

In Vitro Kinase-to-Phosphosite Database (iKiP-DB) Predicts Kinase Activity in Phosphoproteomic Datasets

Tommaso Mari, Kirstin Mösbauer, Emanuel Wyler, Markus Landthaler, Christian Drosten, and Matthias Selbach*



Cite This: *J. Proteome Res.* 2022, 21, 1575–1587



Read Online

ACCESS |

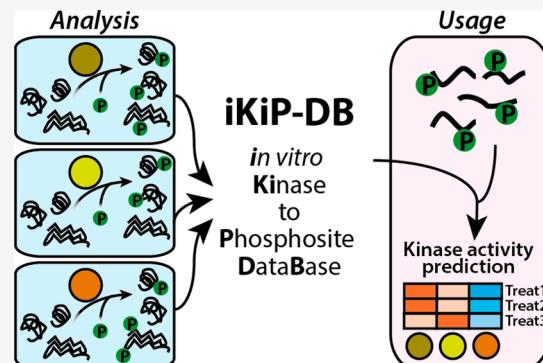
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Phosphoproteomics routinely quantifies changes in the levels of thousands of phosphorylation sites, but functional analysis of such data remains a major challenge. While databases like PhosphoSitePlus contain information about many phosphorylation sites, the vast majority of known sites is not assigned to any protein kinase. Assigning changes in the phosphoproteome to the activity of individual kinases therefore remains a key challenge. A recent large-scale study systematically identified *in vitro* substrates for most human protein kinases. Here, we reprocessed and filtered these data to generate an *in vitro* Kinase-to-Phosphosite database (iKiP-DB). We show that iKiP-DB can accurately predict changes in kinase activity in published phosphoproteomic data sets for both well-studied and poorly characterized kinases. We apply iKiP-DB to a newly generated phosphoproteomic analysis of SARS-CoV-2 infected human lung epithelial cells and provide evidence for coronavirus-induced changes in host cell kinase activity. In summary, we show that iKiP-DB is widely applicable to facilitate the functional analysis of phosphoproteomic data sets.

KEYWORDS: phosphoproteomics, kinase enrichment, phosphosite annotations, phosphosites database, mass spectrometry, tandem mass tags, SARS-CoV-2



INTRODUCTION

Mass spectrometry (MS)-based proteomics provides extensive information about thousands of post-translational modifications (PTMs).^{1–3} Phosphorylation of serine, threonine, or tyrosine residues is the most extensively studied PTM in humans,^{4,5} and over 200 000 phosphorylation sites have been described.⁶ However, the vast majority of these sites have neither an annotated kinase nor a known biological function.^{7,8} Therefore, while phosphoproteomics can now routinely quantify changes in thousands of phosphorylation sites,^{9,10} the functional interpretation of these data remains a major challenge.

Computational methods to predict kinase activity in phosphoproteomic data sets can be broadly divided into two categories. The first category consists of algorithms aiming at assigning kinases to unannotated phosphosites by integrating known kinase-substrate associations with other biological features (e.g., interaction networks or coexpression profiles).^{11–15} The output of such algorithms are scores representing the likelihood of a certain site to be phosphorylated by a specific kinase. The second category of methods seeks to annotate changes in kinase activity in quantitative phosphoproteomics data sets. To this end, observed changes in phosphopeptide levels across conditions are integrated with

kinase-substrate annotations.^{16–18} The output of such algorithms is kinase activity profiles across conditions. Both categories of algorithms share the need for a resource representing kinase-to-phosphosite associations. The most comprehensive resource for phosphosites function is PhosphoSitePlus (PSP).⁶ However, fewer than 3% of known phosphosites have either a reported function or a known regulatory kinase in PSP.⁷ Moreover, the distribution of kinase-to-phosphosite associations in PSP is skewed, with a minority of well-studied kinases making up the majority of entries in the database. This bias remains essentially unchanged, despite tremendous technological developments in the field.¹⁹

Here, we address the challenge of predicting kinase activity in phosphoproteomic data from a different angle: instead of developing algorithms that rely on existing annotation, we expand on the knowledge of kinase-to-phosphosite annotation. To this end, we took advantage of recently published large-

Received: April 6, 2022

Published: May 24, 2022



scale in vitro kinase data for over 300 human protein kinases.²⁰ We reanalyzed and filtered these data to compile it into an in vitro kinase-to-phosphosite database, or iKiP-DB. Using several published phosphoproteomics data sets of kinase activation/inhibition, we show that the iKiP-DB outperforms PSP in its ability to detect changes in kinase activity. Finally, to apply iKiP-DB, we infected lung epithelial cells with SARS-CoV-2 and investigated changes in cellular kinase activity induced by the novel coronavirus.

■ EXPERIMENTAL SECTION

Analysis of In Vitro Kinase Assays

In vitro kinase assay data from the work of Sugiyama and colleagues²⁰ deposited on ProteomeXchange (data set identifier PXD011366) was downloaded and reanalyzed using MaxQuant v1.6.10.43.²¹ Each kinase assay was assigned a specific experimental group in the experimental design of MaxQuant, and, when present, replicate experiments were combined in the same experimental group. The quantification method was set to dimethyl labeling, as described in the original publication. The MS scans were searched against the Human Uniprot database (2018-07, 95 128 entries) using the Andromeda search engine. The false discovery rate (FDR) was calculated based on searches on a pseudoreverse database and set to 0.05. The search included as fixed modifications carbamidomethylation of cysteine and as variable modifications methionine oxidation, N-terminal acetylation, asparagine, and glutamine deamidation as well as Phospho (STY). Trypsin/P was set as protease for in silico digestion of the reference proteome. Phosphosites were filtered for reverse hits, contaminants, and phosphosites with a localization probability lower than 0.75. Additionally, we only considered singly phosphorylated sites for the database. Phosphopeptide intensities were log2-transformed and corrected for the intensity measured in the light dimethyl label channel, corresponding to the control experiment. As in the original publication,²⁰ phosphosites were considered specifically phosphorylated by a recombinant kinase when having a log2-transformed ratio treatment over control higher than 1, a threshold found to efficiently distinguish genuine in vitro kinase substrates from remaining phosphorylation sites in the HeLa lysate that escaped dephosphorylation. From these kinase-substrate lists we removed all sites that were not contained in a large-scale reanalysis of 110 human phosphoproteomics experiments,⁸ which was downloaded from the PRIDE repository (data set identifier PXD012174). As a second filter, we removed all redundant phosphosites assigned to 20 or more kinases. Finally, we excluded all kinase sets (as in all phosphosites assigned to the same kinase) with fewer than five annotated phosphosites. We then extracted the ± 7 amino acid sequence windows around the phosphorylated site to be used as the specific identifier for each site and combined all kinase sets into a single GMT file to be used with the ssGSEA suite to perform a PTM set enrichment analysis (PTM-SEA).¹⁶ Since our database is exclusively composed of kinase-substrate associations, all sites were marked as “up” sites. The database is built to work with the flanking sequence centric mode of PTM-SEA. To compare our database with PSP, we extracted the phosphosites of PTMsigDB¹⁶ annotated as “KINASE-PSP”.

Preparation of Phosphoproteomics Data Set for Benchmarking

All data sets used for the benchmarking of iKiP-DB were downloaded from the supplementary tables containing information on phosphopeptide data of the respective publications.^{22–26} When not already provided, we filtered for single-phosphorylated, high-confidence phosphosites (localization probability >0.75) and calculated log2-transformed ratios of treatment(s) over control(s). We then extracted the ± 7 amino acid sequence windows around the phosphorylation site (± 6 for Rosales et al.²⁶) and removed duplicate entries. For each phosphoproteomics data set, the resulting sites and ratio information were exported as GCT files.¹⁶

Calculation of Kinase Enrichments

We calculated all our kinase enrichments using the ssGSEA2.0 suite^{16,27} using the standard settings: rank sample normalization, weight for the Kolmogorov–Smirnov statistic of 0.75, area under the resulting curve as main statistic and normalized enrichment score (NES) as output score, minimum overlap between query and test set of five sites and 1000 permutation for *p*-value and NES calculations. All benchmarking data sets were analyzed using iKiP-DB and PTMsigDB¹⁶ (v1.9.0) separately. Since PTMsigDB contains annotations for several experimental categories, we considered exclusively sets of kinase-substrate associations (“KINASE-PSP” group). Since Rosales and colleagues²⁶ only provided ± 6 sequence windows for their data set, we adapted iKiP-DB and PTMsigDB by cutting the sequences contained in the databases to ± 6 and removing all duplicate entries that resulted from this cut. Analyses outputs were further processed using in-house made scripts with the R programming language. Kinase-substrate enrichment analysis (KSEA) scores were calculated using in-house R scripts based on the formula described in the original publication.²⁸

Positional Weight Matrices Calculations and Evaluation

To compare positional weight matrices (PWMs) of iKiP-DB, we downloaded the NetworKIN (v3.1) collection of phosphosites from the KinomeXplorer platform.¹² PWMs were calculated with the same procedure for both iKiP-DB and NetworKIN data. First, we filtered out every kinase with fewer than 50 annotated phosphosites. Then for each kinase we calculated a position probability matrix for the occurrence of each possible amino acid at each position (± 6 AA from the phosphorylated amino acid). PWMs were then derived by dividing each amino acid probability by the expected background probabilities of the occurrence of that specific amino acid in the human proteome. To calculate the similarity between different PWMs, we used the Wasserman-Sandelin score²⁹

$$S = 2 - \sum_{b \in \{A, R, N, \dots, V\}} (M_b - N_b)^2$$

where M and N are the compared pair of columns in two different matrix models, and $\{A, R, N, \dots, V\}$ are the 20 canonical amino acids. The significance of the difference between the distribution of the Wassermann-Sandelin scores between same kinase pairs and different kinase pairs from iKiP-DB and NetworKIN was calculated with the Mann–Whitney U-test. Sequence logos were generated with the ggseqlogo R package.³⁰

SARS-CoV-2 Infections of Calu-3 Cells

Experiments were performed and generated in the same context as described previously by Wyler, Mösbauer, Franke, and colleagues.³¹ Briefly, Calu-3 cells (ATCC HTB-55) were cultivated in Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal calf serum, 1% nonessential amino acids, 1% L-glutamine, and 1% sodium pyruvate in a 5% CO₂ atm at 37 °C. SARS-CoV-2 (Patient isolate, BetaCoV/Munich/BavPat1/2020|EPI_ISL_406862) was grown on Vero E6 cells and concentrated using Vivaspin 20 concentrators (Sartorius Stedim Biotech). Virus stocks were stored at -80 °C, diluted in OptiPro serum-free medium supplemented with 0.5% gelatin and phosphate-buffered saline. Mock infected controls were generated with cells inoculated with cell-culture supernatants from uninfected Vero cells in accordance with virus stock preparation. For the infection experiments, Calu-3 cells were seeded at 6 × 10⁵ cells/mL, and after 24 h the cells were infected with SARS-CoV-2 at a multiplicity of infection (MOI) of 0.33 or with Vero E6 medium as control. Samples were harvested in three biological replicates after 4, 8, and 12 h with prewarmed trypsin for 3 min at 37 °C, after removal of the cell culture media. Mock infected controls were similarly harvested in biological triplicates after 4 and 12 h. Samples were lysed and inactivated through boiling in sodium dodecyl sulfate (SDS) sample buffer. All infection experiments were performed under biosafety level three conditions with enhanced respiratory personal protection equipment.

