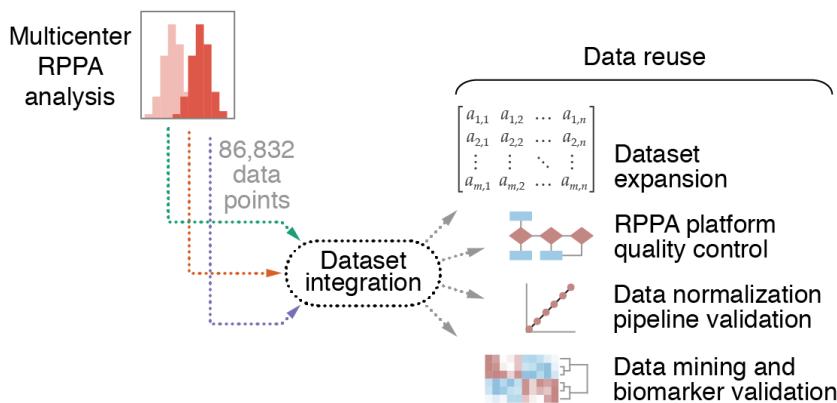


# Multicenter reverse-phase protein array data integration

Leanne de Koning<sup>1</sup>, Stephan Bernhardt<sup>2</sup>, Kenneth G. Macleod<sup>3</sup>, Bérengère Ouine<sup>1</sup>, Aurélie Cartier<sup>1</sup>, Vonick Sibut<sup>4</sup>, Neil O. Carragher<sup>3</sup>, Ulrike Korf<sup>2</sup>, Bryan Serrels<sup>3</sup> & Adam Byron<sup>3\*</sup>

Among the technologies available for protein biomarker discovery and validation, reverse-phase protein array (RPPA) benefits from unequalled sample throughput. Panels of high-quality antibodies enable the quantification by RPPA of protein abundance and posttranslational modifications in biological specimens with high precision and sensitivity. Incorporation of RPPA technology into clinical and drug development pipelines requires robust assays that generate reproducible results across multiple laboratories. We implemented the first international multicenter pilot study to investigate RPPA workflow variability. We characterized the proteomic responses of a series of breast cancer cells to two cancer drugs. This analysis quantified 86,832 sample spots, representing 108 biological samples, arrayed at three independent RPPA platforms. This unique integrated set of data is publicly available as a resource to the proteomic and cancer research communities to catalyse further analysis and investigation. We anticipate that this dataset will form a reference for the comparison of RPPA workflows and reagents, which can be expanded in the future, and will aid the identification of platform-robust treatment-marker antigens in breast cancer cells.



<sup>1</sup>Department of Translational Research, Institut Curie, PSL Research University, Paris, France. <sup>2</sup>Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>3</sup>Cancer Research UK Edinburgh Centre, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom. <sup>4</sup>U900 INSERM, Institut Curie, PSL Research University, Paris, France. Present addresses: S.B., Pfizer Pharma GmbH, Berlin, Germany. A.C., Sederma, Le Perray-en-Yvelines, France. B.S., NanoString Technologies, Inc., Seattle, Washington, United States of America. \*Correspondence should be addressed to A.B. (email: [adam.byron@ed.ac.uk](mailto:adam.byron@ed.ac.uk)).

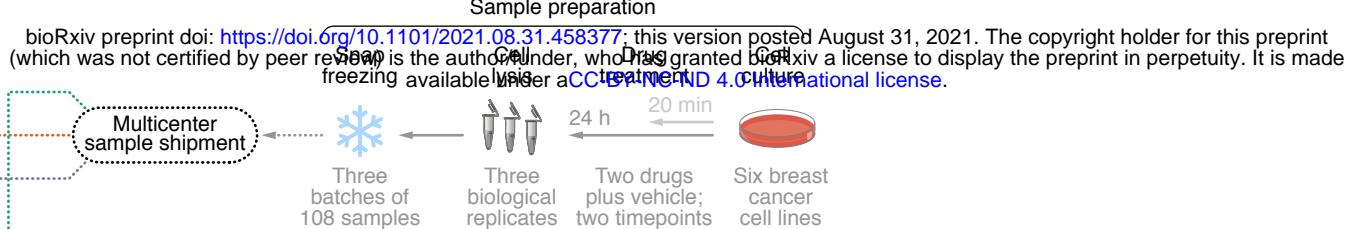
## Background & Summary

In oncology, the therapeutic potential of personalized medicine has been realized for many patients. The targeted therapies imatinib and trastuzumab are examples of drugs that have revolutionized patient survival in certain types of cancer<sup>1,2</sup>. Yet, besides a few success stories, there is little evidence that treatment selection based on genetic alterations is of real benefit to patients<sup>3,4</sup>, possibly because the vast majority of targeted therapies target proteins, and most frequently the active form of the protein. Activation of proteins and their associated signaling pathways often cannot be predicted from genetic alterations. For example, in breast cancer, a PIK3CA mutation is not necessarily associated with activation of the phosphoinositide 3-kinase–Akt pathway<sup>5,6</sup>. Moreover, if parallel signaling pathways are activated in the tumor, the patient will show resistance to the administered targeted therapy. An emerging consensus view has therefore identified the need to include proteomic analyses in drug development pipelines and molecular screening programs<sup>7–11</sup>. However, successful implementation of this requires robust, reproducible, and affordable high-throughput proteomics technologies.

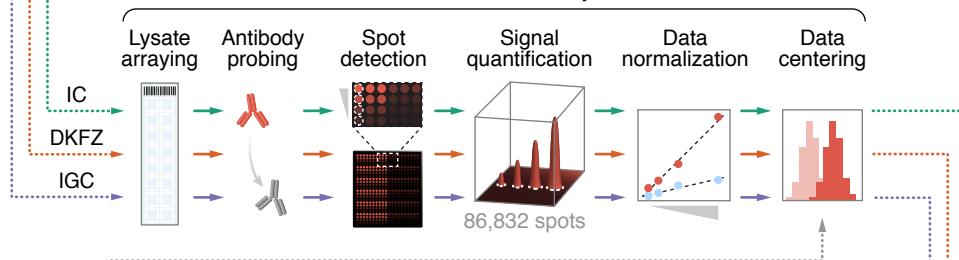
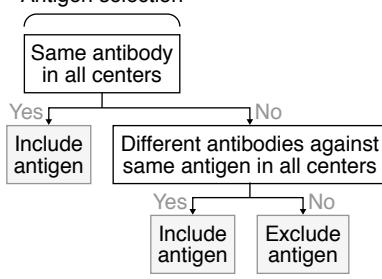
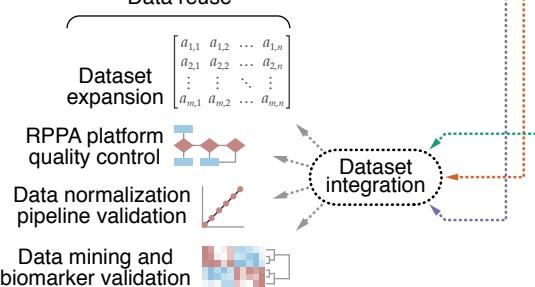
Reverse-phase protein array (RPPA) technology is a high-throughput dot-blot technology that uses panels of high-quality antibodies to enable the quantification of protein abundance, and that of posttranslational modifications, from a limited amount of biological sample<sup>12</sup>. The technology offers precise and sensitive multiplexed quantification of the abundance of proteins and posttranslational modifications in hundreds of arrayed samples, providing the opportunity for high-throughput proteomic profiling of cells and tissues<sup>13–15</sup>. RPPA therefore provides a valuable tool for the targeted analysis of disease-relevant signaling mechanisms, including in cancer<sup>16–35</sup>.

To date, RPPA technology has been, and is currently being, applied in several cancer clinical trials<sup>11,36</sup>, mostly retrospectively to detect markers of response and resistance<sup>37–45</sup>, but in some cases also prospectively to select the most appropriate therapy for a given patient<sup>46,47</sup>. As RPPA technology is readying itself for further implementation in drug development and clinical laboratory settings, there was a demand to evaluate the robustness and reproducibility of RPPA datasets derived from different RPPA platforms.

To address this need for crossplatform validation of RPPA technology, we conducted the first international multicenter pilot study to investigate RPPA workflow variability and potential for multiplatform data integration. With the overarching goal of learning from distinct analysis pipelines and improving RPPA technology, the objectives of this open, collaborative project were to acquire, normalize, and integrate data from multiple RPPA platforms to facilitate crossplatform data comparison. We characterized the protein-level responses of six breast cancer cell lines to two clinically relevant cancer drugs at two treatment timepoints, generating a total of 108 biological samples, incorporating 36 experimental conditions, which were normalized and analyzed at three different research centers (Fig. 1a–c). The dataset consists of original RPPA slide scan images and image quantification of 86,832 sample spots derived from 972 arrayed lysates, incorporating independent biological replicate samples, serial lysate dilution series, and technical replicate sample spots, from three independent RPPA platforms. Integrative analysis of a subset of these data enabled the identification of platform-independent protein markers of breast cancer cell response to drug treatment<sup>48</sup>. We report here the raw and processed data from our assessment of multiple RPPA workflows at different research centers. Our data will serve as a starting point for the appraisal of the reproducibility of RPPA technology and its capacity to identify robust protein markers of response to cancer therapies. The data can be reused by multiple researchers over the world to implement and validate their RPPA quality control, data normalization, and data integration methods. Moreover, the dataset can be expanded

**a****b**

## Multicenter RPPA analysis

**c** Antigen selection**d** Data reuse**Figure 1.** Schematic overview of the sample preparation and data analysis workflows.

(a) Sample preparation workflow. Six breast cancer cell lines were treated with lapatinib, selumetinib, or DMSO (vehicle) for 20 min or 24 h. Cells were lysed ( $n = 3$  biological replicates), and cell lysates were snap frozen and sent to the different research centers.

(b) Multicenter RPPA analysis workflow. Cell lysates were arrayed and processed for RPPA analysis using the platform-specific setups at the respective research centers [IC, Institut Curie (platform 1); DKFZ, Deutsches Krebsforschungszentrum (platform 2); IGC, Institute of Genetics and Cancer, University of Edinburgh (platform 3)].

(c) Antigen selection decision tree. Data derived from validated antibodies targeting the same antigen, including different antibodies from different suppliers, acquired at all three research centers were used for dataset integration. (d) Data reuse exemplar workflow. Normalized RPPA data from each research center were centered and integrated to enable various modes of data reuse. **a** and **b** were adapted from Fig. 1a in our related work<sup>48</sup>.

to support a framework for the integration, interrogation, and interpretation of RPPA data derived from multiple research centers (Fig. 1d).

