Generation of a fluorescently labeled endogenous protein library in living human cells

Alex Sigal[1,4], Tamar Danon[1], Ariel Cohen[1], Ron Milo[2], Naama Geva-Zatorsky[1], Gila Lustig[3], Yuvalal Liron[1], Uri Alon[1] & Natalie Perzov[1]

[1]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel. [4]Present address: Division of Biology, California Institute of Technology, Pasadena, California 91125, USA. Correspondence should be addressed to A.S. (sigal@caltech.edu).

We present a protocol to tag proteins expressed from their endogenous chromosomal locations in individual mammalian cells using central dogma tagging. The protocol can be used to build libraries of cell clones, each expressing one endogenous protein tagged with a fluorophore such as the yellow fluorescent protein. Each round of library generation produces 100–200 cell clones and takes about 1 month. The protocol integrates procedures for high-throughput single-cell cloning using flow cytometry, high-throughput cDNA generation and 3′ rapid amplification of cDNA ends, semi-automatic protein localization screening using fluorescent microscopy and freezing cells in 96-well format.

INTRODUCTION

A quantitative understanding of human protein networks requires the measurement of endogenous protein dynamics in living cells[1]. An ideal measurement system would: (a) work at the protein level, because the regulation of translation, localization and degradation is crucial in mammalian cells, (b) work at the level of individual cells, because experiments that average over cell populations can miss events that occur in only a subset of cells. Furthermore, averaging can miss all-or-none effects, and cell–cell variability. (c) Follow cells over extended periods of time to reveal phenomena such as oscillations and temporal programs and (d) make minimal perturbations to the state of the cells.

Current approaches to proteomics, such as mass spectroscopy and proteome chips, have revolutionized our ability to assay snapshots of the protein content of cells[2–8]. These methods can assay protein modifications, and can be applied to a variety of samples. At present, these methods usually average over many cells, and do not allow quantification of dynamics in individual cells. There have also been advances in high-throughput quantification of protein levels and localizations at the single-cell level using antibody staining and microscopy[9,10]. As staining of internal proteins requires the killing of the cell, it is not possible to follow protein dynamics in the same cell over time. A dynamic proteomics method in individual cells can complement antibody and mass spectrometry-based approaches.

Dynamic measurements in living cells are made possible by the use of fluorescent proteins as genetic tags. Labeling with fluorescent tags often leaves the wild-type localization intact (see for example, ref. 11). A library of cells containing GFP-labeled cDNAs, expressed under an exogenous promoter, has been created to investigate protein localization on the scale of the proteome[12,13]. A disadvantage of this approach is that exogenous expression gives no information about the transcriptional regulation of the gene, and potentially leads to non-physiological levels of expression. To follow wild-type regulation, homologous recombination can be used to integrate sequences of fluorescent proteins into the genome at the wild-type locus. This approach was made high throughput in yeast[14]. High-throughput homologous recombination is also being developed in mouse embryonic stem (ES) cells in the KOMP, EUCOMM and NorCOMM initiatives. To the best of our knowledge, high-throughput homologous recombination has not been achieved in human cells.

Here, we present a protocol for a tagging approach that labels proteins in their native chromosomal locations without the need for homologous recombination. It is based on a tagging method known as central dogma (CD) tagging[15–18]. CD tagging labels genes by integrating a DNA sequence coding for a fluorescent tag into the genome. The tag is inserted in a non-directed manner using a retrovirus (Fig. 1). It is marked as an exon by flanking splice acceptor and donor sequences. If the tag integrates within an
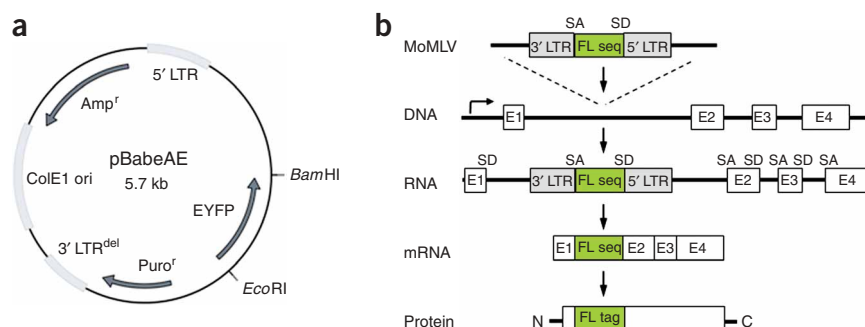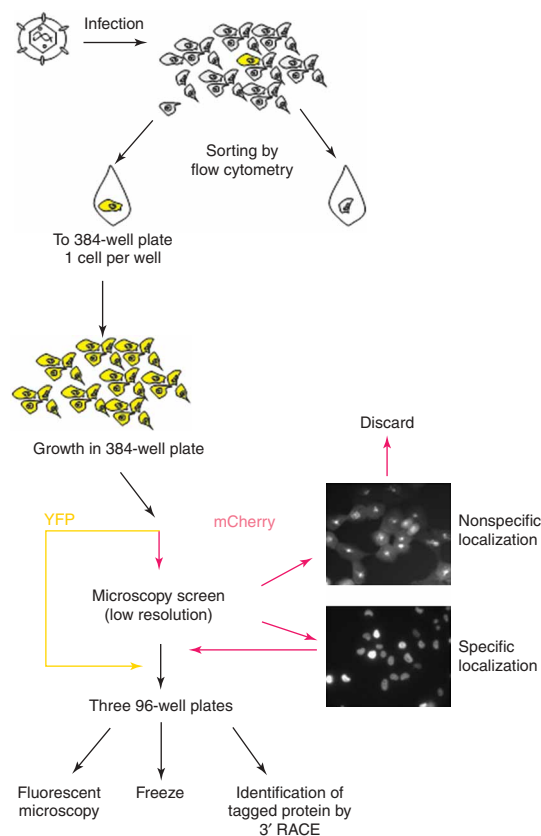

Figure 1 | Generation of a CD-tagged protein library. (**a**) The CD-tagging vector pBabeAE. (**b**) Steps of successful tagging. Splice acceptor (SA) and splice donor (SD) flanked fluorophore sequence (FL seq), with no promoter, no start codon or polyA signal, is inserted into the genome by MoMLV. If the virus inserts into an actively transcribed gene, the fluorophore sequence is retained as a new exon after splicing of the mRNA. Owing to the large size of the first intron and viral preference for integration sites near the start of genes, the first intron is the most common point of insertion. The tagged mRNA translates to an internally labeled protein, with the fluorophore protein tag (FL tag) usually near the N terminus.

Figure 2 | Flowchart of the library generation procedure. Cells that fell into the sorting gate were sorted at one cell per well into 384-well plates and expanded into clones. If the CD tag was YFP, cells were passaged from the 384-well plate into two ordinary tissue culture 96-well plates and one optical 96-well plate. The optical 96-well plate was used to image the proteins tagged, whereas the regular plates were used for 3′ RACE or freezing. If the CD tag was mCherry, an extra screening step was introduced at the 384-well stage to screen out clones that showed a localization signature associated with a low probability of the tagged mRNA to correspond to a known gene, or a new gene with supporting evidence (exon–intron structure, etc.). Modified from ref. 18.

expressed gene, it is then spliced into the gene's mRNA and a fusion protein is translated[18]. The identity of the labeled gene is then determined by rapid amplification of cDNA end (RACE).

The approach can be used in cells permissive for retroviral infection, and we have successfully used this approach on murine NIH 3T3 and C2C12 myoblast cells and H1299 human lung carcinoma cells. It is expected to work in other cell types permissive for retroviral infection provided a viral receptor is present. There is no particular reason why this approach should not work in ES cells, as gene trap approaches (see below) successfully used retroviral vectors to genomically tag mouse ES cells[19]. We did not select ES cells as the cell line for the library, as our primary aim was to image cells and quantify protein behavior in individual cells, and cells that grow in colonies like ES cells pose a challenge for such an approach.

