Quantifying domain-ligand affinities and specificities by high-throughput holdup assay

Renaud Vincentelli1,7, Katja Luck2,6,7, Juline Poirson2, Jolanta Polanowska3,6, Julie Abdat1, Marilyne Blémont1, Jeremy Turchetto1, François Iv1, Kevin Ricquier4, Marie-Laure Straub2, Anne Forster2, Patricia Cassonnet5, Jean-Paul Borg3, Yves Jacob5, Murielle Masson2, Yves Nominé2, Jérôme Reboul3,6, Nicolas Wolff4, Sebastian Charbonnier2 & Gilles Travé2

Many protein interactions are mediated by small linear motifs interacting specifically with defined families of globular domains. Quantifying the specificity of a motif requires measuring and comparing its binding affinities to all its putative target domains. To this end, we developed the high-throughput holdup assay, a chromatographic approach that can measure up to 1,000 domain-motif equilibrium binding affinities per day. After benchmarking the approach on 210 PDZ-peptide pairs with known affinities, we determined the affinities of two viral PDZ-binding motifs derived from human papillomavirus E6 oncoproteins for 209 PDZ domains covering 79% of the human PDZome. We obtained sharply sequence-dependent binding profiles that quantitatively describe the PDZome recognition specificity of each motif. This approach, applicable to many categories of domain-ligand interactions, has wide potential for quantifying the specificities of interactomes.

Many protein–protein interactions are mediated by globular domains1 specifically binding to short linear peptide motifs2, in complex intracellular competitions that depend on individual binding affinities. High-throughput interactomics studies have allowed genome-wide description of protein–protein interaction networks3, including those mediated by domain-motif interactions4. However, these methods have mainly delivered binary data (“interact” or “not interact”) without quantitative information on affinities and specificities. The development of approaches for systematic determination and affinity ranking of domain-motif interacting pairs thus represents a key challenge in systems biology5.

PDZ domains recognize specific linear motifs (dubbed PDZ-binding motifs or PBMs) at the extreme C termini of partner proteins6. The human proteome contains 266 identified PDZ domains (the PDZome) dispersed over 152 proteins, and a few thousand putative PBMs7. PDZ-PBM interactions participate in cell polarity processes critical for cellular differentiation, proliferation and migration across a wide range of epidermal, retinal, endothelial, neuronal or immunological cellular types8,9. PDZ-mediated interactions are recurrently perturbed in carcinogenesis. Several tumor virus proteins, such as human papillomavirus (HPV) E6 oncoprotein, contain PBMs that target cancer-relevant PDZ-containing cellular proteins10.

Here we developed and benchmarked a fast, accurate and versatile chromatographic approach for determination of domain-ligand affinities at high throughput. We used this approach to generate PDZome-binding profiles quantitatively describing the binding specificity of two HPV E6 PBMs toward almost the full complement of human PDZ domains.

RESULTS

Principle of the automated holdup assay

To address PDZ-motif interaction specificities on a large scale, we automated on a liquid handling robot the holdup approach, a chromatographic assay that evaluates domain-ligand interaction affinities at equilibrium11 (Fig. 1a). Soluble bacterial overexpression extracts of domains fused to a solubility-enhancing tag (here, maltose-binding protein) are incubated with aliquots of ligand saturated with ligand molecules (here, biotinylated peptides). At this step, affinity-dependent proportions of domain-ligand complexes are formed, favored by the high ligand concentration (around 50–100 µM) largely exceeding the domain concentration (4 µM). Next the resin-extract mixtures are filtered. The resin, trapped on the filter, retains (‘holds up’) the domain-ligand complexes, whereas the liquid filtrate contains the free domain molecules. Filtrates of domain extracts incubated with

1Unité Mixte de Recherche (UMR) 7257, Centre National de la Recherche Scientifique (CNRS)–Aix-Marseille Université, Architecture et Fonction des Macromolécules Biologiques (AFMB), Marseille, France. 2Équipe labellisée Ligue 2015, UMR 7242, CNRS–Université de Strasbourg, École Supérieure de Médecine, and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA (K.L.); Unité 1104 INSERM – UMR 7280, CNRS–Aix-Marseille Université, Centre d’Immunologie de Marseille-Luminy (CIML), Marseille, France (J. Polanowska and J.R.). 3These authors contributed equally to this work. Correspondence should be addressed to G.T. (trave@unistra.fr).

RECEIVED 18 NOVEMBER 2014; ACCEPTED 19 APRIL 2015; PUBLISHED ONLINE 8 JUNE 2015; DOI:10.1038/NMETH.3438
Figure 1 Principle, readout and data treatment of the high-throughput holdup assay. (a) Soluble extracts of overexpressed domains are incubated with resin saturated with ligands (here, biotinylated peptides) and then filtered. Flow-throughs are analyzed by microfluidic capillary electrophoresis. A decrease in domain signal in the flow-through reveals domain-peptide interactions and allows for determination of binding intensities (BI) using the depicted formula. (b) Top, the readout consists in electropherograms (here displayed as virtual gels) of extracts of the indicated MBP-PDZ constructs filtered through resin coated with negative control (biotin) or peptide (for example, 16E6LV). Bottom, overlaid electropherograms of peptide (red) and biotin (blue) flow-throughs reveal both the occurrence and the strength of binding events. Instances of strong binding (MAGI1-2/6 vs. 16E6LV), filtered through resin saturated with ligands (here, biotinylated peptides) and then filtered. Flow-throughs are analyzed by microfluidic capillary electrophoresis. Holdup and SPR data were in good agreement (correlation coefficient $r = 0.76$; Fig. 2b) and Supplementary Table 1). The best correlation ($r = 0.81$) was obtained for affinities in the 15–100 µM range (Fig. 2b). Furthermore, the holdup assay detected and discriminated very weak affinities (in the submillimolar range) for many PDZ-peptide pairs that had not generated any signal in the SPR study (SCRIB-4/4; Fig. 2a). We estimate that avidin resin–based holdup can detect relevant binding intensities for $K_D$ values up to 300 µM (Online Methods). The weakest affinity detectable in a holdup assay is dictated by the resin capacity, which determines the concentration of resin-attached ligand in the binding reaction.

Validation of automated holdup assay
We previously measured by surface plasmon resonance (SPR; Biacore) 210 interactions involving 42 biotinylated peptides and five MBP-fused PDZ constructs (MAGI1-2/6, MAGI1-3/6, SCRIB-3/4, SCRIB-4/4 and SCRIB-34/4), revealing a variety of affinities ranging from micromolar to submillimolar. We used these 210 interactions as a benchmark for the automated holdup assay, performed using three 96-well plates at a rate of 90 min per plate for the interaction assay, plus 66 min per plate for the analysis by capillary electrophoresis. Holdup and SPR data were in good agreement (correlation coefficient $r = 0.76$; Fig. 2a, b and Supplementary Table 1). The best correlation ($r = 0.81$) was obtained for affinities in the 15–100 µM range (Fig. 2b). Furthermore, the holdup assay detected and discriminated very weak affinities (in the submillimolar range) for many PDZ-peptide pairs that had not generated any signal in the SPR study (SCRIB-4/4; Fig. 2a). We estimate that avidin resin–based holdup can detect relevant binding intensities for $K_D$ values up to 300 µM (Online Methods). The weakest affinity detectable in a holdup assay is dictated by the resin capacity, which determines the concentration of resin-attached ligand in the binding reaction.

For stronger PDZ-peptide complexes in the 1–10 µM affinity range, our standard holdup protocol delivered saturated maximal binding signals and therefore failed to rank affinities in that range in contrast to SPR (Fig. 2b). Additional holdup runs using serial dilutions of both peptide-saturated resin and biotin-saturated resin provide negative controls. All filtrates are then systematically subjected to quantitative microfluidic capillary electrophoresis (Fig. 1b). Whenever a domain binds a ligand, the domain peak intensity in the filtrate of the ligand-saturated resin $I_{\text{lig}}$ decreases as compared to the domain peak intensity in the control filtrate $I_{\text{ctrl}}$ (Fig. 1b). To eliminate scale differences due to discrepancies in sample loading input, we normalize domain peak intensities in the compared filtrates against an internal standard peak (here, lysozyme). Values of the resulting ‘uncorrected’ and ‘corrected’ BIs are indicated.

Because the protocol uses a fixed amount of resin fully saturated with ligand peptide, the concentration of ligand is expected to be the same in all measurements. The fraction of domain bound to ligand-saturated resin can therefore be used to define a binding intensity (BI), which allows us to quantify and rank the affinities of all domain-ligand pairs measured. Equilibrium dissociation constants ($K_D$) can also be estimated from the measurements (see Online Methods).
domain extracts eventually discriminated the individual affinities of the strong complexes (Fig. 2c).

