Molecular Omics

ROYAL SOCIETY OF CHEMISTRY

RESEARCH ARTICLE

View Article Online
View Journal

Cite this: DOI: 10.1039/d3mo00042g

Integrative analysis of cancer dependency data and comprehensive phosphoproteomics data revealed the EPHA2-PARD3 axis as a cancer vulnerability in KRAS-mutant colorectal cancer†

Daigo Gunji, [abc] Ryohei Narumi, [ab] Satoshi Muraoka, [abd] Junko Isoyama, [ab] Narumi Ikemoto, [ab] Mimiko Ishida, [ab] Takeshi Tomonaga, [ab] Yoshiharu Sakai, [c] Kazutaka Obama [c] and Jun Adachi *[abde]

Colorectal cancer (CRC), a common malignant tumour of the gastrointestinal tract, is a life-threatening cancer worldwide. Mutations in KRAS and BRAF, the major driver mutation subtypes in CRC, activate the RAS pathway, contribute to tumorigenesis in CRC and are being investigated as potential therapeutic targets. Despite recent advances in clinical trials targeting KRASG12C or RAS downstream signalling molecules for KRAS-mutant CRC, there is a lack of effective therapeutic interventions. Therefore, understanding the unique molecular characteristics of KRAS-mutant CRC is essential for identifying molecular targets and developing novel therapeutic interventions. We obtained in-depth proteomics and phosphoproteomics quantitative data for over 7900 proteins and 38 700 phosphorylation sites in cells from 35 CRC cell lines and performed informatic analyses, including proteomics-based coexpression analysis and correlation analysis between phosphoproteomics data and cancer dependency scores of the corresponding phosphoproteins. Our results revealed novel dysregulated protein–protein associations enriched specifically in KRAS-mutant cells. Our phosphoproteomics analysis revealed activation of EPHA2 kinase and downstream tight junction signalling in KRAS-mutant cells. Furthermore, the results implicate the phosphorylation site Y378 in the tight junction protein PARD3 as a cancer vulnerability in KRAS-mutant cells. Together, our large-scale phosphoproteomics and proteomics data across 35 steady-state CRC cell lines represent a valuable resource for understanding the molecular characteristics of oncogenic mutations. Our approach to predicting cancer dependency from phosphoproteomics data identified the EPHA2-PARD3 axis as a cancer vulnerability in KRAS-mutant CRC.

Received 28th February 2023,
Accepted 12th May 2023

DOI: 10.1039/d3mo00042g

rsc.li/molomics

Introduction

Colorectal cancer (CRC) is the second most fatal cancer worldwide.[1] Activating mutations in KRAS and BRAF, which are in the Ras pathway, contribute to tumorigenesis in CRC and are being investigated as potential therapeutic targets. As recommended in The National Comprehensive Cancer Network (NCCN) guidelines,[2] anti-epidermal growth factor receptor (EGFR) monoclonal antibodies are a clinically approved treatment for patients with wild-type (wt) RAS/BRAF, and BRAF inhibitors are approved for treating patients with BRAF mutations.[3,4] However, effective strategies targeting KRAS mutations have not yet been developed, and these mutations occur in approximately 45% of CRC patients.[5] KRAS has historically been considered "undruggable" because mutant KRAS protein lacks a pharmacologically targetable pocket.[6] Although selective KRAS G12C inhibitors have recently been developed, KRAS G12C is only present in 3% of patients with CRC,[7] and a recent clinical trial showed limited efficacy with monotherapy with the KRAS G12C inhibitor sotorasib.[8] Furthermore, no effective inhibitors have been identified targeting the same or parallel signalling pathways of the RAS signalling pathway. For example, previous studies have demonstrated that pan-RAF inhibitors, which are inhibitors of signalling molecules downstream of RAS, are not efficient for

[a] Laboratory of Proteome Research, National Institute of Biomedical Innovation, Health and Nutrition, Osaka, 567-0085, Japan. E-mail: jun_adachi@nibiohn.go.jp; Fax: +81-72-641-9861; Tel: +81-72-641-9862
[b] Laboratory of Proteomics for Drug Discovery, Center for Drug Design Research, National Institute of Biomedical Innovation, Health and Nutrition, Osaka, 567-0085, Japan
[c] Department of Surgery, Kyoto University Graduate School of Medicine Faculty of Medicine, Kyoto, 606-8507, Japan
[d] Laboratory of Clinical and Analytical Chemistry, Center for Drug Design Research, National Institute of Biomedical Innovation, Health and Nutrition, Osaka, 567-0085, Japan
[e] Laboratory of Proteomics and Drug Discovery, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto, 606-8501, Japan

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3mo00042g

treating cancers with KRAS mutations.[9] Unacceptable toxicity was reported for a combination therapy of PI3K inhibitors, which target a parallel pathway, and MEK inhibitors, and no response was observed in the CRC patients enrolled.[10] Therefore, understanding the unique molecular characteristics of KRAS-mutant CRC is essential for identifying molecular targets and developing novel therapeutic interventions for KRAS-mutant CRC.

Mass spectrometry (MS)-based proteomics and phosphoproteomics technology is widely used to analyse KRAS oncogenic signalling. Recently, a comprehensive proteomics and phosphoproteomics study of 43 KRAS-mutant cancer cells from 7 tissues revealed diverse molecular signature subtypes in KRAS-mutant cancers.[11] The authors identified three KRAS-mutant subsets across different KRAS-mutant cancer cells and determined the molecular signatures independent of tissue origin. However, they profiled proteomics and phosphoproteomics data in cells from only 11 CRC cell lines out of a total of 43 cell lines, with limited information available on the comprehensive analysis of molecular signatures triggered by oncogenic mutations in CRC. Furthermore, the underlying differences in molecular mechanisms between mutant KRAS and BRAF signalling have not been investigated. Therefore, characterizing differences in molecular signatures among oncogenic KRAS-mutant, BRAF-mutant and KRAS, BRAF wt CRC cells by large-scale proteomics and phosphoproteomics analyses could provide significant insights into developing therapeutic strategies for KRAS-mutant CRC tumours.

Cancer driver mutations are translated to mutated residues, leading to mutation-directed differential protein–protein interactions (PPIs) and pathway rewiring. Previous studies have reported that computational and experimental approaches can detect differentially regulated proteins that directly interact with mutant proteins produced from mutated oncogenes.[12,13] However, generating a comprehensive static protein–protein interaction network catalogue after making a perturbation requires substantial experimental effort, and no information is currently available on a comprehensive map of protein–protein association dynamics. Coexpression analysis based on correlation coefficients of relative protein expression levels across a large panel of cancer cell line proteomics data has been used to address this challenge.[14] Using this method, dysregulated functional protein–protein network associations based on breast cancer subtypes were identified. However, a study to map dysregulated protein–protein associations based on oncogenic mutational status across CRC cells has not been reported. Therefore, applying this coexpression analysis to proteomics data from CRC cell lines may provide information regarding the dysregulation of protein–protein associations specific to the KRAS-mutant cell subgroup.

Protein phosphorylation is an important posttranslational modification (PTM) in biological processes and the regulation of signalling pathways. Based on recent technological advances in MS-based phosphoproteomics analysis, PhosphoSitePlus[15] contains 239 776 human phosphosites to date. However, only 11 927 human phosphorylation sites (less than 5%) have been manually curated as regulatory sites by expert scientists. The functions of the remaining phosphorylation sites are unclear. A machine learning approach has been used to predict and score functionally unknown protein phosphorylation sites. In that study, 112 large-scale publicly available phosphoproteome datasets from the PRIDE database were assessed to rank the functional importance of human phosphorylation sites.[16] Recent studies have shown that this computational prediction of phosphorylation sites allows for ranking and prioritizing the significantly regulated phosphorylation sites.[17] However, this prediction method relies on functional and physical features of the protein fragments, leading to the inclusion of phosphorylation sites that do not impact cancer phenotypes as functionally important phosphorylation sites. The recently provided CRISPR/Cas9 screening data (DepMap[18]), a comprehensive map of cancer dependence, has been utilized to explore genes associated with cancer phenotypes. Previous studies have reported a relationship between high protein expression levels and genetic dependency in cancer cells.[19] Therefore, genes with a high correlation between protein expression and genetic dependency are thought to be potential therapeutic targets for cancer vulnerability. However, there is no comprehensive understanding of the association between cancer dependency and phosphorylation sites that are essential for the regulation of signalling. Assessment of the predictive power of phosphorylation levels for cancer dependencies could provide a unique opportunity to explore novel therapeutic candidate targets based on phosphorylation sites.

In this study, we investigated the effect of KRAS and BRAF oncogenic mutations on phosphorylation signalling and dysregulated protein-association networks by MS-based quantitative proteomics and phosphoproteomics analyses of 35 CRC cell lines. By integrating our phosphoproteomics data with gene dependency data obtained from DepMap CRISPR/Cas9 screening, we predicted phosphorylation sites associated with cancer vulnerability in KRAS-mutant CRC cancer, and these predictions are worth further validation as potential drug targets in the future.

Materials and methods

Cell lines

The 39 CRC cell lines used in this study (C2BBe1, CaR-1, CCK-81, COLO 320HSR, COLO205, COLO320 DM, CW-2, GP5d, HCT116, HCT-15, HT115, HT29, HT55, LIM1215, LoVo, LS1034, LS123, LS180, LS411N, LS513, MDST8, NCI-H508, NCI-H716, NCI-H747, RCM-1, RKO, SK-CO-1, SNU-C1, SNU-C2B, SW1116, SW1417, SW1463, SW48, SW480, SW620, SW837, SW948, T84, and WiDr) were obtained from ATCC, ECACC, JCRB and Riken BRC (Table S1A, ESI†). Cells were grown under standard conditions (5% CO_2, 95% relative humidity and 37 °C) in the recommended medium for each cell source. Experiments were performed with cells at <10 passages from the original vial. Purchase information, culture type, and specific culture conditions are detailed in Table S1A (ESI†).

Preparation of samples for MS-based proteome and phosphoproteome analyses

Cells were cultured to 80% confluence in 15 cm diameter dishes. Cells were lysed in phase transfer surfactant (PTS)

buffer (50 mM ammonium bicarbonate, 12 mM sodium deoxycholate, 12 mM sodium lauroyl sarcosinate) supplemented with PhosSTOP and a cOmplete Protease Inhibitor cocktail (Roche, Basel, Switzerland). Each sample was boiled at 95 °C for 10 min. The lysate was further sonicated three times (15 min per cycle) with a Bioruptor sonicator (Cosmo Bio, Tokyo, Japan). Protein concentrations were determined by DC Protein Assay (Bio-Rad Laboratories, Inc. Hercules, CA, USA). Lysates were stored at −80 °C until use. We used six different genetic background cell line mixtures, namely, LIM1215, COLO320DM, HCT116, SW480, RKO, and WiDr, that were pooled for the global reference internal standard (GIS) for these proteomics analyses. Information on KRAS and BRAF mutational status and microsatellite instability (MSI) status are detailed in Table S1B (ESI†).[20] Each sample lysate containing 1.6 milligrams of protein in 1 mL of PTS buffer (1.6 mg mL^{-1}) was reduced with 10 mM TCEP, alkylated with 20 mM iodoacetamide, and quenched with 21 mM L-cysteine. The samples were digested with trypsin (protein weight ratio: 1/50; Wako) and Lys-C (1 mAU/25 μg; Wako) for 16 h at 37 °C. Samples were acidified with 1% TFA and centrifuged at 20 000 × *g* for 10 min at 4 °C. Supernatants were desalted with OASIS HLB cartridges (Waters, Milford, MA, USA), and 0.1% of each sample was used for proteome analysis. The remaining portions of the samples were subjected to phosphopeptide enrichment *via* an immobilized metal affinity chromatography (IMAC) method.[21]

Peptide/phosphopeptide-enriched samples were labelled with TMTpro 16-plex reagents (Thermo Scientific A44522, Lot No: LV310412). Technical triplicates sampled from each cell line sample were randomly distributed across 7 TMT batches. The 16th channel of each experiment was reserved for the GIS. The TMT channel arrangement information is detailed in Table S1C and Fig. 1A (ESI†). Peptide/phosphopeptide were labelled with TMTpro 16-plex reagents at room temperature for 1 h, and

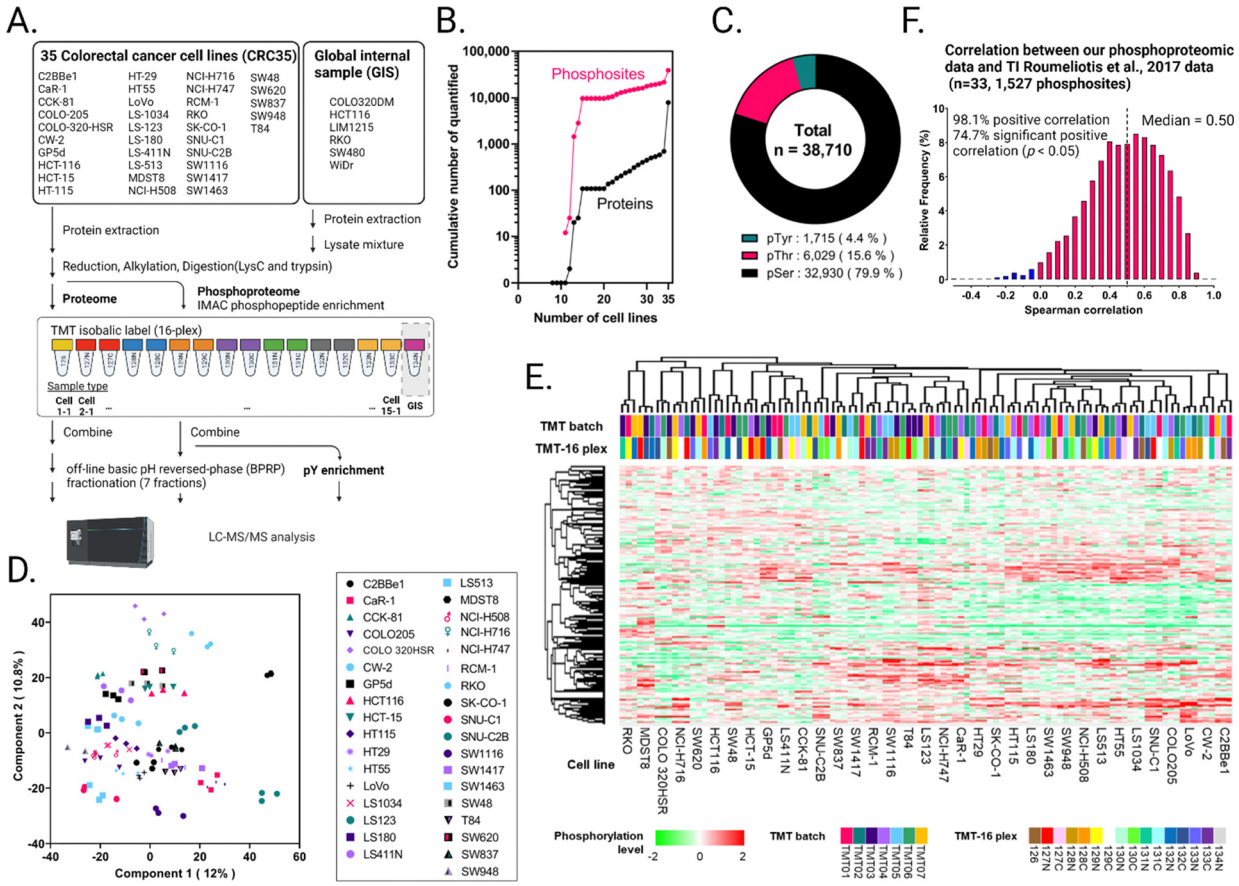


Fig. 1 **Quantitative phosphoproteomics analysis of 35 colorectal cancer cell lines (CRC35)**. (A) Experimental workflow. Three technical replicates prepared from each cell line (105 samples in total) were randomized into seven batches such that each batch contained one global internal standard (GIS) sample. Digestion, desalting, IMAC phosphopeptide enrichment, and TMTpro labelling steps were performed. After TMTpro pooling, a portion of the phosphopeptides was enriched in pY. All samples were fractionated by basic-pH reversed-phase (BPRP) chromatography and analysed *via* LC–MS/MS, and raw data were processed with MaxQuant (version 1.6.14.0). (B) Cumulative number of quantified proteins and phosphosites. (C) Proportions of phosphoserine (pSer, black), phosphothreonine (pThr, red), and phosphotyrosine (pTyr, green). (D) and (E) PCA (D) and unsupervised hierarchical clustering analysis (E) of technical replicates of 35 CRC cell lines. The 8432 phosphorylation sites quantified in all 105 samples were used for analysis. (F) Spearman's correlation analysis between our phosphoproteomics data and data from Theodoros I *et al.* (2017). The 1527 phosphosites across 33 CRC cell lines shared in both datasets were used in this study. Abbreviations: TMT = tandem mass tags pro (16-plex); LC–MS/MS = liquid chromatography-tandem mass spectrometry.

