

## Review Article

# Technologies for Proteome-Wide Discovery of Extracellular Host-Pathogen Interactions

**Nadia Martinez-Martin**

*Department of Microchemistry, Proteomics and Lipidomics, Genentech, South San Francisco, CA, USA*

Correspondence should be addressed to Nadia Martinez-Martin; [martin32@gene.com](mailto:martin32@gene.com)

Received 11 November 2016; Accepted 19 January 2017; Published 22 February 2017

Academic Editor: Frank A. Schildberg

Copyright © 2017 Nadia Martinez-Martin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pathogens have evolved unique mechanisms to breach the cell surface barrier and manipulate the host immune response to establish a productive infection. Proteins exposed to the extracellular environment, both cell surface-expressed receptors and secreted proteins, are essential targets for initial invasion and play key roles in pathogen recognition and subsequent immunoregulatory processes. The identification of the host and pathogen extracellular molecules and their interaction networks is fundamental to understanding tissue tropism and pathogenesis and to inform the development of therapeutic strategies. Nevertheless, the characterization of the proteins that function in the host-pathogen interface has been challenging, largely due to the technical challenges associated with detection of extracellular protein interactions. This review discusses available technologies for the high throughput study of extracellular protein interactions between pathogens and their hosts, with a focus on mammalian viruses and bacteria. Emerging work illustrates a rich landscape for extracellular host-pathogen interaction and points towards the evolution of multifunctional pathogen-encoded proteins. Further development and application of technologies for genome-wide identification of extracellular protein interactions will be important in deciphering functional host-pathogen interaction networks, laying the foundation for development of novel therapeutics.

## 1. Introduction

The plasma membrane constitutes a critical biological interface between the cytosol and the extracellular environment of the cell, and consequently membrane-tethered proteins and secreted molecules (collectively termed extracellular proteins) are essential regulators of cellular communication. From high affinity cytokine-receptor interactions to low affinity cell-cell adhesion contacts, extracellular protein-protein interactions (ePPIs) are key for information processing and coordination of virtually all processes in a living organism. Furthermore, given their fundamental functions and their accessibility to systemically delivered drugs, extracellular proteins are particularly suitable targets for therapeutic intervention. In fact, despite these proteins being encoded by approximately one-fourth of the human genes, at least two-thirds of the existing drugs target either secreted or membrane-bound proteins [1]. Thus, the elucidation of the ePPI networks on a global scale has become crucial for the biomedical research. However, in spite of their relevance

and abundance, ePPIs are remarkably underrepresented in available large-scale datasets. This discrepancy is due to the low sensitivity and limited compatibility of current high throughput technologies to detect extracellular interactions because of the unusual biochemical nature of the membrane proteins and the intractability of their binding partners [2–4]. In particular, transmembrane domain-containing proteins are amphipathic, making it difficult to solubilize them in their native conformation, and often contain posttranslational modifications such as glycans and disulfide bonds, which are not properly added in common heterologous expression systems [5]. In addition, interactions between cell surface proteins are often characterized by fast dissociation rates and therefore weak binding affinities, and in consequence well-established PPI methods such as yeast-two-hybrid or affinity purification-mass spectrometry (AP/MS) largely fail to detect these interactions. Over the last decade, several innovative technologies have been developed to overcome the aforementioned technical challenges and allow for sensitive

detection of ePPIs [2, 6–10]. Nevertheless, the mapping of ePPIs remains a major challenge in biology.

Infectious diseases result in millions of deaths each year and therefore identifying new candidate targets for improved therapeutic development remains a pressing health concern. Pathogens have evolved a myriad of elegant and often complex strategies to invade the host and commandeer host immune responses to allow pathogen replication, spread, and persistence in the infected organism. Many cell surface molecules serve as entry receptors for initial host cell invasion, and concerted responses to the pathogenic challenge critically rely on cell functions mediated by receptors and secreted proteins. To allow host colonization, pathogens encode highly optimized protein modulators, in the form of secreted molecules or receptors expressed on the plasma membrane of the infected cells or the surface of the pathogen [11, 12]. Interactions between these proteins and extracellular host molecules form the foundation of communication between a host and a pathogen and play a vital role in the initiation and outcome of the infection [13, 14]. Characterizing host-pathogen ePPI networks is therefore of utmost importance to gain a better understanding of the infection process and to inform the development of novel or improved therapeutic strategies. Excellent studies on mapping host-pathogen interactions, particularly MS-based analysis of viral infection, have provided a wealth of insight into infectious diseases [15–19]. Nevertheless, similarly to host ePPIs, a significant hurdle to the elucidation of host-pathogen biology has been the shortage of datasets of extracellular interactions between host and pathogen proteins, partly due to the technical challenges that these proteins present. Moreover, an additional consideration when studying pathogen-encoded molecules is that these proteins often lack any recognizable homology with any host molecules, thus precluding prediction of their functions [11, 20]. Robust methodologies that permit unbiased characterization of ePPI in the absence of preexisting hypotheses are thus needed to elucidate the binding partners and molecular functions of most pathogen-encoded molecules.

Excellent reviews have recently revisited the currently available technologies for proteome-wide ePPI discovery [4, 8, 53–55]. Here we discuss the application of some of these technologies to the study of host-pathogen interaction and describe some of the major findings that have recently impacted research in the field of extracellular host-pathogen recognition. Protein microarrays and functional genomics approaches are highlighted here as emerging techniques with unique potential for the elucidation of host-pathogen ePPI networks at a genome-wide scale.

## 2. Biochemical and Biophysical Approaches

Classical biochemical and biophysical approaches are particularly suitable for detection of high affinity host-pathogen interactions, such as those mediated by a viral capsid protein and a host cell surface receptor, or a pathogen-encoded glycoprotein and a secreted host factor. Typically, these approaches have relied on the utilization of recombinant pathogen proteins as baits to probe for host binding partners,

followed by immunoprecipitation and MS, or biophysical techniques for analysis of PPI such as surface plasmon resonance (SPR). SPR requires prior knowledge of the possible interacting partners and is therefore unsuitable for unbiased PPI discovery, whereas immunoprecipitation and MS approaches usually fail to detect weak interactions, which often characterize ePPIs, particularly those that take place on the cell surface. Notwithstanding, the identification of the receptors for some of the most prominent pathogens, such as the severe acute respiratory syndrome coronavirus (SARS-CoV) or the New World Arenaviruses, was made utilizing standard immunoprecipitation techniques [56, 57]. Despite their inherent limitations, biochemical approaches continue to provide relevant insights into host-pathogen interactions, such as the discovery of the dipeptidyl peptidase 4 (DPP4) as the receptor for the MERS-CoV just within a few months after the emergence of this virus [58]. In addition, more recently Kabanova and colleagues identified the cell surface receptor for the trimeric entry complex gHgLgO encoded by human cytomegalovirus (hCMV) [21]. Several studies have shown that the gHgLgO trimer is involved in the infection of fibroblasts, whereas the gHgLUL128L pentameric complex is required for entry into endothelial, epithelial, and myeloid cells [59–61]. In this work, both the trimeric and the pentameric CMV protein complexes were generated as recombinant products and used as baits to perform binding experiments on biotinylated cell surfaces, followed by immunoprecipitation and MS identification of bait-cell receptor complexes. Using this approach, the authors identified the cell surface protein PDGFR $\alpha$  as a high affinity receptor for the gHgLgO trimer and demonstrated that this interaction was required for infection of fibroblasts. Interestingly, in the case of the pentameric CMV complex, multiple bands were detected upon protein immunoprecipitation from epithelial cells, suggesting the existence of multiple receptors on these cells, which so far remain unknown [21] (Table 1).

Different biophysical techniques for detection of PPI, in particular SPR, have also proven valuable in the field of host-pathogen interactions. SPR offers the advantage of label-free, sensitive detection of interactions between a diversity of ligands in real time, thus allowing calculation of kinetic parameters. SPR has been widely utilized to monitor antibody binding to a variety of pathogen antigens, information that has informed vaccine development [62–64]. The SPR technology has also been exploited for discovery of ePPIs. For example, Viejo-Borbolla and coworkers utilized SPR to screen several secreted and membrane-expressed glycoproteins encoded by herpes simplex viruses (HSVs) for binding to chemoattractant cytokines (the chemokine family) and were able to identify a subset of human chemokines that bound to HSV glycoprotein G with high affinity [22]. Recently, Day and colleagues utilized a combination of glycan arrays and SPR and identified over 60 host-bacterial glycan pairs characterized by a wide range of binding affinities, some of which participated in bacterial adherence to host cells in vitro, leading to the hypothesis that bacteria-host surface glycan interactions may mediate initial attachment to the target cell during infection [65]. Despite SPR and related methods offering higher sensitivity for detection of transient

TABLE 1: Overview of the main high throughput approaches utilized to detect ePPI between pathogens and their respective hosts and relevant examples discussed thorough the text.

Pathogen	Identification method	Main findings	References
Human cytomegalovirus (hCMV)	Biochemical and MS	PDGFR $\alpha$ identified as a high affinity cell surface receptor for the CMV gHgLgO protein complex	[21]
Herpes simplex viruses (HSV's)	Biophysical	Secreted and plasma membrane-expressed glycoprotein G targets a specific set of human chemokines with high affinity	[22]
Human immunodeficiency virus type 1 (HIV)	Monoclonal antibodies	CD4 identified as the receptor for HIV infection of T cells	[23, 24]
Rhinovirus	Monoclonal antibodies	ICAM-1 as the common entry receptor for most rhinovirus serotypes	[25, 26]
<i>Plasmodium falciparum</i>	AVEXIS	BASIGIN identified as the cell-surface receptor that mediates erythrocyte infection	[27]
Human adenoviruses	Extracellular human protein microarrays	Elucidation of the extracellular interactome of adenovirus-encoded immunomodulatory proteins	[28]
<i>Pseudomonas aeruginosa</i>	Extracellular pathogen protein microarrays (NAPPA)	Screening of patient sera against all <i>P. aeruginosa</i> extracellular proteins. 12 proteins identified as potent antigens	[29]
Varicella zoster virus (VZV)	Extracellular pathogen protein microarrays (NAPPA)	Identification of 18 extracellular viral proteins that promote humoral responses upon screening of the entire VZV proteome	[30]
<i>Streptococcus</i>	Extracellular pathogen protein microarrays	Identification of new streptococcal proteins that interact with fibronectin, fibrinogen, and C4BP factors	[31]
Hepatitis delta virus LHDAg antigen	Plasma membrane microarrays (MPA)	150 candidate interactions identified between viral antigen and plasma membrane proteins	[32]
Simian virus 40 (SV40)	Plasma membrane microarrays (MPA)	99 candidate interactions between whole particles and plasma membrane proteins identified	[32]
<i>Vaccinia virus</i> (VACV)	TRICEPS (MS)	7 candidate cell surface binding partners identified for VACV	[33]
Viral pathogens	Computational studies	Insights into global principles of virus-host PPI networks	[15–18, 34–36]
Hepatitis C virus (HCV)	cDNA libraries	CD81, Claudin-1, and Occludin as cell surface receptors and some of the players involved in HSV internalization	[37–39]
Adenovirus and coxsackievirus B	Monoclonal antibodies and cDNA libraries	CAR identified as a common entry receptor for adenovirus 2/5 and coxsackievirus B	[40]
Sindbis virus	siRNA screens	NRAMP as cell surface receptor for entry into <i>Drosophila</i> cells	[41]
Murine norovirus	CRISPR/Cas9	CD300lf identified as a cell surface receptor that determines virus tropism	[42]
Bacterial distending toxins ( <i>E. coli</i> )	Haploid cell screens	Sphingomyelin synthase 1 and the putative G protein-coupled receptor TMEM181 identified as toxin receptors	[43, 44]
<i>Clostridium difficile</i> and <i>perfringens</i> toxins	Haploid cell screens	Lipolysis-stimulated lipoprotein and the low-density lipoprotein receptor-related protein 1 identified as the receptors for the bacterial toxins, respectively	[45, 46]
Ebola virus	Haploid cell screens	HOPS proteins and the Niemann-Pick C1 (NPC1) transporter identified as endosomal receptors that mediate cytosol access	[47]
Lassa virus	Haploid cell screens	LAMP1, lysosomal-associated membrane protein 1 identified as an essential host factor mediating virus release to cytosol	[48]
Adeno-associated virus (AAV) serotype 2	Haploid cell screens	46 cell host factors identified, including heparin sulfate proteoglycan biosynthesis and intracellular transport genes. The immunoglobulin domain-containing transmembrane protein KIAA0319L identified as AAV receptor	[49]

TABLE 1: Continued.