(Phospho)-Proteomics Sample Preparation

Proteomics samples were prepared combining single-pot, solid-phase-enhanced sample preparation (SP3)³² and tandem mass tag (TMT) labeling³³ using TMTpro reagents.³⁴ Briefly, lysates were precleared via centrifugation, then reduced and alkylated with dithiothreitol (DTT) and iodoacetamide, respectively. Proteins were then incubated with a 1:1 mix of SeraMag beads A and B at a 10:1 weight/weight bead/protein ratio. Protein binding to the hydrophilic beads was induced by adding acetonitrile (ACN) and washing with 80% EtOH to remove contaminants. Protein digestion was performed on beads in 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) at pH 8, with Trypsin and LysC at a 1:50 protein/enzyme ratio for ~16 h. After digestion, peptides were quantified via a bicinchoninic acid assay (BCA) and directly labeled with TMTpro reagents (Thermo Fisher Scientific; product No. A44520, lot No. UL297970) following the manufacturer's protocol. Samples were randomly assigned to a TMT channel as follows: CoV2 4 h repA → 128C, CoV2 4 h repB → 129N, CoV2 4 h repC → 133N, CoV2 8 h repA → 133C, CoV2 8 h repB → 127N, CoV2 8 h repC → 131C, CoV2 12 h repA → 130C, CoV2 12 h repB → 130N, CoV2 12 h repC → 131N, mock 4 h repA → 132N, mock 4 h repB → 128N, mock 4 h repC → 132C, mock 12 h repA → 129C, mock 12 h repB → 126, mock 12 h repC → 127C. The last TMTpro channel (134) was comprised of a supermix of the other samples. Labeled peptides were pooled in equal amounts and desalting with SepPak columns (Waters), and the resulting peptide mixture was offline-separated with high-pH reverse-phase fractionation^{33,35} on a Dionex 3000 system (Thermo Fisher Scientific) and an XBridge Peptide BEH C18 (130 Å, 3.5 μm; 2.1 mm × 250 mm) column (Waters). Peptides were resuspended in high-pH buffer A (5 mM ammonium formate, 2% ACN) and separated on a multistep gradient from 0 to 60%

high-pH buffer B (5 mM ammonium formate, 90% ACN) 96 min long and collected in 96 fractions (1 fraction/min). The fractions were automatically pooled during collections, where each x th fraction was combined with the $x + 25$ th, $x + 49$ th, $x + 73$ rd fraction for a total of 24 fractions. Of each pooled fraction ~1 μg of peptide was subjected to a mass spectrometric analysis for total proteome measurement. The remaining peptides were further pooled into 12 fractions and used as input for a phosphopeptide enrichment via immobilized metal affinity chromatography (IMAC), which was performed by the Bravo Automated Liquid Handling Platform (Agilent) with AssayMAP Fe(III)-NTA cartridges.

LC-MS/MS Analysis

Proteome and phosphoproteome fractions were online-fractionated on an EASY-nLC 1200 and acquired on an Exploris 480 mass spectrometer (Thermo Fisher Scientific) operated on profile-centroid mode, as previously described.³⁶ Peptide separation was achieved on a fused silica, 25 cm long column packed in-house with C18-AQ 1.9 μm beads kept at a temperature of 45 °C. Mobile phase A consisted of 0.1% formic acid (FA) and 3% ACN in water, while mobile phase B consisted of 0.1% FA and 90% ACN. After column equilibration peptides resuspended in buffer A were separated with a 250 nL/min flow on a 110 min gradient/mobile phase B increased from 4% to 30% in the first 88 min, followed by an increase to 60% in the following 10 min, to then reach 90% in 1 min, which was held for 5 min. The MS was operated in data-dependent acquisition, with MS1 scans from 350 to 1500 m/z acquired at a resolution of 60 000, maximum injection time (IT) of 10 ms, and an automatic gain control (AGC) target value of 3e6. The 20 most intense precursor ion peaks with charges from +2 to +6 were selected for fragmentation, unless present in the dynamic exclusion list (30 s). Precursor ions were selected with an isolation window of 0.7 m/z, fragmented in an higher-energy C-trap dissociation (HCD) cell with a normalized collision energy of 30% and analyzed in the detector with a resolution of 45 000 m/z, AGC target value of 1e5, and maximum injection time of 86 or 240 ms for total proteome and phosphoproteome analysis, respectively.

(Phospho)-Proteomics Data Analysis

RAW files were analyzed using MaxQuant²¹ v1.6.10.43, where TMTpro was manually included as a fixed modification and quantification method. Correction factors for each TMT channel as provided by the vendor were added to account for channel spillage, and minimum reporter precursor intensity fraction was set to 0.5. The MS scans were searched against human and SARS-CoV-2 Uniprot databases (January of 2020 and April of 2020, respectively). The SARS-CoV-2 database was modified to include the D614G mutation on the Spike protein using the Andromeda search engine. FDR was calculated based on searches on a pseudoreverse database and set to 0.05. The search included as fixed modifications carbamidomethylation of cysteine and as variable modifications methionine oxidation, N-terminal acetylation, and asparagine and glutamine deamidation. Trypsin/P was set as protease for the in silico digestion of the proteome database. Total proteome and immobilized metal affinity chromatography (IMAC)-enriched phosphopeptide samples were analyzed in the same MaxQuant run in separate parameter groups with the same settings, except for the IMAC-enriched samples also Phospho (STY) was added as a variable modification. Protein contaminants, hits in the reverse database, only

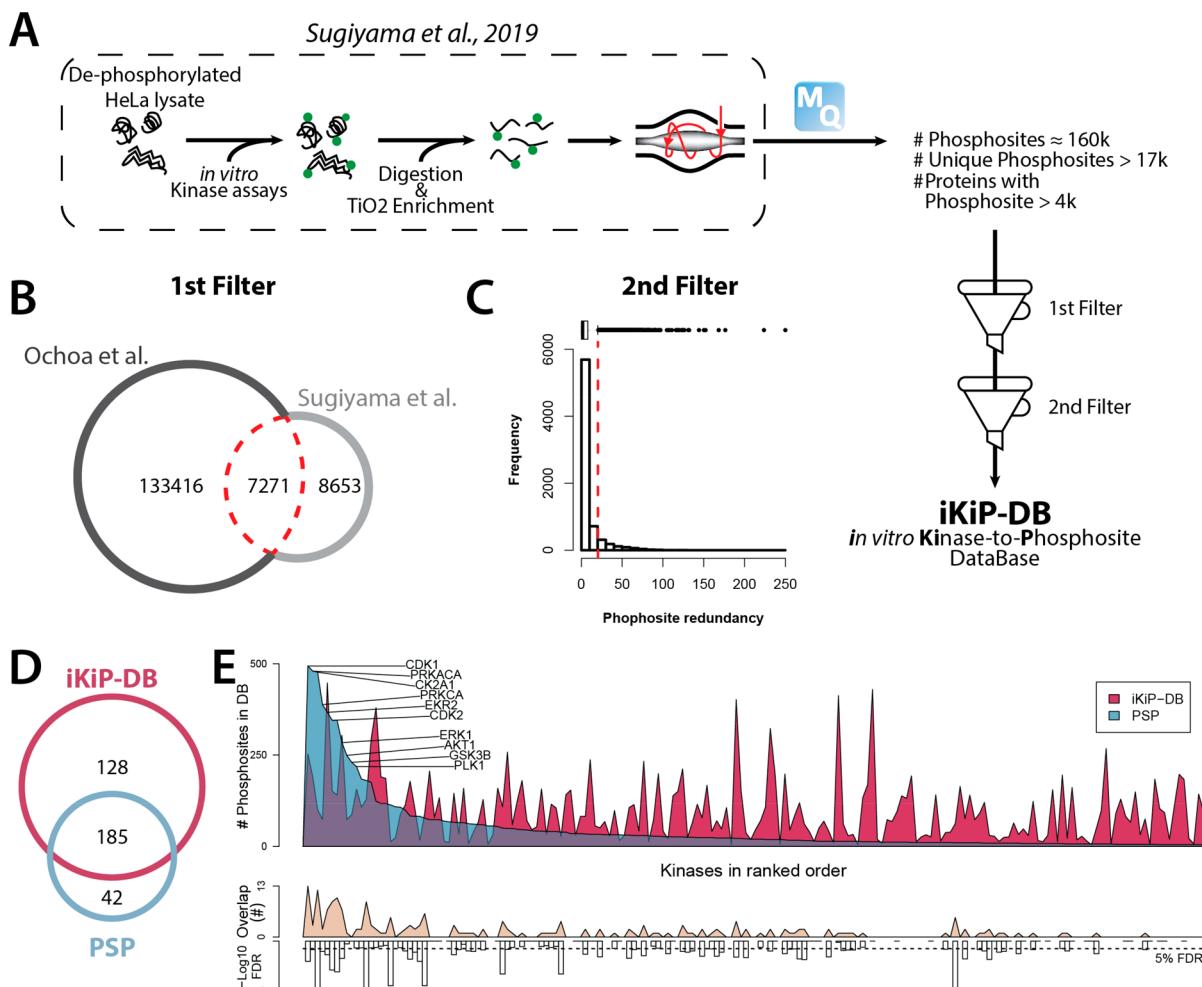


Figure 1. Generating iKiP-DB. (A) Schematic representation of how the kinase assay data were produced, together with the output of our reanalysis and the major steps to derive iKiP-DB. (B) Venn diagram representing the overlap between the sites identified by Sugiyama et al. and the collection of human phosphosites from the work of Ochoa and colleagues. The red line indicates which sites were selected by this first filtering step. (C) Frequency distribution of phosphosites redundancy, as in the number of kinases each phosphosite was assigned to. The red line indicates the cut-off for the second filter we applied to obtain iKiP-DB. (D) Venn diagram showing the overlap of kinases annotated in iKiP-DB and PhosphositePlus (PSP). (E) Kinases present in PSP and iKiP-DB ranked based on the number of sites annotated in PSP. Above, we indicated the number of sites in each database and the names of the kinases with most sites in PSP (higher than 200). Below is depicted the size of the overlap between the two databases and the significance of the overlap (when present) based on a hypergeometric test with BH adjusted *p*-values. The dashed line indicates the 5% FDR cutoff.

