



OPEN

DATA DESCRIPTOR

# Proteomic cellular signatures of kinase inhibitor-induced cardiotoxicity

Yuguang Xiong<sup>1,7</sup>, Tong Liu<sup>2,7</sup>, Tong Chen<sup>2,7</sup>, Jens Hansen<sup>1</sup>, Bin Hu<sup>1</sup>, Yibang Chen<sup>1</sup>, Gomathi Jayaraman<sup>1</sup>, Stephan Schürer<sup>3,4</sup>, Dusica Vidovic<sup>3,4</sup>, Joseph Goldfarb<sup>1</sup>, Eric A. Sobie<sup>1</sup>, Marc R. Birtwistle<sup>1,5</sup>, Ravi Iyengar<sup>1</sup> & Evren U. Azeloglu<sup>1,6</sup>

Drug Toxicity Signature Generation Center (DToxS) at the Icahn School of Medicine at Mount Sinai is one of the centers for the NIH Library of Integrated Network-Based Cellular Signatures (LINCS) program. Its key aim is to generate proteomic and transcriptomic signatures that can predict cardiotoxic adverse effects of kinase inhibitors approved by the Food and Drug Administration. Towards this goal, high throughput shotgun proteomics experiments (308 cell line/drug combinations +64 control lysates) have been conducted. Using computational network analyses, these proteomic data can be integrated with transcriptomic signatures, generated in tandem, to identify cellular signatures of cardiotoxicity that may predict kinase inhibitor-induced toxicity and enable possible mitigation. Both raw and processed proteomics data have passed several quality control steps and been made publicly available on the PRIDE database. This broad protein kinase inhibitor-stimulated human cardiomyocyte proteomic data and signature set is valuable for prediction of drug toxicities.

## Background & Summary

Protein kinase inhibitors (KIs) belong to a class of targeted therapeutics that are being increasingly used in treatment of various cancers<sup>1</sup>. Their use and development have been accelerated in recent years as they could target tumors more effectively than most other chemotherapeutics, and their mechanisms of actions are well defined. In many cases, however, their intended- or off-target kinases serve key biological roles, which when blocked, lead to severe adverse effects<sup>2</sup>. One of the major adverse effects that lead to discontinuation of treatment with KIs is cardiotoxicity<sup>3</sup>. As a part of the NIH Library of Integrated Cellular Signatures (LINCS) program<sup>4</sup>, the main goal of the Mount Sinai Drug Toxicity Signature Generation Center (DToxS) is to better understand mechanisms of KI-associated cardiotoxicity by constructing cellular signatures of drug effects. Since different omics assays show varied sensitivities<sup>5</sup> and offer complementary molecular information on the cellular phenotypic state<sup>6</sup>, in order to develop a comprehensive understanding of the cellular responses to KIs, we use network analyses<sup>7</sup> to combine differential expression of genes and gene products in human cardiomyocytes treated by FDA approved KIs, analyzed using both transcriptomic and proteomic methods. In this dataset, we present the proteomic portion of the effects of KIs on human cardiomyocytes. Kinase inhibitors presented here are grouped and color-coded according to their primary target profile (see Table 1 for this and additional drug metadata).

Quantitative proteomics technologies have evolved to become increasingly effective at identifying differentially expressed proteins among diverse experimental conditions. Currently, two broad strategies are widely used for large-scale quantitative proteomics studies: stable isotope label-based<sup>8,9</sup> and label-free quantification (LFQ) strategies<sup>10,11</sup>. The label-based methods provide fewer experimental variations and better protein quantitation

<sup>1</sup>Department of Pharmacological Sciences and Institute for Systems Biomedicine, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA. <sup>2</sup>Center for Advanced Proteomics Research and Department of Microbiology, Biochemistry and Molecular Genetics, Rutgers University - New Jersey Medical School, Newark, NJ, 07103, USA. <sup>3</sup>Department of Molecular and Cellular Pharmacology, Miller School of Medicine, University of Miami, Miami, FL, 33136, USA. <sup>4</sup>Center for Computational Science, University of Miami, Miami, FL, 33136, USA. <sup>5</sup>Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, SC, 29634, USA. <sup>6</sup>Department of Medicine, Division of Nephrology, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA. <sup>7</sup>These authors contributed equally: Yuguang Xiong, Tong Liu, Tong Chen. ✉e-mail: ravi.iyengar@mssm.edu; liho2@rutgers.edu; evren.azeloglu@mssm.edu

Drug	Acronym	Drug Form	Purity	Concentration ( $\mu\text{M}$ )	Reference
Afatinib	AFA		98.8	0.05	27,28
Axitinib	AXI		98.0	0.20	29,30
Bosutinib	BOS		98.0	0.10	31
Cabozantinib	CAB		99.1	2.00	32
Dabrafenib	DAB		99.6	2.50	33,34
Dasatinib	DAS		98.4	0.10	35
Erlotinib	ERL		99.3	3.00	36,37
Gefitinib	GEF		99.3	1.00	38
Imatinib	IMA		99.5	5.00	39
Lapatinib	LAP		99.0	2.00	40
Nilotinib	NIL		99.8	3.00	41,42
Pazopanib	PAZ		96.1	10.0	43
Ponatinib	PON		97.3	0.10	44
Regorafenib	REG		99.6	1.00	45
Ruxolitinib	RUX		99.4	1.00	46,47
Sorafenib	SOR	Tosylate	99.7	0.50	46,47
Sunitinib	SUN	Malate	97.1	1.00	48,49
Trametinib	TRA		99.2	0.10	50
Tofacitinib	TOF	Citrate	99.9	1.00	50
Vandetanib	VAN		99.7	0.33	51
Vemurafenib	VEM		98.1	2.00	52

