

## A secreted protein microarray platform for extracellular protein interaction discovery

Sree R. Ramani<sup>a,1</sup>, Irene Tom<sup>a,1</sup>, Nicholas Lewin-Koh<sup>b</sup>, Bernd Wranik<sup>a</sup>, Laura DePalatis<sup>a</sup>, Jianjun Zhang<sup>c</sup>, Dan Eaton<sup>a</sup>, Lino C. Gonzalez<sup>a,\*</sup>

<sup>a</sup> Department of Protein Chemistry, Genentech, South San Francisco, CA 94080, USA

<sup>b</sup> Department of Nonclinical Biostatistics, Genentech, South San Francisco, CA 94080, USA

<sup>c</sup> Department of Bioinformatics, Genentech, South San Francisco, CA 94080, USA

### ARTICLE INFO

#### Article history:

Received 10 June 2011

Received in revised form 10 September 2011

Accepted 16 September 2011

Available online 21 September 2011

#### Keywords:

Protein microarray

Protein–protein interaction

Receptor–ligand interaction

Ig receptor

Extracellular matrix

### ABSTRACT


Characterization of the extracellular protein interactome has lagged far behind that of intracellular proteins, where mass spectrometry and yeast two-hybrid technologies have excelled. Improved methods for identifying receptor–ligand and extracellular matrix protein interactions will greatly accelerate biological discovery in cell signaling and cellular communication. These technologies must be able to identify low-affinity binding events that are often observed between membrane-bound coreceptor molecules during cell–cell or cell–extracellular matrix contact. Here we demonstrate that functional protein microarrays are particularly well-suited for high-throughput screening of extracellular protein interactions. To evaluate the performance of the platform, we screened a set of 89 immunoglobulin (Ig)-type receptors against a highly diverse extracellular protein microarray with 686 genes represented. To enhance detection of low-affinity interactions, we developed a rapid method to assemble bait Fc fusion proteins into multivalent complexes using protein A microbeads. Based on these screens, we developed a statistical methodology for hit calling and identification of nonspecific interactions on protein microarrays. We found that the Ig receptor interactions identified using our methodology are highly specific and display minimal off-target binding, resulting in a 70% true-positive to false-positive hit ratio. We anticipate that these methods will be useful for a wide variety of functional protein microarray users.

© 2011 Elsevier Inc. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/3.0/).

Many orphan receptors and ligands remain within the human secretome. Moreover, new interacting partners continue to be

lar protein interactions, these systems offered limited throughput and consumed microgram amounts of protein. To further increase

and similar papers at [core.ac.uk](http://core.ac.uk)

brought to you by  CORE

provided by Elsevier - Publisher Connector

lular protein interaction networks will shed new light and suggest new mechanisms for cellular communication and regulation. Unfortunately, for a number of reasons, methods for identifying secreted protein interactions have remained limited [3,4].

Previously, we used a secreted protein library, the secreted protein discovery initiative (SPDI)<sup>2</sup> [5], to identify coreceptors for the immunoglobulin (Ig) domain-containing receptors BTLA and TIGIT [6,7] using surface plasmon resonance (SPR) and biolayer interferometry technologies. Although amenable to identifying novel extracellu-

in consumption, protein microarrays, [8] and Zhu and

coworkers [9], offer a unique method of depositing very small amounts of protein in a high-density format (>5000 features/slide). A fluorescently labeled, or tagged, analyte protein (the bait) is then used to probe interactions with all of the arrayed proteins simultaneously. Microarrays composed of specific protein domain families have previously been used to identify intracellular protein interactions [10,11]. In vitro transcription/translation capture systems have been developed for direct synthesis of proteins in situ on microarrays [12,13]. Protein microarrays composed of large protein libraries from plant, yeast, and human have also been described [9,14,15]. However, little work has focused on investigating the robustness and broad utility of this approach for identifying interactions among extracellular proteins.

The human Ig receptor family, defined as proteins containing exclusively one or more Ig domains, is composed of more than 200 genes with diverse functions and binding partners. Approximately

\* Corresponding author. Fax: +1 650 225 5945.

E-mail address: [gonzal29@gene.com](mailto:gonzal29@gene.com) (L.C. Gonzalez).

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> Abbreviations used: SPDI, secreted protein discovery initiative; Ig, immunoglobulin; SPR, surface plasmon resonance; CHO, Chinese hamster ovary; SDS-PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; PBS, phosphate-buffered saline; BSA, bovine serum albumin.

half of these Ig receptors have reported binding partners and interact with a wide range of affinities either homotypically, heterotypically (with other Ig receptors), or with other non-Ig-related proteins [16]. Members of this family have previously been used to explore other extracellular protein interaction platforms. Jiang and Barclay screened a panel of 36 Ig receptors for interactions against themselves using a 6 × 6 SPR array [17]. Wright and coworkers developed an enzyme-linked immunosorbent assay (ELISA)-style assay called AVEIXIS (avidity-based extracellular interaction screen) and used it to screen more than 100 zebrafish Ig receptors and leucine-rich repeat proteins against a library of 249 extracellular proteins, identifying a number of novel interactions [18,19].

Here we used a highly diverse secreted protein library in conjunction with a set of 89 Ig receptor baits to develop protein microarrays as an effective tool for large-scale extracellular interaction screening. Our methodology, using a fast and robust multivalent bait approach along with statistical hit calling and nonspecific binding accounting, revealed top hits as known binding partners and several new interactions for functional validation. These results establish protein microarrays as an important technology for characterization of the extracellular protein interactome.

## Materials and methods

### Cloning, protein expression, and purification of bait Ig receptors

Residues encoding the extracellular domain of Ig receptors were amplified by polymerase chain reaction (PCR) using Origene clones or a complementary DNA (cDNA) library as templates. *Clal* and *Ascl* restriction sites were incorporated at the 5' and 3' flanking ends to allow ligation into a C-terminal human Fc-tag pRK vector. Proteins were transiently expressed in Chinese hamster ovary (CHO) cells at a 1-L scale and purified, as described previously [20], over a protein A column, followed by a Mono-Q and/or S200 sizing column to remove degraded or aggregated protein if necessary. Proteins were concentrated using Amicon concentrators (Millipore). For our study, 89 Ig receptor Fc fusions were selected based on expression levels in CHO cells and protein quality after purification (Table 1). Representative sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) gels of several bait proteins are shown in Supplementary Figs. 1F–H (see Supplementary material).

### Bait labeling and protein A microbead–Fc fusion complexes

All Fc fusion bait proteins were labeled with Amersham Cy5 monoreactive dye (GE Healthcare, cat. no. PA25001) and separated from the free dye using desalting columns (Princeton Separation, cat. no. CS-800). Dye to protein ratios, determined by ultraviolet (UV) absorbance at OD<sub>280</sub> and 650 nm, between 2.0 and 4.0 were used. Cy5 conjugates were spun at 100,000g for 15 min in a tabletop ultracentrifuge (Beckman Coulter) at 4 °C before use.

To form protein A microbead–Fc fusion complexes, 200- $\mu$ l aliquots of 20  $\mu$ g/ml Cy5-labeled Ig receptor protein in phosphate-buffered saline (PBS) were individually mixed for 30 min at room temperature with different volumes (0, 10, 20, 40, and 60  $\mu$ l) of stock protein A microbeads (Miltenyi Biotec, cat. no. 130-071-001) on a tube rotator. The remaining uncomplexed Fc-tagged protein was then measured directly from the microbead solution via an Octet biolayer interferometer (ForteBio). The sample containing the minimal saturating volume of beads was selected based on this analysis.

The microbead–Fc fusion complexes were then pelleted in a tabletop centrifuge at 21,000g for 10 min and resuspended in PBS/5% milk. To block binding to Fc-tagged proteins on the microarray, samples were supplemented with 1 mg/ml soluble protein A immediately prior to the binding assay.

**Table 1**

List of 89 Ig receptors screened against the secreted protein microarray.