identified by modified site and identified by less than two peptides of which less than one was unique, were removed from the ProteinGroups result table. Phosphosites were filtered by hits in the reverse database and potential contaminants. Additionally, only sites with a localization probability higher than 75% were considered for further analysis. TMT reporter ion intensities for each sample were then log2-transformed and median-normalized. Treatment-overcontrol ratios for the 4 and 12 h time points were calculated with the matching control (4 h mock and 12 h mock, respectively), while treatment-overcontrol ratios for the 8 h time point were calculated with the 4 h mock infection control sample. Significantly regulated proteins or phosphosites at 12 h postinfection were calculated with a Student *t*-test with Benjamini-Hochberg (BH) corrected *p*-values. Significant cutoff was set at 10% FDR, while we used a data-driven approach for a fold-change cutoff. Specifically, for each side of the distribution of fold-changes (higher or lower than zero), the function describing the density of points on the *x*-axis was calculated. The cut-off was set based on the *x* value

for which $\frac{d}{dx}f(x)=1$ for $x \geq 0$ and $\frac{d}{dx}f(x)=-1$ for $x < 0$. Phosphoproteomics data for a PTM-SEA analysis were prepared as described above for the benchmarking data sets. All statistical analysis was done with the R programming language (v3.6.6). The gene-centric gene ontology analysis of differentially regulated phosphosites was calculated with Metascape.³⁷

Data Availability

The mass spectrometry proteomics data were deposited to the ProteomeXchange Consortium via the PRIDE³⁸ partner repository with the data set identifier PXD030395.

RESULTS AND DISCUSSION

Deriving iKiP-DB from In Vitro Phosphorylation Assays

To develop our database, we took advantage of a recent large-scale discovery study of substrates of the human kinase.²⁰ In this study, dephosphorylated HeLa cell lysate was incubated with 385 recombinantly expressed human kinases to perform

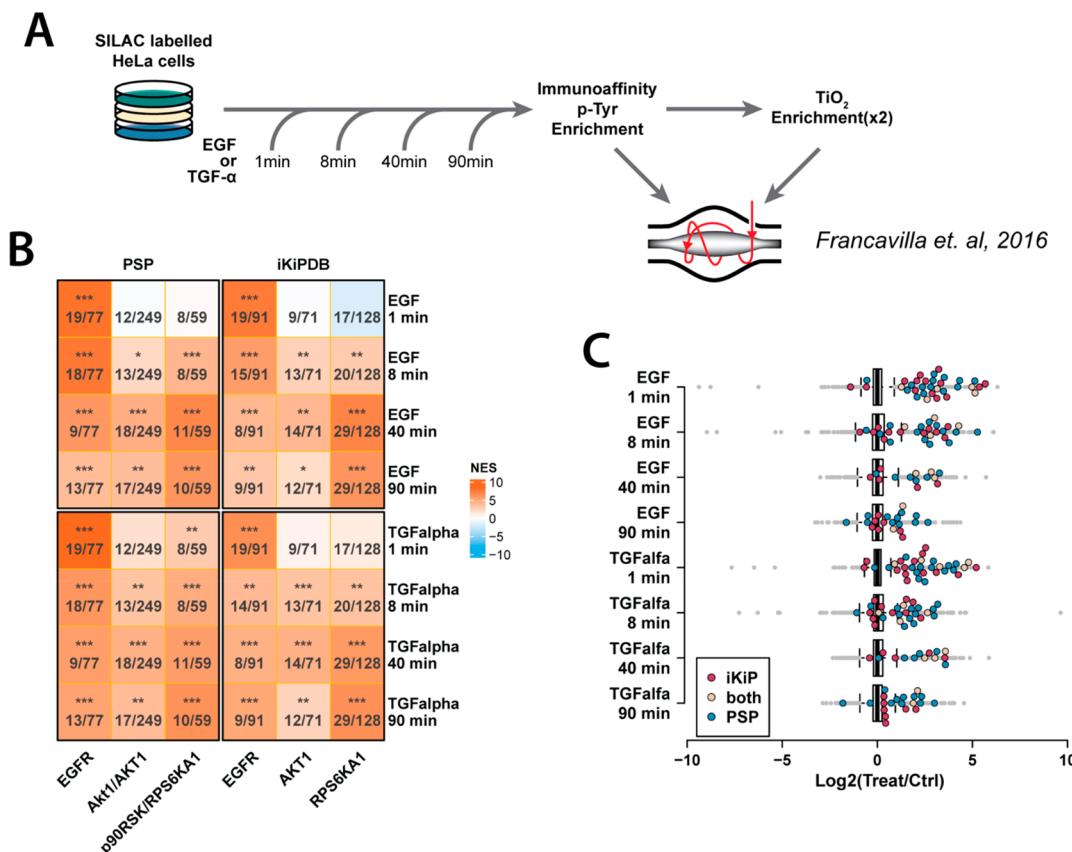


Figure 2. Benchmarking of iKiP-DB. (A) Schematic representation of the experimental design of the data from Francavilla et al. (B) Heatmap depicting the prediction output and colored according to the normalized enrichment score of the PTM-SEA analysis. Each cell reports the significance of the prediction (permutation-based p -values: * $<10\%$, ** $<5\%$, *** $<1\%$) as well as the size of the overlap over the size of the specific kinase set. (C) Distribution of phosphosites ratios for EGF and TGF α treatments. Sites annotated for EGFR in iKiP-DB, PSP, or both are highlighted.

in vitro kinase assays. After proteolytic digestion, the sites phosphorylated by each kinase were identified using a reproducible dimethyl labeling-based quantitative phosphoproteomic workflow (Figure 1A). Intrigued by this study, we retrieved the raw mass spectrometry data (ProteomeXchange data set identifier PXD011366) and reanalyzed it with MaxQuant.²¹ In total, we obtained 159 618 kinase-to-phosphosites associations for 17 225 unique phosphosites, localized on 4032 distinct protein groups (Figure 1A). In vitro kinase assays can induce phosphorylation of sites that are not phosphorylated under physiological conditions (see Conclusions). Therefore, we first filtered the data using a catalogue of 112 manually curated data sets of phospho-enriched proteins from 104 different human cell types or tissues.⁸ More than half of the in vitro sites were not observed in any cell or tissue data set and were therefore excluded (Figure 1B). As a second filter, we excluded all sites assigned to 20 or more kinases, since we reasoned that these highly redundant sites are not well-suited to distinguish between different kinases (Figure 1C). Finally, we removed all kinases with less than five phosphosites, since such a small number would not provide sufficient information to compute robust enrichment scores. After filtering, we obtained an iKiP-DB for 313 kinases (Supporting Information, Table S1). We provide the iKiP-DB in the GMT format as a Supporting Information file to this manuscript, ready to use for PTM-SEA using the ssGSEA suite.^{16,27} In addition to the filtered database, we also provide versions without the first

(overlap with known sites) and the second (kinase redundancy) filter (Supporting Information, Table S1).

iKiP-DB provides kinase-substrate associations for 128 kinases not previously annotated in PSP (Figure 1D). When looking at the 185 kinases shared between both databases, PSP showed a strong bias toward few well-characterized kinases (e.g., CDK1, PRKCA, CK2A1), while most other kinases have only a small number of assigned sites. In contrast, kinase-substrate associations in iKiP-DB do not display such a bias and greatly extend the annotations for many kinases (Figure 1E, upper panel and Supporting Information, Table S1). Nevertheless, although PSP and iKiP-DB were derived in a completely independent fashion, we observed a significant overlap between the number of phosphosites assigned to the same kinases in both databases (Figure 1E, lower panel and Supporting Information Table S1).

iKiP-DB Predicts the Activation of EGFR and Downstream Pathway

To assess if iKiP-DB can predict kinase activity in phosphoproteomic data, we first turned to epidermal growth factor receptor (EGFR) signaling.³⁹ To this end, we took advantage of a recent phosphoproteomic data set.²² In this study, the authors treated HeLa cells with epidermal growth factor (EGF) or transforming growth factor alpha (TGF α) over a time course, enriched for phosphopeptides via combined phospho-tyrosine and titanium-dioxide enrichment, and measured the resulting phosphoproteome by MS (Figure

