Unraveling biological processes and EGFR pathway regulation by the protein tyrosine phosphatase PTPRH in non-small cell lung cancer

Mylena M. O. Ortiz[1,2], Deeya M. Patel[3], Matthew Swiatnicki[2], and Eran Andrechek[2,4]

1 – Genetics and Genomic Sciences Program

2 - Department of Physiology

3 - Lyman Briggs College

4 – Contributing Author

Eran R. Andrechek

Department of Physiology

Michigan State University

2194 BPS Building

567 Wilson Road

East Lansing, MI

48824

andrech1@msu.edu

(517)884-5020

Running title: PTPRH participation in cell signaling and the EGFR pathway

Keywords: tyrosine phosphatase, lung cancer, EGFR, cell signaling, cancer biology, molecular biology

Abstract

The delicate balance of protein phosphorylation is often disrupted in cancers, with hyperactivity of kinases and inactivation of phosphatases driving cell proliferation and survival pathways. PTPRH, a protein tyrosine phosphatase, is deregulated or mutated in certain cancers, including non-small cell lung cancer (NSCLC). However, the biological processes that PTPRH is involved in and how they may contribute to tumorigenesis are unknown. Previous studies have linked PTPRH to the regulation of the EGFR pathway but the full extension of this regulation and the underlying mechanisms remain to be elucidated. We found that PTPRH knockout tends to increase the phosphorylation levels of EGFR at the tyrosine residue 1173 (1197), whereas overexpression of PTPRH wild-type significantly decreases phosphorylation in one of the NSCLC cell lines. Surprisingly, disruption of the extracellular or intracellular domains of the phosphatase leads to EGFR dephosphorylation in a similar manner. Co-immunoprecipitation and proximity-dependent biotinylation (BioID) experiments demonstrated that PTPRH do not directly interact with EGFR, but rather with CDK5RAP3 and EIF2AK2, two proteins reported to regulate the downstream pathway NF-κB. In addition, BioID revealed 55 novel PTPRH interactors in NSCLC cells, which are enrichedin translation and RNA-associated pathways. Guided by RNA sequencing analysis, we observed that overexpression of the phosphatase downregulates multiple oncogenic signature pathways and modulates the gene expression of 34 protein tyrosine phosphatases and 45 tyrosine kinases, EGFR included. Together, these results shed light on the importance of PTPRH in regulating biological and cellular processes and how its inactivation may support cancer progression.

Introduction

Phosphorylation and dephosphorylation are key post-translational modifications that intricately regulate cell signaling cascades[1]. In normal cells, these two events are tightly regulated in time and space by kinases and phosphatases and, together, they orchestrate fundamental cellular processes, such as proliferation, differentiation, and apoptosis[2,3]. However, this delicate equilibrium is often disrupted in multiple diseases, notably cancers[4]. The hyperactivation of kinases and inactivation of phosphatases leads to aberrant cell proliferation, migration, and survival, thereby supporting cancer initiation, progression, and resistance to therapies[5].

In non-small cell lung cancers (NSCLC), activating mutations in the receptor tyrosine kinase (RTK) *EGFR* are observed in approximately 50% of patients from Asian inheritance and 20% in Caucasians[6]. The vast majority are in-frame deletions, point mutations, or insertions in the first few exons of the EGFR kinase domain (exons 18-21), such as the classical L858R substitution and exon 19 deletion[7–9]. These mutations result in ligand-independent activation of the EGFR pathway by stabilizing EGFR dimers in an active conformation[10,11]. Subsequential transphosphorylation of tyrosine residues in the C-terminal tail of EGFR recruits signaling molecules and activates the downstream pathways including PI3K/mTOR/Akt, RAS/RAF/MEK/ERK, and IL-6/JAK/STAT3[12–14]. Therefore, these mutations are generally responsive to EGFR inhibitors (EGFRi) - when not combined with resistance mutations - due to hyperactivity of the RTK. In addition to *EGFR*, other oncogenic kinases are also found to be frequently mutated in NSCLC, including *PIK3CA* (~13%), *SKT11* (~12%), *MET* (~7%), *ALK* (~7%), *BRAF* (~6%), *HER2* (~6%), *ROS* (~5%), *RET* (~4.5%), *AKT1* (~2%), and *MAP2K1* (~1%)[15–17]. Many of these kinases activate similar downstream pathways to EGFR and are actionable drug targets[18].

