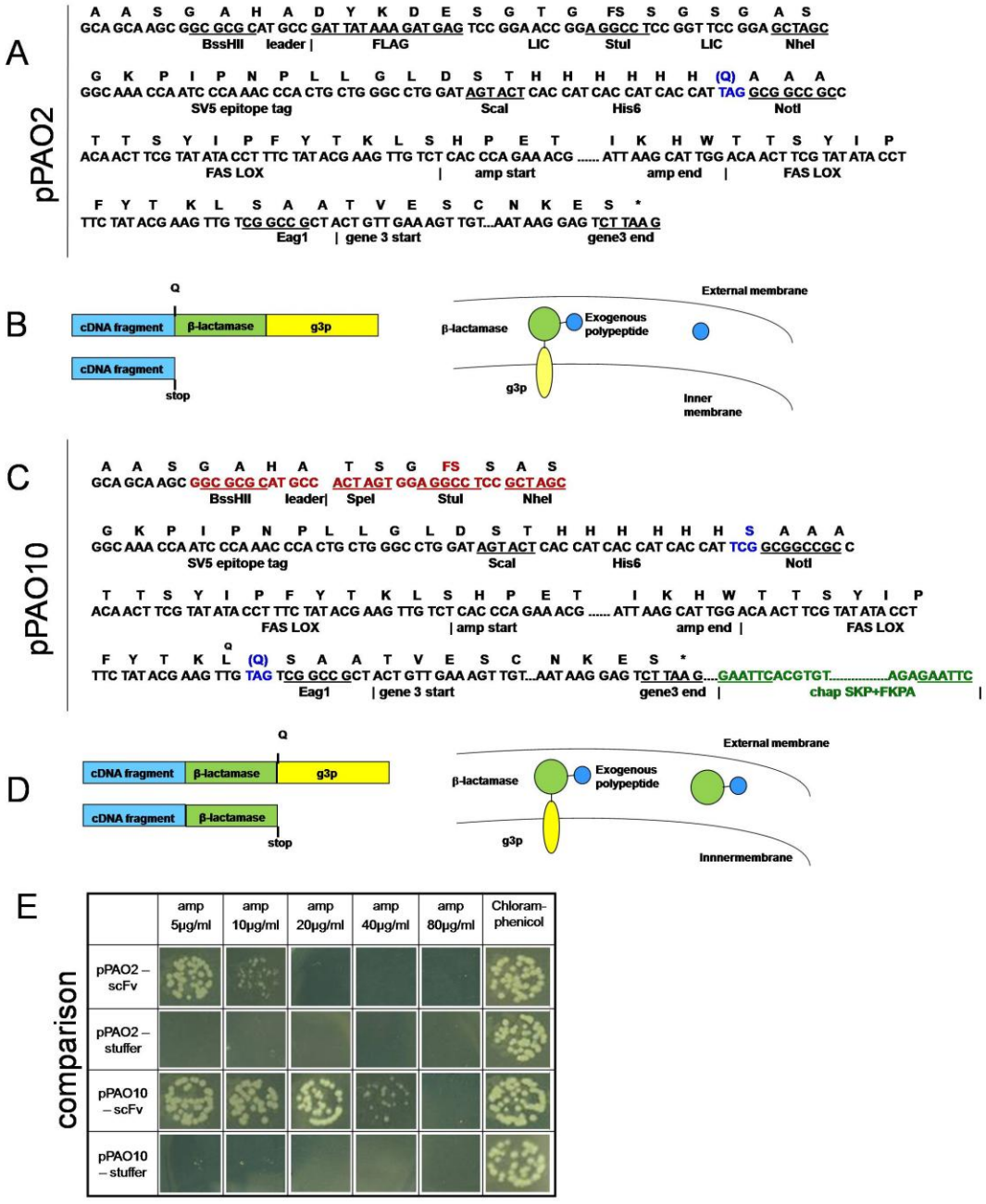
SUPPLEMENTARY INFORMATION

Supplementary Figure 1. Characteristics of the novel pPAO10 vector.