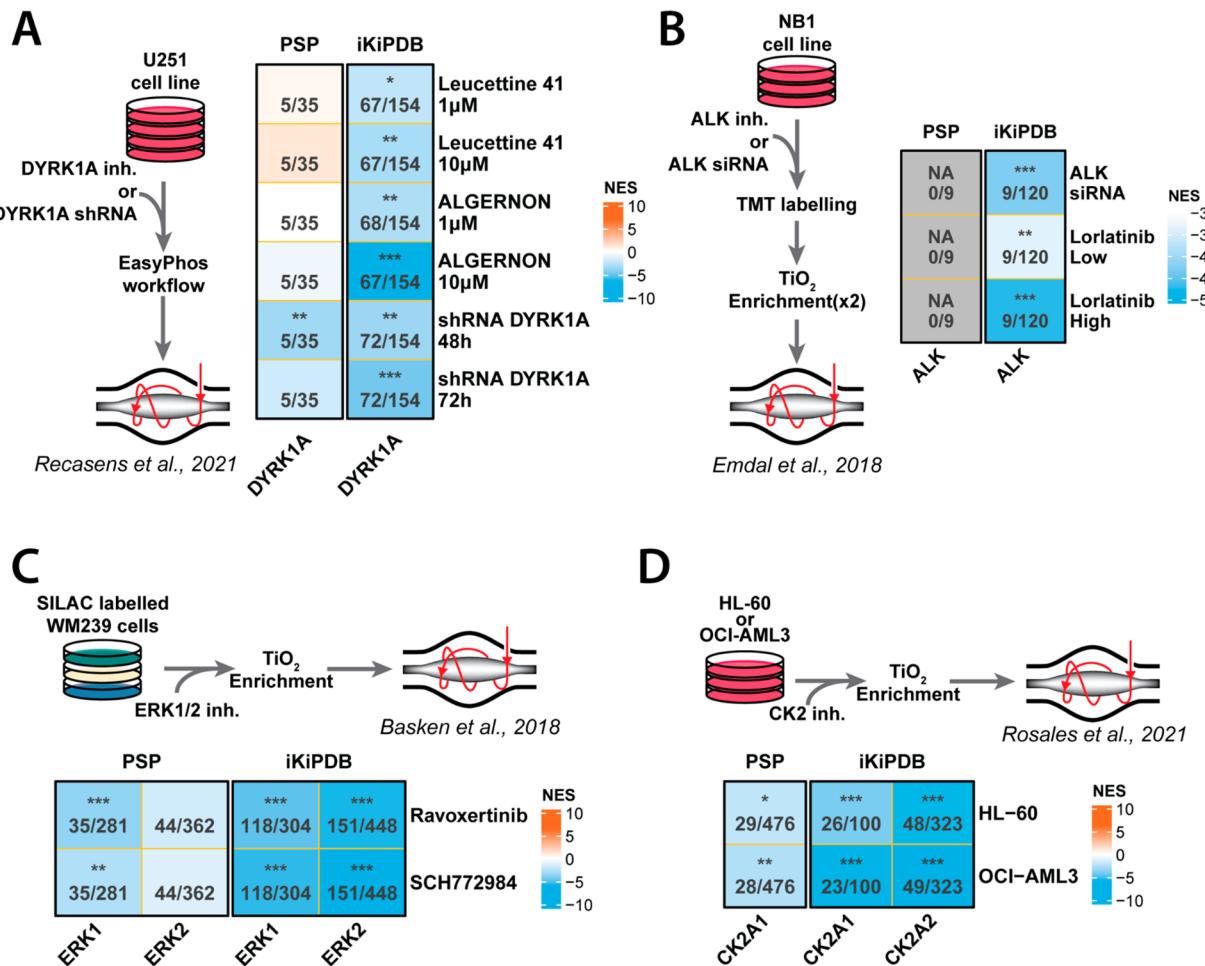


Figure 3. iKiP-DB correctly predicts kinase inhibitions under different experimental conditions. All panels depict the experimental design of the data used on the left or top and a heatmap with the output of the prediction on the right or bottom. Each heatmap is colored according to the NES, and each cell reports the significance (permutation-based *p*-values: * <10%, ** <5%, *** <1%) as well as the size of the overlap over the size of the specific kinase set. We tested our database on the inhibitions of DYRK1A in glioblastoma cells (A), ALK in neuroblastoma cells (B), ERK1/2 in metastatic melanoma cells (C), and CK2 in two acute myeloid leukemia cell lines (D).

2A). This data set provides an excellent benchmark for our database, since (i) the EGFR/PI3K/AKT pathway is well-studied and characterized,³⁹ (ii) the time course provides a longitudinal dimension to evaluate kinase activity annotations over time, and (iii) EGF and TGF α act on the EGFR receptor producing similar but not identical outcomes.^{22,39} To calculate kinase enrichments, we employed the PTM-SEA algorithm,¹⁶ a method recently developed to calculate enrichment of specific PTM sets in MS data. Similarly to a gene set enrichment analysis (GSEA),²⁷ PTM-SEA computes a rank-based statistic for sites present in an annotated PTM set within the overall distribution of all ranked sites (ordered by their abundance ratio or intensity). The normalized enrichment score (NES) reflects the degree to which sites are overrepresented at the top (positive score) or bottom (negative score) of the entire ranked list of sites. For comparison, we computed kinase enrichments with the same algorithm but used a database of kinase sites in the PSP collection of PTMsigDB¹⁶ (Figures 2B and S1A and Supporting Information Table S2). The two databases produced very similar predictions for EGFR and the two downstream kinases AKT1 and RPS6KA1 (also known as p90RSK or MAPKAPK1).^{22,39} In line with previous findings, our analysis reveals that EGFR is rapidly activated upon stimulation with both EGF and TGF α , followed by a decrease

in activity caused by receptor internalization.²² Both analyses indicate a stronger reactivation of EGFR upon TGF α treatment, which has been linked to the more efficient recycling of the receptor to the plasma membrane.²² Of note, both PSP and iKiP-DB predicted the EGFR at time point 1 min as the most activated kinase throughout the experiment, confirming the specificity of the predictions (Supporting Information Table S2). The relevant phosphosites for EGFR from PSP and iKiP-DB showed similar trends across the data set (Figure 2C). Additionally, the delayed and more sustained activation of AKT1 and RPS6KA1 is in line with the more downstream position of these kinases in the EGFR/PI3K/AKT pathway^{40,41} (Figure 2B).

To ensure the robustness of iKiP-DB across different algorithms to calculate kinase enrichments, we also used kinase-substrate enrichment analysis^{17,28} and calculated KSEA scores and statistics, obtaining overall similar results (Figure S1B, Supporting Information Table S2). To assess the consequences of the filters we applied, we repeated the PTM-SEA analysis on iKiP-DB without the first (mapping to Ochoa et al.) and/or the second (remove sites mapping to ≥ 20 kinases) filtering step (Figure S1C). The first filter mainly reduces the total number of sites in the database (from 663 to 309 for EGFR), while the number of overlapping sites is only

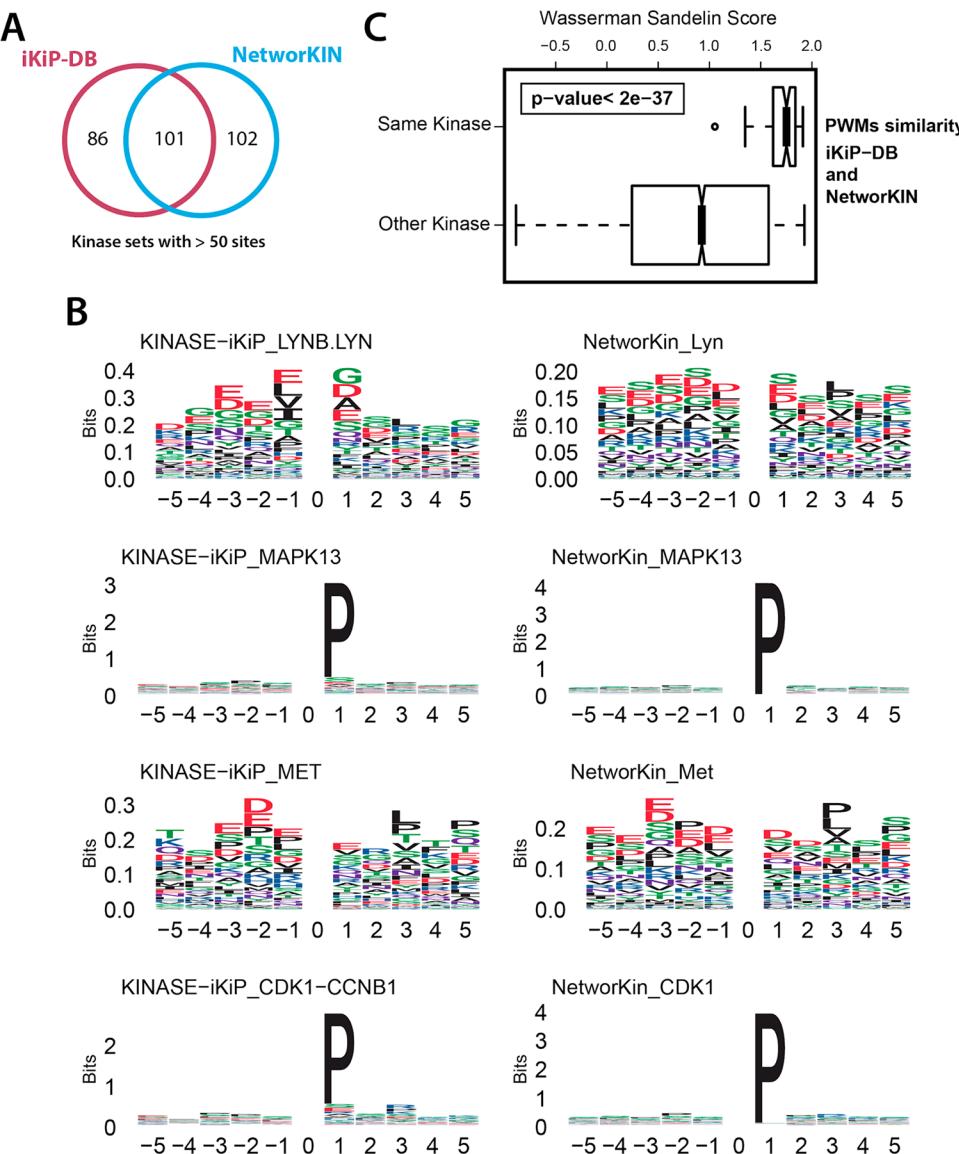


Figure 4. PWMs from iKiP-DB match the ones in NetworkKIN. (A) Overlap between kinase sets present in iKiP-DB and in the kinase site database NetworkKIN. (B) Side-by-side comparison of a selection of sequence logos from iKiP-DB (left) and NetworkKIN (right). (C) Similarity PWMs of the two databases was evaluated with the Wassermann-Sandelin score for all matching (same kinase) or nonmatching (other kinase) between all kinase sets in the two databases. The difference between the two distributions was highly significant ($p < 2e-37$, Mann-Whitney U-test).

mildly reduced (from 40 to 39 for EGFR). Hence, the first filter affects the data in the desired way: it reduces the number of phosphosites observed in the in vitro kinase assays down to sites that are also seen in vivo (that is, in living cells). The redundancy filter (remove sites mapping to ≥ 20 kinases) reduces the number of sites further. Importantly, the filters have an overall minor impact on kinase enrichment scores. Thus, filtering significantly reduces the size of iKiP-DB without negatively affecting its performance. In fact, the fully filtered iKiP-DB better reveals the known activation of RPS6KA1 at later stages of EGF/TGF α stimulation as well as the stronger reactivation of EGFR upon TGF α treatment. For convenience, we also provide iKiP-DB without the first and the second filters (in addition to the filtered version of iKiP-DB recommended for most applications) (Supporting Information Table S2). In summary, these data show that iKiP-DB can be used to predict the activation status of a well-studied kinase as well as a manually curated database.