## Methods

### Experimental design

For this multicenter RPPA study, six breast cancer cell lines were cultured in the presence or absence of two kinase inhibitors for two time periods. For each experimental condition, three biological replicate cell lysates were generated, frozen, and sent to three different research centers for RPPA analysis (Fig. 1a, b). Cell lysates were processed for RPPA analysis using the platform-specific setups at the respective research centers, including incorporation of serial lysate dilution series and technical replicate sample spots (detailed below). Raw RPPA data were processed and normalized according to the respective in-house data analysis pipelines. Each antibody used was assigned a unique antibody identifier linked to a Research Resource Identifier (RRID) (Online-only Table 1). Normalized data derived from validated antibodies targeting the same antigen acquired at all three research centers were integrated to support data reuse (Fig. 1b–d).

Some of the following methods are expanded versions of descriptions in our related work<sup>48</sup>.

### Cell culture and drug treatment

HCC1954 (RRID CVCL\_1259), MCF7 (RRID CVCL\_0031), MDA-MB-231 (RRID CVCL\_0062), MDA-MB-453 (RRID CVCL\_0418), MDA-MB-468 (RRID CVCL\_0419), and SKBR3 (RRID CVCL\_0033) breast cancer cells (Table 1) were purchased from American Type Culture Collection (LGC Standards, Molsheim, France) and were cultured according to the supplied instructions. Cells were routinely tested for mycoplasma and used within three months of recovery from frozen. Lapatinib [Selleck Chemicals, Houston, TX, USA, catalog number (cat. no.) S2111], a dual epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor 2 (HER2; also known as ErbB2) inhibitor, and selumetinib (Selleck Chemicals, cat. no. S1008), a mitogen-activated protein kinase kinase (MEK) inhibitor, were prepared as 10 mM stock solutions in DMSO (Online-only Table 2). Cells were treated with 1  $\mu$ M lapatinib, 1  $\mu$ M selumetinib, or DMSO (vehicle control) in growth medium for 20 min or 24 h.

### Cell lysis

Following treatment with drug or vehicle control, cells were washed twice with ice-cold phosphate-buffered saline (PBS). Cells were lysed in hot Laemmli buffer [50 mM Tris-HCl (pH 6.8), 2% (w/v) sodium dodecyl sulfate, 5% (w/v) glycerol, 2 mM DTT, 2.5 mM EDTA, and 2.5 mM EGTA], supplemented with 4 mM sodium orthovanadate, 20 mM sodium fluoride, 1 $\times$  Perbio Halt phosphatase inhibitor cocktail (Thermo Fisher Scientific, Waltham, MA, USA, cat. no. 78420), and 1 $\times$  Roche cComplete protease inhibitor cocktail (Sigma-Aldrich, Gillingham, UK, cat. no. 11697498001), for 10 min at 100°C. Lysates were passed through a 25-gauge needle five times and then clarified by centrifugation at 18,000  $\times$  g for 10 min at room temperature. Protein concentration was determined for each sample using a Pierce reducing agent-compatible BCA protein assay kit (Thermo Fisher Scientific, cat. no. 23250).

Clarified lysates were aliquoted and snap-frozen in liquid nitrogen prior to shipment to the participating research centers on dry ice. Samples prepared in biological triplicate were analyzed at each research center using the respective in-house RPPA platforms (platform 1, Institut Curie; platform 2, Deutsches Krebsforschungszentrum; platform 3, Institute of Genetics and Cancer, University of Edinburgh) (Table 2).

Cell type	Breast cancer subtype	available under aCC-BY-NC-ND 4.0 International license. Protein expression				<i>ERBB2</i> amplification	Drug treatments
		ER	PR	E-cad.	EGFR-high		
HCC1954	Basal-like HER2-positive	–	–	+	+	+	Lapatinib (20 min, 24 h) Selumetinib (20 min, 24 h) DMSO (20 min, 24 h)
MCF7	Luminal A	+	+	+	–	–	Lapatinib (20 min, 24 h) Selumetinib (20 min, 24 h) DMSO (20 min, 24 h)
MDA-MB-231	Claudin-low	–	–	–	+	–	Lapatinib (20 min, 24 h) Selumetinib (20 min, 24 h) DMSO (20 min, 24 h)
MDA-MB-453	HER2-positive	–	–	–	–	+	Lapatinib (20 min, 24 h) Selumetinib (20 min, 24 h) DMSO (20 min, 24 h)
MDA-MB-468	Basal-like 1	–	–	+	+	–	Lapatinib (20 min, 24 h) Selumetinib (20 min, 24 h) DMSO (20 min, 24 h)
SKBR3	Luminal-like HER2-positive	–	–	–	+	+	Lapatinib (20 min, 24 h) Selumetinib (20 min, 24 h) DMSO (20 min, 24 h)

**Table 1. Breast cancer cell types and drug treatments used to generate samples analyzed in this study.** Receptor status of breast cancer cell types is indicated. The breast cancer subtype for MDA-MB-231 cells is also referred to as mesenchymal stem-like. For MDA-MB-453 cells, the *ERBB2* gene, which encodes HER2, is amplified but HER2 is not overexpressed; its breast cancer subtype is sometimes referred to as luminal androgen receptor. E-cad., E-cadherin; EGFR, epidermal growth factor receptor; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; PR, progesterone receptor.

Platform	Center	Slides	Slide labeling	Signal readout	Scanner	Spot detection	Normalization
1	IC	Grace Bio-Labs ONCYTE SuperNOVA nitrocellulose film	DAKO autostainer	Alexa Fluor 647	GenePix 400B	MicroVigene	SYPRO Ruby and NormaCurve
2	DKFZ	Grace Bio-Labs ONCYTE AVID nitrocellulose film	By hand	Alexa Fluor 680	Odyssey	GenePix Pro	Fast Green FCF and RPPAnalyzer
3	IGC	Grace Bio-Labs ONCYTE AVID nitrocellulose film	By hand	DyLight 800	InnoScan 710	Mapix	Fast Green FCF

**Table 2. RPPA workflows used at each center.** All research centers used the 2470 Arrayer for slide printing. Only platform 1 used signal amplification (Bio-Rad Amplification Reagent). For further details, including manufacturer information, see the Methods section. IC, Institut Curie; DKFZ, Deutsches Krebsforschungszentrum; IGC, Institute of Genetics and Cancer, University of Edinburgh. This table was adapted from Supplementary Table 2 in our related work<sup>48</sup>.

## RPPA sample handling

**Platform 1.** The protein concentration of all samples was adjusted to 1.0 mg/ml. Samples were then serially diluted in lysis buffer to produce four serial two-fold dilutions of each sample (1 mg/ml, 0.5 mg/ml, 0.25 mg/ml, and 0.125 mg/ml concentrations). Samples were printed onto Grace Bio-Labs ONCYTE SuperNOVA nitrocellulose film slides (Sigma-Aldrich, cat. no. 705170) under conditions of constant 70% relative humidity using an Aushon BioSystems 2470 microarrayer (Quanterix, Billerica, MA, USA). Two technical replicate spots were printed per sample dilution.

**Platform 2.** Vehicle control samples only were serially diluted in lysis buffer to produce six serial two-fold dilutions of each sample (1.8 mg/ml, 0.9 mg/ml, 0.45 mg/ml, 0.225 mg/ml, 0.1125 mg/ml, and 0.05625 mg/ml concentrations). The protein concentration of all samples was then adjusted to 1.5 mg/ml. Samples were printed onto Grace Bio-Labs ONCYTE AVID nitrocellulose film slides (Sigma-Aldrich, cat. no. GBL305108) under conditions of constant 70% relative humidity using an Aushon BioSystems 2470 microarrayer. Three technical replicate spots were printed per sample dilution.

**Platform 3.** The protein concentration of all samples was adjusted to 1.5 mg/ml. Samples were then serially diluted in PBS containing 10% (w/v) glycerol to produce four serial two-fold dilutions of each sample (1.5 mg/ml, 0.75 mg/ml, 0.375 mg/ml, and 0.1875 mg/ml concentrations). Samples were printed onto Grace Bio-Labs ONCYTE AVID nitrocellulose film slides under conditions of constant 70% relative humidity using an Aushon BioSystems 2470 microarrayer. Three technical replicate spots were printed per sample dilution.

## RPPA sample analysis

**Platform 1.** Slides were hydrated in deionized water (four 15-min washes), washed with Tris-buffered saline (TBS) containing 0.1% (w/v) Tween 20 (TBS-T) (three 5-min washes), and incubated with TBS-T containing 5% (w/v) bovine serum albumin (TBS-T-BSA) for 10 min. Slides were washed with TBS-T (three 5-min washes), incubated with Dako biotin (cat. no. X0590) and peroxidase (cat. no. S200389) blocking reagents (both Agilent, Santa Clara, CA, USA), and incubated with primary antibodies (diluted in TBS-T-BSA; Online-only Table 1) for 60 min. Slides were washed with TBS-T (three 5-min washes), incubated with TBS-T-BSA for 10 min, and washed with TBS-T (three 5-min washes). Bound antibodies were detected by incubation with horseradish peroxidase-conjugated anti-mouse (cat. no. 115-035-062; RRID AB\_2338504) or anti-rabbit (cat. no. 111-035-045; RRID AB\_2337938) secondary antibodies (both Jackson ImmunoResearch Europe, Ely, UK) for 60 min. To amplify the signal, slides were incubated with Bio-Rad Amplification Reagent (Bio-Rad, Watford, UK, cat. no. 1708230) for 15 min. The arrays were washed with TBS-T (three 5-min washes), probed with Alexa Fluor 647-coupled streptavidin (Thermo Fisher Scientific, cat. no. S32357), incubated with TBS-T-BSA for 60 min, and washed in TBS-T (three 5-min washes). Slides were washed with deionized water (one 5-min wash), spun at 2,000 r.p.m. for 5 min, and stored in the dark.

For normalization against total protein, an additional slide was incubated with destain solution 1 [7% (v/v) acetic acid, 10% (v/v) methanol in deionized water] for 15 min and washed with deionized water (two 5-min washes). The slide was then incubated with SYPRO Ruby staining solution (Thermo Fisher Scientific, cat. no. S11791) for 5 min, washed with deionized water (five 5-min washes), spun at 2,000 r.p.m. for 5 min, and stored in the dark. All steps were performed at room temperature with agitation on a rocking platform.

**Platform 2.** Slides were hydrated in deionized water (four 15-min washes), washed with TBS-T (three 5-min washes), and incubated with 50% (v/v) blocking buffer for fluorescent western blotting (Rockland Immunochemicals, Limerick, PA, USA, cat. no. MB-070) in TBS containing 5 mM sodium fluoride and 1 mM sodium orthovanadate (hereafter, Rockland blocking solution) for 2 h. Slides were washed with TBS-T (four 5-min washes) and incubated with primary antibodies (diluted in Rockland blocking solution; Online-only Table 1) overnight at 4°C with agitation on a rocking platform. Slides were washed with TBS-T (four 5-min washes). Bound antibodies were detected by incubation with Alexa Fluor 680-conjugated F(ab')<sub>2</sub> fragments of anti-mouse IgG (cat. no. A-21059; RRID AB\_2535725) or anti-rabbit IgG (cat. no. A-21077; RRID AB\_2535737) secondary antibodies (both Thermo Fisher Scientific) (diluted 1:8,000 in TBS-T) for 1 h. Slides were washed with TBS-T (four 5-min washes), washed with deionized water (one 5-min wash), air dried for 10 min at 30°C in the dark, and stored at room temperature in the dark.