Number	Name	Number	Name
1	ASAM	46	KIR3DL3
2	BSG	47	LAIR1
3	BTLA	48	LILRB2
4	BTN2A1	49	LNIR
5	BTN3A1	50	LSAMP
6	BTN3A2	51	LSR
7	BTN3A3	52	LY6G6D
8	BTNL2	53	MFAP3
9	BTNL8	54	MFAP3L
10	CADM3	55	MOG
11	CD160	56	MPZ
12	CD2	57	MPZL1
13	CD200	58	MPZL3
14	CD200R1	59	MXRA8
15	CD226	60	NCR3
16	CD244	61	NEGR1
17	CD274	62	NPTN
18	CD276	63	NTM
19	CD300C	64	PDCD1
20	CD300LD	65	PSG11
21	CD300LF	66	PSG4
22	CD33	67	PSG5
23	CD4	68	PSG7
24	CD80	69	PVRL1
25	CD84	70	PVRL2
26	CEACAM1	71	PVRL3
27	CEACAM20	72	PVRL4
28	CEACAM4	73	SCN1B
29	CEACAM6	74	SEMA4A PSI
30	CTLA4	75	SIGLEC5
31	CXADR	76	SIGLEC8
32	ERMAP	77	SIGLEC9
33	ESAM	78	SIRPA
34	F11R	79	SIRPB2
35	FCRL1	80	SLAMF1
36	FCRL2	81	SLAMF7
37	FCRL4	82	SLAMF8
38	GPA33	83	TAPBPL
39	HEPACAM2	84	TIGIT
40	HIDE1	85	TMIGD1
41	ICAM	86	TREM1
42	ICOS	87	TREM2
43	IGSF11	88	TREML4
44	JAM2	89	VSIG4
45	JAM3		

### Secreted protein microarray production

We compiled 1851 protein samples from the Genentech secreted protein library (SPDI) [5]. The samples represent mostly human proteins (>99%) and were expressed from either CHO, baculovirus, or *Escherichia coli* systems. The majority were poly-His (948 proteins), poly-His/Gln (246 proteins), or C-terminally Fc tagged (613 proteins). The remaining proteins were untagged. Prey proteins were purified by standard affinity purification methods. Representative SDS–PAGE gels of several prey proteins are shown in Supplementary Fig. 1A–E. Protein concentrations were adjusted to between 200 and 400  $\mu$ g/ml when possible. Protein stocks were diluted 1:1 with PBS/80% glycerol for long-term storage at –20 °C in 96-well plates. Working plates (384 wells) containing 10  $\mu$ l of samples in each well were generated from the stock plates. Proteins were spotted with quill-type spotting pins (Arrayit, cat. no. 946MP3) onto epoxysilane slides (SCHOTT, Nexterion slide E, cat. no. 1064016) using a NanoPrint LM60 48-pin microarrayer (Arrayit) at 60% relative humidity. To visualize the array for mask fitting and to aid in identifying sample carryover, bovine serum albumin (BSA)–Cy3 (5  $\mu$ g/ml in PBS/40% glycerol) was spotted in duplicate between each protein sample. Next, 1 h after printing, microarray slides were removed from the humidified environment

and immediately blocked overnight with PBS/5% milk (OXOID, skim milk powder, cat. no. LP0031) at 4 °C. Slides were stored at –20 °C in PBS/40% glycerol to prevent freezing.

The relative immobilization level for each Fc-, His/Gln-, and His-tagged protein was determined by probing two replicate slides with a Cy5-labeled anti-Fc antibody (Jackson ImmunoResearch, cat. no. 109-176-170), a Cy5-labeled anti-His/Gln antibody (Genentech), or a multivalent anti-His microbead (Miltenyi Biotec, cat. no. 130-091-124) detected with Cy5-labeled goat anti-mouse F(ab')<sub>2</sub> (Jackson ImmunoResearch, cat. no. 115-176-072). Data for anti-Fc and anti-polyHis/Gln were normalized relative to anti-polyHis using the mean background subtracted fluorescence (F635 – B635) value for each tag subset. Protein with background subtracted fluorescence values below 200, including 296 polyHis-, 135 poly-His/Gln-, and 86 Fc-tagged proteins, were considered to have minimal immobilization and are not counted in the total number of unique proteins represented on the microarray. The remaining 1334 protein samples and their relative immobilization levels are summarized in [Supplementary Table 1](#), and the amino acid sequences for nonpurchased samples are provided in [Supplementary Table 2](#) (see [Supplementary material](#)). In a small fraction of cases, spots showed anomalous anti-tag signal (defined here as having >20% signal from another tag). Based on our BSA-Cy3 printing controls, the majority of these instances (15 of 24) could be attributed to protein carryover ([Supplementary Table 1](#)). For the entire Ig receptor set, we identified 9 false-positive hits due to carryover. Although relatively minimal, these observations suggest that carryover should still be controlled for.

#### Protein microarray processing

Slides were allowed to warm to room temperature in PBS/40% glycerol and rinsed with PBST (PBS + 0.1% Tween 20) before loading onto an automated a-Hyb hybridization station (Miltenyi Biotec) for binding. The a-Hyb staining protocol was run at 15 °C as follows: wash with PBST for 1 min (step 1), load 200 µl of 1.0 mg/ml protein A (Sigma, cat. no. P7837) in 5% milk/PBST and incubate for 30 min to prevent uncomplexed protein A microbeads from binding Fc tags on the microarray (step 2), wash 5 times with PBST for 1 min (step 3), load 200 µl of bait microbead complex in 5% milk/PBST in the presence of 1.0 mg/ml protein A and incubate for 30 min (step 4), wash 5 times with PBST for 1 min (step 5), and wash with PBS for 1 min (step 6). The slides were then immediately placed in individual 50-ml Falcon tubes and dried by spinning at 900g for 5 min in a tabletop centrifuge. Slides were scanned with a GenePix 4000B scanner (Molecular Devices). The Cy3 (532 nm) and Cy5 (635 nm) emissions for each slide were measured using a photomultiplier tube (PMT) setting that avoided signal saturation. GenePix Pro 6.0 software (Molecular Devices) was used for analysis. Representative images are shown in [Supplementary Fig. 2](#).

#### Data analysis

The scanned and intensity integrated data were saved as GPR files in GenePix Pro 6.0 (Molecular Devices) and processed in R using the limma package [21]. Preprocessing steps included background correction and within-chip normalization. For background correction, a local background estimate based on the normal exponential convolution model [22] was used. The “normexp” method models the observed pixel intensities as a mixture of two random variables, one normally distributed and the other exponentially distributed, representing background noise and signal, respectively. After background correction, the log<sub>2</sub> transformation of the background corrected signal was applied to correct the skew. The data were then normalized by subtracting the mean and dividing by

the standard deviation. Further quantile normalization was applied between the two replicate microarrays to put them on the same scale. The normalized log<sub>2</sub> signals were then used to score hits as a function of the two replicate spots on each microarray according to the following equation:

$$\text{score} = \frac{\frac{\text{spot}_1 + \text{spot}_2}{2} - \delta}{\sqrt{|\text{spot}_1 - \text{spot}_2| + 2}} \quad (1)$$

The parameter  $\delta$  designates a lower signal threshold empirically set at the 25th percentile of the total array fluorescence. In the denominator, the addition of 2 to the signal difference inflates the variance measure for low signal spots and helps to normalize the overall distribution of scores. Although other values for this factor could be used, empirically the number 2 worked well. With scores calculated for each slide, the results were then analyzed for the intersection of high-scoring candidates between slides. To control for slide variability, duplicate microarrays from separate spotting runs were analyzed. An upper tail probability from the approximate normal distribution of 0.0001 was used as the hit cutoff. Initial hits were assigned to each microarray replicate separately, and the intersection of these hits from both replicates was used to call final hits. Finally, an additional level of filtering was needed to identify and exclude proteins on the microarray that might bind nonspecifically. To address this issue, final hits for the complete Ig receptor set were compiled ([Supplementary Table 1](#)). From this list, the cumulative prey hit rate was determined and a data-driven elimination threshold of 10% was used to remove nonspecific binders.