iKiP-DB Predicts the Inhibition of Well and Poorly Characterized Kinases

To further validate iKiP-DB as a useful tool to predict kinase activity, we searched the literature for phosphoproteomics data sets with the following characteristics: (i) treatments with kinase inhibitors or knockdowns, as they would provide an easy ground truth to test our predictions, (ii) studies focusing on a range of kinases, both well-characterized and less-well characterized, and (iii) data sets derived from samples other than HeLa cells, to test predictions across different model systems. According to these criteria, we selected four additional data sets for further benchmarking of iKiP-DB (Figure 3). Again, we used the PTM-SEA algorithm to compute kinase enrichment scores.¹⁶

For the selected data sets on DYRK1A²³ and ALK²⁴ kinase inhibition, PSP provides poor predictions of activity (Figure 3A, Supporting Information Table S3) or no prediction at all because no observed sites overlapped with the database

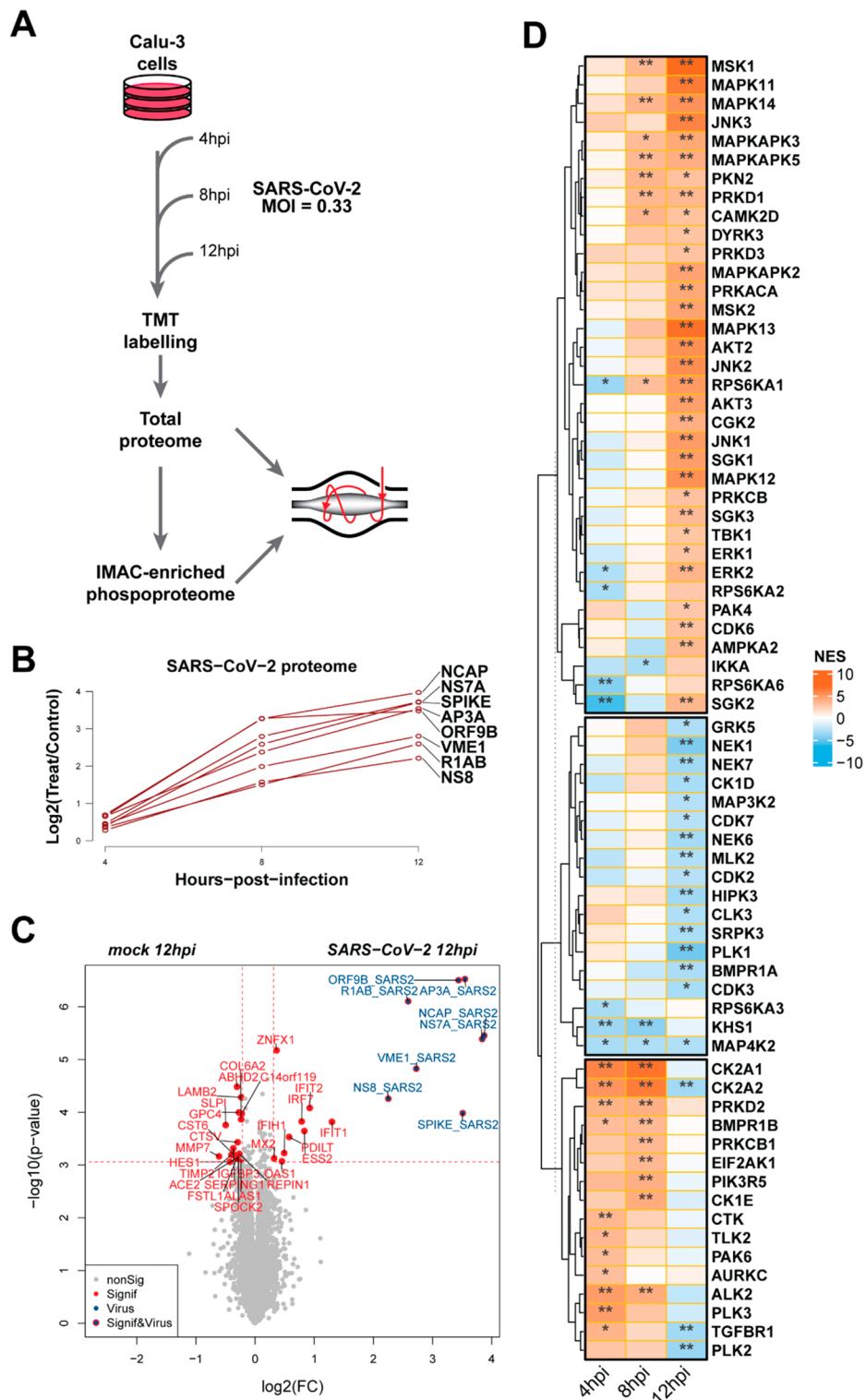


Figure 5. Changes in kinase activities upon infection with SARS-CoV-2. (A) Schematic representation of our infection experiments of lung epithelial cells with SARS-CoV-2. (B) Variation in abundance of all quantified SARS-CoV-2 proteins throughout the course of the infection. (C) Volcano plot depicting the significantly regulated proteins (host and viral) at the latest time point measured. The red lines indicate the cutoff used for significance calling. (D) Variation in kinase activity caused by the infection, as predicted by iKiP-DB. The heatmap is colored based on the NES of the PTM-SEA analysis, and each cell reports the significance cutoff (BH-corrected *p*-values: * < 10%, ** < 5%, *** < 1%). Kinases were divided into three clusters by k-means clustering based on the NES values.

(Figure 3B, Supporting Information Table S3). In contrast, the predictions made with iKiP-DB reflect the treatments the cells were subjected to: U251 glioblastoma cells treated with two DYRK1A inhibitors at two different concentrations show an

inhibition for the correct kinase, with the NES reflecting the increased concentrations of the inhibitors with lower scores. A similar behavior is observed for DYRK1A knockdowns, with lower NES for the longer time point of knockdown, providing

an excellent confirmation of the validity of our kinase activity annotation (Figures 3A and S2A). Additionally, iKiP-DB indicated DYRK1A among, when not the most, inhibited kinase (Figure S2B). Similarly, analysis with iKiP confirms the knockdown of ALK in neuroblastoma NB1 cells as well as a dose-dependent inhibition with lorlatinib (Figures 3B, S2C, and S2D).

The next two data sets we selected focused on the inhibition of ERK1/2 and CK2, which are well-studied and annotated kinases in PSP (Figure 1E). ERK1/2 inhibitors are of particular interest due to their role in the clinic, for example, for the treatment of melanoma, especially after the development of metastases.⁴² In their work, Basken and colleagues tested two potent ERK1/2 inhibitors on WM239 metastatic melanoma cells and measured their effect on the phosphoproteome.²⁵ Kinase activity prediction with iKiP-DB correctly indicated the strong inhibition of both ERK1 and 2, while prediction with PSP could only detect the significant depletion of ERK1 (Figure 3C, Supporting Information Table S3). Next, we wanted to test whether our database could be used to unbiasedly identify the correct target of a kinase inhibitor. For this, we assessed which kinases were predicted as most inhibited by both drug treatments. Analysis with iKiP-DB correctly identified ERK2 as the most inhibited kinase by raxoxatinib and SCH772984 (Figure S2E), which is well in line with the lower half-maximal inhibitory concentration (IC_{50}) values of both compounds for ERK2 compared to ERK1.^{43,44} On the contrary, an analysis with PSP did not indicate ERK as the most inhibited kinase by either treatment (Figure S2E, Supporting Information Table S3). Finally, in the last study we used for benchmarking, Rosales and colleagues treated two acute myeloid leukemia cell lines with a peptide-based kinase inhibitor targeting CK2.²⁶ While both PSP and iKiP-DB correctly predicted the inhibition of CK2A1, our database could also provide information regarding the inhibition of CK2A2, the other catalytic subunit of the kinase (Figure 3D, Supporting Information Table S3). Again, we assessed which kinases were predicted as most inhibited to evaluate the potential of our database for unbiased target discovery. Also here, iKiP-DB could be used to correctly assess the target of the inhibitor in all conditions (Figure S2F, Supporting Information Table S3).

The analyses above are based on five different data sets from four different laboratories. However, this still cannot comprehensively capture known kinase substrate relationships. We therefore also asked whether the sequences of phosphorylation sites in iKiP-DB match to annotated NetworKIN motifs in KinomeXplorer.¹² To this end, we computed position weight matrices for kinases in iKiP-DB with more than 50 sites (Figure S3). For the 101 kinases shared between iKiP-DB and NetworKIN (Figure 4A) we plot the corresponding sequence logos next to each other for direct visual inspection (see Figure 4B for selected examples and Figure S4 for the whole comparison). To more systematically assess the similarity between iKiP-DB and NetworKIN motifs we computed Wassermann-Sandelin scores.²⁹ The distribution of scores for the same kinase was significantly higher than for different kinases, confirming motif similarity in a more global and unbiased method (Figure 4C and Supporting Information Table S4).

In conclusion, a reanalysis of published phosphoproteomic data sets and the comparison of substrate motifs highlights the ability of iKiP-DB to predict the activity of different kinases in

various cell lines and with data generated via different quantification and enrichment strategies (SILAC, TMT, EasyPhos, TiO₂, anti-p-Tyr antibodies). Hence, iKiP-DB can robustly predict kinase activity in diverse phosphoproteomic data sets.

iKiP-DB Identifies Activated Kinases upon SARS-CoV-2 Infection

As proof of concept, we applied iKiP-DB to a new data set of SARS-CoV-2 infected lung epithelial cells. Briefly, we infected lung epithelial cells (Calu-3) with SARS-CoV-2 (Patient isolate, BetaCoV/Munich/BavPat1/2020/EPI_ISL_406862) at a multiplicity of infection of 0.33 and collected samples at 4, 8, and 12 h postinfection (hpi) in triplicates, along with mock infected controls. For quantitation and multiplexing we employed isobaric labeling with tandem mass tags (TMTpro).^{33,34} From these samples we acquired total proteome and IMAC-enriched phosphopeptide data (Figure 5A), which yielded over 8000 proteins and ~13 000 phosphosites with no missing values and high reproducibility (Figure S6A). The impact of the infection on host proteome and phosphoproteome was modest, with few proteins or phosphosites differentially regulated at the latest time point postinfection (Figures 5 and S6B). We measured a strong increase of all SARS-CoV-2 proteins and phosphosites, confirming the presence of a productive infection in Calu-3 cells (Figure 5B, Supporting Information Table S5). At the latest time point, we identified several differentially regulated host proteins (Figure 5C, Supporting Information Table S5). Among these, many interferon-stimulated genes such as *IRF7*, *IFIT1*, *IFIT2*, *IFIH1*, *OAS1*, and *MX2* were enriched upon infection. These proteins are well-characterized for their antiviral function,^{45,46} also in the context of SARS-CoV-2 infection.^{31,47,48} We also identified downregulation of ACE2, the most important entry factor of SARS-CoV-2,⁴⁹ which is in line with the reported shedding of the receptor from infected cells.⁵⁰

Of the ~13 000 phosphosites measured, 169 were upregulated, and 147 were downregulated upon SARS-CoV-2 infection at the latest time point, confirming our evaluation of the overall effect of the infection (Figure S5B). To have an overview of the affected proteins, we performed a gene-centric gene ontology (GO) analysis of the significantly regulated phosphosites, considering separately up- and downregulated sites. Very interestingly, we found two coronavirus-related GO terms (from the WikiPathways collection) upregulated upon infection. The upregulation of coronavirus-related interferon induction is well in line with the upregulation of interferon-stimulated genes we observed at the total protein level. We also observed a downregulation of cell cycle-related processes, again in line with the ongoing infection (Figure S5B).