Mutations in phosphatases and their consequences, on the other hand, are much less well understood. While kinases generally act as oncogenes, phosphatases can exhibit characteristics of either tumor suppressors or oncogenes in lung cancers[19]. A well-known example is *PTEN*, an important tumor-suppressor gene mutated in approximately 7% of NSCLC cases. The inactivation of this phosphatase leads to an increase in the PI3K/mTOR/Akt signaling pathway and lower survival rates[20,21]. Conversely, activating mutations in the tyrosine phosphatase *PTPN11*/Shp2 increases phospho-ERK 1/2 levels and activates the MAPK pathway, thus

behaving as an oncogene[22]. More recently, our group has shown that point mutations in the receptor-like protein tyrosine phosphatase (RPTP) *PTPRH* are observed in 6% of NSCLC patients and over 80% of MMTV-*PyMT* mice, a murine model of breast cancer[23,24]. A mutation that results in inactivation of this phosphatase leads to elevated phosphorylation of EGFR at tyrosine 1173 (alternatively numbered as 1197) and results in activation of the Akt pathway[23–25]. This suggests that PTPRH contributes to negatively regulating the EGFR pathway; nonetheless, the precise underlying mechanism is unclear.

PTPRH, also known as SAP-1, is a transmembrane protein and part of the tyrosine phosphatase family, which encompasses over 100 members[26]. Structurally, PTPRH is composed of an extracellular region rich in fibronectin type 3 (FN3) repeats, a hydrophobic transmembrane domain, and a single intracellular phosphatase domain[27,28]. Little is known about PTPRH kinetics and interactors; however, a few associations with tyrosine kinase signaling and cellular processes have been found in different cell types. By dephosphorylating tyrosine residues, PTPRH is suggested to negatively regulate T-cell function[29], intestinal immunity[30], integrin-associated pathways[31], and, more recently, EGFR signaling[23–25,32]. Furthermore, PTPRH has been found to induce apoptotic cell death pathways[33] and to stimulate c-Src activity[34].

The participation of this phosphatase in cancer, although not fully elucidated, seems to be largely related to the tissue of origin. PTPRH expression varies across and within different cancer types and conflicting findings are observed[27,35–37]. For instance, in colorectal cancers, PTPRH plays a causal role in cancer progression, but not initiation, and overexpression is observed in some cases[29,38]. On the other hand, as hepatocarcinoma advances to poorly differentiated stages, a decrease in the phosphatase level is observed[39]. In lung cancers, *PTPRH* mutations happen in ~6% of the patients, and *PTPRH* and *EGFR* mutations were found to be mutually exclusive[23]. Therefore, it is evident that we do not understand the full spectrum of cell signaling regulation by PTPRH and how it stimulates or suppresses tumorigenesis. To bridge this gap, we built a comprehensive map of PTPRH interactors and associated cell signaling networks. We started to investigate the mechanisms by which PTPRH regulates the EGFR pathway by assessing the physical interaction between the two proteins and the ability of different PTPRH constructs to dephosphorylate the kinase in NSCLC cell lines. Following this, we expanded our search to cover all proteins and signaling pathways regulated by the phosphatase. To achieve this, we used

a proximity-dependent biotinylation assay (BioID), tandem mass spectrometry (MS/MS), and RNA sequencing. Our study revealed novel PTPRH interactors and functions, shedding light on the role of this phosphatase in the cellular context of lung cancers.

Results

Investigation of EGFR phosphorylation status in lung cancer cells in the absence of PTPRH

As PTPRH is a protein tyrosine phosphatase and published observations suggest that it regulates EGFR signaling[23–25], we therefore investigated how the inactivation of PTPRH affects EGFR phosphorylation status in lung cancer. We hypothesized that PTPRH negatively regulates EGFR; thus, phospho-EGFR levels (pEGFR) would increase in the absence of the phosphatase. To confirm this relationship, we used our previously generated human lung adenocarcinoma cell line NCI-H23 *PTPRH* knockout (H23 KO), whose parental cell line is wild-type for *PTPRH* and *EGFR*. As previous data from our lab suggests that PTPRH affects phosphorylation of the EGFR tyrosine residue 1173 (pEGFR 1173, also known as pEGFR 1197), we first screened pEGFR 1173 levels in two knockout clones following EGF stimulation for 15 min. Although densitometry of the western blot bands revealed approximately a two-fold increase in one of the clones (H23 KO#13) when pEGFR 1173 was corrected by the total EGFR levels, none of them presented a significant change compared to the parental cells although a trend was evident (Supplemental Fig. 1A). Interestingly, reduction of total EGFR was observed in H23 KO#2 cells, whereas in H23 KO#13 clone this was not as evident.