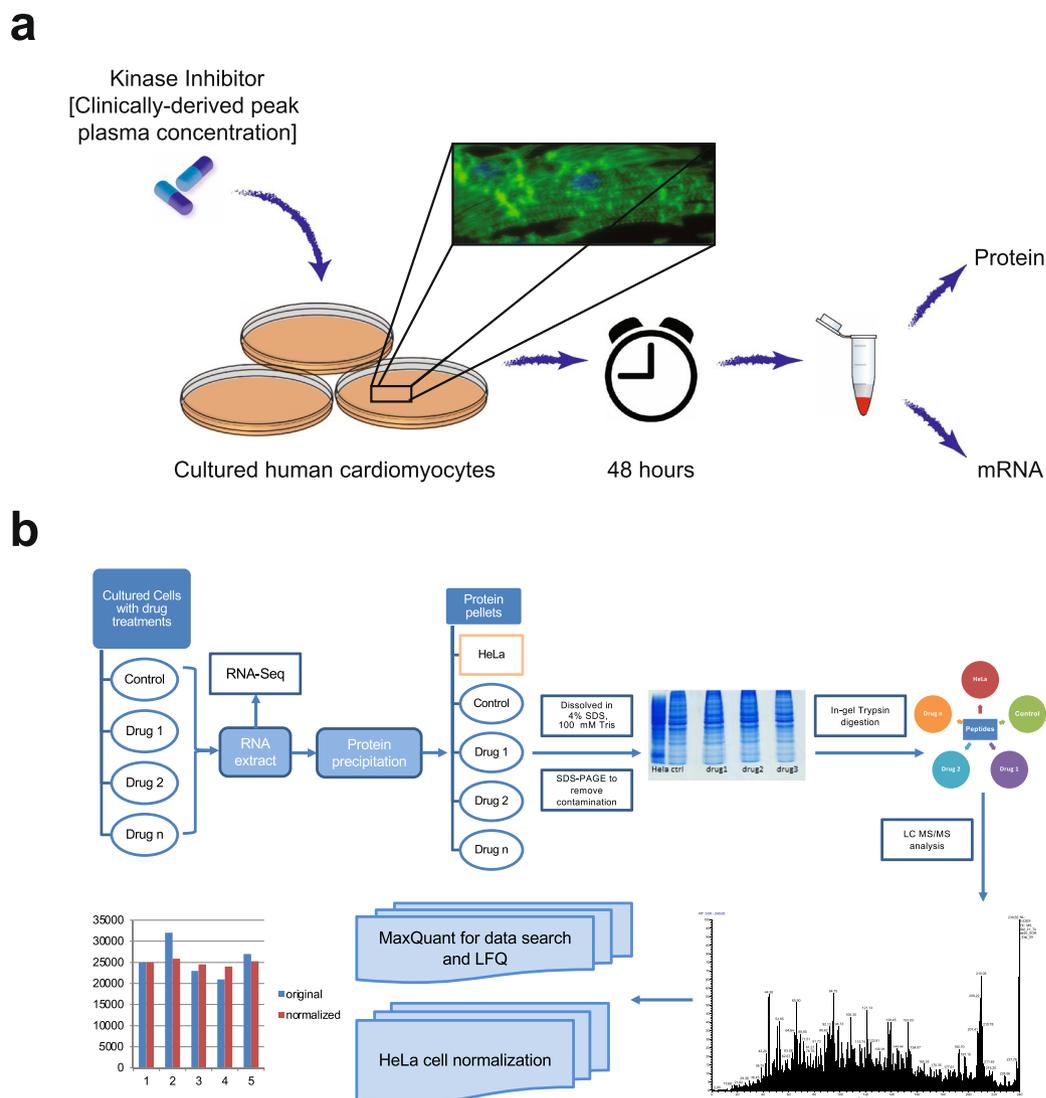
**Table 1.** Drug metadata. Purity was verified for each lot number. The concentration used for each drug is equivalent to the clinically-observed median peak plasma concentration as per the cited reference. Drugs are coded into different shades of a given color based on their primary target: purple (EGFR), orange (BCR-ABL), gray (VEGFR), red (JAK1/2), pink (MEK1/2), blue (multi-kinase).

precision<sup>12,13</sup>; however, the cost for label-based proteomics reagents is prohibitively expensive for analyzing hundreds of samples. LFQ methods, on the other hand, do not involve expensive labeling reagents; thus, they can be used for clinically relevant studies of hundreds and thousands of samples if proper controls and quality guidelines are followed. LFQ methods have more comprehensive dynamic ranges and more flexibility in study design<sup>14–16</sup>. However, while LFQ sample throughput can be increased through this approach, it should be noted that this approach can also lead to a smaller number of identified proteins, due to limited peptide fractionation.

In this DToxS dataset, we have chosen the LFQ method in order to economically compare the proteomics signatures from over 300 samples (Supplementary Table 1) that have been collected over an extended period of time, some with limited protein yields. The transcriptomes of these samples have been previously assessed using 3' digital gene expression RNA sequencing (RNAseq)<sup>17,18</sup>. To ensure both proteomic and transcriptomic data can be compared from the same drug-treated cells, we have optimized a method to extract proteins from the samples after RNA extraction at the Center for Advanced Proteomics Research (CAPR, <http://njms.rutgers.edu/proweb/>) at Rutgers University - New Jersey Medical School (Fig. 1). All the sample preparation, analysis procedures and data have gone through careful quality control steps to ensure reproducibility (Fig. 2); standard operating procedures and data descriptors are publicly available in the DToxS website ([www.dtox.org](http://www.dtox.org)) and both raw and MaxQuant-analyzed data are publicly available via the PRIDE database (PXD014791) as well as the LINCS Project data portal (<http://lincsportal.ccs.miami.edu/dcic-portal/>). We think that this large KI-induced proteomics dataset will be a valuable resource for cancer, cardiology and drug development communities.