the reaction was quenched with hydroxylamine to a final concentration of 0.3% (v/v). The TMT-labelled phosphopeptide samples were pooled and divided into 10% for global phosphoproteome and 90% for phosphotyrosine (pY) proteome analyses. pY enrichment was performed using a pY1000 antibody as previously reported.[22] All samples were vacuum centrifuged to near dryness.

The TMT-labelled peptide/phosphopeptides used for total proteome and global phosphoproteome analysis were subjected to off-line basic pH reversed-phase (BPRP) fractionation.[23] We used a Thermo Scientific UltiMate 3000 UHPLC system equipped with a dual wavelength detector (set at 220 nm and 280 nm). The mobile phases were BPRP-A (5 mM ammonium bicarbonate pH 9.2) and BPRP-B (5 mM ammonium bicarbonate pH 9.2 and 60% acetonitrile). The LC gradient for total proteomics samples was carried out as follows: 5–40% BPRP-B for 7 min, 40–60% BPRP-B for 19 min and then 60–99% BPRP-B for 7 min. The gradient for phospho-proteomics samples was carried out as follows: 1–19.5% BPRP-B for 7 min and then 19.5–64% BPRP-B for 27 min. Digested peptides equivalent to 10 micrograms of protein were separated on an L-column3 C18 column (5 μm particles, 0.3 mm ID and 150 mm in length, Chemicals Evaluation and Research Institute, Japan) at a flow rate of 2 μL min^{-1}. The peptide mixture was fractionated into a total of 21 fractions, which were consolidated into 7 fractions. Samples were subsequently vacuum centrifuged to near dryness. Each fraction was reconstituted in 2% acetonitrile and 0.1% trifluoroacetic acid for LC–MS/MS processing.

MS data acquisition

Liquid chromatography-tandem mass spectrometry (LC–MS/MS) was performed with an Ultimate 3000 U-HPLC system (Thermo Fisher Scientific, Waltham, MA, USA) and an HTC-PAL autosampler (CTC Analytics, Zwingen, Switzerland) coupled to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). Peptides were delivered to an analytical column (75 μm × 30 cm, packed in-house with ReproSil-Pur C18-AQ, 1.9 μm resin, Dr Maisch, Ammerbuch, Germany) and separated at a flow rate of 280 nL min^{-1} using a 145-min gradient from 5% to 30% solvent B (solvent A, 0.1% FA; solvent B, 0.1% FA and 99.9% acetonitrile). The Orbitrap Fusion Lumos mass spectrometer was operated in the data-dependent mode.

For global proteome and phosphoproteome analyses, survey full-scan MS spectra (m/z 375 to 1500) were acquired with the Orbitrap with 120 000 resolution after accumulation of ions to a 4×10^5 target value. The maximum injection time was set to 50 ms, and dynamic exclusion was set to 30 s. MS2 analysis consisted of higher-energy collisional dissociation (HCD); AGC 1×10^5; normalized collision energy (NCE) 32; maximum injection time 105 ms; 50 000 resolution and isolation window of 0.7 Da. For phosphotyrosine proteome analysis, survey full-scan MS spectra (m/z 350 to 1500) were acquired with the Orbitrap with 120 000 resolution after accumulation of ions to a 4×10^5 target value. The maximum injection time was set to 100 ms, and dynamic exclusion was set to 5 s. MS2 analysis consisted of higher-energy collisional dissociation (HCD); AGC 1×10^5; normalized collision energy (NCE) 32; maximum injection time 315 ms; 120 000 resolution and isolation window of 0.7 Da.

We performed a total of 154 LC–MS/MS measurements for 18.6 days, which included 98 measurements for total proteome analysis, 49 measurements for phosphoproteome analysis, and 7 measurements for tyrosine-enriched phosphoproteome analysis.

MS data analysis

Raw MS data were processed with MaxQuant (version 1.6.14.0) supported by the Andromeda search engine for peak detection and quantification.[24] The MS/MS spectra were searched against the UniProt[25] human database (Homo Sapiens, organism ID: 9606, from March 2020) with the following search parameters: full tryptic specificity, up to two missed cleavage sites, carbamidomethylation of cysteine residues set as the fixed modification, and N-terminal protein acetylation and methionine oxidation set as variable modifications. Reporter ion intensities were processed by PSM-level ratio normalization.[26] For phosphoproteome analysis, phosphorylation of serine, threonine, and tyrosine was added as a variable modification. The search results were filtered to a maximum false discovery rate of 0.01 for the protein groups, PSMs, and phosphorylation sites. We required two or more unique/razor peptides for protein identification and a ratio count of two or more for protein quantification. PTM sites with a measured localization probability ≥ 0.75 were considered to be localized. MaxQuant provides three quantification values (multiplicity) with the phosphorylation sites being quantified with a single, doubly, and triply-phosphorylated peptide. We refer to these three values as phosphorylation events. Quantitative values of the phosphorylation sites across different fractions were automatically integrated and summarized in "Phospho (STY) Sites.txt" by MaxQuant.

Bioinformatic analysis

Statistical analysis and data visualization were performed using Perseus 1.6.14.0,[27] R software (version 4.2.0), and GraphPad Prism version 9.0.0 for Windows [GraphPad Software, San Diego, CA, USA, **https://www.graphpad.com**]. The quantitative ratio was log 2-transformed, and the values were normalized by the median-centring of each sample. We used the phosphorylation site with a quantified value of at least one sample. Imputation was not performed for missing values. Quantitative values for each technical replicate were averaged and used for further analysis. Comparison of a continuous variable in 2 or more than 2 groups was performed using either a parametric test (Welch's t test or one-way ANOVA followed by Tukey's multiple comparisons test). We used phosphorylation sites and proteins for which at least half of the samples in each mutational group (BRAF/KRAS wt (5/9), BRAF mutation (4/7), KRAS mutation (9/18)) had quantitative values. All tests were 2-tailed, and $p < 0.05$ was considered significant. Correlation analysis between continuous variables was performed using Spearman's rank-order correlation. Twenty-eight cell lines from CRISPR–Cas9 knockout screen data (DepMap Public 22Q2 CRISPR_gene_dependency.csv) and 18 cell lines from CTD2 cell line pharmacological profiling data (CTRPv2, CTRPv2.0_2015_ctd2_ExpandedDataset.zip) were common to the

35 cell lines used in our study. For Spearman correlation analysis between phosphoproteomic data and dependency scores, minimum cells being used for correlation was 14. Over-representation analyses (ORAs), including Gene Ontology (GO) terms (BP, CC, and MF), the protein complex database CORUM,[28] and Kyoto Encyclopedia of Genes and Genomes (KEGG)[29] pathway enrichment analysis, were performed with WebGestalt 2017 (WEB-based GEne SeT AnaLysis Toolkit).[30] Kinase activity prediction was performed using site-centric posttranslational modification-signature enrichment analysis (PTM-SEA)[31] with the GenePattern[32] platform, a seven-amino-acid sequence flanking the phosphosite as an identifier and the human kinase/pathway definitions of PTMsigDB (v.1.9.0) according to the following parameters: (gene.set.database = ''ptm.sig.db.all.flanking.human.v1.9.0.gmt'', sample.norm.type = ''rank'', weight = 0.75, statistic = ''area.under. RES'', output.score.type = ''NES'', nperm = 1000, global.fdr = TRUE, min.overlap = 5, correl.type = ''z.score''). For protein–protein association network construction, proteomics profiles were correlated using Spearman's rank correlation coefficients. We calculated correlation coefficients for all possible protein pairs among these 7777 proteins in the R environment using the function rcorr() found in the Hmisc package. Constructed networks were visualized in Cytoscape. To compare two correlation coefficients, we used the R package cocor.[33] The hotspot mutation information data (DepMap Public 22Q2 CCLE_mutations_bool_hotspot.csv) were downloaded from the DepMap website (**https://depmap.org/portal/**). High-throughput CRISPR–Cas9 knockout screen data (DepMap Public 22Q2 CRISPR_gene_dependency.csv) were obtained from DepMap (**https://depmap.org/portal/**). The literature-derived phosphoproteome dataset was extracted from the curated site-centric databases in PTMsigDB, version 1.9.0.[31] This dataset contains >8000 phosphorylation sites across three signature categories. To investigate the functional importance of phosphorylation sites in signal transduction, we extracted the 1139 phosphorylation sites of PTMsigDB annotated as ''PATH-NP'' and ''PATH-WP''. The cell line pharmacological profiling data from the Cancer Therapeutics Response Portal (CTRPv2, CTRPv2.0_2015_ctd2_ExpandedDataset.zip) dataset were downloaded from the DepMap website (**https://depmap.org/portal/**).[34]

Data availability

All raw data files generated in this study were deposited into jPOST, a public proteome database certified by the ProteomeXchange Consortium,[35] under the accession number PXD040404 (**https://repository.jpostdb.org/preview/157215138363fd430e179d0**, Access key: 2594).

Results

Quantitative proteomics and phosphoproteomics analyses of 35 CRC cell lines (CRC35)

We performed proteomics and phosphoproteomics analyses of 35 CRC cell lines. The proteomics and phosphoproteomics data were acquired using high-resolution MS. Proteomics and phosphoproteomics quantification across the samples was achieved by a tandem mass tag (TMT) labelling approach (Fig. 1A). Each cell line was acquired from three replicates and randomly labelled based on the batch arrangement shown in Table S1C (ESI†). One TMT channel in each batch was dedicated to the GIS, representing a mixture of 6 different genetic background cell lines (Fig. 1A and Table S1B, C, ESI†). The resulting dataset contained the total number of quantitative data for 7913 proteins, 38 710 phosphorylation sites and 6626 phosphorylated leading proteins (Fig. 1B and Table S1D, E, ESI†). The 38 710 phosphorylation sites included 30 930 phosphoserine sites (79.9%), 6029 phosphothreonine sites (15.6%), and 1715 pY sites (4.4%) (Fig. 1C). To evaluate whether there were batch effects from TMT 16-plex labelling, we performed unsupervised hierarchical clustering and principal component analyses (PCAs) on the phosphorylation site levels of each sample relative to the GIS (Fig. 1D and E). We observed that three technical replicates from the same cell line clustered together despite being distributed in different TMT batches. PCA of three technical replicate data points indicated no observable batch effect. We further compared our phosphoproteomics data with those of a previously reported large-scale phosphoproteomics study of CRC cell lines.[36] In a previous study, a panel of 50 CRC cell lines was used for phosphoproteomics based on isobaric peptide labelling (TMT-10-plex) and MS3 quantification. They identified 11 647 phosphopeptides in total, whereas in the current study, we identified 44 165 phosphorylation sites in total, approximately 3.8 times as many as in the previous results. We also assessed the correlation of phosphorylation levels between our data and two previously reported large-scale phosphoproteomic datasets obtained from independent laboratories using different sample processing protocols.[36,37] We performed phosphosite-wise Spearman's correlation analyses of the phosphoproteomic data using complete phosphoproteomic data from all overlapping cells between our data and previously reported data. The median Spearman's correlation coefficients were 0.50 (n = 33, 1527 phosphosites (*vs.* the data obtained by TI Roumeliotis *et al.* (2017)) and 0.44 (n = 23, 1827 phosphosites (*vs.* the data obtained by Frejno, M., Meng, C., Ruprecht B, *et al.* (2020)) (Fig. 1F and Fig. S1, ESI†). We observed that over 95% of the phosphorylation sites were positively correlated with both datasets. These data demonstrate that our large-scale phosphoproteome data are deeply quantitative and highly reproducible. The quantitative values for each technical replicate were averaged and used for further analysis.

