Protein interaction discovery using parallel analysis of translated ORFs (PLATO)

Jian Zhu[1–3,10], H Benjamin Larman[1–5,10], Geng Gao[1,3], Romel Somwar[6], Zijuan Zhang[7], Uri Laserson[2,4,8], Alberto Ciccia[1–3], Natalya Pavlova[1–3], George Church[2,4], Wei Zhang[7], Santosh Kesari[9] & Stephen J Elledge[1–3]

Identifying physical interactions between proteins and other molecules is a critical aspect of biological analysis. Here we describe PLATO, an *in vitro* method for mapping such interactions by affinity enrichment of a library of full-length open reading frames displayed on ribosomes, followed by massively parallel analysis using DNA sequencing. We demonstrate the broad utility of the method for human proteins by identifying known and previously unidentified interacting partners of LYN kinase, patient autoantibodies, and the small-molecules gefitinib and dasatinib.

Several methods have been developed to characterize the specificities of protein-binding molecules. Display technologies that utilize cDNA libraries are often limited by the small fraction of in-frame polypeptides and the highly skewed clonal abundances that reflect differences in gene expression[1]. Two-hybrid and split-reporter techniques[2] are limited to analyses of bait molecules that can be presented within the cell, and thus are not suitable for drug or antibody target identification. More recently, protein microarrays have been used for these purposes[3], but their construction typically requires individual proteins to be purified and arrayed, resulting in substantial costs and various degrees of protein denaturation.

To address these limitations, we developed PLATO (parallel analysis of translated open reading frames (ORFs)), a method that combines *in vitro* display of full-length proteins with analysis by high-throughput DNA sequencing. We demonstrate the utility of PLATO by performing diverse interaction screens against the human ORFeome, a normalized collection of 15,483 human cDNAs in the Gateway cloning system[4]. To express an ORF library *in vitro*, we used ribosome display, a technique used to prepare a library of mRNA molecules that, lacking stop codons, remain tethered to the proteins they encode[5].