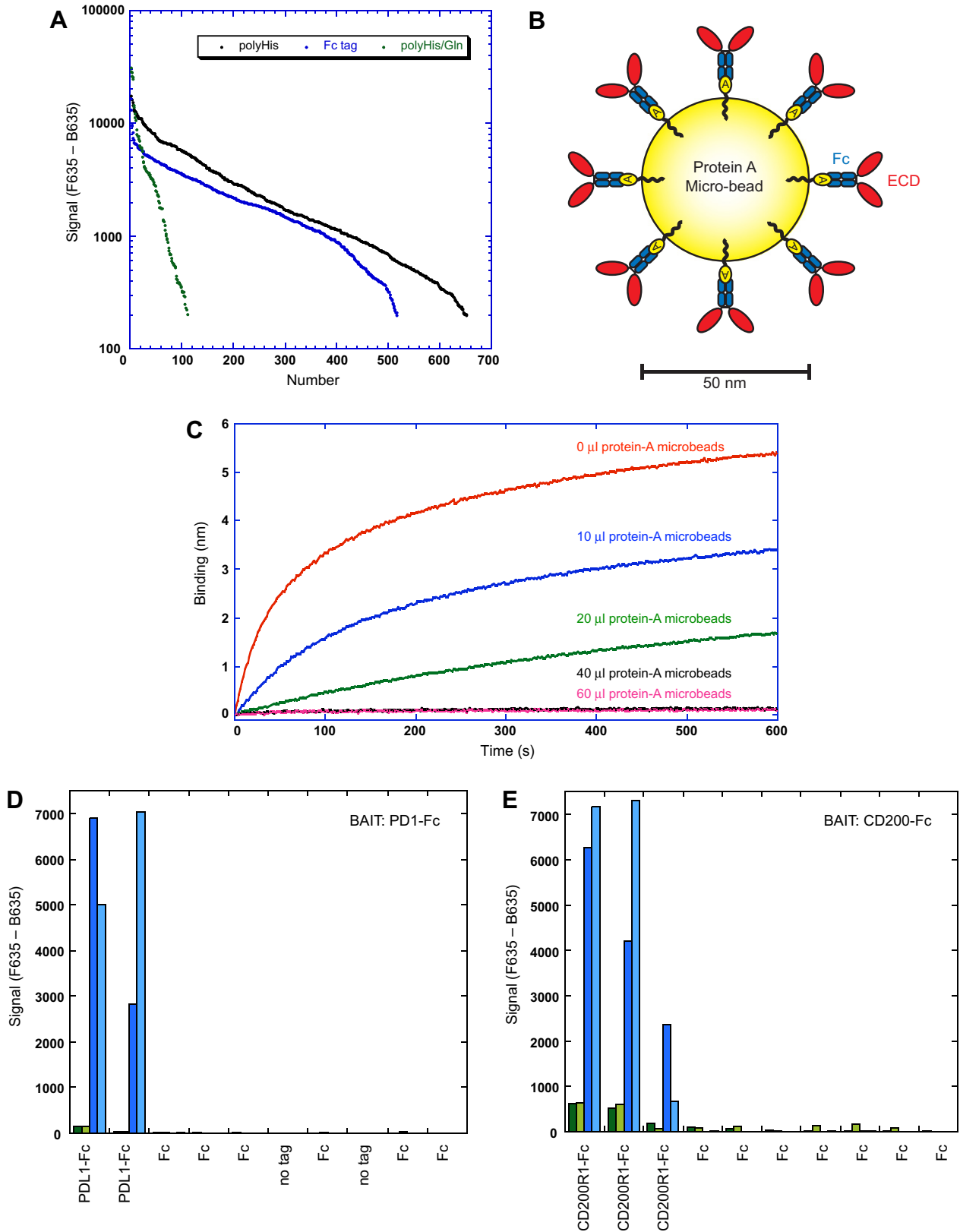
#### SPR validation

Hits were validated by SPR using a Biacore 3000 instrument ([Supplementary Figs. 3–8](#)). Proteins were immobilized (or captured with immobilized anti-Fc antibody) on a CM5 chip at more than 1000 resonance units (RU). Analytes were generally run at 50 µg/ml protein concentration in HBS-P buffer (0.01 M Hepes [pH 7.4], 0.15 M NaCl, and 0.005% surfactant P20). Interaction pairs were tested twice, once with each partner immobilized. In each case, negative control proteins were immobilized on separate flow cells as specificity controls. Furthermore, negative control analytes were run for each immobilized protein. An interaction was considered as validated if binding was detected in both orientations and was not observed for negative controls.

## Results

#### Generating a functional secreted protein microarray

The SPDI library [5] is composed of more than 700 secreted or extracellular domains from single transmembrane proteins individually purified using Fc, polyHis, or polyHis/Gln tags. SPDI proteins were spotted at 60% relative humidity from 40% glycerol in PBS buffer to allow proteins to remain hydrated during the print run. Importantly, glycerol also allows the printing plates to be stored at –20 °C and transferred to room temperature without freeze–thawing that would risk protein inactivation. Keeping slides hydrated at all times was critical to maintaining protein functional activity (data not shown). The relative levels of proteins immobilized on the microarray were determined by probing slides with either anti-Fc, anti-polyHis, or anti-polyHis/Gln antibodies, confirming immobilization for 1334 protein samples representing 686 genes ([Fig. 1A](#) and [Supplementary Table 1](#)).