Comparative phosphoproteomics analysis of CRC cells stratified based on the BRAF and KRAS mutational status revealed distinct signalling pathways

To identify phosphoproteomics features associated with the KRAS and BRAF mutational status, we classified cell lines into three groups based on the KRAS and BRAF hotspot mutational status (BRAF/KRAS wt, BRAF mutation, and KRAS mutation). Many studies have indicated that mutations in KRAS and BRAF are mutually exclusive in CRC.[38] Out of 35 CRC cell lines, 34 cell lines (except COLO320-HSR) were defined into 3 subgroups

based on the hotspot mutational status indicated in the Cancer Dependency Map.[39] To identify phosphorylation sites that exhibited significant upregulation across the three different mutational statuses, we performed one-way ANOVA and Tukey's HSD *post hoc* test. The 343 differentially regulated phosphorylation sites were grouped into 3 clusters (p value $<$ 0.05, FDR $<$ 0.1; Fig. 2A and Table S2A, ESI†). To capture the molecular features of each subgroup, we performed an overrepresentation analysis of the phosphoproteins of differentially regulated phosphorylation sites in each cluster (Fig. 2B). The results of the KEGG pathway analysis with p values less than 0.05 are summarized in Fig. 2C and Table S2C (ESI†). The ErbB signalling pathway was significantly enriched in all three subgroups. To compare the effects of KRAS and BRAF mutations on the downstream ErbB signalling pathway, we illustrated the phosphoproteomics alterations in the ErbB

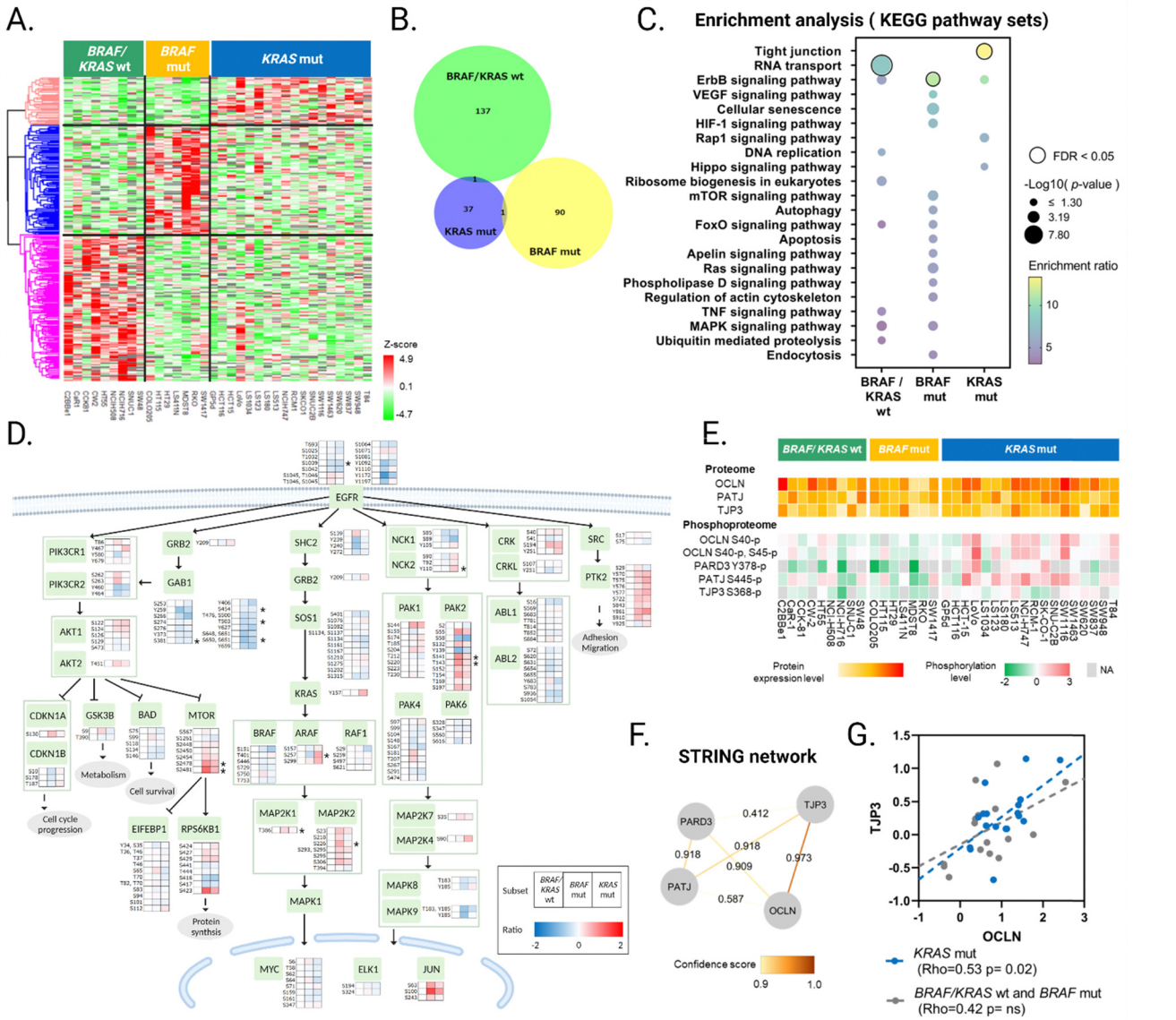


Fig. 2 **Molecular features of each subset of KRAS and BRAF mutational status**. (A) Hierarchical clustering of 343 significantly regulated phosphorylation sites according to the KRAS and BRAF mutational status (p value $<$ 0.05, FDR $<$ 0.1, one-way ANOVA followed by Tukey's *post hoc* test). (B) Venn diagram of significantly upregulated phosphorylated proteins in each cluster. The phosphoproteins identified from each of the three subgroups were used for further analysis. (C) KEGG pathway enrichment analysis of differentially regulated phosphoproteins in each CRC subgroup. A bubble chart shows the enrichment of differentially expressed genes in signalling pathways. The colour intensity reflects the enrichment ratio, and the node size reflects the $-\log 10$ p value. (D) Phosphorylation-level alterations in the ErbB signalling pathway. The colour represents the ratio of the phosphorylation level in BRAF mut or KRAS mut *versus* BRAF/KRAS wt. Significantly differentially upregulated phosphosites (p $<$ 0.05, FDA $<$ 0.1 one-way ANOVA followed by Tukey's test for multiple comparisons) are depicted by asterisks. The colour indicates upregulation (red) or downregulation (blue). (E) Protein expression and phosphorylation level of the key components of tight junctions. (F) String network analysis of the key components of tight junctions. The edge label shows the confidence score based on the STRING database. (G) Correlation between OCLN and TJP3 protein expression levels in KRAS-mut cell lines (Spearman's test).

signalling pathway in each subgroup (Fig. 2D). Asterisks denote significantly different phosphorylation sites among the three groups (one-way ANOVA, p value < 0.05, FDR < 0.1). A variety of components were significantly different among the three subgroups. In the BRAF/KRAS wt subgroup, we identified upregulation of the phosphorylation level of EGFR S1039 and GAB1 S381, S454, T503, and S650/S651. The phosphorylation levels of other GAB1 proteins also tended to be upregulated in the BRAF/KRAS wt subgroup. In the BRAF mutation subgroup, the phosphorylation levels of MAP2K1 T386, MAP2K2 S222, MTOR S2478/S2481, and PAK2 S141/T143 were significantly upregulated. In the KRAS-mutant subgroup, the phosphorylation levels of ARAF S257 and NCK2 Y110 were significantly upregulated (Fig. S2A, ESI†). Interestingly, these alterations were not observed in protein expression levels (Fig. S2B and Table S2B, ESI†). These results showed that the phosphorylation levels of downstream components of ErbB signalling are differentially upregulated based on the BRAF and KRAS mutational status.

To investigate the KRAS-mutant-specific upregulated signature, we focused on the tight junction pathway (Fig. 2C). The phosphorylation levels of OCLN, PARD3, PATJ, and TJP3, components of tight junctions, were significantly upregulated in KRAS-mutant cells (Fig. 2E and Fig. S2C, ESI†). However, the protein expression levels of these genes, except for PARD3, which was not identified in this study, showed no statistically significant differences among the three CRC subgroups (Fig. 2E and Fig. S2D, ESI†). To examine the protein–protein association of upregulated phosphoproteins in the KRAS-mutant subgroup, we performed STRING network analysis. The ''tight junction'' and ''cell–cell junction'' terms were found to be most highly enriched in this analysis (Table S2D, ESI†). Fig. 2F shows the protein–protein interaction of the tight junction components and their confidence scores from the STRING database.[40] In this network, the OCLN and TJP3 proteins showed the highest confidence score. Indeed, we observed a significant positive correlation between both the protein expression levels and phosphorylation levels of OCLN and TJP3 across all CRC cell lines (Fig. S2E and F, ESI†). We next compared the correlation coefficients based on the KRAS mutational status at the proteomics level. A positive correlation was found between these two proteins across KRAS-mutant cells. In contrast, we found no statistically significant positive correlation across cell lines without KRAS mutations (Fig. 2G and Fig. S2G, ESI†). These results suggest that a functional tight junction complex is assembled in KRAS-mutant cells.

Proteomics-based protein coexpression analysis revealed unique protein–protein association network dysregulation in KRAS-mutant cells

Coexpression analysis allows the identification of functional protein–protein associations. Previous studies have demonstrated that the ratio of relative protein expression levels across samples is consistent in protein complexes.[14] To identify commonly occurring protein complexes, we performed a coexpression analysis of all 34 CRC cell lines. We then performed a coexpression analysis of 18 KRAS-mutant CRC cell lines and compared these results with those obtained with all 34 cell lines to characterize the dysregulated protein–protein association networks in KRAS-mutant cells (Fig. 3A).

Coexpression was defined using Spearman's rank correlation coefficients. Spearman's rank correlation coefficients for 30 240 864 possible protein pairs among 7777 proteins with expression data across 34 CRC cell lines were calculated for each subset. We first applied a positive correlation rate $R \geq 0.9$ filter to the KRAS-mutant cell subset. In total, we obtained 503 differentially coexpressed protein pairs among 672 proteins (Table S3A, ESI†). We then measured the difference in correlation coefficients between two subsets for each protein pair (subtracting the Rho value of all 34 CRC cell subsets from the Rho value of the KRAS-mutant cell subset) and ranked protein pairs based on correlation differences (Fig. 3A). We highlighted the top 10% and the bottom 10% of absolute values of the correlation differences between protein pairs. A larger absolute value of the correlation difference indicated KRAS-mutant cell line-specific network dysregulation (KRAS-specific protein pairs (top 10% of the correlation difference)). In contrast, if the correlation difference was low, proteins were coexpressed in all 34 CRC cell lines (commonly occurring protein pairs (bottom 10% of the correlation difference)).

To estimate the accuracy of the protein-derived networks, we used both functional and physical protein associations (the full STRING network) in the STRING database[40] as a benchmark. In the commonly occurring protein pairs (bottom 10%), we found 178 protein-based associations confirmed by the STRING database and 20 novel associations identified in our study (Fig. 3B and Fig. S3A, B, ESI†). The median Spearman's rank correlation coefficient between the protein–protein association of these 178 commonly occurring protein pairs (bottom 10%) was 0.78 in both the KRAS-mutant and all 34 CRC cell subsets (Fig. S3C and Table S3B, ESI†), indicating that the ratio of relative protein expression levels in protein complexes based on the STRING database is highly preserved across 34 CRC cell lines. We then performed overrepresentation analyses of the commonly occurring protein pairs of gene names using the CORUM sets (Fig. 3C). We identified well-known human protein complexes that were significantly enriched, such as ''ribosome'' and ''respiratory chain complex'' (Fig. 3B, C and Table S3C, ESI†). An example of coexpression for the two ribosome subunits RBL17 and RBL26 is shown in Fig. S3D (ESI†). We observed a high correlation between the protein expression levels for these known interactors in both subsets (R = 0.91 in KRAS-mutant cells, R = 0.93 in all 34 CRC cell types).

To investigate protein–protein association dysregulation in the KRAS-mutant cell lines, we compared the KRAS-specific protein pairs (top 10% of the correlation difference) of protein pairs of the correlation difference to protein associations (the full STRING network) in the STRING database. We identified 48 novel protein-based associations with significantly different correlation coefficients between the KRAS-mutant and all 34 CRC cell subsets (p value < 0.05, implemented in the R package cocor,[33] Fig. 3D and Fig. S3E, F, Table S3B, ESI†).

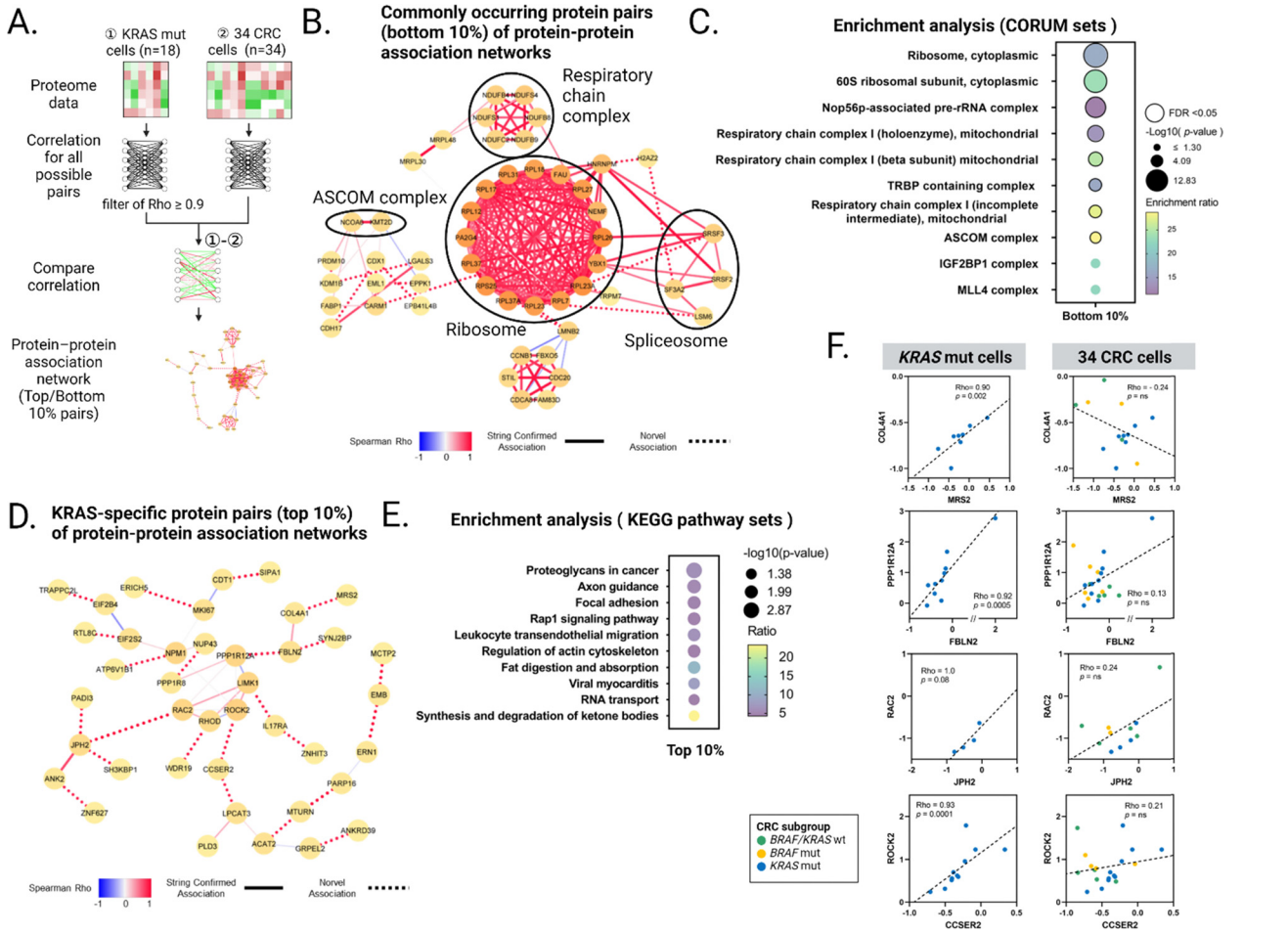


Fig. 3 **Protein coexpression analysis to identify protein–protein association dysregulation in KRAS-mutant cells**. (A) Computational workflow for the differential coexpression analysis. (B) Protein–protein association network of the commonly occurring protein pairs (bottom 10% of the correlation difference). The largest network is shown. The rest of the network is shown in Fig. S3B (ESI†). The solid lines indicate STRING-confirmed associations. The dotted lines indicate novel associations found in this work. (C) Bubble plot of CORUM enrichment analysis using gene names in the commonly occurring protein pairs (bottom 10% of the correlation difference). The bubble colour and size correspond to the *p* value and number of genes enriched in the pathway, respectively. (D) Protein–protein association network of the KRAS-specific protein pairs (top 10% of the correlation difference). The largest network is shown. The rest of the network is shown in Fig. S3F (ESI†). The solid lines indicate STRING-confirmed associations. The dotted lines indicate novel associations found in this study. (E) Bubble plot of KEGG enrichment analysis using gene names in the KRAS-specific protein pairs (top 10% of the correlation difference). The bubble colour and size correspond to the *p* value and number of genes enriched in the pathway, respectively. (F) Four correlation plots with KRAS-mut cells and 34 CRC cells. The protein expression of COL4A1, PPP1R12A, RAC2, and ROCK2, which are components of the focal adhesion, are on the *y*-axis.