Ribosome display imposes minimal constraints on the length or composition of proteins that can be efficiently displayed.
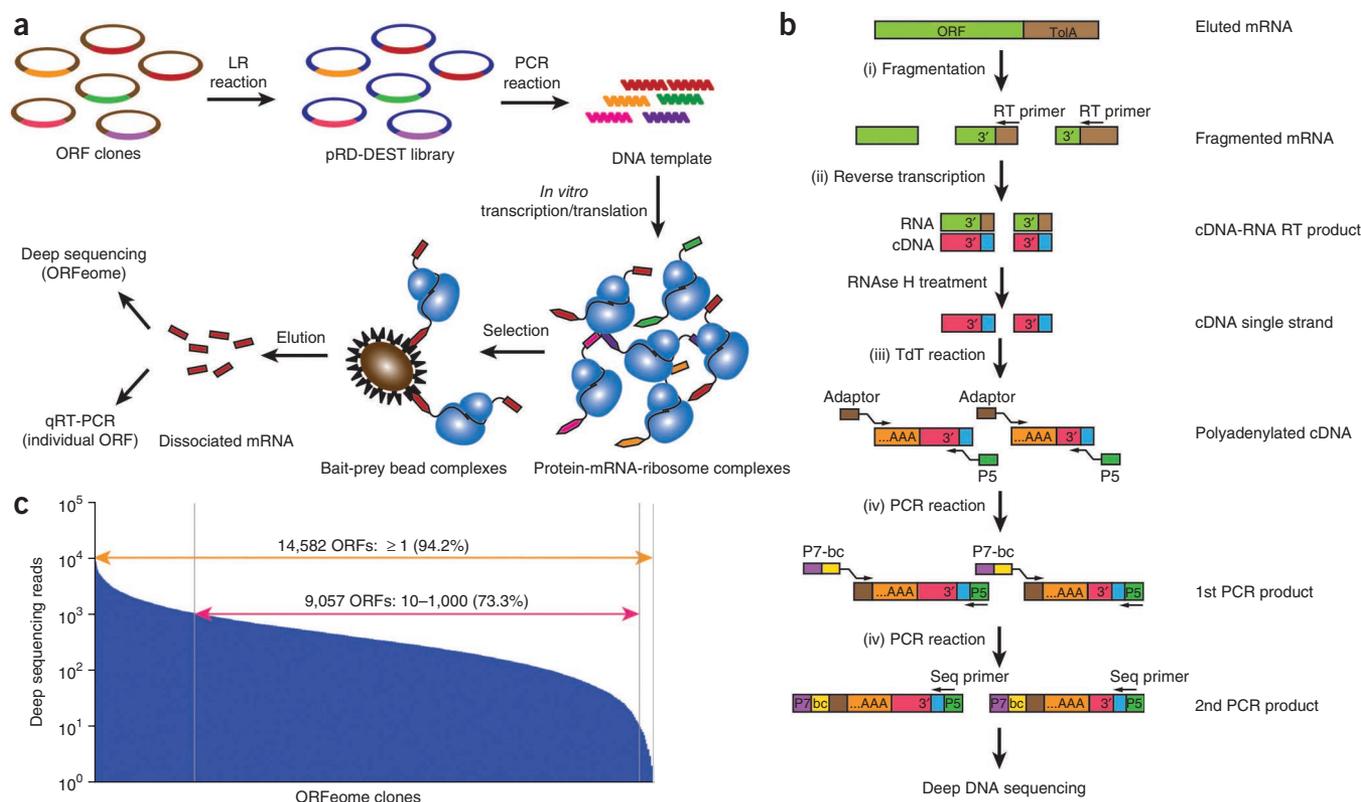
We constructed a ribosome display vector (pRD-DEST; **Supplementary Fig. 1** and **Supplementary Methods**), compatible with Gateway cloning, to be used as a recipient for a concentration-normalized pool of ORF clones. After recombination, DNA was amplified by PCR, yielding linear templates lacking stop codons. Following *in vitro* transcription and translation, the ribosome-displayed ORFeome was screened for binding to immobilized bait(s). Enrichment of candidate binding proteins can be rapidly assessed using quantitative real-time PCR (qPCR) with ORF-specific primers; alternatively, unbiased discovery of interactions can be achieved by deep sequencing of the enriched mRNAs (**Fig. 1a**). Sequencing libraries can additionally be highly multiplexed, thereby reducing the cost of each screen. All steps required for PLATO are compatible with automation using standard liquid-handling robotics.

Our strategy for deep sequencing of enriched display libraries employs recovery of the ORF 3′ termini, which minimizes interference from RNA degradation and ensures stoichiometric correlation between tag counts and transcript abundance. To this end, we adopted the following protocol: (i) chemically fragment enriched-mRNAs; (ii) reverse-transcribe fragments using a common primer; (iii) poly-adenylate cDNAs; (iv) add sample barcodes and sequencing adaptors using two-stage PCR amplification (**Fig. 1b**). Subsequent, multiplex, deep-sequencing analysis of pooled display libraries is reproducible and quantitative (**Supplementary Fig. 2**). We sequenced a sample of the unenriched human ORFeome display library mRNA (input) and detected the transcripts of 14,582 unique ORFs out of 15,483 total cDNAs in the entry clone library (94%; **Fig. 1c**).

To test the ability of PLATO to identify protein-protein interactions, we used LYN kinase, which contains common structural components of the SRC family, including SH3, SH2 and kinase domains[6], and has been extensively characterized for its interaction partners. After affinity enrichment of the human ORFeome using GST-LYN, GST alone or an unrelated GST-fused protein (GST-MUTED), we used Illumina sequencing to identify proteins specifically bound by GST-LYN (**Fig. 2a**, **Supplementary Table 1** and **Supplementary Fig. 3a**). A number of established LYN binding partners were among those identified, and we validated two by qPCR (**Fig. 2b**)[7,8]. We ranked candidate LYN interactors by their degree of enrichment on GST-LYN, and confirmed five of seven tested by western blot analysis (**Fig. 2c**). Of the two candidates not validated, one bound nonspecifically to GST, whereas the other was a true negative. Among the highly enriched ORFs, SH2 domain–containing proteins were over-represented ($P < 0.01$, Fisher's test). Consistent with