CD tagging is similar in its mechanism but not in its aim or final result to gene trapping approaches. These use random reporter insertions to track transcriptional kinetics[20,21]. Gene trapping inserts a splice acceptor followed by a promoterless reporter and a polyA termination signal. This is usually followed by a selection marker under its own promoter. The tagged protein is truncated by the polyA signal, and its expression depends on the endogenous promoter of the gene into which the reporter is integrated. The protein is therefore mutated on purpose, and this is usually carried out in mouse ES cells with the final aim of generating mice with two mutated alleles. In contrast, the goal of CD tagging is to preserve the full length of the tagged protein. The CD tag does not truncate the protein, but instead integrates as a new exon (**Fig. 1**).

CD tagging was previously used by several groups to tag genes with GFP in mouse and *Drosophila* cells[16,22,23]. We used yellow fluorescent protein (YFP) and the red fluorophore mCherry[24], integrated into the genome by the Moloney murine leukemia retrovirus (MoMLV) to CD-tag human cells[18]. We found that our CD-tagging strategy resulted in most proteins tagged near the N terminus[18,25]. This bias for the N terminus is probably due to the generally large size of the first intron, and the known preference of MoMLV to integrate near the start of genes[26]. No C-terminal

integration bias was observed, indicating that instability of centrally labeled proteins may not be the reason for their underrepresentation. About two-thirds of CD-tagged proteins showed localizations, which matched the localizations reported in other studies[18].

The present protocol was used to build our current library (can be viewed at http://www.dynamicproteomics.net/). Our main library, designated d7, currently contains about 700 different proteins. The aim was to make CD tagging a high-throughput method to tag a significant fraction of the proteins expressed by a cell (**Fig. 2**). Combined with image analysis techniques[18,25], the CD-tagging approach may become a powerful tool to investigate the dynamics of the proteome in individual living cells. Each cycle of library generation yields about 100 different tagged proteins, takes about 1 month (see **Table 1** for detailed timing information) and involves retroviral infection with the tag sequence, cloning positive cells by flow cytometry, tagged protein identification by 3′ RACE, recording of the localizations of CD-tagged proteins and clone freezing.

MATERIALS
REAGENTS
· pBabeAE CD-tagging vector (available from us in all reading frames for YFP, and frame 0 for mCherry) or other CD-tagging vector
· Cell line expressing the MoMLV ecotropic receptor to enable murine retrovirus entry, and culture medium for these cells. Pseudotyping the virus with the VSVg envelope protein should also work
· Phoenix virus packaging cells (developed by Gary Nolan and available from the ATCC)
· Culture medium for packaging cells: DMEM with 2 mM L-glutamine, supplemented with 10% (v/v) fetal bovine serum (FBS), 100 U ml^{-1} penicillin/streptomycin

· FuGene 6 transfection reagent (Invitrogen)
· Polybrene (AL-118, Sigma-Aldrich)
· Optical medium with no phenol red or riboflavin (Biological Industries, or an alternative supplier such as Sigma-Aldrich), supplemented for normal cell growth
· ZR-96 Mini RNA isolation I Kit (Zymo Research)
· Omniscript RT kit (Qiagen)
· QIAquick PCR purification and gel extraction kits (Qiagen)
· Primers for first-strand cDNA synthesis and gene identification by 3′ RACE (see **Table 2**). First-strand adaptor primer can either be purchased (Invitrogen) or synthesized and cleaned to high purity with SDS-PAGE (Sigma Genosys)

TABLE 1 | Timing information.

Day	Procedure	Time	Work involved	Work intensity
Day 1	Plating virus packaging cells	30 min	Tissue culture	Low
Day 2-3	Virus production, plating target cells	48 h	Tissue culture, incubations	Low
Day 4-6	Infection and incubation	72 h	Incubations	Low
Day 7	Plating of infected cells on 15 cm dishes	30 min	Tissue culture	Low
Day 8	Preparation for sorting	1 h	Tissue culture	High
Day 8	Sort	3-4 h	Flow cytometry	High
Day 8-20 (approximate[a])	Clonal expansion from single cells	10-14 days	Incubations with occasional tissue culture	Low
Day 16/23 (384/96 plate) (approximate[a])	Microscopy screen of 384- or 96- well plates	1-2 h per plate	Semi-automated microscopy	High
Day 25 (approximate[a])	Freezing	30 min per plate	Tissue culture	High
Flexible	Thawing	30 min per plate	Tissue culture	High
Day 25 (approximate[a])	Isolation of total RNA from 96-well plates	60 min per plate	Molecular biology	Medium-High
Day 26 (approximate[a])	First-strand cDNA synthesis	2 h per plate	Molecular biology	Medium-High
Day 27 (approximate[a])	PCR and nested PCR	1 day	Molecular biology	Medium

[a]Depends on cell growth.

- Accudrop beads (BD Biosciences) or ordinary beads such as FluoSpheres yellow-green beads (Molecular Probes) for fluorescence-activated cell sorting (FACS) calibration
- RNaseZap (Ambion)
- Ready-Mix PCR mix (9597; Bio-Lab)
- RNaseOUT diluted to 10 U μl^{-1} in water (10777-019, Invitrogen)
- Omniscript RT (205111, Qiagen)
- Trypsin 0.05% solution (e.g., 03-053-1, Biological Industries)

EQUIPMENT
- 32 °C incubator
- Cooled centrifuge with plate holders (Eppendorf 5810R)
- Becton Dickenson FACS Vantage machine (BD Biosciences) equipped with Cloncyt motorized stage control software, or an equivalent system such as the Becton Dickenson FACS Aria
- 50 μm nylon mesh
- Sterile FACS tubes with lids (Falcon)
- 10 ml syringes
- Extension tube with luer fitting to syringe (MG918485, Multimedical)
- DMIRE2 inverted fluorescence microscope (Leica) or equivalent
- Plate holder (e.g., Martzhaeuser 96-well plate holder with top screw removed to enable CO_2 cover placement)
- ×20, ×40 or ×63 Plan Apochromat air objectives, highest possible numerical aperture (usually about 0.7 for ×20, 0.85 for ×40–63)
- Filter cubes for YFP (Chroma set 41028) or mCherry (Chroma set 41043)
- Motorized stage (e.g., Martzhaeuser)
- Automated shutters such as Uniblitz (Vincent Associates)
- Cooled CCD camera (some options: ORCA ER by Hamamatsu Photonics, Retiga SRV by Qimaging, Photometrics CoolSnapHQ by Roper, iXon by Andor, although we have not come across an ImagePro driver for the last

option). It is important to test the camera and its ability to be controlled by the imaging software before buying
- Microscope-mounted temperature-controlled incubator, with an internal chamber for regulated CO_2 and humidity (PeCon). Alternatively, replacement of normal bicarbonate-buffered medium with medium buffered by HEPES will keep unincubated cells alive for several hours
- C-mount to couple camera to microscope
- A PC computer with a windows XP or 2000 operating system and sufficient serial (RS232), USB2 and firewire ports to control the various hardware and enable data transfer. See EQUIPMENT SETUP
- ImagePro 5.1 and Scope Pro (Media Cybernetics) or equivalent
- ScanPlate macro (available from our laboratory and adapted to ImagePro5.1)
- 384-well plastic optical plates (781091, Greiner Bio-One)
- 96-well optical CVG plate (164588, Nunc)
- 24- and 6-well plates and 6, 10 and 15 cm dishes for tissue culture
- Cooled centrifuge for RNA extraction and PCR cleaning in 96-well format (Sigma-Aldrich)
- PCR machine (e.g., Biometra)
- Thermo-Fast non-skirted 96-well PCR plate (AB-0600, Abgene)
- Cryo 1 °C freezing containers (Nalgene)
- Pads for liquid absorption

Note that lower-throughput applications can be performed without much of the specialized kits and equipment. Infection can be performed without spinning (see alternative protocol in PROCEDURE), and RNA extraction can be performed with standard kits (e.g., RNeasy by Qiagen)

EQUIPMENT SETUP
Filtration of cells for sort The filtering syringe can be assembled by discarding the needle if present, folding a small piece of 50 μm mesh in two, then placing it over the syringe opening and securing tightly with the luer fitting of the

TABLE 2 | Oligonucleotide primers used for 3′ RACE and sequencing.