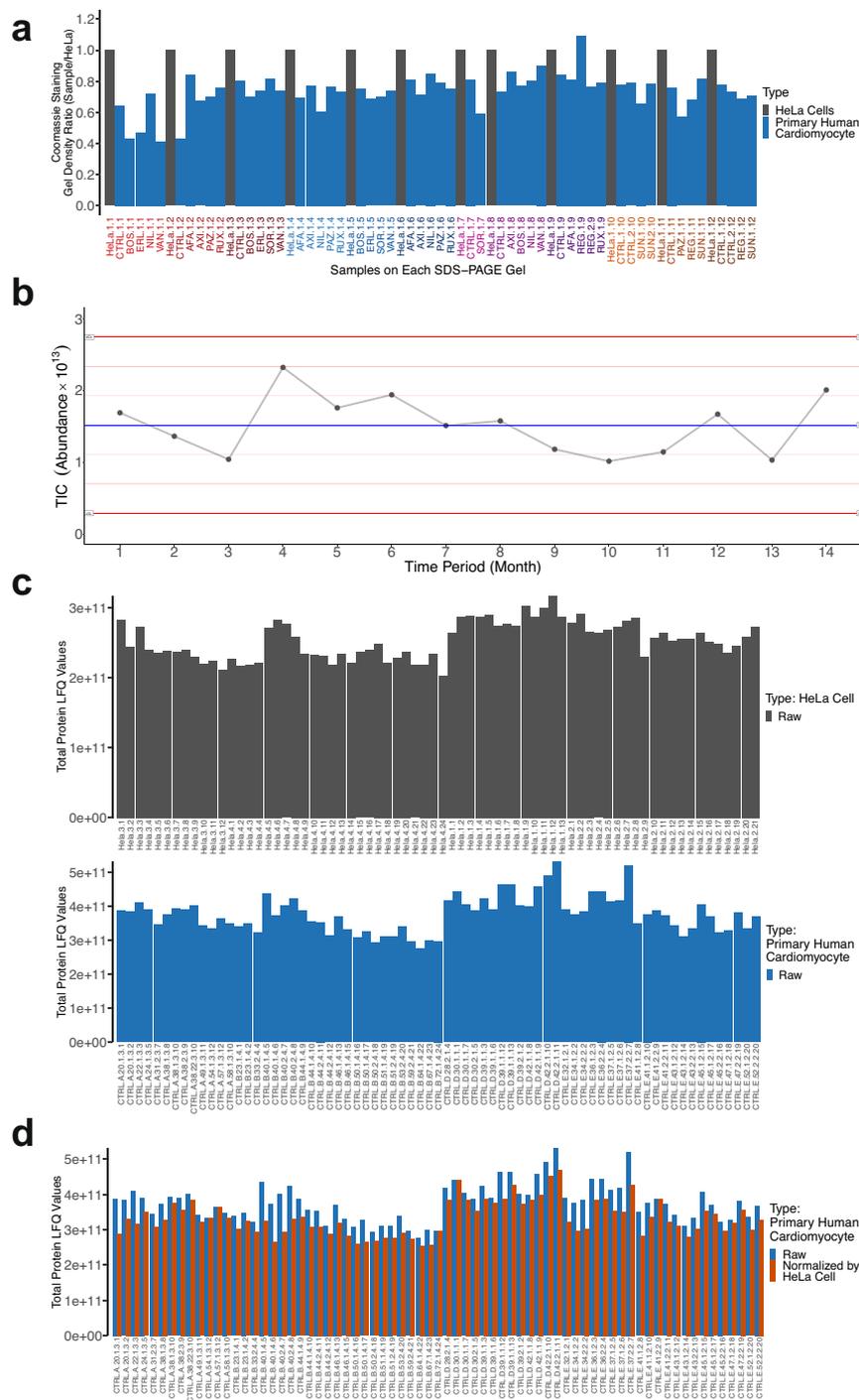
## Methods

**Cell culture, drug treatments and RNA extraction.** Detailed materials and methods for all the experiments in this study can be accessed as version- and quality-controlled standard operating procedures on DToxS.org. Briefly, four commercially available cell lines of primary adult human cardiomyocytes were purchased from PromoCell GmbH (Cat #: C12810; Heidelberg, Germany), expanded and differentiated under serum-free conditions for 28 days per manufacturer's instructions. The four cell lines used in this study (Lot #: 3042901.2, 4031101.3, 2082801.2, 2120301.2) were isolated from two Caucasian male and two Caucasian female subjects, aged 54, 62, 61, 56, respectively (see cell lines PMC-A, B, D, E in Table 2). Although these cell lines were originally derived from the heart and have many cardiomyocyte-like properties, they are not excitable or contractile; hence, we refer to them as cardiomyocyte-like throughout the dataset. All four cell lines were subjected to rigorous standardized quality control metrics that included regular testing for mycoplasma contamination and confirmation of cardiac origin through RT-PCR screening of key cardiac genes (ACTN2, TNNT2, KCHN2, MEF2C, NKX2-5 and GATA4), immunofluorescent validation of cardiac-specific actinin crossfiber formation, and measurement of calcium activity. Details of extended cellular metadata for each line can also be found at DToxS.org. Briefly, cells were subcultured at 37 °C under 5% CO<sub>2</sub> with the manufacturer-supplied and serum-supplemented growth media (Cat #: C22270, C39275) up until passage four. Once they reached 100% confluence, cells were trypsinized, counted, and replated in serum-free growth media (Cat #: C22270), at a concentration of 40,000 cells/cm<sup>2</sup> in



60 mm tissue culture dishes. Cells were then differentiated under serum-free conditions for 28 days, whereby 50% of the media was replenished every other day. All experiments were performed on sixth or earlier passage cells. After 28 days of differentiation, cells were treated with individual KIs for 48 hours at a concentration equivalent to clinically-observed median peak plasma concentration. Forty-eight hours was chosen to mimic the *in situ* chronic conditions and to detect consistent, non-transient, proteomic changes. We note that no extensive cell death or apoptosis were observed. Before and after treatment cell counts and bright field images were obtained to quantify changes in cell numbers. After 48 hours of KI treatment, cells were lysed on ice using TRIzol (Thermo Fisher, Cat #: 15596026) for five minutes, scraped off the dish, and mixed. TRIzol lysates were mixed with chloroform according to manufacturer's instructions and the RNA-protein fractions were separated. From these organic partitions, RNA samples were processed, sequenced and analyzed as previously reported<sup>18</sup>. Results of the transcriptomic signatures are reported elsewhere<sup>17</sup>.

**Protein extraction.** After isolating RNA, the remaining TRIzol solutions were processed for protein extractions (Fig. 1a). Each TRIzol extract were first separated into two 1.5 mL Eppendorf tubes; each tube contained ~600  $\mu$ L of the TRIzol chloroform and methanol solvent mixture. Then to each tube, 900  $\mu$ L of pre-chilled cold acetone was added and mixed well. Five hundred  $\mu$ L of the mixture was transferred from each tube to a third new tube, and 500  $\mu$ L pre-chilled  $-20^{\circ}\text{C}$  acetone was added to each of the three tubes and mixed well. All three tubes were put on ice for at least 12 hours inside a cold room, and centrifuged at 16,000 rpm ( $\sim 23,469$  g) for 30 minutes,



**Fig. 2** MS/MS quality control for the DToxS dataset. A series of quality control steps have been implemented to obtain high quality data for this dataset. **(a)** A gel-based method was developed for protein estimation. In each SDS-PAGE gel, 25  $\mu\text{g}$  of HeLa cell protein extract was run as a reference to estimate the total protein amount of each sample (blue verticle lines). Based on the density of the CBB stain of each lane in relation to the HeLa stain density, the protein amount can be calculated (see examples in Online-only Table 1). After the tryptic digestions, the resulting peptides were diluted into 0.5  $\mu\text{g}/\text{ml}$  for LC-MS/MS analysis (orange verticle lines), based on the estimated protein amount in each sample. **(b)** Levey-Jennings performance chart showing instrument stability. Two hundred nanograms of commercial HeLa protein digest was analyzed regularly during the time period of the DToxS LINC data acquisition. Total ion current (TIC) from each month is shown. Red lines demarcate one, two and three times the standard deviation; LCL/UCL = lower/upper control limit. **(c)** Deep proteome coverage indicates good instrument sensitivity. Total protein groups identified from each cell line in this study are shown, with over  $\sim 5,000$  unique proteins identified per cell line. **(d)** The HeLa protein digests were used for LFQ normalization. Based on the MS1 LFQ counts of the HeLa cells from each gel (top panel), a normalization factor was calculated and applied to the MS1 counts of the samples. After the HeLa cell normalization, the LFQ variation was more stable.

Cell Name	Vendor	Cat No	Lot No	Race	Sex	Age	Disease
PMC-A	PromoCell	C-12810	3042901.2	Caucasian	Male	54	None
PMC-B	PromoCell	C-12810	4031101.3	Caucasian	Female	62	None
PMC-D	PromoCell	C-12810	2082801.2	Caucasian	Female	61	None
PMC-E	PromoCell	C-12810	2120301.2	Caucasian	Male	56	None

**Table 2.** Cellular metadata.

in a Beckman Coulter Allegra 64 R centrifuge. After centrifugation, the supernatants were removed and 1 mL of pre-chilled acetone was added to each tube to wash the protein pellets. The protein pellets were sonicated for five bursts at level-3 on a Branson Ultrasonic cleaner, repeated once. The protein solutions were placed on ice for four hours and pelleted *via* centrifugation as described above. The wash procedure for each protein pellet was repeated twice for a total of three times. The final protein pellet was washed with 50  $\mu$ L of pre-chilled acetone and centrifuged at 16,000 rpm for 10 minutes, and the remaining pellet was air dried at room temperature for 10 to 15 minutes. The protein pellet was then resolubilized with a vortex mixer, with 70  $\mu$ L of 8 M urea (Fisher Cat #: U15500) in 50 mM ammonium bicarbonate (Sigma Cat#: 09830-500 G), and stored at  $-80^{\circ}\text{C}$  before proteomics analysis. See additional protocol details at <https://martip03.u.hpc.mssm.edu/sop.php>.

**Sample preparation for proteomics analysis.** The protein solutions may contain TRIzol and other low mass contaminants from the RNA extraction buffer, which can confound accurate protein estimation via conventional biochemical assays. Consequently, we ran the SDS-PAGE gels to remove the low mass contaminants from the high mass proteins, and estimated the protein amounts via the stain intensities from the gels stained with the Coomassie brilliant blue (CBB) dye (Fig. 1b and Supplementary Fig. 1). In each gel, 25  $\mu$ g of a HeLa cell extract produced at the CAPR (aliquots of the same preparation were used for all gels) was run in one lane as the normalization reference, along with the proteins isolated from the drug-treated cell lines. After CBB staining of the gels, the density of each sample lane was measured using the ImageJ software (NIH), and the protein amount for each drug-treated sample was calculated using the following equation: protein ( $\mu$ g) = HeLa protein 25  $\mu$ g  $\times$  (Sample density/HeLa density). The protein quantity information was later used to dilute the tryptic peptides into 0.5  $\mu$ g protein equivalents per  $\mu$ L for the LC-MS/MS analysis (Online-only Table 1).