We also tested holdup measurements using purified domain constructs instead of crude soluble extracts. The results of both approaches were strongly correlated ($r = 0.92$; Supplementary Fig. 1). Binding intensities were in general slightly stronger for unpurified domains, indicating that the extracts may protect the constructs from unfolding or aggregation and/or stabilize the complexes. Therefore, for constructs such as MBP-PDZ fusions that are mostly well folded and soluble, purification can be skipped.

Probing the binding of viral peptides to the human PDZome

We cloned and expressed as MBP-fusions the 266 known PDZ proteins corresponding to the extreme C termini of the HPV16 and HPV18 E6 oncoproteins (sequences SRTRRRETQV and RLQRRRETQV, respectively). The assay was performed in two alternative formats using either 384-well plates (Fig. 3a) or 96-well plates (not shown). Both within and between these setups, results were highly reproducible and comparable (Fig. 3b,c) with s.d. of BI below 0.1 in 90% of all cases. By running a triplicate 384-well holdup screen, we generated—in a single shot—high-quality interaction data for 209 PDZ constructs (79% of the PDZome) versus two PBMs and two negative controls (Supplementary Figs. 3 and 4), representing more than 1,000 domain-peptide interactions measured over 1 d.

Choice of control and threshold for defining binders

We tested two alternative controls in the PDZome interaction screen (Fig. 3a): biotin and an irrelevant ‘none’ peptide presenting a neutral hydrophilic sequence (GSNSGNGNS) devoid of any PBM. The results obtained using either control were highly correlated ($r = 0.97$; Fig. 3d), and neither control displayed relevant binding to any PDZ domain (Fig. 4a and Supplementary Fig. 3). Remarkably, 100% and 98% of the BI values obtained in the negative-control experiments were lower than 0.2 and 0.1, respectively (Supplementary Table 3), and 100% and 95% of BI values measured in the 384-well format had s.d. values lower than 0.2 and 0.1, respectively.
respectively (Fig. 3b). Therefore, we propose two alternative BI threshold values ($B_{\text{min}}$) for defining binders when quantitative holdup data must be converted to binary interaction data. $B_{\text{min}} = 0.2$ represents a very stringent threshold to retain only high-confidence binding pairs, whereas $B_{\text{min}} = 0.1$ represents a more relaxed threshold to retain a number of additional pairs that display weak but relevant binding signal.

Quantifying the PDZome recognition specificities of PBMs

For each peptide, the results of the holdup screen were plotted by ranking decreasing PDZ binding intensities from left to right (Fig. 4, Supplementary Fig. 4 and Supplementary Table 3). This plotting mode produces PDZome binding profiles, which depict the binding specificity of each motif at a glance. The steeper the slope of the profile, the more specific the motif. The HPV18 E6 PBM appears slightly more specific than the HPV16 E6 PBM (Fig. 4a).

We estimated $K_D$ values for all the PDZ-peptide pairs investigated (Online Methods and Supplementary Table 3). We observed a continuum of affinities, spanning from micromolar to sub-millimolar. Roughly, each E6 PBM bound 1%, 4% or 20% of the PDZome with a $K_D$ below 5 µM, 25 µM or 250 µM, respectively. The continuity of the profiles indicates that human PDZ domains are sufficiently diverse to sample a whole range of affinities for any PBM, including viral ones. These data confirmed predictions from Stiffler et al., who suggested that PDZ domains evenly sample a continuous selectivity space rather than clustering into discrete classes.

HPV16 E6 and HPV18 E6 shared 70% of common binders (using $B_{\text{min}} = 0.1$) (Fig. 4b and Supplementary Fig. 5), but their respective affinities for these binders could vary considerably, leading to a substantial reshuffling of the binding profiles (Fig. 4a,b). Given the high similarity of the two E6 PBMs, this shows that the profiles represent a signature sharply dependent on the motif sequence. This signature
and its variations are likely to be functionally relevant. For instance, a recent study indicated that HPV16 E6 and HPV18 E6 preferentially degrade SCRIB or DLG proteins, respectively, whereas both of them efficiently degrade MAGI1. Accordingly, the profiles show that both HPV16 and HPV18 E6 preferentially bind to MAGI1 PDZ2 domains, closely followed by SCRIB PDZ3 and PDZ1 domains for HPV16 E6, and followed instead by PDZ2 and PDZ3 domains of DLG proteins for HPV18 E6 (Fig. 4b).

Holdup-based heat maps of E6-binding PDZ domains in the context of full-length proteins are also informative (Fig. 5). They show that some proteins, such as SCRIB, DLGs or MAGI1, can contain several E6-binding domains. This, combined with the ability of E6 to dimerize, may create avidity effects prone to potentiate the capture of some host multi-PDZ proteins by E6.

Identifying sequence determinants of PDZ-E6 recognition
An alignment of 209 PDZ domains ranked according to their highest E6 binding intensity (Supplementary Fig. 6) revealed, at defined PDZ domain positions, residues displaying a distribution bias toward the top or the bottom of the alignment, which we interpreted as residues preferred or disliked for E6 binding, respectively (Online Methods, Supplementary Figs. 6 and 7 and Supplementary Tables 4 and 5). We found, by highlighting them on the structure of E6-bound MAGI1 PDZ2 (Fig. 6a), that the PDZ domain positions carrying the residues best correlated to E6 binding were proximal to E6 PBM. Two of the best-correlated residues, His119 and Val123, are well-known determinants for recognition of class I PBMs displaying a threonine at position −2, with which they establish atomic contacts (Fig. 6a); and Glu63 contributes to E6 binding by contacting a stretch of arginines present in both E6 PBMs.

We also exploited the profiles to identify critical PDZ positions dictating a preference for HPV18 E6 over HPV16 E6 or vice versa. 60 E6-binding PDZ domains (BI ≥ 0.1 for HPV16 and/or HPV18) were aligned (Supplementary Fig. 8) and ranked according to the differences ABI observed between their intensities of binding to HPV18 and HPV16 E6 (Supplementary Table 6). We observed substantial ABI values spanning from −0.54 to 0.4. We identified 13 PDZ domain positions in the alignment whose physico-chemical properties of the respective residues (evaluated by four indices developed by Achley et al.) significantly correlated with ABI values (Online Methods, Supplementary Fig. 9 and Supplementary Table 7). The most highly correlated domain positions were proximal to peptide position p0 in PDZ-PBM complexes (Fig. 6b). Remarkably, the most notable difference between HPV16 and HPV18 E6 PBMs (see sequences in Fig. 4a) is at peptide position p0 (L vs. V). These results demonstrate the power of quantitative holdup data for identifying sequence determinants of domain-motif interaction specificities.

Comparison to previous studies and orthogonal validation
The holdup screen identified 50 binders for HPV16 E6 and 52 for HPV18 E6 (using BImin = 0.1). Comparing our screen to previous low-throughput studies, we confirmed 11 among 12 and 8 among 10 of the published binders (Fig. 4) of HPV16 E6 and HPV18 E6, respectively (see also Supplementary Fig. 5 and Supplementary Table 8). We identified about 40 new E6-binding PDZ domains, including strong binders such as MAST2, DLG2_2, DLG2_3, DLG3_2, DLG3_3, PDZK1_1, PDZR3N3_1 and SYNJ2BP (Fig. 4).

Recently, Belotti et al. identified 36 HPV16 E6–binding PDZs from a high-throughput yeast two-hybrid (Y2H) screen of the

**Figure 5** | Domain maps of human proteins containing E6-binding PDZ domains. PDZ domains are colored in heat-map mode according to E6 binding intensities. Proteins are ranked according to the binding intensities of their best E6-binding PDZ domain.
human PDZome. The holdup assay confirmed most of the binders detected by Y2H: 26/36 (72%) using $B_{\text{min}} = 0.2$, and up to 29/36 (81%) using $B_{\text{min}} = 0.1$ (Fig. 4b and Supplementary Table 9). Using $B_{\text{min}} = 0.2$, the holdup screen returned 37 HPV16 E6 binders. Using $B_{\text{min}} = 0.1$ instead, the holdup identified 13 additional hits, of which only 3 were detected by Y2H (Supplementary Table 9). These additional holdup hits displayed a weak but significant E6-binding signal, with a low s.d.