A) Scheme of the original phagemid vector pPAO2; **B**) the presence of the amber stop codon TAG results in the production of two polypeptides, only one of which in fusion with β-lactamase; **C**) scheme of the novel pPAO10 vector; significative differences from the original pPAO2 vector are marked with different colors; **D**) in the novel vector 100% of expressed polypeptides carries β-lactamase; **E)** Comparison of the growth of bacteria harboring either pPAO2 or pPAO10 on agar plates containing serial dilutions of ampicillin, when carrying an ORF (represented by a single chain antibody – scFv) or an out-of-frame sequence (stuffer). Bacteria plated on chloramphenicol only (constitutively resistance carried by all plasmids) represent 100% of growth.

SUPPLEMENTARY METHODS

Creation and validation of the pPAO10 vector

The pPAO2 phagemid vector [3] allows selection of ORFs, by expressing polypeptides in fusion with the enzyme β-lactamase (Supplementary Fig. 1A and 1B). In the DH5α *E.coli* strain, when an exogenous DNA fragment is cloned in frame, the TAG *amber* stop codon (which can be translated as a glutamine or recognized as a stop codon) determines the expression of two polypeptides (Supplementary Fig. 1B), one of them in fusion with β-lactamase, resulting in the positive selection of the clone on ampicillin-containing agar plates. The original pPAO2 vector was improved by creating a new version, named pPAO10 (Supplementary Fig 1C and 1D), with three main new features:

i) repositioning of the amber stop codon. The original *amber* stop codon was substituted by a codon encoding for a serine, and re-introduced after the second *lox* site. As a result polypeptides are always expressed in fusion with β-lactamase (Supplementary Fig. 1D) both in the soluble or G3P fused form. The replacement of the stop codon was done by PCR mutagenesis, by amplifying the original pPAO2 polylinker with the primer sense VLPT2 5'-tac cta ttg cct acg gca gcc gct gga ttg tta tta ctc-3', annealing in correspondence of the BssHII site, and the antisense primer G3P-stop 5'-agt tgt ggc ggc cgc cga atg gtg atg gtg agg tg-3', introducing the new stop codon at the 5' of the EagI restriction site. The resulting PCR product was cloned as BssHII-EagI fragment in the backbone of the original pPAO2 vector. The polylinker of this intermediate vector was PCR amplified again with the primer sense VLPT2 and the antisense His-nostop-NotI 5'-agt tgt ggc ggc cgc cga atg gtg atg gtg atg gtg ag-3', with a single A→C mutation in the original stop codon, resulting in a serine codon. This was cloned as BssHII-NotI fragment into the vector carrying the new stop codon at the 3' end of the beta-lactamase gene.

ii) Modification of the polylinker. The FLAG tag at the N-terminus was removed and the polylinker was streamlined; the new polylinker reported in Supplementary Figure 1C carries BssHII and NheI restriction sites for directional cloning, SpeI site for random cloning with SpeI adapters, and StuI site for blunt-ended fragment cloning.

iii) Over-expression of chaperon proteins. Skp/ompH and FkpA have been described as important elements to improve the folding of proteins in the periplasmic space [13, 14]. These genes were cloned downstream of geneIII, reported in green in the scheme. The Skp/OmpH chaperon gene was PCR amplified from *E.coli* genomic DNA, with its own promoter, by using the primer Skp-sense 5'-acg tga att cac gtg tgt tct cca caa agg aat-3', introducing the EcoRI site, and Skp-antisense 5'-agc tga att ctc tag atc ttt att taa cct gtt tca-3', introducing BglII and EcoRI sites, and cloned in the EcoRI site of the vector at the 3' end of the geneIII. The BglII site introduced by the Skp-antisense primer was used for the cloning of the FkpA chaperon

gene, amplified from *E.coli* genomic DNA with primers FkpA-sense 5'-agc tag atc tct gcc cgc act cat ttc gcg gtc atc-3' and Fkp-antisense 5'-ggt cag atc ttt att ttt tag cag aat ctg cg-3', introducing the BglII site. The correct orientation of both Skp and FkpA genes, as well as the confirmation that all the introduced mutations were correct, were assessed by DNA sequencing. The resulting vector is called pPAO10.