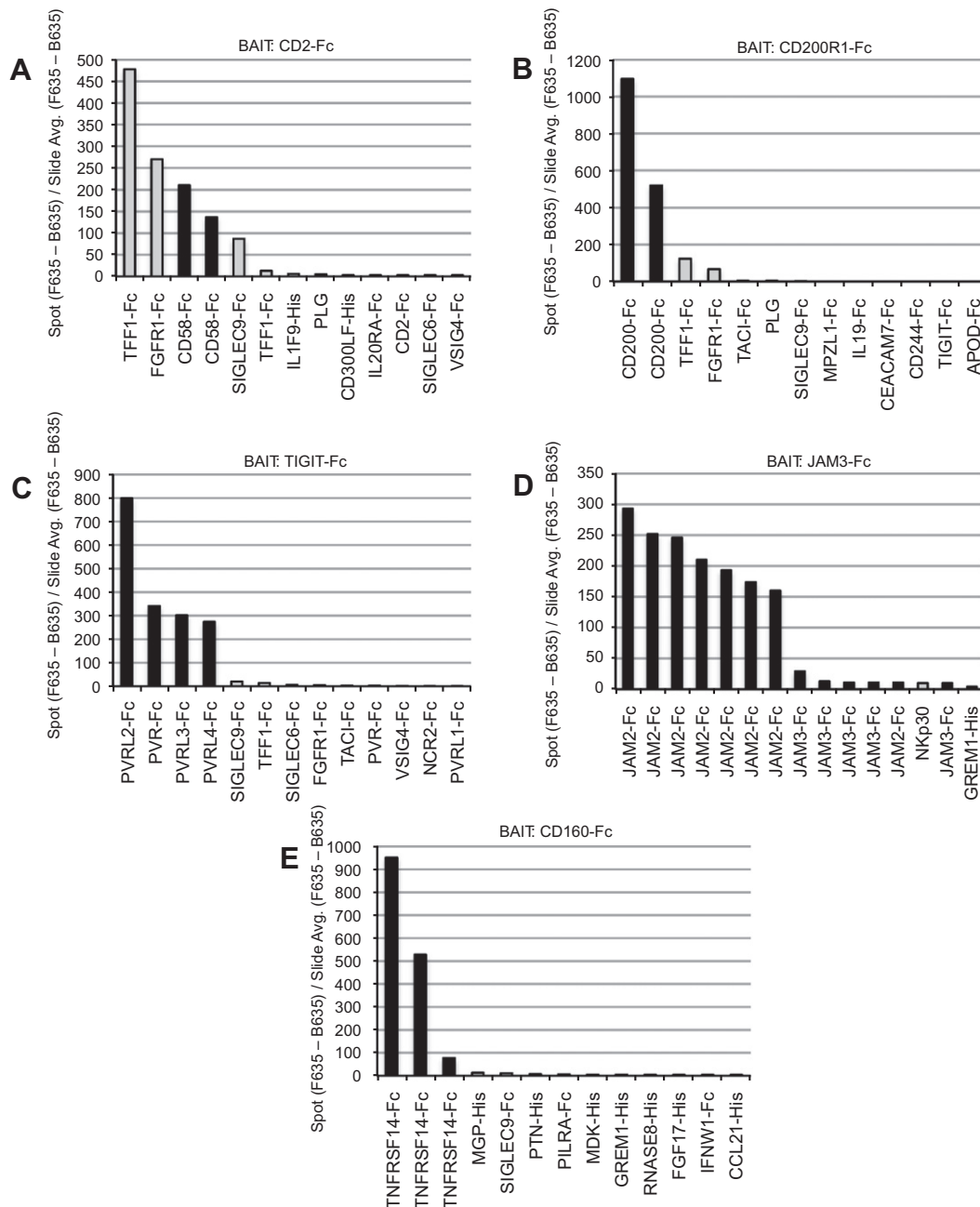


**Fig. 1.** Secreted protein microarray immobilization and multivalent analysis using Fc fusion proteins. (A) Log plot of the relative background subtracted fluorescence signal for His (black)-, Fc (blue)-, and His/Gln (green)-tagged proteins immobilized on epoxy-coated slides detected with respective anti-tag antibodies having a signal above 200. (B) Schematic model of a protein A microbead complex with ECD-Fc fusion protein. Protein A is shown attached to the microbead and binding the Fc (blue) domain of the captured bait protein. (C) Identification of optimal protein A microbead to Fc fusion protein ratios. A representative ForteBio Octet sensorgram using protein A sensors is shown. The association curves represent a titration of protein A microbeads (as indicated) against a constant amount of Fc fusion protein (4  $\mu$ g). The minimal bead concentration where no free Fc fusion remains is selected as optimal (the black curve in this case). (D and E) Screens of PD1-Fc-Cy5 (D) and CD200-Fc-Cy5 (E) against the secreted protein microarray. The green bars show two independent replicates for soluble bait. The blue bars show two independent replicates for protein A microbead complexes. The top 10 hits are shown sorted based on data from replicate 1 of the microbead complex screens (dark blue bars). Relevant hits are labeled. Tags (if present) on all other proteins are indicated.

Multivalent microbead complexes for enhanced protein–protein interaction signal detection on protein microarrays

Although soluble ligands generally bind to cell surface receptors with high affinity, coreceptor interactions between two cell surface proteins can bind with much lower affinity and, therefore, must be a technical consideration [17,18]. Voulgaraki and coworkers [23] demonstrated microarray detection of a low-affinity ( $K_D \sim 1 \mu\text{M}$ ) coreceptor interaction between CD200 and CD200R1 by generating a multivalent analyte. We have developed an extension of this approach, using Fc fusion constructs, that allows for the fast and efficient formation of multivalent bait particles. Protein A-coated microbeads are used to capture Cy5-labeled Fc fusion protein from solution (Fig. 1B). The optimal microbead to protein ratio is

determined by measuring the amount of free Fc fusion in solution via bilayer interferometry (Fig. 1C). As proof of principle, two low-affinity receptors, PD1-Fc-Cy5 and CD200-Fc-Cy5 (either as free soluble protein or in complex with protein A microbeads), were used to probe the secreted protein microarray (Fig. 1D and E). The two PDL1 and three CD200R1 protein lots present on the microarray were identified with significantly higher signal with the microbead complexes compared with screens performed with the soluble baits. The enhanced signal from the microbead complexes ranged from 10 to more than 150 times the signal from the soluble protein alone despite using the same amount of protein in each assay. Importantly, because of our soluble protein A blocking protocol, the microbead complexes did not show any off-target binding to unrelated Fc fusion proteins present on the array.



**Fig. 2.** Specific binding of Fc fusion–protein A microbead complexes. Bar plots displaying the background subtracted fluorescent signal normalized against the average background subtracted fluorescent signal for the whole slide are shown. The top ranking interactors are shown for screens of CD2-Fc (A), CD200R1-Fc (B), TIGIT-Fc (C), JAM3-Fc (D), and CD160-Fc (E) baits. Black bars represent known interactions, and gray bars represent background or unexpected binding.

### Testing interaction specificity on the secreted protein microarray

Using the microbead complex method described above, we selected an additional five Ig receptors with known ligands on the microarray to evaluate specific versus nonspecific binding. The Ig receptors selected have varying affinities to their cognate ligands, ranging from nanomolar (TIGIT and CD160) [2,7] to micromolar (CD2 and CD200) [23,24], whereas JAM3 binds homotypically with low affinity and JAM2 with much higher affinity [25]. Baits were screened in duplicate, and the average fluorescent signals for the top hits for each screen were determined (Fig. 2). In each case except for CD2, the top hits were the expected ligands and little nonspecific binding was observed. JAM3 showed strong binding signal to the JAM2 protein lots on the microarray and much weaker, but detectable, signals to several JAM3 samples. In addition to binding the two protein lots of CD58 on the microarray, CD2 displayed unexpected binding to TFF1, FGFR1, and SIGLEC9. Incidentally, these three proteins were also present as top hits in several of the other screens, but with much lower fluorescent signal, suggesting that binding to these proteins may represent general nonspecific interactions.

### Screen of an extended Ig receptor set and identification of specific protein–protein interaction hits

To further validate the secreted protein microarray and investigate the rate of nonspecific binding, we selected an additional 82 Ig receptors, composed of both orphan receptors and receptors with known binding partners present on the microarray, to screen (Table 1). To eliminate user bias and develop a more standardized method for hit determination, we developed a statistical scoring scheme (see Materials and Methods). An upper tail probability from a normal distribution fit of the data was used as the hit cutoff (Fig. 3A). The results were then analyzed for the intersection of hits between two replicate slides (Fig. 3B). The intersection method represents a more stringent methodology relative to taking a simple average where deviations on a single array can skew the results. To identify promiscuous binders on the microarray, we hypothesized that these proteins could be identified and excluded from consideration by determining the hit rate across the 89 independent screens (Fig. 3C). A data-driven elimination threshold of a 10% prey hit rate was used and identified five protein samples having highly nonspecific characteristics (Fig. 3D). Interestingly, two of the five proteins (SIGLEC6 and SIGLEC9) are known sialic acid binding proteins [26].

### Evaluating true-positive and false-positive hit rates

We reexamined the data for CD200 and the screens represented in Fig. 2 with our statistical scoring criteria. The results showed distinctly called hits versus the lower scoring false positives that were identified on a single replicate only (Fig. 4). Using this methodology, we proceeded to analyze the entire dataset of 89 bait receptors and identified 151 hits (Supplementary Table 1). Scores from multiple hits deriving from different lots of the same protein were averaged together, resulting in a total of 105 bait/prey interactions. The hits sorted according to their mean score are plotted in Fig. 5. Table 2 summarizes known or expected hits and novel interactions validated by SPR. The majority of high-scoring interactions identified represent true binding partners. For example, of the 53 top-half scoring hits in Fig. 5, forty-five (85%) were known or expected. Of the 8 unexpected hits in this group, 4 were validated positively by SPR. Of the 52 lower scoring interactions, only 18 (35%) were known or expected, and only 7 of the 34 unexpected hits were validated positively by SPR. A total of 38 baits had no hits, and these were generally orphans or did not have binding partners included on the microarray. Moreover, 16 homotypic interactions were

detected, with 3 more (IGSF4B, JAM1, and HEPACAM2) identified at a lower probability threshold. Of these 19 interactions, 2 have not been reported in the literature (HEPACAM2 and LAIR1). The only expected homotypic interactions not observed were for PVRL4 and PVRL3. The fact that we were able to identify known homotypic interactions in this set without significant false positives suggests that the microbead protein microarray approach represents a robust method for identifying this class of interaction. The microarray scoring system we have developed appears to be sufficient as a general qualitative confidence measure.

