Figure 1. Development of targeted proteomics assays using enriched *in vitro* synthesized full length proteins
(**a**) Proteins for which targeted assays were built. (**b**) Schematic of the synthesis, enrichment, digestion and analysis of proteins to identify proteotypic peptides and their fragmentation patterns. (**c**) Protein samples were highly-enriched and full-length, as detected by silver-staining and immunodetection with an anti-schistosomal GST antibody. (**d**) SRM chromatographic traces from the NFIA peptide EDFVLTVTGK were readily detected over background. (**e**) Proteotypic peptides for EWSR1 were identified by comparing the signal intensity of all of the tryptic peptides monitored.

Figure 2. Targeted assays can be efficiently developed using *in vitro* synthesized proteins and applied to measure proteins *in vivo*
(**a**) The absolute quantity of each *in vitro*-synthesized protein sample, as measured using a tryptic peptide contained within the c-terminal schistosomal GST tag. (**b**) The number of peptides per protein empirically assessed with salient features to accurately detect and quantify the target proteins (peptides with a quality score of either 1 or 2). (**c**) Proteotypic peptides identified using *in vitro*-synthesized CTCF were monitored in K562 nuclear extracts. The relative contribution of each fragment ion to each peptide peak is displayed as different colors. (**d**) For each proteotypic peptide from CTCF, the relative signal intensity observed using in vitro synthesized protein is displayed alongside the relative signal intensity observed using K562 nuclear extract Peptides not observed (n.o.) in K562 nuclear extracts are indicated. (**e**) The measured relative abundance of four transcription factors between the fibroblast (BJ), hepatic carcinoma (HepG2), erythroleukemia (K562) and neuroblastoma (SKNSH) human cell lines. Data points are mean ± s. d. (n = 6).

SUPPLEMENTARY METHODS

Clones and Plasmids
All of the clones used are from the pANT7_cGST clone collection distributed by the DNASU plasmid repository. These full-length cDNA clones contain a T7 transcriptional start sequence as well as an internal ribosome entry site (IRES) which is compatible with *in vitro* transcription/translation reagents[11]. Additionally, each clone contains an in-frame fused c-terminal *Schistosoma japonicum* GST tag. Each bacterial stock clone was grown overnight in 5mL of LB-amp (100μg/mL). Plasmid DNA was extracted using the manufacture mini-prep protocol with the exception of an additional wash with PE buffer (Qiagen). All plasmid stocks were Sanger sequenced (UW High Throughput Genomics Unit) using an M13 priming site upstream of the T7 promoter to confirm the identity of the insert and to ensure that there was no contamination of the plasmid stocks.

Peptides
We obtained ~0.1mg of FasTrack crude heavy $^{13}C_6$$^{15}N_2$ L-Lysine labeled LLLEYLEEK and IEAIPQIDK peptides to use as internal standards (Thermo). The LLLEYLEEK peptide was resuspended in 75% Acetonitrile, 0.1% Formic acid in H_2O. The IEAIPQIDK peptide was resuspended in 5% Acetonitrile

in H_2O. Unlabeled peptides provided at a concentration of 5 pmol/µl (assessed by the manufacturer by amino acid analysis) of AQUA Ultimate light LLLEYLEEK and IEAIPQIDK peptides were obtained to use as calibration standards (Thermo).

Protein production and purification
Protein production and purification was optimized to be performed in 96-well plate format. Different protein production conditions, capture conditions, wash conditions and digestion conditions were tested to identify a protocol that gave maximal protein yield at the highest possible purity. The final protocol takes one person 2 days to transform a 96-well plate of plasmids into desalted peptide samples ready for mass-spectrometry analysis with a cost of less than $20 per protein **(Supplementary Fig. 7)**.

Protein Production Conditions
Proteins were synthesized from plasmid DNA using the Pierce Human *In Vitro* Protein Expression Kit (Thermo) according to the manufacturer's protocol with some slight modifications. Briefly, 1µg of plasmid DNA was transcribed at 32°C for 70 minutes in a 20µl transcription reaction supplemented with 0.3µl RNase Inhibitors (Thermo). 2µl of the transcription reaction was then added to a 23µl translation reaction mix and incubated at 30°C for 2 hours. The translation reaction was then spiked with an additional 2 µl of the transcription reaction and incubated at 30°C for an additional 2 hours.