We validated the holdup screen in mammalian cells using the split *Gaussia princeps* luciferase protein complementation assay (GPCA)\(^{18}\). For PDZ domains identified as binders of HPV16 or HPV18 E6 PBMs by the holdup assay (threshold $B_{\text{min}} = 0.1$), 33/45 (73%) and 45/45 (100%) of the domains were found to bind full-length HPV16 E6 or HPV18 E6, respectively, by GPCA (Fig. 4b and Supplementary Table 10). Next, we retrieved and subcloned the cDNAs of 17 full-length cellular proteins containing E6-binding PDZ domains identified by the holdup assay, and we probed by GPCA their interaction with full-length HPV16 and HPV18 E6 proteins. 13/17 (76%) and 15/17 (88%) of these proteins were seen to interact with HPV16 E6 and HPV18 E6, respectively (Fig. 4c and Supplementary Table 10). Performing holdup assays on protein fragments (i.e., peptides vs. domains) thus appears to be a valuable strategy for identifying interactions between full-length proteins.

Attogether, we observed a very high agreement between holdup, GPCA and Y2H assays for domain-peptide interaction discovery. Furthermore, the holdup approach displayed a high sensitivity for low-affinity pairs and presented the unique advantage of providing affinity-based ranking of the identified binders.

**Novel E6 targets potentially involved in cancer**

To investigate whether genes encoding potential E6-binding PDZ proteins identified by holdup are significantly deleted or amplified in tumor samples, we queried the Tumorscape data set containing the sequences of more than 3,000 tumor genomes\(^{19}\) (Supplementary Table 11). The genes of five putative E6 targets (MAGI2, MAGI3, DLG2, PDZRN3 and ARHGEF12) were found to be deleted in at least two different tumor types, and those of four other E6 putative targets (SCRIB, DLG1, SNTB1 and SNX27) were found to be amplified in at least three different tumor types. The cell junction scaffolding MAGI proteins have already been implicated in tumor suppressor functions\(^{20-23}\). The SCRIB and DLG cell polarity proteins are known to display tumor suppressor and oncogene functions depending on the cellular context\(^{24,25}\). ARHGEF12 has also been recently noticed as a potential tumor suppressor\(^{26}\). Whereas MAGI and DLG proteins, SCRIB and—as was very recently discovered—PDZRN3 (ref. 27) are known E6 targets, ARHGEF12, SNTB1 and SNX27 are novel putative E6 targets with a possible link to cancer.

**DISCUSSION**

The automated holdup assay described herein detects complexes at equilibrium, in contrast to several other high-throughput methods for domain-motif binding specificity evaluation such as ELISA\(^{28}\), SPOT assays (synthesis of peptides on cellulose membranes)\(^{29}\) or protein arrays\(^{13,30}\), which all analyze the survival of complexes after washing steps. Because it operates at equilibrium in the presence of high ligand concentrations, the holdup assay is particularly suitable for analyzing low-affinity domain-ligand complexes in the submillimolar range, which may escape detection by methods that require protein concentrations in the range of the $K_D$ investigated, such as SPR, calorimetry or fluorescence polarization. Indeed, many domain-motif interactions display low affinity\(^{12,13,30,31}\) as transient interactions are essential for cell signaling\(^{32}\). The holdup assay, with its ability to accurately rank low-affinity interactions, will thus represent a valuable approach to quantitatively investigate the specificity of cell signaling mediated by large numbers of domains and motifs.

Whereas most approaches for affinity measurements evaluate complexes via indirect properties such as fluorescence\(^{30,33}\), fluorescence polarization\(^{34}\), calorimetry\(^{35}\), SPR\(^{36}\) or antibody-mediated signal\(^{28,29,37}\), the holdup assay directly visualizes the bound and unbound proteins by means of electropherograms that provide, for each individual data point, highly precise qualitative and quantitative information (molecular weight, construct integrity, concentration and purity).

The holdup assay is easy to implement and highly versatile. The protocol demonstrated here uses a multitask liquid-handling robot and a microfluidics capillary instrument, both found in many laboratories where they serve other purposes. An alternative cost-effective protocol can still address up to a few hundreds of interactions per day without using these instruments (Online Methods).

Holdup profiles depict motif specificities, quantify sequence determinants of motif-domain recognition and deliver functionally relevant information. Observed differences in the E6-PDZome binding profiles of HPV types 16 and 18 likely play a role in their distinctive biological and pathological traits. The tight sequence dependency of the profiles sets the stage for
sequence-function studies, which will analyze the functional impact of natural variations or site-directed mutations of motifs with respect to the resulting alterations of their domain-ome binding profiles. The affinity-based ranking contained in binding profiles can also serve as a rational criterion to select a subset of interactions for further functional characterization. Although we focused in this work on motif-domain profiles, we envision future adaptations of the holdup assay for analyzing, at a proteomic scale, interactomes of various domain families with different categories of resin-attachable ligands including all types of modified peptides, sugars, nucleic acids, lipids or small-molecule inhibitors.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**ACKNOWLEDGMENTS**

This work was supported by institutional support from CNRS, Université de Strasbourg, Institutions Biologie Santé et Agronomie (Ibisa) and grants from the Association pour la Recherche contre le Cancer (ARC) (no. 3171), Agence Nationale de la Recherche (ANR-MIME-2007 EPI-HPV-3D), US National Institutes of Health (NIH grant R01CA134737), Ligue contre le Cancer and Alsace against cancer. J.L. was supported by grants of Région Alsace and ARC. J. Poisson was supported by a grant of the Ligue Régionale contre le Cancer. J.-P.B. is supported as a scholar of Institut Universitaire de France. We thank the CSSE (Harvard University) for kindly providing clones of the human ORFeome, and T. Rolland, M.A. Delsuc, D. Altschuh, B. Kieffer, A. Dejaegere and all members of the "Oncoproteins" team for helpful discussions and advice.

**AUTHOR CONTRIBUTIONS**

G.T. conceived the project and raised funding for it. G.T., S.C. and R.V. supervised the project. R.V. and S.C. developed the robotized holdup assay. K.L. designed the project and raised funding for it. G.T., S.C. and R.V. supervised the project. J.R. and J.-P.B. provided the ‘A’ PDZome resource. Y.N. provided the data. J. Poirson, A.F., P.C., Y.J. and M.M. performed the GPCA experiments. S.C., M.-L.S. and K.L. curated and performed all computational data treatment and bioinformatics analysis. T. Rolland, M.A. Delsuc, D. Altschuh, B. Kieffer, A. Dejaegere and all members of the "Oncoproteins" team for helpful discussions and advice.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Reprints and permissions information is available online at [http://www.nature.com/reprints/index.html](http://www.nature.com/reprints/index.html).