Pathogen	Identification method	Main findings	References
<i>Plasmodium falciparum</i>	Population genomics analysis	<i>P. falciparum</i> EBL-1 protein binds to the erythrocyte receptor glycophorin B, a highly polymorphic gene in malaria-endemic regions	[50]
<i>Streptococcus</i>	Phage display	19 bacterial proteins identified as potential fibronectin-binding proteins	[51]
<i>Fusobacterium nucleatum</i>	Transposon-based mutant libraries	The bacterial protein Fap2 binds to the receptor TIGIT and downregulates NK-mediated killing of tumor cells	[52]

AVEXIS, Avidity-based extracellular interaction screen; NAPPA, nucleic-acid programmable protein array; MPA, microfluidic-based comprehensive protein array; CRISPR, clustered regularly interspaced short palindromic repeat; MS, mass spectrometry; PPI, protein-protein interaction.

PPIs than most biochemical approaches, these biophysical techniques have not been exploited for large-scale ePPI discovery, possibly due to the low throughput of the available instrumentation and the overall difficulties for generation of the relevant protein libraries.

These studies, among many others, have demonstrated the power of the classical biochemical and biophysical techniques for detection of host-pathogen interactions. Nevertheless, these approaches require previous knowledge of the pathogen-encoded proteins responsible for binding and the ability to produce such proteins as recombinant reagents, which may be challenging, as exemplified by the production of hCMV entry complexes [21, 66]. Alternative methods have been utilized in those cases where there is no previous knowledge of the pathogen proteins required for interaction with the host cells. In this regard, the screening of large collections of monoclonal antibodies raised against membrane proteins has proven particularly useful to identify receptors that mediate viral entry. Back in the early 80s, the discovery of CD4 as an entry receptor for the human immunodeficiency virus type 1 (HIV) significantly impacted our understanding of viral pathogenesis and subsequent development of therapeutics [23, 24]. In this case, the well-defined tropism of the virus determined the choice of over 100 antibodies directed against human leukocyte differentiation antigens, of which only antibodies that recognized the surface receptor CD4 blocked viral infection [23]. It is worth noting that similar monoclonal antibody screens have also been utilized for unbiased characterization of viral blockers. For example, Colonno and colleagues performed a screen of more than 2,000 hybridomas from mice immunized with preparations of plasma membranes from human cells and were able to find one antibody that blocked rhinovirus binding to its cell surface receptor [25], identified as the intercellular cell adhesion molecule 1 (ICAM-1) in subsequent studies [26].

Despite the undoubted importance of the biochemical and biophysical approaches to the study of host-pathogen interactions, the aforementioned limitations have motivated the development of alternative technologies for large-scale analysis of ePPIs.

### 3. Protein Microarrays

From the initial utilization of microarrays for detection of PPI over a decade ago, human proteome chips containing thousands of recombinant proteins have been generated, some of which are now commercially available. Protein microarrays offer the unique advantage of requiring minimal consumption of protein reagents, fast readouts, and relatively more affordable instrumentation. Typically, a fluorescently labeled or tagged protein of interest (the bait) is generated as a recombinant product and screened against all proteins in the array [10, 53]. Despite the increased availability of high-coverage protein arrays, very few are focused on extracellular proteins and therefore are not suitable for study of host-pathogen ePPIs. Most existing microarray-based methodologies rely on multimerization of the bait protein for increased avidity and detection of weak ePPIs, mimicking the way these interactions occur in vivo, where proteins are arrayed

in the crowded molecular environment of apposing plasma membranes. Different multimerization strategies and protein microarray libraries have been developed and utilized for host-pathogen interaction discovery, some of which are described in more detail below.

#### 3.1. Avidity-Based Extracellular Interaction Screen (AVEXIS).

The Wright lab developed a novel method for detection of low affinity ePPIs, termed avidity-based extracellular interaction screen (AVEXIS) [2]. In brief, AVEXIS consists of the expression of the extracellular domain (ECD) of the bait of interest as a recombinant protein, which retains its binding properties while removing the insoluble transmembrane region of the protein. These ECDs are tagged with a coiled-coil sequence from the rat cartilage oligomeric matrix protein to allow for pentamerization of the bait and therefore increased binding avidity, alongside a  $\beta$ -lactamase tag for detection of bait-prey interactions upon incubation with the colorimetric substrate nitrocefin. This multivalent strategy has been used for medium-scale screens, allowing detection of weak interactions between human receptors with low false-positive rates [2]. Notably, Crosnier and colleagues applied AVEXIS to search for the plasma membrane receptor responsible for *Plasmodium falciparum* infection of erythrocytes [27]. The authors compiled a library consisting of most secreted or cell surface-expressed proteins in erythrocytes and systematically assayed more than 40 red blood cell proteins for binding to *P. falciparum* protein PfRh5, a parasite protein essential for blood stage growth, expressed as an AVEXIS pentameric bait. Notably, the Ok blood group antigen BASIGIN was identified as a unique receptor for PfRh5, and inhibition of this interaction was shown to be sufficient to block parasite invasion of the erythrocyte, findings that may importantly inform antimalarial therapies [27]. In later studies, AVEXIS was miniaturized making this approach compatible with the protein microarray format, thus permitting more comprehensive and lower resource-intensive screenings [67]. Although this technique should allow for high throughput and sensitive determination of ePPIs, this approach has not yet been applied to elucidation of pathogen-host interactions.

**3.2. Extracellular Protein Microarray Platforms.** Over a decade ago, fueled by the recent completion of the human genome, Genentech pioneered a significant effort to identify novel secreted or transmembrane domain-containing proteins, upon careful bioinformatics assessment and high throughput protein purification [68]. These efforts resulted in the generation of a comprehensive human protein library, which was subsequently utilized to develop an extracellular protein microarray platform, consisting of over 1,500 secreted or single-transmembrane domain containing proteins [10]. For the generation of this human protein library, secreted proteins or the ECD of single-transmembrane receptors were fused to different affinity tags and subsequently purified from cell culture supernatants by size-exclusion chromatography. Mammalian cells or baculovirus-insect cells were preferentially used as expression systems, to maximize the likelihood of proper folding and glycosylation of the extracellular protein collection [10, 69]. SDS-PAGE and multiangle laser



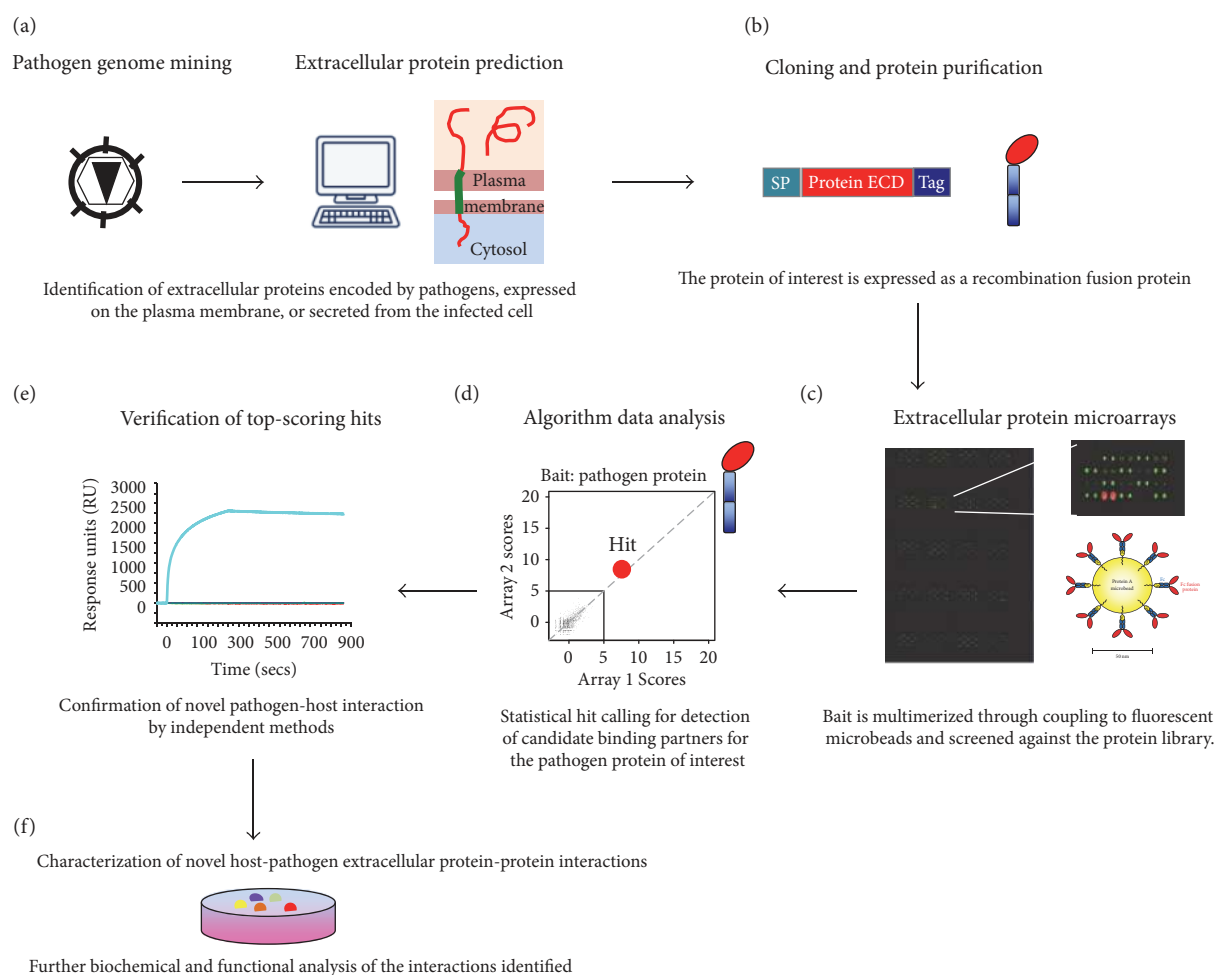
light scattering were used to analyze noncovalent aggregation and ensure high-quality protein production. Subsequently, the purified proteins were spotted on epoxysilane slides using a NanoPrint Arrayer, and protein immobilization on the microarray was determined by probing the slides with the relevant anti-tag antibodies [10]. To enhance detection of low affinity interactions, a rapid method to assemble bait proteins (whose ECD was expressed as a Fc tag-fusion protein) into multivalent complexes using fluorescently labeled protein A microbeads was developed. Proof-of-concept assays showed high sensitivity for detection of weak ePPIs characterized by micromolar  $K_D$ , a minimal off-target binding, and more than 70% true-positive to false-positive detection ratio [10, 69]. Over the years, this extracellular protein microarray has successfully identified counterreceptors for a number of human molecules, providing relevant insights into novel pathways that coordinate a multitude of cell functions [70–72].