Protein Capture Conditions
To enrich the GST-fusion protein, we used 2mL of glutathione sepharose 4B beads (GE), washed 3 times with 15 mL 1xDPBS (GIBCO) and resuspended in 12.5mL of 1xDPBS. A 125µl aliquot of the washed bead slurry was added to each well of 8 12-well strip-tubes such that each well received 20µl of packed beads. Completed translation reactions were added to the beads and the bead-protein mixture was rocked end-over-end for 16 hours at 4°C.

Bead Wash Conditions
Bead washing was staggered to ensure that only 2 12-well strip-tubes were washed at a given time. By limiting the number of tubes washed at a time, it enabled the total wash time for each reaction to be reduced to less than 25 minutes. The bead-protein mixture was sedimented at 500 g for 2 minutes using a swinging plate rotor. The supernatant was removed and 150 µl of Wash buffer (1xDPBS supplemented with 863 mM NaCl) was added to the beads. The beads were mixed by inverting several times and sedimented at 500 g for 2 minutes. The beads were washed a total of 2 times with 150 µl Wash buffer each and 2 times with 150 µl 50 mM Ammonium Bicarbonate pH 7.8 each. After the last wash, the beads were resuspended in 100 µl Elution Buffer (0.05% PPS (Protein Discovery), ~5 nM heavy labeled GST peptide LLLEYLEEK(+8 Da) (Thermo), ~5 nM heavy labeled GST peptide IEAIPQIDK(+8 Da) (Thermo) and 50mM Ammonium Bicarbonate pH 7.8) and stored at 4°C until all 8 12-well strip-tubes had been washed. 10 µl of each enriched protein sample was added to 4 µl 4x LDS buffer (Invitrogen) and saved for silver-staining and western blotting.

Protein Digestion
Bead bound protein samples were boiled at 95°C for 5 minutes, reduced with 5 mM DTT at 60°C for 30 minutes and alkylated with 15 mM iodoacetic acid (IAA) at 25°C for 30 minutes in the dark. Proteins were then digested with 400 ng Trypsin (Promega) at 37°C for 2 hours while shaking. Beads were then sedimented at 500 g for 2 minutes and the supernatant, which contained the digested peptides, was transferred to a new 96-well collection plate. The beads were washed once with 150 µl 50mM Ammonium Bicarbonate pH 7.8 and the supernatant from this wash was combined with the previous supernatant. The pH of the supernatant sample was adjusted to <3.0 by

5 µl of 5 M HCl and incubated at 25°C for 20 minutes. The digested samples were desalted using a 96-well Oasis MCX plate 30 mg/60 µm (Waters) following the manufacturer protocol with minor modifications. Briefly, the cartridge was conditioned using 1 mL methanol, 1 mL 10% ammonium hydroxide in H_2O, 2 mL methanol and finally 3 mL 0.1% formic acid in H_2O. The samples were then loaded onto the cartridge and washed with 1 mL 0.1% formic acid in H_2O and 1 mL of 0.1% formic acid in methanol. The peptides were eluted from the cartridge with 600µl 10% ammonium hydroxide in methanol, collected in a 1 mL round bottom 96-well collection plate and evaporated using a SpeedVac (Labconco) set to 50°C. Peptide samples were evaporated down to 10-30µl of volume then resuspended in 50µl 0.1% formic acid in H_2O. These peptide samples were stored at -20°C until injected on the mass spectrometer.

Silver staining and Immunoblotting
Undigested protein extract from each of the fractions was boiled in 1x LDS buffer (Invitrogen) and separated on a 4-12% Bis-Tris denaturing and reducing SDS-PAGE gel (Invitrogen). Gels were then subjected to either silver-staining (Invitrogen) or transferred onto a nitrocellulose membrane (Bio-Rad) for immunoblotting. Membranes were blocked with 5% non-fat dry milk (Safeway) in TBS-tween buffer and probed for schistosomal GST (GE 27-4577-01). All primary incubations were done at 4°C overnight using a 1:1,000 dilution. Secondary incubations were performed in 5% non-fat dry milk in TBS-tween using 1:10,000 diluted peroxidase conjugated Rabbit Anti-Goat IgG (H+L) (Pierce). Membranes were visualized using an ECL plus western blotting kit (Amersham) and detected with radiographic film (Thermo).