Primer name	Use	Sequence	Alignment in YFP or mCherry
AP first-strand	First-strand cDNA synthesis	GGCCACGCGTCGACTAGTAC(T)$_{17}$	
AP 92	RACE first and nested reaction 3′ primer	GGCCACGCGTCGACTAGTAC	
YFP 90	RACE first reaction 5′ primer for YFP-tagged genes	GCAGAAGAACGGCATCAAGG	Bases 471–490
YFP 85	RACE-nested reaction 5′ primer for YFP-tagged genes	CGCGATCACATGGTCCTGCTG	Bases 646–666
Cherry 45	RACE first reaction 5′ primer for mCherry-tagged genes	GTGGTGACCGTGACCCAGGA	Bases 322–341
Cherry 46	RACE-nested reaction 5′ primer for mCherry-tagged genes	GCGGATGTACCCCGAGGACG	Bases 456–475
Cherry 56	Sequencing of mCherry RACE product	GACTACACCATCGTGGAACA	Bases 586–605
YFP 906	Sequencing of YFP RACE product	GGATCACTCTCGGCATGGAC	Bases 686–705

extension tube. The extension tube should then be cut so that 2–3 cm remains attached to the syringe. To sterilize, fill and leave overnight in 70% EtOH. Before use, empty off ethanol, then rinse three times in different tubes of PBS without calcium or magnesium.

Flow cytometry basic setup   These instructions are given for a FACSVantage cell sorter with a tunable argon laser. They may need to be modified for other sorter types. The sorter must be presterilized with bleach, then 70% EtOH. For sorting of YFP- or GFP-positive cells, use the 488 nm band of an argon laser (see **Fig. 3a** for filter configuration). For sorting of mCherry-positive cells, tune the argon laser to 514 nm (see **Fig. 3b** for filter configuration). Check that the observation point is not obscured. Check that side streams are not split and adjust vibration phase if they are. Calibrate drop delay using Accudrop, if available: select a left sorting gate that includes all Accudrop beads and determine the drop delay that deflects the most beads to the left (in FACS Vantage, use the "Enrich" sorting mode). If the sorter is not equipped with Accudrop, use ordinary beads and choose a sorting gate to include all beads. Deflect 20 beads to the left and onto a slide at various drop delays. Examine the slide by fluorescence microscopy and choose the drop delay with the most beads deflected. After this calibration, sterilize the sample fluid path by running through it 1% bleach (a 1:3 dilution of off-the-shelf bleach) for 3–5 min, followed by 70% EtOH for 3–5 min to wash away the bleach, then sterile PBS for 3–5 min to wash away the EtOH. Between each fluid change, place the machine on run mode and let several drops pass into the waste to remove residual bleach or alcohol.

Calibration of sorting streams before sort   Even if the sorter is deflecting the correct drops, it may not deflect them into the correct well. This may lead to two positive cells falling into the same well, and apparent behavior such as bistability would actually be an artifact of two cell clones mixed together. We recommend to sort cells into every second well, since this greatly reduces the chance that two cells will be deflected to the same well, since the deviations in stream angle for this to occur would have to be much larger. To determine how the actual streams will behave during the sort, test run the sort by using the cells to be sorted (not beads) and programming the sorter and mechanical stage to sort 100 cells every fourth well on a covered 384-well plate. This will enable you to see the locations where cells are deposited. Sort for 3–4 rows (do not let the drops dry on the cover by sorting too long). Remove the plate and check the location of the sorted drops on the cover (**Fig. 3c**). The possible outcomes and responses are as follows: (i) Most drops land in the appointed well and no visible drops land more than one well away. Response: start sort. (ii) Most drops land in a localized area but not in the designated wells. Response: adjust stream angle until the drops land in the designated wells, and start sort. (iii) Drops fall in a smear that spans less than three wells. Response: smear indicates a split in the streams. If sorting is into every second well, some cell loss will occur. Either accept the loss and start sort, or adjust vibration frequency to unify streams. (iv) Drop smear spans more than three wells. Response: this type of severe stream split may lead to two cells sorted into the same well, even with a one well separation between target wells. Unify streams by adjusting vibration frequency. If this is unsuccessful, there may be a blockage in the fluidics, which requires clearing.

Microscopy: setup overview   This section describes high-throughput screening using our software. The setup for high-throughput microscopy can

be assembled using off-the-shelf components from microscope, camera, stage and shutter manufacturers, together with commercially available microscope control software such as ImagePro or Metamorph. We use a semi-automatic approach, which we wrote as a macro in ImagePro, where the user manually focuses on the cells, while the screening program enables one-step image acquisition and storage with an automatically derived name that includes the well number.

Microscope computer   A PC computer with a windows XP or 2000 operating system and sufficient serial (RS232), USB2 and firewire ports to control the various hardware and enable data transfer is required. We recommend buying a powerful computer, with a fast processor, high rapid access memory and a large hard drive. If our software is used, the default storage location is disk D (see "Configuring our ScanPlate program to run correctly on a new microscope setup" below for how to change this). If the number of serial ports is insufficient, a PCI card with 2–8 serial ports can be installed (Digi International), or USB-to-serial converters (Prolific Technology) can be used.

Control of microscope, motorized stage and shutters from ImagePro   To be controlled by the ImagePro software, components need to be connected to the computer and located for ImagePro. Microscope, stage and shutters generally connect to the computer by serial (RS232) nine-pin connections. A successfully installed serial port should show up in the Device Manager (located in Start menu/Settings/Control Panel/System/Hardware/Device Manager). The next step is to enable the ScopePro component of ImagePro to locate the devices. For this, open ScopePro (StagePro for stage) and configure the ports (named communication ports or COMM ports). It may be hard to tell which COMM ports correspond to which physical serial connections on the computer. If a Prolific USB-to-serial adaptor is used, the COMM port it represents will disappear from the device manager if the device attached to it is unplugged. For serial ports from a PCI card, we use trial and error to define the correct port. The data transfer rates should also be set correctly. Our settings are as follows: microscope: 19,200 baud; stage and shutters: 9,600 baud. For all devices, stop bits = 1; parity: none. Once you have located the devices for ScopePro or StagePro, they become controllable by these programs.

Controlling the camera from ImagePro   Inducing ImagePro to communicate with the camera takes a different course. For the Hamamatsu and Qimaging cameras, data is transferred to the computer by firewire (IEEE 1394). The Roper camera uses a specialized PCI card. It is best to test whether the camera is functioning by using the acquisition software that comes with the camera. If the camera works, install the camera driver for the appropriate version of ImagePro. We suggest testing camera compatibility with the program before buying the camera.

Configuring our ScanPlate program to run correctly on a new microscope setup   Once the microscope system is set up, the ScopePro designations of the components may not match our designations in our semi-automatic scanning program "ScanPlate". This is especially likely for the shutters. One example for this kind of problem is that pressing the button for transmitted light ("Phase" in ScanPlate) has no effect, or causes the fluorescence shutter to open. To solve, either switch the shutter cables, or find the correct shutter designation in your setup and paste it into ScanPlate. For this, use the "Record Macro" feature in the macros menu of ImagePro. Name the macro and start recording. Change the
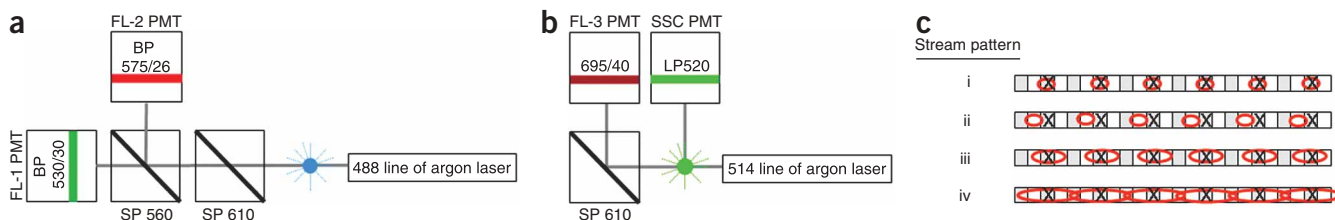