The in-gel digestions of proteins were performed following a procedure similar to the one described by the Mann lab<sup>19</sup>. In brief, the entire gel lane of each sample was excised into  $\sim 1\text{ cm}^2$  gel blocks, and washed four times with 10 mL each of a Wash Buffer containing 30% acetonitrile (ACN) and 70% of 100 mM  $\text{NH}_4\text{HCO}_3$  to remove the contaminants from the RNA extraction buffer. Subsequently, one mL of 25 mM dithiothreitol (DTT) solution was added to the gel blocks for disulfide reduction at  $55^{\circ}\text{C}$  for 30 minutes, and 1 mL of 50 mM iodoacetamide solution was then added for thiol alkylation at  $37^{\circ}\text{C}$  in the dark for 30 minutes. After alkylation, the gel blocks were dehydrated with 9 mL of ACN to remove both DTT and iodoacetamide. For in-gel trypsin digestion, one mL of trypsin solution (5  $\mu$ g/mL in 50 mM  $\text{NH}_4\text{HCO}_3$ ) was added into each sample, and incubated at  $37^{\circ}\text{C}$  for 16 hours. Resulting peptides were extracted, desalted with Pierce  $\text{C}_{18}$  spin columns (Thermo Scientific) based on the manufacturer's protocol and concentrated in a Speed Vac prior to LC-MS/MS analysis.

**LC-MS/MS analysis.** According to the gel-based protein amount estimates described above, the peptides from each sample were resuspended in Solvent A (2% ACN in 0.1% formic acid (FA)) at a final concentration of 0.5  $\mu$ g/ $\mu$ L (see examples in Online-only Table 1). Two micrograms of peptides from each sample were subjected to LC-MS/MS analysis on a Q Exactive Mass Spectrometer coupled with an UltiMate 3000 RSLCnano Duo LC system (Thermo Scientific). The peptides were first loaded onto a trapping column (Acclaim PepMap 100  $\text{C}_{18}$  trap column 75  $\mu$ m  $\times$  2 cm, 3  $\mu$ m, 100  $\text{\AA}$ ) and then separated on an Acclaim PepMap  $\text{C}_{18}$  column (75  $\mu$ m  $\times$  50 cm, 2  $\mu$ m, 100  $\text{\AA}$ ), using a 4-h binary gradient from 2–95% of Solvent B (85% ACN in 0.1% FA), at a flow rate of 250 nL/min. The eluted peptides were directly introduced into the MS system for data-dependent MS/MS analysis in the positive ion mode. The MS full scans were acquired in an m/z range of 400 to 1750 in the profile mode, with the AGC value specified at 3E6, and the injection time set at 100 msec. The resolution of the full MS scan was set to 140,000 at m/z 400. Following each full MS scan, the 15 most intense ions with charge states between  $2^+$  to  $5^+$  were selected within an isolation window of 2 m/z for the subsequent MS/MS analysis. The AGC of MS/MS analysis was set to 5E4 and the dynamic exclusion was 45 sec. The peptide ions were fragmented using higher energy collision dissociation at a NCE of 27. A total of 308 LC-MS/MS raw files were obtained for the proteins isolated across four different cell lines for 60 different types of drug treatment conditions. The reference HeLa digest from each SDS-PAGE gel was analyzed using the same method to normalize the LFQ intensities from the other cell digests. To quality control the LC-MS/MS system performance, we also ran a Pierce HeLa Protein Digest Standard (Pierce, Cat# 88329) on the instrument, prior to running each set of drug-treated samples.

**MaxQuant for protein identification and quantification.** In order to evaluate the quality of the data and compare the quantitative proteomic signatures among the drug-treated samples across different cell lines, the entire 372 raw LC-MS/MS dataset (308 samples and 64 HeLa cell controls,  $\sim 825$  GB of data) was submitted for database search using the Andromeda search engine on the MaxQuant platform (Version 1.6.0.13). The raw data files were loaded with “No fractions” option selected. Trypsin was selected as enzyme with two miss cleavages. Methionine oxidation (+15.9949 Da) and protein N-terminal acetylation (+42.0106 Da) were selected as various modifications and cysteine carbamidomethyl modification (+57.0215 Da) was set as a fixed modification. Initial search peptide mass tolerance was set to 20 ppm, and the main search peptide mass tolerance was set to

Group	PMC-A	PMC-B	PMC-D	PMC-E
Total Identified Protein Groups	4,288	4,635	4,881	5,227
Total Identified Peptides	1,552,043	3,176,801	1,980,185	3,141,618
Median LFQ Intensity	14,366,500	15,757,000	13,341,000	12,479,000

**Table 3.** Summary of MaxQuant analytics.

4.5 ppm. LFQ was selected for label-free quantification, with the minimal LFQ ratio count set at 2. The MS/MS spectra were searched against both UniProt human FASTA database (downloaded from <https://www.uniprot.org/proteomes/UP000005640> with the last modification date of 10/22/2018, containing 73,101 human protein sequences) and the MaxQuant default contaminants FASTA database (containing 245 protein sequences). Match between runs was selected to maximize protein identification and quantitation with a match time window of 0.7 minutes. The protein false discovery rate (FDR) was estimated using the decoy databases containing reversed sequences of the original proteins. Proteins identified (Table 3) with both protein and peptide FDRs at or less than 1% were included in the final results for the subsequent analyses.