To gain more insight on how the virus changes the activity of host kinases, we analyzed the phosphoproteomic data using iKiP-DB (Figure 4D, Supporting Information Table S4). This analysis divided kinases into three broad groups: early activated, late inhibited, and late activated, with the last group being the largest (Figure 4D). Of particular interest is the presence of TANK binding kinase 1 (TBK1) among the kinases activated at the latest time point postinfection. This is well in accordance with the role of TBK1 as a downstream effector of RIG-I-like receptor signaling⁵¹ and also with the induction of interferon-stimulated genes we observed at 18 hpi. Additionally, many of the late-activated kinases are known to

be involved in the response to viral infections, such as MAP kinases⁵² and stress-activated kinases downstream of the MAP/ERK pathway like RPS6KA1.⁵³ Upregulation of MAPK signaling is also in line with the gene-centric GO analysis of phosphosites upregulated upon infection (Figure S5B). Among the kinases predicted to be downregulated, several are involved in cell cycle progression, including Polo-like kinases (PLKs) and NimA related kinases (NEKs),^{54,55} in line with the decreased proliferation induced by the infection. Overall, our analysis is consistent with current knowledge of the effects of SARS-CoV-2 infection at the cellular and phosphoproteome levels,^{31,56} reflecting the cellular response to infection. Similarly, the analysis with the PSP kinase set showed enriched activation of some stress-activated kinases of the MAP/ERK pathway and downregulation of CDKs but, overall, could provide prediction only for a small number of kinases (Figure S5C, Supporting Information Table S5).

CONCLUSIONS

Protein phosphorylation is a fundamental and ubiquitous process in biological systems,⁵⁷ and as such studying the phosphoproteome is highly informative for basic biology and disease.^{22,58} While novel technologies are rapidly increasing the depth and coverage of MS-derived phosphoproteomics data, their interpretation remains challenging, partially due to the scarcity of functional annotations of phosphosites.⁷

In this work, we described the reanalysis of published in vitro kinase assay data, which we compiled into a database of kinase-substrate associations we named iKiP-DB. Compared to PSP as the current gold standard for phosphosite annotation,⁶ our database increases the range of annotated kinases as well as the number of substrate sites annotated. Additionally, since our database is derived from experimental data, it is not biased toward the most well-studied portion of the kinome. We employed iKiP-DB in conjunction with PTM-SEA¹⁶ to predict the activation or inhibition of kinases in published phosphoproteomics data. A direct comparison demonstrated an equal or higher degree of accuracy of iKiP-DB compared to the manually annotated data of PSP. Interestingly, we also observed a superior prediction by iKiP-DB for CK2 and ERK1/2 inhibitor experiments, even though both CK2 and ERK1/2 are well-annotated kinases in PSP. Finally, analysis of newly generated data of SARS-CoV-2 infected lung epithelial cells recapitulates the phenotypic effects of viral infections as well as the known biology of the novel coronavirus.

One important limitation of iKiP-DB is that it is based on in vitro kinase experiments. While the catalytic activity of each recombinant kinase was individually tested in the original paper (Table S5),²⁰ the in vitro activity of a kinase does not necessarily reflect its in vivo activity. Indeed, many sites that are phosphorylated in vitro do not seem to be phosphorylated in vivo. We alleviate this problem by restricting the database to sites that have been observed in cells (Figure 1B). Nevertheless, individual phosphosites in iKiP-DB are not necessarily phosphorylated by the corresponding kinase under physiological conditions. Therefore, iKiP-DB is more useful for the large-scale analysis of kinase signatures, rather than at the level of individual phosphosites. It is actually surprising how well the in vitro kinase data alone (without any manual annotation) reflects cellular kinase activity (Figures 2 and 3). This point also highlights the main difference between PSP and iKiP-DB: PSP is a manually curated database with information about the topology, biological function, and regulatory significance of

modification sites that assembles different types of information from many sources. Hence, PSP is a “one-stop shop” for functional information about individual phosphorylation sites. In contrast, iKiP-DB is based on a single source, is not manually curated, and was assembled with the only purpose to predict kinase activity in large-scale (thousands of sites) phosphoproteomic data sets. Thus, PSP and iKiP-DB are complementary and serve different purposes.

In summary, we demonstrate that integrating phosphoproteomic data sets with in vitro kinase data via iKiP-DB greatly facilitates detection of altered kinase activity. We believe this tool will be broadly useful for phosphoproteomic data analysis. In the future, performing in vitro kinase assays using protein lysates from other cell lines/tissues could broaden the range of phosphosites covered, which could further improve iKiP-DB predictions. We provide iKiP-DB in GMT format as a supplementary file to this manuscript, ready to use for PTM-SEA using the ssGSEA suite.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.2c00198>.

Extended analyses of Francavilla et al., related to Figure 2. Evaluation of specificity of iKiP-DB, related to Figure 3. Sequence logos of kinase sets with more than 50 sites in iKiP-DB. Sequence logo comparison of iKiP-DB and NetworKIN. Extended analyses of SARS-CoV-2-infected Calu-3 cells, related to Figure 5 (PDF)

Filtering steps from Sugiyama et al. to iKiP-DB and iKiP-DB in table format (XLSX)

PTM-SEA and KSEA output of the Francavilla et al. data set, Figure 2 (XLSX)

PTM-SEA outputs for data sets analyzed in Figure 3 (XLSX)

Wassermann-Sandelin scores for the comparison between PWMs of iKiP-DB and NetworKIN (XLSX)

Proteomics, phosphoproteomics, and PTM-SEA outputs of the analysis of SARS-CoV-2-infected Calu-3 cells (XLSX)

Compressed iKiP-DB.gmt database (ZIP)

AUTHOR INFORMATION

Corresponding Author

Matthias Selbach — Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, 13092 Berlin, Germany; Charité—Universitätsmedizin, 10117 Berlin, Germany; orcid.org/0000-0003-2454-8751; Phone: +49 30 9406 3574; Email: matthias.selbach@mdc-berlin.de

Authors

Tommaso Mari — Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, 13092 Berlin, Germany; orcid.org/0000-0001-6011-9386

Kirstin Mösbauer — Institute of Virology, Charité—Universitätsmedizin, 10117 Berlin, Germany
Emanuel Wyler — Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, 13092 Berlin, Germany

Markus Landthaler — Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, 13092 Berlin, Germany

**Christian Drosten — Institute of Virology,
Charité—Universitätsmedizin, 10117 Berlin, Germany**

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jproteome.2c00198>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank E. Ramberger for the critical reading of the manuscript as well as M. Haji of the mass spectrometry core facility of the MDC for his invaluable help in sample preparation. Since this work would have not been possible without all the published data we used for the database and the benchmarking, we express our gratitude to all the researchers involved in these studies that generated high-quality phosphoproteomics data we could use. We especially mention N. Sugiyama, Y. Ishihama, and H. Imamura, since the core of this work was based on their dataset. This work was supported by the German Ministry of Education and Research (BMBF) via the national research node for mass spectrometry in systems medicine MSTARS (031L0220B to M.S.).