The novel pPAO10 and the original pPAO2 vectors were compared in their ability to grow with increasing concentrations of ampicillin, with either an ORF (represented in the example reported here by a single chain antibody fragment – scFv –, a relatively large structure of approximately 300 aminoacids) or a not-ORF (reported here as "stuffer") cloned into the vectors (Supplementary Fig. 1E). The four resulting constructs were transformed into *E.coli* DH5α strain and the bacteria assayed for growth on increasing concentrations of ampicillin (0 to 80 μg/ml). The pPAO10 vector showed a 4-fold increased resistance to ampicillin compared to the pPAO2 vector when carrying an ORF; neither vector conferred resistance to ampicillin with a not-ORF sequence.

Supplementary Figure 2. Focussing of selected clones

Genomic context of LAP3 (upper panel) and MYO18A (lower panel) genes, two of the top ranked ones, showing the gene aligned to the supporting sequences obtained before (NS) and after selection (BIO, SP). Blue bars on top of each panel represent genes, red boxes correspond to exons (RefSeq mRNAs) of the gene, blue lines depict introns. Rectangles in the lanes labeled NS, BIO and SP are colored according to the color code in the legend and show sequencing depth obtained from the three libraries. Regions with high sequencing depth represent enriched fragments, are only observed in the selected libraries (i.e. blue region in the SP selection in the LAP3 panel, and purple region in the BIO selection in the MYO18A panel) and correspond to specific interacting domains. A similar panel can be obtained from www.interactomeataglance.org for all the genes represented in the libraries.

Supplementary Table 1. Identity of the first 50 ranked genes in the NS, BIO and SP libraries.

A) Non selected library, first 50 ranked genes

NS			
hugo	**Reads**	**Depth**	**Focus_index**
MYH9	680	130	0,19
PLEC1	430	52	0,119
TNKS1BP1	316	183	0,576
HSPA1A	232	59	0,25
NUMA1	187	33	0,171
RPL4	175	157	0,891
MYH14	172	55	0,314
EIF4G1	170	122	0,712
FLNA	167	20	0,114
SLC38A10	162	68	0,414
MAP3K13	158	144	0,905
RRBP1	148	34	0,223
H19	145	55	0,372
ACTN4	144	43	0,292
HSD17B10	140	119	0,843
NUCB1	138	51	0,362
COL6A2	135	43	0,311
LAMB2	131	34	0,252
SPTBN1	130	25	0,185
PRDX5	130	101	0,769
LRP1	128	17	0,125
SPTAN1	119	26	0,21
FLOT1	119	66	0,546
DCTN1	112	58	0,509
PPL	110	23	0,2
CHGA	110	48	0,427
TLN1	107	18	0,159
LRPAP1	107	41	0,374
FLNB	102	16	0,147
KRT8	101	32	0,307
RNF40	101	34	0,327
MYO18A	101	36	0,347
MVP	101	43	0,416
LTBP2	99	64	0,636
CLIP2	95	26	0,263
TNIP1	94	43	0,447
PLXNB2	91	25	0,264
ASPH	90	27	0,289
FLNC	87	18	0,195
COL4A2	85	12	0,129
KIF1C	82	35	0,415
ST14	82	62	0,744
SART1	81	27	0,321
LMNA	81	28	0,333
EEF1D	81	33	0,395