### Data Records

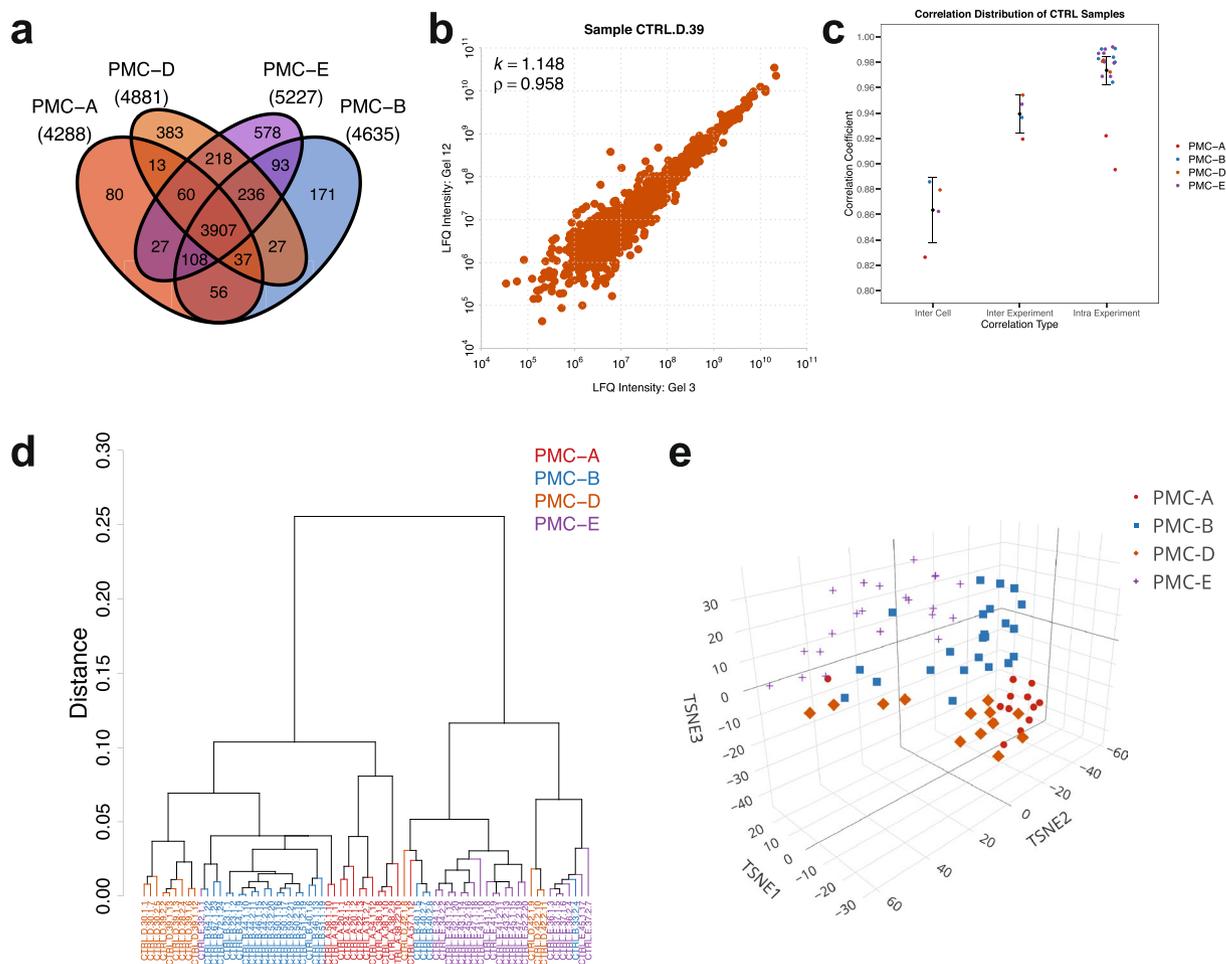
All the raw data described in this study have been uploaded to the PRIDE repository at the ProteomeXchange website (<https://identifiers.org/pride.project.PXD014791>) with the project accession PXD014791, and they are freely available for the research community<sup>20</sup>. Online-only Table 2 provides the metadata information of drug treatments, cell lines and replication numbers for all raw mass spectrometry data files. In addition, the processed higher level data is available through LINC Data Portal (<http://lincportal.ccs.miami.edu/dcic-portal/>) with the accession LDG-1444. Data include metadata, raw files of LC-MS/MS analysis, protein FASTA database, and protein identification and quantitation results from MaxQuant<sup>21</sup>.

### Technical Validation

We have developed a three-tiered quality control strategy to ensure the high throughput production of deep and reproducible DToxS proteomics datasets. This strategy, which focuses on samples, instrument and analytical pipeline, has enabled us to ensure high quality at (1) the protein sample level, (2) the LC-MS/MS instrument performance level, and (3) the overall quantitative data processing variations.

**Quality validation of the protein samples for LC-MS/MS.** Due to both the cost constraints of treating the human primary cardiomyocyte-like cells with KI drugs and the scientific necessity to compare proteomics data with RNAseq data from the same samples, the same drug-treated cells were used to analyze protein expression, after the RNA extraction. The challenges from this approach included occasionally uneven or poor protein yields, and imprecise protein concentration estimates with the established Bradford or BCA assays, due to the confounding components in the RNA extraction buffers. While the protein yield of our combined RNA-protein extraction approach was lower and more variable compared to standard “protein-only” protocols, we note that all of the samples investigated here had more than adequate protein yield necessary for LC-MS/MS. In order to rigorously compare drug effects with the given sample-to-sample variability in experimental processing, we performed an additional SDS-PAGE in-gel quantification step with a consistent positive control sample and internally-prepared HeLa protein lysates. These gels further served as an additional quality control step, as they were checked for signs of protein degradation (such as smearing), which were not observed in any of the samples. Seventy SDS-PAGE gels were run to increase the estimation accuracy of the proteins derived from the cells (see examples in Supplementary Fig. 1); in each gel, 25 µg of HeLa cell protein lysate was included in the first lane, as the reference for both accurate protein estimation and efficient in-gel digestion, and later for LC-MS/MS data normalization. In brief, the densities of the CBB for each KI-treated cell line in each gel lane were recorded (see examples in Online-only Table 1). The protein amounts from each sample were extrapolated by comparing its CBB density to that of the 25 µg of the internally-prepared HeLa proteins; the estimated protein amounts were later used to obtain 0.5 µg/µL equivalent of the tryptic peptides from each sample for LC-MS/MS analysis (see examples in Fig. 2a). For tryptic digestion, the drug-treated samples were digested along with the internally-prepared HeLa proteins separated on the same SDS-PAGE gels, which were later used for LC-MS/MS data normalization as well.

**Quality validation of the LC-MS/MS system.** To ensure that LC-MS/MS system could deliver overall satisfactory performance, 200 ng of commercial HeLa cell digest (Pierce, Cat# 88329) was run prior to each gel batch of the biological samples. To confirm well-behaved LC separations, MS1 ions for select peptides were carefully monitored for retention time shifts and peak widths. In order to minimize the run-to-run carry-over of peptides, a duo-LC system was configured in parallel for this study, so while one LC column was used for peptides separation and LC-MS/MS analysis, the other LC column could be washed to remove any residual peptides. For the overall LC-MS/MS quality control, the system is capable of identifying >3,800 unique proteins and >14,000 unique peptides from ~200 ng of the commercial HeLa cell digest. Indeed, the Levey-Jennings quality assurance plot showing the monthly total ion current (TIC) of the HeLa digest indicates consistent LC-MS/MS system performance for LFQ quantification (Fig. 2b). Qualitatively, the LC-MS/MS system enabled the deep proteome coverages (~5,000 unique protein groups/cell line) of all four DToxS cell lines (Fig. 3a and Table 3). Total 372 raw LC-MS/MS files, including the files for 308 drug treated cardiomyocyte-like cell line and 64 HeLa cell line samples, were submitted to the MaxQuant analysis (Table 3). In total, 83,674 unique peptides that mapped onto 5,994



**Fig. 3** Coverage and reproducibility of detected proteins in control samples. **(a)** Venn diagram of the overlapping and unique protein groups across all samples for the four cardiomyocyte lines PMC-A, B, D, E. **(b)** Normalized MS1 LQ intensities for the same sample measured from two separate SDS-PAGE gels in two different MS/MS runs had strong agreement. **(c)** The correlation (mean and variation) of LQ intensity between the control samples from the same experiment for each experiment versus the samples from different experiments in each cell line (cell lines PMC-A, B, D, E shown in color coordination). **(d)** The clustering of all biological replicates from four cell lines under control conditions based on the Euclidean distances of LQF intensities of replicates. **(e)** Accordingly, proteomic signatures of the four cell lines under control conditions across 71 biological replicates showed strong clustering based on the source cell line.

unique protein groups were identified with the FDR of less than 1%, among the four different cell lines. These proteins were mapped onto 5,784 unique genes. Given no multi-dimensional LC fractionation was performed, these deep proteome coverages confirm the effectiveness of the LC-MS/MS quality control procedures used for this study.