For normalization against total protein, an additional slide was washed with TBS for 1 min and incubated with 0.005% (w/v) Fast Green FCF (Sigma-Aldrich, cat. no. F7258) in destain solution 2 [10% (v/v) acetic acid, 30% (v/v) ethanol in deionized water] for 45 min. The slide was then incubated with destain solution 2 (two 15-min incubations), washed with deionized water (two 5-min washes), air dried for 10 min at 30°C in the dark, and stored at room temperature in the dark. Unless otherwise stated, all steps were performed at room temperature with agitation on a rocking platform.

**Platform 3.** Slides were hydrated in deionized water (four 15-min washes), incubated with ReBlot Plus strong antigen retrieval agent (Merck, Watford, UK, cat. no. 2504) for 15 min, washed with PBS containing 0.1% (w/v) Tween 20 (PBS-T) (two 5-min washes), and incubated with SuperBlock T20 (TBS) blocking buffer (Thermo Fisher Scientific, cat. no. 37536) for 10 min. Slides were washed with TBS-T (two 5-min washes) and incubated with primary antibodies (diluted in SuperBlock T20 blocking buffer; Online-only Table 1) for 60 min. Slides were washed with TBS-T (two 5-min washes), incubated with SuperBlock T20 blocking buffer for 10 min, and washed with TBS-T (three 5-min washes). Bound antibodies were detected by incubation with DyLight 800-conjugated anti-mouse IgG (cat. no. 5257; RRID AB\_10693543) or anti-rabbit IgG (cat. no. 5151; RRID AB\_10697505) secondary antibodies (both Cell Signaling Technology) (diluted 1:2,500 in SuperBlock T20 blocking buffer) for 30 min. Slides were washed with TBS-T (two 5-min washes), washed briefly with deionized water, spun at 2,000 r.p.m. for 5 min, and stored in the dark.

For normalization against total protein, an additional slide was washed with deionized water (one 5-min wash), incubated with 1% (w/v) NaOH for 15 min, rinsed with deionized water (10 rinses), washed with deionized water (one 10-min wash), and incubated with destain solution 3 [7% (v/v) acetic acid, 30% (v/v) methanol in deionized water] for 15 min. The slide was then incubated with 0.0025% (w/v) Fast Green FCF in destain solution 3 for 3 min, rinsed with deionized water (10 rinses), incubated with destain solution 3 for 15 min, rinsed with deionized water (10 rinses), spun at 2,000 r.p.m. for 5 min, and stored in the dark. All steps were performed at room temperature with agitation on a rocking platform.

### RPPA data analysis

**Platform 1.** Slides were visualized using a GenePix 4000B microarray scanner (Molecular Devices, San Jose, CA, USA) at an excitation wavelength of 635 nm and a resolution of 10 μm, and images were acquired as tagged image file format (TIFF) files (stored in Images1\_IC.zip)<sup>49</sup>. Scanner settings were chosen to obtain around 10–15% of

saturated spots, which improves subsequent data normalization. Signals were quantified using MicroVigene microarray image analysis software (VigeneTech, Carlisle, MA, USA). Nonspecific signals were determined by omitting the primary antibody incubation step. Means of technical replicate sample spots were calculated, and data were written to a comma-separated values (CSV) file (SampleQuantification1\_IC.csv)<sup>49</sup>. Data were normalized using NormaCurve<sup>50</sup>, which normalizes for fluorescence background per spot and total protein stain and defines a log<sub>2</sub>-transformed single expression value for each biological sample, based on the entire serial dilution curve and all technical replicates. Next, each RPPA slide was median centered and scaled (divided by median absolute deviation). We then corrected for remaining sample loading effects individually for each array by correcting the dependency of the data for individual arrays on the median value of each sample over all the arrays using linear regression. Normalized, log<sub>2</sub>-transformed data were written to a CSV file (NormalizedData1\_IC.csv)<sup>49</sup>.

**Platform 2.** Slides were visualized using an Odyssey near-infrared microarray scanner (LI-COR Biotechnology, Bad Homburg, Germany) at an excitation wavelength of 685 nm and a resolution of 21  $\mu$ m, and images were acquired as TIFF files (stored in Images2\_DKFZ.zip)<sup>49</sup>. Scanner settings were chosen to enable the highest gain without saturation of the signal. Signals were quantified using GenePix Pro microarray image analysis software (version 7.0) (Molecular Devices). Nonspecific signals were determined by omitting the primary antibody incubation step. The linear fit of the relative fluorescence intensities for the dilution series of each vehicle control-treated sample was determined for each primary antibody, and >99% of signals derived from neat samples (1.5 mg/ml) were within the linear range of detection<sup>48</sup>. Medians of technical replicate sample spots were calculated, background was corrected, and quality control was performed using RPPAnalyzer<sup>51</sup>, and data were written to a CSV file (SampleQuantification2\_DKFZ.csv)<sup>49</sup>. Data were normalized for protein loading using RPPAnalyzer, log<sub>2</sub> transformed, and written to a CSV file (NormalizedData2\_DKFZ.csv)<sup>49</sup>.

**Platform 3.** Slides were visualized using an InnoScan 710-IR infrared microarray scanner (Innopsys, Carbone, France) at an excitation wavelength of 785 nm and a resolution of 10  $\mu$ m, and images were acquired as TIFF files (stored in Images3\_IGC.zip)<sup>49</sup>. Following a preview scan to assess signal saturation, scanner settings were chosen to enable the highest laser power and gain with no saturation of the signal. Signals were quantified using Mapix microarray image analysis software (version 6.5.0) (Innopsys). Nonspecific signals were determined by omitting the primary antibody incubation step. The linear fit of the relative fluorescence intensities for the dilution series of each sample was determined for each primary antibody, to assess whether signals were within the linear range of detection with  $R^2 > 0.9$ . Mean pixel values of sample spots were corrected by subtracting the median pixel value of the local background surrounding each spot (excluding a 2-pixel region at the spot–background border) using Mapix. Means of technical replicate sample spots were calculated, and data were written to a CSV file (SampleQuantification3\_IGC.csv)<sup>49</sup>. Data were normalized for protein loading using the corresponding total protein stain. After data normalization, two data points were negative [cleaved PARP (antibody identifier PARP\_cleaved\_a) probing of MDA-MB-231 cells treated with vehicle control for 24 h (biological replicate 2) and with selumetinib for 20 min (biological replicate 3)]; to prevent an error upon log<sub>2</sub> transformation, these data points were imputed with  $\sim 0.98 \times$  column minimum (0.0078125). Imputed, normalized data were log<sub>2</sub> transformed and written to a CSV file (NormalizedData3\_IGC.csv)<sup>49</sup>.

## Dimensionality reduction

For t-distributed stochastic neighbor embedding (t-SNE), RPPA data were scaled and subjected to principal component analysis, from which the leading 50 principal components were retained for t-SNE. The Barnes–Hut tradeoff parameter,  $\theta$ , was 0.1, the exaggeration factor was 4, perplexity was 10, and the maximum number of optimization iterations was 5,000. For reproducibility, random seed 1 was used.

For uniform manifold approximation and projection (UMAP), RPPA data were scaled and Euclidean distances between sample points were computed. The local neighborhood used for manifold approximation was 15 neighboring sample points. Fuzzy set operation was 1.0 (pure fuzzy union). Local connectivity was 1, and negative sample rate was 5, with weight  $\gamma = 1.0$ . The effective minimum distance between embedded points was 0.1, and the effective scale of embedded points (spread function) was 1.0. The number of training epochs for optimizing the low-dimensional embedding was 1,000. For reproducibility, random seed 1 was used.

## Multicenter dataset integration

The  $\log_2$  transforms of data normalized using the methods employed by the respective research center were median-subtracted antibody wise and concatenated into a single matrix. To enable comparison across all three research centers, data were filtered for antibodies against the same target (i.e. protein or phosphoprotein) probed at all centers. To consolidate antibody crossreactivity, antibodies recognizing related or multiple protein isoforms (e.g. MEK1 and MEK1/2) were considered to recognize the same target family (e.g. MEK1/2) and were thus included in the integrated dataset if the respective target family was probed at all centers. The final concatenated matrix was written to a CSV file (IntegratedData.csv)<sup>49</sup>.

## Cluster analysis

Agglomerative hierarchical cluster analysis of Spearman rank correlation coefficients was performed using Cluster 3.0 (C Clustering Library, version 1.54)<sup>52</sup>. Spearman rank correlation coefficient-based distance matrices were constructed using pairwise average linkage. Clustering results were visualized using Java TreeView (version 1.1.5r2)<sup>53</sup>.

## Statistical analysis

Three independent biological replicate samples of 36 experimental conditions were analyzed at each research center. No statistical methods were used to predetermine sample size. The median absolute deviation (MAD) was defined as  $MAD = \text{median}[|x_i - \text{median}(x)|]$  for a set of biological replicate samples or correlation coefficients  $x_1, x_2, \dots, x_n$ . Robust coefficient of variation based on the median absolute deviation ( $RCV_M$ ) was defined as  $RCV_M = 1.4826 \times MAD / \text{median}(x)$ , where the multiplier 1.4826 is a scaling factor to adjust for asymptotically normal consistency. Pearson correlation coefficient, coefficient of determination, Spearman rank correlation coefficient, and Euclidean distance were calculated for every pairwise combination of biological samples measured by RPPA.