**3.3. Protein Microarrays for Viral Immunomodulatory Protein Receptor Discovery.** Recently, we applied this ePPI discovery platform to the study of extracellular viral proteins (Figure 1), with a focus on human adenovirus- (HAdV-) encoded immunomodulatory proteins [28]. Despite the increasing relevance of HAdV as both pathogens and therapeutic vectors, information on the interaction of these viruses with the host immune system remains scarce [73, 74]. Interestingly, the immunomodulatory proteins encoded by these viruses, termed E3 proteins, show substantial diversity in their ECDs across and within viral species and constitute one of the most divergent regions of the HAdV genome [75, 76]. Given this striking variability, the E3 proteins have been suggested to play a role in viral tropism and pathogenesis, yet the functions of virtually all E3 proteins have remained unknown [73]. In our study, we took advantage of such unique variability to evaluate the effect of viral immunomodulatory protein diversity in extracellular host targeting. Screening of a substantial number of E3 proteins encoded by different HAdV species using the extracellular protein microarray platform allowed identification of over 50 novel virus-host interactions encompassing 5 viral species, which were fully validated by orthogonal methods [28]. These findings revealed significant diversity in extracellular host targeting and, moreover, allowed identification of semiconserved host targets, pointing towards specific human receptors that may represent previously unrecognized hubs for viral perturbation. Furthermore, most of the E3 immunomodulators were identified as multifunctional proteins, suggesting that viruses have evolved proteins capable of interfering with several cellular functions, a strategy consistent with the optimization of limited genomic resources. Such economic targeting has been often observed in intracellular targeting [15–17], but so far few examples of widespread targeting in the extracellular environment have been reported [28, 77–80], let alone a global elucidation of ePPI networks, in part due to the technical challenges associated with ePPI detection.

Remarkably, many of the HAdV E3 proteins preferentially interacted with host receptors that exert known or predicted inhibitory functions during the immune response (as defined

by the presence of intracellular immunoreceptor tyrosine-based inhibitory motif, ITIM), including LILRB1 [81, 82], LAIR1 [83], and MPZL1 [84], suggesting previously unrecognized strategies of immunosuppression that may be utilized by other human pathogens [28]. Moreover, several of the receptors identified as targets for the viral proteins in this study (including the prominent cell surface molecule CD45) do not have known counterreceptors in the host, supporting the longstanding hypothesis that pathogen molecules drive the evolution of immune receptors and in many instances may represent the most relevant modulators of host receptor function [85–87]. In summary, such unbiased, microarray-based study of immunomodulatory proteins represented the first large-scale analysis of the PPI landscape of a collection of extracellular immunomodulators encoded by viruses. Future investigation of other pathogen-encoded molecules using similar extracellular protein microarrays will likely shape our understanding of the pathogen imprint in our immune system.

**3.4. In Vitro Transcription and Translation- (IVTT-) Based Microarrays.** One of the main limitations of any protein microarray platform is the lower protein coverage relative to other genome-wide methods for PPI identification, due to the costs and difficulties for generation of comprehensive protein libraries to be deposited onto the microarrays. In an attempt to address this caveat, Ramachandran and colleagues developed a method called nucleic-acid programmable protein array (NAPPA), in which DNAs are directly deposited onto the array followed by protein synthesis in situ using an in vitro transcription and translation (IVTT) system, thus avoiding the need for protein purification [88]. Although this promising approach has proven superior in generating transmembrane-containing molecules as soluble proteins, it still remains to be systematically addressed if the extracellular human proteins produced in this manner present the folding and posttranslational modifications necessary for protein functionality. Nevertheless, emerging data support the utility of NAPPA as a useful tool for the study of bacterial proteins. For example, Montor and colleagues used a bioinformatics approach to predict the *Pseudomonas aeruginosa* proteins that reside in the outer membrane of the bacteria or are secreted to the extracellular environment of the infected cell [29]. In this work, the authors utilized the NAPPA approach to screen all predicted extracellular gene products for interaction with sera from cystic fibrosis patients, where *P. aeruginosa* establish a life-threatening lung infection. From 266 bacterial proteins initially selected, 12 proteins were recognized by antibodies in the sera, indicating that these bacterial proteins represent major antigens that trigger adaptive immune responses in humans. Interestingly, robust antibody responses against three previously uncharacterized proteins were detected, suggesting this approach could help identify new extracellular proteins that exert unknown functions during the infection [29]. These results confirmed the utility of the microarrays to detect immune responses against membrane proteins encoded by pathogens, and supported the use of this methodology for diagnosis applications. In this regard, several groups have developed microarrays composed



**FIGURE 1: Overview of the application of the protein microarrays technology for extracellular pathogen-host protein-protein interaction discovery.** (a) Identification of genes that encode for secreted factors or cell surface-expressed proteins, based on published data or bioinformatics analysis. (b) Cloning, expression, and purification of the pathogen-encoded proteins of interest. The full-length protein (secreted proteins) or the ECD (transmembrane-containing proteins) is fused to a tag for subsequent expression in the heterologous system of choice followed by affinity purification. Mammalian or baculovirus-based systems are preferred to allow for introduction of posttranslational modifications. (c) Screening of the selected pathogen-encoded proteins (baits) against extracellular human protein libraries using protein microarray technologies. Different strategies for bait multimerization have been developed to allow for detection of lower affinity interactions (see text for details). A multimerization strategy based on the coupling of Fc-tagged baits to fluorescent protein A microbeads is shown. Additional microarray-based technologies have been developed to avoid the need for extensive protein purification associated with library generation (see text for details). (d) Algorithm analysis of the protein microarray data. Frequent nonspecific binders in the human library are filtered out, and binding partners for the pathogen protein under study are depicted as high-scoring, intersecting hits. (e) Validation of the interaction between the pathogen-encoded protein of interest and the novel receptor(s) identified in the screens. Experimental validation of the protein-protein interactions may be performed using orthogonal approaches, such as surface plasmon resonance, immunoprecipitation, and flow cytometry. (f) Selected binding partners may be further characterized biochemically and functionally to assess the relevance of the novel pathogen-host interactions identified. SP, signal peptide; ECD, extracellular domain.

of pathogen-encoded proteins [30, 89–93]. Such pathogen protein arrays have so far been exploited mainly for diagnosis purposes, to allow screening of antibodies present in patient sera for binding to extracellular bacterial or viral antigens on the array. Nevertheless, their inherent high throughput and compatibility with multivalent bait approaches makes them a powerful tool for ePPI discovery. For example, Margarit et al. developed a *Streptococcus* microarray to find novel microbial proteins capable of binding to the human proteins fibronectin, fibrinogen, and C4BP and were able to identify

a set of streptococcal proteins that interacted with these factors [31]. Nevertheless, despite such pathogen protein-based arrays offering great promise, this methodology remains to be systematically analyzed for ePPI discovery.

More recently, Yu and colleagues applied NAPPA technology in combination with the HaloTag-Halo ligand detection system to elucidate the interaction network of two effector proteins (SidM and LidA) encoded by *Legionella pneumophila*, a highly pathogenic bacteria that is the causative agent of Legionnaire's pneumonia [94]. Similarly to many

pathogen proteins, these virulence factors lack significant homology to host molecules therefore complicating the assessment of their host targets and biological functions. In this work, the bacterial proteins of interest were tagged with a HaloTag, a modified haloalkane dehalogenase that covalently binds to synthetic Halo-ligands (haloalkanes) that can be fluorescently labeled, thus allowing more robust detection of bait protein binding to interactors present on the array. In this study, more than 10,000 human proteins were expressed on the NAPPA array using different IVTT techniques, leading to identification of 20 and 18 binding partners for the LidA and SidM effectors, respectively, most of them experimentally verified by pull-down [94]. Although this study focused on identification of intracellular PPI, the applicability of the NAPPA-HaloTag technology for ePPI determination should be explored in the future. Moreover, bait multimerization strategies should be implemented in order to make this approach more suitable for detection of transient PPIs.

Related to the NAPPA technology, Glick and colleagues recently built a miniaturized platform focused on human membrane proteins. By integrating the microfluidics technology, protein microarrays, and an IVTT system, this group built a new device named microfluidic-based comprehensive human membrane protein array (MPA) [32]. A notable improvement introduced by these investigators was the addition of microsomal membranes to the IVTT system to allow for improved folding and posttranslational modifications in plasma membrane proteins, both common limitations of IVTT systems. In this work, a library of 2,700 human genes encoding for membrane proteins was built and subsequently utilized to screen the large-form delta antigen (L-HDAg) encoded by the hepatitis delta virus (HDA) and whole viral particles of the simian virus 40 (SV40), a nonenveloped human pathogen. Proof-of-concept assays showed encouraging results, with over 75% true-positive rate within a small set of proteins with known interactors and, more importantly, indicated the feasibility of this approach for expression of multitransmembrane-containing proteins, a protein type that has proven challenging given their high hydrophobicity. The MPA screens identified 99 and over 150 interactions for SV40 particles and L-HDAg viral protein, respectively, and around 35 interactions were validated by coimmunoprecipitation or protein-fragment complementation assays [32]. To our knowledge, this is the first study to assess ePPIs using a comprehensive human protein library and whole viral particles (SV40) as baits, a valuable approach that may provide important insights into pathogen tropism, alongside a molecular explanation for the cell surface receptors engaged by the pathogen. Further utilization of this platform followed by a more systematic analysis of the candidate hits, including nonspecific binder determination, will be needed to assess the overall performance of the MPA technology. Regardless, this platform provides an extended version of the NAPPA approach that focuses on mammalian ePPIs and may therefore provide relevant insights into extracellular host-pathogen interactions.

The protein microarrays have represented one of the most fruitful approaches for unbiased determination of ePPIs, including host-pathogen interactions. Nevertheless, one of

the main limitations of this technology is the need to generate comprehensive libraries, a process that is resource consuming and often not available to many researchers [53, 93]. Consequently, although some of the available arrays were designed to cover a significant fraction of the human proteome, any discoveries made using these platforms are limited to the proteins present in each array. The current libraries are likely to continue expanding alongside innovative approaches to facilitate sensitive detection of ePPI using protein microarray formats.

#### 4. Mass Spectrometry-Based and Computational Approaches

Over the last decade, MS-based technologies have emerged as a versatile, powerful approach to decipher many aspects of the human proteome, including the characterization of protein complexes. Excellent reviews on current MS-based technologies, recent improvements, and future prospects for elucidation of PPI networks are available [95–98]. In this review, we briefly discuss the applications of some of these techniques to the study of extracellular host-pathogen interactions.

*4.1. Mass Spectrometry-Based Characterization of Membrane Proteins and Interacting Partners.* The proteins expressed on the surface of pathogens mediate functions necessary for survival, replication, immunoevasion, and transmission and therefore are logical candidates for therapeutic and vaccine design. However, the study of the surface proteome in pathogens, particularly in bacteria is constrained by the fact that commonly used prediction algorithms fail to correctly predict the location of several proteins [29, 99–101]. Despite the characterization of the extracellular proteins and their interactions still representing the Achilles heel of most proteomics methods, MS has emerged as an invaluable approach to characterize the protein composition of plasma membranes [5]. To date, several studies have exploited MS-based techniques to gain insights into the extracellular protein composition of bacterial pathogens [101–104]. For example, Palmer and colleagues studied the surface proteome of the tick-borne intracellular pathogen *Anaplasma marginale* (Rickettsiales: Anaplasmataceae) using liquid chromatography and tandem MS [105]. Interestingly, the authors found that the surface proteome of *A. marginale* isolated from tick cells, despite being less complex than that of bacteria isolated from human erythrocytes, contained a novel protein, which the authors hypothesized to play a function in human cell invasion in spite of its human counterreceptor remaining uncharacterized. This interesting observation suggests a remodeling of the bacteria surface proteome during the transition between mammalian and arthropod hosts, an aspect of the infection that could be targeted to block transmission. Similarly, several studies have pursued the identification of the proteins present in viral particles utilizing MS. Although these analyses suffer from several drawbacks associated with membrane protein characterization, particularly the poor solubility of these proteins and the low abundance of many plasma membrane



proteins, these studies have revealed a complex composition for most of the viruses studied, alongside incorporation of many host proteins in the virions, in most cases with undetermined functions [106–108].