1. Nash, P.D. Why modules matter. *FEBs Lett.* **586**, 2572–2574 (2012).
2. Diella, F. et al. Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.* **13**, 6580–6603 (2008).
3. Lievens, S., Eckerman, S., Lemmens, I. & Tavernier, J. Large-scale protein interactome mapping: strategies and opportunities. *Expert Rev. Proteomics* **7**, 679–690 (2010).
4. Tonikian, R. et al. Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins. *PLoS Biol.* **7**, e1000218 (2009).
5. Jadwin, J.A., Ogule-Ikeda, M. & Machida, K. The application of modular protein domains in proteomics. *FEBs Lett.* **586**, 2586–2596 (2012).
6. Doyle, D.A. et al. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell 85*, 1067–1076 (1996).
7. Luck, K., Charbonnier, S. & Trévé, G. The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBs Lett.* **586**, 2648–2661 (2012).
8. Noury, C., Grant, S.G. & Borg, J.P. PDZ domain proteins: plug and play? *Sci. STKE* **2003**, RE7 (2003).
9. Linder, S. & Collard, J.G. The link between small GTPases and polarity proteins in cell polarization. *Nat. Rev. Mol. Cell Biol.* **9**, 846–859 (2008).
10. Javier, R.T. & Rice, A.P. Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. *J. Virol.* **85**, 11544–11556 (2011).
11. Charbonnier, S., Zanier, K., Masson, M. & Travé, G. Capturing protein-protein complexes at equilibrium: the holdup comparative chromatographic retention assay. *Protein Expr. Purif.* **50**, 89–101 (2006).
12. Luck, K. et al. Putting in vitro practice domain-in-line motif interaction predictions for exploration of protein networks. *PLoS ONE* **6**, e25376 (2011).
13. Stiffer, M.A. et al. PDZ domain binding selectivity is optimized across the mouse proteome. *Science* **317**, 364–369 (2007).
14. Kranjec, C. & Banks, L. A systematic analysis of human papillomavirus (HPV) E6 PDZ substrates identifies MAGI-1 as a major target of HPV type 16 (HPV-16) and HPV-18 whose loss accompanies disruption of tight junctions. *J. Virol.* **85**, 1757–1764 (2011).
15. Zanier, K. et al. Solution structure analysis of the HPV16 E6 oncoprotein reveals a self-association mechanism required for E6-mediated degradation of p53. *Structure* **20**, 604–617 (2012).
16. Atchley, W.R., Zhao, J., Fernandes, A.D. & Driüke, T. Solving the protein sequence metric problem. *Proc. Natl. Acad. Sci. USA* **102**, 6395–6400 (2005).
17. Belotti, E. et al. The human PDZome: a gateway to PDZ mediated functions. *Mol. Cell. Proteomics* **12**, 2587–2603 (2013).
18. Cassonnet, P. et al. Benchmarking a luciferase complementation assay for detecting protein complexes. *Nat. Methods* **8**, 990–992 (2011).
19. Beraoukhim, R. et al. The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
20. Zarić, J. et al. Identification of MAGI1 as a tumor-suppressor protein induced by cyclooxygenase-2 inhibitors in colorectal cancer cells. *Oncogene* **31**, 48–52 (2012).
21. Lee, S.J. et al. MAGI-3 competes with NHERF-2 to negatively regulate LPA2 receptor signaling in colon cancer cells. *Gastroenterology* **140**, 924–934 (2011).
22. Valiente, M. et al. Binding of Pten to specific PDZ domains contributes to Pten protein stability and phosphorylation by microtubule-associated serine/threonine kinases. *J. Biol. Chem.* **280**, 28936–28943 (2005).
23. Kotelevets, L. et al. Implication of the MAGI-1b/Pten signalosome in stabilization of adherens junctions and suppression of invasiveness. *FASEB J.* **19**, 115–117 (2005).
24. Ellenbroek, S.I., Iden, S. & Collard, J.G. Cell polarity proteins and cancer. *Semin. Cancer Biol.* **22**, 208–215 (2012).
25. Roberts, S., Delury, C. & Marsh, E. The PDZ protein discs-large (DLC): the 'Jekyll and Hyde' of the epithelial polarity proteins. *FEBS J.* **279**, 3549–3558 (2012).
26. Öng, D.C. et al. LARG at chromosome 11q23 has functional characteristics of a tumor suppressor in human breast and colorectal cancer. *Oncogene* **28**, 4189–4200 (2009).
27. Thomas, M. & Banks, L. PDZRN3/LNK3 is a novel target of human papillomavirus type 16 (HPV-16) and HPV-18 E6. *J. Virol.* **89**, 1439–1444 (2015).
28. Hu, H. et al. A map of WW domain family interactions. *Proteomics* **4**, 643–655 (2004).
29. Landgraf, L. et al. Protein interaction networks by proteome peptide scanning. *PLoS Biol.* **2**, E14 (2004).
30. Kaušansky, A. et al. Quantifying protein-protein interactions in high throughput using protein domain microarrays. *Nat. Protoc.* **5**, 773–790 (2010).
31. Babault, N. et al. Peptides targeting the PDZ domain of PTPN4 are efficient inducers of glioblastoma cell death. *Structure* **19**, 1518–1524 (2011).
32. Perkins, J.R., Dilboun, I., Dessayl, B.H., Lees, J.G. & Orengo, C. Transient protein-protein interactions: structural, functional, and network properties. *Structure* **18**, 1233–1243 (2010).
33. Gerber, D., Maerkl, S.J. & Quake, S.R. An in vitro microfluidic approach to generating protein-interaction networks. *Nat. Methods* **6**, 71–74 (2009).
34. Hause, R.J. Jr. Comprehensive binary interaction mapping of SH2 domain proteins via fluorescence polarization reveals novel functional diversification of ErbB receptors. *PLoS ONE* **7**, e44471 (2012).
35. Vediaez, A. et al. ITC in the post-genomic era…? Priceless. *Biophys. Chem.* **115**, 115–124 (2005).
36. Rich, R.L. & Myszka, D.G. Advances in surface plasmon resonance biosensor analysis. *Curr. Opin. Biotechnol.* **11**, 54–61 (2000).
37. Weiser, A.A. et al. SPOD synthesis: reliability of array-based measurement of peptide binding affinity. *Anal. Biochem.* **342**, 300–311 (2005).
ONLINE METHODS

Cloning of human PDZ domains. PDZ domains of benchmark data set. For benchmarking the automated holdup assay, we used five human PDZ domain constructs (MAGI1-2/6, MAGI1-3/6, SCRIB-3/4, SCRIB-4/4, SCRIB-34/4), which we had previously cloned into PETM41 bacterial expression vectors (EMBL Heidelberg and described in ref. 12; for sequence information, see category “C” in Supplementary Table 2).

Control PDZ domains. For each holdup experiment performed in 96- or 384-well-plate format in the context of the PDZome screen, we ran control measurements including six PDZ domain constructs whose diverse binding behavior to HPV16 screen, we ran control measurements including six PDZ domain constructs used in the benchmarking (see previous paragraph) and the two PDZ constructs SCRIB-1/4 and SCRIB-2/4, which have been cloned identically to the others (for sequence information see category “C” in Supplementary Table 2).

PDZ domains of PDZome. To clone an expression library of the human PDZome (266 PDZs in total7), we retrieved 224 out of 246 human PDZ domain constructs previously cloned by Belotti et al17. The sequences of these constructs are provided in Supplementary Table 2 (category “A”). We completed the collection by cloning synthetic genes (NZYtech, Portugal) of 42 PDZ domains (category “B” in Supplementary Table 2). The design of the N-terminal and C-terminal boundaries of the “A” and “B” constructs has been previously described in ref. 17 and ref. 7, respectively. Both A and B types of constructs were initially cloned into the pDONRZeo entry vector using Gateway BP reactions and then transferred using Gateway LR reaction into the His-pKM596 vector38.

Both His-pKM596 and PETM41 vectors allow bacterial overexpression of the cloned domains as fusions to His-tagged maltose-binding protein (His-MBP). The MBP tag is known to optimize the folding and the solubility of passenger proteins39, and it also increases the molecular weight of the constructs, facilitating their detection and quantification by microfluidic capillary electrophoresis.

Note that in this study we used for each of the six PDZ domains MAGI1-2/6, MAGI1-3/6, and PDZ 1-4 of SCRIB two clones (“A” and “C”) differing in their N- and C-terminal domain boundaries. Both clones have been used in the PDZome screening experiments with the C clones being used as control (see previous paragraph). If measurements of several clones of the same PDZ domain have been available, final data analysis and plotting was performed on clone of type B if available, on C if B data were not available, or on A if neither C nor B data were available.

The full PDZome clone collection (266 clones) was sequenced in a thermostated multiplate shaker. The cells were harvested by centrifugation. To increase biomass, we grew each clone in triplicate (three 4-ml cultures). The 798 cultures for expressing the full PDZome were all performed simultaneously. Each pellet was resuspended in 600 µl of lysis buffer (buffer A: 50 mM Tris (pH 8.0), 300 mM NaCl, 10 mM imidazole, 5 mM DTT, supplemented with 17.25 µM lysozyme). The three resuspended pellets of each His-MBP-PDZ domain production (1.8 ml in total) were pooled into three 96–deep-well plates (DW 96) labeled PDZ 1-96, PDZ 97-192 and PDZ 193-246. An 80-µl aliquot of each suspension was transferred into separate 96-well PCR plates. Cell pellet resuspensions were frozen at −20 °C until further processing.

Lysis of the 80-µl aliquot and quantification of His-MBP-PDZ domains. For bacterial lysis, the frozen cell suspensions of the 80-µl aliquots of each clone were first thawed by incubation for 20 min at 37 °C and were then supplemented with 10 µg/ml DNase I and 20 mM MgSO4 and further incubated for 15 min at 17 °C with agitation. The lysed cell suspension was transferred into Eppendorf tubes and cleared by centrifugation. The lysates were diluted 4× with lysis buffer, and His-MBP-PDZ protein concentration was measured by quantitative capillary gel electrophoresis (LabChip GXII, Caliper, PerkinElmer). A dilution matrix was generated, the lysates were adjusted accordingly with lysis buffer to a final His-MBP-PDZ concentration of 5 µM.

Lysis of the expression cultures and quantification of His-MBP-PDZ domains. Lysis of the His-MBP-PDZ expressions was performed using the protocol mentioned above. Cleared lysates were diluted to a final concentration of 5 µM His-MBP-PDZ using the dilution matrix based on the 80-µl aliquots. The PDZome was divided into aliquots of 40 µl into series of 96-well plates (representing 45 plates of PDZ 1-96, PDZ 97-192 and PDZ 193-246 for one PDZome culture) and frozen at −20 °C before being used for holdup experiments. For each new PDZome, one series of PDZ 1-96, PDZ 97-192 and PDZ 193-246 adjusted at 5 µM was thawed and quantified a second time using capillary gel electrophoresis, and a new dilution matrix was applied to adjust all His-MBP-PDZ samples precisely to 4 µM before the holdup experiment. This allowed for removal of any discrepancies in His-MBP-PDZ concentration due to slight differences in the initial cleared lystate preparation. This second dilution matrix was applied on the robot systematically at the beginning of each of the 44 holdup experiments for a given PDZome expression.
Purification of the His-MBP-PDZ domains. All the PDZ domains (benchmarking set, control PDZ and the 246 PDZ domains of the PDZome) were expressed and purified by using strictly the high-throughput protocol previously described\(^\text{40,41}\) in deep-well block 24. Briefly, after harvesting the cells by centrifugation (4-mL cultures), we resuspended each pellet in 1,000 µl of lysis buffer (buffer A: 50 mM Tris (pH 8.0), 300 mM NaCl, 10 mM imidazole, 5 mM DTT, supplemented with 17.25 µM lysozyme) and froze the suspensions at −20 °C until further processing.