## Data Records

RPPA data records generated at each research center and the integrated RPPA dataset are summarized in Table 3. RPPA slide images generated at each research center are available as TIFF files via figshare<sup>49</sup>. Associated slide descriptor metadata, including antibody position details, and sample descriptor metadata, including spot position details, are available as tab-delimited text files<sup>49</sup>. Sample quantification data files (CSV files) contain background-

Platform	Center	Samples	Serial dilutions	Technical replicates	Abs	Data files	DOI
1	IC	108	4	2	42	61 TIFF files stored in Images1_IC.zip (slide scan images) <sup>49</sup> SlideDescriptor1_IC.txt (slide descriptor) <sup>49</sup> SampleDescriptor1_IC.txt (sample descriptor) <sup>49</sup> SampleQuantification1_IC.csv (sample quantification) <sup>49</sup> NormalizedData1_IC.csv (normalized data) <sup>49</sup> CenteredData1_IC.csv (centered data) <sup>49</sup>	10.6084/m9.figshare.14754069
2	DKFZ	108	1	3	44	18 TIFF files stored in Images2_DKFZ.zip (slide scan images) <sup>49</sup> SlideDescriptor2_DKFZ.txt (slide descriptor) <sup>49</sup> SampleDescriptor2_DKFZ.txt (sample descriptor) <sup>49</sup> SampleQuantification2_DKFZ.csv (sample quantification) <sup>49</sup> NormalizedData2_DKFZ.csv (normalized data) <sup>49</sup> CenteredData2_DKFZ.csv (centered data) <sup>49</sup>	10.6084/m9.figshare.14754069
3	IGC	108	4	3	28	17 TIFF files stored in Images3_IGC.zip (slide scan images) <sup>49</sup> SlideDescriptor3_IGC.txt (slide descriptor) <sup>49</sup> SampleDescriptor3_IGC.txt (sample descriptor) <sup>49</sup> SampleQuantification3_IGC.csv (sample quantification) <sup>49</sup> NormalizedData3_IGC.csv (normalized data) <sup>49</sup> CenteredData3_IGC.csv (centered data) <sup>49</sup>	10.6084/m9.figshare.14754069

Continued on next page.

1–3	Integrated dataset	108	–	–	53	IntegratedData.csv (integrated data) <sup>49</sup>	10.6084/m9.figshare.14754069
-----	--------------------	-----	---	---	----	--	------------------------------

**Table 3. Data files generated at each center.** A total of 108 biological samples were analyzed at each of the research centers using their in-house RPPA platforms. The dataset comprises RPPA slide scan images and image quantification and normalization of 86,832 sample spots derived from 972 arrayed lysates in total, incorporating independent biological replicate samples, serial lysate dilution series, and technical replicate sample spots, from three independent RPPA platforms. For platform 2, a control dilution series of vehicle control-treated samples only was used (not included in sample spot sum total). Abs, antibodies used; IC, Institut Curie; DKFZ, Deutsches Krebsforschungszentrum; IGC, Institute of Genetics and Cancer, University of Edinburgh.

corrected RPPA intensity data for all biological samples<sup>49</sup>. Normalized data files (CSV files) contain log<sub>2</sub>-transformed data normalized using the methods employed by the respective research center<sup>49</sup>. Centered data files (CSV files) contain antibody-wise median-centered normalized data<sup>49</sup>. The integrated dataset, compiled from the centered data files, is available as a CSV file<sup>49</sup>. A README markdown file (README.md), also available as a plain-text file (README.txt), accompanies the data records<sup>49</sup>.

## Technical Validation

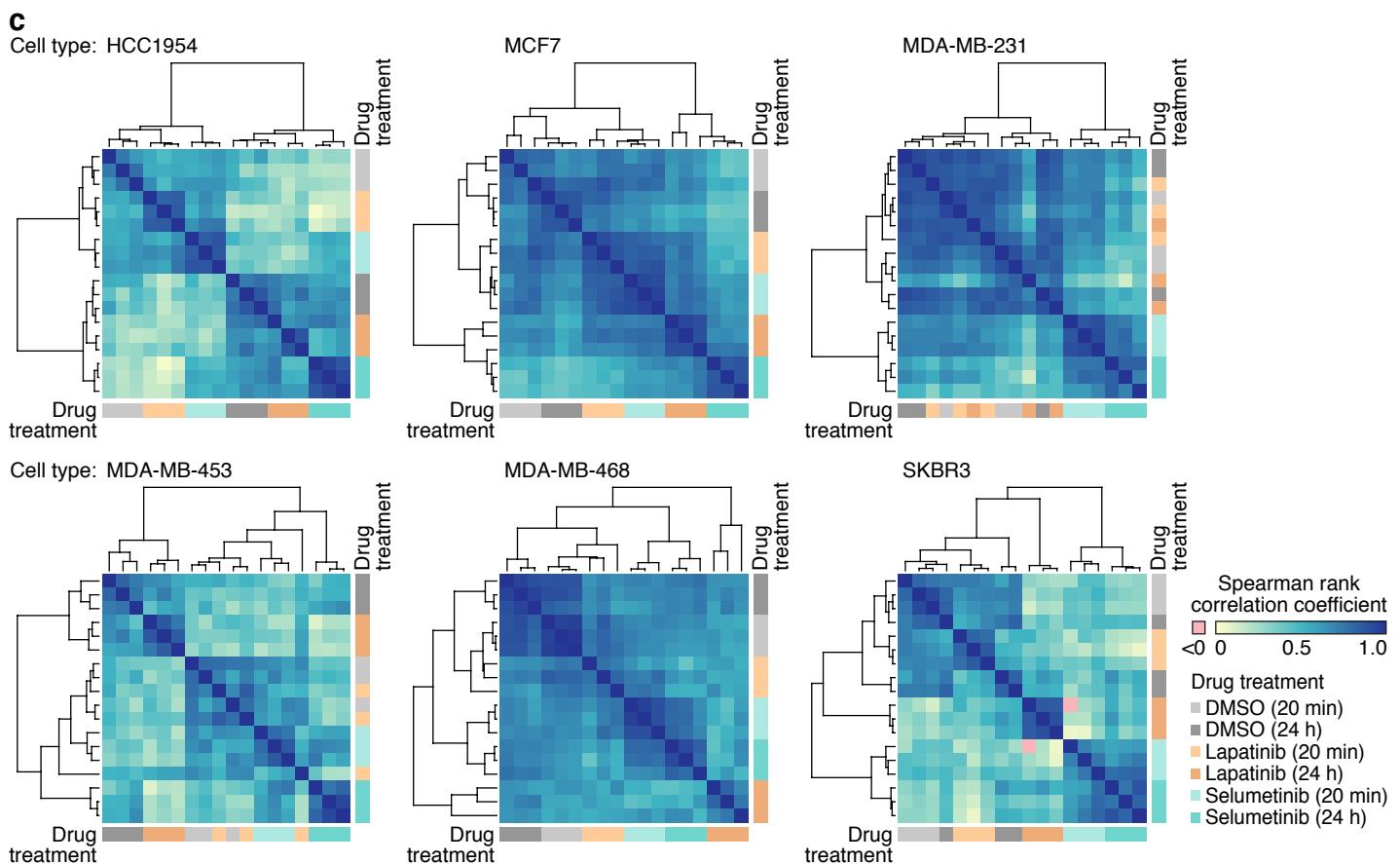
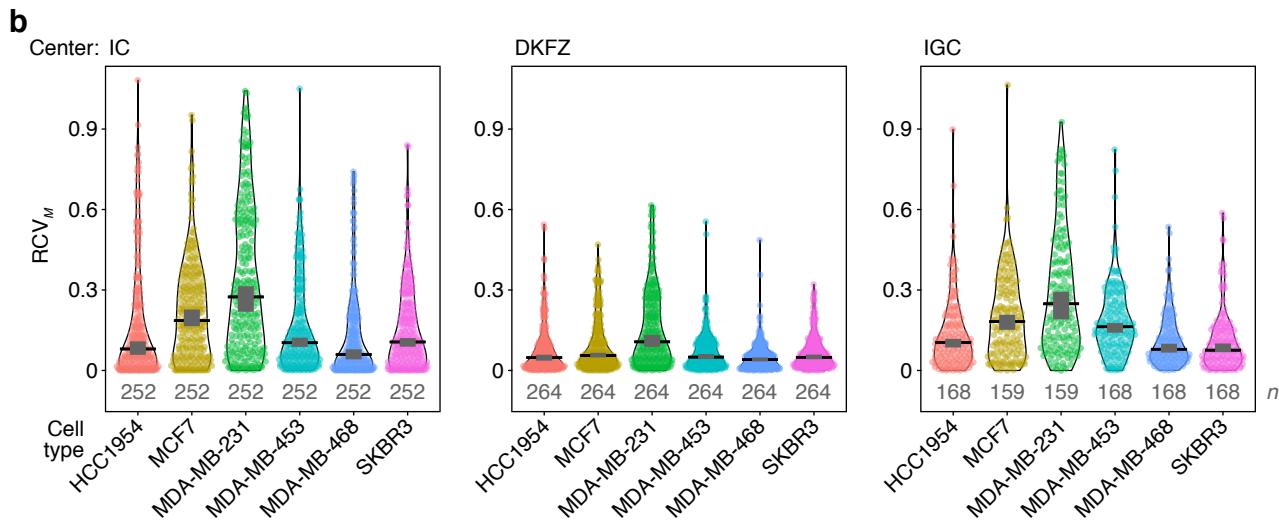
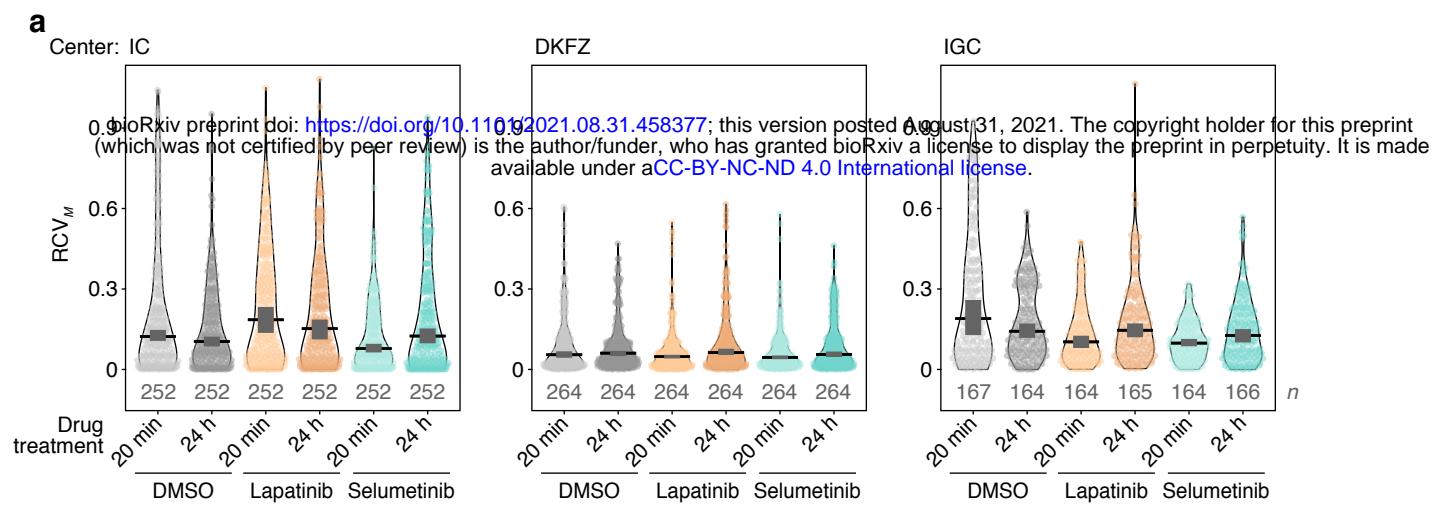
### Experimental design

Robust, high-quality data are crucial for successful RPPA workflows, especially as RPPA pipelines are continuing to be developed for use in clinical settings<sup>54–58</sup>. Thus, there is a growing need for approaches to record and evaluate the reproducibility of different RPPA platform outputs<sup>59</sup>. To control biological and technical variability in this multicenter study, each cell line was cultured at one center and then lysates were distributed to all centers, so that all platforms analyzed exactly the same batch of protein extract. Sample shipping and transit time have been shown previously to have limited effects on RPPA quantification accuracy<sup>54</sup>. Importantly, the same protein extraction buffer was used for all experiments reported here to avoid differences in efficiency of extraction of nuclear, DNA-bound or membrane-bound proteins. When printing the samples, negative control spots (lysis buffer) were printed in parallel to enable evaluation of background levels. Technical replicate sample spots were printed on different areas of the slides to avoid potential spatial bias.