Figure 3 | Optical configurations and stream alignments for YFP- and mCherry-based cell sorting. Simplified configurations of the essential laser lines, filters and dichroic mirrors in nm for YFP (**a**) and mCherry (**b**) are shown. BP: bandpass, where number after / indicates bandpass width; LP: longpass; SP: shortpass; PMT: detector (photomultiplier tube); SSC: side scatter; FL-1,2,3: usual designations for the PMTs at the indicated positions. Blue or green circle indicates observation point. Note that our SSC LP 520 filter gives a low but detectable signal from the 514 nm argon line, but a filter that has transmission at 514 nm is more suitable. (**c**) Locations of cells deposited by the sorting stream. A total of 100 cells are targeted in the calibration to wells marked with an X. Wells that are targeted during actual sort are shaded gray. Red ellipses represent the cell dispersal patterns, visualized as drops. Several situations may arise: (i) most drops land in the appointed well and no visible drops land more than one well away. (ii) Most drops land in a localized area but not in the designated wells. (iii) Drops fall in a smear that spans less than three wells. (iv) Drop smear spans more than three wells and will cause some wells to have two sorted clones if every second well is targeted during sort. See EQUIPMENT SETUP for correction procedure.

selected shutter from the reflected to the transmitted shutter. End macro recording, then view the recorded macro by choosing "Edit macro" and scrolling to the bottom of the macro code. The recorded macro should contain the command for selecting the transmitted light shutter, of the form ret = IpScopeControl(SCP_SETCURRSHUTTER, 0, 1, "", IPNULL). Go to the subroutine "Private Sub TakePicPh" included in the ScanPlate macro code (the equivalent subroutine for reflected light is "Private Sub TakePicFl") and

look for the same type of command (for transmitted light, ret = IpScopeControl (SCP_SETCURRSHUTTER, 0, 0, "", IPNULL). Note the difference in this example: the shutter named 0 is now named 1. Replace the old command with the new command. To change the location at which image files are saved, find the command FileName="D:\current.scan\"+DlgText$("WellIndex")+"-phase. TIF" in the same subroutine and change the location from D:\current.scan to the desired location.

PROCEDURE
Production of retrovirus and infection (days 1–7)
1| Day 1: plate Phoenix packaging cells at 40% confluence in 6 cm cell culture dishes. After plating, incubate at 37 °C overnight.

2| Day 2: transfect virus-packaging cells to produce virus containing the CD tag. First, mix 9 µl of FuGene with 3 µg of CD-tagging vector DNA (if pBabeAE, mix 1 µg of each of the three reading frames) in 100 µl of DMEM media without FBS. Flick the tube and incubate for 30 min at room temperature (20–25 °C). Add the DNA/FuGene mixture into culture medium of the Phoenix cell dish dropwise and then incubate cells at 37 °C overnight without changing the medium.
▲ **CRITICAL STEP** Avoid FuGene touching the plastic wall of the tube.

3| Day 3: replace the medium on the Phoenix cells with 4 ml of fresh medium and incubate at 32 °C for about 24 h without changing the medium. Culturing transfected Phoenix cells at 37 °C is also possible, but reduces viral titer. Note that stable lines producing virus can also be generated, but are not covered here.

4| Day 3: plate target cells in all wells of a six-well plate at a concentration of 10^5 cells per well for spinfection (infection step 6, option A below) or to a 20% confluence on a 10 cm dish for regular infection (Step 6, option B below). Incubate target cells overnight at 37 °C.

5| Day 4: collect the culture supernatant from the Phoenix packaging cells and centrifuge it at 450*g* for 5 min at room temperature to sediment cell debris. Retain the supernatant.

6| Day 4: infection. Cells can be infected via spinfection (option A) or by a simpler, less effective regular infection (option B).
(A) Spinfection
  (i) Remove the medium from the target cells and discard. Add 2 ml fresh media without FBS to each well.
  (ii) Add 8 µg ml^{-1} polybrene to the virus-containing supernatant from Step 5. Add 2 ml per well of the virus supernatant to the target cells.
  (iii) Centrifuge the target cell plate at 1,000*g* for 2 h at room temperature (about 24 °C).
  (iv) Incubate cells at 32 °C for 24 h without changing the medium.
(B) Alternative infection protocol: regular infection
  (i) Remove all but 2 ml of the target cell medium in the 10 cm dish and add 4 ml of centrifuged viral supernatant from Step 5 and 4 µg ml^{-1} polybrene.
  (ii) Incubate cells overnight at 32 °C without changing the medium.

7| Day 5: remove medium and replace with fresh culture medium and incubate cells at 37 °C. Incubate cells for 48 h, splitting cells if necessary.
▲ **CRITICAL STEP** Cells that express fluorescent fusion proteins should start appearing after the end of the incubation period. Allow an extra day for protein accumulation, and examine the cell population by FACS to detect if positive cells are present. With experience, the rare positive cells can also be detected by microscopy. To verify that infectious virus was produced, check for puromycin resistance conferred by the virus. Negative control: uninfected cells. Positive control: NIH 3T3 cells, which readily undergo infection with MoMLV.
■ **PAUSE POINT** Cells can be frozen in liquid nitrogen after 48-h incubation indefinitely.
? **TROUBLESHOOTING**

Preparation of infected cells for FACS sorting
8| Set up cells for production of conditioned medium. Conditioned medium is medium from exponentially growing cells, which provides survival signals, helping cells survive the single-cell stage after cell sorting. We found that incubation in conditioned medium was required for incubation of all cell types following sorting (Step 11). We observed that the confluence of the H1299 non-small lung cell carcinoma cell line, which produces the best conditioned medium, is 80%. Optimal conditioned medium is therefore produced by plating cells of the same type on day 6 and harvesting the medium when they reach 80% confluence on

day 8. However, time can be saved by using the conditioned medium from the infected cells on day 8. Either way, use medium from at least one 15 cm dish. Collect the supernatant (for adherent cells) or spin down and collect the supernatant (for cells in suspension) on day 8 and filter using a 0.45 μm filter.

■ PAUSE POINT Excess conditioned medium can be stored at 4 °C for 1 day.

9| Day 7: split the infected target cells into 15 cm plates.
▲ CRITICAL STEP Use the minimal number of passages between infection and sorting. We observed that the clonal diversity of the library is reduced with each passage owing to drift or selection effects.
▲ CRITICAL STEP To reduce the occurrence of aggregates that may block the cell sorter, confluence of infected cells should be preferably 50%, and not be higher than 70% on the day of the sort.

10| Day 8: harvest conditioned medium from either the specially grown cells or from the infected cells. If conditioned medium is derived from infected cells, replace with fresh medium to prevent drying of cells while subsequent steps are performed.

11| Day 8: add 25 μl conditioned medium to each second well of the tissue culture optical 384-well plate using a multipipette. Place the plate in the incubator.

12| Day 8: if the infected cell line being used is adherent, trypsinize the cells. After the cells detach, add 4 volumes of calcium- and magnesium-free PBS.

13| Spin at 250g at room temperature for 2 min. Discard the supernatant.
▲ CRITICAL STEP Fast or long spins may create aggregates.

14| Resuspend the infected cells in 1 ml calcium- and magnesium-free PBS for each 10% of confluence (e.g., use 5 ml calcium- and magnesium-free PBS to resuspend a 15 cm dish of cells at 50% confluence). Cells can be concentrated more if they are suspension cells or do not present aggregation problems. Filter cells through a 10 ml syringe with two layers of 50 μm sterile mesh (see EQUIPMENT SETUP for filter assembly).
▲ CRITICAL STEP Using PBS with calcium and magnesium will create a severe aggregate problem.

15| Aliquot 2–3 ml of mesh-filtered cell suspension into each sterile FACS tube directly from the filtering syringe and keep cells on ice until sorting.

Cell sorting (day 8)
16| Configure laser line and optical components for the fluorophore used (see EQUIPMENT SETUP and **Fig. 3a,b**).

17| Calibrate drop delay and sorting stream direction (see EQUIPMENT SETUP and **Fig. 3c**).