The genes of KRAS-specific protein pairs were subjected to overrepresentation analyses of KEGG pathway sets (Fig. 3E). We observed that the RAC2 and ROCK2 genes were shared by multiple KEGG pathway enriched terms, including ''proteoglycans in cancer'', ''axon guidance'', ''focal adhesion'', and ''Rap1 signalling pathway'' (Table S3D, ESI†). Fig. 3F illustrates novel protein–protein association pairs for four genes (ROCK2, RAC2, PPP1R12A, COL4A1) enriched in the focal adhesion pathway as an example. Previous studies have reported that some of the selected genes are involved in the progression of CRC. COL4A1, a tumour angiogenesis indicator, has been identified to play a key role in CRC progression among all consensus molecular subtypes (CMSs) in a weighted gene-based protein–protein interaction network analysis.[41] MRS2, an essential magnesium transporter in the mitochondria, has been observed to have relatively high mRNA expression levels in colorectal adenocarcinoma and rectum adenocarcinoma samples compared to corresponding normal tissues.[42] Although further investigations are needed to confirm these findings, the protein complex of the magnesium transporter and extracellular matrix component could contribute to the phenotypic characteristics of KRAS-mutant CRC.

Phosphorylation sites crucial to the regulation of signal transduction were associated with cancer vulnerability

The gene dependency score profiles obtained from the CRISPER-Cas9 screens[43] in a large number of cell lines constitute a comprehensive catalogue of which genes are needed for cancer cell growth or viability. The prediction of specific phosphorylation sites that are important for cell growth and

viability by utilizing these gene dependency score profiles could contribute to the development of therapeutic opportunities. We undertook the following analysis based on the following assumption: if the phosphorylation levels of phosphorylation sites showed a high positive correlation to the gene dependency of the corresponding protein, these phosphorylation sites were associated with cancer vulnerability. We performed a pairwise Spearman's correlation analysis of phosphorylation levels and gene-dependency scores (DepMap CRISPR/Cas9 screening) of the corresponding genes for each phosphorylation site (Table S4, ESI†). We used phosphoproteomics data of 29 585 phosphorylation sites and the corresponding gene dependency scores of 5366 genes across 28 CRC cell lines.

To investigate whether phosphorylation sites involved in oncogenic signalling pathway activation are correlated with corresponding dependency scores, we focused on functionally annotated phosphorylation sites in PTMsigDB, which is a recently published database of curated resources for phosphosite-specific analysis.[31] In the signature of molecular signalling pathways (PATH) category in the PTMsigDB, 1137 phosphorylation sites were manually curated regardless of whether their functions were activated (upregulated) or deactivated (downregulated) in signalling pathways. Of these 1137 literature-annotated phosphorylation sites, 417 and 57 were annotated as upregulated and downregulated in our phosphoproteomics data, respectively. We compared the correlation coefficients of site-gene pairs between upregulated and downregulated pairs and found that the upregulated site-gene pairs showed substantially higher correlations than the downregulated site-gene pairs (p value = 0.0009, Welch's t test; Fig. 4A and Table S2A, ESI†). The top 5% significantly positively correlated site-gene pairs in literature-annotated phosphorylation sites are shown in Fig. 4B. For the 23 site-gene pair datasets, the proportions of pSer/Thr/Tyr phosphorylation sites were 26.0%, 4.4% and 69.6%, respectively (Fig. 4C). We noted that the fraction

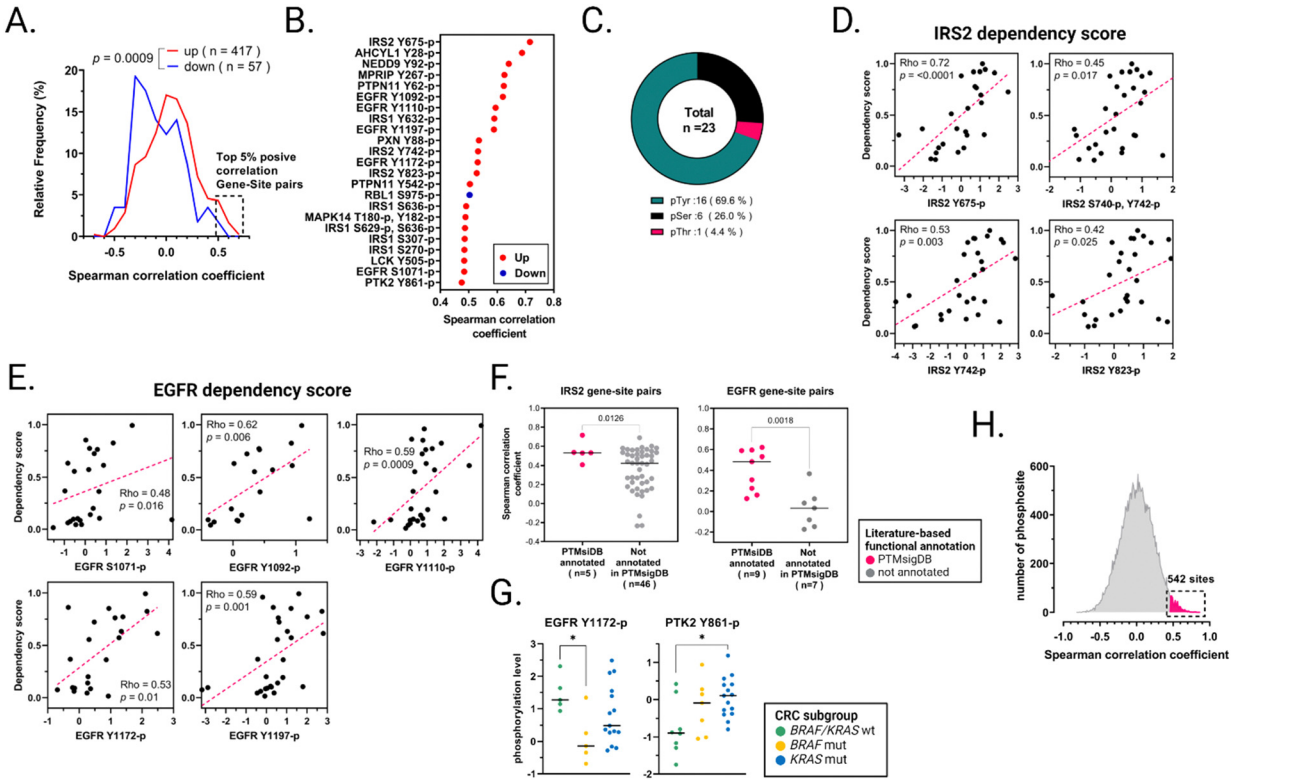


Fig. 4 **Functionally annotated phosphorylation sites that are positively correlated with genetic dependency play essential roles in protein function**. (A) Distribution of phosphosite-gene dependency correlation comparing "up" annotated phosphosites (n = 417, red) with "down" annotated phosphosites (n = 57, Blue) from PTMsigDB (p value = 0.0009, Welch's t test). The top 5% correlation gene-site pairs are shown in the black dotted frame. (B) List of Spearman's correlation coefficients of the top 5% positive correlation site-gene pairs. The red dots indicate "up" annotated phosphosites, and the blue dots indicate "down" annotated phosphosites (p value < 0.05). (C) Proportions of phosphoserine (pSer, black), phosphothreonine (pThr, red), and phosphotyrosine (pTyr, green). (D) Correlation between the phosphorylation level of 4 phosphorylation sites of IRS2 and their corresponding gene dependency scores. Rho and p values were obtained by Spearman's correlation test. (E) Correlation between the phosphorylation level of 5 phosphorylation sites of EGFR and their corresponding gene dependency scores. Rho and p values were obtained by Spearman's correlation test. (F) Beeswarm plot of the Spearman's correlation coefficients for annotated and non-annotated gene-site pairs for EGFR and IRS2 curated from the PTMsigDB. The colours of the dots indicate phosphorylation sites annotated in different databases (deep pink, PTMsigDB; grey, not annotated) (Welch's t test). (G) Beeswarm plots of the phosphorylation levels of EGFR Y1172 and PTK2 Y861 differentially regulated between CRC subgroups. The horizontal line represents the median. (H) Distribution of phosphosite-gene dependency correlation of all 29 585 phosphorylation sites. The 542 phosphorylation sites distributed in the black dotted frame have the same correlation coefficient (Rho > 0.475) as the top 5% or more of the correlation distribution of functionally annotated phosphorylation sites in PTMsigDB.

of tyrosine residues (pTyr) was 16.6-fold higher than that of the total quantitative phosphoproteomics data (Fig. S4A, ESI†).

To further investigate whether phosphorylation sites crucial to the regulation of signal transduction were associated with cancer vulnerability, we focused on the top two genes (IRS2 and EGFR) with the highest number of highly correlated site-gene pairs among the 23 site-gene pairs (Fig. 4D and E). We obtained 51 and 16 correlation coefficients of site-gene pairs for IRS2 and EGFR in our phosphoproteomics dataset, respectively. We compared the correlation coefficients of site-gene pairs between annotated and nonannotated pairs in PTMsigDB (PATH) and found that the correlation coefficients of literature annotated site-gene pairs were significantly higher than those for the unannotated site-gene pairs in both IRS2 (p value = 0.0126, Welch's t test) and EGFR (p value = 0.0018, Welch's t test) (Fig. 4F). We also identified EGFR Y1172 and PTK2 Y861, which were significantly upregulated in the BRAF/KRAS wt and KRAS-mutant subgroups, respectively, in this top 5% site-gene correlation list (Fig. 4B, G). The correlation coefficients for phosphorylation sites of IRS2, EGFR, and PTK2 are illustrated in Fig. S4B (ESI†). Taken together, these data suggest that phosphorylation sites crucial to the regulation of signal transduction are associated with cancer vulnerability and are potential drug targets.

To explore phosphorylation sites as potential drug targets, we filtered the correlation coefficients of all 29 585 phosphoproteomics data by Rho $>$ 0.475, which is the same coefficient as the top 5% or more of the correlation distribution of literature-annotated phosphorylation sites. We identified 542 sites with a high correlation coefficient (Fig. 4H and Table S2A, ESI†).

EPHA2 kinase is activated in KRAS-mutant CRC cells

Inference of protein kinase activity helps in better understanding kinase-regulated signalling pathways and is used to identify novel therapeutic targets. To infer upregulated kinase activities in KRAS-mutant cells, we first applied PTM-SEA to the phosphoproteomics data for each cell on a case-by-case basis. Then, the normalized enrichment scores (NESs) of each kinase were compared between the KRAS-mutant subgroup and the other subgroup, and the

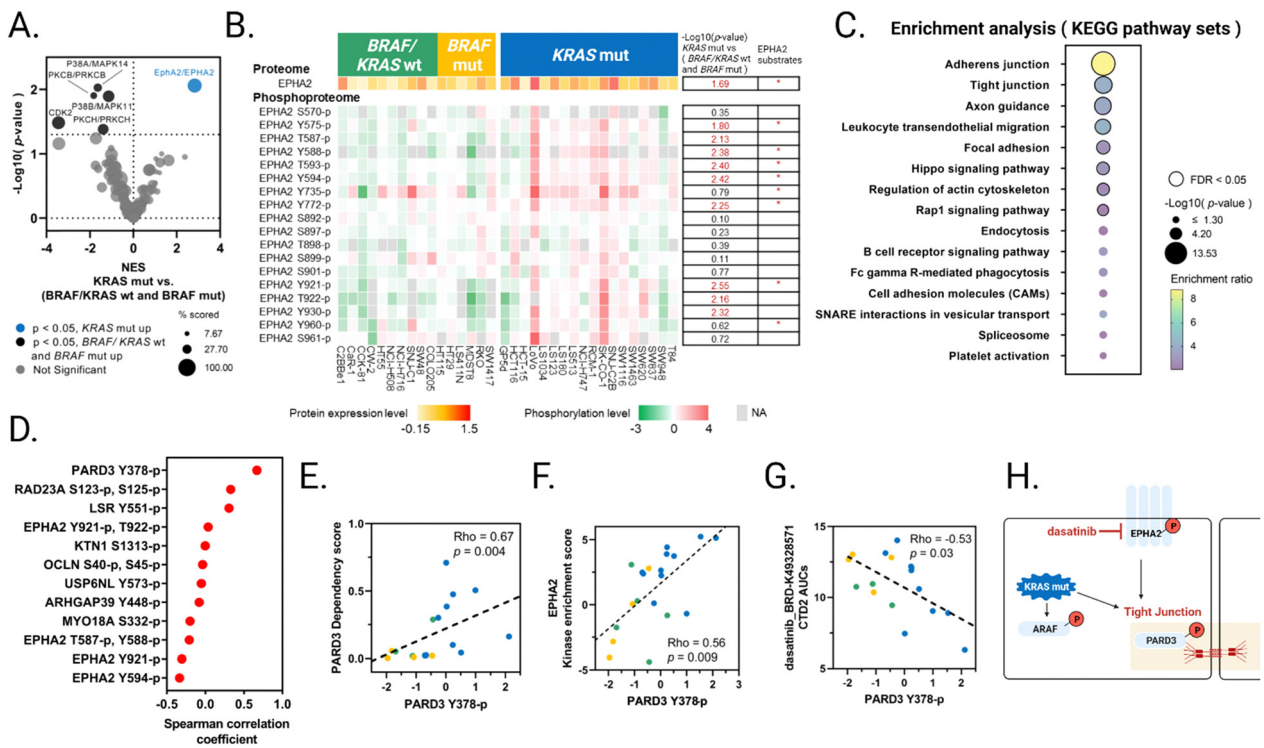


Fig. 5 **PARD3 Y378 is involved in KRAS-mutant cell viability**. (A) Volcano plot of the enrichment of various phosphoproteome signatures. The x-axis indicates the normalized enrichment scores between KRAS-mut and BRAF/KRAS wt and BRAF mut, and a positive value indicates upregulated phosphoproteome signatures in KRAS-mut cells. The dot size indicates the percent overlap of phosphorylation sites associated with the signature and quantified phosphorylation sites in CRC35. The blue colour indicates significantly upregulated signatures in KRAS-mut cells. (B) Heatmap of the protein expression levels of EPHA2 and the phosphorylation levels of each phosphorylation site of EPHA2. $-\log 10$ p values were determined by Welch's t test between KRAS-mut and BRAF/KRAS wt and BRAF mut. EPHA2 substrates are shown as asterisks. (C) Bubble plot of KEGG enrichment analysis using gene names in the top 5% of phosphorylation sites correlated with EPHA2 kinase NES. The bubble colour and size correspond to the p value and number of genes enriched in the pathway, respectively. (D) List of the phosphorylation sites that are significantly upregulated in KRAS-mutant cells among the top 5% of phosphorylation sites that are positively correlated with the EPHA2 kinase activity score. The x-axis shows the Spearman's correlation coefficients between the phosphorylation levels of phosphorylation sites and gene-dependency scores of the corresponding genes for each phosphorylation site. (E) Correlation plot of the phosphorylation level of PARD3 Y378 and PARD3 dependency scores. (F) Correlation plot of the phosphorylation level of PARD3 Y378 and EPHA2 kinase enrichment scores. (G) Correlation plot of the phosphorylation level of PARD3 Y378 and CTD2 AUCs of dasatinib. (H) Model of signalling pathways in KRAS-mut cells and novel therapeutic target candidates.