For protein purification, the frozen cell pellets were thawed by incubation for 20 min at 37 °C and then supplemented with 10 µg/ml DNase I and 20 mM MgSO\(_4\) and further incubated for 15 min at 17 °C with agitation. 50 ml of Ni Sepharose 6 FF beads (GE-Healthcare: 17-5318-02) were added into the lysed cell suspension, incubated and transferred into a filter plate (Novagen 71196-3). The beads were washed and eluted with 200 µl of buffer A containing 250 mM imidazole. The concentration of the His-MBP-PDZ proteins was measured by quantitative microfluidic capillary gel electrophoresis (LabChip GXII, Caliper, PerkinElmer), and a dilution matrix was calculated in order to adjust each lysate to a final His-MBP-PDZ concentration of 4 µM.

Preparation of peptides. Biotinylated peptides were synthesized by either JPT Innovative Peptide Solutions (https://www.jpt.com/) with 70–80% purity or by the peptide synthesis service at IGBMC (http://www.igbmc.fr/) with 80–90% purity. All peptides are N-terminally coupled to biotin. Peptides provided by JPT are coupled to biotin via a TTDS (N-(13-amino-4,7,10-trioxatriacyley)succinamic acid) linker. Peptides provided by IGBMC are directly chemically linked to biotin via a peptide bond. The lyophilized peptides were taken up into water to a stock concentration of 250 mM. After that the pH was adjusted to roughly pH 8 (with NaOH), and the solutions were sampled as 100 µl aliquots and stored at −20 °C.

The holdup assay. Principle. The holdup assay is based on the principle of comparative chromatographic retention of ligand-analyte pairs at equilibrium conditions. Two equivalent batches of affinity chromatographic beads are coated with either a ligand and unbound analyte amounts of the ligand and negative-control peptide (GSNSGN)GNS), which should not display any interaction to any analyte. Both batches are incubated with analyte protein in identical experimental conditions until binding equilibrium between ligand and analyte is reached in solution. Now washing steps are performed, and therefore the binding equilibrium is not disturbed. Liquid phases are rapidly collected by filtration, and unbound analyte amounts of the ligand and negative-control experiment are compared to each other. This comparison reveals the proportion of free and ligand-bound analyte at equilibrium conditions and thus allows us to obtain an estimate of the binding intensity of tested ligand-analyte pairs.

This method can be performed using purified analyte or cleared analyte expression extracts. Samples can be tested in a high-throughput automated format or manually at lab scale using either 96- or 384-well-plate format.

Automated holdup assay protocol. Automated holdup assays were performed on a Tecan Evo200 robot with a 96-tip pipetting head and an eight-needle pipetting arm using MSDVN6B or MZHVNOW plates from Millipore for 96- or 384-well-plate formats (HU96 and HU384), respectively. 1.5 ml (HU96 and HU384) of streptavidin Sepharose high-performance beads (17-5113-01, GE Healthcare) were equilibrated in 45 ml buffer A (50 mM Tris (pH 8.0), 300 mM NaCl, 10 mM imidazole, 5 mM DTT). The beads were resuspended by vigorous up-and-down pipetting before being transferred by means of a multipipette or the 96-needle robotic pipetting arm into the appropriate filter plate placed on a vacuum station (MultiScreen vacuum manifold (MAVM 096 OR, Millipore) for the manual protocol or a robot-driven T-Vac vacuum system (Tecan)). Two aliquots of 150 µl (2 x 150 µl) of bead suspension (representing 10 µl of beads (HU96)) or 75 µl of suspension (representing 2.5 µl of beads (HU384)) were transferred into the wells of the filter plate. The liquid phase was discarded by vacuum filtration. Beads were equilibrated by adding 30 volumes of buffer A per volume of beads, followed by immediate removal of the buffer by vacuum filtration. This washing step ensures that the beads are perfectly equilibrated. Eight volumes per bead volume of a 42 µM stock solution of biotinylated peptide were transferred into 48 wells (HU96) or 288 wells (HU384) of a filter plate (see Fig. 3a). Eight volumes per bead volume of a 42 µM stock solution of biotinylated peptide were transferred into 48 wells (HU96) or 288 wells (HU384) of a filter plate (see Fig. 3a). Eight volumes per bead volume of a 42 µM stock solution of biotin was transferred into the remaining wells of the filter plate serving as negative control. The 96-well protocol requires two 96-well filter plates and two separate holdup assays for testing 96 His-MBP-PDZ domains against one peptide. In case of a 384-well plate holdup assay, each of the 4 × 96 His-MBP-PDZ domains was transferred into three wells containing beads saturated with biotinylated peptides and one well containing beads saturated with biotin (negative control), allowing us to test 96 PDZ domains against three peptides on one filter plate (see Fig. 3a). According to the manufacturer’s specifications, streptavidin Sepharose high-performance beads have a capacity of 150–300 µM of biotin binding sites (considering equilibrated bedded resin without any extra liquid). With this protocol we assume the beads are saturated with either biotinylated peptides or biotin.

The filter plate containing the biotinylated peptide-resin and biotin-resin mixes was transferred onto a multiwell plate shaker and incubated under vigorous agitation at 1,200 r.p.m. for 15 min. After saturation of the beads with biotinylated peptide or biotin, the liquid phase was removed by vacuum filtration. Remaining free streptavidin sites on the beads were blocked by incubating a 1 mM biotin solution in every well under agitation at 1,200 r.p.m. for 15 min. Liquid phases were extracted by vacuum filtration.

Two volumes per resin volume of a 4 µM His-MBP-PDZ domain solution (analytes) were transferred into the filter plate. Analytes can be provided either as purified samples or as cleared soluble E. coli overexpression extracts, both adjusted to 4 µM with buffer A. The bacterial overexpression extracts all contain 17.25 µM lysozyme provided by the lysis buffer (see section "Lysis of the 80-µl aliquot and quantification of His-MBP-PDZ domains"). Successive dilutions for adjusting the His-MBP-PDZ concentration in the extracts to 4 µM were performed with lysis buffer also containing 17.25 µM lysozyme, thereby providing a reference lysozyme peak for internal normalization during data treatment. Experiments using purified His-MBP-PDZ domains had not been provided with a lysozyme reference for internal normalization.

At this step we assume that the biotinylated peptide concentration in the resin-PDZ mix is about 75 µM and therefore in
a 20-fold excess compared to the His-MBP-PDZ domain concentration. In such conditions, assuming that the His-MBP-PDZ domains in the mix are active to 100%, the fraction of His-MBP-PDZ domain, which will bind to the biotinylated peptides and therefore be retained on the resin, will be more than 90% for any domain–peptide pair displaying a dissociation constant ($K_D$) better than 8 μM.

The plate with the His-MBP-PDZ-peptide mixes was incubated under vigorous agitation at 1,200 r.p.m. for 15 min. During this incubation step the interaction between domain and peptide reaches its equilibrium in solution. Longer incubation times (30 or 60 min) do not change the result of the assay (Charbonnier et al., and data not shown). After this incubation the liquid phase is rapidly collected by filtration and collected into a 384- or 96-well PCR plate according to the format of the used filter plate. To ensure precise and reproducible volumes of filtered liquid, we performed an extra centrifugation step for Hu384. The filtered flow-throughs were blocked with LabChip GX II sample buffer according to manufacturer instructions. Alternatively, they can be blocked with SDS-PAGE sample buffer, which can be used either for qualitative SDS-PAGE analysis but could alternatively be diluted with LabChip GX II sample buffer for quantitative microfluidic capillary gel electrophoresis analysis. Blocked samples were boiled at 95 °C for 5 min and samples were either subjected directly to LabChip GX II capillary gel electrophoresis or stored at −20°C before migration.