18| Run cells and determine forward and side scatter settings, then fluorescent signal settings. For the YFP signal, use log instrument settings and a dot plot with the 535 nm signal on the x axis and the 575 nm signal on the y axis. Use compensation on the 575 nm signal to view positive cells more easily (emission at 575 nm minus X% emission at 535 nm, where usually $X = 15$–40%). For the mCherry signal, use linear instrument settings and a dot plot with the 695 nm signal on the x axis and the side scatter signal on the y axis (**Fig. 4**).
▲ CRITICAL STEP Use uninfected cells as a negative control to determine if CD-tagged proteins are present, and the gating region to be used for positive cells. Draw the region so that a small fraction of the negative control cells (0.01–0.02% for YFP, up to 0.03–0.1% for mCherry) fall within the gate. The number of positives when infected cells are run with the same gate should be three- to tenfold higher. Cells transfected with YFP or mCherry can be used as a positive control. Note that fluorescence of these cells may be much higher and may not guarantee that CD-tag protein fluorescence will be detected.
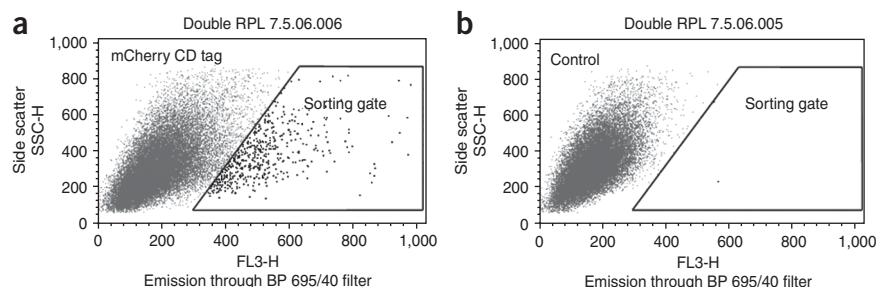

Figure 4 | Cell sorting of positive clones by flow cytometry. Shown are cells CD-tagged with mCherry and selected for mCherry fluorescence. Positive cells were selected based on a high BP 695/40 nm to side scatter ratio. (**a**) Clone 280705p1f13b11 with RPL7 CD-tagged with YFP and super-infected with mCherry. (**b**) The same clone not super-infected with mCherry (control). The number of cells positive for mCherry ranged from 0.1% to 0.5%. Adapted from ref. 25.

19| Start sorting at 1 cell per well into every second well.
▲ **CRITICAL STEP** Do not leave plates in the sorter for more than 20–25 min if the plate chamber is not cooled or incubated with CO_2. If the time it takes to fill a plate is longer, fill half-plates at a time, or use medium with HEPES buffer instead of CO_2-dependent bicarbonate.
▲ **CRITICAL STEP** Adherent cells tend to be sensitive to the flow rate within the sorter. Keep the flow rate low for such cells, at about 300–600 cells per second if the concentration used is as described in Step 14. This will also reduce blockage of the nozzle tip.
? TROUBLESHOOTING

20| After sorting cells, incubate the 384-well plates for 2–3 days before changing the medium.
? TROUBLESHOOTING

Clonal expansion from single cells (days 8–20)
21| Two to three days after the sort, remove the medium from the wells by decanting and replace with 25 μl of 0.45 μm pore filtered conditioned medium from subconfluent cells (generated as described in Step 8). Decant old medium by uncovering the plate, turning upside-down, and flicking medium onto a 15 cm dish.

22| After the first medium change, incubate the cells for 2–3 days before changing the medium again.

23| After the additional incubation time, remove the medium from the wells by decanting and replace with 25 μl of 0.45 μm pore filtered conditioned medium from subconfluent cells diluted with 2 volumes of fresh medium.
▲ **CRITICAL STEP** Small cell colonies should be visible by the end of the second incubation. Check for their presence.
? TROUBLESHOOTING

24| Incubate the cells until the cells are confluent (usually 10–14 days after sort).
▲ **CRITICAL STEP** If the mCherry tag was used, or the researcher is interested in recovering only specific localizations, screen the optical 384-well plate using microscopy (Step 26, option A) about 1–2 population doublings before most clones reach confluence.

25| When cells reach 90–100% confluence, split them to three 96-well plates: two regular tissue culture plates for freezing and RNA isolation, and one optical plate for microscopy.

26| For microscopy (option A below), wait for 2 days or until cells in 96-well optical plate reach 30–40% confluence, whichever is longer. For both freezing (option B below) and RNA isolation and tagged protein identification using 3' RACE (option C below), wait until cells become confluent. The first stage of option C is isolation of total RNA from 96-well plates. Note that multiple kits are available for this, but we describe here what worked best for us.
(A) Microscopy screen of 384- or 96-well plates
  (i) Decant medium from the optical plate and replace with 25 μl (for 384-well plate from Step 24) or 100 μl (for 96-well plate) optical medium.
  (ii) Insert the plate in the plate holder and place in the microscope, making sure that it sits properly.
  (iii) Next steps assume our screening program is used. In ImagePro, press the camera icon to open the camera interface. In the Acquire menu, open ScopePro, then StagePro.
  (iv) After StagePro window opens, select to calibrate stage by joystick (first option). Follow prompts and move stage to the lower right and upper left corners of the plate. The StagePro dialog box should now open.
  (v) In StagePro, open the Sample Pattern tab. Choose the plate used (96- or 384-well plate), then move to the middle of the last well (H-12, 96-well plate; P-24, 384-well plate). Press "Set sample pattern origin". Well locations are now calibrated. Start preview, choose a well with cells and focus on them.
  (vi) In the Macro menu, go to "ScanPlate". A dialog box will open. Options include moving to specific wells or to the next well.
  (vii) Create a folder named "current.scan" in drive D so that the location reads D/current.scan (drive D is the default for saving the image files using our program. To change location, see the microscopy section in EQUIPMENT SETUP).
    By pressing on the Phase or Fluorescence bars, transmitted light or fluorescent images will be acquired and stored in the "Phase" or "Fluorescence" subfolders in the current.scan folder. This macro enables semi-automatic scanning, where the investigator chooses the well and focuses on the cells, while the operations of acquisition and storage are automated.
    ! CAUTION As the plate may be sloped even when inserted properly, making very large stage moves may result in the objective colliding into the plate.
    ? TROUBLESHOOTING
(B) Cell freezing and thawing
  (i) For adherent cells, aspirate off all of the media and wash once with PBS.
    ▲ **CRITICAL STEP** Clones should be mostly confluent.

(ii) For adherent cells, add 50 µl of trypsin using a multichannel pipette to each of the wells, and incubate for 5–10 min at 37 °C. Trypsin incubation times may need to be calibrated for different adherent cell types.

▲ **CRITICAL STEP** A high trypsin concentration or long incubation times may kill the cells. We use a 0.05% trypsin solution and check for cell detachment after the first 5 min incubation, before proceeding with an additional 5 min incubation.

(iii) For adherent cells, add 50 µl of ice-cold 2× freezing medium (66% growth medium, 20% FCS and 14% DMSO, v/v) to each well and resuspend cells by pipetting five times. Place the plate on ice for 5–10 min. For non-adherent cells, add 100 µl of 2× freezing medium (76% growth medium, 10% FCS and 14% DMSO, v/v) to 100 µl of unchanged growth medium in each well.

(iv) Cover the plate in styrofoam or air bubbles and wrap the covered plate with two layers of pads for liquid absorption, and freeze at −80 °C.

▲ **CRITICAL STEP** After freezing, wait to thaw until tagged proteins are identified and localizations determined. Continue working with clones for which identification was successful.

■ **PAUSE POINT** The plate can be stored at −80 °C for up to 2 months.

(v) To thaw cell clones, place the 96-well plate from the −80 °C freezer directly into the 37 °C incubator. Allow all of the wells to thaw (5–7 min for a 100 µl volume per well), then place the plate on ice to prevent DMSO and trypsin toxicity. Transfer the clones for which tagged protein identification was successful to wells in 24-well plates containing 2 ml of culture medium to dilute DMSO and trypsin.

(vi) After 24 h, replace old medium with fresh medium.

(vii) Continue expanding successfully identified clones to six-well plate wells, then to 10 cm dishes, until a confluent 10 cm dish per clone is obtained.

(viii) To freeze clones from 10 cm dishes, precool a Cryo 1 °C freezing container to 4 °C. Freeze down cell clones by adding 1 ml per aliquot of a solution of 83% complete growth medium, 10% FCS and 7% DMSO, v/v. Immediately place in the freezing chamber. Once the chamber is full, transfer to −80 °C for at least 2 h, then to liquid nitrogen. For the H1299 non-small lung carcinoma cell line, four aliquots can be frozen from each 10 cm dish.