Figure 1 Parallel analysis of translated ORFs (PLATO). (**a**) ORF display scheme. The pooled human ORFeome v5.1 entry vector library is cloned by means of attL-attR (LR) recombination into the pRD-DEST expression vector. Expression plasmids are PCR amplified to generate the DNA templates for *in vitro* transcription. After *in vitro* translation, the protein-mRNA-ribosome complexes are incubated with protein, antibody or small-molecule bait immobilized on beads. The enriched mRNA library is recovered from bait-prey bead complexes for further analysis. (**b**) Processing of mRNA samples for deep DNA sequencing. After fragmentation and reverse transcription (RT) using a universal primer to recover the 3′ end of ORFeome transcripts, cDNA is polyadenylated with terminal deoxynucleotide transferase (TdT) and amplified for multiplex deep sequencing using primers containing a sample barcode (bc) and the P5 and P7 Illumina sequencing adaptors. (**c**) Sequencing reads of the unenriched human pRD-ORFeome mRNA library (the 'input' library). Most ORFs were sequenced at least once.

a role for LYN autophosphorylation in mediating interactions with SH2 domains, phosphatase treatment of immobilized GST-LYN abolished binding of SH2D1A and SH2D4A, but only partly diminished PIK3R3 binding, suggesting the presence of an additional interaction domain besides SH2 (**Supplementary Fig. 3b**). These proteins have not previously been reported to interact with LYN.

We next asked whether PLATO could be used to identify protein targets of antibodies from patients with autoimmune disease. We first examined target enrichment using affinity-purified P53 and PDCD4 antibodies immobilized on protein A/G beads for library immunoprecipitation. As shown by qPCR, P53 and PDCD4 transcripts were robustly enriched by their cognate antibodies, but not by control antibodies (**Supplementary Fig. 4**).

In previous work, we synthesized an oligonucleotide library encoding a 36-residue overlapping human peptidome for display on bacteriophage T7 (T7-Pep). Deep sequencing of affinity-enriched T7-Pep using autoimmune cerebrospinal fluid from three individuals with paraneoplastic neurological disorder (PND) uncovered known and novel autoantigens[9]. We screened these samples using PLATO. Unlike T7-Pep, the human ORFeome is an incomplete collection of full-length proteins, and our findings reflect the inherent complementarity of these libraries. For example, neuro-oncological ventral antigen 1 (NOVA1) is absent from the human ORFeome v5.1, and so PLATO did not detect this known autoreactivity in patient A, whereas it was robustly identified with T7-Pep. Conversely, PLATO identified numerous autoantigens

for each patient that were missed in our peptidome screens (**Supplementary Table 2**). For example, PLATO analysis of patients A and B revealed immunoreactivity with known cancer autoantigens not detected with T7-Pep. Several of these reactive antigens were confirmed by means of immunoprecipitation and western blot analysis (**Fig. 2d** and **Supplementary Fig. 5a–d**). In addition, we had previously established that antibodies from patient C recognized the tripartite motif containing proteins TRIM9 and TRIM67. PLATO considerably expanded the members of the TRIM family recognized by antibodies in this patient's cerebrospinal fluid to include TRIM1/MID2, TRIM18/MID1, TRIM54 and TRIM55 (**Fig. 2e**). Notably, multiple sequence alignment results in tight clustering of this precise subset of the extended TRIM family, suggesting the presence of shared, conformational epitopes not represented in T7-Pep[10]. As an alternative PLATO readout, hybridization of autoantibody-enriched libraries to custom oligonucleotide microarrays revealed a similar list of autoantigens (**Supplementary Fig. 6**).