An interesting observation from some of the studies referred above is the fact that certain bacterial proteins, predicted cytoplasmic by consensus, can be found in the extracellular environment of the cell, where they may play alternative functions. In fact, the number of proteins that are secreted through noncanonical signal sequence pathways is increasingly appreciated [99, 100, 109, 110]. Little is known about these bacterial proteins originally described as cytosolic proteins but capable of exerting functions on the cell surface, which some authors have named moonlight proteins, in reference to their potential to exert multiple functions [111]. There is emerging evidence that protein moonlighting contributes to virulence of important bacterial pathogens including *Staphylococcus aureus* or *Mycobacterium tuberculosis*, sometimes in fascinating ways. For example, *M. tuberculosis* is known to encode two molecular chaperones, Cpn60.1 and Cpn60.2, which function as modulators of myeloid cells among other regulatory functions [112]. Despite these chaperones being by definition cytosolic, Cpn60.2 has been detected in significant amounts on the bacterial surface, and either recombinant Cpn60.2 or antibodies against this protein efficiently block binding of *M. tuberculosis* to macrophages [112], through a potential interaction with the receptor CD43 [113]. In addition, the protein DnaK, a Hsp70-related protein encoded by *M. tuberculosis*, can locate to the bacterial surface and functionally interact with CD40 [114] and with the HIV coreceptor CCR5 [115]. Notably, DnaK appears to block HIV binding to CCR5 in vitro, an interesting observation given the co-occurrence of *M. tuberculosis* and HIV infection [116]. Although a more detailed revision is out of the scope of this review, bacterial protein moonlighting, excellently revisited by Henderson and Martin [111], is a thought-provoking phenomenon that suggests a much more complex extracellular landscape than anticipated. Moreover, such protein moonlighting is in line with the hypothesis that pathogens have evolved multifunctional proteins as a prominent strategy for efficient use of limited genomic resources [17, 28].

Another MS-based approach that holds great promise for host-pathogen ePPI detection is the recently developed TRICEPS [33]. TRICEPS is a chemoproteomic reagent that consists of three moieties, one that binds the ligand of interest through its amino groups, a second one that binds glycosylated receptors on the cell surface, and a biotin tag for purifying the receptor peptides for subsequent identification by MS. Notably, in the initial description of the method, TRICEPS was successfully applied to the identification of receptors for extracellular ligands of diverse nature, such as secreted glycoproteins, small peptide ligands for G protein-coupled receptors, and therapeutic antibodies. Importantly, this approach has also been utilized to study cell surface molecules targeted by vaccinia virus (VACV). Interestingly, the analysis of VACV binding to HeLa cells revealed seven candidate binding partners, including the previously identified receptors AXL, chondroitin sulfate proteoglycan 4, and

laminin binding protein dystroglycan 1. Further, downregulation of five out of the seven candidates using short interfering RNA reduced VACV infection by 40–60%, supporting the functionality of the interactions identified, at least in vitro [33]. Although this technology is still developing and no studies on other pathogens have been published yet, future TRICEPS-based studies promise relevant insights into pathogen interaction with distinct components of the cell surface.

**4.2. Computational Approaches for Characterization of Pathogen-Host Interactions.** As an addendum to the vast amount of knowledge acquired using MS approaches and some of the additional methodologies discussed in this review, bioinformatics offers an in silico systems biology approach that reveals a global perspective on host-pathogen interactions. Advances in computation have been fundamental to dissect the complex datasets generated in many genome-wide MS-based studies and have enabled the reconstruction of large-scale host-pathogens PPI networks, providing fundamental insights into viral disease and hence host biology [15–18, 34–36, 117]. Although the computational tools available for analysis of large datasets, in some cases developed in association with some of the high throughput screens mentioned in this review, certainly deserve a focused chapter, a couple of observations are specially notable. Commonly observed in these studies is that the intracellular viral effectors preferentially target host proteins that act as hubs (proteins with many interacting partners) or bottlenecks (proteins central to many pathways in the network) [15, 16, 36]. For example, Dyer and colleagues built a network of host-pathogen PPI by integrating published information from 190 pathogens [36]. Supporting previous findings, this analysis indicated that pathogen-encoded proteins preferentially interfere with host molecules that control critical cellular processes, such as cell death or nuclear transportation, possibly as a strategy to maximize control of the host machinery given limited genomic resources. Interestingly, this study highlighted a small set of extracellular host proteins recurrently targeted by several of the viral and bacterial pathogens analyzed, including cell surface receptors such as VEGFR2/KDR and collagen, possibly indicating previously unrecognized roles in the immune response against pathogens. Although informative, the analysis performed by Dyer and collaborators was skewed towards viruses, with a prominent enrichment in HIV strains [36]. More extensive analyses encompassing other human viruses and bacterial pathogens may reveal general strategies of immunomodulation and potential human targets suitable to therapeutic intervention. Interestingly, increasing evidence suggests that virus-host interactions are governed by principles distinct to those that dictate within-host interactions [20, 28, 85, 87, 118]. Notably, detailed analyses carried out by the Xia group highlighted significant differences between virus-host and within-host (also called endogenous) interactions, such as the tendency of viral proteins to compete with host proteins for binding to a given receptor in the absence of sequence similarity with the host counterpart or the observation that viral molecules have evolved multiple short linear motifs capable of mediating a number of diverse interactions

[20, 118], features that are consistent with the multifunctional capabilities of some pathogen-encoded proteins [20, 22, 28, 77–80, 118]. Altogether, bioinformatics analysis of virus-host interactions suggest that virus-mediated targeting of host proteins is characterized by signatures of pleiotropy, economy, and convergent evolution, conclusions that are supported by emerging experimental data. Followed by thorough biological experimentation such computational-based systems biology approaches will provide a unique tool to help decipher basic global principles of pathogen-host interaction and may reveal novel ePPIs amenable to therapeutic intervention.

## 5. Genetic Screens

**5.1. Complementary DNA Libraries and RNA-Interference-Based Approaches.** Alongside protein microarrays-based technologies, MS, and computational analysis, the explosion of the functional genomics field in the last years has revolved the avenues to study pathogen interactions with their hosts, often in high throughput. In brief, genetic screens comprise gain-of-function and loss-of-function strategies, represented by complementary DNA (cDNA) libraries and RNA-interference- (RNAi-) based approaches, respectively. These methods were developed more than two decades ago and have been widely utilized by the scientific community, providing fundamental insights into the infection process. In particular, the cDNA libraries have proven extremely successful in identifying viral receptors through a gain-of-function approach, upon transduction of the cDNA library from a susceptible cell line into nonpermissive cell lines. The use of cDNA libraries is not reviewed in detail here in the interest of a more comprehensive revision of relative newer genomics-based approaches, such as the clustered regularly interspaced short palindromic repeat/CRISPR-associated protein 9 (CRISPR/Cas9) or the haploid cell screens. Nevertheless, these libraries have represented one of the most significant technologies to further our understanding of the pathogen-host interaction. For example, early studies made use of cDNA libraries to shed light on the complex mechanism exploited by hepatitis C virus for initial invasion of the cell [37–39], identified CAR as a common receptor for adenovirus 5 and coxsackievirus B [40], and were instrumental to identify SLAMF1 and PVR as receptors for measles and poliovirus, respectively [119, 120].

In turn, the RNAi technology has yielded significant insights into virus-host interactions, such as the identification of the ion transporter NRAMP as the receptor for the mosquito-borne Sindbis virus colonization of *Drosophila* cells [41]. The main power of the RNAi technology is that it allows high throughput genome-wide screens and therefore potential identification of essential factors that play roles in different aspects of the pathogen life cycle, including initial interaction with the host cell. Although RNAi screens have provided tremendous insights into host-pathogen interactions and remain widely utilized [121], inefficient gene depletion and off-target effects are important limitations of this methodology [122].

**5.2. CRISPR/Cas9-Based Screening Technology.** The increasingly popular CRISPR/Cas9 technology overcomes some of the caveats often associated with genetic manipulation and holds enormous promise for genome editing and downstream applications, including host-pathogen interaction discovery [123]. Although still early days, high throughput CRISPR/Cas9 screens for genome-wide studies have already displayed remarkable results, with high levels of genomic modification, hit confirmation, and strong phenotypic effects [124]. The development of the CRISPR/Cas9 technology has undoubtedly transformed the functional genetic analysis in mammals. Recent studies have applied the CRISPR/Cas9 technology to ablate expression of previously identified receptors for viral entry, such as the HIV coreceptors CXCR4 and CCR5, leading to resistance to infection in primary cells [125, 126]. An interesting additional application of CRISPR/Cas9 is the direct editing of viral genes important for viral fitness. This approach has recently been used to target HSV-1, CMV, and Epstein-Barr virus (EBV) essential genes, leading to a significant decrease of viral replication [127]. These studies suggest the potential use of CRISPR/Cas9 as an innovative therapeutic strategy, as aspect that will surely be further explored in the near future.

Another prominent example published recently is the identification of host factors that confer susceptibility to the evolutionary related type III secretion systems, T3SS1 and T3SS2, encoded by *Vibrio parahaemolyticus* [128]. The T3SSs are highly complex nanomachines utilized by gram-negative pathogens to inject a variable repertoire of virulence factors into the cytosol of the eukaryotic cells, enabling pathogen adhesion and internalization of modulation of host processes. Interestingly, using genome-wide CRISPR/Cas9 screens, sulfation and fucosylation of cell surface components were identified as host determinants of T3SS1- and T3SS2-mediated cytotoxicity, respectively. The authors hypothesized that interactions between sulfated cell surface molecules such as host proteoglycans and bacterial adhesins act as facilitators of T3SS1 activity, whereas fucosylated glycans on the surface may serve as receptors for T3SS2 components necessary for insertion of the complex in the host membrane [128]. The CRISPR/Cas9 approach has just started to reveal its power as a tool for unbiased identification of novel ePPIs, elegantly exemplified by the identification of CD300lf as the cell surface receptor for noroviruses, which, strikingly, was identified as the main determinant for the tropism of the murine norovirus [42]. Further optimization of this technology will unequivocally signify a tremendous advance for the discovery of extracellular host-pathogen PPIs, the processes underlying host-pathogen interactions and its possible therapeutic applications.

**5.3. Haploid Genetic Screens.** Haploid cells, in turn, allow the study of recessive phenotypes that can be masked in diploid cells, due to the difficulties of creating true genetic knockouts in mammalian cells. Despite yeast being a useful tool due to the simplicity of obtaining relevant mutants at its haploid life stage, the majority of human pathogens do not replicate in yeast therefore limiting the applicability of this approach [129]. In recent years, human haploid cells have

been increasingly utilized for genome-wide loss-of-function genetic screens using insertional mutagenesis [43, 44, 47]. In initial studies, Carette and colleagues took advantage of the KBM7 cell line, a derivative of the chronic myeloid leukemia cell line (CML) with a haploid karyotype except for chromosome 8 [130]. Using gene-trap retroviruses for efficient insertional mutagenesis, the authors generated a genome-wide collection of null mutants for most nonessential genes [43]. This approach was successfully utilized to identify host factors essential for the functions of the distending toxins or CDTs, potent virulence factors secreted by a number of pathogenic bacteria. In particular, mutagenized KBM7 cells were treated with *Escherichia coli*-derived CDTs and resistant clones were isolated, leading to identification of insertions in the sphingomyelin synthase 1 and the putative G protein-coupled receptor TMEM181, suggesting that this molecule may serve as a surface receptor for the toxin [43, 44]. Similar haploid screens have identified novel receptors for a number of bacterial toxins, including the lipolysis-stimulated lipoprotein receptor for the *Clostridium difficile* transferase [45], or the low-density lipoprotein receptor-related protein 1 as a host receptor of the *Clostridium perfringens* TpeL toxin [46].

In a later study, Carette et al. generated a KBM7-derived cell line named HAP1, haploid for all chromosomes [47]. Similarly to previous studies, the authors used the retroviral gene-trap approach to mutagenize HAP1 cells followed by deep sequencing to map more than 800,000 insertions. In this study, using a replication competent vesicular stomatitis virus (VSV) carrying the Ebola virus glycoprotein, a previously unknown entry receptor for Ebola virus was identified. Notably, these haploid cell screens identified six members of the HOPS complex, proteins known to play functions in endosomal/lysosomal trafficking, as well as the Niemann-Pick C1 (NPC1) transporter as the most prominent hit of the assay. It is worth noting that NPC1 is not a surface molecule but rather an endosomal receptor. These findings led the authors to propose a novel mechanism of entry by which Ebola virus is internalized into the endocytic pathway, followed by endosome maturation and cleavage of the surface glycoprotein of the virus. Endosome fusion, mediated by the HOPS complex, would allow interaction with NPC1 containing endosomes, triggering fusion and release of the viral genome into the cytosol. Multiple cell surface receptors can lead to internalization of the Ebola virus into the endocytic pathway [131]; such redundancy in receptor usage likely explains why these receptors were not identified in the haploid cell screen [47]. Notably, independent studies have confirmed that NPC1 acts an intracellular receptor for Ebola, including a chemical screen approach, a study showing NPC1 dependence for infection of otherwise nonsusceptible cells, and more recently the elucidation of the crystal structure of this receptor bound to the Ebola virus glycoprotein [132–134].

