Supplementary Information

State-based discovery: a multidimensional screen for small-molecule modulators of EGF signaling

Mark Sevecka & Gavin MacBeath

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, 02138, USA.

Correspondence should be addressed to Gavin MacBeath (macbeath@chemistry.harvard.edu)

Nature Methods - **3**, 825 - 831 (2006)
Published online: 21 September 2006; | doi:10.1038/nmeth931

This files contains the following supplementary information:

Supplementary Figure 1
Evaluation of antibodies for lysate microarrays.

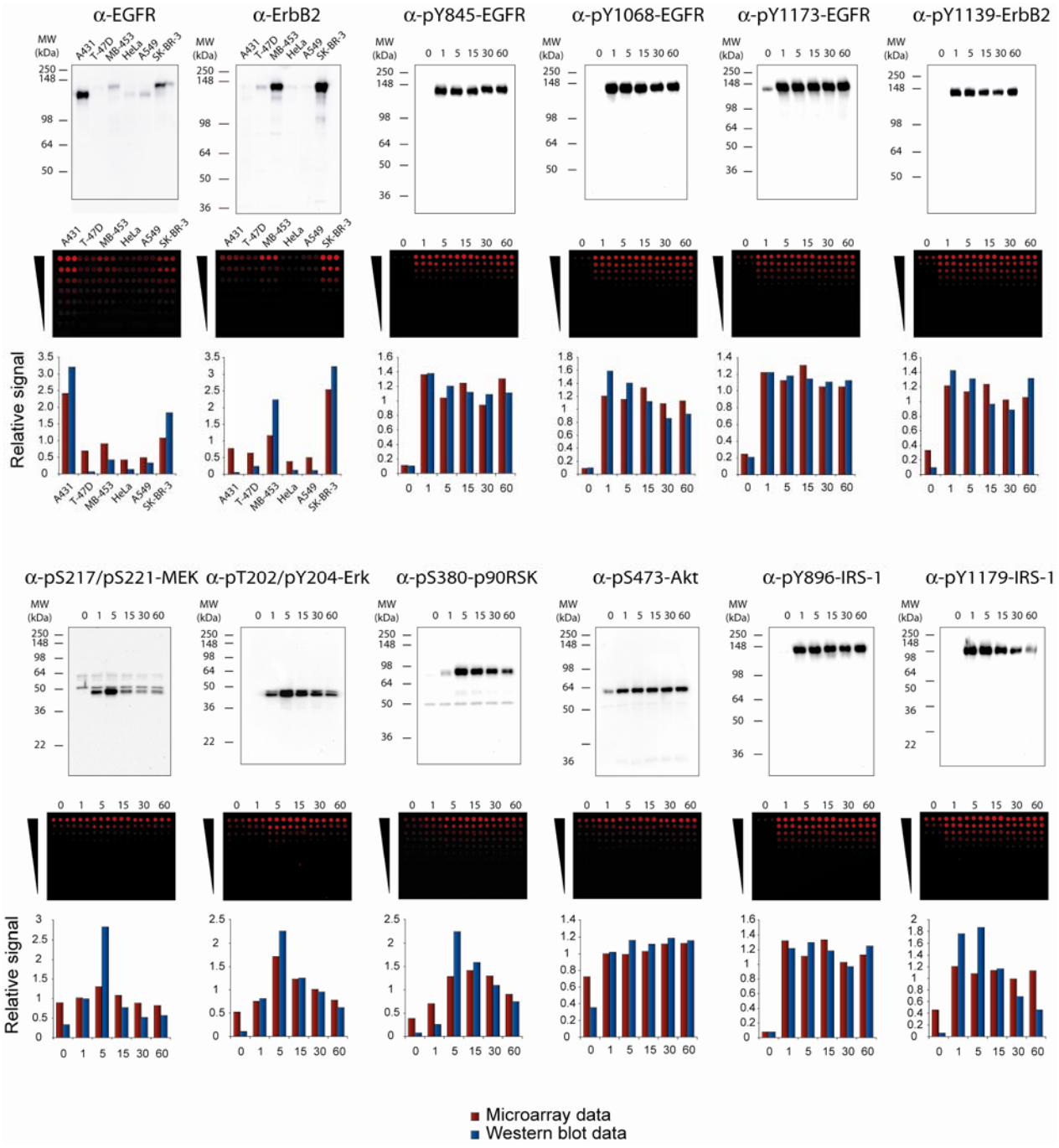
Supplementary Figure 2
Verification of compound activities using traditional western blotting.