### Sample variability

The high-throughput nature of RPPA permitted the multiplexed quantification of 108 biological samples at each research center in this study, which yielded quantitative data from a total of 86,832 arrayed sample spots. This large-scale dataset enables the assessment and comparison of variability within and between independent RPPA platforms. Normalized RPPA data for all drug treatments and cell types were similarly distributed for all biological replicate samples (Supplementary Fig. 1). To measure relative variability among biological replicates, robust coefficient of variation based on the median absolute deviation ( $RCV_M$ ) was calculated using unnormalized RPPA data for all samples across drug treatments (Fig. 2a) and cell types (Fig. 2b). Data from all three RPPA platforms exhibited positively skewed distributions of  $RCV_M$  values, with median  $RCV_M \leq 0.190$  for all drug treatments [ranges 0.078–0.185 (platform 1), 0.045–0.063 (platform 2), and 0.098–0.190 (platform 3)] (Fig. 2a). Data for all cell types had median  $RCV_M \leq 0.274$  [ranges 0.059–0.274 (platform 1), 0.041–0.107 (platform 2), and 0.076–0.249 (platform 3)], which was no greater than 0.186 if the slightly more variable MDA-MB-231 cell type was excluded (Fig. 2b). Median absolute deviations (MADs) for centered RPPA data from all three RPPA platforms were similarly distributed, with median MADs no greater than 0.153 for all drug treatments (range 0.043–0.153 across platforms) and no greater than 0.211 for all cell types (range 0.037–0.211 across platforms) (Supplementary Fig. 2). These data indicate that biological replicate samples were generally concordant at all research centers.

To quantify the extent of agreement between all biological samples tested, multiple similarity metrics were calculated for each of the 11,556 pairwise combinations of RPPA samples (Online-only Table 3). Unsupervised cluster analysis of intersample Spearman rank correlation coefficients for each cell type partitioned sets of biological replicate samples together for most cell types, indicating that they were better correlated than distinct biological samples (Fig. 2c). Indeed, pairwise combinations of like biological replicate samples exhibited strong positive correlations (median  $\geq 0.770$ ) with low variability (MAD  $\leq 0.066$ ), whereas pairwise combinations of



**Figure 2.** Comparison of RPPA data acquired at each research center. **(a, b)** Distributions of robust coefficient of variation based on the median absolute deviation ( $RCV_M$ ) for biological replicate sample data derived from all antibodies used at each research center ( $n RCV_M$  values, from  $n = 3$  biological replicates, are indicated in gray text below each respective probability density).  $RCV_M$  values derived from negative data points were omitted. Unnormalized RPPA data are compared across drug treatments **(a)** and cell types **(b)**. Black bar, median; dark gray box, 95% confidence interval; black silhouette outline, probability density **(c)**. Hierarchical cluster analysis of pairwise Spearman rank correlation coefficients between all drug treatments ( $n = 3$  biological replicates). Annotation bar color indicates drug treatment and treatment timepoint.

nonreplicate samples were more weakly correlated over a substantially broader range of Spearman rank correlation coefficients (lower limit 0.485 for replicate samples, -0.006 for nonreplicated samples across cell types) (Table 4).

### Cell type-specific RPPA profiles

To explore the protein-level relationships between the different breast cancer cell types examined in this study, we analyzed all centered RPPA data from each research center using t-SNE (Fig. 3a). This dimensionality reduction approach enabled the projection of the high-dimensional RPPA data to two-dimensional space, facilitating the assessment of the structure of each dataset. Visualization using t-SNE projection identified clusters of samples that were partitioned by cell type for each research center (Fig. 3a), supporting the notion that the RPPA data from each platform captured cell type-specific molecular profiles. We confirmed the clustering of cell types in the RPPA data from each platform using an alternative dimensionality reduction approach, UMAP (Supplementary Fig. 3a). In addition to analyses of centered RPPA data, which were used for multicenter dataset integration, we performed t-SNE and UMAP analyses of uncentered normalized RPPA data, which resulted in very similar projections to the centered data (Supplementary Fig. 3b, c), verifying that antibody-wise centering did not distort the data structure.

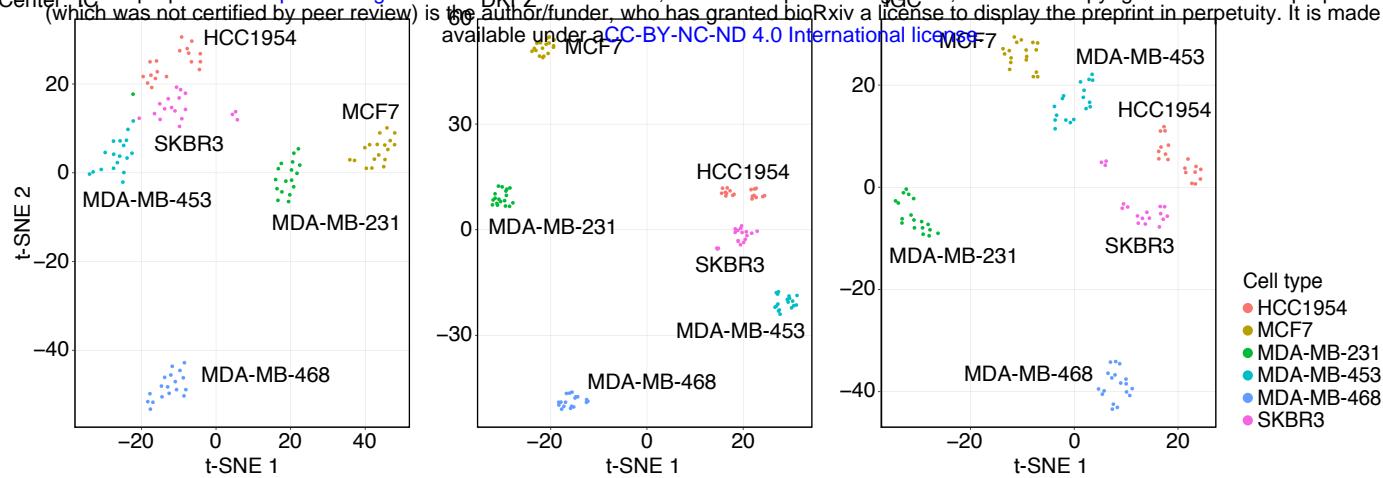
Samples from some cell types, such as HCC1954 and SKBR3, tended to collocate in t-SNE and UMAP space (Fig. 3a, Supplementary Fig. 3a-c), suggesting that their RPPA profiles were more similar compared to the other cell types. Indeed, the six breast cancer cell lines were chosen to reflect different molecular subtypes of breast cancer (Table 1). Overexpression of the cell-surface receptor HER2 was correctly detected in HCC1954, MDA-MB-453, and SKBR3 cell lines (Fig. 3b); MCF7 was correctly identified as the only estrogen receptor (ER)-positive cell line (Fig. 3c); MDA-MB-468 was correctly identified as the highly EGFR-overexpressing cell line (Fig. 3d). These data imply that cell type-specific protein levels quantified by RPPA can distinguish distinct molecular profiles in breast cancer cell types. Moreover, similar coclustering of these cell types was detected in data from all research centers (Fig. 3a, Supplementary Fig. 3a-c), confirming the capacity for all three RPPA platforms to identify related cell type-specific data structures.

### Cell type-specific drug sensitivities

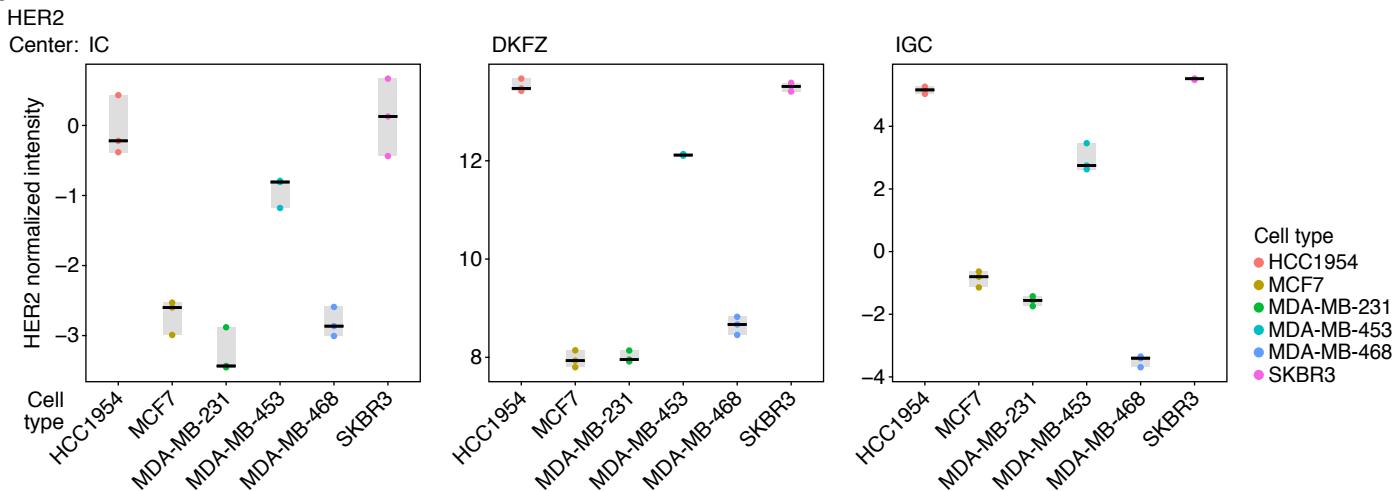
The panel of breast cancer cell lines examined in this study were treated with two clinically relevant drugs, selumetinib (a MEK inhibitor) and lapatinib (a HER2/EGFR inhibitor) (Online-only Table 2). The sensitivity of these cell lines towards these drugs and the expected changes in major cell signaling pathways are known (<http://www.cancerrxgene.org>; refs 60–62), which thus served as an internal quality control. We assessed the protein-level responses of the different cell lines following drug treatment. All platforms readily detected the effect of selumetinib on the MEK–extracellular signal-regulated kinase (Erk) pathway, and in particular the blockade between phospho-MEK and phospho-Erk, resulting in the accumulation of phospho-MEK and decrease of phospho-Erk (ref. 63) (Fig. 4a). The two cell lines known to be sensitive to lapatinib (HCC1954 and SKBR3) showed, as expected, downregulation of phospho-HER2 and phospho-EGFR, as well as downregulation of downstream phospho-Akt and phospho-S6 ribosomal protein (Fig. 4b). Cluster analysis of intersample correlations also revealed subclustering of drug-treated samples that was specific to different cell types (Fig. 2c). For example, for MDA-MB-231 cells, Spearman rank correlation coefficients of samples treated with selumetinib (20 min and 24 h) partitioned from the other drug and control treatments, whereas for SKBR3 cells, samples treated with lapatinib (24 h) and selumetinib (20 min and 24 h) partitioned separately from the other treatment conditions (Fig. 2c).