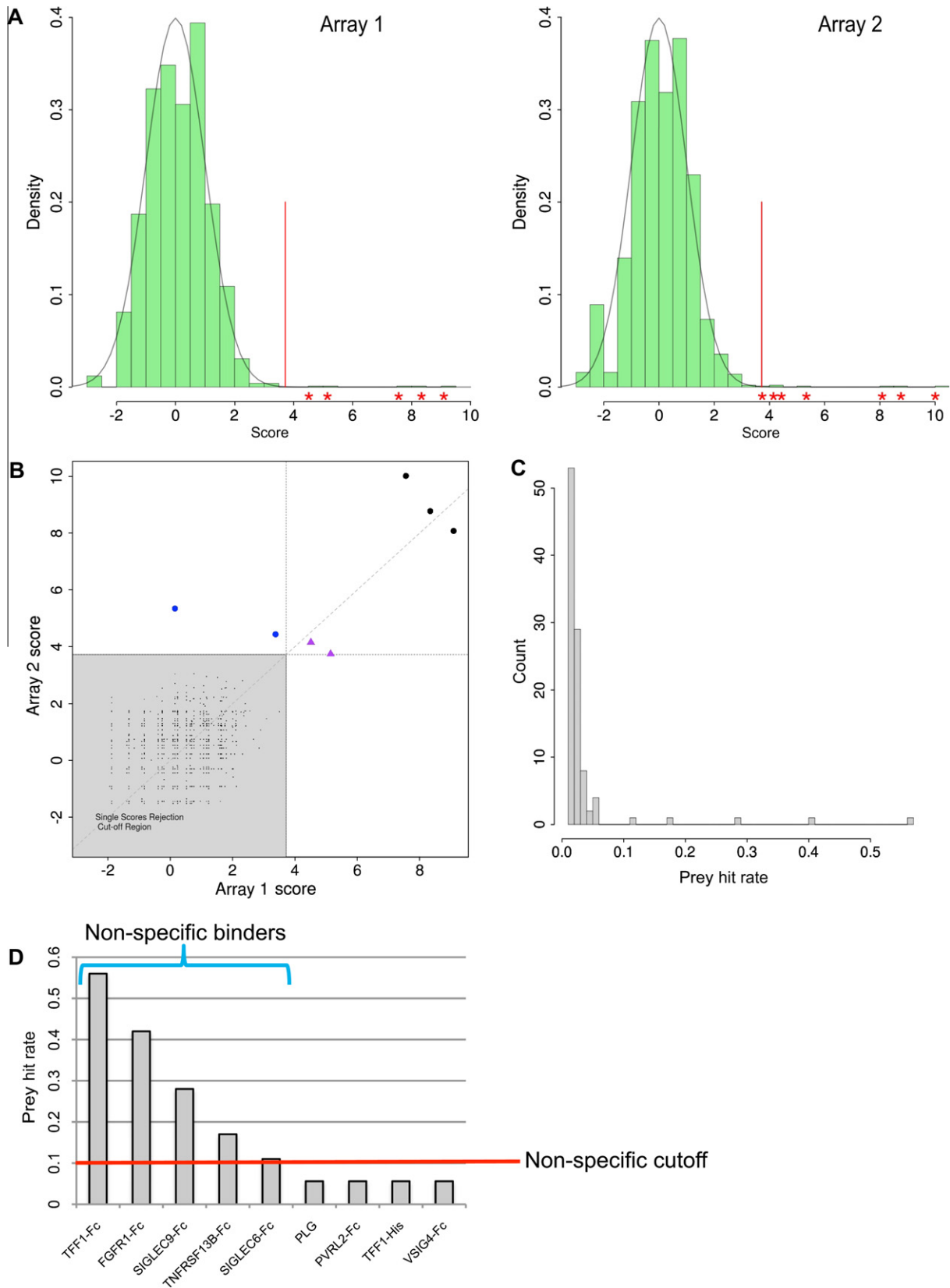
### Evaluating the influence of immobilization levels on false negatives

We investigated whether we could derive any general conclusions on the influence of immobilization level on the false-negative rate. Although our study was not designed to answer the limit of sensitivity relative to immobilization levels, we can analyze the data in aggregate by considering only hits that had more than a single lot on the microarray. This compiled subset is summarized in Supplementary Table 3. Analysis of false negatives and true positives for Fc-tagged proteins within this set shows that there is little difference in the distribution of immobilization levels (Supplementary Fig. 9A). The same is true for the His-tagged proteins (Supplementary Fig. 9B). Moreover, the relative immobilization levels do not show an obvious correlation with the mean hit score derived from our statistical analysis, suggesting that above a certain immobilization level there may be other significant factors that contribute to the binding signal. Interestingly, the false-negative to true-positive hit ratio for Fc- or His-tagged proteins (Supplementary Fig. 9D) suggests that Fc-tagged prey proteins are more effective for identifying hits with the class of proteins used in this study.

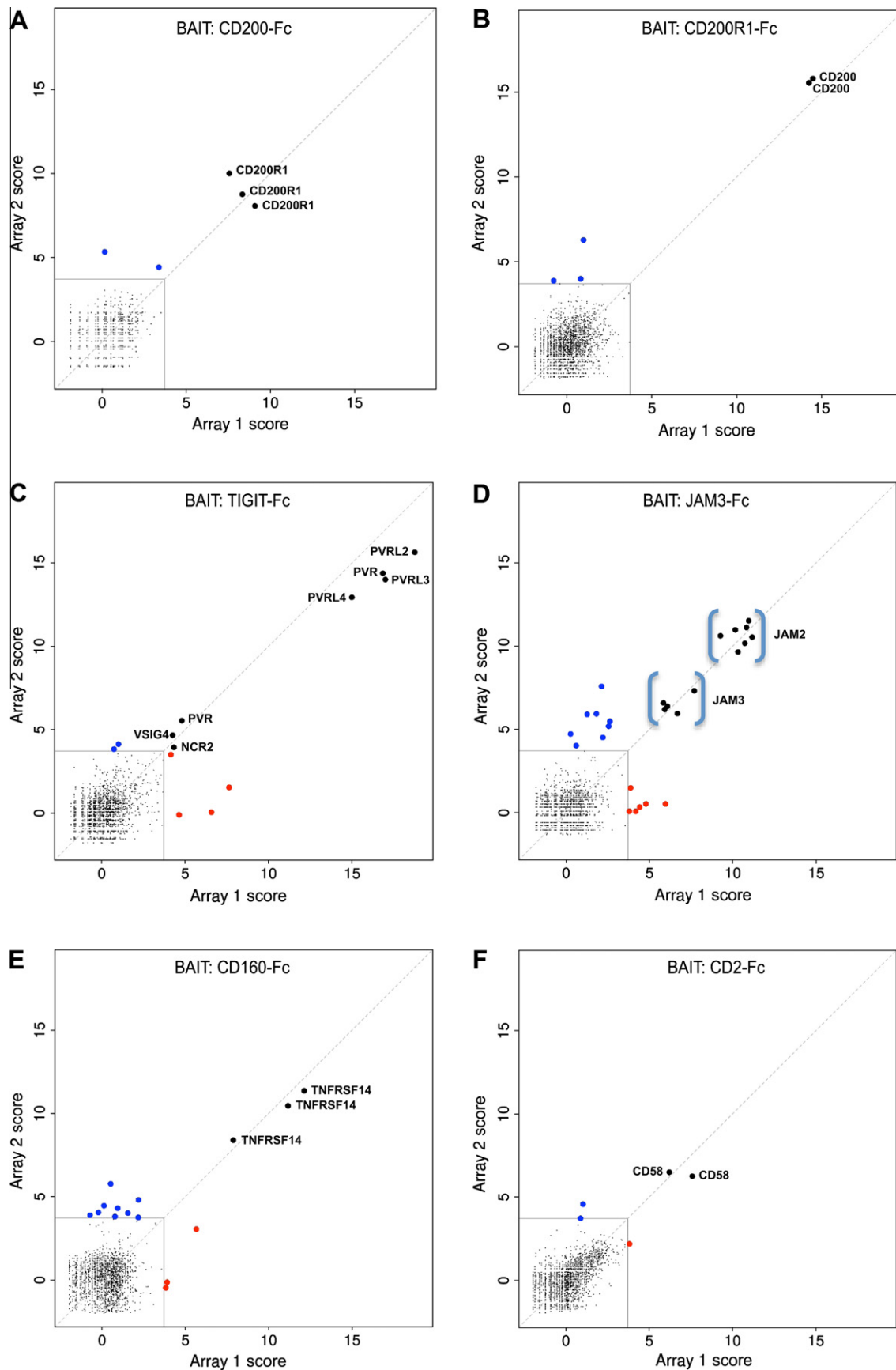
## Discussion

Our aim in this study was to test the performance of protein microarrays in identifying extracellular protein interactions. We used a set of Ig receptors, with known and unknown binding partners, and an unbiased statistical hit identification algorithm to evaluate the ability of the platform to detect known and novel interactions. With 89 receptors screened against 686 represented genes, effectively a total of 61,054 potential interactions were probed. Based on the hit selection criteria applied and our SPR validation results, we obtained an overall 70% true-positive to false-positive hit ratio. The interactions identified largely represent the expected interactions for this receptor set. Of the positive interactions identified, 11 were novel. These results demonstrate that protein microarrays are an effective and robust technology for rapid extracellular protein interaction screening.