Since EGFR phosphorylation and activation happen rapidly after ligand binding and dimerization[40,41], we reasoned that an increase in pEGFR1173 in H23 KO cells could reach a maximum at other time points following EGF stimulation. To test this, we stimulated the parental and knockout cells with EGF for 5, 15, 30, 60, and 120 minutes. As demonstrated in Fig. 1A, the 15-minute time point presents the biggest difference in pEGFR1173/total EGFR ratios between the two groups. However, concordant with our initial observations, none of the time points were statistically significant although an increasing trend was still evident. Concomitantly, we also evaluated the activation of AKT, a pathway downstream of EGFR. Following EGF stimulation, phospho-AKT increases gradually and reaches maximum activation in both groups around the 30-minute time-point, but no difference in pAKT/total AKT ratios was seen between the H23 and H23 KO groups within the same time point.

We then began to consider whether PTPRH might dephosphorylate a tyrosine residue other than 1173. To investigate potential differences in phosphorylation at other EGFR tyrosine residues

between the groups, we immunoprecipitated EGFR from both H23 and H23 KO cells stimulated with EGF for 15 minutes, and quantified phospho-tyrosine residues using tandem mass spectrometry (MS/MS). Our analysis revealed that two tyrosine residues, Y1086 and Y1173, were phosphorylated in most replicates from both groups (Fig. 1B). Consistent with our western blot findings, H23 KO cells exhibited an approximately two-fold increase in pEGFR1173 levels; however, the differences in phosphorylation levels of Y1086 and Y1173 did not quite reach significance. Notably, phosphorylation of Y992 and Y1068 was detected exclusively in the knockout cells, but only in 1 out of 3 replicates. Together, these results demonstrate a trend to increase pEGFR 1173 levels in lung adenocarcinoma cells when PTPRH is absent.

Investigation of EGFR phosphorylation status in lung cancer cells overexpressing PTPRH wild-type or various deletions

Considering that we did not see significant changes in pEGFR when using the *PTPRH* knockout cells, we reasoned that one explanation could be the low PTPRH protein levels in the parental cells (Supplemental Fig. 2A-B). To address this issue, we took the opposite strategy and stably overexpressed PTPRH wild-type containing a C-terminal HA tag in the H23 KO cells. In addition to WT, we also generated different groups of cells overexpressing various *PTPRH* in-frame deletions in different domains of the protein. The purpose of this approach was to investigate how these deletions altered PTPRH regulation of EGFR phosphorylation status. Since the extracellular portion of PTPRH is composed of multiple fibronectin 3 (FN3) domains, we started by deleting blocks of two consecutive FN3 (constructs Δ1-3) using site-directed mutagenesis. The intracellular catalytic site of the phosphatase domain was deleted in the Δ4 construct. A schematic of the deletions is presented in Fig. 2A. We were unable to generate an *in-frame* deletion of the last FN3 domain due to the repetitive sequences situated in the extracellular portion of the phosphatase. To validate the successful generation of our constructs, we sequenced their *PTPRH* regions and aligned the sequences against the human *PTPRH* wild-type coding sequence. All constructs contained in-frame deletions in the desired region (Supplemental Fig. 3B). We also noticed the presence of a synonymous mutation across all constructs (c.717 G>C). Moreover, as a consequence of the site-directed mutagenesis method, a mutation was introduced in the c.1157 C>A position of the Δ1 construct, resulting in a P386H substitution despite multiple attempts to revert it to the wild-type. Immunoblot using antibodies against PTPRH and the HA

tag revealed similar banding patterns between the two blots, but different sizes of bands across the addback groups, confirming the overexpression on the various PTPRH deletions (Fig. 2B).