■ **PAUSE POINT** Cells can be stored in liquid nitrogen indefinitely.

(C) Tagged protein identification using 3′ RACE

(i) Isolation of total RNA from 96-well plates: aspirate off the media from a confluent 96-well plate of tagged clones.

(ii) Wash the wells with 200 µl of PBS and add 150 µl of ZR-96 Mini RNA isolation I Kit RNA extraction buffer. Shake the plate vigorously for 1 min.

(iii) Incubate the lysate on ice for 10 min.

(iv) Add 150 µl of 100% ethanol to each well and mix by pipetting up and down three times, then incubate on ice for 10 min.

(v) Transfer the sample mixture to the wells of a Zymo-spin I-96 filtration plate mounted on a collection plate.

(vi) Centrifuge at 2,000g for 5 min. Discard the flow-through.

(vii) Add 0.2 ml RNA wash buffer to each well of the Zymo-spin I-96 filtration plate. Centrifuge at 2,000g for 5 min. Repeat this wash step once.

(viii) Elute RNA onto a sterile collection plate by adding 25 µl of RNase-free water to each well of the Zymo-spin I-96 plate. Incubate the plate for 1 min at room temperature and centrifuge with a collection plate at 2,000g for 5 min. Isolated total RNA is ready for first-strand cDNA synthesis.

■ **PAUSE POINT** RNA can be stored at −70 °C for several months.

(ix) First-strand cDNA synthesis: in a 96-well PCR Thermo-Fast plate, denature 12 µl of the total RNA by incubation for 5 min at 65 °C. Place on ice.

(x) Prepare master mix on ice. For each reaction (all concentrations are of working stocks), use 2 µl of 10× RT buffer, 2 µl dNTP mix (5 mM each), 2 µl of 10 µM AP first-strand primer, 1 µl RNase inhibitor (10 U µl^{-1}) and 1 µl Omniscript reverse transcriptase (4 U µl^{-1}).

(xi) Add 8 µl master mix into tubes containing denatured template. Briefly centrifuge to collect the contents at the bottom.

(xii) Incubate for 90 min at 37 °C in a PCR machine. This cDNA is the template used in the subsequent nested PCRs.

■ **PAUSE POINT** DNA can be stored at −20 °C for several months.

(xiii) Nested PCR for amplification of cDNA containing the CD-tag sequence: for the first PCR, make a PCR mix containing 60–70 pmol of first reaction primers (**Table 2**) and PCR mix such as Ready-Mix (contains 1.25 U DNA polymerase, 37 mM Tris-HCl (pH 8.8), 10 mM (NH$_4$)$_2$SO$_4$, 1.5 mM MgCl$_2$, 0.01% Tween 20, 0.2 mM each dNTPs, glycerol and red dye). Let the reaction mix reach 80 °C (hot start) and add 2.5 µl of first-strand-synthesized cDNA as template. Use the following PCR program: 10 cycles: 94 °C for 15 s, annealing at 66 °C for 30 s and elongation for 2 min at 72 °C. After the tenth cycle, continue with 20 cycles of 94 °C for 15 s, annealing at 66 °C for 30 s and elongation for 2 min, plus 5 s time increase per cycle at 72 °C. After the last cycle, add a final elongation for 7 min at 72 °C.

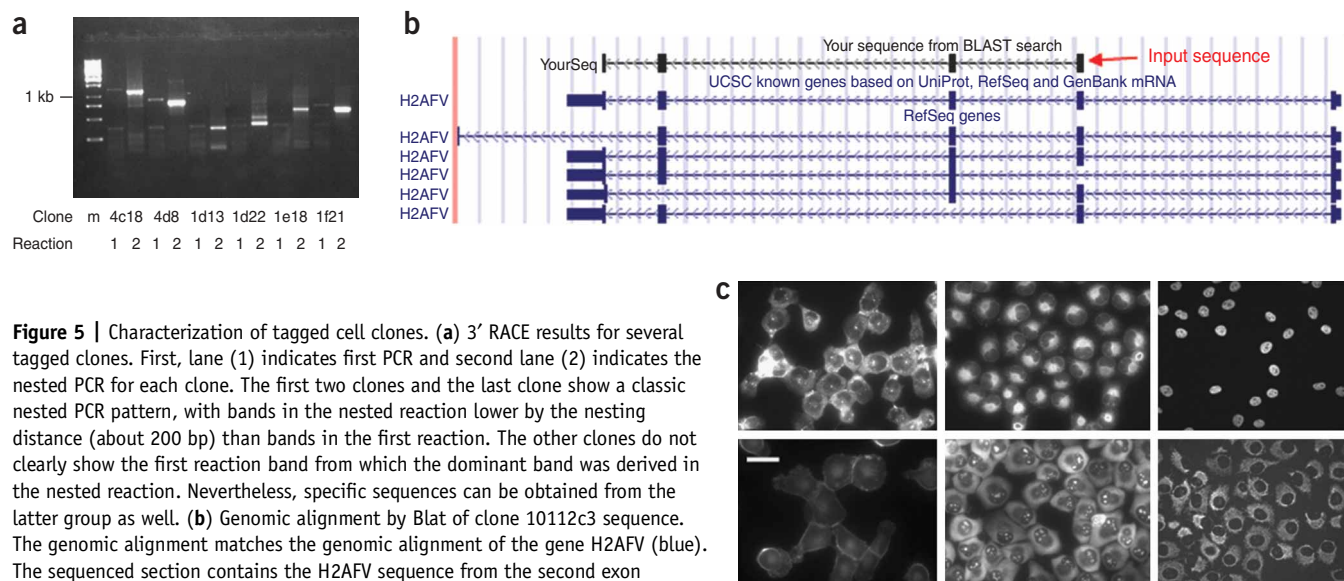■ **PAUSE POINT** The PCR product can be stored at 4 °C for months.

Figure 5 | Characterization of tagged cell clones. (**a**) 3' RACE results for several tagged clones. First, lane (1) indicates first PCR and second lane (2) indicates the nested PCR for each clone. The first two clones and the last clone show a classic nested PCR pattern, with bands in the nested reaction lower by the nesting distance (about 200 bp) than bands in the first reaction. The other clones do not clearly show the first reaction band from which the dominant band was derived in the nested reaction. Nevertheless, specific sequences can be obtained from the latter group as well. (**b**) Genomic alignment by Blat of clone 10112c3 sequence. The genomic alignment matches the genomic alignment of the gene H2AFV (blue). The sequenced section contains the H2AFV sequence from the second exon onward, indicating that the integration occurred in the first intron. (**c**) Examples of localizations of YFP CD-tagged proteins from our CD-tag library. First row, left to right: MYH9 (clone 010506d7p1d11, cytoplasm), PDIA6 (clone 010506d7p1b6, endoplasmic reticulum), HNRPD (clone 010806d7p3f9, nucleus). Second row, left to right: BSG (clone 210206d7d9, plasma membrane), RPL7 (clone 310506p13f10, ribosome), C1QBP (clone 010806d7p3e6, mitochondria). Scale bar, 25 μm.