Cell type	Replicate sample Spearman rank correlation coefficients					Nonreplicate sample Spearman rank correlation coefficients				
	Min.	Max.	Median	MAD	P	Min.	Max.	Median	MAD	P
HCC1954	0.611	0.930	0.838	0.053	$\leq 3.18 \times 10^{-10}$	0.023	0.791	0.449	0.153	$\leq 8.35 \times 10^{-1}$
MCF7	0.738	0.941	0.877	0.024	$\leq 3.67 \times 10^{-16}$	0.355	0.895	0.666	0.084	$\leq 7.41 \times 10^{-4}$
MDA-MB-231	0.485	0.917	0.841	0.045	$\leq 1.88 \times 10^{-6}$	0.126	0.920	0.680	0.131	$\leq 2.46 \times 10^{-1}$
MDA-MB-453	0.500	0.925	0.770	0.046	$\leq 8.20 \times 10^{-7}$	0.088	0.771	0.437	0.117	$\leq 4.16 \times 10^{-1}$
MDA-MB-468	0.665	0.981	0.889	0.066	$\leq 2.04 \times 10^{-12}$	0.397	0.895	0.650	0.073	$\leq 1.43 \times 10^{-4}$
SKBR3	0.618	0.918	0.814	0.052	$\leq 1.74 \times 10^{-10}$	-0.006	0.819	0.414	0.122	$\leq 9.60 \times 10^{-1}$

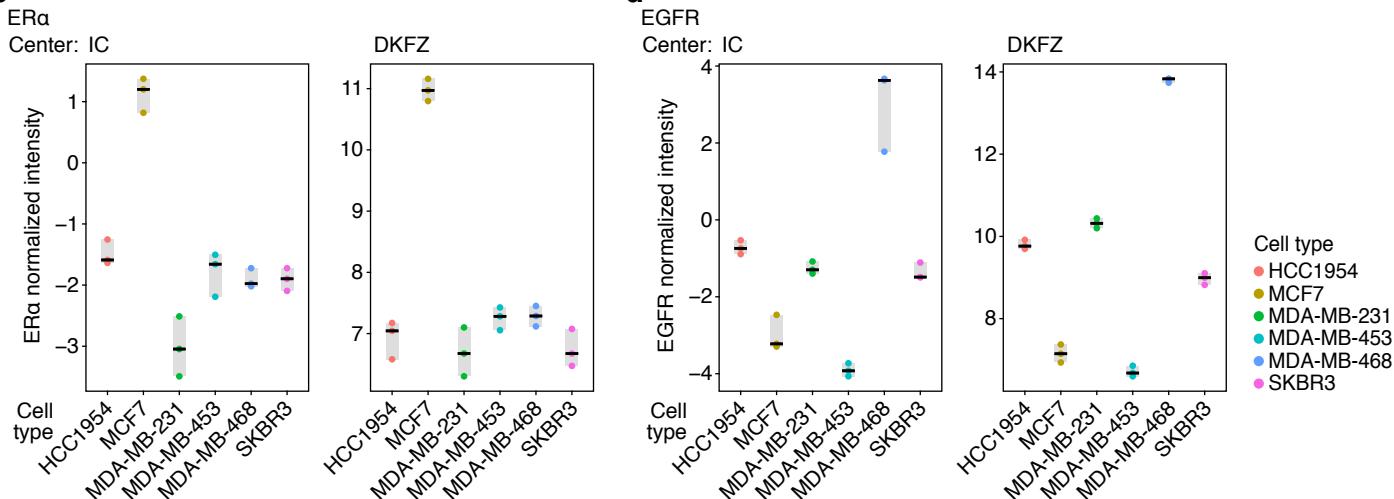
**Table 4. Correlations between replicate and nonreplicate samples.** Pairwise sample Spearman rank correlation coefficients were calculated for like biological replicate samples and for distinct (nonreplicate) samples for each cell type. Minimum (min.) and maximum (max.) values define the range of Spearman rank correlation coefficients for corresponding pairwise sample comparisons.



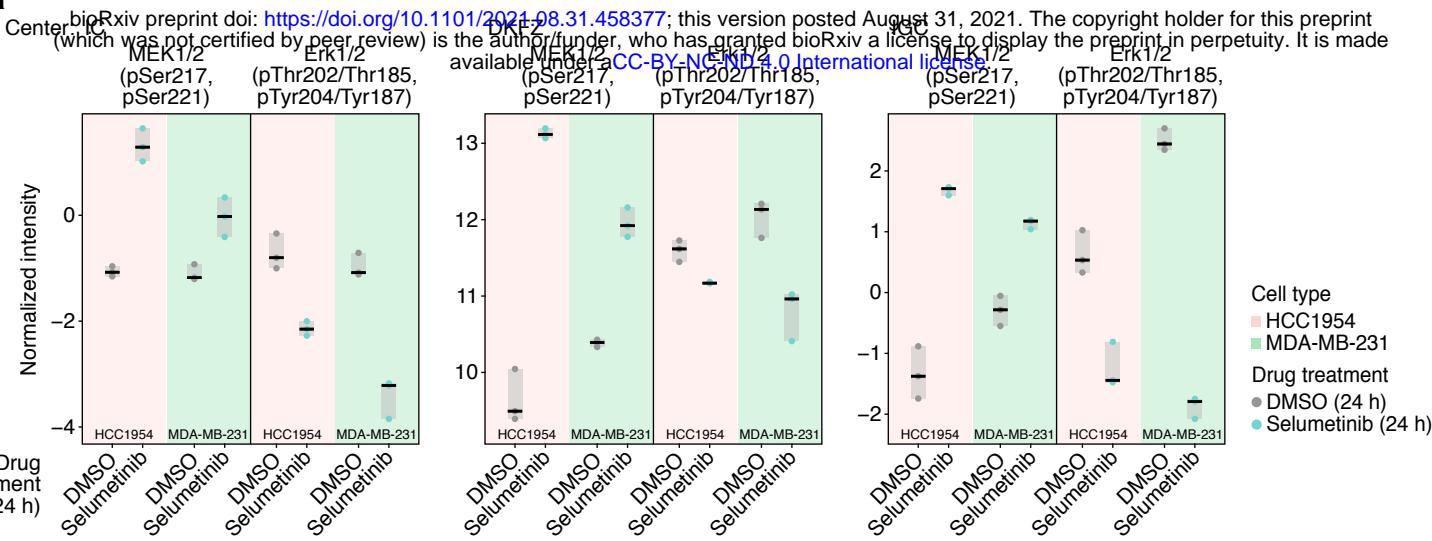
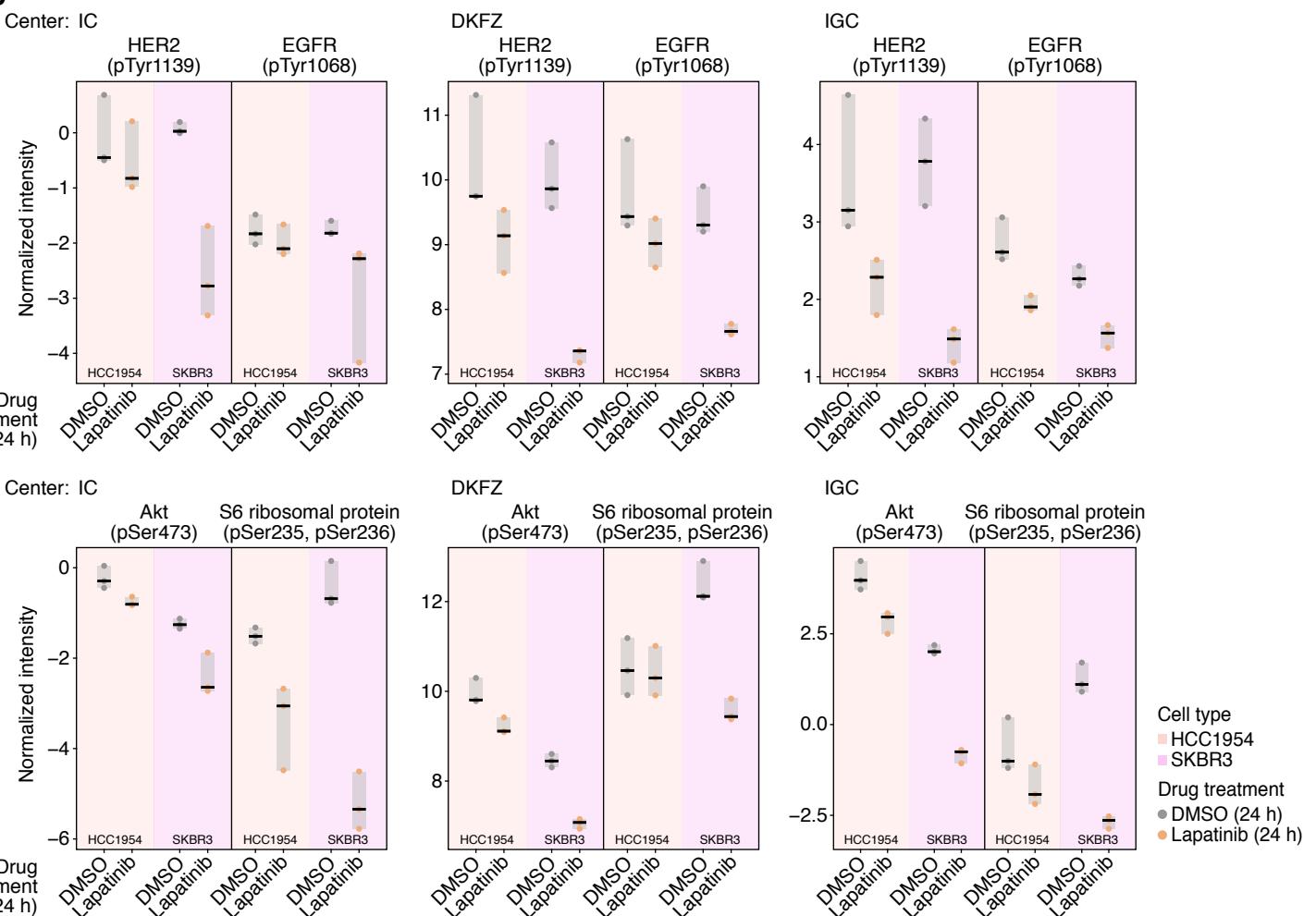
b