EPHA2 kinase signature was most significantly upregulated in the KRAS-mutant subgroup (Fig. S5A, Fig. 5A and Table S5A, ESI†). The protein expression and phosphorylation levels of EPHA2 itself were significantly upregulated in KRAS-mutant cells (Fig. S5B, Fig. 5B and Table S2A, B, ESI†). Furthermore, the protein expression level of EPHA2 and the phosphorylation levels of all EPHA2 substrates showed a significantly positive correlation with EPHA2 kinase activity (Fig. S5C and D, ESI†).

To investigate whether phosphorylation sites of EPHA2 are therapeutic candidate targets, we highlighted the correlation coefficients of each phosphorylation level of EPHA2 substrates and gene dependency scores of EPHA2. We found that all phosphorylation levels of EPHA2 substrates were inversely or not correlated with EPHA2 gene dependency scores (Fig. S5E, ESI†), suggesting that the EPHA2 phosphorylation sites themselves did not indicate cancer vulnerability in KRAS-mutant cells.

To further understand the downstream molecules of EPHA2 kinase-regulated signalling pathways, we performed a Spearman's correlation analysis of the phosphorylation levels of each phosphorylation site against the EPHA2 kinase activity scores. We identified 1095 phosphorylation sites among the top 5% positively correlated sites against EPHA2 kinase activity scores (Table S2A, ESI†), which are candidates for EPHA2 downstream components. The 478 phosphoproteins with these 1095 phosphorylation sites were subjected to functional annotation using the KEGG pathway database. These proteins were significantly enriched in pathways including "tight junction" and "focal adhesion" (Fig. 5C and Table S5B, ESI†). The results of this functional analysis were consistent with the characteristics of KRAS-mutant cells based on the comparative phosphoproteomics analysis and the proteomics-based protein coexpression analysis.

The EPHA2-PARD3 axis is involved in KRAS-mutant cell viability

To explore potential therapeutic target phosphorylation sites for KRAS-mutant cells in a downstream molecule in EPHA2 signalling, we filtered phosphorylation sites that were significantly upregulated in the KRAS-mutant subgroups from the top 5% of phosphorylation sites that were positively correlated with the EPHA2 kinase activity score. Fig. 5D shows the Spearman's correlation coefficients between the phosphorylation levels of phosphorylation sites and gene-dependency scores of the corresponding genes for each phosphorylation site for 12 phosphorylation sites resulting from this filtering. Among these 12 phosphorylation sites, the phosphorylation level of PARD3 Y378 showed the highest positive correlation coefficient with the corresponding gene dependency (Rho = 0.67, $p = 0.004$; Fig. 5D and E) and a positive correlation coefficient with EPHA2 kinase activity (Rho = 0.56, $p = 0.009$; Fig. 5F). These data indicate that PARD3 Y378 is involved in KRAS-mutant cell viability.

To predict therapeutic candidates showing sensitivity to the phosphorylation level of PARD3 Y378, we searched for kinase inhibitors that showed an inverse correlation with the phosphorylation level of PARD3 Y378 and ACU scores for drug sensitivity among the 18 FDA-approved kinase inhibitors in the CTD2 dataset from the Cancer Target Discovery and Development Network.[44] We found that dasatinib was the most inversely correlated drug (Rho = −0.53, $p = 0.03$; Fig. 5G and Fig. S5F, Table S5C, ESI†). Dasatinib is known to be a multitarget kinase inhibitor that also potently targets EPHA2. Taken together, these findings suggest that PARD3 Y378, which is upregulated in KRAS-mutant cells, is a downstream target phosphorylation site of EphA2 signalling and a potential predictive biomarker for dasatinib sensitivity (Fig. 5H).

The molecular features of KRAS mutant cell lines match those of KRAS mutant human tumour samples

To evaluate the degree to which the results of the CRC35 phosphoproteomic data matched the clinical phosphoproteomics data, we performed a comparative analysis of the publicly available phosphoproteomic data of tumour samples from the CPTAC colon cancer study.[45] We curated a characterized phosphoproteomics dataset of 96 cases of CRC. This study used isotopic labelling to quantify the relative abundance of phosphosites. We compiled available clinical information on the mutational status: BRAF/KRAS wt ($n = 46$), BRAF mut ($n = 33$), and KRAS mut ($n = 17$). The phosphoproteomics data of tumour samples among three different mutation subgroups were then subjected to one-way ANOVA and Tukey's HSD *post hoc* test, which revealed a total of 4103 differentially regulated phosphorylation sites among the three subgroups (p value < 0.05, FDR < 0.1; Fig. S6A, ESI†), and 1586 (1004) phosphorylation sites (phosphorylated proteins) were significantly upregulated in the KRAS mut samples compared to the other two groups (p value < 0.05, FDR < 0.1 Fig. S6B, ESI†). Notably, we found that phosphorylation changes in the KRAS mut group were markedly greater than those in the BRAF mut group. The results from the KEGG pathway overrepresentation analysis are shown in Fig. S6C (ESI†). Consistent with the CRC35 proteomics and phosphoproteomic data, we observed upregulation of both focal adhesion and tight junction pathways in the KRAS mut subgroup.

We then found a 34.5% overlap of phosphosites (11 578 phosphosites) in our CRC35 phosphoproteomic data (Fig. S6D, ESI†). Among these phosphosites, the ARAF-T253, S257 phosphosites were the only significantly upregulated phosphosites in the KRAS mut group in both the CRC35 and CPTAC datasets (Fig. S6E, ESI†). Although we quantified 89% of EPHA2 substrates (8 quantified phosphosites/9 annotated phosphosites) in the CRC35 data, only 22% (2 quantified phosphosites/9 annotated phosphosites) were quantified in the CPTAC data. The phosphorylation levels of EPHA2 Y772, which is significantly upregulated in KRAS mut CRC cells in the CRC35 data, tended to be increased, albeit not significantly, in the KRAS mut patient tumours in the CPTAC data (Fig. S6F, ESI†). The quantified numbers of EPHA2 Y772 substrates each subset in the CPTAC data were 11 (23.9%) in the BRAF/KRAS wt subgroup, 2 (11.7%) in the BRAF mut subgroup, and 7 (21.2%) in the KRAS mut subgroup. Furthermore, PARD3 Y378 was not quantified in the CPTAC dataset.

Discussion

We performed comprehensive proteomic, phosphoproteomic and pY-enriched phosphoproteomic analyses across 35 CRC

cell lines. We quantified over 1700 tyrosine phosphorylation sites, and this number is markedly higher than those previously reported for large-scale phosphoproteomics data of CRC (86 sites in the data found by TI Roumeliotis *et al.* (2017) and 1208 sites in the data reported by Frejno, M., Meng, C., Ruprecht B, *et al.* (2020)). These results provide a highly valuable resource for understanding important posttranslational modifications that regulate receptor tyrosine kinase-mediated signalling in CRC cells. The correlation between our phosphoproteomic data and the other two datasets obtained by independent laboratories was approximately 0.4–0.5. Furthermore, the correlation between the data reported by TI Roumeliotis *et al.* (2017) and the data found by Frejno, M., Meng, C., Ruprecht B, *et al.* (2020) was 0.38 ($n = 23,682$ phosphosites), suggesting that differences in cell culture conditions (medium condition, confluence at cell collection) and phosphoproteomic methods may influence the results. Further studies will be needed to improve the correlation between independent laboratories that utilise different sample processing protocols.

Although large-scale proteomics and phosphoproteomics analyses of cell lines have been performed, a comprehensive understanding of the molecular differences between mutant KRAS and BRAF signalling in CRC cells has not been established. In this study, we performed a deep proteomics and phosphoproteomics analysis of a global steady-state of 35 CRC cell lines and characterized phosphoproteomics features and dysregulated protein–protein association networks associated with the KRAS oncogenic mutation. We characterized the upregulation of tight junction signalling and novel dysregulated protein–protein association networks in KRAS-mutant cells. Furthermore, the EPHA2-PARD3 axis has been suggested to confer cancer vulnerability to KRAS-mutant cells by integrated analysis of phosphoproteomics data and gene dependency scores. Our results provide novel insights into the biological and therapeutic information of KRAS-mutant CRC.

Our data have demonstrated that phosphoproteomics data are more valuable for understanding distinct activated signalling pathways based on KRAS and BRAF mutational status than proteomics data. We observed that the ErbB signalling pathway was activated in all mutational subgroups. Interestingly, the phosphorylation levels of direct downstream targets of each mutated gene were significantly upregulated for each mutational status. These findings were not recapitulated at the protein expression level. We observed that the MAPK signalling pathway was upregulated in the BRAF mutation subgroup. Our results are consistent with previous studies showing that BRAF mutations constitutively activate the MAPK pathway in cancer cells.[46,47] Previous studies have shown that oncogenic KRAS activates several downstream phosphorylation signalling pathways, including the RAF/MEK/ERK, PI3K/AKT, and RALGDS/RAL pathways.[48] However, we did not observe the activation of any of these pathways in KRAS-mutant cells. Our findings may reflect differences in the degree of activation of these downstream pathways in different KRAS-mutant cells.

Interestingly, we found that the tight junction KEGG signalling pathway was most significantly activated in the KRAS-mutant subgroup. Indeed, we observed that the phosphorylation levels of tight junction scaffolding proteins, including OCLN, TJP, and PARD3, were significantly upregulated in KRAS-mutant cells. Previous studies have demonstrated that oncogenic Ras signalling is involved in deficiencies in tight junctions, leading to loss of apicobasal cell polarity and disruption of epithelial architecture.[49] The phosphorylation of PARD3 has been reported to change its affinity for its interaction partners, leading to loss of cell polarity and induction of migration.[50] The phosphorylation of OCLN has been implicated in tight junction disassembly and assembly.[51] Taken together, our data may provide phosphorylation sites that contribute to tight junction regulation. However, the function of these phosphorylation sites has not been functionally annotated in the literature. Therefore, further studies are needed to confirm whether these phosphorylation sites contribute to tight junction regulation.

To the best of our knowledge, this is the first time that we have identified protein–protein association networks that are dysregulated in KRAS-mutant cells by comparing coexpression data between KRAS-mutant cells and cells of 34 other cell lines. Previous studies have demonstrated that the high level of correlation using relative expression levels across samples allows the identification of protein complexes.[14] Consistent with this view, we confirmed that well-studied multiprotein complexes were enriched in the commonly occurring protein pairs (the bottom 10% of correlation coefficients of differential gene pairs). These results imply that our proteomics data were reliable for investigating the KRAS-specific protein pairs (top 10% of correlation coefficients of differential gene pairs), which reflect protein–protein association networks that are dysregulated in KRAS-mutant cells. We focused on two novel protein association pairs between PPP1R12A and FBLN2 and between COL4A1 and MRS2 in the focal adhesion pathway. A functional association between COL4A1 and FBLN2 was identified based on the STRING database. Fibulin-2 (FBLN2) is a secreted extracellular matrix glycoprotein that contributes to basement membrane (BM) stability. Previous studies have demonstrated that FBLN2 promotes tumour cell adherence to collagen in KRAS(G12D) and p53 mutation lung adenocarcinoma cell line models.[52] PPP1R12A, which is mainly involved in the Rho/ROCK signalling pathway, regulates focal adhesion formation by actomyosin contraction. Thus, the two novel protein associations mediated by the interaction between FBLN2 and COL4A1 may represent features of focal adhesion in KRAS-mutant CRC cells. Although further validation is needed to investigate whether these proteins functionally comprise complexes, our proteomics-based protein coexpression analysis correlation analysis may provide insight into the dysregulated protein–protein association landscape of each mutation.

To the best of our knowledge, this is the first analysis of cancer vulnerability prediction by a pairwise correlation between phosphorylation level and gene dependency across a large number of CRC cell lines. Our data showed that phosphorylation sites correlated with corresponding gene dependency could play an important role in cancer cell growth or viability, and these sites are worth further validation as potential drug targets.

We found that the phosphorylation levels of EGFR Y1172 and PTK2 Y861, which were upregulated in the BRAF/KRAS wt and KRAS-mutant subgroups, respectively, showed a strong correlation with the corresponding gene dependency scores. Previous studies have reported that both phosphorylation sites contribute to kinase activation. It has been experimentally validated that the phosphorylation level of EGFR Y1172 is decreased by treatment with cetuximab, an anti-EGFR inhibitor, in cetuximab-sensitive CRC cell lines.[53] The phosphorylation level of PTK2 Y861 is decreased by treatment with TAE226, a focal adhesion kinase inhibitor, in ovarian cancer cells.[54] Therefore, applying this correlation analysis to all quantified phosphorylation sites may lead to exploring potential therapeutic target phosphorylation sites.