Supplementary Table 1
Antibodies used in the state-based screen.

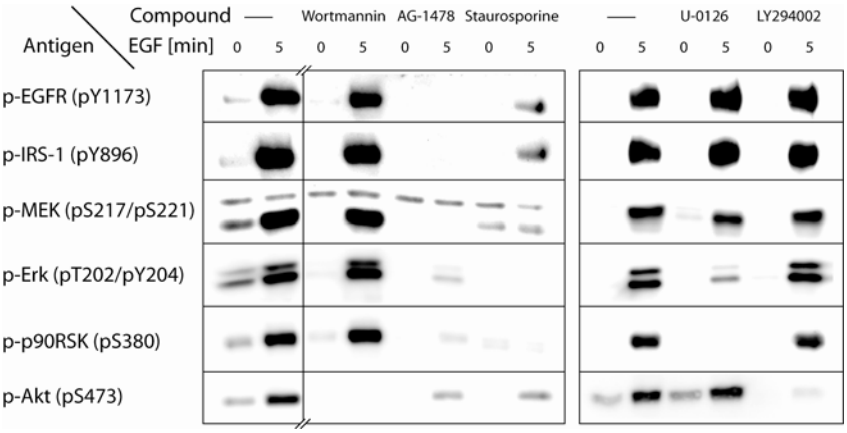
Supplementary Table 2
Identity of compounds in the model library.

Supplementary Methods
Reagents, cell culture, immunoblotting, and determination of dose-response behavior.

Supplementary Note
Data processing steps used in the state-based screen and how these steps affect uncertainty in the data.



Supplementary Figure 1 | Evaluation of antibodies for lysate microarrays. Cellular lysates were analyzed at one concentration on immunoblots (top row for each antibody) and at seven different concentrations on lysate microarrays (two-fold dilution series; center row for each antibody). For pan-specific antibodies, the lysates were derived from six cancer cell lines; for phospho-specific antibodies, the lysates were derived from A431 cells stimulated for different times with 200 ng/ml of EGF. Quantitative data from the immunoblots and the lysate microarrays were normalized relative to their respective mean values and compared (bottom row for each antibody). Shown here are data for all antibodies that produced specific signals on the lysate microarrays and were used in the state-based screen. Anti-pY845-EGFR, anti-pY1068-EGFR, anti-pY1173-EGFR, and anti-pY896-IRS-1 antibodies produced signal ratios on the lysate microarrays that were identical (within error) to those observed on the immunoblots. The other eight antibodies produced compressed signal ratios on the lysate microarrays relative to the immunoblots.



Supplementary Figure 2 | Verification of compound activities using traditional Western blotting. A431 cells were serum-starved for 24 h, pre-incubated with small molecules for 30 min and lysed immediately (0 min) or after treatment with 200 ng/ml of EGF for 5 min. Lysates were analyzed on Western blots using a representative subset of antibodies. '—' indicates no compound. The vertical line separating lanes 2 and 3 on the leftmost blots indicates that intervening lanes on the membrane that were not relevant to this study were removed for clarity of presentation.

Supplementary Table 1 | Antibodies used in the state-based screen.