**Quality control of the quantitative data analysis.** To minimize the LC-MS/MS quantification noise derived from the sample preparation steps, the HeLa proteins used for the in-gel protein estimation were digested along with the drug-treated samples from the same gels. Thus, these HeLa cell protein digests were used to control for the trypsin digestion efficiencies, LC-MS/MS variations and for the quantitative normalization of the LQF data. For example, the TIC analyses indicates sample-to-sample variations among both HeLa and control (i.e., no drug treatment) cell digests (Fig. 2c, upper and lower panels, respectively); with HeLa TIC signal normalization, the variations in the control cell digests were further reduced (Fig. 2d). We also randomized all samples into separate gels and included multiple technical replicates of control samples into the gels to prevent time-dependent artifacts that may arise from instrumental drift or lot-to-lot variations in reagents.

Overall, our method for protein count normalization showed good reproducibility (Fig. 3b), which can be quantitatively assessed using Pearson correlation across two technical replicates that were analyzed on different days from two different HeLa signal normalization protocols (i.e., different SDS-PAGE gels). This was consistent for all technical replicates in this dataset (Supplementary Fig. 2). In fact, the average slope for sample-to-sample comparison was remarkably close to unity:  $1.03 \pm 0.16$ . Similarly, the average coefficient of determination for all

technical replicates in the study was  $0.97 \pm 0.03$ . We also note that both the depth of protein quantification for each replicate and the overlap of detected proteins between different biological replicates were consistently high for all cardiomyocyte-like cell lines (Fig. 3c), suggesting that our experimental pipeline was robust in identification of proteomic signatures. The quantification strategy was also successful in reproducible identification of subtle proteomic differences between individual cardiomyocyte cell lines PMC-A, B, D and E, which clustered consistently across most biological replicates using either Pearson (Fig. 3d) or Euclidean (Fig. 3e) distance metrics. Furthermore, we used overlap analysis, to show that most of the identified protein groups were common to either all four cell lines (3,907 or 65%) or detected in at least three out of four lines (4,348 or 73%) with only 1,212 (or 20%) protein groups detected in uniquely one of the lines (Fig. 3a).

**Identification of differentially expressed proteins.** The normalized LFQ intensities of identified proteins were processed by a custom computational pipeline for identifying differentially expressed proteins (DEPs) that consists of four parts: (1) general logarithmic transformation and variance stabilization; (2) linear model fitting for treatment comparison; (3) empirical Bayes test for differential statistical significance; and (4) Fisher's test for combining experiment-wise differential comparison results (Supplementary Fig. 3).

First, for each experiment on each cell line for each drug treatment, the LFQ intensity measurements of drug-treated and control conditions are transformed with a general logarithm formula based on inverse hyperbolic sine function using the *vsnMatrix* and *predict* function of the *vsn* R package such that the variance of transformed data becomes independent of its mean value. Then the transformed intensities of the samples are fitted to the linear model and the fitting coefficients are estimated using the *lmFit* and *contrasts.fit* functions from the *limma* R package. Next, these fitting parameters are given to the empirical Bayes test using the *eBayes* function with its intensity-based trend parameter turned on (for the purpose of dynamic prior estimation) to calculate various statistical metrics for the differential comparison between the drug-treated samples and the control samples. Finally, the p-values from all individual independent differential comparisons are combined using the Fisher's method with the *sumlog* function of the *metap* R package to obtain a single meta p-value for the comparison between all the drug-treated samples and the control samples for a cell line and a drug treatment. This procedure is then repeated for all the combinations of cell lines and drug treatments. This produces a list of reference proteins ranked by their p-values (from most to least significant). These ranked protein lists together with their cell lines and treated drugs are passed to the downstream analysis.

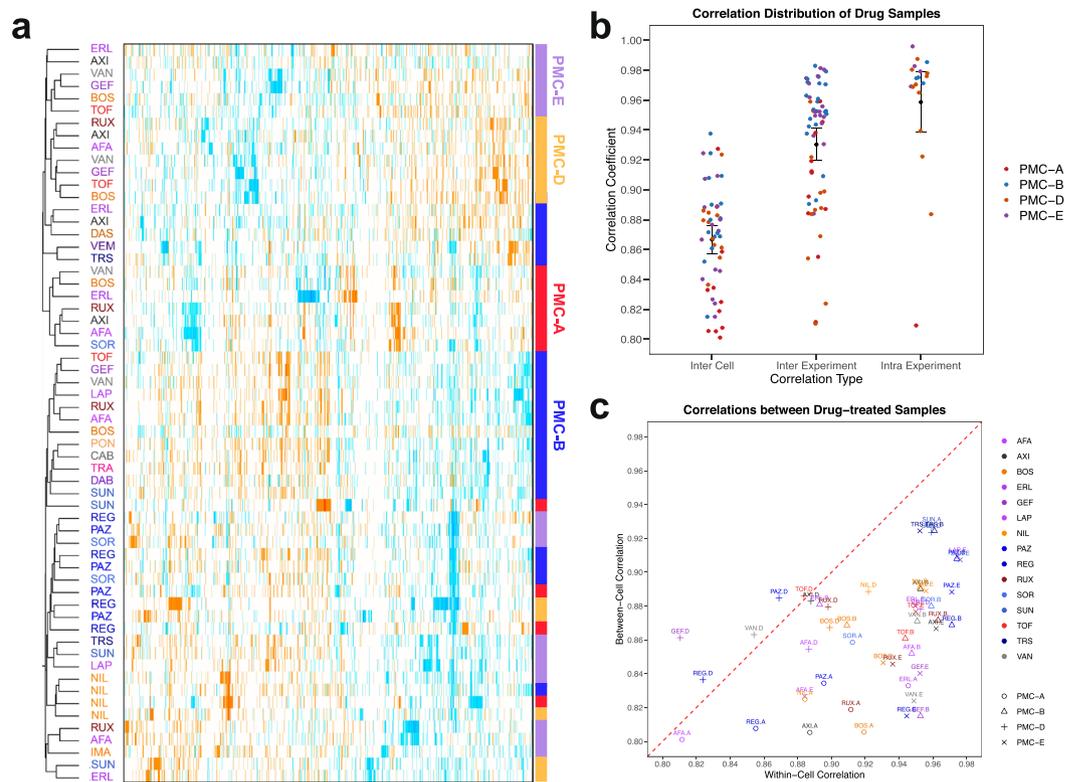
**Downstream analysis of differentially expressed proteins.** To better collate the cell line and drug specific responses, we first merged all DEPs of the same cell line treated with the same drug to one new list of DEPs. For each DEP list, we then calculated a combined p-value defined as the geometric mean of the individual p-values of the different treatments and a combined  $\log_2$ (fold changes) defined as the average mean of the individual  $\log_2$ (fold changes). For all downstream analyses, we considered the top 100 DEPs (based on the combined p-values). Combined drug treatments of the different cell lines were hierarchically clustered based on the Euclidean distance of the combined  $\log_2$ (fold changes) using Cluster 3.0<sup>22,23</sup> (Fig. 4a). In order to investigate the consistency between the different experiments for the same drug treatment across the four cell lines, we ran successive Pearson correlations between different replicates for a given drug treatment of the same cell compared to the correlation among different cell types (Fig. 4b).