C



**Figure 3.** Cell type-specific differences identified by RPPA. **(a)** Dimensionality reduction of centered RPPA data derived from all antibodies used at each research center using t-SNE ( $n = 4,536, 4,752$ , and  $3,024$  arrayed lysates for IC, DKFZ, and IGC research centers, respectively, from  $n = 108$  biological samples). Cell type classes were colored as indicated. **(b-d)** Quantification of receptor expression in vehicle control-treated cells (DMSO, 24 h) analyzed by RPPA ( $n = 3$  biological replicates). HER2 antibody identifiers for IC, Her2\_b; DKFZ, Her2\_d; IGC, Her2\_a **(b)**. ER $\alpha$  antibody identifier for IC and DKFZ, ER-alpha\_b **(c)**. EGFR antibody identifiers for IC, EGFR\_a; DKFZ, EGFR\_b **(d)**. ER $\alpha$  and EGFR not analyzed at IGC. Black bar, median; light gray box, range.

**a****b**

**Figure 4.** Cell type-specific drug sensitivities identified by RPPA. (a) Quantification of phospho-MEK and phospho-Erk in selumetinib- and vehicle control-treated HCC1954 and MDA-MB-231 cells (24 h) analyzed by RPPA ( $n = 3$  biological replicates). MEK1/2 (pSer217, pSer221) antibody identifier for IC, DKFZ, and IGC, MEK1/2\_pSer217,pSer221. Erk1/2 (pThr202/Thr185, pTyr204/Tyr187) antibody identifier for IC, DKFZ, and IGC, Erk1/2\_pThr202/Thr185,pTyr204/Tyr187\_b. (b) Quantification of phospho-HER2, phospho-EGFR, phospho-Akt, and phospho-S6 ribosomal protein in lapatinib- and vehicle control-treated HCC1954 and SKBR3 cells (24 h) analyzed by RPPA ( $n = 3$  biological replicates). HER2 (pTyr1139) antibody identifier for IC, DKFZ, and IGC, Her2\_pTyr1139. EGFR (pTyr1068) antibody identifiers for IC and IGC, EGFR\_pTyr1068\_a; DKFZ, EGFR\_pTyr1068\_b. Akt (pSer473) antibody identifier for IC, DKFZ, and IGC, Akt\_pSer473\_a. S6 ribosomal protein (pSer235, pSer236) antibody identifiers for IC and IGC, S6 ribosomal protein\_pSer235,pSer236\_a; DKFZ, S6 ribosomal protein\_pSer235,pSer236\_b. Black bar, median; light gray box, range.

Together, these data suggest that RPPA data from each platform have the capacity to reveal inherent differences in drug sensitivity and pharmacodynamics between cell types, including those based on their response to targeted therapies.

In summary, the dataset represents high-quality RPPA data derived from multiple distinct RPPA platforms, characterizing a range of human breast cancer cell lines and their response to two clinically relevant cancer drugs. Our collaborative multicenter proteomic approach will enable further dataset integration for the analysis of breast cancer therapeutic response and data reuse for the assessment of interplatform reproducibility.

## Usage Notes

The data reported here serve as a valuable resource for the examination of protein-level responses to pharmacological inhibition in a panel of breast cancer cells, including the search for protein markers of response to clinically relevant therapeutics (Fig. 1d). Such investigations may be facilitated by the interrogation of the high-dimensional RPPA data using statistical and machine learning approaches, the functional analysis of signaling pathway changes, or the modeling of protein complexes with protein interaction networks<sup>64</sup>. Indeed, RPPA technology is well placed to elucidate the mechanism-of-action of novel hit compounds or candidate drugs discovered by modern phenotypic screening approaches<sup>65</sup>. Robust and reproducible RPPA platforms are required for profiling large numbers of small-molecule hits from cell-based phenotypic screening assays across dose-response and time series. Building a repository of publicly available RPPA data associated with pharmacological perturbation, such as that exemplified here, will provide valuable datasets for reuse in emerging machine learning approaches which aim to predict mechanism-of-action or differentiate novel compounds based on RPPA signature similarity with well characterized compounds. This, for example, could extend the Connectivity Map concept – the cataloging and connecting of transcriptional responses to genetic, chemical, and disease perturbations<sup>66</sup> – by providing additional, orthogonal data at the dynamic posttranslational pathway level.

Moreover, our multicenter analysis provides an important reference for the assessment of RPPA platform robustness and technology reproducibility. Individual RPPA platforms may use the dataset to quality control their workflows, by reproducing the experiments outlined here and comparing their results with our data (Fig. 1d). For interplatform comparison, only antigens analyzed by RPPA at all research centers were compiled in the integrated dataset in this study (Fig. 1c), so future data reuse may include the reintegration of data for antibodies targeting excluded antigens, which are available in the raw data files<sup>49</sup>, to enable analysis of data acquired on different combinations of RPPA platforms. In addition, as no standard tools exist to normalize RPPA data, the raw data reported here may be used to test and validate new data normalization pipelines (Fig. 1d). The data also complement ongoing efforts to assess the effects of sample handling on RPPA quantification in the setting of multisite clinical trials<sup>54</sup>. Finally, the integrated dataset can act as a framework for future interplatform comparisons, and this study opens the door for crossplatform validation of RPPA data to identify robust markers of disease and response to therapy. Expansion of the current dataset to include additional samples, antibodies, or RPPA platforms will improve the applicability of such projects and expand the scope of analysis to, for example, other drug treatments or disease models (Fig. 1d).

The data reported here thus provide a valuable springboard for the assessment of reproducibility or benchmarking of platform-specific RPPA datasets, which will support the refinement of technology best practice and may aid development toward standardization of data processing procedures and improvements in RPPA platform interoperability.

## Code Availability

The code to implement RPPA data analysis using NormaCurve<sup>50</sup> is freely available at [http://microarrays.curie.fr/publications/U900-RPPA\\_PLT/Normacurve](http://microarrays.curie.fr/publications/U900-RPPA_PLT/Normacurve). The code to implement RPPA data analysis using RPPanalyzer<sup>51</sup> is freely available at <https://cran.r-project.org/package=RPPanalyzer>. The code to visualize data using PlotsOfData<sup>67</sup> is freely available at <https://github.com/JoachimGoedhart/PlotsOfData>. The code to visualize data using UMAP and t-SNE is freely available at <https://github.com/atamaianalytics/DimensionReduction>.

## Acknowledgments

We thank Stéphane Liva and Patrick Poulet for development and maintenance of data management tools at the Institut Curie research center and Henry Beetham, Philippe Hupé, and Andy Sims for helpful discussions. RPPA platform 1 (Institut Curie research center) is supported by Cancéropôle Ile-de-France. K.G.M. is supported by the Cancer Research UK Edinburgh Centre award (C157/A25140). N.O.C. is supported by Cancer Research UK (C42454/A28596) and the Brain Tumour Charity (GN-000676). A.B. was supported by Cancer Research UK.

## Author Contributions

L.d.K., N.O.C., U.K., B.S., and A.B. conceptualized the study and designed the experiments. S.B., K.G.M., B.O., and A.C. generated the data. L.d.K., S.B., K.G.M., V.S., B.S., and A.B. analyzed the data. A.B. designed and implemented multicenter data integration, curated the data, and prepared the figures. L.d.K. and A.B. and wrote the manuscript. L.d.K., S.B., K.G.M., N.O.C., and A.B. edited the manuscript. All authors critically reviewed and approved the final manuscript.

## Competing Interests

The authors declare no competing interests.

## References

1. Kantarjian, H. M. *et al.* Nilotinib versus imatinib for the treatment of patients with newly diagnosed chronic phase, Philadelphia chromosome-positive, chronic myeloid leukaemia: 24-month minimum follow-up of the phase 3 randomised ENESTnd trial. *Lancet Oncol.* **12**, 841–851 (2011).
2. Pegram, M. D. *et al.* Phase II study of receptor-enhanced chemosensitivity using recombinant humanized anti-p185HER2/neu monoclonal antibody plus cisplatin in patients with HER2/neu-overexpressing metastatic breast cancer refractory to chemotherapy treatment. *J. Clin. Oncol.* **16**, 2659–2671 (1998).
3. Tannock, I. F. & Hickman, J. A. Limits to personalized cancer medicine. *N. Engl. J. Med.* **375**, 1289–1294 (2016).
4. Hickman, J. A., Tannock, I. F., Meheus, L. & Hutchinson, L. The European Union and personalised cancer medicine. *Eur. J. Cancer* **150**, 95–98 (2021).
5. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
6. Michaut, M. *et al.* Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. *Sci. Rep.* **6**, 18517 (2016).