Given that we demonstrated maximal phosphorylation at 15 minutes after EGF stimulation in Figure 1, we used this time point to collect lysates from cells expressing the various PTPRH constructs. As anticipated, while the H23 KO group tended to increase pEGFR 1173 compared to H23 cells, the difference was not statistically significant. However, as expected upon overexpression of PTPRH WT addback in the PTPRH knockout line, a significant decrease of approximately 2-fold was observed (Fig. 2B). The pEGFR levels in the various deletion addbacks were then compared to theparental cells. Surprisingly, all PTPRH deletions yielded similar results to the wild-type addback, even the one lacking the catalytic site. Therefore,it seems that neither the FN3 domains nor the catalytic site of PTPRH are necessary for EGFR Y1173 dephosphorylation.

We then investigated the effect of PTPRH WT overexpression in a different lung adenocarcinoma cell line, NCI-H2023. Curiously we observed an opposite trend: total EGFR expression decreases following overexpression and EGF stimulation, whereas phosphorylation of EGFR at Y1173 is sustained. Consequently, pEGFR 1173 levels seem to be increased after normalization by total EGFR. Noteworthy, only a few replicates were used (n=2); hence, no strong conclusions can be drawn. Taken together, these findings suggest that PTPRH affects EGFR levels and phosphorylation in a cell-specific manner. Further investigation is required to fully understand the underlying mechanism for the differences herein observed.

Evaluation of protein-protein interaction between PTPRH and EGFR

To further clarify whether PTPRH directly interacts with EGFR and regulates tyrosine kinase signaling, we conducted a protein-protein interaction experiment using co-immunoprecipitation (Co-IP). For this purpose, we stimulated H23 parental cells with EGF for 15 minutes, followed by immunoprecipitation of EGFR and immunoblot for EGFR and PTPRH. H23 KO cells were used as a negative Co-IP control, whereas whole-cell lysates (WCL) from the colorectal adenocarcinoma cell line HT-29, known to express PTPRH at elevated levels[37], were used as a positive control for immunoblotting. As demonstrated in Fig. 3A (top panel), a substantial amount of EGFR was recovered from all cell lines, however, PTPRH co-immunoprecipitation was not observed (bottom panel).

To overcome the difficulties associated with low PTPRH protein levels in the H23 cells or the possibility of a transient interaction between the two proteins, we repeated the experiment using our previously generated cell line in which wild-type PTPRH was added back to the H23 KO cells and overexpressed. Protein lysates were collected at 2, 5, 10, and 15 min post-EGF stimulation. Despite varying EGF stimulation, PTPRH was not observed in the immunoprecipitated fraction (Fig. 3B). Together, these results demonstrate that PTPRH does not directly interact with EGFR in any of the time points in a manner detectable by Co-IP. This raises several possibilities, including that either PTPRH-EGFR interactions may happen rapidly following EGFR phosphorylation or that other proteins or phosphatases are involved in EGFR dephosphorylation.

BioID reveals novel PTPRH interactors in lung cancer cells

Based on the results that EGFR levels are modulated by an unknown mechanismwhen PTPRH is overexpressed or knocked out, we next wanted to establish this mechanistic association between the two. Considering that little is known about PTPRH interaction partners and how this phosphatase may regulate different biological processes, we explored those interactions through a comprehensive protein-protein interaction assay. We used BioID2 for this purpose, an *in-vitro* labeling technique that adds a biotin tag to proteins interacting stably or transiently with the protein of interest in a labeling radius of ~10 nm[42]. BioID2 followed by an HA tag was cloned into the C-terminal of the PTPRH wild-type coding sequence, in proximity to its intracellular phosphatase domain. Two different lung adenocarcinoma cell lines (NCI-H2023 and H23 KO) were infected with lentiviruses containing wild-type PTPRH-BioID2 or BioID2-only, the latter being a control to assess non-specific biotinylation. To validate the expression of the constructs, an immunoblot against the HA tag was performed. All the cell lines infected with BioID2-only or wild-type PTPRH-BioID2 presented bands at the expected molecular weight, while no bands were seen in the mock lanes (Fig. 4A). We then performed immunofluorescence using antibodies against HA (Fig. 4B) and BioID2 (Supplemental Fig. 4) to evaluate whether the BioID2 tag would affect PTPRH subcellular localization. While BioID2 is highly expressed in the cytoplasm and nucleus of the BioID2-only cells, the modified PTPRH was found to be expressed at lower levels in the plasma membrane and cytoplasm, in agreement with the subcellular localization of the endogenous counterpart.