Interestingly, after the aforementioned haploid genetic screens identified NPC1 as a noncanonical entry receptor (given its intracellular localization), other filoviruses have been shown to take advantage of this receptor [135]. The relevance of this intriguing mechanism of viral entry is further reinforced by recent work on Lassa virus, an Old

World Arenavirus that, similarly to Ebola virus, causes severe to fatal hemorrhagic disease in humans [48, 136]. A genome-wide haploid screen using VSV pseudotyped with Lassa glycoprotein was performed in order to identify host factors essential for viral entry. Although  $\alpha$ -dystroglycan (DAG1) was long recognized as the cell surface receptor for Lassa virus, additional factors were suspected, given the observation that certain DAG1-expressing cells are resistant to infection. The authors elegantly demonstrated that at a neutral pH, the Lassa virus glycoprotein was bound to DAG1, whereas upon exposure to lower pH (resembling the lysosome environment), a receptor switch occurred leading to strong association with the lysosomal-associated membrane protein 1 (LAMP1) [48]. Thus, similarly to Ebola virus, in the model suggested the virus would be incorporated into the endocytic pathway after interaction with its surface receptor DAG1, followed by increasingly acidic conditions that would result in interaction with LAMP1 in the lysosomal membrane, triggering membrane fusion and release of the virus in the cytosol [48].

More recently, Pillay and colleagues applied the haploid cell screening approach to the identification of host factors essential for the adeno-associated virus (AAV) serotype 2 infection, one of the leading vectors for virus-based genes therapies [49]. Notably, the most significantly enriched gene in these screens was KIAA0319L, a poorly characterized type I immunoglobulin domain-containing transmembrane protein named hereafter as the AAV receptor. Among the 46 host factors identified as hits, many were implicated in heparin sulfate proteoglycan biosynthesis as well as a number of proteins that participate in intracellular transport processes. AAV is known to attach to the cells using heparin sulfate proteoglycans and hijacks endosomal trafficking to travel to the nucleus upon invasion of the cell; thus the authors hypothesized that these additional factors may influence virus tropism [49].

Altogether, these studies elegantly demonstrate the power of genome-wide screens in human haploid cells and the power of this approach to study virus-host interactions. Future studies should further assess the applicability of this method for general detection of interactions that take place at the pathogen-plasma membrane interface. It will also be important to generate additional haploid cell lines, in order to broaden the range of pathogens and pathogen-derived molecules that can be studied using these genetic tools. In this regard, a number of haploid cell lines have been generated in mammals [137], unique tools to elucidate the basic aspects of human genetics.

**5.4. Population Genomics for Pathogen-Host Interaction Discovery.** Pathogens are among the most intriguing and prominent drivers of human evolution. Humans have adapted to the pressure imposed by microorganisms through genomic diversification, particularly through variation of genes involved in immune system function, constantly challenged by the rapidly coevolving pathogen genomes. The advent of new technologies such as next-generation sequencing and the computational tools associated have opened new avenues for the study of human genetics, making it



possible to evaluate the contribution of genetic diversity to susceptibility to infection at the genomic level. The emergence of datasets of genomic variation in multiple human populations, as well as pathogen genomes, allows detection of signatures of selection, which can be exploited to identify genes with major roles in immunity (for an excellent review see [138]). Remarkably, the cell surface-expressed receptors are among the most polymorphic gene families in mammals, subjected to strong positive selection and rapid evolution, in many instances possibly driven by pathogen molecules that remain unknown [87, 139–141]. Polymorphisms in receptors and immunomodulatory genes contribute to the natural susceptibility of different individuals to infection [142–144], as illustrated by protection against HIV infection in individuals carrying homozygous polymorphisms in the viral coreceptor CCR5 [145]. The identification and further study of genes under positive selection may represent a mainstream approach to dissect novel genes involved in disease and host-pathogen interaction. A notable example is the identification of glycophorin B as the erythrocyte receptor for *P. falciparum* protein EBL-1 through examination of highly polymorphic genes in populations from malaria-endemic regions [50]. Further population genetics studies promise key insights into novel immunological mechanisms and have the potential to provide molecular details that will ultimately help design effective therapies.

**5.5. Phagemic and Transposon Library-Based Screens.** In addition to these encouraging technologies, the generation of phagemic libraries has also represented an important tool for deciphering PPIs, in this case between particular binding partners and the whole genome of specific pathogens [146–148]. Typically, pathogen-encoded molecules are expressed as fusions with phage envelope proteins, a method known as phage display that has been widely exploited to identify peptides with specific binding properties. For example, Beckmann and colleagues built a phage display library to identify novel group B streptococci proteins capable of mediating adherence to fibronectin, a major component of the extracellular matrix often exploited for colonization of the host [51]. From this analysis, the authors identified 19 genes with homology to known bacterial adhesin proteins, genes involved in virulence, transport, or metabolic processes, along with genes with uncharacterized functions. Interestingly, one of these genes showed significant homology with the ScpB protein, a peptidase found in other streptococci that inactivates the member of the human complement system C5a, suggesting that this bacterial molecule acts as a bifunctional protein, similarly to other examples of multifunctional proteins discussed above [51].

More recently, a transposon-based insertion-inactivation mutant library was elegantly utilized to identify a bacterial protein capable of targeting the surface receptor TIGIT, an inhibitory molecule present in natural killer (NK) cells and T cells [52]. *Fusobacterium nucleatum* is a common oral bacterium that has been associated with colon adenocarcinoma and rheumatoid arthritis among other malignancies. In this study, Gur and colleagues showed that different strains of *F. nucleatum* blocked NK-mediated killing of human

tumors. Using a library of *F. nucleatum* mutants, the authors identified Fap2 as the bacterial protein that directly interacted with TIGIT, leading to inhibition of NK cytotoxicity and downregulation tumor-infiltrating T lymphocytes activation. Immuno-evasion is a hallmark of cancer; however whether members of the microbiome found within the tumor provide cancer cells with immunoregulatory properties has remained a major matter of debate [149]. These interesting findings suggest that *F. nucleatum* present in the tumor niche may enhance tumor escape by inactivating NK-mediated killing upon interaction of the fusobacterial Fap2 with the inhibitory receptor TIGIT. Of note, transposon-based mutant libraries are readily available for other pathogenic bacteria and have been successfully applied to identification of bacterial genes implicated in bacterial physiology [150–152]. It would be of interest to employ these libraries for unbiased identification of ePPIs. Notably, as mentioned above, we found that HAdV immunomodulators preferentially target other immunoreceptors that, similarly to TIGIT, also play inhibitory functions [28], suggesting this might represent a common immunosuppressive tactic evolved by pathogens. In fact, there is emerging evidence suggesting that may be the case, as a number of extracellular proteins from unrelated human pathogens have already been shown to target diverse immune receptors with inhibitory functions [28, 52, 82, 87, 153, 154]. Further exploration of inhibitory receptor targeting by other pathogens warrants exciting biological discoveries.

## 6. Concluding Remarks

Deciphering the human genome made possible the categorization of genes that encode for the human secretome; now, the challenge of the postgenomic era is to annotate the functions of those genes and their expression patterns during health and disease. A lot has been learnt from painstaking, highly focused experiments using classical biochemistry. In recent years, the impressive technological advances in proteomics, functional genomics, and computation have revolved our understanding of cell communication and function and have collectively created a versatile platform to enable biological discoveries, from mechanistic explorations to big data and systems biology analysis. Nevertheless our understanding of the molecules and mechanisms of extracellular immunomodulation and pathogen invasion remains remarkably limited.

Extracellular PPIs between host- and pathogen-encoded molecules orchestrate an enormous diversity of cellular processes, from initial colonization of the target cell to subsequent immune responses. The elucidation of these extracellular interactomes is integral to understanding the molecular basis of infection and will guide the development of more efficient or innovative therapeutics. Improvements in proteomics and genomics approaches have exponentially increased our understanding of how pathogens, particularly viruses, modulate the intracellular environment of the cell. Concomitantly, we and several other groups have implemented technologies directed towards elucidation of extracellular interactomes [2, 9, 10, 28, 32, 155], which have begun to reveal fundamental principles of extracellular

host-pathogen interactions. Notably, recent studies have revealed extensive ePPI networks in model organisms such as *Drosophila* or zebrafish [2, 9]. These undertakings predict that, similarly to intracellular PPIs, extracellular networks will be highly connected, with secreted and plasma membrane-expressed proteins having multiple binding partners. However, as discussed in this review, the identification of the host factors and in many cases the pathogen molecules that mediate ePPIs have largely defied molecular identification, in part due to the technical difficulties inherent to the study of these extracellular proteins. The elucidation of the global principles dictating extracellular pathogen-host PPIs will require a coordinated effort to bring together the areas of biology and technology. There are now considerable opportunities for integrating multiple disciplines for ePPI discovery, particularly proteomics and CRISPR/Cas9 genome-wide screens, which should be powered by commensurable advances in bioinformatics and computation for big data analysis. The integration of orthogonal datasets coming from multiple “omics” approaches will be advantageous for elucidating the intricacies of the host-pathogen extracellular interactomes and will further enhance the rational identification of novel therapeutic targets by uncovering fundamental principles of biology.

The journey from classical biochemical studies towards a systems biology approach has just begun and promises major technological breakthroughs and surprising biological findings. The development of powerful technologies for ePPI discovery has already illuminated sophisticated and sometimes unexpected molecular mechanisms by which pathogens interact with their hosts and has provided unique opportunities to increase our understanding of viral and bacterial pathogenesis. Further improvement of these technologies is warranted and will surely provide the scientific community with unprecedented insights into pathogen and host biology.

## Abbreviations

AVEXIS:	Avidity-based extracellular interaction screen
AAV:	Adeno-associated virus
CAR:	Coxsackievirus receptor
CMV:	Cytomegalovirus
CRISPR/Cas9:	Clustered regularly interspaced short palindromic repeat/CRISPR-associated protein 9
EBV:	Epstein-Barr virus
ECD:	Extracellular domain
ePPI:	Extracellular protein-protein interaction
HAdV:	Human adenoviruses
HCV:	Hepatitis C virus
HIV:	Human immunodeficiency virus
HSV:	Herpes simplex virus
ITIM:	Immunoreceptor tyrosine-based inhibitory motif
IVTT:	In vitro transcription/translation
LAIR1:	Leukocyte-associated immunoreceptor like 1

LILRB1:	Leukocyte immunoglobulin-like receptor 1
MPA:	Microfluidic-based comprehensive human membrane protein array
MPZL1:	Myelin protein zero-like protein 1
MS:	Mass spectrometry
NAPPA:	Nucleic-acid programmable protein array
NCPI:	Niemann-Pick C1
PDGFR $\alpha$ :	Platelet-derived growth factor receptor $\alpha$
PPI:	Protein-protein interaction
PVR:	Poliovirus receptor
SARS-CoV:	Severe acute respiratory syndrome coronavirus
SV40:	Simian virus 40
SLAM:	Signaling-lymphocytic activation molecule
SPR:	Surface plasmon resonance
T3SS:	Type III secretion system
VAVC:	Vaccinia virus.

## Competing Interests

The author declares that they have no competing interests.

## Acknowledgments

The author is grateful to Lino Gonzalez, Jennie Lill, Erik Verschuere, and Yvonne Franke for critical reading of the manuscript and insightful comments and suggestions.