Manual holdup assay protocol. The above automated holdup assay protocol using 96-well or 384-well formats on Tecan robots can be performed at laboratory scale using eight-channel micro-pipettes with variable tip spacing, a vacuum manifold device (MultiScreen vacuum manifold (MAVM 096 OR, Millipore) and an MS3 digital multiwell plate shaker (ika Werke GmbH). At lab scale, experimental samples can be analyzed qualitatively using SDS-PAGE or quantitatively with microfluidic capillary gel electrophoresis.

Benchmarking the holdup assay (Fig. 2a,b). The 210 benchmark PDZ–peptide pairs were analyzed using the standard 96-well automated holdup assay protocol. The experiments were performed without replicates in two different conditions (using purified proteins or extracts).

Holdup experiment using serial dilutions to improve differentiation of high-affinity interactions. This experiment (Fig. 2c) was performed in 96-well format using crude extracts of the MBP-SCRIB-34/4 tandem construct from the benchmarking experiments adjusted to 4 μM stock concentration. Ten batches of 50 μl of streptavidin beads were batch equilibrated in buffer, saturated with 200 μl of one of ten different biotinylated-peptide 125 μM stock solutions (16E6_10, 16E6_05, 16E6L/V, TAX1_10, TAX1_05, CLTR2_10, CLTR2_05, TANC1_10, LPP and ZO2). These peptides were selected for their binding intensities to MBP-SCRIB-34/4 ranging from 0.5–0.9. The peptide-saturated resin batches were resuspended in buffer, and varying amounts of beads were transferred into a 96-well microtiter filter plate (bead volumes: 25 ml, 6.25 ml, 3.2 ml, 1.9 ml). The beads were saturated with biotin, washed and equilibrated with 2 × 300 μl of buffer as previously described. The different resin aliquots were incubated with 40 μl of varying dilutions of MBP-SCRIB-34/4 extracts (4 μM, 1.3 μM, 0.8 μM, and 0.3 μM, respectively). The remaining steps of the holdup experiment were performed as described above for a standard Hu96 assay.

Results have been analyzed using quantitative microfluidic capillary gel electrophoresis. Binding intensities (BI) have been calculated as described in section “Principles and automation of binding intensity calculation” (see Supplementary Data 1 and 2 (“dilution” folder)).

PDZome profiling experiments (Fig. 4). PDZome profiling experiments in 96-well format were reproduced up to nine times, allowing us to obtain the first profiles and to standardize the assay. The PDZome profiling experiments in 384-well format were systematically performed in triplicate. The total triplicate 384-well experiment was reproduced thrice for HPV16 and twice for HPV18. Figure 4 show the results of one such triplicate experiment. For each PDZ–peptide pair, the number of repeated measurements (3, 2 or 1) that were retained after data curation (see below) for calculating the average BI value is indicated in Supplementary Table 3.

Quantitative analysis using microfluidic capillary gel electrophoresis. Preparation of samples for microfluidic capillary gel electrophoresis measurements. 96-well or 384-well PCR plates containing the blocked and boiled holdup assay samples were directly loaded and measured on a LabChip GX II machine (Caliper, PerkinElmer) equipped with an HT protein express LabChip (760499). The high-sensitivity 10- to 100-kDa protocol was applied, according to the manufacturer instructions.

Principles of microfluidic capillary gel electrophoresis. Filtrates obtained from holdup experiments are denatured and labeled with fluorescent dyes and subjected to microfluidic capillary electrophoresis, resulting in a separation of the proteins of the injected sample based on their molecular mass. During separation, labeled proteins pass a laser that excites the attached fluorescent dyes. Emitted fluorescence signals of labeled proteins can be displayed versus the time in electropherograms produced by the LabChip GX II software. Using internal standard markers, the LabChip GX II software calculates the molecular mass of proteins in the sample and quantifies their concentration. Thus, the time axis can be converted into an axis of protein sizes (see Fig. 1b for an example electropherogram of two overlaid injections).

Manual curation of data before binding intensity calculations. Each holdup plate injected in the LabChip GX II device was visually inspected using the LabChip GX II software. During this step incorrect internal standard markers were excluded and replaced by correct ones when applicable. Invalid holdup measurements were identified and excluded. Main criteria for exclusion were distorted or unrecognizable electropherograms; absence of correct internal standard markers; bands of His-MBP-PDZ and/or lysozyme absent, unrecognized or degraded; signal of MBP too close to background noise in the negative-control samples. After this visual curation the data points were labeled either “yes” (accepted for treatment) or “no” (excluded). On the accepted data points, windows of protein sizes containing His-MBP-PDZ peaks (usually between 50 and 70 kDa) and standard peaks (usually between 10 and 20 kDa) (see next paragraph) were defined. Finally, all electropherogram data of that holdup plate containing the fluorescence signals versus time and protein size were exported in simple text format for further data treatment (see Supplementary Data 1).
Principle and automation of binding intensity calculation. Processing of data exported from LabChip GX II software, as well as binding intensity calculation, has been fully automated and implemented in Python (https://www.python.org/). For automatic peak identification, we used exported protein size data determined by the LabChip GX II software, including corrections for signal shifts occurring during LabChip GX II measurements. To automatically estimate the amount of unbound His-MBP-PDZ domain or standard peaks from exported electropherogram data, we used the maximal height of the fluorescence signals of the corresponding peaks instead of the areas of the peaks. Maximal heights of peaks have been identified by searching the maximal fluorescence signal within the defined protein size windows (see previous paragraph).

We extracted the maximal fluorescence signals of peaks resulting from incubations of His-MBP-PDZ domains with either biotinylated peptide or a negative control (biotin or none peptide). The more His-MBP-PDZ domain retained on the peptide-coated beads, the smaller the fluorescence peak of the His-MBP-PDZ domain in the collected peptide-incubated filtrate as compared to the corresponding fluorescence peak in the negative-control filtrate. This difference in signal is directly correlated with the binding strength of His-MBP-PDZ to the biotinylated peptide. This difference in PDZ peak height is normalized with division by the height of the PDZ peak resulting from the negative-control run to take into account variations in PDZ amounts loaded on the well plate. The binding intensity (BI) is defined as

\[ BI = \frac{I_{ctrl} - I_{lig}}{I_{ctrl}} = 1 - \frac{I_{lig}}{I_{ctrl}} \]

where \(I_{lig}\) and \(I_{ctrl}\) correspond to the maximal fluorescence signals of the His-MBP-PDZ peaks of the biotinylated peptide sample and the negative-control sample, respectively.

Input correction: the standard-peak method. Slight variations in input quantities between the peptide sample and the corresponding negative-control sample (visible by shifts of the fluorescence signals when overlaying peptide- and control-sample electropherograms; see Fig. 1c) can distort calculated binding intensities. We developed a simple strategy, the standard-peak method, which allows input variations to be corrected before binding intensity calculation. This method is based on the determination of an \(\alpha\) factor to rescale the peptide sample to the negative-control signals:

\[ BI = 1 - \frac{\alpha I_{lig}}{I_{ctrl}} \]

Here, \(\alpha\) has been calculated from the fluorescence signal of the protein standard lysozyme that has been added in high and equal amounts to all His-MBP-PDZ extracts:

\[ \alpha = \frac{I_{ctrl, st}}{I_{lig, st}} \]

where \(I_{ctrl, st}\) and \(I_{lig, st}\) are the maximal fluorescence signals of the standard peak of the negative-control and the peptide sample, respectively.

If peptide or negative-control samples displayed negative fluorescence signals, the multiplication with \(\alpha\) would have led to incorrect scaling. Therefore, the minimal fluorescence signals \(m_{lig}\) and \(m_{ctrl}\) determined within a window between 12.5 and 90 kDa were subtracted from the peptide sample and negative-control sample, respectively, resulting in the final equation for binding intensity calculation when incorporating input correction:

\[ BI = 1 - \frac{\alpha (I_{lig} - m_{lig})}{I_{ctrl} - m_{ctrl}} \]

Benchmarking of automated holdup and the standard-peak method. Automated holdup assay and the protocol of binding intensity calculations have been benchmarked on a previously published data set. This data set consists of 210 interactions and noninteractions between five His-MBP-PDZ domain constructs and 42 biotinylated C-terminal peptides for which binding intensities had been obtained in a previous study using surface plasmon resonance (Biacore)\(^{12}\). Like in that study, normalized response units (RUs) obtained for the interaction experiments (see Supplementary Table 1) were used instead of dissociation constants to interpret binding intensities of measured interactions.

Uncorrected binding intensities obtained for crude protein samples using the holdup assay correlate well with the reference binding intensities determined using Biacore \((r = 0.69)\). Better correlations can even be obtained when comparing input-corrected holdup assay binding data with the Biacore data \((r = 0.76)\). This analysis indicates that input correction using standard peaks results in more reliable binding intensities (see Supplementary Data 1 and 2 (folder “benchmarkHU”) as well as Supplementary Table 1 for exported caliper data, parameters and binding intensities).