(xiv) For the nested reaction, make a PCR mix containing 60–70 pmol of nested reaction primers (**Table 2**) and Ready-Mix. Hot start the reaction mix with primers at 80 °C and add 1 μl of PCR product from the first reaction.
  ▲ **CRITICAL STEP** Run the PCR products (see also Step xv) to ensure that the reaction worked.
  ■ **PAUSE POINT** The PCR product can be stored at 4 °C for months.
  ? TROUBLESHOOTING

(xv) Direct sequencing of nested PCR products to obtain gene identity: run the PCR products on a 1% agarose gel. If specific bands are visible in the first reaction, they should be higher by the nesting distance than the dominant bands in the nested reaction (**Fig. 5a**). However, we suggest sequencing bands whether or not they show the correct nesting, unless they are nonspecific. In the case of primers used here, the nested YFP band should be 180 bp lower and the nested mCherry band 130 bp lower on the gel. If a single dominant nested band appears, purify using the QIAquick PCR purification kit and sequence directly using the 906 primer (YFP) or 56 primer (mCherry). If multiple bands appear, use the QIAquick gel extraction kit and sequence each band separately with the primers above. As often multiple bands indicate splice variants of the same RNA, consider using the much quicker PCR purification kit on multiple band samples to save time. Bands derived from multiple labeling can be reduced by decreasing the multiplicity of infection.
  ? TROUBLESHOOTING

(xvi) To obtain gene identity from the sequencing result, take the sequence from the end of the CD-tag sequence to the end of the region where the sequence remains good, and align to the genome using the Blat genomic alignment tool[27] (http://genome.ucsc.edu/cgi-bin/hgBlat). Choose the alignment with the best score, highest percent identity and especially longest span, indicative of exon–intron structure of almost all known genes (**Fig. 5b**). The first aligned exon indicates the integration point (e.g., if the first aligned region is a complete exon 2, the integration occurred in intron 1). Note that the sequence immediately after the tag sequence may be problematic. In this case, the alignment in this region will fail owing to the bad sequence and will start only when the sequence improves inside the first aligned exon. Thus, tag integration will seem to be inside the first exon, when in fact it occurs in the intron before it.

● **TIMING**
Timing information can be found in **Table 1**.

? TROUBLESHOOTING
Troubleshooting advice can be found in **Table 3**.

TABLE 3 | Troubleshooting table.

Step	Problem	Possible reasons	Solutions
7	No fluorescent cells	Infection failed	Use selection marker to check infection (for pBabeAE, puromycin). If less than 20% of cells are resistant, redo infection, and use fresh FuGene and low-passage packaging cells. For further infection troubleshooting, see the Nolan lab website at http://www.stanford.edu/group/nolan/
19	Frequent blockage of sorter	Aggregates	Filter cells as described in PROCEDURE. Make sure that PBS used to resuspend cells is calcium- and magnesium-free and prefiltered through a 0.22 μm filter. Reduce spin speed in cell preparation for sorting. Reduce cell density before harvesting and make sure cells are in good condition. Reduce cell concentration and sorting speed to sort about 200 cells per s
20	Plates contaminated with bacteria after sorting	Contamination of sample fluid path	Assume all beads used for calibration are nonsterile. Therefore, sterilize sample fluid path as described in EQUIPMENT SETUP
23	Few or no cell colonies on 384-well plates	Cell line does not tolerate single cell dilution. Sorting speed too high. Cells did not land in the correct wells. Plate kept outside the incubator for too long. Incorrect drop delay	Change cell type. Reduce sorting speed. Do stream calibration as outlined in EQUIPMENT SETUP. Reduce the time that plates spend outside the incubator by filling half plates at a time. Redo calibration of drop delay
26A(vii)	Microscopy shows no fluorescent signal for most or all of the clones	Incorrectly set sorting gate. Optical medium or plate not used. Poor microscope sensitivity. Camera problems	The sorting gate should contain cells that have normal forward and side scatter and fall within the positive region (e.g., **Fig. 4a**). Use optical plates and medium, as regular medium or plates greatly increase background fluorescence. To increase microscope sensitivity, use higher numerical aperture objectives and a cooled CCD camera, and increase exposure times to several seconds and/or use binning and gain. The camera may not be synchronized with the shutter. Since this should also be a problem with transmitted light, it is detectable by comparing to images with the same exposure while the shutter is open manually. To solve, obtain a different camera driver
26A(vii)	Preview slow to respond to stage movement or other changes	"Workspace preview" option is selected in ImagePro	Uncheck the "workspace preview" in the preview tab of the camera control interface in ImagePro
26A(vii)	Shutters do not respond when using the ScanPlate program, or open the wrong lightpath	Shutter designation is different in the ScanPlate program from the designation in ScopePro in your microscopy setup	To solve, see the "Configuring our ScanPlate program to run correctly on a new microscope setup" part of the EQUIPMENT SETUP section
26C(xiv)	No PCR product	Low amounts of RNA or RNA degradation. Problems with RT or PCRs. Hard-to-amplify sequences	Start procedure with wells at 100% confluence on average. Run RNA on gel to determine if degradation occurred. In non-degraded RNA, the 18S and 28S RNA bands should be clearly visible. Prevent degradation by using RNase-free reagents and clean surfaces with RNaseZap. No product in any reaction may indicate RT or PCR problems. In contrast, no product in occasional reactions does not indicate errors in the procedure and may be the result of hard-to-amplify sequences under the conditions used
26C(xv)	Bad sequence	Several sequences mixed together. Low DNA purity	If two or more bands appear on gel, clean and sequence each one. Otherwise redo PCR and purification of PCR product

ANTICIPATED RESULTS

The fraction of fluorescent CD-tagged cells

In a typical library generation cycle, about 3×10^6 cells are infected. The fraction of positive fluorescent cells tends to range between 0.1% and 1%. This relatively low fraction can be explained if most integration events are either in noncoding parts of the genome, in genes not detectably expressed in the cell line, or in the wrong frame or orientation. We estimate the probability of fluorescently tagging a protein in a cell to be $P_t = \text{MOI } P_g P_o P_e$, where MOI is the multiplicity of infection, or how many viruses infect one cell, P_g is the probability that the virus will enter a gene rather than an intergene region, P_o is the probability that the tag sequence will be in the correct frame and orientation relative to the gene and P_e is the probability that the gene is expressed in the cell line used.

Our MOI was on the order of 1 virus per cell, based on a rough estimate from the fact that about 80% of cells were puromycin resistant (the puromycin-resistance marker is included in pBabeAE). P_g can be estimated as the proportion of the genome occupied by genes, which is about 20% (20% introns and less than 1% exons[28]). The number of possible frames is 3 and number of orientations 2; therefore P_o, the probability of randomly inserting in both the correct frame and orientation, is 1 in 6 or about 15%. Assuming 20% of the total genes are expressed in a particular cell line, $P_t = 1 \times 0.2 \times 0.15 \times 0.2 \sim 0.01$. This fraction of about 1% corresponds well with the actual fraction of tagged cells observed. The probability of labeling two proteins in the same cell, assuming that tagging events are independent and 1% of cells are tagged, is therefore $0.01 \times 0.01 = 0.0001$, that is, 1 in 10,000 cells, or one in a hundred tagged clones will have two fluorescently tagged genes. This number is reasonably small but can be reduced further by decreasing the MOI, which results in a lower fraction of positive cells. A comfortable percentage of positive cells to work with is 0.3%, where we expect that 1 in about 300 positive fluorescent cells will have two fluorescently labeled proteins.

Factors determining the number of clones produced per plate by sorting

We found that cell survival and growth after single cell sorting is highly cell type dependent. With the H1299 non-small lung cell carcinoma cell line, we observed that about 30% of cells survived the single cell sorting stage to form clones. Empty wells may be the result of cell loss from death owing to sorting and single-cell density stress, and misses by the sorting streams. The murine NIH 3T3 and C2C12 cell lines had slightly lower survival. However, survival was only about 5% for the human Jurkat T-cell line, and no survival was seen for the mouse M1 myoblast cell line under the same conditions (A.S., unpublished observations).

Detection of positive cells and screening for specific localizations

Detection of YFP- or mCherry-positive cells by flow cytometry enables detection below autofluorescence values, as autofluorescence is characterized by a constant 535 to 575 nm ratio for YFP[18], or a constant side scatter to 695 nm ratio for mCherry (**Fig. 4b**) The cells near autofluorescence levels when viewed by flow cytometry may still be clearly visible under the microscope because of the closer match of excitation wavelength made possible by filter sets specific for the fluorophore. The signal to noise level can be further increased in a microscope image if the fluorescence is localized to a small area.

Microscopy of YFP-tagged proteins in cell clones showed diverse localizations (**Fig. 5c**). We observed that some mCherry CD-tagged proteins seemed to show what we term a nonspecific localization when compared to YFP, where such localization patterns were rarely seen. For reasons we do not understand, this type of localization appeared in about 50% of mCherry-positive clones, and consisted of the labeled protein localizing to the nucleus, nucleoli and cytoplasm, and highly fluorescent speckles (**Fig. 2**). The clones showing this localization gave RACE products that did not align to either characterized genes or uncharacterized genes with supporting evidence (e.g., exon–intron structure, EST alignment). As a working assumption, we treated clones with this localizational signature as problematic and did not select them for further studies. However, many mCherry proteins showed specific localizations. To pick them out, we screened the optical 384-well plates at ×20 magnification (**Fig. 2**). Such screening can also be performed to obtain particular localization patterns[25].