REFERENCES

- (1) Doll, S.; Burlingame, A. L. Mass Spectrometry-Based Detection and Assignment of Protein Posttranslational Modifications. *ACS Chem. Biol.* **2015**, *10* (1), 63–71.
- (2) Olsen, J. V.; Mann, M. Status of Large-Scale Analysis of Post-Translational Modifications by Mass Spectrometry. *Mol. Cell. Proteomics* **2013**, *12* (12), 3444–3452.
- (3) Ramberger, E.; Suarez-Artiles, L.; Perez-Hernandez, D.; Haji, M.; Popp, O.; Reimer, U.; Leutz, A.; Dittmar, G.; Mertins, P. A Universal Peptide Matrix Interactomics Approach to Disclose Motif-Dependent Protein Binding. *Mol. Cell. Proteomics* **2021**, *20*, 100135.
- (4) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schlüter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlén, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschläger, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B. How Many Human Proteoforms Are There? *Nat. Chem. Biol.* **2018**, *14* (3), 206–214.
- (5) Roux, P. P.; Thibault, P. The Coming of Age of Phosphoproteomics—from Large Data Sets to Inference of Protein Functions. *Mol. Cell. Proteomics* **2013**, *12* (12), 3453–3464.
- (6) Hornbeck, P. V.; Zhang, B.; Murray, B.; Kornhauser, J. M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and Recalibrations. *Nucleic Acids Res.* **2015**, *43* (D1), D512–D520.
- (7) Needham, E. J.; Parker, B. L.; Burykin, T.; James, D. E.; Humphrey, S. J. Illuminating the Dark Phosphoproteome. *Sci. Signal.* **2019**, *12*, 565.
- (8) Ochoa, D.; Jarnuczak, A. F.; Viéitez, C.; Gehre, M.; Soucheray, M.; Mateus, A.; Kleefeldt, A. A.; Hill, A.; Garcia-Alonso, L.; Stein, F.; Krogan, N. J.; Savitski, M. M.; Swaney, D. L.; Vizcaíno, J. A.; Noh, K.-M.; Beltrao, P. The Functional Landscape of the Human Phosphoproteome. *Nat. Biotechnol.* **2020**, *38* (3), 365–373.
- (9) Cao, L.; Huang, C.; Cui Zhou, D.; Hu, Y.; Lih, T. M.; Savage, S. R.; Krug, K.; Clark, D. J.; Schnaubelt, M.; Chen, L.; da Veiga Leprevost, F.; Eguez, R. V.; Yang, W.; Pan, J.; Wen, B.; Dou, Y.; Jiang, W.; Liao, Y.; Shi, Z.; Terekhanova, N. V.; Cao, S.; Lu, R. J.-H.; Li, Y.; Liu, R.; Zhu, H.; Ronning, P.; Wu, Y.; Wyczalkowski, M. A.; Easwaran, H.; Danilova, L.; Mer, A. S.; Yoo, S.; Wang, J. M.; Liu, W.; Haibe-Kains, B.; Thiagarajan, M.; Jewell, S. D.; Hostetter, G.; Newton, C. J.; Li, Q. K.; Roehrl, M. H.; Fenyö, D.; Wang, P.; Nesvizhskii, A. I.; Mani, D. R.; Omenn, G. S.; Boja, E. S.; Mesri, M.; Robles, A. I.; Rodriguez, H.; Bathe, O. F.; Chan, D. W.; Hruban, R. H.; Ding, L.; Zhang, B.; Zhang, H.; et al. Proteogenomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cell* **2021**, *184* (19), 5031–5052.
- (10) Hogreve, A.; von Stechow, L.; Bekker-Jensen, D. B.; Weinert, B. T.; Kelstrup, C. D.; Olsen, J. V. Benchmarking Common Quantification Strategies for Large-Scale Phosphoproteomics. *Nat. Commun.* **2018**, *9* (1), 1045.
- (11) Patrick, R.; Horin, C.; Kobe, B.; Cao, K.-A. L.; Bodén, M. Prediction of Kinase-Specific Phosphorylation Sites through an Integrative Model of Protein Context and Sequence. *Biochim. Biophys. Acta* **2016**, *1864* (11), 1599–1608.
- (12) Horn, H.; Schoof, E. M.; Kim, J.; Robin, X.; Miller, M. L.; Diella, F.; Palma, A.; Cesareni, G.; Jensen, L. J.; Linding, R. KinomeXplorer: An Integrated Platform for Kinome Biology Studies. *Nat. Methods* **2014**, *11* (6), 603–604.
- (13) Wang, C.; Xu, H.; Lin, S.; Deng, W.; Zhou, J.; Zhang, Y.; Shi, Y.; Peng, D.; Xue, Y. GPS 5.0: An Update on the Prediction of Kinase-Specific Phosphorylation Sites in Proteins. *Genomics Proteomics Bioinformatics* **2020**, *18* (1), 72–80.
- (14) Nováček, V.; McGauran, G.; Matallanas, D.; Vallejo Blanco, A.; Conca, P.; Muñoz, E.; Costabello, L.; Kanakaraj, K.; Nawaz, Z.; Walsh, B.; Mohamed, S. K.; Vandebussche, P.-Y.; Ryan, C. J.; Kolch, W.; Fey, D. Accurate Prediction of Kinase-Substrate Networks Using Knowledge Graphs. *PLoS Comput. Biol.* **2020**, *16* (12), e1007578.
- (15) Xue, B.; Jordan, B.; Rizvi, S.; Naegle, K. M. KinPred: A Unified and Sustainable Approach for Harnessing Proteome-Level Human Kinase-Substrate Predictions. *PLoS Comput. Biol.* **2021**, *17* (2), e1008681.
- (16) Krug, K.; Mertins, P.; Zhang, B.; Hornbeck, P.; Raju, R.; Ahmad, R.; Szucs, M.; Mundt, F.; Forestier, D.; Jane-Valbuena, J.; Keshishian, H.; Gillette, M. A.; Tamayo, P.; Mesirov, J. P.; Jaffe, J. D.; Carr, S.; Mani, D. R. A Curated Resource for Phosphosite-Specific Signature Analysis. *Mol. Cell. Proteomics* **2019**, *18* (3), 576–593.
- (17) Wiredja, D. D.; Koyutürk, M.; Chance, M. R. The KSEA App: A Web-Based Tool for Kinase Activity Inference from Quantitative Phosphoproteomics. *Bioinformatics* **2017**, *33* (21), 3489–3491.
- (18) Mischnik, M.; Sacco, F.; Cox, J.; Schneider, H.-C.; Schäfer, M.; Hendlich, M.; Crowther, D.; Mann, M.; Klabunde, T. IKAP: A Heuristic Framework for Inference of Kinase Activities from Phosphoproteomics Data. *Bioinformatics* **2016**, *32* (3), 424–431.
- (19) Edwards, A. M.; Isserlin, R.; Bader, G. D.; Frye, S. V.; Willson, T. M.; Yu, F. H. Too Many Roads Not Taken. *Nature* **2011**, *470* (7333), 163–165.
- (20) Sugiyama, N.; Imamura, H.; Ishihama, Y. Large-Scale Discovery of Substrates of the Human Kinome. *Sci. Rep.* **2019**, *9* (1), 10503.
- (21) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics. *Nat. Protoc.* **2016**, *11* (12), 2301–2319.
- (22) Francavilla, C.; Papetti, M.; Rigbolt, K. T. G.; Pedersen, A.-K.; Sigurdsson, J. O.; Cazzamali, G.; Karemire, G.; Blagoev, B.; Olsen, J. V. Multilayered Proteomics Reveals Molecular Switches Dictating Ligand-Dependent EGFR Trafficking. *Nat. Struct. Mol. Biol.* **2016**, *23* (6), 608–618.
- (23) Recasens, A.; Humphrey, S. J.; Ellis, M.; Hoque, M.; Abbassi, R. H.; Chen, B.; Longworth, M.; Needham, E. J.; James, D. E.; Johns, T. G.; Day, B. W.; Kassiou, M.; Yang, P.; Munoz, L. Global Phosphoproteomics Reveals DYRK1A Regulates CDK1 Activity in Glioblastoma Cells. *Cell Death Discov* **2021**, *7* (1), 81.
- (24) Emdal, K. B.; Pedersen, A.-K.; Bekker-Jensen, D. B.; Lundby, A.; Claeys, S.; De Preter, K.; Speleman, F.; Francavilla, C.; Olsen, J. V. Integrated Proximal Proteomics Reveals IRS2 as a Determinant of Cell Survival in ALK-Driven Neuroblastoma. *Sci. Signal.* **2018**, *11* (557), eaap9752.