## References

- [1] J. P. Overington, B. Al-Lazikani, and A. L. Hopkins, “How many drug targets are there?” *Nature Reviews Drug Discovery*, vol. 5, no. 12, pp. 993–996, 2006.
- [2] K. M. Bushell, C. Söllner, B. Schuster-Boeckler, A. Bateman, and G. J. Wright, “Large-scale screening for novel low-affinity extracellular protein interactions,” *Genome Research*, vol. 18, no. 4, pp. 622–630, 2008.
- [3] G. J. Wright, S. Martin, K. M. Bushell, and C. Söllner, “High-throughput identification of transient extracellular protein interactions,” *Biochemical Society Transactions*, vol. 38, no. 4, pp. 919–922, 2010.
- [4] G. J. Wright, “Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions,” *Molecular BioSystems*, vol. 5, no. 12, pp. 1405–1412, 2009.
- [5] N. M. Griffin and J. E. Schnitzer, “Overcoming key technological challenges in using mass spectrometry for mapping cell surfaces in tissues,” *Molecular and Cellular Proteomics*, vol. 10, no. 2, 2011.
- [6] J. Petschnigg, J. Snider, and I. Stagljar, “Interactive proteomics research technologies: recent applications and advances,” *Current Opinion in Biotechnology*, vol. 22, no. 1, pp. 50–58, 2011.
- [7] J. A. Maynard, N. C. Lindquist, J. N. Sutherland et al., “Surface plasmon resonance for high-throughput ligand screening of membrane-bound proteins,” *Biotechnology Journal*, vol. 4, no. 11, pp. 1542–1558, 2009.
- [8] M. H. Y. Lam and I. Stagljar, “Strategies for membrane interaction proteomics: no mass spectrometry required,” *Proteomics*, vol. 12, no. 10, pp. 1519–1526, 2012.
- [9] E. Özkan, R. A. Carrillo, C. L. Eastman et al., “An extracellular interactome of immunoglobulin and LRR proteins reveals



- receptor-ligand networks," *Cell*, vol. 154, no. 1, pp. X228–239, 2013.
- [10] S. R. Ramani, I. Tom, N. Lewin-Koh et al., "A secreted protein microarray platform for extracellular protein interaction discovery," *Analytical Biochemistry*, vol. 420, no. 2, pp. 127–138, 2012.
  - [11] A. Alcamí, "Viral mimicry of cytokines, chemokines and their receptors," *Nature Reviews Immunology*, vol. 3, no. 1, pp. 36–50, 2003.
  - [12] P. G. Fallon and A. Alcamí, "Pathogen-derived immunomodulatory molecules: future immunotherapeutics?" *Trends in Immunology*, vol. 27, no. 10, pp. 470–476, 2006.
  - [13] B. B. Finlay and G. McFadden, "Anti-immunology: evasion of the host immune system by bacterial and viral pathogens," *Cell*, vol. 124, no. 4, pp. 767–782, 2006.
  - [14] D. Bumann, "Heterogeneous host-pathogen encounters: act locally, think globally," *Cell Host and Microbe*, vol. 17, no. 1, pp. 13–19, 2015.
  - [15] P. Uetz, Y.-A. Dong, C. Zeretzke et al., "Herpesviral protein networks and their interaction with the human proteome," *Science*, vol. 311, no. 5758, pp. 239–242, 2006.
  - [16] A. Pichlmair, K. Kandasamy, G. Alvisi et al., "Viral immune modulators perturb the human molecular network by common and unique strategies," *Nature*, vol. 487, no. 7408, pp. 486–490, 2012.
  - [17] Z. H. Davis, E. Verschueren, G. M. Jang et al., "Global mapping of herpesvirus-host protein complexes reveals a transcription strategy for late genes," *Molecular Cell*, vol. 57, no. 2, pp. 349–360, 2015.
  - [18] B. De Chassey, V. Navratil, L. Tafforeau et al., "Hepatitis C virus infection protein network," *Molecular Systems Biology*, vol. 4, no. 1, article 230, 2008.
  - [19] O. Rozenblatt-Rosen, R. C. Deo, M. Padi et al., "Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins," *Nature*, vol. 487, no. 7408, pp. 491–495, 2012.
  - [20] S. Garamszegi, E. A. Franzosa, and Y. Xia, "Signatures of pleiotropy, economy and convergent evolution in a domain-resolved map of human-virus protein-protein interaction networks," *PLoS Pathogens*, vol. 9, no. 12, Article ID e1003778, 2013.
  - [21] A. Kabanova, J. Marcandalli, T. Zhou et al., "Platelet-derived growth factor- $\alpha$  receptor is the cellular receptor for human cytomegalovirus gHgLgO trimer," *Nature Microbiology*, vol. 1, no. 8, article no. 16082, 2016.
  - [22] A. Viejo-Borbolla, N. Martínez-Martín, H. J. Nel et al., "Enhancement of chemokine function as an immunomodulatory strategy employed by human herpesviruses," *PLoS Pathogens*, vol. 8, no. 2, Article ID e1002497, 2012.
  - [23] A. G. Dalgleish, P. C. L. Beverley, P. R. Clapham, D. H. Crawford, M. F. Greaves, and R. A. Weiss, "The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus," *Nature*, vol. 312, no. 5996, pp. 763–767, 1984.
  - [24] D. Klatzmman, E. Champagne, S. Chamaret et al., "T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV," *Nature*, vol. 312, no. 5996, pp. 767–768, 1984.
  - [25] R. J. Colonno, P. L. Callahan, and W. J. Long, "Isolation of a monoclonal antibody that blocks attachment of the major group of human rhinoviruses," *Journal of Virology*, vol. 57, no. 1, pp. 7–12, 1986.
  - [26] J. M. Greve, G. Davis, A. M. Meyer et al., "The major human rhinovirus receptor is ICAM-1," *Cell*, vol. 56, no. 5, pp. 839–847, 1989.
  - [27] C. Crosnier, L. Y. Bustamante, S. J. Bartholdson et al., "Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*," *Nature*, vol. 480, no. 7378, pp. 534–537, 2011.
  - [28] N. Martínez-Martín, S. R. Ramani, J. A. Hackney et al., "The extracellular interactome of the human adenovirus family reveals diverse strategies for immunomodulation," *Nature Communications*, vol. 7, Article ID 11473, 2016.
  - [29] W. R. Montor, J. Huang, Y. Hu et al., "Genome-wide study of *Pseudomonas aeruginosa* outer membrane protein immunogenicity using self-assembling protein microarrays," *Infection and Immunity*, vol. 77, no. 11, pp. 4877–4886, 2009.
  - [30] A. Ceroni, S. Sibani, A. Baiker et al., "Systematic analysis of the IgG antibody immune response against varicella zoster virus (VZV) using a self-assembled protein microarray," *Molecular BioSystems*, vol. 6, no. 9, pp. 1604–1610, 2010.
  - [31] I. Margarit, S. Bonacci, G. Pietrocola et al., "Capturing host-pathogen interactions by protein microarrays: identification of novel streptococcal proteins binding to human fibronectin, fibrinogen, and C4BP," *FASEB Journal*, vol. 23, no. 9, pp. 3100–3112, 2009.
  - [32] Y. Glick, Y. Ben-Ari, N. Drayman et al., "Pathogen receptor discovery with a microfluidic human membrane protein array," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 16, pp. 4344–4349, 2016.
  - [33] A. P. Frei, H. Moest, K. Novy, and B. Wollscheid, "Ligand-based receptor identification on living cells and tissues using TRICEPS," *Nature Protocols*, vol. 8, no. 7, pp. 1321–1336, 2013.
  - [34] M. D. Dyer, T. M. Murali, and B. W. Sobral, "Supervised learning and prediction of physical interactions between human and HIV proteins," *Infection, Genetics and Evolution*, vol. 11, no. 5, pp. 917–923, 2011.
  - [35] S.-L. Tan, G. Ganji, B. Paepers, S. Proll, and M. G. Katze, "Systems biology and the host response to viral infection," *Nature Biotechnology*, vol. 25, no. 12, pp. 1383–1389, 2007.
  - [36] M. D. Dyer, T. M. Murali, and B. W. Sobral, "The landscape of human proteins interacting with viruses and other pathogens," *PLoS Pathogens*, vol. 4, no. 2, article e32, 2008.
  - [37] P. Pileri, Y. Uematsu, S. Campagnoli et al., "Binding of hepatitis C virus to CD81," *Science*, vol. 282, no. 5390, pp. 938–941, 1998.
  - [38] M. J. Evans, T. Von Hahn, D. M. Tscherné et al., "Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry," *Nature*, vol. 446, no. 7137, pp. 801–805, 2007.
  - [39] A. Ploss, M. J. Evans, V. A. Gaysinskaya et al., "Human occludin is a hepatitis C virus entry factor required for infection of mouse cells," *Nature*, vol. 457, no. 7231, pp. 882–886, 2009.
  - [40] J. M. Bergelson, J. A. Cunningham, G. Droguett et al., "Isolation of a common receptor for coxsackie B viruses and adenoviruses 2 and 5," *Science*, vol. 275, no. 5304, pp. 1320–1323, 1997.
  - [41] P. P. Rose, S. L. Hanna, A. Spiridigliozzi et al., "Natural resistance-associated macrophage protein is a cellular receptor for Sindbis virus in both insect and mammalian hosts," *Cell Host and Microbe*, vol. 10, no. 2, pp. 97–104, 2011.
  - [42] R. C. Orchard, C. B. Wilen, J. G. Doench et al., "Discovery of a proteinaceous cellular receptor for a norovirus," *Science*, vol. 353, no. 6302, pp. 933–936, 2016.
  - [43] J. E. Carette, C. P. Guimaraes, I. Wuethrich et al., "Global gene disruption in human cells to assign genes to phenotypes by deep sequencing," *Nature Biotechnology*, vol. 29, no. 6, pp. 542–546, 2011.
  - [44] J. E. Carette, C. P. Guimaraes, M. Varadarajan et al., "Haploid genetic screens in human cells identify host factors used by pathogens," *Science*, vol. 326, no. 5957, pp. 1231–1235, 2009.

