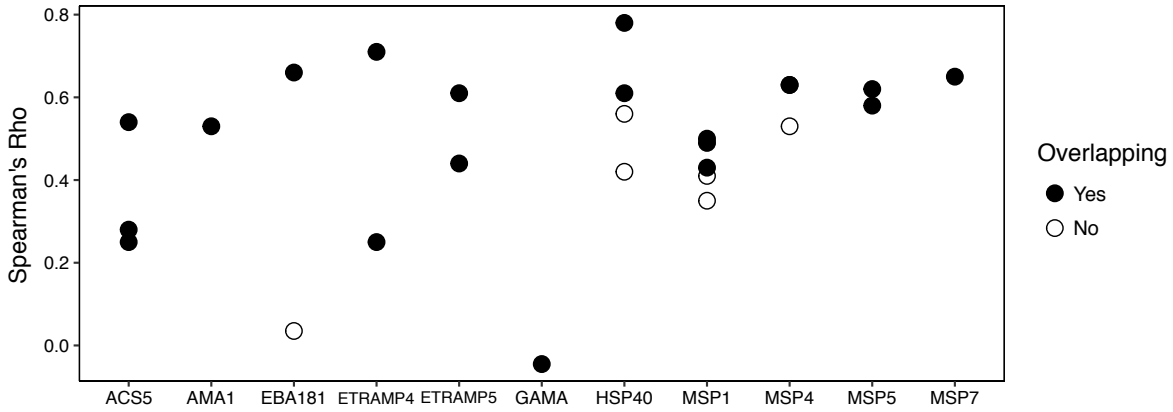

ETRAMP4
Full length
1
136
IVTT_1
r_s = 0.69
r_s = 0.24
1
136
Pure_2
r_s = 0.30
76
136
Pure_1
28
52
Log2(MFI)
Log2(MFI)

Figure 4


Spearman's Rho
0.8
0.6
0.4
0.2
0.0
ACS5
AMA1
EBA181
ETRAMP4
ETRAMP5
GAMA
HSP40
MSP1
MSP4
MSP5
MSP7
Overlapping
Yes
No