Antibody	**Vendor**	**Catalog number**	**Antibody incubation time (h)**
anti-EGFR	Santa Cruz Biotechnology	sc-03	1
anti-p-EGFR (pY845)	Biosource International	44-784ZG	1
anti-p-EGFR (pY1068)	Biosource International	44-788ZG	1
anti-p-EGFR (pY1173)	Santa Cruz Biotechnology	sc-12351-R	1
anti-ErbB2	Labvision Corporation	MS-599	1
anti-p-ErbB2 (pY1139)	Abcam	ab5653	1
anti-p-MEK (pS217/pS221)*	Cell Signaling Technology	2354	48
anti-p-Erk (pT202/pY204)*	Cell Signaling Technology	9106	1
anti-p-p90RSK (pS380)	Cell Signaling Technology	9341	48
anti-p-IRS-1 (pY896)	Biosource International	44-818G	1
anti-p-IRS-1 (pY1179)	Biosource International	44-822G	48
anti-p-Akt (pS473)*	Cell Signaling Technology	4058	48
anti-β-actin	Sigma-Aldrich	A1978	48

* numbering of phosphorylation sites is for isoform 1 of the protein. These antibodies also recognize the corresponding epitopes on the other isoforms of the protein.

Supplementary Table 2 | Identity of compounds in the model library.

Compound number	Compound name
1	PD-98059
2	U-0126
3	SB-203580
4	H-7
5	H-9
6	Staurosporine
7	AG-494
8	AG-825
9	Lavendustin A
10	RG-14620
11	Tyrphostin 23
12	Tyrphostin 25
13	Tyrphostin 46
14	Tyrphostin 47
15	Tyrphostin 51
16	Tyrphostin 1
17	Tyrphostin AG 1288
18	Tyrphostin AG 1478
19	Tyrphostin AG 1295
20	Tyrphostin 9
21	Hydroxy-2-naphthalenylmethylphosphonic acid
22	Damnacanthal
23	Piceatannol
24	PP1
25	AG-490
26	AG-126
27	AG-370
28	AG-879
29	LY 294002
30	Wortmannin
31	GF 109203X
32	Hypericin
33	Ro 31-8220
34	Sphingosine
35	H-89
36	H-8
37	HA-1004
38	HA-1077
39	2-Hydroxy-5-(2,5-dihydroxybenzylamino)benzoic acid
40	KN-62
41	KN-93
42	ML-7

43	ML-9
44	2-Aminopurine
45	N9-Isopropyl-olomoucine
46	Olomoucine
47	Iso-olomoucine
48	Roscovitine
49	5-Iodotubercidin
50	LFM-A13
51	SB-202190
52	PP2
53	ZM 336372
54	SU 4312
55	AG-1296
56	GW 5074
57	Palmitoyl-DL-carnitine Cl
58	Rottlerin
59	Genistein
60	Daidzein
61	Erbstatin analog
62	Quercetin dihydrate
63	SU1498
64	ZM 449829
65	BAY 11-7082
66	5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole
67	2,2',3,3',4,4'-Hexahydroxy-1,1'-biphenyl-6,6'-dimethanol dimethyl ether
68	SP 600125
69	Indirubin
70	Indirubin-3'-monooxime
71	Cantharidic acid
72	Cantharidin
73	Endothall
74	Benzylphosphonic acid
75	L-p-Bromotetramisole oxalate
76	RK-682
77	RWJ-60475
78	Levamisole HCl
79	Tetramisole HCl
80	Cypermethrin
81	Deltamethrin
82	Fenvalerate
83	Tyrphostin 8
84	CinnGEL

Supplementary Methods

Reagents. Pan- and phospho-specific antibodies for signalling proteins were purchased from Abcam (Cambridge, MA), BD Biosciences (San Jose, CA), Biosource International (Camarillo, CA), Cell Signaling Technology (Beverly, MA), Lab Vision/Neomarkers (Fremont, CA), Santa Cruz Biotechnology (Santa Cruz, CA) and Upstate (Charlottesville, VA). Mouse monoclonal anti-β-actin antibody (clone AC-15) was purchased from Sigma-Aldrich (Saint Louis, MO). Horseradish peroxidase (HRP) conjugates of secondary antibodies for Western blotting were purchased from Amersham Biosciences (Buckinghamshire, England). Biotinylated secondary antibodies for microarray experiments were obtained from Kirkegaard & Perry Laboratories (Gaithersburg, MD). Streptavidin-HRP and biotinyl-tyramide were part of the tyramide signal amplification (TSA) kit from Perkin Elmer (Boston, MA). Streptavidin-Alexafluor647 was purchased from Molecular Probes/Invitrogen (Carlsbad, CA). Human recombinant EGF was purchased from Upstate.