Application of automated holdup. Exported Caliper files as well as parameters and binding intensities of measurements of the PDZome vs. HPV16 or 18 E6 done in 96- or 384-well-plate format can be found in Supplementary Data 1 and 2 (folders “PDZome_16E6_96” and “PDZome_16_18E6_384”).

Estimation of equilibrium dissociation constants \((K_D)\) from holdup binding intensities. In principle, the equilibrium dissociation constant \(K_D\) for PDZ-peptide complex formation is calculated as follows

\[ K_D = \frac{C_{pdz} \times C_{pept}}{C_{pdz-pept}} \]

where \(C_{pdz}\) and \(C_{pept}\) are the concentrations of free PDZ and free peptide, respectively, and \(C_{pept-pdz}\) is the concentration of the complex.

According to the announced capacity of the streptavidin resin (17-5113-01, GE Healthcare) and to our own resin-saturation experiments (data not shown), the total peptide concentration \((C_{0,pept})\) in our assay conditions is around 50–100 µM. The concentration of PDZ \((C_{0, pdz})\) is adjusted to approximately 4 µM. Therefore, peptide is in large excess relatively to PDZ so that the free peptide concentration will remain practically unaltered by the formation of PDZ-peptide complexes, even if these complexes use up the full amount of PDZ present in the reaction. Therefore, the free peptide concentration \((C_{pept})\) is approximately equal to the total peptide concentration \((C_{0, pept})\):

\[ C_{pept} = C_{0,pept} \]

The \(K_D\) can thus be approximated as follows:

\[ K_D = \frac{C_{pdz}}{C_{pdz-pept}} C_{0,pept} \]
In electropherograms, for a given molecular size (horizontal axis) the concentration of protein is supposed to be proportional to the signal (vertical axis) after subtraction of the background intensity. The concentration of free PDZ, \( C_{\text{pdz}} \), can thus be obtained from the experiment as follows:

\[
C_{\text{pdz}} = \lambda (I_{\text{lig}} - I_{\text{bg}})
\]

where \( I_{\text{lig}} \) is the PDZ signal detected in the flow-through after incubation with ligand-saturated resin, \( I_{\text{bg}} \) is the background intensity contained within the PDZ signal, and \( \lambda \) is the proportionality factor between PDZ concentration and background-subtracted PDZ signal.

The concentration of bound PDZ, \( C_{\text{pept-pdz}} \), can be obtained from the experiment as follows:

\[
C_{\text{pept-pdz}} = \frac{\lambda (I_{\text{biot}} - I_{\text{bg}})}{I_{\text{lig}}} C_{\text{pept}}^0
\]

where \( I_{\text{biot}} \) is the signal of the full PDZ present in the reaction, detected in the flow-through of the negative-control experiment.

Therefore, each single \( K_D \) value can be approximated as follows:

\[
K_D = \frac{I_{\text{lig}} - I_{\text{bg}}}{I_{\text{biot}} - I_{\text{lig}}} C_{\text{pept}}^0
\]

On the other hand, the holdup binding intensity (BI) is defined as follows:

\[
\text{BI} = \frac{I_{\text{biot}} - I_{\text{lig}}}{I_{\text{biot}}}
\]

Therefore, we have

\[
I_{\text{lig}} = (1 - \text{BI}) \times I_{\text{biot}}
\]

and

\[
I_{\text{biot}} - I_{\text{lig}} = \text{BI} \times I_{\text{biot}}
\]

so that equation (1) can also be written:

\[
K_D = \frac{1 - \text{BI} - I_{\text{bg}} / I_{\text{biot}}}{\text{BI}} C_{\text{pept}}^0
\]

Whereas the BI is directly obtained for each PDZ-peptide pair in the experiment, the values of the background correction factor \( I_{\text{bg}} / I_{\text{biot}} \) and the peptide concentration \( C_{\text{pept}}^0 \) must be evaluated for the entire experiment. \( C_{\text{pept}}^0 \) should be between 50 and 100 \( \mu \text{M} \), and \( I_{\text{bg}} / I_{\text{biot}} \) should be between 0 and 0.1 as the best binder in our assay has a BI of ~0.9. For a precise estimation of these two values, we included in all our experiments three standard PDZ-peptide pairs, for which we had previously published \( K_D \) values obtained with Biacore: MAG11-2/6–HPV16 E6 (\( K_D = 2.5 \) \( \mu \text{M} \))\(^{12,42,43} \), SCRIB-3/4–HPV16 E6 (\( K_D = 3 \) \( \mu \text{M} \))\(^12 \) and SCRIB-3/4–HPV18 E6 (\( K_D = 12 \) \( \mu \text{M} \))\(^12 \). The best fit between these three \( K_D \) values and the BIs that we measured for these three standards in the triplicate 384-well-plate experiment is obtained with a correction factor \( I_{\text{bg}} / I_{\text{biot}} = 0.097 \) and a peptide concentration \( C_{\text{pept}}^0 = 86.75 \) \( \mu \text{M} \). Therefore, these values of \( I_{\text{bg}} / I_{\text{biot}} \) and \( C_{\text{pept}}^0 \) were used for all \( K_D \) estimations of triplicate holdup experiments (Supplementary Table 3).

**Statistics.** Correlations and significances were determined using the stats module of SciPy (http://www.scipy.org/). Linear correlations between data sets were assessed using Pearson correlation coefficients. Significance of correlations was calculated using the two-tailed Student’s \( t \)-test. Data sets were not normally distributed (assessed with a combination of the skewtest and kurtosistest, function normaltest of stats module of SciPy) but of sufficiently large size (between 192 and 208 data points) to justify the application of this test.

**Box-and-whisker plot.** Box-and-whisker plots (see Fig. 3b) were used to analyze the reproducibility of three categories of automated holdup experiments: repetitions of measurements on different 96-well plates, repetitions of measurements on different 384-well plates, and comparison of measurements performed in 96- and 384-well-plate format. For the latter comparison, the correlation value was \( r = 0.92 \). Measurements for this analysis were obtained from screening of the PDZome vs. HPV16 E6 on 96- and 384-well plates excluding all measurements involving the six control PDZ clones (see “Control PDZ domains”). Standard deviations of repeated measurements of all PDZ-peptide pairs per category were determined and plotted in box-whisker mode, thereby allowing the range and distribution of s.d. values obtained within a given category to be visualized. The number of repetitions available for all PDZ-peptide pairs within a given category varied (mostly either two or three repetitions). To increase comparability, we determined s.d. values of all possible pairs of repeated measurements obtained for a given PDZ-peptide pair within a given category. In box-and-whisker plots we have used the following scheme: lower whisker (error bar), min s.d. observed (note that this lower whisker is hidden under the lower box margin in all plots of Fig. 3b); lower box margin, 10th percentile; black line, median; upper box margin, 90th percentile; upper whisker (error bar), max s.d. observed.

**Validation of interactions in mammalian cells. Principle.** The split *G. princeps* luciferase protein complementation assay (GPCA)\(^18 \) is based on reconstitution of luciferase activity upon interaction of two tested proteins fused each to a fragment of luciferase. A normalized luminescence ratio (NLR) is computed from the luminescence signals generated by the potential interactors and two controls. According to a benchmarking study of the GPCA approach, assayed protein pairs with an NLR >3.5 have a probability greater than 97.5% to be true interactors\(^18 \). This cutoff has been applied in the present study to confirm interactions. It is worth noting, however, that NLR <3.5 does not mean “no interaction.” It means that “according to the published benchmark of the GPCA\(^18 \), the probability that this is a true interaction is below 97.5%.”

**Cell lines.** HeLa WS cells were obtained from the cell culture platform of the IGBMC (Illkirch), which regularly checks the authenticity of cell lines. Absence of mycoplasms in the cells was regularly verified.

**Cell culture.** HeLa WS cells were grown and maintained in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% FCS and 50 \( \mu \text{g/ml} \) of gentamycin at 37 °C with 5% CO\(_2\) and 95% humidity.
Cloning and plasmids. The two ORFs encoding HPV16 E6 and HPV18 E6 were amplified by PCR and cloned into vector pDONR207 by recombinational cloning system (Gateway recombinational cloning system, Invitrogen). The resulting Entry clones were then transferred into Gateway-compatible GPCA destination vector pSPICA-N2. ORFs encoding full-length PDZ proteins or PDZ domains were obtained from the Human ORFeome resource (Center for Cancer Systems Biology, Dana-Farber Cancer Institute) and PDZ domains collections, respectively, and directly transferred into the Gateway-compatible GPCA destination vector pSPICA-N1.