Discovering the targets of small molecules typically involves the use of cell extracts containing a wide distribution of protein abundances. Analysis by mass spectrometry is thus biased toward highly expressed proteins. Normalized ORF libraries and quantitative DNA sequencing might therefore offer greater power to detect interactions of proteins with small molecules. We tested this idea with gefitinib (Iressa), an inhibitor of epidermal growth factor receptor's (EGFR's) tyrosine kinase domain. Gefitinib interacts with the ATP-binding pocket of
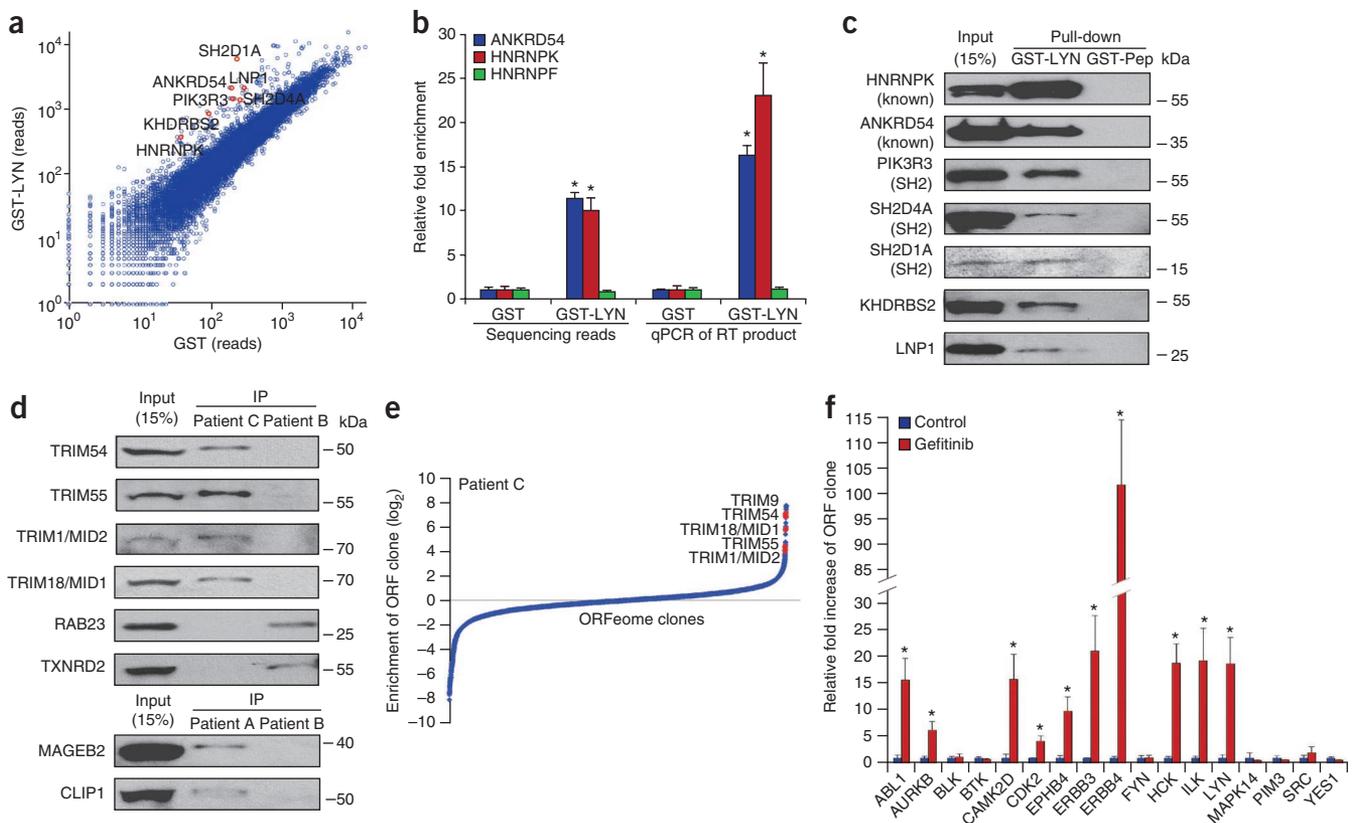
Figure 2 Identification of known and previously undescribed interactions using PLATO. (**a**) Interactions with LYN tyrosine-protein kinase. Scatter plot of each ORF's sequencing reads after enrichment on GST-LYN or GST. Several known and undescribed LYN binding candidates are highlighted in red. (**b**) Enrichment of two known interactors of LYN (HNRNPF was included as a negative control). Data were normalized to the GST-enriched libraries ($n = 3$, mean ± s.d.; *, $P < 0.01$; t-test). (**c**) Confirmation of known and predicted LYN binding proteins by affinity precipitation–western blot analysis of lysates from HEK293T cells transiently overexpressing the individual V5-His-tagged candidate proteins. (**d**) Interactions with autoantibodies. Confirmation of previously unidentified autoantigens from PND patients. (**e**) Enrichment ranking of PND autoantigens identified using cerebrospinal fluid from patient C. (**f**) Interactions with a small molecule. Enrichment of previously identified targets of gefitinib. Data were normalized to the control-enriched libraries ($n = 3$, mean ± s.d.; *, $P < 0.05$; t-test). Primers used for qPCR measurement are listed in **Supplementary Table 4**.