Nuclear Protein Extraction
Nuclear proteins from K562, HepG2 and SKNSH cancer cell lines and the BJ fibroblast cell line were isolated in three biological replicates as previously described[16]. BJ cells were grown in MEM (GIBCO) supplemented with 10% Fetal Bovine Serum (PAA), non-essential amino acids (GIBCO), sodium pyruvate (GIBCO), 1.5mg/mL $NaHCO_3$, penicillin and streptomycin (GIBCO). HepG2 cells were grown in MEM supplemented with 10% FBS, non-essential amino acids, sodium pyruvate, penicillin and streptomycin. K562 and SKNSH cells were grown in RPMI (GIBCO) supplemented with 10% FBS, sodium pyruvate, L-glutamine (GIBCO), penicillin and streptomycin. SKNSH cells were treated with 6 µM retinoic acid for 48 hours before harvesting. K562 nuclear extraction was performed by resuspending cells at $2.5*10^6$ cells/mL in Buffer A (15mM Tris pH 9.0, 15mM NaCl, 60mM KCl, 1mM EDTA pH 8.0, 0.5mM EGTA pH 8.0, 0.5mM Spermidine) containing 0.05% NP-40 (Roche). After an 8 minute incubation on ice, nuclei were pelleted at 400 rcf for 7 minutes and washed once with Buffer A. SKNSH, HepG2 and BJ nuclei were isolated in a similar fashion, but with the use of cell line specific NP-40 concentrations and cytoplasmic lysis times (SKNSH was 0.05% NP-40 for 5 minutes; HepG2 was 0.1% NP-40 for 8 minutes; and BJ was 0.5% NP-40 for 40 minutes). Nuclei were then resuspended in Buffer A containing 0.2% NP-40, sonicated at setting output 3 for 30 seconds, digested with Benzonase for 15 minutes at 4°C, digested with DNaseI for 15 minutes at 37°C and finally digested with trypsin. Samples were brought to 6mM $MgCl_2$ before digestion with 0.375U/µl Benzonase (Fisher Scientific). Samples were brought to 6mM $CaCl_2$ and 90mM NaCl before digestion with DNaseI (Sigma). DNaseI digestion reactions were stopped using 50mM EDTA. Nuclear protein samples were digested with trypsin as described above using a 50:1 protein:trypsin ratio. After digestion and MCX cleanup, each sample was resuspended 0.1% formic acid in H_2O to a final concentration of ~10,000 nuclei/µl.

Targeted Proteomic Mass Spectrometry
Peptide samples were analyzed with a TSQ-Vantage triple-quadrupole instrument (Thermo) using either an Eksigent nanoLC separation system (Eksigent) or a nanoACQUITY UPLC (Waters). A 5µl aliquot of each sample was separated on a 16cm long 75µm I.D. packed column (Polymicro

Technologies) using Jupiter 4u Proteo 90A reverse-phase beads (Phenomenex). Peptides were separated using a 27.5 minute gradient from 2% acetonitrile in 0.1% formic acid to 23% acetonitrile in 0.1% formic acid. The gradient was followed by a wash for 10.5 minutes at 80% acetonitrile in 0.1% formic acid and a column re-equilibration at 2% acetonitrile in 0.1% formic acid for 12 minutes. Ions were isolated in both Q1 and Q3 using 0.7 FWHM resolution. Peptide fragmentation was performed at 1.5mTorr in Q2 using calculated peptide specific collision energies[17]. Data was acquired using a scan width of 0.002 m/z and a dwell time of 10ms.

Each protein sample was injected separately. For the target protein, all monoisotopic, +2 charge state, fully tryptic peptides from 7 to 23 amino acids in length were tested. In addition, the heavy and light forms of the schistosomal GST peptides LLLEYLEEK and IEAIPQIDK and the light form of the endogenous glutathione-binding protein GSTM3 peptide IAAYLQSDQFCK were tested. Peptides that flanked the target and fusion protein were not tested. For all peptides, the monoisotopic, +1 charge state y_3 to y_{n-1} fragment ions were monitored. All cysteines were monitored as carbamidomethyl cysteines. All methods were designed such that no more than 240 transitions were monitored in a given run. Quality control runs were staggered every ~8 injection to monitor column stability.