HDLBP	80	30	0,363
GSN	79	25	0,304
TPM3	79	71	0,886
EMILIN1	78	20	0,244
ROBLD3	78	71	0,897

B) BIO selected library, first 50 ranked genes

BIO			
hugo	**Reads**	**Depth**	**Focus_index**
ALDOB	4975	4935	0,992
TTC31	1974	1898	0,961
SMOC1	1286	1210	0,94
ZNF23	536	318	0,591
MYO18A	526	463	0,878
MYH9	522	381	0,728
TOP1	511	499	0,975
TAF3	488	460	0,941
EIF2B4	466	420	0,899
RBMX2	440	431	0,977
CARS	386	361	0,933
MAP1A	370	311	0,838
NOP56	364	319	0,874
ANKRD11	354	276	0,777
TCEB3	321	290	0,9
COL12A1	272	254	0,93
TBC1D10B	262	239	0,908
CCDC124	239	222	0,925
PDIA4	230	212	0,917
DNAJC17	228	203	0,886
LRRC59	191	163	0,848
CLU	186	174	0,93
FN1	181	149	0,818
ZNF407	175	164	0,931
H19	164	124	0,75
GIGYF2	157	155	0,981
TRAF3IP1	134	112	0,828
CREB3L1	132	125	0,939
CCDC80	130	116	0,885
ZNF320	127	123	0,961
ABCF1	125	96	0,76
NIPBL	103	87	0,835
UBAP2L	103	93	0,893
PLEC1	100	12	0,11
PPP1R9B	100	89	0,88
PCBD2	99	56	0,556
AP3D1	99	88	0,879
SERTAD2	97	94	0,959
STAU1	95	62	0,642
MLF2	90	84	0,922
EIF5B	85	69	0,8

PITX1	85	70	0,812
RRBP1	83	27	0,313
PBXIP1	82	52	0,622
CCDC55	81	77	0,938
SND1	76	47	0,605
CIZ1	75	60	0,787
GHITM	73	56	0,753
MAP1B	72	31	0,417
PCDHGA1	72	50	0,681

C) SP selected library, first 50 ranked genes

SP			
hugo	**Reads**	**Depth**	**Focus_index**
LAP3	918	826	0,899
GSTO2	850	815	0,958
FN1	685	599	0,873
H19	372	269	0,72
HSPA1A	223	164	0,731
HOXB6	215	197	0,912
SRRT	204	172	0,838
HNRNPUL2	154	107	0,688
MYH9	150	24	0,153
CSRP1	141	113	0,794
PLEC1	137	26	0,182
CDKN1A	129	119	0,915
CDC37	126	102	0,802
DDX21	125	119	0,944
ALDOB	125	122	0,968
CLU	124	110	0,879
TNKS1BP1	122	75	0,607
RAB6B	121	104	0,851
SAFB	120	95	0,783
PRDX4	114	109	0,947
LOC100131484	108	107	0,981
G6PC3	104	101	0,962
UBAP2L	101	90	0,881
GHITM	97	80	0,814
SERTAD2	97	93	0,948
FLNB	96	82	0,844
ZNF23	95	53	0,547
PRDX5	90	80	0,878
COL6A2	89	21	0,225
TAF15	89	42	0,461
TBC1D9B	88	54	0,602
SPTBN1	87	44	0,494
ISYNA1	87	68	0,77
CHGB	86	49	0,558
ACACA	85	84	0,976
EIF4G1	84	69	0,81
STOML2	83	76	0,904

FLNA	82	29	0,341
POLR2L	82	71	0,854
LNX1	74	27	0,351
COL4A2	73	29	0,384
FBRSL1	73	55	0,74
MMP14	71	34	0,465
EEF2	70	34	0,471
CCDC124	70	45	0,629
UBXN1	69	59	0,841
ATAD3B	67	32	0,463
AZI1	67	65	0,955
PCDHGA1	65	31	0,462
ABCA2	65	47	0,708

Top 50 genes for each library. A) NS; B) BIO; C) SP. Genes are ranked according the number of supporting reads. Depth and focus indexes are also shown.