## Usage Notes

**Use case 1: Identification of KI drug-induced protein expression profiles.** This DToxS dataset can be mined to identify protein expression changes triggered by each KI drug treatment of the four cardiac cell lines derived from human tissues. For example, we note that the DEP signatures from different cell lines and KIs clustered together based on the impact of the drug. While KIs that had modest proteomic changes clustered along the cell lines (top clusters of Fig. 4a), those that have more substantial differential protein expression signatures (bottom clusters in Fig. 4a that include nilotinib, sorafenib, pazopanib and regorafenib, all of which are reported to have cardiotoxic adverse associations including QT prolongation and myocardial infarction) clustered on their own irrespective of the cell line. The proteomic changes identified in this study can be used to understand the cell signal transduction events that may underlie KI-induced cardiotoxicities. Also, with the LC-MS/MS information provided in this dataset, follow-up proteomics assays can be developed to validate the protein expression changes in larger number of samples; such assays may include either targeted assays, e.g., selected or parallel reaction monitoring or untargeted proteomics assays, e.g., data-independent analysis.

**Use case 2: Variability of differential KI drug-induced protein expression profiles.** This dataset can be analyzed to compare and cluster the proteomic changes induced by different KIs, thus ascertaining the similarities and idiosyncrasies of different drugs and patient (or cell line) specific responses. One can also quantify the variability and consistency of drug-induced proteomic changes within a given subject (i.e., within a given cell line) compared across a population (i.e., between lines) to determine the generalizability of a given drug effect on cardiomyocytes (Fig. 4c).

**Use case 3: Integration of transcript and protein expression profiles for comprehensive network studies.** Drugs may induce complex changes on either gene or protein expression. Since proteomic and transcriptomic changes in complex human cells are not always in agreement (due to technical bias, differential regulation or turnover rates), it may be desirable to integrate both gene and protein expression datasets from the same samples, in order to obtain more comprehensive understanding of the drug-induced changes among the signaling networks. For this purpose, the DToxS proteomics dataset presented here can be integrated with the DToxS transcriptomics dataset<sup>17</sup> to fully understand the effects of KI-induced cardiotoxicities. We note that 97% of the protein groups in our dataset map to individual genes (Supplementary Fig. 4).



**Fig. 4** Reproducibility of differential protein expression measurements for drug-treated samples. **(a)** Unsupervised clustering of top 100 differentially expressed proteins in KI treated cardiomyocyte-like cell lines. Color-coding on the left reflects the primary target of the drugs (see Table 1 for further details), whereas the bars on the right highlight the cell line. While KIs with weak proteomic signatures cluster along cell lines, those that have strong differential protein expression cluster along the primary target profile of the drugs. **(b)** The similarity and variability of drug-treated samples within and between cell lines as determined by Pearson correlation between drug treated replicates and controls. **(c)** A summary of the drugs showing similar profiles between cells versus those showing different profiles within cells.

**Use case 4: Identification of drug-induced post-translational modifications (PTMs) and regulated proteolysis.** Drugs may induce changes among signaling pathways not only *via* the changes of gene and protein expressions, but also *via* the rapid regulations of protein functions by altering specific PTMs (*e.g.*, phosphorylation) and regulated proteolytic events (*e.g.*, caspase 3 activation)<sup>24–26</sup>. Therefore, alternative protein database search schemes can be carried out to analyze the raw DToxS proteomics dataset (Supplementary Fig. 5a), and identify KI-induced changes among specific PTMs (Supplementary Fig. 5b,c) or semi-tryptic or non-tryptic peptides derived from regulated proteolysis. Since the current workflow did not include the enrichments of the subproteomes specific for PTMs, the yield for peptides containing PTMs or non-tryptic cleavages will likely be modest. As such, the results from the alternative analysis of the DToxS dataset may provide information for future proteomics studies that focus on the enrichments of specific PTMs, *e.g.* acetylation and phosphorylation.

### Code availability

All of the annotated custom code and scripts used for the generation of this dataset is publicly available at the DToxS GitHub repository at <https://github.com/DToxS>.

Received: 11 January 2021; Accepted: 14 December 2021;

Published online: 20 January 2022

### References

- Gross, S., Rahal, R., Stransky, N., Lengauer, C. & Hoeflich, K. P. Targeting cancer with kinase inhibitors. *The Journal of clinical investigation* **125**, 1780–1789 (2015).
- Calizo, R. C. *et al.* Disruption of podocyte cytoskeletal biomechanics by dasatinib leads to nephrotoxicity. *Nat Commun* **10**, 2061 (2019).
- Sharma, A. *et al.* High-throughput screening of tyrosine kinase inhibitor cardiotoxicity with human induced pluripotent stem cells. *Science translational medicine* **9** (2017).
- Keenan, A. B. *et al.* The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations. *Cell Syst* (2017).
- Schwanhauss, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
- Cancer Genome, Atlas Research, N. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315–322 (2014).