- [45] P. Papatheodorou, J. E. Carette, G. W. Bell et al., "Lipolysis-stimulated lipoprotein receptor (LSR) is the host receptor for the binary toxin *Clostridium difficile* transferase (CDT)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 39, pp. 16422–16427, 2011.
- [46] B. Schorch, S. Song, F. R. van Diemen et al., "LRP1 is a receptor for *Clostridium perfringens* TpeL toxin indicating a two-receptor model of clostridial glycosylating toxins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 17, pp. 6431–6436, 2014.
- [47] J. E. Carette, M. Raaben, A. C. Wong et al., "Ebola virus entry requires the cholesterol transporter Niemann-Pick C1," *Nature*, vol. 477, no. 7364, pp. 340–343, 2011.
- [48] L. T. Jae, M. Raaben, A. S. Herbert et al., "Lassa virus entry requires a trigger-induced receptor switch," *Science*, vol. 344, no. 6191, pp. 1506–1510, 2014.
- [49] S. Pillay, N. L. Meyer, A. S. Puschnik et al., "An essential receptor for adeno-associated virus infection," *Nature*, vol. 530, no. 7588, pp. 108–112, 2016.
- [50] D. C. G. Mayer, J. Cofie, L. Jiang et al., "Glycophorin B is the erythrocyte receptor of *Plasmodium falciparum* erythrocyte-binding ligand, EBL-1," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 13, pp. 5348–5352, 2009.
- [51] C. Beckmann, J. D. Waggoner, T. O. Harris, G. S. Tamura, and C. E. Rubens, "Identification of novel adhesins from group B streptococci by use of phage display reveals that C5a peptidase mediates fibronectin binding," *Infection and Immunity*, vol. 70, no. 6, pp. 2869–2876, 2002.
- [52] C. Gur, Y. Ibrahim, B. Isaacson et al., "Binding of the Fap2 protein of *Fusobacterium nucleatum* to human inhibitory receptor TIGIT protects tumors from immune cell attack," *Immunity*, vol. 42, no. 2, pp. 344–355, 2015.
- [53] L. C. Gonzalez, "Protein microarrays, biosensors, and cell-based methods for secretome-wide extracellular protein-protein interaction mapping," *Methods*, vol. 57, no. 4, pp. 448–458, 2012.
- [54] J. Snider, M. Kotlyar, P. Saraon, Z. Yao, I. Jurisica, and I. Stagljar, "Fundamentals of protein interaction network mapping," *Molecular Systems Biology*, vol. 11, no. 12, article no. 484, 2015.
- [55] T. Bock, D. Bausch-Fluck, A. Hofmann, and B. Wollscheid, "CD proteome and beyond - technologies for targeting the immune cell surfaceome," *Frontiers in Bioscience*, vol. 17, no. 5, pp. 1599–1612, 2012.
- [56] W. Li, M. J. Moore, N. Vasilieva et al., "Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus," *Nature*, vol. 426, no. 6965, pp. 450–454, 2003.
- [57] S. R. Radoshitzky, J. Abraham, C. F. Spiropoulou et al., "Transferrin receptor 1 is a cellular receptor for New World haemorrhagic fever arenaviruses," *Nature*, vol. 446, no. 7131, pp. 92–96, 2007.
- [58] V. S. Raj, H. Mou, S. L. Smits et al., "Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC," *Nature*, vol. 495, no. 7440, pp. 251–254, 2013.
- [59] B. J. Ryckman, M. C. Chase, and D. C. Johnson, "HCMV gH/gL/UL128-131 interferes with virus entry into epithelial cells: evidence for cell type-specific receptors," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 37, pp. 14118–14123, 2008.
- [60] D. Wang and T. Shenk, "Human cytomegalovirus virion protein complex required for epithelial and endothelial cell tropism," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 50, pp. 18153–18158, 2005.
- [61] B. J. Ryckman, M. A. Jarvis, D. D. Drummond, J. A. Nelson, and D. C. Johnson, "Human cytomegalovirus entry into epithelial and endothelial cells depends on genes UL128 to UL150 and occurs by endocytosis and low-pH fusion," *Journal of Virology*, vol. 80, no. 2, pp. 710–722, 2006.
- [62] M. I. Roche, Z. Lu, J. H. Hui, and J. Sharon, "Characterization of monoclonal antibodies to terminal and internal O-antigen epitopes of *Francisella tularensis* lipopolysaccharide," *Hybridoma*, vol. 30, no. 1, pp. 19–28, 2011.
- [63] F. Broecker, J. Aretz, Y. Yang et al., "Epitope recognition of antibodies against a yersinia pestis lipopolysaccharide trisaccharide component," *ACS Chemical Biology*, vol. 9, no. 4, pp. 867–873, 2014.
- [64] J. Häyrynen, S. Haseley, P. Talaga, M. Mühlenhoff, J. Finne, and J. F. G. Vliegthart, "High affinity binding of long-chain polysialic acid to antibody, and modulation by divalent cations and polyamines," *Molecular Immunology*, vol. 39, no. 7-8, pp. 399–411, 2002.
- [65] C. J. Day, E. N. Tran, E. A. Semchenko et al., "Glycan:glycan interactions: high affinity biomolecular interactions that can mediate binding of pathogenic bacteria to host cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 52, pp. E7266–E7275, 2015.
- [66] C. Ciferri, S. Chandramouli, D. Donnarumma et al., "Structural and biochemical studies of HCMV gH/gL/gO and Pentamer reveal mutually exclusive cell entry complexes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 6, pp. 1767–1772, 2015.
- [67] Y. Sun, M. Gallagher-Jones, C. Barker, and G. J. Wright, "A benchmarked protein microarray-based platform for the identification of novel low-affinity extracellular protein interactions," *Analytical Biochemistry*, vol. 424, no. 1, pp. 45–53, 2012.
- [68] H. F. Clark, A. L. Gurney, E. Abaya et al., "The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment," *Genome Research*, vol. 13, pp. 2265–2270, 2003.
- [69] I. Tom, N. Lewin-Koh, S. R. Ramani, and L. C. Gonzalez, "Protein microarrays for identification of novel extracellular protein-protein interactions," in *Current Protocols in Protein Science*, chapter 27, unit 27.3, 2013.
- [70] X. Yu, K. Harden, L. C. Gonzalez et al., "The surface protein TIGIT suppresses T cell activation by promoting the generation of mature immunoregulatory dendritic cells," *Nature Immunology*, vol. 10, no. 1, pp. 48–57, 2009.
- [71] A. Jaworski, I. Tom, R. K. Tong et al., "Operational redundancy in axon guidance through the multifunctional receptor Robo3 and its ligand NELL2," *Science*, vol. 350, no. 6263, pp. 961–965, 2015.
- [72] F. L. Yeh, Y. Wang, I. Tom, L. C. Gonzalez, and M. Sheng, "TREM2 binds to apolipoproteins, including APOE and CLU/APOJ, and thereby facilitates uptake of amyloid-beta by microglia," *Neuron*, vol. 91, no. 2, pp. 328–340, 2016.
- [73] H.-G. Burgert, Z. Ruzsics, S. Obermeier, A. Hilgendorf, M. Windheim, and A. Elsing, "Subversion of host defense mechanisms by adenoviruses," *Current Topics in Microbiology and Immunology*, vol. 269, pp. 273–318, 2002.
- [74] B. Ghebremedhin, "Human adenovirus: viral pathogen with increasing importance," *European Journal of Microbiology and Immunology*, vol. 4, no. 1, pp. 26–33, 2014.

- [75] C. M. Robinson, G. Singh, J. Y. Lee et al., "Molecular evolution of human adenoviruses," *Scientific Reports*, vol. 3, article no. 1812, 2013.
- [76] G. Singh, C. M. Robinson, S. Dehghan et al., "Homologous recombination in E3 genes of human adenovirus species D," *Journal of Virology*, vol. 87, no. 22, pp. 12481–12488, 2013.
- [77] H. Arase, E. S. Mocarski, A. E. Campbell, A. B. Hill, and L. L. Lanier, "Direct recognition of cytomegalovirus by activating and inhibitory NK cell receptors," *Science*, vol. 296, no. 5571, pp. 1323–1326, 2002.
- [78] T. L. Chapman, A. P. Heikema, A. P. West Jr., and P. J. Bjorkman, "Crystal structure and ligand binding properties of the D1D2 region of the inhibitory receptor LIR-1(ILT2)," *Immunity*, vol. 13, no. 5, pp. 727–736, 2000.
- [79] J. R. U. Cabrera, A. Viejo-Borbolla, N. Martinez-Martín, S. Blanco, F. Wandosell, and A. Alcamí, "Secreted herpes simplex virus-2 glycoprotein G modifies NGF-TrkA signaling to attract free nerve endings to the site of infection," *PLoS Pathogens*, vol. 11, no. 1, Article ID e1004571, 2015.
- [80] B. K. Kaiser, J. C. Pizarro, J. Kerns, and R. K. Strong, "Structural basis for NKG2A/CD94 recognition of HLA-E," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 18, pp. 6696–6701, 2008.
- [81] D. Brown, J. Trowsdale, and R. Allen, "The LILR family: modulators of innate and adaptive immune pathways in health and disease," *Tissue Antigens*, vol. 64, no. 3, pp. 215–225, 2004.
- [82] T. L. Chapman, A. P. Heikema, and P. J. Bjorkman, "The inhibitory receptor LIR-1 uses a common binding interaction to recognize class I MHC molecules and the viral homolog UL18," *Immunity*, vol. 11, no. 5, pp. 603–613, 1999.
- [83] T. H. C. Brondijk, T. de Ruiter, J. Ballering et al., "Crystal structure and collagen-binding site of immune inhibitory receptor LAIR-1: unexpected implications for collagen binding by platelet receptor GPVI," *Blood*, vol. 115, no. 7, pp. 1364–1373, 2010.
- [84] D. Jia, Y. Jing, Z. Zhang et al., "Amplification of MPZL1/PZR promotes tumor cell migration through Src-mediated phosphorylation of cortactin in hepatocellular carcinoma," *Cell Research*, vol. 24, no. 2, pp. 204–217, 2014.
- [85] K. Kuroki, A. Furukawa, and K. Maenaka, "Molecular recognition of paired receptors in the immune system," *Frontiers in Microbiology*, vol. 3, article 429, 2012.
- [86] L. C. Filip and N. I. Mundy, "Rapid evolution by positive Darwinian selection in the extracellular domain of the abundant lymphocyte protein CD45 in primates," *Molecular Biology and Evolution*, vol. 21, no. 8, pp. 1504–1511, 2004.
- [87] M. Akkaya and A. N. Barclay, "How do pathogens drive the evolution of paired receptors?" *European Journal of Immunology*, vol. 43, no. 2, pp. 303–313, 2013.
- [88] N. Ramachandran, J. V. Raphael, E. Hainsworth et al., "Next-generation high-density self-assembling functional protein arrays," *Nature Methods*, vol. 5, no. 6, pp. 535–538, 2008.
- [89] P. A. Beare, C. Chen, T. Bouman et al., "Candidate antigens for Q fever serodiagnosis revealed by immunoscreening of a *Coxiella burnetii* protein microarray," *Clinical and Vaccine Immunology*, vol. 15, no. 12, pp. 1771–1779, 2008.
- [90] J. E. Lopez, P. A. Beare, R. A. Heinzen et al., "High-throughput identification of T-lymphocyte antigens from *Anaplasma marginale* expressed using in vitro transcription and translation," *Journal of Immunological Methods*, vol. 332, no. 1–2, pp. 129–141, 2008.
- [91] S. L. Keasey, K. E. Schmid, M. S. Lee et al., "Extensive antibody cross-reactivity among infectious gram-negative bacteria revealed by proteome microarray analysis," *Molecular and Cellular Proteomics*, vol. 8, no. 5, pp. 924–935, 2009.
- [92] Y. Xu, J. F. Bruno, and B. J. Luft, "Profiling the humoral immune response to *Borrelia burgdorferi* infection with protein microarrays," *Microbial Pathogenesis*, vol. 45, no. 5–6, pp. 403–407, 2008.
- [93] F. X. Sutandy, J. Qian, C. S. Chen, and H. Zhu, "Overview of protein microarrays," in *Current Protocols in Protein Science*, chapter 27, unit 27.1, 2013.
- [94] X. Yu, K. B. Decker, K. Barker et al., "Host-pathogen interaction profiling using self-assembling human protein arrays," *Journal of Proteome Research*, vol. 14, no. 4, pp. 1920–1936, 2015.
- [95] R. Aebersold and M. Mann, "Mass-spectrometric exploration of proteome structure and function," *Nature*, vol. 537, no. 7620, pp. 347–355, 2016.
- [96] C. E. Bakalarski and D. S. Kirkpatrick, "A biologist's field guide to multiplexed quantitative proteomics," *Molecular and Cellular Proteomics*, vol. 15, no. 5, pp. 1489–1497, 2016.
- [97] A. H. Smits and M. Vermeulen, "Characterizing protein-protein interactions using mass spectrometry: challenges and opportunities," *Trends in Biotechnology*, vol. 34, no. 10, pp. 825–834, 2016.
- [98] M. Rebsamen, R. K. Kandasamy, and G. Superti-Furga, "Protein interaction networks in innate immunity," *Trends in Immunology*, vol. 34, no. 12, pp. 610–619, 2013.
- [99] B. E. Gewurz, H. L. Ploegh, and D. Tortorella, "US2, a human cytomegalovirus-encoded type I membrane protein, contains a non-cleavable amino-terminal signal peptide," *The Journal of Biological Chemistry*, vol. 277, no. 13, pp. 11306–11313, 2002.
- [100] J. D. Bendtsen, L. J. Jensen, N. Blom, G. Von Heijne, and S. Brunak, "Feature-based prediction of non-classical and leaderless protein secretion," *Protein Engineering, Design and Selection*, vol. 17, no. 4, pp. 349–356, 2004.
- [101] K. T. Sears, S. M. Ceraul, J. J. Gillespie et al., "Surface proteome analysis and characterization of surface cell antigen (Sca) or autotransporter family of *Rickettsia typhi*," *PLoS Pathogens*, vol. 8, no. 8, Article ID e1002856, 2012.
- [102] T. Myers-Morales, C. Cowan, M. E. Gray et al., "A surface-focused biotinylation procedure identifies the *Yersinia pestis* catalase KatY as a membrane-associated but non-surface-located protein," *Applied and Environmental Microbiology*, vol. 73, no. 18, pp. 5750–5759, 2007.
- [103] Y. Ge and Y. Rikihisa, "Identification of novel surface proteins of *Anaplasma phagocytophilum* by affinity purification and proteomics," *Journal of Bacteriology*, vol. 189, no. 21, pp. 7819–7828, 2007.
- [104] Y. Ge and Y. Rikihisa, "Surface-exposed proteins of *Ehrlichia chaffeensis*," *Infection and Immunity*, vol. 75, no. 8, pp. 3833–3841, 2007.
- [105] S. M. Noh, K. A. Brayton, W. C. Brown et al., "Composition of the surface proteome of *Anaplasma marginale* and its role in protective immunity induced by outer membrane immunization," *Infection and Immunity*, vol. 76, no. 5, pp. 2219–2226, 2008.
- [106] L. M. Kattenhorn, R. Mills, M. Wagner et al., "Identification of proteins associated with murine cytomegalovirus virions," *Journal of Virology*, vol. 78, no. 20, pp. 11187–11197, 2004.
- [107] E. Johannsen, M. Luftig, M. R. Chase et al., "Proteins of purified Epstein-Barr virus," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 46, pp. 16286–16291, 2004.