EGFR and additional tyrosine kinases[11]. Analysis after ORFeome affinity enrichment on gefitinib-coupled beads revealed significant ($P < 0.05$) enrichment of 10 out of the 17 predicted targets tested (**Fig. 2f**). This experiment demonstrates the relative ease by which candidate protein interactions can be assayed with PLATO; the binding of any ORF can be rapidly assessed using qPCR without the need for cloning or western blot analysis. ORFeome libraries affinity-enriched by the SRC family tyrosine kinase inhibitor dasatinib (Sprycel) exhibited overrepresentation of protein kinases (9 out of 75; $P = 0.0003$; Fisher's test), including the known target LCK and several targets not previously associated with this compound (**Supplementary Table 3**).

PLATO's limitations include incomplete ORFeome collections and a lack of protein post-translational modifications. However, the quality, completeness and availability of these libraries will continue to improve over time. In addition, very large ORF proteins may be displayed with low efficiency and proteins containing membrane-spanning or aggregation-prone domains that normally require host cellular machinery for proper folding may aggregate; these factors may complicate data analysis. Finally, ribosome display imposes certain limitations on the conditions under which affinity enrichments can be done (e.g., low temperature and absence of RNase contamination are essential), and using proteins containing nucleic acid–binding domains as baits may result in nonspecific binding. When the required conditions for PLATO are met, however, this method provides

three advantages as a tool for proteomic investigations. First, protein size and composition should minimally affect display efficiency. Second, its cost and instrument requirements are low. Finally, the rapidly declining cost of DNA sequencing will make PLATO an ideal platform for projects involving large numbers of samples, such as cohort-scale autoantibody profiling or structure-activity relationship analyses of small-molecule compounds.

Note: Supplementary information is available in the online version of the paper.

AUTHOR CONTRIBUTIONS
S.J.E. and H.B.L. conceived and supervised the project. pRD human ORFeome library was constructed by J.Z., and characterized by J.Z. and H.B.L. The PLATO protocol was developed by H.B.L. and J.Z. Clinical evaluations and patient sample acquisitions were performed by S.K. Statistical analysis was performed

1. Faix, P.H. *et al. Biotechniques* **36**, 1018–1022, 1024, 1026–1029 (2004).
2. Zhu, H., Bilgin, M. & Snyder, M. *Annu. Rev. Biochem.* **72**, 783–812 (2003).
3. Jeong, J.S. *et al. Mol. Cell Proteomics* **11**, O111.016253 (2012).
4. Lamesch, P. *et al. Genomics* **89**, 307–315 (2007).
5. Amstutz, P., Binz, H.K., Zahnd, C. & Plückthun,, A. Ribosome display: *in vitro* selection of protein-protein interactions in *Cell Biology—A Laboratory Handbook* (ed. Celis, J.) Vol. 1, 3rd Ed., 497–509 (Elsevier Academic Press, 2006).
6. Boggon, T.J. & Eck, M.J. *Oncogene* **23**, 7918–7927 (2004).
7. Weng, Z. *et al. Mol. Cell Biol.* **14**, 4509–4521 (1994).
8. Samuels, A.L., Klinken, S.P. & Ingley, E. *Blood* **113**, 3845–3856 (2009).
9. Larman, H.B. *et al. Nat. Biotechnol.* **29**, 535–541 (2011).
10. Carthagena, L. *et al. PLoS ONE* **4**, e4894 (2009).
11. Brehmer, D. *et al. Cancer Res.* **65**, 379–382 (2005).

npg