The IgLON (NTM, NEGR1, and LSAMP), PVR, and CEACAM Ig receptor subfamilies have members that are known to interact within their own Ig subgroups [7,27,28]. We were able to recapitulate the majority of these interactions. For example, all three IgLON family members screened interacted with each other as well as homotypically, and the six members of the PVR family screened generally interacted with the expected specificity [7]. Of the novel interactions identified, one of the highest scoring was between the T cell costimulatory molecule CD80 and nerve growth factor receptor NGFR. CARTPT, a hypothalamus-expressed secreted protein, was found to interact with two CEACAM family members. MPZL3 bound to MPZL2, both of which are broadly expressed and belong to the MPZ subfamily (this interaction was also recently identified in a screen by Bushell and coworkers [18]). PSG5 bound TIE1, an endothelial cell receptor regulating vascular development. The trefoil protein TFF1 bound to five different orphan Ig receptors

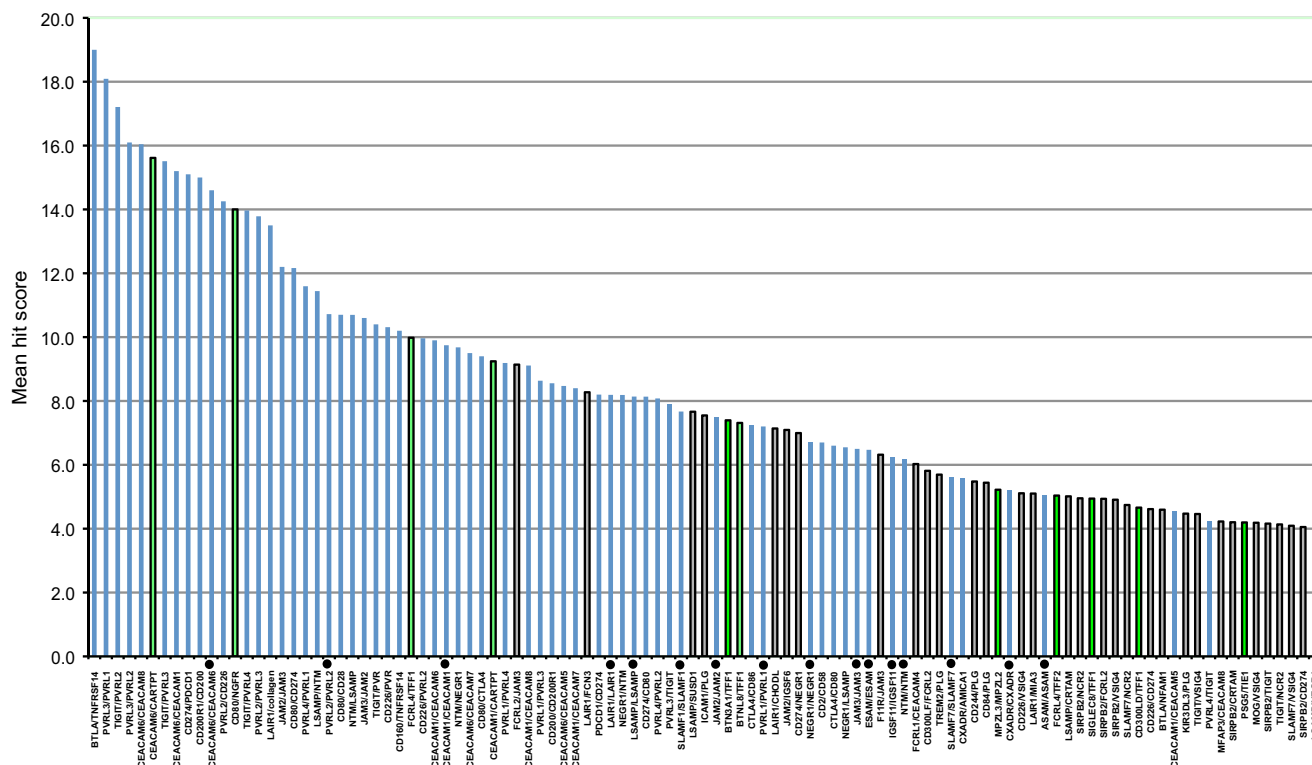


**Fig.3.** Statistical methodology for hit determination. (A) Histograms of scores for two array replicates with the fitted normal distribution and the 0.0001 probability cutoff (vertical red line) indicated. Data for the CD200 screen is shown here as a representative example. Hits are represented by asterisks plotted above the x axis. (B) Representative intersection plot for hit identification. The histograms for arrays 1 and 2 shown in panel A can be considered as one-dimensional projections along the x and y axes of the intersection plot, respectively. The dashed diagonal line represents equality. The horizontal and vertical lines are the individual 0.0001 probability cutoffs. There are no array 1-only hits. Blue circles are array 2-only hits. Purple triangles represent hits against nonspecific binders. Black circles are intersection hits. (C) Histogram showing the distribution of prey hit rates compiled from the screen of 89 Ig receptors. The majority of prey proteins have a hit rate of less than 5%. (D) Bar plot showing the top promiscuous binders. Five proteins were found to be highly nonspecific, appearing in more than 10% of screens.



**Fig. 4.** Intersection plots for representative screens. Black circles (labeled) represent intersection hits scored as described in Materials and Methods. Red and blue circles represent hits called on only a single array. The lower left square of each plot represents the 0.0001 percentile cutoff and contains all non-hit proteins. Results from screens of CD200-Fc (A), CD200R1-Fc (B), TIGIT-Fc (C), JAM3-Fc (D), CD160-Fc (E), and CD2-Fc (F) baits are shown.





**Fig. 5.** Hit summary for the complete Ig receptor screen. A bar plot ranking of 105 interactions identified from the screen of 89 Ig receptors is shown. The interactions are labeled in the bait/prey orientation (e.g., BTLA/TNFRSF14, where BTLA is the bait and TNFRSF14 was the hit). Blue bars represent known or expected interactions. Green bars represent unexpected interactions that were validated by SPR. Interactions represented by gray bars were negative by SPR. Black circles along the lower axis indicate homotypic interactions. If multiple lots of the same protein in the SPDI library were hit, their scores were averaged.

(BTN3A1, SIGLEC8, CD300D, FCRL4, and BTNL8). FCRL4 bound both TFF1 and TFF2. Trefoil proteins are expressed in the gastrointestinal mucosa and are suggested to play a role in the maintenance of epithelial integrity. These proteins may be involved in carbohydrate recognition, which might explain the broad binding specificity [29]. Further studies will be needed to address the physiological relevance of each of these interactions.

The ability to form multivalent bait particles to enhance signal strength, especially for low-affinity interactions, was an important aspect of our ability to identify these interactions. It is less clear whether multivalency of the immobilized prey is important. Presumably at high enough immobilization levels, proteins may be in close enough proximity to act in a multivalent manner. Interestingly, however, relative immobilization levels did not show a correlation with mean hit score, suggesting that above a certain threshold the amount of protein immobilized does not dominantly contribute to the score. Moreover, in a few cases we found that interactions appeared to be tag dependent. For example, ASAM hit with an Fc-tagged ASAM lot but not with three other His-tagged ASAM samples. Similarly, LSAMP, NEGR1, and NTM all hit against NTM-Fc but not with four other lots of NTM-His despite having similar relative immobilization levels. Correspondingly, the false-negative to true-positive hit ratio for prey proteins with more than one lot on the microarray was higher for His-tagged preys when compared with Fc-tagged proteins. It is possible that the dimeric Fc tag confers additional avidity, which allows these interactions to be identified as hits. A more systematic study is needed to fully evaluate the importance of prey multimerization on microarrays; however, our current data suggest that C-terminally tagged Fc fusions (or other multimerizing tag) for extracellular domains of single-transmembrane receptors may be beneficial for microarray screens involving potentially low-affinity coreceptor interactions.

We also found that in certain instances not all lots of a protein scored sufficiently high, even with the same tag, to be identified as a hit under our methodology. There are several potential, and possibly confounding, reasons for this. For instance, protein activity or quality for a particular purification lot may be compromised.