Cell culture. A431, T-47D, MDA-MB-453, HeLa, A549 and SK-BR-3 cell lines were obtained from American Type Culture Collection (Manassas, VA). Cells were cultured in DMEM (Mediatech, Herndon, VA) supplemented with 10% fetal bovine serum (HyClone, Logan, UT), 2 mM glutamine (Mediatech) and penicillin-streptomycin (Mediatech). For analysis of protein expression across cell lines, cells were grown to approximately 50% confluence, washed twice with ice-cold phosphate-buffered saline (PBS), and lysed in 2% SDS lysis buffer for 30 min on ice. 2% SDS lysis buffer containing protease and phosphatase inhibitors was prepared as described[9]. Samples were

cleared by filtration using 0.2 μm microcentrifuge filter tubes (Millipore, Billerica, MA) and stored at -80 ºC. For time-course stimulations, A431 cells were grown to approximately 50% confluence and serum-starved for 24 h. 200 ng/ml EGF was added and the cells were incubated for 1, 5, 15, 30 or 60 min at 37 ºC. Unstimulated cells served as the 0 min time-point. After incubation, cells were washed and lysed as described above.

Immunoblotting. In order to normalize lysate concentrations across different samples, equal volumes of lysates were loaded onto SDS gels, separated electrophoretically, transferred to PVDF membranes (Millipore) and blocked with PBS supplemented with 5% dry milk (w/v). β-Actin was detected using a mouse anti-β-actin antibody, followed by an HRP-conjugated goat anti-mouse antibody. Blots were developed using the SuperSignal West Femto™ kit (Pierce Biotechnology) and imaged with a ChemiImager 5500 (Alpha Innotech, San Leandro, CA). The bands were quantified using ImageJ (available as freeware from the National Institutes of Health), and the lysates diluted accordingly to equalize concentrations. For analysis of signalling proteins, immunoblotting was carried out by the same method, using lysates with normalized concentrations. For immunoblotting, all primary antibodies were used at 1:1000 dilution in PBS supplemented with 5% BSA (w/v) and 0.1% Tween-20 (v/v), while HRP-conjugates of secondary antibodies were diluted 1:5000 in PBS supplemented with 5% dry milk (w/v) and 0.1% Tween-20 (v/v).

Determination of dose-response behavior. A431 cells were grown, serum-starved and EGF-stimulated in a 96-well tissue culture plate as described above. Instead of a library of compounds, however, 12 different concentrations of each compound (U-0126 and LY294002) were used, ranging from 0.32 nM to 100 μM. Lysate microarrays were fabricated as described above and probed with anti-pT202/pY204-Erk antibody, anti-pS380-p90RSK antibody, anti-pS473-Akt antibody, and anti-β-actin antibody. Data from replicate spots and replicate microarrays were averaged, corrected for non-linearity, and normalized to β-actin signals as described above. Average signals and standard deviations were determined for each concentration of compound from three replicate wells. Data were fit to a Boltzmann sigmoidal curve, and EC_{50} values were determined.

Supplementary Notes

The following is a detailed description of the data processing steps used in the pilot small molecule screen and a treatment of how these steps affect uncertainty in the data.

Data analysis includes the following seven steps (in order):

(1) Data are corrected for nonlinearity introduced by the tyramide signal amplification procedure. This is done using a series of two-fold serial dilutions of a control lysate, printed in quadruplicate at the bottom of each array. For each array, the four replicate spots at each concentration are averaged, and the averaged values are fit to eq. 1:

$$I = -\frac{a}{1+e^{bx-c}} + d \tag{1}$$

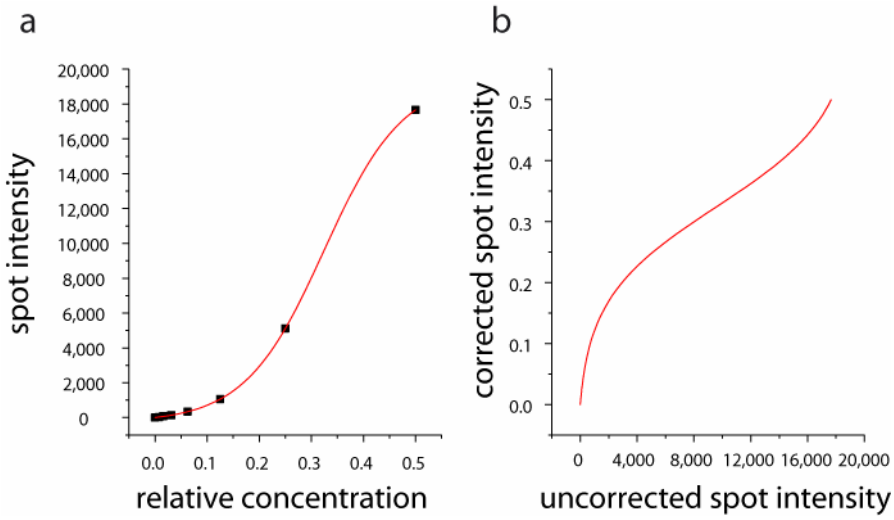
yielding parameters *a*, *b*, *c* and *d*. The inverse function, eq. 2

$$I_{corrected} = \frac{\ln\left(\frac{a}{d - I_{uncorrected}} - 1\right) + c}{b} \tag{2}$$

is then applied to all spots in the array, using parameters *a*, *b*, *c* and *d*. An example is provided below (graph 1), using the dilution series of control lysate for the phospho-Erk array shown in figure 3 of the paper.

(2) All signal intensities in an array are normalized to the mean signal intensity of that array.
(3) All eight replicate spots from two arrays probed with the same antibody are averaged.
(4) Averaged signal intensities are divided by the averaged signal intensity from the duplicate arrays probed with the anti-β-actin antibody.
(5) Edge effects are corrected by scaling the mean intensity of edge wells to the mean intensity of interior wells.
(6) All three replicate biological experiments are averaged.
(7) Relative inhibition values are calculated using eq. 3:

$$relative\ inhibition = \frac{I_{only\ EGF} - I_{compound+EGF}}{I_{only\ EGF} - I_{no\ EGF}} \tag{3}$$



Graph 1 | (**a**) Relative concentrations and spot intensities for the dilution series of control lysate shown in Fig. 2b (phospho-Erk1/2). The data were fit to eq. 1 to obtain parameters *a*, *b*, *c* and *d*. (**b**) Calibration curve for correction of non-linearity, obtained by plotting eq. 2 using parameters *a*, *b*, *c* and *d*.

A treatment of uncertainty in the data.

In the following discussion, propagation of error and potential additional sources of error are discussed for each data processing step:

(1) Taking into account all of the data used in this study, the average coefficient of variation (c.v.) for replicate spots within an array, before applying the correction for non-linearity, is 5.6%. This c.v. was calculated as the ratio of the standard deviation of spot intensity to the absolute spot intensity:

$$c.v.(I) = \frac{\Delta I}{I} \tag{5}$$

Interestingly, the c.v. for replicate spots after correcting for the non-linearity of signal amplification was reduced to, on average, 3.5%. This is because the non-linearity in the signal amplification procedure tends to exaggerate differences in signal intensity, particularly at the extremes of the dynamic range. Correcting for this non-linearity places all spot intensities closer to the mean intensity of the array, thus also reducing the average c.v. for replicate spots.

This does not, however, take into account the unknown error associated with parameters *a*, *b*, *c* and *d*. The main source of error for these values lies in pipetting inaccuracies during the preparation of the control dilution series and manifests itself only when comparing different biological experiments (when printing each

experiment, a fresh dilution series is prepared). Overall, the c.v. of 3.5% reflects the accuracy of the microarrayer in printing consistent volumes of the same sample.