We found that EPHA2 kinase is activated in KRAS-mutant CRC cells. Previous studies have shown that high phosphorylation levels of EPHA2 S897 and Y772 as well as the protein expression level of EPHA2 were upregulated in invasive CRC cell lines with KRAS mutation by western blotting of 14 CRC cell line models.[55] Our large-scale proteomics and phosphoproteomics data across 35 CRC cell lines quantified 18 phosphorylation sites for EPHA2, including 8 EPHA2 kinase substrate phosphorylation sites annotated in the PhosphoSitePlus database. We demonstrated consistent upregulation of EPHA2 protein expression and phosphorylation levels of EPHA2 kinase substrate sites in KRAS-mutant CRC cells compared to BRAF mutant and BRAF/KRAS wt cells.

Our data indicated that inhibition of direct phosphorylation sites of EPHA2 alone did not show cancer vulnerability in KRAS-mutant CRC cells. Previous studies have proposed that EPHA2 is a therapeutic target in diverse cancers, including CRC.[56–58] Additionally, several compounds targeting EPHA2 have been evaluated and tested in clinical studies with limited clinical success.[59] One possible interpretation for this finding is that EPHA2 exhibits both pro-oncogenic functions in a ligand-independent manner and anti-oncogenic functions in a ligand-dependent manner.[60] A recent report has shown that EPHA2 Y772 has both ligand-dependent and ligand-independent functions.[61] Due to its ligand-independent role, EPHA2 Y772 is responsible for cell proliferation and could be a target of the EPHA2 tyrosine kinase inhibitor ALW-II-41-27 in nasopharyngeal carcinoma (NPC).[62] However, our approach for predicting CRC cancer dependency from phosphoproteomic data demonstrated that all phosphorylation levels of EPHA2 kinase substrate sites, including Y772, showed an inverse or no correlation with EPHA2 gene dependency scores. Thus, we speculated that common downstream molecules mediated by EPHA2 and KRAS are direct therapeutic candidate targets in KRAS-mutant CRC.

We identified 1244 phosphorylation sites with a significantly positive correlation against corresponding gene dependency scores (p value < 0.05), indicating that these phosphorylation sites play roles in cancer cell vulnerability. One of the narrowed-down cancer vulnerability candidate phosphosites in KRAS-mutant CRC is phosphorylation at Y378 of PARD3, which is a cell polarity protein. A recent study demonstrated that PARD3 Tyr residues are a substrate of EPHA4 and modulate downstream signalling.[63] The EPHA2-EPHA4 interaction has been previously observed using BioID data.[64] Based on these data, PARD3 Y378 is also a potential downstream candidate of EPHA2. Although the phosphorylation site of Y378 on PARD3 has not been functionally annotated previously, aberrant tyrosine phosphorylation of Par3 is thought to be associated with dysfunction of growth factor receptors, leading to pathological tight junction assembly and disassembly in tumour cells.[65,66] Previous reports have shown that RAS activation disrupts epithelial homeostasis and leads to loss of tissue polarity.[67] Taken together, the results indicate that the phosphorylation of PARD3 Y378, which is a downstream component candidate of EPHA2, may contribute to loss of cell polarity in KRAS-mutant CRC cells. Taken together, our data suggest that the EPHA2-PARD3 axis is a cancer vulnerability in KRAS-mutant CRC. This study has one limitation. Previous studies demonstrated that up to 65% of phosphosites might be nonfunctional.[68] It will be important to assess whether PARD3 Y378 is a functional phosphorylation site involved in cell viability through experiments, including experiments involving phosphorylated amino acid substitutions.

We also demonstrated that the molecular features of KRAS mutant cell lines, including tight junctions and focal adhesion, were upregulated in KRAS mutant clinical samples based on phosphoproteomic data from the CPTAC colon cancer study. In contrast, we did not observe any upregulation of EPHA2 substrates or quantification of PARD3 Y378. One possible explanation for this result is the low number of pY quantifications in CPTAC data across samples. The quantified sample numbers of EPHA2 Y772 in the CPTAC data was 20 (20.8%), indicating that the limited number of sample quantifications of this phosphosite may not lead to reliable evaluation. It will be possible to increase the quantitative depth of the tyrosine phosphoproteome from clinical tissue samples by collecting high-freshness tissue samples and performing phospho-tyrosine enrichment.

Conclusion

In summary, our large-scale phosphoproteomics and proteomics data across 35 steady-state CRC cell lines represent a valuable resource for understanding the oncogenic signalling pathways and dysregulated protein complexes associated with different oncogenic mutations. Our integrated analysis of phosphorylation levels of phosphorylation sites and cancer dependency scores of corresponding proteins across CRC cell lines provides a systematic assessment of the importance of phosphorylation sites to cancer vulnerability. In the future, we envision that matching the profile of phosphorylation sites involved in cancer cell viability with phosphoproteomic data of CRC patients will become an important tool for predicting phosphorylation sites that play a central role in cancer viability.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

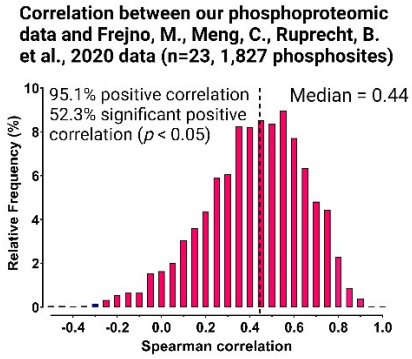
This work was supported by Japan Society for the Promotion of Science KAKENHI Grant Number 19H03530 awarded to J.A.

References

1 E. Dekker, P. J. Tanis, J. L. A. Vleugels, P. M. Kasi and M. B. Wallace, *Lancet*, 2019, **394**, 1467–1480.
2 A. B. Benson, A. P. Venook, M. M. Al-Hawary, L. Cederquist, Y. J. Chen, K. K. Ciombor, S. Cohen, H. S. Cooper, D. Deming, P. F. Engstrom, I. Garrido-Laguna, J. L. Grem, A. Grothey, H. S. Hochster, S. Hoffe, S. Hunt, A. Kamel, N. Kirilcuk, S. Krishnamurthi, W. A. Messersmith, J. Meyerhardt, E. D. Miller, M. F. Mulcahy, J. D. Murphy, S. Nurkin, L. Saltz, S. Sharma, D. Shibata, J. M. Skibber, C. T. Sofocleous, E. M. Stoffel, E. Stotsky-Himelfarb, C. G. Willett, E. Wuthrick, K. M. Gregory and D. A. Freedman-Cass, *J. Natl. Compr. Cancer Network*, 2018, **16**, 359–369.
3 K. K. Ciombor, J. H. Strickler, T. S. Bekaii-Saab and R. Yaeger, *J. Clin. Oncol.*, 2022, **40**, 2706–2715.
4 F. Pietrantonio, C. Cremolini, F. Petrelli, M. Di Bartolomeo, F. Loupakis, C. Maggi, C. Antoniotti, F. de Braud, A. Falcone and R. Iacovelli, *Crit. Rev. Oncol. Hematol.*, 2015, **96**, 156–166.
5 J. L. Bos, E. R. Fearon, S. R. Hamilton, M. Verlaan-de Vries, J. H. van Boom, A. J. van der Eb and B. Vogelstein, *Nature*, 1987, **327**, 293–297.
6 S. R. Punekar, V. Velcheti, B. G. Néel and K.-K. Wong, *Nat. Rev. Clin. Oncol.*, 2022, **19**, 637–655.
7 J. T. Henry, O. Coker, S. Chowdhury, J. P. Shen, V. K. Morris, A. Dasari, K. Raghav, M. Nusrat, B. Kee, C. Parseghian, S. Pant, N. Jeyakumar, L. Zhu, Y. Nishioka, D. Fogelman, R. A. Wolff, D. Hong, M. J. Overman, J. Vauthey, S. Kopetz and B. Johnson, *JCO Precis. Oncol.*, 2021, **5**, PO.20.00256.
8 M. G. Fakih, S. Kopetz, Y. Kuboki, T. W. Kim, P. N. Munster, J. C. Krauss, G. S. Falchook, S. W. Han, V. Heinemann, K. Muro, J. H. Strickler, D. S. Hong, C. S. Denlinger, G. Girotto, M. A. Lee, H. Henary, Q. Tran, J. K. Park, G. Ngarmchamnanrith, H. Prenen and T. J. Price, *Lancet Oncol.*, 2022, **23**, 115–124.
9 D. Tang, G. Kroemer and R. Kang, *Mol. Cancer*, 2021, **20**, 128.
10 A. Bardia, M. Gounder, J. Rodon, F. Janku, M. P. Lolkema, J. J. Stephenson, P. L. Bedard, M. Schuler, C. Sessa, P. LoRusso, M. Thomas, H. Maacke, H. Evans, Y. Sun and D. S. W. Tan, *Oncologist*, 2020, **25**, e160–e169.
11 Z. Liu, Y. Liu, L. Qian, S. Jiang, X. Gai, S. Ye, Y. Chen, X. Wang, L. Zhai, J. Xu, C. Pu, J. Li, F. He, M. Huang and M. Tan, *Mol. Cell*, 2021, **81**, 4076–4090.
12 R. Mosca, J. Tenorio-Laranga, R. Olivella, V. Alcalde, A. Céol, M. Soler-López and P. Aloy, *Nat. Methods*, 2015, **12**, 167–168.
13 X. Mo, Q. Niu, A. A. Ivanov, Y. H. Tsang, C. Tang, C. Shu, Q. Li, K. Qian, A. Wahafu, S. P. Doyle, D. Cicka, X. Yang, D. Fan, M. A. Reyna, L. A. D. Cooper, C. S. Moreno, W. Zhou, T. K. Owonikoko, S. Lonial, F. R. Khuri, Y. Du, S. S. Ramalingam, G. B. Mills and H. Fu, *Cell*, 2022, **185**, 1974–1985.
14 J. D. Lapek, P. Greninger, R. Morris, A. Amzallag, I. Pruteanu-Malinici, C. H. Benes and W. Haas, *Nat. Biotechnol.*, 2017, **35**, 983–989.
15 P. V. Hornbeck, B. Zhang, B. Murray, J. M. Kornhauser, V. Latham and E. Skrzypek, *Nucleic Acids Res.*, 2015, **43**, D512–20.
16 D. Ochoa, A. F. Jarnuczak, C. Viéitez, M. Gehre, M. Soucheray, A. Mateus, A. A. Kleefeldt, A. Hill, L. Garcia-Alonso, F. Stein, N. J. Krogan, M. M. Savitski, D. L. Swaney, J. A. Vizcaíno, K. M. Noh and P. Beltrao, *Nat. Biotechnol.*, 2020, **38**, 365–373.
17 K. B. Emdal, N. Palacio-Escat, C. Wigerup, A. Eguchi, H. Nilsson, D. B. Bekker-Jensen, L. Rönnstrand, J. U. Kazi, A. Puissant, R. Itzykson, J. Saez-Rodriguez, K. Masson, P. Blume-Jensen and J. V. Olsen, *Cell Rep.*, 2022, **40**, 111177.
18 C. Pacini, J. M. Dempster, I. Boyle, E. Gonçalves, H. Najgebauer, E. Karakoc, D. van der Meer, A. Barthorpe, H. Lightfoot, P. Jaaks, J. M. McFarland, M. J. Garnett, A. Tsherniak and F. Iorio, *Nat. Commun.*, 2021, **12**, 1–14.
19 E. Gonçalves, R. C. Poulos, Z. Cai, S. Barthorpe, S. S. Manda, N. Lucas, A. Beck, D. Bucio-Noble, M. Dausmann, C. Hall, M. Hecker, J. Koh, H. Lightfoot, S. Mahboob, I. Mali, J. Morris, L. Richardson, A. J. Seneviratne, R. Shepherd, E. Sykes, F. Thomas, S. Valentini, S. G. Williams, Y. Wu, D. Xavier, K. L. MacKenzie, P. G. Hains, B. Tully, P. J. Robinson, Q. Zhong, M. J. Garnett and R. R. Reddel, *Cancer Cell*, 2022, **40**, 835–849.
20 E. Medico, M. Russo, G. Picco, C. Cancelliere, E. Valtorta, G. Corti, M. Buscarino, C. Isella, S. Lamba, B. Martinoglio, S. Veronese, S. Siena, A. Sartore-Bianchi, M. Beccuti, M. Mottolese, M. Linnebacher, F. Cordero, F. Di Nicolantonio and A. Bardelli, *Nat. Commun.*, 2015, **6**, 1–10.
21 Y. Abe, H. Hirano, H. Shoji, A. Tada, J. Isoyama, A. Kakudo, D. Gunji, K. Honda, N. Boku, J. Adachi and T. Tomonaga, *Theranostics*, 2020, **10**, 2115–2129.
22 Y. Abe, M. Nagano, A. Tada, J. Adachi and T. Tomonaga, *J. Proteome Res.*, 2017, **16**, 1077–1086.
23 H. Mizuta, K. Okada, M. Araki, J. Adachi, A. Takemoto, J. Kutkowska, K. Maruyama, N. Yanagitani, T. Oh-hara, K. Watanabe, K. Tamai, L. Friboulet, K. Katayama, B. Ma, Y. Sasakura, Y. Sagae, M. Kukimoto-Niino, M. Shirouzu, S. Takagi, S. Simizu, M. Nishio, Y. Okuno, N. Fujita and R. Katayama, *Nat. Commun.*, 2021, **12**(1), 1261.
24 J. Cox and M. Mann, *Nat. Biotechnol.*, 2008, **26**, 1367–1372.
25 A. Bateman, *Nucleic Acids Res.*, 2019, **47**, D506–D515.
26 S.-H. Yu, P. Kyriakidou and J. Cox, *J. Proteome Res.*, 2020, **19**, 3945–3954.
27 S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann and J. Cox, *Nat. Methods*, 2016, **13**, 731–740.
28 M. Giurgiu, J. Reinhard, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone and A. Ruepp, *Nucleic Acids Res.*, 2019, **47**, D559–D563.
29 M. K and S. Goto, *Oncol. Lett.*, 2020, **19**, 3316–3332.
30 J. Wang, S. Vasaikar, Z. Shi, M. Greer and B. Zhang, *Nucleic Acids Res.*, 2017, **45**, W130–W137.