Localizations and stability of CD-tagged proteins

In our experience, about two-thirds of YFP CD-tagged proteins had similar localization patterns to those reported in studies using either antibody staining or end labeling with GFP[18]. The mislocalization of the remaining one-third may have several explanations. The most probable is that the inserted YFP tag disturbed the protein sufficiently to affect its localization. It could also have resulted from wrong identification, which may occur if more than one expressed gene is labeled (the probability of this is three times higher than the probability of a cell actually containing two fluorescent proteins, as the frame is irrelevant to identification). Another possibility is that localizations differ between cell types and conditions.

Our CD-tagging vector contains a puromycin-resistance gene. However, we found that it was not necessary to culture cells with puromycin as the expression levels of the tagged proteins were not observed to decrease with time owing to selection against the tagged proteins on the timescale tested (about 30 population doublings). However, puromycin resistance may be useful to rid clones of cells that have randomly lost the tag, and to evaluate infection efficiency (see **? TROUBLESHOOTING**).

Integration sites of YFP and functions of CD-tagged proteins

Integrations sites of the CD tag had a bias to the 5' end of tagged genes[18], consistent with a known integration preference for the murine leukemia virus[26,29], and with the fact that the first intron tends to be very large in proportion to the rest of the gene[18]. This bias caused the fluorophore sequence insertion point to be close to the N terminus in about half of the tagged proteins[18]. We did not see an obvious correlation between YFP integration site and mislocalization, but we intend to re-examine this issue with a larger protein subset.

The function of the tagged protein in parameters besides localization is difficult to assay on a high-throughput level, because one (or more than one in non-diploid cell lines) untagged allele remains, and may be sufficient to maintain normal cell growth.

Figure 6 | Timescale of library construction. The generation curve for the d7 library initiated in 2006 in our laboratory is shown. Approximate dates for the production and identification of each batch of clones are indicated, and a second-order polynomial curve is added as a guide for the eye.

Time of library generation

Each experiment results in a pool of approximately 10^4 detectably labeled proteins (assuming approximately 10^6 infected cells and a 1% fraction of positives). Out of these, about 100–200 can be cloned, identified and stored per library generation cycle by two researchers, if high-throughput techniques are used (**Fig. 6**). Given that each cycle lasts about 1 month, a 1,000-clone library can be reasonably generated by two researchers in a year, once the protocol becomes routine.

Limits to the number of proteins that can be CD-tagged

Some of the proteins are tagged in multiple clones, reducing the number of different tagged proteins produced in a library generation cycle. The percentage of repeated proteins was about 20% in the first two generation cycles and increased to about 40% by the last cycle analyzed here. One constant source of multiple clones expressing the same tagged protein is multiple cells derived from the same tagged cell clone. We also observed repeated tagging of the same genes in independent infections. These tagging hotspots may be the result of viral preference for either specific genes or highly transcribed genes (A.S., R.M., A.C. & U.A., unpublished observations). This would provide a second constant source of repeated proteins. However, the fact that the number of repeatedly tagged genes increases may indicate that the pool of proteins that can be tagged is limited by the number of proteins one cell line expresses under the conditions assayed (in our case, normal growth), and which can be detected with the fluorophore and equipment used. It may be further constrained by the number of expressed proteins that can tolerate the tag. Based on our current results, we roughly estimate that the number of proteins that can be labeled by CD tagging before saturation would be one to several thousand in a given somatic cell line under the conditions described here, and can be possibly increased by using different fluorophores, infection and sorting conditions, and viral vectors.
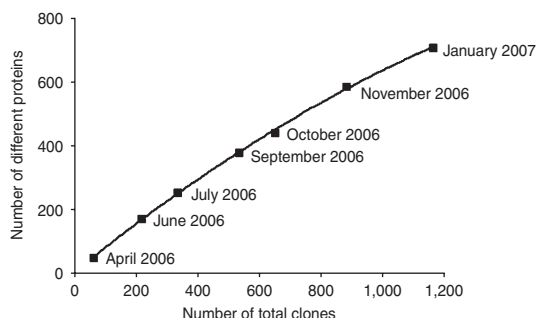
1. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman & Hall/CRC press, Boca Raton, FL, 2006).
2. Andersen, J.S. *et al.* Nucleolar proteome dynamics. *Nature* **433**, 77–83 (2005).
3. Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
4. Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
5. Gavin, A.C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
6. Butland, G. *et al.* Interaction network containing conserved and essential protein complexes in *Escherichia coli. Nature* **433**, 531–537 (2005).
7. Yi, E.C. *et al.* Increased quantitative proteome coverage with (13)C/(12)C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme. *Proteomics* **5**, 380–387 (2005).
8. Whitney, A.R. *et al.* Individuality and variation in gene expression patterns in human blood. *Proc. Natl. Acad. Sci. USA* **100**, 1896–1901 (2003).
9. Perlman, Z.E. *et al.* Multidimensional drug profiling by automated microscopy. *Science* **306**, 1194–1198 (2004).
10. Mayer, T.U. *et al.* Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* **286**, 971–974 (1999).
11. Chen, D. & Huang, S. Nucleolar components involved in ribosome biogenesis cycle between the nucleolus and nucleoplasm in interphase cells. *J. Cell Biol.* **153**, 169–176 (2001).
12. Bannasch, D. *et al.* LIFEdb: a database for functional genomics experiments integrating information from external sources, and serving as a sample tracking system. *Nucleic Acids Res.* **32** Database issue, D505–D508 (2004).
13. Simpson, J.C., Wellenreuther, R., Poustka, A., Pepperkok, R. & Wiemann, S. Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep.* **1**, 287–292 (2000).
14. Huh, W.K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
15. Jarvik, J.W., Adler, S.A., Telmer, C.A., Subramaniam, V. & Lopez, A.J. CD-tagging: a new approach to gene and protein discovery and analysis. *Biotechniques* **20**, 896–904 (1996).
16. Jarvik, J.W. *et al. In vivo* functional proteomics: mammalian genome annotation using CD-tagging. *Biotechniques* **33**, 852–854 856, 858–860 passim (2002).
17. Jarvik, J.W. & Telmer, C.A. Epitope tagging. *Annu. Rev. Genet.* **32**, 601–618 (1998).
18. Sigal, A. *et al.* Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. *Nat. Methods* **3**, 525–531 (2006).

19. Friedrich, G. & Soriano, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* **5**, 1513–1523 (1991).
20. Gossler, A., Joyner, A.L., Rossant, J. & Skarnes, W.C. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463–465 (1989).
21. Stanford, W.L. *et al.* Expression trapping: identification of novel genes expressed in hematopoietic and endothelial lineages by gene trapping in ES cells. *Blood* **92**, 4622–4631 (1998).
22. Morin, X., Daneman, R., Zavortink, M. & Chia, W. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **98**, 15050–15055 (2001).
23. Clyne, P.J., Brotman, J.S., Sweeney, S.T. & Davis, G. Green fluorescent protein tagging *Drosophila* proteins at their native genomic loci with small P elements. *Genetics* **165**, 1433–1441 (2003).
24. Shaner, N.C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572 (2004).
25. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).
26. Wu, X., Li, Y., Crise, B. & Burgess, S.M. Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**, 1749–1751 (2003).
27. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
28. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
29. Mitchell, R.S. *et al.* Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* **2**, E234 (2004).

Corrigendum: Generation of a fluorescently labeled endogenous protein library in living human cells

Alex Sigal, Tamar Danon, Ariel Cohen, Ron Milo, Naama Geva-Zatorsky, Gila Lustig, Yuvalal Liron, Uri Alon & Natalie Perzov

In the version of this article initially published, the authors omitted the following acknowledgment: "We thank the Kahn Family Foundation and the Israel Science Foundation for support." The acknowledgment has been added to the PDF version of the article.