7. Barbieri, R. *et al.* Proteogenomics: key driver for clinical discovery and personalized medicine. *Adv. Exp. Med. Biol.* **926**, 21–47 (2016).
8. Carr, K. M., Rosenblatt, K., Petricoin, E. F. & Liotta, L. A. Genomic and proteomic approaches for studying human cancer: prospects for true patient-tailored therapy. *Hum. Genomics* **1**, 134–140 (2004).
9. Masuda, M. & Yamada, T. Signaling pathway profiling by reverse-phase protein array for personalized cancer medicine. *Biochim. Biophys. Acta* **1854**, 651–657 (2015).
10. Swanton, C. *et al.* Consensus on precision medicine for metastatic cancers: a report from the MAP conference. *Ann. Oncol.* **27**, 1443–1448 (2016).
11. van Dijk, A. D., de Bont, E. S. J. M. & Kornblau, S. M. Targeted therapy in acute myeloid leukemia: current status and new insights from a proteomic perspective. *Expert Rev. Proteomics* **17**, 1–10 (2020).
12. Liotta, L. A. *et al.* Protein microarrays: meeting analytical challenges for clinical applications. *Cancer Cell* **3**, 317–325 (2003).
13. Akbani, R. *et al.* Realizing the promise of reverse phase protein arrays for clinical, translational, and basic research: a workshop report: the RPPA (Reverse Phase Protein Array) society. *Mol. Cell. Proteomics* **13**, 1625–1643 (2014).
14. Baldelli, E. *et al.* Reverse phase protein microarrays. *Methods Mol. Biol.* **1606**, 149–169 (2017).
15. Pawlak, M. & Carragher, N. O. Reverse phase protein arrays elucidate mechanisms-of-action and phenotypic response in 2D and 3D models. *Drug Discov. Today Technol.* **23**, 7–16 (2017).
16. Aslan, O. *et al.* Preclinical evaluation and reverse phase protein array-based profiling of PI3K and MEK inhibitors in endometrial carcinoma in vitro. *BMC Cancer* **18**, 168 (2018).
17. Bernhardt, S. *et al.* Functional proteomics of breast cancer metabolism identifies GLUL as responder during hypoxic adaptation. *J. Proteome Res.* **18**, 1352–1362 (2019).
18. Blum, Y. *et al.* Dissecting heterogeneity in malignant pleural mesothelioma through histo-molecular gradients for clinical applications. *Nat. Commun.* **10**, 1333 (2019).
19. Coussy, F. *et al.* Response to mTOR and PI3K inhibitors in enzalutamide-resistant luminal androgen receptor triple-negative breast cancer patient-derived xenografts. *Theranostics* **10**, 1531–1543 (2020).
20. Dawson, J. C. *et al.* A synergistic anticancer FAK and HDAC inhibitor combination discovered by a novel chemical-genetic high-content phenotypic screen. *Mol. Cancer Ther.* **19**, 637–649 (2020).
21. Faham, N., Zhao, L. & Welm, A. L. mTORC1 is a key mediator of RON-dependent breast cancer metastasis with therapeutic potential. *NPJ Breast Cancer* **4**, 36 (2018).
22. Ferrarotto, R. *et al.* Proteogenomic analysis of salivary adenoid cystic carcinomas defines molecular subtypes and identifies therapeutic targets. *Clin. Cancer Res.* **27**, 852–864 (2021).
23. Hoff, F. W. *et al.* Recognition of recurrent protein expression patterns in pediatric acute myeloid leukemia identified new therapeutic targets. *Mol. Cancer Res.* **16**, 1275–1286 (2018).
24. Lallemand, F. *et al.* The high protein expression of FOXO3, but not that of FOXO1, is associated with markers of good prognosis. *Sci. Rep.* **10**, 6920 (2020).
25. McDaniel, N. K. *et al.* AXL mediates cetuximab and radiation resistance through tyrosine 821 and the c-ABL kinase pathway in head and neck cancer. *Clin. Cancer Res.* **26**, 4349–4359 (2020).
26. Osumi, H. *et al.* Evaluation of the RAS signaling network in response to MEK inhibition using organoids derived from a familial adenomatous polyposis patient. *Sci. Rep.* **10**, 17455 (2020).
27. Ozkan-Dagliyan, I. *et al.* Low-dose vertical inhibition of the RAF-MEK-ERK cascade causes apoptotic death of KRAS mutant cancers. *Cell Rep.* **31**, 107764 (2020).

28. Scholl, S. M. *et al.* Genetic markers and phosphoprotein forms of beta-catenin p $\beta$ -Cat552 and p $\beta$ -Cat675 are prognostic biomarkers of cervical cancer. *EBioMedicine* **61**, 103049 (2020).
29. Subramonian, D. *et al.* Regorafenib is effective against neuroblastoma in vitro and in vivo and inhibits the RAS/MAPK, PI3K/Akt/mTOR and Fos/Jun pathways. *Br. J. Cancer* **123**, 568–579 (2020).
30. Teo, K. *et al.* E-cadherin loss induces targetable autocrine activation of growth factor signalling in lobular breast cancer. *Sci. Rep.* **8**, 15454 (2018).
31. van Dijk *et al.* Loss of H3K27 methylation identifies poor outcomes in adult-onset acute leukemia. *Clin. Epigenetics* **13**, 21 (2021).
32. Wagner, M. J. *et al.* Combined VEGFR and MAPK pathway inhibition in angiosarcoma. *Sci. Rep.* **11**, 9362 (2021).
33. Wang, J. *et al.* AKT isoform-specific expression and activation across cancer lineages. *BMC Cancer* **18**, 742 (2018).
34. Xue, L. *et al.* Metformin and an insulin/IGF-1 receptor inhibitor are synergistic in blocking growth of triple-negative breast cancer. *Breast Cancer Res. Treat.* **185**, 73–84 (2021).
35. Yang, B. *et al.* MEK inhibition remodels the immune landscape of mutant KRAS tumors to overcome resistance to PARP and immune checkpoint inhibitors. *Cancer Res.* **81**, 2714–2729 (2021).
36. Mueller, C., Liotta, L. A. & Espina, V. Reverse phase protein microarrays advance to use in clinical trials. *Mol. Oncol.* **4**, 461–481 (2010).
37. Bauman, J. E. *et al.* Randomized, placebo-controlled window trial of EGFR, Src, or combined blockade in head and neck cancer. *JCI Insight* **2**, e90449 (2017).
38. Garrido-Castro, A. C. *et al.* Phase 2 study of buparlisib (BKM120), a pan-class I PI3K inhibitor, in patients with metastatic triple-negative breast cancer. *Breast Cancer Res.* **22**, 120 (2020).
39. Gonzalez-Angulo, A. M. *et al.* Open-label randomized clinical trial of standard neoadjuvant chemotherapy with paclitaxel followed by FEC versus the combination of paclitaxel and everolimus followed by FEC in women with triple receptor-negative breast cancer. *Ann. Oncol.* **25**, 1122–1127 (2014).
40. Hoff, F. W. *et al.* Heat shock factor 1 (HSF1-pSer326) predicts response to bortezomib-containing chemotherapy in pediatric AML: a COG report. *Blood* **137**, 1050–1060 (2021).
41. Posadas, E. M. *et al.* A phase II and pharmacodynamic study of gefitinib in patients with refractory or recurrent epithelial ovarian cancer. *Cancer* **109**, 1323–1330 (2007).
42. Ragon, B. K. *et al.* Oral MEK 1/2 inhibitor trametinib in combination with AKT inhibitor GSK2141795 in patients with acute myeloid leukemia with RAS mutations: a phase II study. *Clin. Lymphoma Myeloma Leuk.* **19**, 431–440.e13 (2019).
43. Wolf, D. M. *et al.* Mechanism of action biomarkers predicting response to AKT inhibition in the I-SPY 2 breast cancer trial. *NPJ Breast Cancer* **6**, 48 (2020).
44. Wulffkuhle, J. D. *et al.* Evaluation of the HER/PI3K/AKT family signaling network as a predictive biomarker of pathologic complete response for patients with breast cancer treated with neratinib in the I-SPY 2 TRIAL. *JCO Precis. Oncol.* **2**, PO.18.00024 (2018).
45. Xing, Y. *et al.* Phase II trial of AKT inhibitor MK-2206 in patients with advanced breast cancer who have tumors with PIK3CA or AKT mutations, and/or PTEN loss/PTEN mutation. *Breast Cancer Res.* **21**, 78 (2019).

46. Jameson, G. S. *et al.* A pilot study utilizing multi-omic molecular profiling to find potential targets and select individualized treatments for patients with previously treated metastatic breast cancer. *Breast Cancer Res. Treat.* **147**, 579–588 (2014).
47. Pierobon, M. *et al.* Pilot phase I/II personalized therapy trial for metastatic colorectal cancer: evaluating the feasibility of protein pathway activation mapping for stratifying patients to therapy with imatinib and panitumumab. *J. Proteome Res.* **13**, 2846–2855 (2014).
48. Byron, A. *et al.* Integrative analysis of multi-platform reverse-phase protein array data for the pharmacodynamic assessment of response to targeted therapies. *Sci. Rep.* **10**, 21985 (2020).
49. de Koning, L. *et al.* Data records for multicenter reverse-phase protein array data integration. *figshare* <https://doi.org/10.6084/m9.figshare.14754069> (2021).
50. Troncale, S. *et al.* NormaCurve: a SuperCurve-based method that simultaneously quantifies and normalizes reverse phase protein array data. *PLoS One* **7**, e38686 (2012).
51. Mannsperger, H. A., Gade, S., Henjes, F., Beissbarth, T. & Korf, U. RPPAnalyzer: analysis of reverse-phase protein array data. *Bioinformatics* **26**, 2202–2203 (2010).
52. de Hoon, M. J. L., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).
53. Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
54. Horton, T. M. *et al.* The effects of sample handling on proteomics assessed by reverse phase protein arrays (RPPA): functional proteomic profiling in leukemia. *J. Proteomics* **233**, 104046 (2021).
55. Labrie, M. *et al.* Using reverse phase protein array (RPPA) to identify and target adaptive resistance. *Adv. Exp. Med. Biol.* **1188**, 251–266 (2019).
56. Masuda, M. & Yamada, T. Utility of reverse-phase protein array for refining precision oncology. *Adv. Exp. Med. Biol.* **1188**, 239–249 (2019).
57. Petricoin, E. 3rd *et al.* RPPA: origins, transition to a validated clinical research tool, and next generations of the technology. *Adv. Exp. Med. Biol.* **1188**, 1–19 (2019).
58. Siwak, D. R., Li, J., Akbani, R., Liang, H. & Lu, Y. Analytical platforms 3: processing samples via the RPPA Pipeline to generate large-scale data for clinical studies. *Adv. Exp. Med. Biol.* **1188**, 113–147 (2019).
59. Byron, A. Reproducibility and crossplatform validation of reverse-phase protein array data. *Adv. Exp. Med. Biol.* **1188**, 181–201 (2019).
60. Garon, E. B. *et al.* Identification of common predictive markers of in vitro response to the Mek inhibitor selumetinib (AZD6244; ARRY-142886) in human breast cancer and non-small cell lung cancer cell lines. *Mol. Cancer Ther.* **9**, 1985–1994 (2010).
61. Konecny, G. E. *et al.* Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells. *Cancer Res.* **66**, 1630–1639 (2006).
62. O'Neill, F. *et al.* Gene expression changes as markers of early lapatinib response in a panel of breast cancer cell lines. *Mol. Cancer* **11**, 41 (2012).
63. Friday, B. B. *et al.* BRAF V600E disrupts AZD6244-induced abrogation of negative feedback pathways between extracellular signal-regulated kinase and Raf proteins. *Cancer Res.* **68**, 6145–6153 (2008).
64. Byron, A. Clustering and network analysis of reverse phase protein array data. *Methods Mol. Biol.* **1606**, 171–191 (2017).

65. Hughes, R. E., Elliott, R. J. R., Dawson, J. C. & Carragher, N. O. High-content phenotypic and pathway profiling to advance drug discovery in diseases of unmet need. *Cell Chem. Biol.* **28**, 338–355 (2021).
66. Subramanian, A. *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452.e17 (2017).
67. Postma, M. & Goedhart, J. PlotsOfData—a web app for visualizing data together with their summaries. *PLoS Biol.* **17**, e3000202 (2019).