7. Azeloglu, E. U. *et al.* Interconnected network motifs control podocyte morphology and kidney function. *Science signaling* **7**, ra12 (2014).
8. Trinh, H. V. *et al.* iTRAQ-Based and Label-Free Proteomics Approaches for Studies of Human Adenovirus Infections. *Int J Proteomics* **2013**, 581862 (2013).
9. Johnson, E. C. B. *et al.* Deep proteomic network analysis of Alzheimer's disease brain reveals alterations in RNA binding proteins and RNA splicing associated with disease. *Mol Neurodegener* **13**, 52 (2018).
10. Beer, L. A., Liu, P., Ky, B., Barnhart, K. T. & Speicher, D. W. Efficient Quantitative Comparisons of Plasma Proteomes Using Label-Free Analysis with MaxQuant. *Methods Mol Biol* **1619**, 339–352 (2017).
11. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513–2526 (2014).
12. Anand, S., Samuel, M., Ang, C. S., Keerthikumar, S. & Mathivanan, S. Label-Based and Label-Free Strategies for Protein Quantitation. *Methods Mol Biol* **1549**, 31–43 (2017).
13. Megger, D. A. *et al.* Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines. *Biochimica et biophysica acta* **1844**, 967–976 (2014).
14. Patel, V. J. *et al.* A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *J Proteome Res* **8**, 3752–3759 (2009).
15. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **389**, 1017–1031 (2007).
16. Bantscheff, M., Lemeer, S., Savitski, M. M. & Kuster, B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **404**, 939–965 (2012).
17. van Hasselt, J. G. C. *et al.* Transcriptomic profiling of human cardiac cells predicts protein kinase inhibitor-associated cardiotoxicity. *Nat Commun* **11**, 4809 (2020).
18. Xiong, Y. *et al.* A Comparison of mRNA Sequencing with Random Primed and 3'-Directed Libraries. *Sci Rep* **7**, 14626 (2017).
19. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**, 2856–2860 (2006).
20. Xiong, Y., Azeloglu, E. U. & Li, H. Proteomic cellular signatures of kinase inhibitor-induced cardiotoxicity. *PRIDE* <https://identifiers.org/pride.project:PX0014791> (2021).
21. Iyengar, S. R. Proteomic cellular signatures of kinase inhibitor-induced cardiotoxicity (LDG-1444: LDS-1575, LDS-1576, LDS-1577, LDS-1578, LDS-1579) *LINCS Data Portal* <http://identifiers.org/lincs.data/LDG-1444> (2020).
22. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**, 14863–14868 (1998).
23. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).
24. Liu, T. *et al.* Identification of differentially expressed proteins in experimental autoimmune encephalomyelitis (EAE) by proteomic analysis of the spinal cord. *J Proteome Res* **6**, 2565–2575 (2007).
25. Grant, J. E. *et al.* Post-translational modifications in the rat lumbar spinal cord in experimental autoimmune encephalomyelitis. *J Proteome Res* **6**, 2786–2791 (2007).
26. Jain, M. R. *et al.* Proteomic identification of immunoproteasome accumulation in formalin-fixed rodent spinal cords with experimental autoimmune encephalomyelitis. *J Proteome Res* **11**, 1791–1803 (2012).
27. Schnell, D. *et al.* Pharmacokinetics of afatinib in subjects with mild or moderate hepatic impairment. *Cancer Chemother Pharmacol* **74**, 267–275 (2014).
28. Stopfer, P. *et al.* Afatinib pharmacokinetics and metabolism after oral administration to healthy male volunteers. *Cancer Chemother Pharmacol* **69**, 1051–1061 (2012).
29. Pithavala, Y. K. *et al.* Evaluation of the effect of food on the pharmacokinetics of axitinib in healthy volunteers. *Cancer Chemother Pharmacol* **70**, 103–112 (2012).
30. Rugo, H. S. *et al.* Phase I trial of the oral antiangiogenesis agent AG-013736 in patients with advanced solid tumors: pharmacokinetic and clinical results. *J Clin Oncol* **23**, 5474–5483 (2005).
31. Abbas, R. *et al.* A phase I ascending single-dose study of the safety, tolerability, and pharmacokinetics of bosutinib (SKI-606) in healthy adult subjects. *Cancer Chemother Pharmacol* **69**, 221–227 (2012).
32. Kurzrock, R. *et al.* Activity of XL184 (Cabozantinib), an oral tyrosine kinase inhibitor, in patients with medullary thyroid cancer. *J Clin Oncol* **29**, 2660–2666 (2011).
33. Ouellet, D. *et al.* Population pharmacokinetics of dabrafenib, a BRAF inhibitor: effect of dose, time, covariates, and relationship with its metabolites. *J Clin Pharmacol* **54**, 696–706 (2014).
34. Falchook, G. S. *et al.* Dabrafenib in patients with melanoma, untreated brain metastases, and other solid tumours: a phase 1 dose-escalation trial. *Lancet* **379**, 1893–1901 (2012).
35. McCaig, A. M., Cosimo, E., Leach, M. T. & Michie, A. M. Dasatinib inhibits CXCR4 signaling in chronic lymphocytic leukaemia cells and impairs migration towards CXCL12. *PLoS one* **7**, e48929 (2012).
36. Herbst, R. S. *et al.* TRIBUTE: a phase III trial of erlotinib hydrochloride (OSI-774) combined with carboplatin and paclitaxel chemotherapy in advanced non-small-cell lung cancer. *J Clin Oncol* **23**, 5892–5899 (2005).
37. Soulieres, D. *et al.* Multicenter phase II study of erlotinib, an oral epidermal growth factor receptor tyrosine kinase inhibitor, in patients with recurrent or metastatic squamous cell cancer of the head and neck. *J Clin Oncol* **22**, 77–85 (2004).
38. Hirano, S. *et al.* The pharmacokinetics and long-term therapeutic effects of gefitinib in patients with lung adenocarcinoma harboring the epidermal growth factor receptor (EGFR) mutation. *Gan To Kagaku Ryoho* **39**, 1501–1506 (2012).
39. Demetri, G. D. *et al.* Imatinib plasma levels are correlated with clinical benefit in patients with unresectable/metastatic gastrointestinal stromal tumors. *J Clin Oncol* **27**, 3141–3147 (2009).
40. LoRusso, P. M. *et al.* Phase I and pharmacokinetic study of lapatinib and docetaxel in patients with advanced cancer. *J Clin Oncol* **26**, 3051–3056 (2008).
41. Trent, J. & Molimard, M. Pharmacokinetics and pharmacodynamics of nilotinib in gastrointestinal stromal tumors. *Semin Oncol* **38**(Suppl 1), S28–33 (2011).
42. Tanaka, C. *et al.* Clinical pharmacokinetics of the BCR-ABL tyrosine kinase inhibitor nilotinib. *Clin Pharmacol Ther* **87**, 197–203 (2010).
43. Hurwitz, H. I. *et al.* Phase I trial of pazopanib in patients with advanced cancer. *Clin Cancer Res* **15**, 4220–4227 (2009).
44. Narasimhan, N. I., Dorer, D. J., Niland, K., Haluska, F. & Sonnichsen, D. Effects of ketoconazole on the pharmacokinetics of ponatinib in healthy subjects. *J Clin Pharmacol* **53**, 974–981 (2013).
45. Rey, J. B., Launay-Vacher, V. & Tournigand, C. Regorafenib as a single-agent in the treatment of patients with gastrointestinal tumors: an overview for pharmacists. *Target Oncol* **10**, 199–213 (2015).
46. Shi, J. G. *et al.* The pharmacokinetics, pharmacodynamics, and safety of orally dosed INCB018424 phosphate in healthy volunteers. *J Clin Pharmacol* **51**, 1644–1654 (2011).
47. Chen, X. *et al.* Pharmacokinetics and pharmacodynamics of orally administered ruxolitinib (INCB018424 phosphate) in renal and hepatic impairment patients. *Clin Pharmacol Drug Dev* **3**, 34–42 (2014).
48. Minkin, P. *et al.* Quantification of sunitinib in human plasma by high-performance liquid chromatography-tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **874**, 84–88 (2008).

49. Faivre, S. *et al.* Safety, pharmacokinetic, and antitumor activity of SU11248, a novel oral multitarget tyrosine kinase inhibitor, in patients with cancer. *J Clin Oncol* **24**, 25–35 (2006).
50. Kim, K. B. *et al.* Phase II study of the MEK1/MEK2 inhibitor Trametinib in patients with metastatic BRAF-mutant cutaneous melanoma previously treated with or without a BRAF inhibitor. *J Clin Oncol* **31**, 482–489 (2013).
51. Martin, P. *et al.* Pharmacokinetic drug interactions with vandetanib during coadministration with rifampicin or itraconazole. *Drugs R D* **11**, 37–51 (2011).
52. Grippo, J. F. *et al.* A phase I, randomized, open-label study of the multiple-dose pharmacokinetics of vemurafenib in patients with BRAF V600E mutation-positive metastatic melanoma. *Cancer Chemother Pharmacol* **73**, 103–111 (2014).

## Acknowledgements

The authors are grateful for the support from the NIH Common Fund's Library of Integrated Cellular Signatures (LINCS) grant U54 HG008098 to R.I., M.R.B., E.A.S. In addition, H.L. is supported in part by NIH R01 GM112415; and E.U.A. is supported in part by NIH R01 DK118222. Mass spectrometer purchases were supported by NIH grant P30 NS046593 from the National Institute of Neurological Disorders and Stroke and another NIH grant 1S10 OD025047 from the Office of the Director of the National Institutes of Health. S.S. and D.V. are supported by NIH U54 HL127624. The content is sole responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Author contributions

R.I., E.U.A. and H.L. conceived the project. Y.X. developed the analytical pipeline for proteomic data processing and tabulated all data and metadata. T.L. executed the LC-MS/MS workflow and data analysis. T.C. performed protein SDS-PAGE analysis and in-gel tryptic digestions. B.H., Y.C., G.J. performed cell culture experiments along with protein isolation. J.H. developed the downstream bioinformatic pipeline. S.S., D.V. and J.G. assisted in the development and organization of quality control and metadata standards. H.L. designed the proteomics sample preparation and proteomics analysis workflow from expert input from M.R.B., E.A.S., R.I., and E.U.A. The manuscript was written by H.L., T.L. and E.U.A. with critical input from all the authors. E.U.A. provided overall supervision for all phases of the project.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-021-01114-3>.

**Correspondence** and requests for materials should be addressed to R.I., H.L. or E.U.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2022