Similarly, there are several contributing factors that might result in nonspecific binding. For example, proteins may be naturally highly charged or hydrophobic. Some proteins may interact with general carbohydrate motifs. There may be issues of protein quality (e.g., some level of protein degradation, denaturation, or aggregation). Therefore, it is likely that in any large set of proteins, some nonspecific binders will be present. The approach we describe here accounts for nonspecific interactors by tracking their hit behavior over many unrelated screens. As with any threshold method, an appropriate cutoff must be applied. Fortunately, we were able to account for the majority of the nonspecific binding events by discounting hits from five protein samples on the array (TFF1-Fc, FGFR1-Fc, SIGLEC9-Fc, TNFRSF13B-Fc, and SIGLEC6-Fc). Interestingly, TFF1, FGFR1, and TNFRSF13B each had two or three protein lots in the SPDI library, and in each case only one lot was highly nonspecific. This observation suggests that there was a protein quality problem for these particular lot preparations and that these proteins did not display general nonspecific binding characteristics. Interestingly, there was not an obvious correlation between high immobilization levels and nonspecific binding activity. For example, the anti-Fc-Cy5 signal for nonspecific FGFR1-Fc was 2948, somewhat higher than the other two lots with signals of 1670 and 1808. However, 140 other Fc-tagged proteins had higher immobilization signals ranging from 2950 to 16,540. Similarly, TFF1-Fc had an immobilization signal of only 1912. In addition, SDS-PAGE gel analysis did not reveal any obvious deficiencies in these nonspecific protein samples. Overall, we were encouraged by the fact that significant nonspecific binding

**Table 2**  
Complete list of interactions identified from Ig receptor set.

Screen	Mean Score	Hit Name	Screen	Mean Score	Hit Name
ASAM	5.1	ASAM	IGSF11	6.3	IGSF11
CD80	14.0	NGFR	JAM2	12.2	JAM3
CD80	12.2	CD274	JAM2	7.5	JAM2
CD80	10.7	CD28	JAM3	10.6	JAM2
CD80	9.4	CTLA4	JAM3	6.5	JAM3
CD274	15.1	PDCD1	LAIR1	13.5	collagen1
CD274	8.1	CD80	LAIR1	8.2	LAIR1
BTLA	19.0	TNFRSF14	PVRL4	11.6	PVRL1
BTN3A1	7.4	TFF1	PVRL4	8.1	PVRL2
BTNL8	7.3	TFF1	PVRL4	4.2	TIGIT
CD160	10.2	TNFRSF14	LSAMP	11.4	NTM
CD2	6.7	CD58	LSAMP	8.1	LSAMP
CD200	8.6	CD200R1	LSAMP	4.0	NEGR1
CD200R1	15.0	CD200	MPZL3	5.2	MPZL2
CD226	10.3	PVR	NEGR1	8.2	NTM
CD226	10.0	PVRL2	NEGR1	6.7	NEGR1
CD300LD	4.7	TFF1	NEGR1	6.5	LSAMP
CEACAM1	9.9	CEACAM6	PDCD1	8.2	CD274
CEACAM1	9.7	CEACAM1	PSG5	4.2	TIE1
CEACAM1	9.2	CARTPT	PVRL1	9.2	PVRL4
CEACAM1	9.1	CEACAM6	PVRL1	8.6	PVRL3
CEACAM1	8.4	CEACAM7	PVRL1	7.2	PVRL1
CEACAM1	4.6	CEACAM5	PVRL2	14.3	CD226
CEACAM6	16.0	CEACAM6	PVRL2	13.8	PVRL3
CEACAM6	15.6	CARTPT	PVRL2	10.7	PVRL2
CEACAM6	15.2	CEACAM1	PVRL3	18.1	PVRL1
CEACAM6	14.6	CEACAM6	PVRL3	16.1	PVRL2
CEACAM6	9.5	CEACAM7	PVRL3	7.9	TIGIT
CEACAM6	8.5	CEACAM5	SIGLEC8	4.9	TFF1
CTLA4	7.3	CD80	SLAMF1	7.7	SLAMF1
CTLA4	6.6	CD80	SLAMF7	5.6	SLAMF7
CXADR	5.6	AMICA1	TIGIT	17.2	PVRL2
CXADR	5.2	CXADR	TIGIT	15.5	PVRL3
ESAM	6.5	ESAM	TIGIT	14.0	PVRL4
FCRL4	10.0	TFF1	TIGIT	10.4	PVR
FCRL4	5.0	TFF2			
NTM	10.7	LSAMP			
NTM	9.7	NEGR1			
NTM	6.2	NTM			

Green = novel & SPR validated interaction  
Blue = known or expected interaction

Note. Green: novel and SPR validated interaction. Blue: known or expected interaction. (For interpretation of the references to color in this table note, the reader is referred to the web version of this article.)

occurred in well under 1% of the protein preparations used in our study.

Each of the five nonspecific binding proteins appeared as hits in more than 10% of screens. In actuality, this 10% cutoff was fairly lenient. A cutoff of 5% would have served to eliminate 10 additional false positives. Both VSIG4-Fc and mPLG-His had a hit frequency of 6% and accounted for 10 false positives combined. However, this threshold would have also eliminated some cross-interacting Ig subfamily interactions. For example, PVRL2 (binding to CD226, PVRL2, PVRL3, PVRL4, and TIGIT) and TFF-His (binding to BTN3A1, BTN1L8, CD300LD, FCRL4, and SIGLEC8) also had hit frequencies of 6%. As more screens are conducted against the secreted protein microarray, true nonspecific interactors would be expected to maintain their hit frequency scores, whereas the values for specific interactions such as PVRL2-Fc and TFF1-His would be expected to decrease.

Nonspecific binding may also derive from the bait proteins themselves, but these are much easier to identify due to high background binding on the microarrays. In our Ig receptor set, only two baits (CEACAM4 and SIGLEC5) showed significant background levels and needed to be eliminated from the analysis. Interestingly, SIGLEC5 is a sialic acid binding protein, similar to the nonspecific prey proteins SIGLEC6 and SIGLEC9, suggesting that high background binding may be due to general sialic acid recognition.

Although it has been suggested that protein interaction networks may contain a number of “noisy” or nonfunctional interactions [30], the results presented here suggest that extracellular protein interactions are quite specific. It is clear that even between members of the Ig receptor family, where there is significant structure and sequence homology, truly selective interactions have evolved and can be identified outside the cellular context. Nevertheless, *in vivo* it is likely that temporal and spatial expression differences serve to regulate the interactions of cross-reacting Ig subfamily members such as IgLON, PVR, and CEACAM. In these cases, identification of positive interactors *in vitro* can provide a starting place for functional and spatiotemporal expression studies [19,31].

Our data demonstrate the power of protein microarrays for identifying extracellular protein interactions. Although establishing a large protein library may initially require significant resources, the small amount of protein required to generate microarrays is a great advantage. Relatively small-scale purifications can produce enough material to theoretically print thousands of microarrays. For example, with our current protocol, 10  $\mu$ g of protein would be sufficient to print more than 5000 microarrays. We anticipate that, with the development of more efficient and high-throughput methods for mammalian protein production and purification, extracellular protein microarrays can readily be expanded to cover a larger fraction of the secretome and will provide an especially powerful and rapid platform for identifying extracellular protein interactions.

## Acknowledgments

We acknowledge M. Nakamura, R. Tong, and P. Hass for management and assistance with the SPDI library, E. Christensen for purification assistance, D. Reilly and A. Wong for CHO protein expressions, Jerry Tang for compiling prey sequences, and A. Bruce for illustration assistance.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.ab.2011.09.017](https://doi.org/10.1016/j.ab.2011.09.017).