GPCA vectors pSPICA-N1 and pSPICA-N2 (both derived from the pciNeo mammalian expression vector) respectively express the Gluc1 and Gluc2 complementary fragments of the G. princeps luciferase linked to the N-terminal ends of tested proteins by a flexible hinge polypeptide of 20 amino acid residues. To normalize expression levels, a Kozak consensus translation start sequence was included at the N-terminal end of the fusion protein

GPCA experiment. HeLa WS cells were seeded in white 96-well plates at a concentration of 2 × 10^4 cells per well. After 24 h, cells were transfected using JetPEI (Polyplus transfection) with 100 ng of pSPICA-N2 expressing HPV16 or HPV18 E6 and 100 ng of pSPICA-N1 expressing full-length PDZ protein or PDZ domain. At 24 h post-transfection, cells were washed with 50 µl of PBS and harvested with 40 µl per well of Renilla lysis buffer (Promega, E2820) for 30 min. G. princeps luciferase enzymatic activity was measured using a Berthold Centro LB960 luminometer by injecting 50 µl per well of luciferase substrate reagent (Promega, E2820) and counting luminescence for 10 s. GPCA experiments were performed in duplicate. Results were expressed as a fold change normalized over the sum of controls, specified herein as NLR

For a given protein pair A-B, the “pSPICA-N1-A+pSPICA-N2-B” couple luminescence activity was divided by the sum of luminescence activity for negative controls “pSPICA-N1-A+empty pSPICA-N2” and “empty pSPICA-N1+ pSPICA-N2-B.”

The GPCA experiments (Fig. 4 and Supplementary Table 10) were performed on 52 PDZ domains and 17 human proteins. These experiments were performed in duplicate with one single experiment performed the same day with the same batch of cells. The data shown correspond to one such duplicated experiment. The resulting s.d. values are reported in Supplementary Table 10. The entire experiment was repeated twice and yielded highly comparable results each time.

Bioinformatics analysis. Search for residues of PDZ domains that are either preferred or disliked for E6 binding. For this analysis we considered all PDZ domains that gave a reliable measurement for both HPV16 E6 and HPV18 E6 (209 PDZs in total). We assigned to each of these 209 PDZs the greater BI that this PDZ obtained from its binding to 18 and 16 E6 (Supplementary Table 4). Negative BIs were set to 0. Next, the 209 PDZ domains were aligned using ClustalW (http://www.clustal.org/). The resulting alignment was manually corrected and refined using Jalview (http://www.jalview.org/) and available structural information of PDZ domains from the PDB. The aligned sequences were then ranked from highest to lowest BI (Supplementary Fig. 6 and Supplementary Data 3). For each amino acid at each alignment position, we built a cumulative count by adding 1 if the given PDZ domain had that specific amino acid at that particular alignment position, or adding 0 otherwise. The count was performed by starting from the bottom of the alignment (lowest BIs). This cumulative count was then normalized by division with the total number of PDZ domains that had this amino acid at this alignment position. If an amino acid x at alignment position y is strongly supporting binding to E6, one would expect a distribution of the cumulative count staying close to 0 for PDZ domains that did not bind to E6 and then rising for PDZ domains that bound E6. This distribution can be plotted as a function f(x, y) versus the ordered PDZ domains (Supplementary Fig. 7a). The stronger the selection of amino acid x for E6 binding, the lower the area under the plotted curve (AUC). Amino acids that do not play a direct role for E6 binding would result in an AUC around 104 (209 × 1/2). Amino acids that are strongly disliked for E6 binding would result in an AUC greater than 104.

Which AUCs are significantly different from random? The random distribution of such AUCs depends on the residue count (i.e., a distribution of random AUCs is different for residue counts of 10 versus 200). For each possible count between 1 and 209, we performed 10,000 randomizations (randomly assigning 1 n times on 209 positions), resulting in 10,000 AUCs defining a random distribution of AUCs for a specific count of residues. To determine the significance of an AUC found for a real residue x at alignment position y, we compared its AUC to the corresponding random distribution and calculated a P value representing the likeliness of randomly finding an AUC as extreme as the one observed (Supplementary Fig. 7b). With a P-value cutoff of 0.001, 21 residues (including a gap) at 20 specific alignment positions were found to be significantly correlated with either preferred or disliked binding to E6 (Supplementary Table 5; the z score indicates whether observed AUCs had been lower (preference for E6 binding) than random distributions or greater).

Search for residue positions in PDZ domains that correlate with a preferred binding to the C-terminal peptide of HPV18 E6 (18E6) vs. HPV16 E6 (16E6). All PDZ domains, which bound to HPV16 or HPV18 E6 with a BI ≥0.1 were taken into consideration for this analysis (60 PDZs in total). For each of these PDZ domains, the difference in BI for HPV18 vs. HPV16 E6 (∆BI) was determined (Supplementary Table 6). Negative BIs were set to 0 before difference calculation. The protein sequences of these PDZ domains were aligned (see Supplementary Fig. 8) as described before. The aligned sequences were ranked from highest to lowest ∆BI. We used four of five amino acid indices published by Atchley et al. (we excluded index 4) to transform amino acids in the PDZ alignment into numerical values representing different physicochemical amino acid properties. Index 1 represents polarity, index 2 secondary structure propensity, index 3 molecular size, and index 5 electrostatic charge.

We systematically searched for correlations between differences in BI of PDZ domains and each of the four amino acid indices for each column in the PDZ alignment where at least half of the PDZ domains did not have a gap. Pearson correlation coefficients and P values were calculated for each set of data points using the stats module of Scipy. P values were corrected for multiple testing. With a P-value cutoff of 0.01, 13 different alignment positions were found to significantly correlate with differences in BI values using at least one of the four indices (see Supplementary Table 7).
We controlled for the significance of these correlations by performing 100,000 randomizations using the random module of Python. We randomly assigned, to all the PDZ domains used in this analysis, BI values selected from within the range of BIs observed from the real data. Significantly correlated domain positions had been determined as described above.

PyMOL (http://www.pymol.org/) was used to create figures of PDZ structures with highlighted domain positions.

Search for occurrences of PDZ proteins binding 18E6 or 16E6 in oncogene and tumor suppressor data. We determined all PDZ proteins that had at least one PDZ domain binding to 18E6 or 16E6 with a BI ≥ 0.2 (30 PDZ proteins in total out of 118 in this study; see Supplementary Table 11).

We used two lists of genes annotated to be drivers in cancerogenesis from Vogelstein et al.44 and that were observed to be modified in cancer either via nonsynonymous mutations or copy-number variations. None of the PDZ proteins found to bind 18E6 or 16E6 in our assay occurred in these two lists.

We also used a list of driver genes in cancer as annotated from the Sanger Cancer Gene Consensus (http://cancer.sanger.ac.uk/census/). ARHGEF12 was the only PDZ protein found to bind E6 and occurring in this list of drivers. From this annotation it is unclear whether it belongs to the tumor suppressor or oncogene class.

Last, we used annotations of genes found to be significantly altered in different cancer types via somatic copy-number alterations as made publicly available by Tumorscape19. We focused on those genes that were significantly altered (q value ≤ 0.25) and were predicted to be in a peak region (as defined by Beroukhim et al.19). This allowed us to obtain, for the 118 PDZ proteins of our assay (binding HPV 18 or 16 E6 or not), a list (Supplementary Table 11) of the cancer types in which the genes of these proteins have been observed to be significantly altered (separated into amplified or deleted) and to occur within a peak region.

38. Tropea, J.E., Cherry, S., Nallamsetty, S., Bignon, C. & Waugh, D.S. A generic method for the production of recombinant proteins in Escherichia coli using a dual hexahistidine-maltose-binding protein affinity tag. Methods Mol. Biol. 363, 1–19 (2007).
39. Nominé, Y. et al. A strategy for optimizing the monodispersity of fusion proteins: application to purification of recombinant HPV E6 oncoprotein. Protein Eng. 14, 297–305 (2001).
40. Vincentelli, R. et al. High-throughput protein expression screening and purification in Escherichia coli. Methods 55, 65–72 (2011).
41. Saez, N.J. & Vincentelli, R. High-throughput expression screening and purification of recombinant proteins in E. coli. Methods Mol. Biol. 1091, 33–53 (2014).
42. Fournane, S. et al. Surface plasmon resonance analysis of the binding of high-risk mucosal HPV E6 oncoproteins to the PDZ1 domain of the tight junction protein MAGI-1. J. Mol. Recognit. 24, 511–523 (2011).
43. Charbonnier, S. et al. The structural and dynamic response of MAGI-1 PDZ1 with noncanonical domain boundaries to the binding of human papillomavirus E6. J. Mol. Biol. 406, 745–763 (2011).
44. Vogelstein, B. et al. Cancer genome landscapes. Science 339, 1546–1558 (2013).