- [108] S. M. Varnum, D. N. Streblow, M. E. Monroe et al., "Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome," *Journal of Virology*, vol. 78, no. 20, pp. 10960–10966, 2004.
- [109] J.-M. Revest, L. DeMoerloozee, and C. Dickson, "Fibroblast growth factor 9 secretion is mediated by a non-cleaved amino-terminal signal sequence," *Journal of Biological Chemistry*, vol. 275, no. 11, pp. 8083–8090, 2000.
- [110] K. Miyakawa and T. Imamura, "Secretion of FGF-16 requires an uncleaved bipartite signal sequence," *The Journal of Biological Chemistry*, vol. 278, no. 37, pp. 35718–35724, 2003.
- [111] B. Henderson and A. Martin, "Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease," *Infection and Immunity*, vol. 79, no. 9, pp. 3476–3491, 2011.
- [112] T. B. M. Hickey, L. M. Thorson, D. P. Speert, M. Daffé, and R. W. Stokes, "Mycobacterium tuberculosis Cpn60.2 and DnaK are located on the bacterial surface, where Cpn60.2 facilitates efficient bacterial association with macrophages," *Infection and Immunity*, vol. 77, no. 8, pp. 3389–3401, 2009.
- [113] T. B. M. Hickey, H. J. Ziltener, D. P. Speert, and R. W. Stokes, "Mycobacterium tuberculosis employs Cpn60.2 as an adhesin that binds CD43 on the macrophage surface," *Cellular Microbiology*, vol. 12, no. 11, pp. 1634–1647, 2010.
- [114] Y. Wang, C. G. Kelly, J. T. Karttunen et al., "Cd40 is a cellular receptor mediating mycobacterial heat shock protein 70 stimulation of CC-chemokines," *Immunity*, vol. 15, no. 6, pp. 971–983, 2001.
- [115] R. A. Floto, P. A. MacAry, J. M. Boname et al., "Dendritic cell stimulation by mycobacterial Hsp70 is mediated through CCR5," *Science*, vol. 314, no. 5798, pp. 454–458, 2006.
- [116] K. Babaahmady, W. Oehlmann, M. Singh, and T. Lehner, "Inhibition of human immunodeficiency virus type 1 infection of human CD4<sup>+</sup> T cells by microbial HSP70 and the peptide epitope 407–426," *Journal of Virology*, vol. 81, no. 7, pp. 3354–3360, 2007.
- [117] M. A. Calderwood, K. Venkatesan, L. Xing et al., "Epstein-Barr virus and virus human protein interaction maps," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 18, pp. 7606–7611, 2007.
- [118] E. A. Franzosa and Y. Xia, "Structural principles within the human-virus protein-protein interaction network," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 26, pp. 10538–10543, 2011.
- [119] H. Tatsuo, N. Ono, K. Tanaka, and Y. Yanagi, "Slam (CDw150) is a cellular receptor for measles virus," *Nature*, vol. 406, no. 6798, pp. 893–897, 2000.
- [120] C. L. Mendelsohn, E. Wimmer, and V. R. Racaniello, "Cellular receptor for poliovirus: molecular cloning, nucleotide sequence, and expression of a new member of the immunoglobulin superfamily," *Cell*, vol. 56, no. 5, pp. 855–865, 1989.
- [121] S. Cherry, "What have RNAi screens taught us about viral-host interactions?" *Current Opinion in Microbiology*, vol. 12, no. 4, pp. 446–452, 2009.
- [122] A. L. Jackson, S. R. Bartz, J. Schelter et al., "Expression profiling reveals off-target gene regulation by RNAi," *Nature Biotechnology*, vol. 21, no. 6, pp. 635–637, 2003.
- [123] J. E. Garneau, M.-È. Dupuis, M. Villion et al., "The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA," *Nature*, vol. 468, no. 7320, pp. 67–71, 2010.
- [124] O. Shalem, N. E. Sanjana, and F. Zhang, "High-throughput functional genomics using CRISPR-Cas9," *Nature Reviews Genetics*, vol. 16, no. 5, pp. 299–311, 2015.
- [125] P. Hou, S. Chen, S. Wang et al., "Genome editing of CXCR4 by CRISPR/cas9 confers cells resistant to HIV-1 infection," *Scientific Reports*, vol. 5, Article ID 15577, 2015.
- [126] J. F. Hultquist, K. Schumann, J. M. Woo et al., "A Cas9 ribonucleoprotein platform for functional genetic studies of HIV-host interactions in primary human t cells," *Cell Reports*, vol. 17, no. 5, pp. 1438–1452, 2016.
- [127] F. R. van Diemen, E. M. Kruse, M. J. G. Hooykaas et al., "CRISPR/Cas9-mediated genome editing of herpesviruses limits productive and latent infections," *PLoS Pathogens*, vol. 12, no. 6, Article ID e1005701, 2016.
- [128] C. Blondel, J. Park, T. Hubbard et al., "CRISPR/Cas9 screens reveal requirements for host cell sulfation and fucosylation in bacterial Type III secretion system-mediated cytotoxicity," *Cell Host & Microbe*, vol. 20, no. 2, pp. 226–237, 2016.
- [129] P. D. Nagy, J. Pogany, and J.-Y. Lin, "How yeast can be used as a genetic platform to explore virus-host interactions: from 'omics' to functional studies," *Trends in Microbiology*, vol. 22, no. 6, pp. 309–316, 2014.
- [130] M. Kotecki, P. S. Reddy, and B. H. Cochran, "Isolation and characterization of a near-haploid human cell line," *Experimental Cell Research*, vol. 252, no. 2, pp. 273–280, 1999.
- [131] S. Moller-Tank and W. Maury, "Ebola virus entry: a curious and complex series of events," *PLoS Pathogens*, vol. 11, no. 4, Article ID e1004731, 2015.
- [132] M. Côté, J. Misasi, T. Ren et al., "Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection," *Nature*, vol. 477, no. 7364, pp. 344–348, 2011.
- [133] E. H. Miller, G. Obernosterer, M. Raaben et al., "Ebola virus entry requires the host-programmed recognition of an intracellular receptor," *EMBO Journal*, vol. 31, no. 8, pp. 1847–2059, 2012.
- [134] H. Wang, Y. Shi, J. Song et al., "Ebola viral glycoprotein bound to its endosomal receptor niemann-pick C1," *Cell*, vol. 164, no. 1-2, pp. 258–268, 2016.
- [135] M. Ng, E. Ndungo, R. K. Jangra et al., "Cell entry by a novel European filovirus requires host endosomal cysteine proteases and Niemann-Pick C1," *Virology*, vol. 468–470, pp. 637–646, 2014.
- [136] H. Cohen-Dvashi, N. Cohen, H. Israeli, and R. Diskin, "Molecular mechanism for LAMP1 recognition by Lassa virus," *Journal of Virology*, vol. 89, no. 15, pp. 7584–7592, 2015.
- [137] I. Sagi, G. Chia, T. Golan-Lev et al., "Derivation and differentiation of haploid human embryonic stem cells," *Nature*, vol. 532, no. 7597, pp. 107–111, 2016.
- [138] L. Quintana-Murci and A. G. Clark, "Population genetic tools for dissecting innate immunity in humans," *Nature Reviews Immunology*, vol. 13, no. 4, pp. 280–293, 2013.
- [139] J. C. Sun and L. L. Lanier, "The natural selection of herpesviruses and virus-specific NK cell receptors," *Viruses*, vol. 1, no. 3, pp. 362–382, 2009.
- [140] R. Cagliani and M. Sironi, "Pathogen-driven selection in the human genome," *International Journal of Evolutionary Biology*, vol. 2013, Article ID 204240, 6 pages, 2013.
- [141] E. J. Vallender and B. T. Lahn, "Positive selection on the human genome," *Human Molecular Genetics*, vol. 13, no. 2, pp. R245–R254, 2004.
- [142] K. G. Andersen, I. Shylakhter, S. Tabrizi, S. R. Grossman, C. T. Happi, and P. C. Sabeti, "Genome-wide scans provide evidence

- for positive selection of genes implicated in Lassa fever," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1590, pp. 868–877, 2012.
- [143] S. Rahman, M. Shering, N. H. Ogden, R. Lindsay, and A. Badawi, "Toll-like receptor cascade and gene polymorphism in host-pathogen interaction in Lyme disease," *Journal of Inflammation Research*, vol. 9, pp. 91–102, 2016.
- [144] G. T. J. van Well, M. S. Sanders, S. Ouburg, V. Kumar, A. M. van Furth, and S. A. Morré, "Single nucleotide polymorphisms in pathogen recognition receptor genes are associated with susceptibility to meningococcal meningitis in a pediatric cohort," *PLoS ONE*, vol. 8, no. 5, Article ID e64252, 2013.
- [145] R. Liu, W. A. Paxton, S. Choe et al., "Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection," *Cell*, vol. 86, no. 3, pp. 367–377, 1996.
- [146] K. Jacobsson and L. Frykberg, "Cloning of ligand-binding domains of bacterial receptors by phage display," *BioTechniques*, vol. 18, no. 5, pp. 878–884, 1995.
- [147] J. Vasi, L. Frykberg, L. E. Carlsson, M. Lindberg, and B. Guss, "M-like proteins of *Streptococcus dysgalactiae*," *Infection and Immunity*, vol. 68, no. 1, pp. 294–302, 2000.
- [148] M. Nilsson, L. Frykberg, J.-I. Flock, L. Pei, M. Lindberg, and B. Guss, "A fibrinogen-binding protein of *Staphylococcus epidermidis*," *Infection and Immunity*, vol. 66, no. 6, pp. 2666–2673, 1998.
- [149] W. S. Garrett, "Cancer and the microbiota," *Science*, vol. 348, no. 6230, pp. 80–86, 2015.
- [150] A. Fajardo, N. Martínez-Martín, M. Mercadillo et al., "The neglected intrinsic resistome of bacterial pathogens," *PLoS ONE*, vol. 3, no. 2, Article ID e1619, 2008.
- [151] L. A. Gallagher, E. Ramage, R. Patrapuvich, E. Weiss, M. Brittnacher, and C. Manoil, "Sequence-defined transposon mutant library of *Burkholderia thailandensis*," *mBio*, vol. 4, no. 6, Article ID e00604-13, 2013.
- [152] M. A. Jacobs, A. Alwood, I. Thaipisuttikul et al., "Comprehensive transposon mutant library of *Pseudomonas aeruginosa*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 2, pp. 14339–14344, 2003.
- [153] K. R. Chan, E. Z. Ong, H. C. Tan et al., "Leukocyte immunoglobulin-like receptor B1 is critical for antibody-dependent dengue," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 7, pp. 2722–2727, 2014.
- [154] T. Satoh and H. Arase, "HSV-1 infection through inhibitory receptor, PILRalpha," *Virus*, vol. 58, no. 1, pp. 27–36, 2008.
- [155] J. Petschnigg, B. Groisman, M. Kotlyar et al., "The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells," *Nature Methods*, vol. 11, no. 5, pp. 585–592, 2014.