SRM data analysis
Targeted proteomic data was analyzed using the software package Skyline[12]. Skyline can be downloaded at http://proteome.gs.washington.edu/software/skyline/. Chromatographic data from each peptide was manually analyzed to determine the quality of the peptide signal. Scoring of peptide quality was done by assessing the following requirements:
A) A prominent chromatographic peak with a signal intensity of at least 60,000;
B) Two or more data points were collected across the peak;
C) Three or more fragment ions not including y_3 co-eluted to contribute to this peak signal;
D) The chromatographic peak had a Gaussian elution profile.
Based on these requirements, peptides were given a quality score between 1 and 4 with 1 being the highest score.
Quality Score 1 peptides: Peptides that had chromatographic traces that met all of these requirements were given a quality score of 1.
Quality Score 2 peptides: Peptides were given a quality score of 2 if they met requirement A with a signal intensity of at least 20,000 and requirement B but either had only three fragment ions including y_3 contributing to the peak or had an abnormal peak shape.
Quality Score 3 peptides: Peptides were given a quality score of 3 if; a) more than one chromatographic peak was detected that met requirements B, C and D; b) requirement B was not met; c) or if requirements A and D were not met.
Quality Score 4 peptides: Peptides not classified as having a quality score of 1, 2 or 3 were given a quality score of 4.
Chromatographic peak intensities from all monitored transitions of a given peptide were integrated and summed to give a final peptide peak height. Fragment ion chromatographic traces that were clearly contaminated by some other ion were removed from this analysis (these fragment ions are noted in **Supplementary Data 1** as an absence of a monitored fragment ion for a given peptide).

Absolute Quantification
Absolute quantification of the GST peptides LLLEYLEEK and IEAIPQIDK was performed using a calibration curve of light labeled peptides with each sample containing the same amount of heavy peptide. The calibration points used were 40nM, 12.5nM, 5nM, 2.5nM, 1nM, 0.5nM, 0.25nM and 0.1nM each of the light LLLEYLEEK and IEAIPQIDK peptides. All peptide standards were mixed with identical quantities of heavy labeled LLLEYLEEK and IEAIPQIDK peptides (~5nM each) and a Bovine QC standard mix (25nM) (Michrome) in 2% acetonitrile and 0.1% formic acid in water.

Peptide standards were measured in triplicate and a linear regression of the data points was used to calibrate the GST peptide light-to-heavy ratio of all other samples.

Relative Quantification Between Nuclei
6 replicate measurements comprising 2 technical replicates each of 3 biological replicates were made for each protein in each of the four cell types. The peptides monitored were the HIST4 peptide DNIQGITKPAIR, the GATA2 peptide GAECFEELSK, the CTCF peptide CPDCDMAFVTSGELVR, the CREB1 peptide ILNDLSSDAPGVPR and the EZH2 peptide EFAAALTAER. For each replicate, the intensity of the target peptide was normalized to the intensity of the HIST4 peptide to control for any variance in the autosampler and/or chromatography. Additionally, as the amount of Histone 4 protein should be constant between the four cell types, this normalization should correct for any errors in the measurement of nuclei used for digestion. The mean and standard deviation of these normalized intensities were then calculated for each protein in each of the cell types.

Shotgun Proteomic Mass Spectrometry
Peptide samples were analyzed with an LTQ-VELOS instrument (Thermo) using an Agilent 1100 binary pump and autosampler (Agilent). 5μl of each sample was separated on a 16cm long 75μm I.D. packed column (Polymicro Technologies) using Jupiter 4u Proteo 90A reverse-phase beads (Phenomenex). Peptides were separated using a 25 minute gradient from 8.75% acetonitrile in 0.1% formic acid to 33% acetonitrile in 0.1% formic acid. The gradient was followed by a wash for 15 minutes at 65% acetonitrile in 0.1% formic acid and a column re-equilibration at 8.75% acetonitrile in 0.1% formic acid for 15 minutes. Spectra were acquired in Data Dependent Acquisition (DDA) mode. Raw spectral files were searched using the Sequest algorithm[18] and spectra identified using Percolator[19] with an FDR cut off of 1% were used for analysis.

Database spectra analysis
The 2011_05_26 release of the *H. sapiens* Ion Trap library of peptide tandem mass spectra was downloaded from the National Institute of Standards and Technology (NIST)[20]. Out of the 1421 peptides monitored in our dataset, 189 had spectra available in the NIST database. Dot-products were calculated using Skyline for all peptides with four or more monitored fragment ions (186 peptides)[21].

ESPPredictor scores
ESPPredictor scores[6] were calculated using the Gene Pattern web-tools interface (http://www.broadinstitute.org/cancer/software/genepattern/modules/ESPPredictor.html). Proteins that had 8 or more peptides, with one third of these peptides having a quality score of 1 or 2 were analyzed using the ESPPredictor algorithm (75 total proteins).

REFERENCES

1. Lange, V., Picotti, P., Domon, B. & Aebersold, R. *Mol. Syst. Biol.* **4**, 222 (2008).
2. Carr, S. A. & Anderson, L. *Clin. Chem.* **54**, 1749-1752 (2008).