(2) For a function y of n variables $x_1,\ldots,x_n$,

$$y = f\left(x_1,...,x_n\right) \tag{6}$$

the standard deviation of y (Δy) can be obtained from the standard deviations of the basic variables x_i (Δx_i) assuming Gaussian error propagation:

$$\Delta y = \sqrt{\sum_{i=1}^{n}\left(\frac{\partial y}{\partial x_i}\right)^2 \cdot \left(\Delta x_i\right)^2} \tag{7}$$

Using eq. 7, the c.v. associated with the mean value of each array can be calculated as

$$c.v.\left(I_{mean}\right) = \frac{1}{\sqrt{384}} \cdot c.v.\left(I_{spot}\right) \tag{8}$$

After dividing each spot intensity by the mean spot intensity, we obtain, using eq. 7:

$$c.v.\left(I_{spot,\,normalized}\right) = \sqrt{c.v.\left(I_{spot}\right)^2 + c.v.\left(I_{mean}\right)^2} = \sqrt{1+\frac{1}{384}} \cdot c.v.\left(I_{spot}\right) \cong 3.5\% \tag{9}$$

(3) The c.v. for spots in both replicate arrays depends largely on pipetting inaccuracies during the probing steps, and can therefore not be calculated from the previous errors. In our data set, we determined this value empirically to be 5.0%. The c.v. of the mean of all eight spots is then given by:

$$c.v.\left(I_{mean\ of\ 8}\right) = \frac{1}{\sqrt{8}} \cdot c.v.\left(I_{spot}\right) \cong 1.8\% \tag{10}$$

(4) Coefficients of variation after dividing by the mean β-actin signals can be calculated as:

$$c.v.\left(I_{mean\ of\ 8,\ corrected\ for\ actin}\right) = \sqrt{\left(c.v.\left(I_{mean\ of\ 8}\right)\right)^2 + \left(c.v.\left(I_{actin,\,mean\ of\ 8}\right)\right)^2} \cong 2.5\% \tag{11}$$

(5) For correction of edge effects, the average of 6 or 12 edge wells is calculated, and scaled to the average of the 60 interior wells. The resulting c.v.'s are thus:

$$c.v.\left(I_{edge,\,group\ of\ 6}\right) = \sqrt{\left(\frac{1}{\sqrt{6}} \cdot c.v.\left(I_{well}\right)\right)^2 + \left(\frac{1}{\sqrt{60}} \cdot c.v.\left(I_{well}\right)\right)^2} \cong 4.3\% \tag{12}$$

and

$$c.v.\left(I_{edge,\,group\,of\,12}\right)=\sqrt{\left(\frac{1}{\sqrt{12}}\cdot c.v.\left(I_{well}\right)\right)^2+\left(\frac{1}{\sqrt{60}}\cdot c.v.\left(I_{well}\right)\right)^2}\cong 3.2\% \qquad (13)$$

The c.v. for the interior wells remains unchanged (2.5%).

(6) The c.v. for replicate biological experiments cannot be calculated from the previous error values, since most variation between experiments arises from slight differences in pipetting, temperature, incubation time, etc. We therefore empirically determined the c.v. of three replicate biological experiments to be 12.4% (an average over the entire data set).

(7) Uncertainty for relative inhibition values is highly dependent on the degree of signal compression of each antibody. For highly cross-reactive antibodies, the two terms $I_{only\,EGF}$ and $I_{no\,EGF}$ in eq. 3 are similar, resulting in a high error, while highly selective antibodies show a wide spread between $I_{only\,EGF}$ and $I_{no\,EGF}$, and hence a low error for the resulting relative inhibition values. Average standard deviations over all 84 compounds ranged from 0.19 for pT202/pY204-Erk1/2 to 0.79 for pY845-EGFR. Coefficients of variation are not meaningful for relative inhibition values, as these lie around zero for inactive compounds, resulting in potentially infinitely high c.v.'s.

In summary, we conclude that, for the known initial errors (spot-to-spot and array-to-array variation), these errors have very little impact on the final relative inhibition values, even though the data are subjected to a substantial number of processing steps. This is because each step averages over a progressively larger number of previous values, thus effectively reducing (rather than increasing) c.v.'s. In contrast, initial errors of pipetting and biological variation, which are only apparent when comparing replicate experiments, are substantially larger. This emphasizes the importance of basing our clustering algorithm on the standard deviations among replicate experiments, to ensure conclusions are based on high-quality data points rather than assay noise.