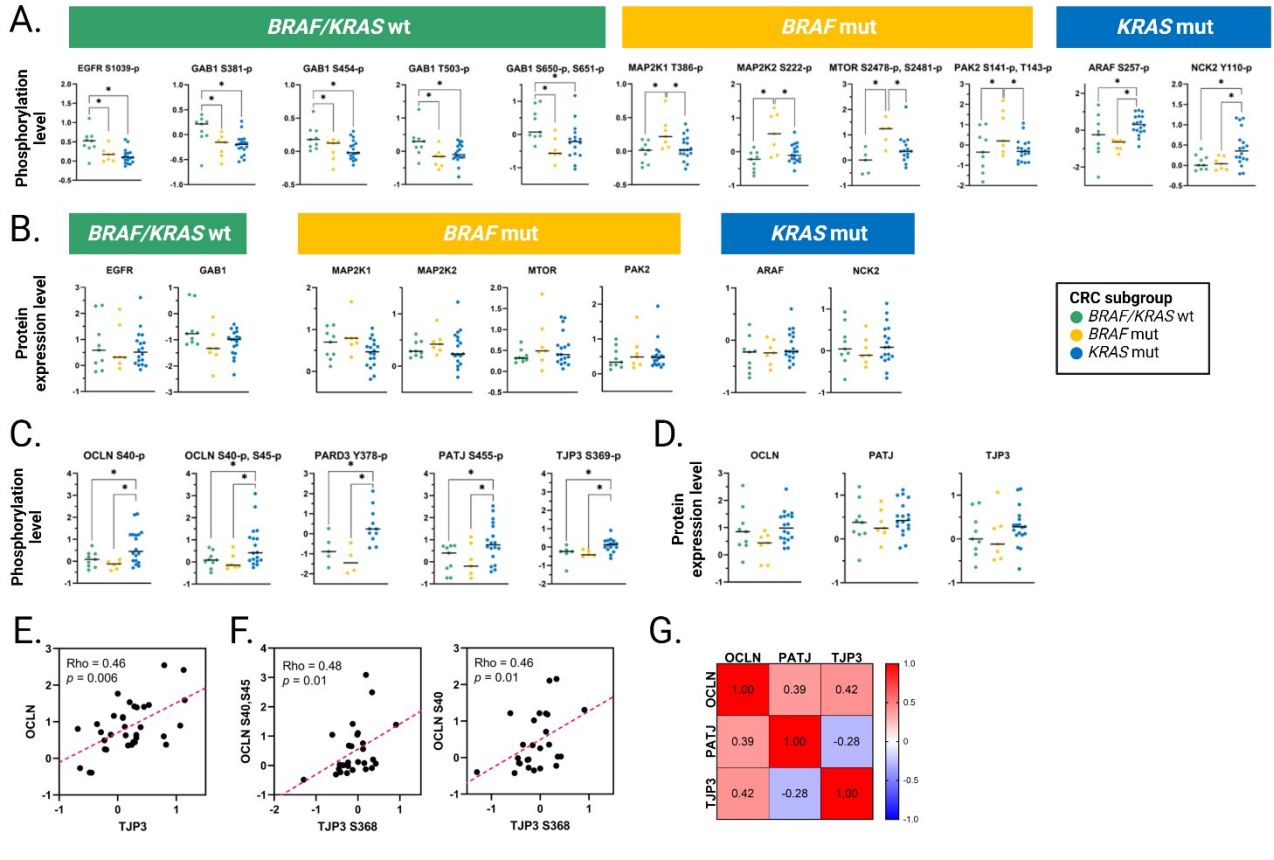
31 K. Krug, P. Mertins, B. Zhang, P. Hornbeck, R. Raju, R. Ahmad, M. Szucs, F. Mundt, D. Forestier, J. Jane-Valbuena, H. Keshishian, M. A. Gillette, P. Tamayo, J. P. Mesirov, J. D. Jaffe, S. A. Carr and D. R. Mani, *Mol. Cell. Proteomics*, 2019, **18**, 576–593.
32 M. Reich, T. Liefeld, J. Gould, J. Lerner, P. Tamayo and J. P. Mesirov, *Nat. Genet.*, 2006, **38**, 500–501.
33 J. Diedenhofen and B. Musch, *PLoS One*, 2015, **10**, e0131499.
34 B. Seashore-Ludlow, M. G. Rees, J. H. Cheah, M. Coko, E. V. Price, M. E. Coletti, V. Jones, N. E. Bodycombe, C. K. Soule, J. Gould, B. Alexander, A. Li, P. Montgomery, M. J. Wawer, N. Kuru, J. D. Kotz, C. Suk-Yee Hon, B. Munoz, T. Liefeld, V. Dančik, J. A. Bittker, M. Palmer, J. E. Bradner, A. F. Shamji, P. A. Clemons and S. L. Schreiber, *Cancer Discov.*, 2015, **5**, 1210–1223.
35 S. Okuda, Y. Watanabe, Y. Moriya, S. Kawano, T. Yamamoto, M. Matsumoto, T. Takami, D. Kobayashi, N. Araki, A. C. Yoshizawa, T. Tabata, N. Sugiyama, S. Goto and Y. Ishihama, *Nucleic Acids Res.*, 2017, **45**, D1107–D1111.
36 T. I. Roumeliotis, S. P. Williams, E. Gonçalves, C. Alsinet, M. Del Castillo Velasco-Herrera, N. Aben, F. Z. Ghavidel, M. Michaut, M. Schubert, S. Price, J. C. Wright, L. Yu, M. Yang, R. Dienstmann, J. Guinney, P. Beltrao, A. Brazma, M. Pardo, O. Stegle, D. J. Adams, L. Wessels, J. Saez-Rodriguez, U. McDermott and J. S. Choudhary, *Cell Rep.*, 2017, **20**, 2201–2214.
37 M. Frejno, C. Meng, B. Ruprecht, T. Oellerich, S. Scheich, K. Kleigrewe, E. Drecoll, P. Samaras, A. Hogrebe, D. Helm, J. Mergner, J. Zecha, S. Heinzlmeir, M. Wilhelm, J. Dorn, H.-M. Kvasnicka, H. Serve, W. Weichert and B. Kuster, *Nat. Commun.*, 2020, **11**, 3639.
38 M. Morkel, P. Riemer, H. Bläker and C. Sers, *Oncotarget*, 2015, **6**, 20785–20800.
39 J. S. Boehm, M. J. Garnett, D. J. Adams, H. E. Francies, T. R. Golub, W. C. Hahn, F. Iorio, J. M. McFarland, L. Parts and F. Vazquez, *Nature*, 2021, **589**, 514–516.
40 D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez, T. Doerks, M. Stark, J. Muller, P. Bork, L. J. Jensen and C. Von Mering, *Nucleic Acids Res.*, 2011, **39**, 561–568.
41 R. Chen, A. Sugiyama, H. Seno and M. Sugimoto, *PLoS One*, 2019, **14**, e0221772.
42 J. Auwercx, P. Rybarczyk, P. Kischel, I. Dhennin-Duthille, D. Chatelain, H. Sevestre, I. Van Seuningen, H. Ouadid-Ahidouch, N. Jonckheere and M. Gautier, *Nutrients*, 2021, **13**(1), 210.
43 C. Pacini, J. M. Dempster, I. Boyle, E. Gonçalves, H. Najgebauer, E. Karakoc, D. van der Meer, A. Barthorpe, H. Lightfoot, P. Jaaks, J. M. McFarland, M. J. Garnett, A. Tsherniak and F. Iorio, *Nat. Commun.*, 2021, **12**, 1661.
44 B. A. Aksoy, V. Dancík, K. Smith, J. N. Mazerik, Z. Ji, B. Gross, O. Nikolova, N. Jaber, A. Califano, S. L. Schreiber, D. S. Gerhard, L. C. Hermida, S. Jagu, C. Sander, A. Floratos and P. A. Clemons, *Database*, 2017, 1–10.
45 S. Vasaikar, C. Huang, X. Wang, V. A. Petyuk, S. R. Savage, B. Wen, Y. Dou, Y. Zhang, Z. Shi, O. A. Arshad, M. A. Gritsenko, L. J. Zimmerman, J. E. McDermott, T. R. Clauss, R. J. Moore, R. Zhao, M. E. Monroe, Y.-T. Wang, M. C. Chambers, R. J. C. Slebos, K. S. Lau, Q. Mo, L. Ding, M. Ellis, M. Thiagarajan, C. R. Kinsinger, H. Rodriguez, R. D. Smith, K. D. Rodland, D. C. Liebler, T. Liu and B. Zhang, *Cell*, 2019, **177**, 1035–1049.
46 T. S. Niault and M. Baccarini, *Carcinogenesis*, 2010, **31**, 1165–1174.
47 A. Ahmadzadeh, S. Shahrabi, K. Jaseb, F. Norozi, M. Shahjahani, T. Vosoughi, S. Hajizamani and N. Saki, *Oncol. Rev.*, 2014, **8**, 22–25.
48 M. B. Ryan and R. B. Corcoran, *Nat. Rev. Clin. Oncol.*, 2018, **15**, 709–720.
49 T. A. Tervonen, S. M. Pant, D. Belitškin, J. I. Englund, K. Närhi, C. Haglund, P. E. Kovanen, E. W. Verschuren and J. Klefström, *Cancer Res.*, 2021, **81**, 1513–1527.
50 X. Chen and I. G. Macara, *Nat. Cell Biol.*, 2005, **7**, 262–269.
51 C. M. Van Itallie and J. M. Anderson, *Tissue barriers*, 2018, **6**, e1382671.
52 B. N. Baird, M. J. Schliekelman, Y.-H. Ahn, Y. Chen, J. D. Roybal, B. J. Gill, D. K. Mishra, B. Erez, M. O'Reilly, Y. Yang, M. Patel, X. Liu, N. Thilaganathan, I. V. Larina, M. E. Dickinson, J. L. West, D. L. Gibbons, D. D. Liu, M. P. Kim, J. M. Hicks, I. I. Wistuba, S. M. Hanash and J. M. Kurie, *PLoS One*, 2013, **8**, e67054.
53 Y. Abe, M. Nagano, T. Kuga, A. Tada, J. Isoyama, J. Adachi and T. Tomonaga, *Sci. Rep.*, 2017, **7**, 1–12.
54 J. Halder, Y. G. Lin, W. M. Merritt, W. A. Spannuth, A. M. Nick, T. Honda, A. A. Kamat, L. Y. Han, T. J. Kim, C. Lu, A. M. Tari, W. Bornmann, A. Fernandez, G. Lopez-Berestein and A. K. Sood, *Cancer Res.*, 2007, **67**, 10976–10983.
55 P. D. Dunne, S. Dasgupta, J. K. Blayney, D. G. McArt, K. L. Redmond, J.-A. Weir, C. A. Bradley, T. Sasazuki, S. Shirasawa, T. Wang, S. Srivastava, C. W. Ong, K. Arthur, M. Salto-Tellez, R. H. Wilson, P. G. Johnston and S. Van Schaeybroeck, *Clin. cancer Res. an Off. J. Am. Assoc. Cancer Res.*, 2016, **22**, 230–242.
56 G. Martini, C. Cardone, P. P. Vitiello, V. Belli, S. Napolitano, T. Troiani, D. Ciardiello, C. M. Della Corte, F. Morgillo, N. Matrone, V. Sforza, G. Papaccio, V. Desiderio, M. C. Paul, V. Moreno-Viedma, N. Normanno, A. M. Rachiglio, V. Tirino, E. Maiello, T. P. Latiano, D. Rizzi, G. Signoriello, M. Sibilia, F. Ciardiello and E. Martinelli, *Mol. Cancer Ther.*, 2019, **18**, 845–855.
57 K. R. Amato, S. Wang, A. K. Hastings, V. M. Youngblood, P. R. Santapuram, H. Chen, J. M. Cates, D. C. Colvin, F. Ye, D. M. Brantley-Sieders, R. S. Cook, L. Tan, N. S. Gray and J. Chen, *J. Clin. Invest.*, 2014, **124**, 2037–2049.
58 P. D. Dunne, S. Dasgupta, J. K. Blayney, D. G. McArt, K. L. Redmond, J. A. Weir, C. A. Bradley, T. Sasazuki, S. Shirasawa, T. Wang, S. Srivastava, C. W. Ong, K. Arthur, M. Salto-Tellez, R. H. Wilson, P. G. Johnston and S. Van Schaeybroeck, *Clin. Cancer Res.*, 2016, **22**, 230–242.
59 T. Xiao, Y. Xiao, W. Wang, Y. Y. Tang, Z. Xiao and M. Su, *J. Hematol. Oncol.*, 2020, **13**, 1–17.

60 H. Miao, D.-Q. Li, A. Mukherjee, H. Guo, A. Petty, J. Cutter, J. P. Basilion, J. Sedor, J. Wu, D. Danielpour, A. E. Sloan, M. L. Cohen and B. Wang, *Cancer Cell*, 2009, **16**, 9–20.
61 M. Locard-Paulet, L. Lim, G. Veluscek, K. McMahon, J. Sinclair, A. van Weverwijk, J. D. Worboys, Y. Yuan, C. M. Isacke and C. Jørgensen, *Sci. Signaling*, 2016, **9**, ra15.
62 Y.-P. Xiang, T. Xiao, Q.-G. Li, S.-S. Lu, W. Zhu, Y.-Y. Liu, J.-Y. Qiu, Z.-H. Song, W. Huang, H. Yi, Y.-Y. Tang and Z.-Q. Xiao, *Cell Death Dis.*, 2020, **11**, 709.
63 S. L. Banerjee, F. Lessard, F. J. M. Chartier, K. Jacquet, A. I. Osornio-Hernandez, V. Teyssier, K. Ghani, N. Lavoie, J. N. Lavoie, M. Caruso, P. Laprise, S. Elowe, J. P. Lambert and N. Bisson, *Cell Rep.*, 2022, **40**(1), 111031.
64 K. Salokas, X. Liu, T. Öhman, I. Chowdhury, L. Gawriyski, S. Keskitalo and M. Varjosalo, *EMBO Rep.*, 2022, **23** 1–28.
65 Y. Wang, D. Du, L. Fang, G. Yang, C. Zhang, R. Zeng, A. Ullrich, F. Lottspeich and Z. Chen, *EMBO J.*, 2006, **25**, 5058–5070.
66 C. Jakubowski, N. B. Collins, E. A. Sugar, M. Berg, H. Cao, M. Giannakis, E. M. Jaffee and N. S. Azad, *J. Clin. Oncol.*, 2020, **38**, TPS4114.
67 K. Magudia, A. Lahoz and A. Hall, *J. Cell Biol.*, 2012, **198**, 185–194.
68 C. R. Landry, E. D. Levy and S. W. Michnick, *Trends Genet.*, 2009, **25**, 193–197.



Supplemental Figure 1. Comparison between our phosphoproteomics data and data from a previously reported large-scale phosphoproteomics study of CRC cell lines, related to Figure 1.

Spearman's correlation between our phosphoproteomics data and data from Frejno, M., Meng, C., Ruprecht B, et al. (2020). The 1,827 phosphosites across 23 CRC cell lines shared in both datasets were used in this study.



Supplemental Figure 2. Boxplots and correlation matrix of proteins and phosphorylation sites in the enriched pathway that showed differential expression between the CRC subgroups, related to Figure 2.

A. Beeswarm plots of the 11 phosphosites of the ErbB signalling pathway showing differential regulation among CRC subgroups. The horizontal line represents the median.

B. Beeswarm plots of the 8 proteins of the ErbB signalling pathway among CRC subgroups. The horizontal line represents the median.