- (25) Basken, J.; Stuart, S. A.; Kavran, A. J.; Lee, T.; Ebmeier, C. C.; Old, W. M.; Ahn, N. G. Specificity of Phosphorylation Responses to Mitogen Activated Protein (MAP) Kinase Pathway Inhibitors in Melanoma Cells. *Mol. Cell. Proteomics* **2018**, *17* (4), 550–564.
- (26) Rosales, M.; Pérez, G. V.; Ramón, A. C.; Cruz, Y.; Rodríguez-Ulloa, A.; Besada, V.; Ramos, Y.; Vázquez-Bloomquist, D.; Caballero, E.; Aguilar, D.; González, L. J.; Zettl, K.; Wiśniewski, J. R.; Yang, K.; Perera, Y.; Perea, S. E. Targeting of Protein Kinase CK2 in Acute Myeloid Leukemia Cells Using the Clinical-Grade Synthetic-Peptide CIGB-300. *Biomedicines* **2021**, *9* (7), 766.
- (27) Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (43), 15545–15550.
- (28) Casado, P.; Rodriguez-Prados, J.-C.; Cosulich, S. C.; Guichard, S.; Vanhaesebroeck, B.; Joel, S.; Cutillas, P. R. Kinase-Substrate Enrichment Analysis Provides Insights into the Heterogeneity of Signaling Pathway Activation in Leukemia Cells. *Sci. Signal.* **2013**, *6* (268), rs6.
- (29) Sandelin, A.; Wasserman, W. W. Constrained Binding Site Diversity within Families of Transcription Factors Enhances Pattern Discovery Bioinformatics. *J. Mol. Biol.* **2004**, *338* (2), 207–215.
- (30) Wagih, O. Ggseqlogo: A Versatile R Package for Drawing Sequence Logos. *Bioinformatics* **2017**, *33* (22), 3645–3647.
- (31) Wyler, E.; Mösbauer, K.; Franke, V.; Diag, A.; Gottula, L. T.; Arsié, R.; Klironomos, F.; Koppstein, D.; Hönzke, K.; Ayoub, S.; Buccitelli, C.; Hoffmann, K.; Richter, A.; Legnini, I.; Ivanov, A.; Mari, T.; Del Giudice, S.; Papies, J.; Praktikno, S.; Meyer, T. F.; Müller, M. A.; Niemeyer, D.; Hocke, A.; Selbach, M.; Akalin, A.; Rajewsky, N.; Drosten, C.; Landthaler, M. Transcriptomic Profiling of SARS-CoV-2 Infected Human Cell Lines Identifies HSP90 as Target for COVID-19 Therapy. *iScience* **2021**, *24* (3), 102151.
- (32) Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsfeld, J. Single-Pot, Solid-Phase-Enhanced Sample Preparation for Proteomics Experiments. *Nat. Protoc.* **2019**, *14* (1), 68–85.
- (33) Mertins, P.; Tang, L. C.; Krug, K.; Clark, D. J.; Gritsenko, M. A.; Chen, L.; Clouser, K. R.; Clauss, T. R.; Shah, P.; Gillette, M. A.; Petyuk, V. A.; Thomas, S. N.; Mani, D. R.; Mundt, F.; Moore, R. J.; Hu, Y.; Zhao, R.; Schnaubelt, M.; Keshishian, H.; Monroe, M. E.; Zhang, Z.; Udeshi, N. D.; Mani, D.; Davies, S. R.; Townsend, R. R.; Chan, D. W.; Smith, R. D.; Zhang, H.; Liu, T.; Carr, S. A. Reproducible Workflow for Multiplexed Deep-Scale Proteome and Phosphoproteome Analysis of Tumor Tissues by Liquid Chromatography-Mass Spectrometry. *Nat. Protoc.* **2018**, *13* (7), 1632–1661.
- (34) Li, J.; Van Vranken, J. G.; Pontano Vaites, L.; Schewpke, D. K.; Huttlin, E. L.; Etienne, C.; Nandhikonda, P.; Viner, R.; Robitaille, A. M.; Thompson, A. H.; Kuhn, K.; Pike, I.; Bomgard, R. D.; Rogers, J. C.; Gygi, S. P.; Paulo, J. A. TMTpro Reagents: A Set of Isobaric Labeling Mass Tags Enables Simultaneous Proteome-Wide Measurements across 16 Samples. *Nat. Methods* **2020**, *17* (4), 399–404.
- (35) Battth, T. S.; Francavilla, C.; Olsen, J. V. Off-Line High-pH Reversed-Phase Fractionation for in-Depth Phosphoproteomics. *J. Proteome Res.* **2014**, *13* (12), 6176–6186.
- (36) Wendisch, D.; Dietrich, O.; Mari, T.; von Stillfried, S.; Ibarra, I. L.; Mittermaier, M.; Mache, C.; Chua, R. L.; Knoll, R.; Timm, S.; Brumhard, S.; Krammer, T.; Zauber, H.; Hiller, A. L.; Pascual-Reguant, A.; Mothes, R.; Bülow, R. D.; Schulze, J.; Leipold, A. M.; Djudjaj, S.; Erhard, F.; Geffers, R.; Pott, F.; Kazmierski, J.; Radke, J.; Pergantis, P.; Baßler, K.; Conrad, C.; Aschenbrenner, A. C.; Sawitzki, B.; Landthaler, M.; Wyler, E.; Horst, D.; Hippensiel, S.; Hocke, A.; Heppner, F. L.; Uhrig, A.; Garcia, C.; Machleidt, F.; Herold, S.; Elezkurtaj, S.; Thibeault, C.; Witzenrath, M.; Cochaint, C.; Suttorp, N.; Drosten, C.; Goffinet, C.; Kurth, F.; Schultz, J. L.; Radbruch, H.; Ochs, M.; Eils, R.; Müller-Redetzky, H.; Hauser, A. E.; Luecken, M. D.; Theis, F. J.; Conrad, C.; Wolff, T.; Boor, P.; Selbach, M.; Saliba, A.-E.; Sander, L. E. SARS-CoV-2 Infection Triggers Profibrotic Macrophage Responses and Lung Fibrosis. *Cell* **2021**, *184*, 6243–6261.
- (37) Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K. Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat. Commun.* **2019**, *10* (1), 1523.
- (38) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.; Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47* (D1), D442–D450.
- (39) Wee, P.; Wang, Z. Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. *Cancers* **2017**, *9* (5), 52.
- (40) Shimamura, A.; Ballif, B. A.; Richards, S. A.; Blenis, J. Rsk1Mediates a MEK-MAP Kinase Cell Survival Signal. *Curr. Biol.* **2000**, *10* (3), 127–135.
- (41) Richards, S. A.; Dreisbach, V. C.; Murphy, L. O.; Blenis, J. Characterization of Regulatory Events Associated with Membrane Targeting of p90 Ribosomal S6 Kinase 1. *Mol. Cell. Biol.* **2001**, *21* (21), 7470–7480.
- (42) Savoia, P.; Fava, P.; Casoni, F.; Cremona, O. Targeting the ERK Signaling Pathway in Melanoma. *Int. J. Mol. Sci.* **2019**, *20* (6), 1483.
- (43) Blake, J. F.; Burkard, M.; Chan, J.; Chen, H.; Chou, K.-J.; Diaz, D.; Dudley, D. A.; Gaudino, J. J.; Gould, S. E.; Grina, J.; Hunsaker, T.; Liu, L.; Martinson, M.; Moreno, D.; Mueller, L.; Orr, C.; Pacheco, P.; Qin, A.; Rasor, K.; Ren, L.; Robarge, K.; Shahidi-Latham, S.; Stults, J.; Sullivan, F.; Wang, W.; Yin, J.; Zhou, A.; Belvin, M.; Merchant, M.; Moffat, J.; Schwarz, J. B. Discovery of (S)-1-(1-(4-Chloro-3-Fluorophenyl)-2-Hydroxyethyl)-4-(2-((1-Methyl-1H-Pyrazol-5-Yl)-amino)pyrimidin-4-Yl)pyridin-2(1H)-One (GDC-0994), an Extracellular Signal-Regulated Kinase 1/2 (ERK1/2) Inhibitor in Early Clinical Development. *J. Med. Chem.* **2016**, *59* (12), 5650–5660.
- (44) Morris, E. J.; Jha, S.; Restaino, C. R.; Dayananth, P.; Zhu, H.; Cooper, A.; Carr, D.; Deng, Y.; Jin, W.; Black, S.; Long, B.; Liu, J.; Dinunzio, E.; Windsor, W.; Zhang, R.; Zhao, S.; Angagaw, M. H.; Pinheiro, E. M.; Desai, J.; Xiao, L.; Shipps, G.; Hruza, A.; Wang, J.; Kelly, J.; Paliwal, S.; Gao, X.; Babu, B. S.; Zhu, L.; Daublain, P.; Zhang, L.; Lutterbach, B. A.; Pelletier, M. R.; Philipp, U.; Siliphaiwanh, P.; Witter, D.; Kirschmeier, P.; Bishop, W. R.; Hicklin, D.; Gilliland, D. G.; Jayaraman, L.; Zawel, L.; Fawell, S.; Samatar, A. A. Discovery of a Novel ERK Inhibitor with Activity in Models of Acquired Resistance to BRAF and MEK Inhibitors. *Cancer Discovery* **2013**, *3* (7), 742–750.
- (45) Schoggins, J. W.; Rice, C. M. Interferon-Stimulated Genes and Their Antiviral Effector Functions. *Curr. Opin. Virol.* **2011**, *1* (6), 519–525.
- (46) Crosse, K. M.; Monson, E. A.; Beard, M. R.; Helbig, K. J. Interferon-Stimulated Genes as Enhancers of Antiviral Innate Immune Signaling. *J. Innate Immun.* **2018**, *10* (2), 85–93.
- (47) Blanco-Melo, D.; Nilsson-Payant, B. E.; Liu, W.-C.; Uhl, S.; Hoagland, D.; Möller, R.; Jordan, T. X.; Oishi, K.; Panis, M.; Sachs, D.; Wang, T. T.; Schwartz, R. E.; Lim, J. K.; Albrecht, R. A.; tenOever, B. R. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **2020**, *181* (5), 1036–1045.
- (48) Sa Ribero, M.; Jouvenet, N.; Dreux, M.; Nisole, S. Interplay between SARS-CoV-2 and the Type I Interferon Response. *PLoS Pathog.* **2020**, *16* (7), e1008737.
- (49) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; Müller, M. A.; Drosten, C.; Pöhlmann, S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *181* (2), 271–280.
- (50) Patel, S. K.; Juno, J. A.; Lee, W. S.; Wragg, K. M.; Hogarth, P. M.; Kent, S. J.; Burrell, L. M. Plasma ACE2 Activity Is Persistently Elevated Following SARS-CoV-2 Infection: Implications for COVID-

19 Pathogenesis and Consequences. *Eur. Respir. J.* **2021**, *57* (5), 2003730.

(51) Onomoto, K.; Onoguchi, K.; Yoneyama, M. Regulation of RIG-I-like Receptor-Mediated Signaling: Interaction between Host and Viral Factors. *Cell. Mol. Immunol.* **2021**, *18* (3), 539–555.

(52) Kumar, R.; Khandelwal, N.; Thachamvally, R.; Tripathi, B. N.; Barua, S.; Kashyap, S. K.; Maherchandani, S.; Kumar, N. Role of MAPK/MNK1 Signaling in Virus Replication. *Virus Res.* **2018**, *253*, 48–61.

(53) Dalby, K. N.; Morrice, N.; Caudwell, F. B.; Avruch, J.; Cohen, P. Identification of Regulatory Phosphorylation Sites in Mitogen-Activated Protein Kinase (MAPK)-Activated Protein Kinase-1a/p90rsk That Are Inducible by MAPK. *J. Biol. Chem.* **1998**, *273* (3), 1496–1505.

(54) Lee, S.-Y.; Jang, C.; Lee, K.-A. Polo-like Kinases (plks), a Key Regulator of Cell Cycle and New Potential Target for Cancer Therapy. *Dev Reprod* **2014**, *18* (1), 65–71.

(55) Fry, A. M.; O'Regan, L.; Sabir, S. R.; Bayliss, R. Cell Cycle Regulation by the NEK Family of Protein Kinases. *J. Cell Sci.* **2012**, *125*, 4423–4433.

(56) Bouhaddou, M.; Memon, D.; Meyer, B.; White, K. M.; Rezelj, V. V.; Correa Marrero, M.; Polacco, B. J.; Melnyk, J. E.; Ulferts, S.; Kaake, R. M.; Batra, J.; Richards, A. L.; Stevenson, E.; Gordon, D. E.; Rojc, A.; Obernier, K.; Fabius, J. M.; Soucheray, M.; Miorin, L.; Moreno, E.; Koh, C.; Tran, Q. D.; Hardy, A.; Robinot, R.; Vallet, T.; Nilsson-Payant, B. E.; Hernandez-Armenta, C.; Dunham, A.; Weigang, S.; Knerr, J.; Modak, M.; Quintero, D.; Zhou, Y.; Dugourd, A.; Valdeolivas, A.; Patil, T.; Li, Q.; Hüttenhain, R.; Cakir, M.; Muralidharan, M.; Kim, M.; Jang, G.; Tutuncuoglu, B.; Hiatt, J.; Guo, J. Z.; Xu, J.; Bouhaddou, S.; Mathy, C. J. P.; Gaulton, A.; Manners, E. J.; Félix, E.; Shi, Y.; Goff, M.; Lim, J. K.; McBride, T.; O'Neal, M. C.; Cai, Y.; Chang, J. C. J.; Broadhurst, D. J.; Klippsten, S.; De Wit, E.; Leach, A. R.; Kortemme, T.; Shoichet, B.; Ott, M.; Saez-Rodriguez, J.; tenOever, B. R.; Mullins, R. D.; Fischer, E. R.; Kochs, G.; Grosse, R.; García-Sastre, A.; Vignuzzi, M.; Johnson, J. R.; Shokat, K. M.; Swaney, D. L.; Beltrao, P.; Krogan, N. J. The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell* **2020**, *182* (3), 685–712.

(57) Aebersold, R.; Mann, M. Mass-Spectrometric Exploration of Proteome Structure and Function. *Nature* **2016**, *537* (7620), 347–355.

(58) Mertins, P.; Mani, D. R.; Ruggles, K. V.; Gillette, M. A.; Claußer, K. R.; Wang, P.; Wang, X.; Qiao, J. W.; Cao, S.; Petralia, F.; Kawaler, E.; Mundt, F.; Krug, K.; Tu, Z.; Lei, J. T.; Gatza, M. L.; Wilkerson, M.; Perou, C. M.; Yllapantula, V.; Huang, K.-L.; Lin, C.; McLellan, M. D.; Yan, P.; Davies, S. R.; Townsend, R. R.; Skates, S. J.; Wang, J.; Zhang, B.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Ding, L.; Paulovich, A. G.; Fenyö, D.; Ellis, M. J.; Carr, S. A. Proteogenomics Connects Somatic Mutations to Signalling in Breast Cancer. *Nature* **2016**, *534* (7605), 55–62.