## References

- [1] M.J. Butte, V. Pena-Cruz, M.J. Kim, G.J. Freeman, A.H. Sharpe, Interaction of human PD-L1 and B7-1, *Mol. Immunol.* 45 (2008) 3567–3572.
- [2] G. Cai, G.J. Freeman, The CD160, BTLA, LIGHT/HVEM pathway: a bidirectional switch regulating T-cell activation, *Immunol. Rev.* 229 (2009) 244–258.
- [3] G.J. Wright, Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions, *Mol. Biosyst.* 5 (2009) 1405–1412.
- [4] G.J. Wright, S. Martin, K.M. Bushell, C. Söllner, High-throughput identification of transient extracellular protein interactions, *Biochem. Soc. Trans.* 38 (2010) 919–922.
- [5] H.F. Clark, A.L. Gurney, E. Abaya, K. Baker, D. Baldwin, J. Brush, J. Chen, B. Chow, C. Chui, C. Crowley, B. Currell, B. Deuel, P. Dowd, D. Eaton, J. Foster, C. Grimaldi, Q. Gu, P.E. Hass, S. Heldens, A. Huang, H.S. Kim, L. Klimowski, Y. Jin, S. Johnson, J. Lee, L. Lewis, D. Liao, M. Mark, E. Robbie, C. Sanchez, J. Schoenfeld, S. Seshagiri, L. Simmons, J. Singh, V. Smith, J. Stinson, A. Vagts, R. Vandlen, C. Wantanabe, D. Wieand, K. Woods, M.H. Xie, D. Yansura, S. Yi, G. Yu, J. Yuan, M. Zhang, Z. Zhang, A. Gaddard, W.I. Wood, P. Godowski, A. Gray, The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment, *Genome Res.* 13 (2003) 2265–2270.
- [6] L.C. Gonzalez, K.M. Loyet, J. Calemine-Fenaux, V. Chauhan, B. Wranik, W. Ouyang, D.L. Eaton, A coreceptor interaction between the CD28 and TNF receptor family members B and T lymphocyte attenuator and herpesvirus entry mediator, *Proc. Natl. Acad. Sci. USA* 102 (2005) 1116–1121.
- [7] X. Yu, K. Harden, L.C. Gonzalez, M. Francesco, E. Chiang, B. Irving, I. Tom, S. Ivelja, C.J. Refino, H. Clark, D. Eaton, J.L. Grogan, The surface protein TIGIT suppresses T cell activation by promoting the generation of mature immunoregulatory dendritic cells, *Nat. Immunol.* 10 (2009) 48–57.
- [8] G. MacBeath, S.L. Schreiber, Printing proteins as microarrays for high-throughput function determination, *Science* 289 (2000) 1760–1763.
- [9] H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R.A. Dean, M. Gerstein, M. Snyder, Global analysis of protein activities using proteome chips, *Science* 293 (2001) 2101–2105.
- [10] A. Kaushansky, J.E. Allen, A. Gordus, M.A. Stiffler, E.S. Karp, B.H. Chang, G. MacBeath, Quantifying protein–protein interactions in high throughput using protein domain microarrays, *Nat. Protoc.* 5 (2010) 773–790.
- [11] J.R. Newman, A.E. Keating, Comprehensive identification of human bZIP interactions with coiled–coil arrays, *Science* 300 (2003) 2097–2101.
- [12] S.C. Tao, H. Zhu, Protein chip fabrication by capture of nascent polypeptides, *Nat. Biotechnol.* 24 (2006) 1253–1254.
- [13] N. Ramachandran, J.V. Raphael, E. Hainsworth, G. Demirkan, M.G. Fuentes, A. Rolfs, Y. Hu, J. LaBaer, Next-generation high-density self-assembling functional protein arrays, *Nat. Methods* 5 (2008) 535–538.
- [14] S.C. Popescu, G.V. Popescu, S. Bachan, Z. Zhang, M. Seay, M. Gerstein, M. Snyder, S.P. Dinesh-Kumar, Differential binding of calmodulin-related proteins to their targets revealed through high-density *Arabidopsis* protein microarrays, *Proc. Natl. Acad. Sci. USA* 104 (2007) 4730–4735.
- [15] N. Goshima, Y. Kawamura, A. Fukumoto, A. Miura, R. Honma, R. Satoh, A. Wakamatsu, J. Yamamoto, K. Kimura, T. Nishikawa, T. Andoh, Y. Iida, K. Ishikawa, E. Ito, N. Kagawa, C. Kaminaga, K. Kanehori, B. Kawakami, K. Kenmochi, R. Kimura, M. Kobayashi, T. Kuroita, H. Kuwayama, Y. Maruyama, K. Matsuo, K. Minami, M. Mitsubori, M. Mori, R. Morishita, A. Murase, A. Nishikawa, S. Nishikawa, T. Okamoto, N. Sakagami, Y. Sakamoto, Y. Sasaki, T. Seki, S. Sono, A. Sugiyama, T. Sumiya, T. Takayama, Y. Takayama, H. Takeda, T. Togashi, K. Yahata, H. Yamada, Y. Yanagisawa, Y. Endo, F. Imamoto, Y. Kisu, S. Tanaka, T. Isogai, J. Imai, S. Watanabe, N. Nomura, Human protein factory for converting the transcriptome into an *in vitro*-expressed proteome, *Nat. Methods* 5 (2008) 1011–1017.
- [16] A.R. Aricescu, E.Y. Jones, Immunoglobulin superfamily cell adhesion molecules: Zippers and signals, *Curr. Opin. Cell Biol.* 19 (2007) 543–550.
- [17] L. Jiang, A.N. Barclay, Identification of leucocyte surface protein interactions by high-throughput screening with multivalent reagents, *Immunology* 129 (2010) 55–61.
- [18] K.M. Bushell, C. Söllner, B. Schuster-Boeckler, A. Bateman, G.J. Wright, Large-scale screening for novel low-affinity extracellular protein interactions, *Genome Res.* 18 (2008) 622–630.
- [19] S. Martin, C. Söllner, V. Charoensawan, B. Adryan, B. Thisse, C. Thisse, S.A. Teichmann, G.J. Wright, Construction of a large extracellular protein interaction network and its resolution by spatiotemporal expression profiling, *Mol. Cell. Proteomics* 9 (2010) 2654–2665.
- [20] A.A. Lobito, S.R. Ramani, I. Tom, J.F. Bazan, E. Luis, W.J. Fairbrother, W. Ouyang, L.C. Gonzalez, Murine insulin growth factor-like (IGFL) and human IGFL1 proteins are induced in inflammatory skin conditions and bind to a novel tumor necrosis factor receptor family member, IGFLR1, *J. Biol. Chem.* 286 (2011) 18969–18981.
- [21] G.K. Smyth, *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Springer, New York, 2005, pp. 397–420.
- [22] J. Silver, M.E. Ritchie, G.K. Smyth, Microarray background correction: maximum likelihood estimation for the normal–exponential convolution, *Biostatistics* 10 (2009) 352–363.
- [23] D. Voulgaraki, R. Mitnacht-Kraus, M. Letarte, M. Foster-Cuevas, M.H. Brown, A.N. Barclay, Multivalent recombinant proteins for probing functions of

- leucocyte surface proteins such as the CD200 receptor, *Immunology* 115 (2005) 337–346.
- [24] P.A. van der Merwe, A.N. Barclay, D.W. Mason, E.A. Davies, B.P. Morgan, M. Tone, A.K. Krishnam, C. Lanelli, S.J. Davis, Human cell-adhesion molecule CD2 binds CD58 (LFA-3) with a very low affinity and an extremely fast dissociation rate but does not bind CD48 or CD59, *Biochemistry* 33 (1994) 10149–10160.
- [25] M.P. Arrate, J.M. Rodriguez, T.M. Tran, T.A. Brock, S.A. Cunningham, Cloning of human junctional adhesion molecule 3 (JAM3) and its identification as the JAM2 counter-receptor, *J. Biol. Chem.* 276 (2001) 45826–45832.
- [26] S. von Gunten, B.S. Bochner, Basic and clinical immunology of siglecs, *Ann. N.Y. Acad. Sci.* 1143 (2008) 61–82.
- [27] J. Reed, C. McNamee, S. Rackstraw, J. Jenkins, D. Moss, Diglons are heterodimeric proteins composed of IgLON subunits, and Diglon-CO inhibits neurite outgrowth from cerebellar granule cells, *J. Cell Sci.* 117 (2004) 3961–3973.
- [28] S.D. Gray-Owen, R.S. Blumberg, CEACAM1: contact-dependent control of immunity, *Nat. Rev. Immunol.* 6 (2006) 433–446.
- [29] E.P. Reeves, T. Ali, P. Leonard, S. Hearty, R. O’Kennedy, F.E. May, B.R. Westley, C. Josenhans, M. Rust, S. Suerbaum, A. Smith, B. Drumm, M. Clyne, *Helicobacter pylori* lipopolysaccharide interacts with TFF1 in a pH-dependent manner, *Gastroenterology* 135 (2008) 2043–2054.
- [30] E.D. Levy, C.R. Landry, S.W. Michnick, How perfect can protein interactomes be?, *Sci. Signal.* 2 (2009) pe11.
- [31] V. Charoensawan, B. Adryan, S. Martin, C. Söllner, B. Thisse, C. Thisse, G.J. Wright, S.A. Teichmann, The impact of gene expression regulation on evolution of extracellular signalling pathways, *Mol. Cell. Proteomics* 9 (2010) 2666–2677.