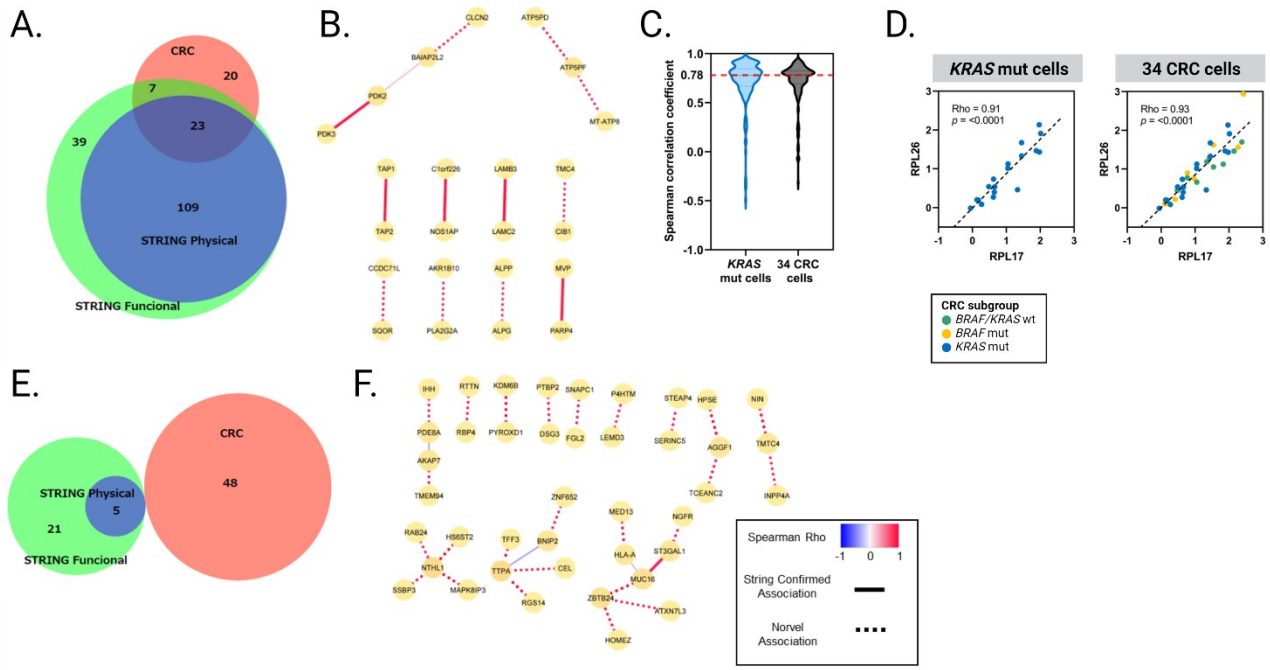
C. Beeswarm plots of the 5 phosphorylation sites of tight junction signalling showing differential regulation among CRC subgroups. The horizontal line represents the median.

D. Beeswarm plots of the 3 protein components of tight junction signalling among CRC subgroups. The horizontal line represents the median.

E. Correlation plot of the protein expression levels of TJP3 and OCLM.

F. Correlation plot of the phosphorylation levels of TJP3 S368 and OCLM S40 and S45.

G. Correlation matrix of the proteome expression levels of the key components of tight junctions in KRAS-mut cell lines (Spearman's test).



Supplemental Figure 3. Protein coregulation analysis for the identification of protein–protein associations in KRAS-mutant cells, related to Figure 3.

A. Venn diagram of the commonly occurring protein pairs (the bottom 10% of protein–protein interactions of the correlation difference) among STRING physical dataset, STING functional dataset and our CRC dataset.

B. Protein–protein association network of the commonly occurring protein pairs (bottom 10%), related to Fig. 3B.

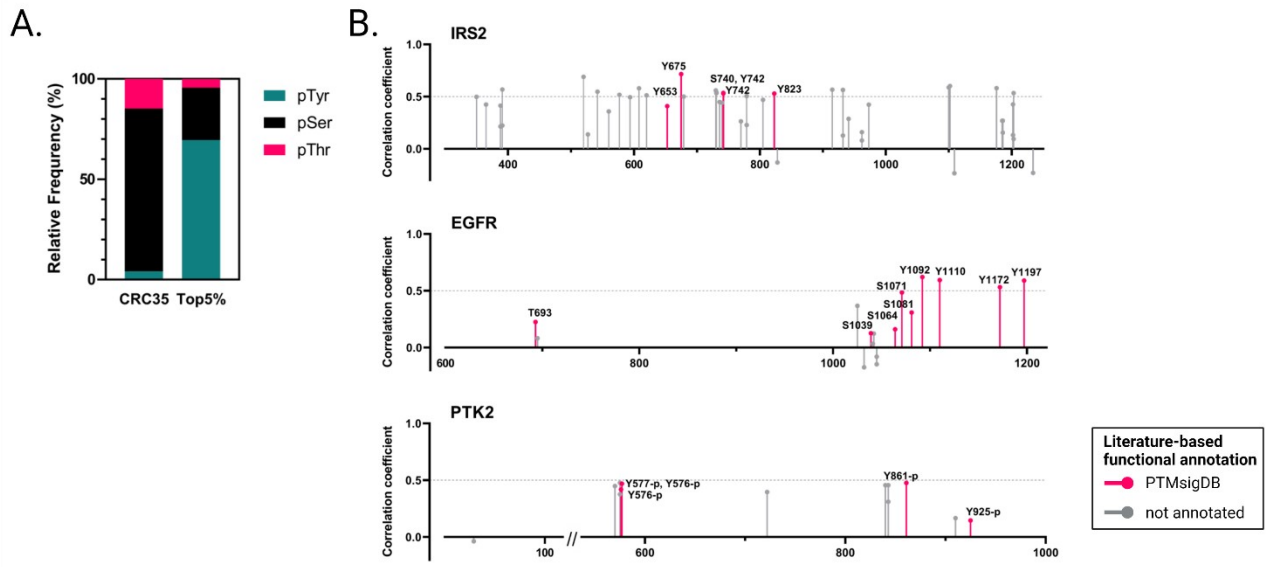
C. Violin plot of the Spearman's correlation coefficient of functional protein–protein associations based on the STRING database between KRAS-mutant cells and 34 CRC cell types.

D. Correlation plot of the protein expression levels of RPL17 and RPL26 in KRAS-mutant cells and 34 other CRC cell lines.

E. Venn diagram of the KRAS-specific protein pairs (the top 10% of protein–protein interactions

of the correlation difference) among the STRING physical dataset, STING functional dataset and our CRC dataset.

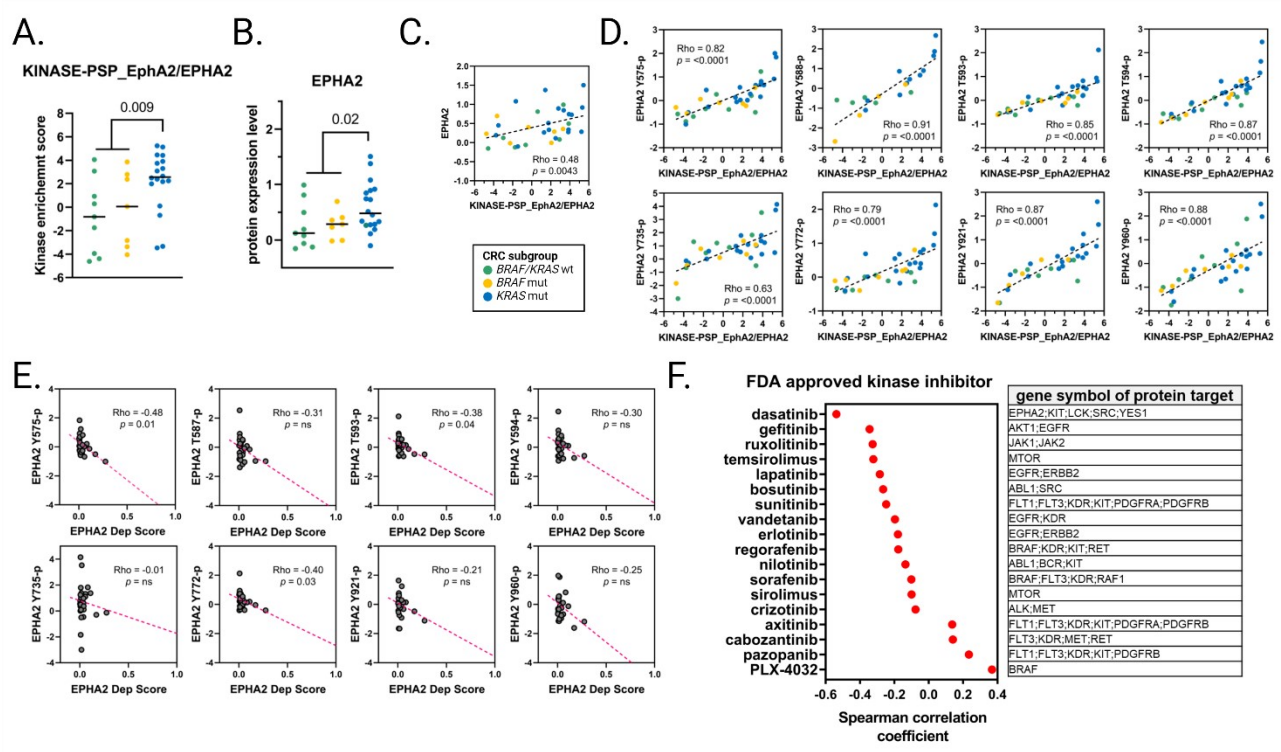
F. Protein–protein association network of KRAS-specific protein pairs (top 10%) of the correlation difference, related to Fig. 3D.



Supplemental Figure 4. Features of phosphorylation sites showing a high correlation between phosphorylation levels and gene dependency scores, related to Figure 4.

A. Distribution comparing the relative frequency of phosphoserine (pSer, black), phosphothreonine (pThr, red), and phosphotyrosine (pTyr, green) in the top 5% highly correlated phosphosites with gene dependency scores with that in all quantified phosphosites in CRC35.

B. Spearman's correlation coefficient of phosphorylation sites identified in IRS2 (top), EGFR (middle) and PTK2 (down) with unknown (grey) and known activation in PTMsigDB (deep pink).



Supplemental Figure 5. Potential therapeutic opportunity for KRAS-mutant cells, related to Figure 5.

A. Beeswarm plots of EPHA2 kinase enrichment scores showing differential enrichment between the CRC subgroups (p value = 0.009, Welch's t test between KRAS mut vs. (BRAF mut and BRAF/KRAS wt)).

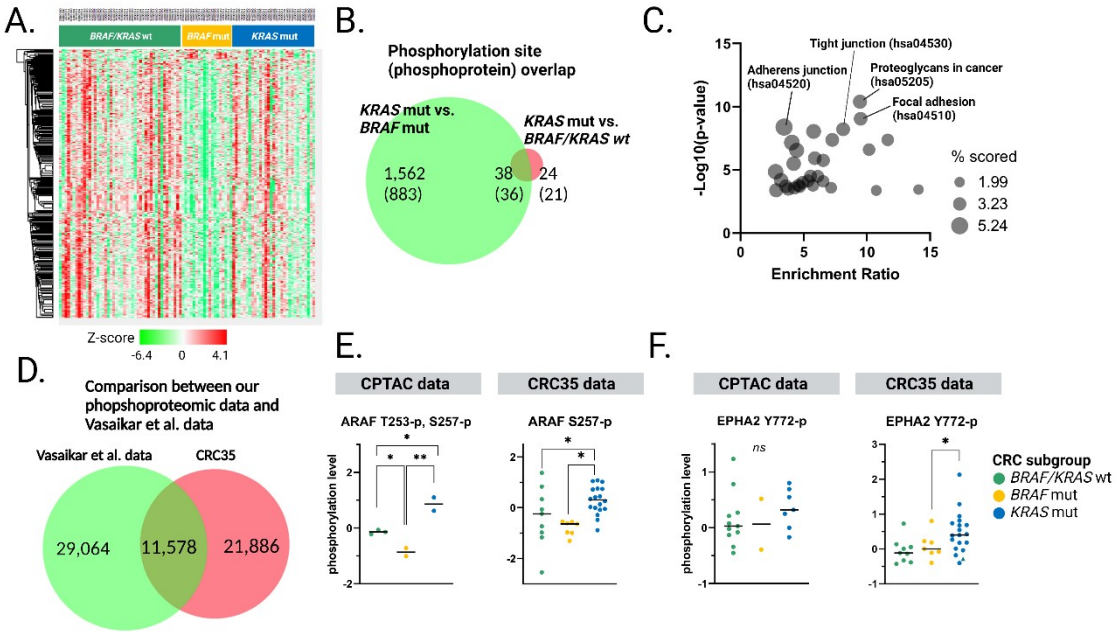
B. Beeswarm plots of EPHA2 protein expression levels showing differential expression between the CRC subgroups (p value = 0.02, Welch's t test between KRAS mut vs. BRAF mut and BRAF/KRAS wt).

C. Correlation plots between EPHA2 kinase enrichment scores and protein expression levels of EPHA2.

D. Correlation plots between EPHA2 kinase enrichment scores and phosphorylation levels of substrates on EPHA2.

E. Correlation plots between EPHA2 dependency scores and phosphorylation levels of substrates on EPHA2.

F. List of Spearman's correlation coefficients between phosphorylation levels of PARD3 Y378 and CTD2 AUCs of FDA-approved kinase inhibitors. Gene symbols of the protein targets for each inhibitor are shown in the right table.



Supplemental Figure 6. Comparison of the phosphoproteomic data of CRC35 with data from the CPTAC Colon Cancer Study

A. Hierarchical clustering of 4,103 significantly regulated phosphorylation sites according to the KRAS and BRAF mutational status (p value < 0.05, FDR < 0.1, one-way ANOVA followed by Tukey's post hoc test).

B. Venn diagram of significantly upregulated phosphorylation sites (phosphoproteins) in the KRAS mut vs. BRAF mut or KRAS mut vs. BRAF/KRAS wt subgroups.

C. Bubble plot of the KEGG pathway enrichment of differentially expressed genes (DEGs) in the KRAS mut group. The bubble size corresponds to the % scored in the pathway. The % score indicates the ratio of the number of DEGs mapped to a certain pathway to the total number of genes mapped to the pathway.

D. Venn diagram demonstrating the overlap between in-house CRC35 phosphoproteomic data and CPTAC phosphoproteomic data (Vasaikar, et al. 2019).

E. and F. Beeswarm plots of the phosphorylation levels of ARAF-T253, S257 (E), and EPHA2-Y772 (F) showing differential regulation among the various KRAS and BRAF mutational status subgroups in the CPTAC data (left) and CRC35 data (right). The horizontal line represents the median.