We hypothesized that PTPRH regulates different cell signaling pathways, including those emanating from other tyrosine kinases. To begin to test this, all groups of cells were serum-starved for 24 hours, followed by FBS stimulation with a complete medium supplemented with biotin for 18 hours. Immunoblot against biotin in the WCL revealed promiscuous biotinylation in the cells infected with BioID2-only (Fig. 5A). Importantly, some biotinylated proteins showed up exclusively in the wild-type PTPRH-BioID2 group, suggesting that these proteins are specifically interacting with PTPRH. Biotinylation was not observed in the mock group, validating the activity of the BioID2 enzyme and the experimental results.

Biotinylated proteins were immunoprecipitated, followed by their identification using MS/MS. When analyzed individually, H2023 and H23 KO cells expressing wild-type PTPRH-BioID2 revealed 10 and 42 candidate interactors, respectively (Fig. 5C and 5D, and Supplemental Table 1). When analyzed in combination, 12 candidate interactors achieved significant values. These proteins exhibited above 2-fold increase in wild-type PTPRH-BioID2 cells compared to their corresponding BioID2-only cells and $p < 0.05$, composing a total of 55 PTPRH candidate interactors. In agreement with a previous report that employed BioID to screen the phosphatase PTP1B interactors, few proteins overlapped between the two cell lines, highlighting that protein-protein interactions are fluid and depend on the cell type and intrinsic cell signaling (Fig. 5D)[43]. The top 3 potential interacting proteins were PTPRH itself (130,000-fold change, $p < 0.0001$), followed by CDK5RAP3 (760-fold change, $p = 0.00012$), and the protein kinase EIF2AK2 (2.1-fold change, $p = 0.00047$). Considering that receptor tyrosine kinase phosphatases, PTPRH included, are thought to form dimers, this data suggests that PTPRH maybe found as a homodimer in our infected cells[34,44]. Interestingly, both CDK5RAP3 and EIF2AK2 regulate the NF-κB pathway, a transcription factor downstream of the EGFR pathway[45–48]. Although not significant, NFKB2 and the NF-κB activating protein NKAP presented a fold-change of 7.9 ($p = 0.096$) and 11 ($p = 0.2$), respectively (Supplemental Table 1), suggesting that PTPRH may act on this pathway. Furthermore, THBS-1, which mediates cell-cell and cell-matrix interactions, and the myosins MYO1C and MYO1B, which are actin motors, were significantly enriched in our analysis, in agreement with PTPRH participation in the actin-based cytoskeleton processes[31]. BRAP, which negatively regulates the activation of the MAP kinase, was also listed as a potential interactor[49].

Since the proposed role for PTPRH is to dephosphorylate tyrosine residues, we then evaluated how many proteins from this list are reported to have at least one phospho-site in the PhosphoSitePlus database. 83.6% (n=46) of PTPRH candidate interactors are reported to have at least one phospho-tyrosine site and 100% (n=55) present at least one phospho-serine or phospho-threonine site (Fig. 5E). We then sought to understand whether the candidate proteins interact with each other and form complexes of biological relevance. STRING analysis revealed several interacting proteins forming one major network and two small ones (Fig. 5B and Supplemental Fig. 7). To elucidate the biological processes associated with these proteins, we conducted a DAVID functional annotation analysis and subsequently reported annotations with a significant threshold of $p < 0.05$ after multi-hypothesis correction (Table 1). Despite our initial hypothesis that PTPRH regulates tyrosine kinase pathways, we found an enrichment of biological processes related to translation, RNA binding, helicases, and ATP binding/ATPase activity. Post-translational modifications related to phosphoproteins were also significant as expected; however, we have also observed the enrichment of acetylation and isopeptide bond. In summary, these results highlight PTPRH's participation in novel biological pathways, raising the possibility that PTPRH also accumulates non-canonical roles.

Transcriptome analysis of lung cancer cells overexpressing PTPRH

To further validate our BioID2 results and gain a deeper understanding of PTPRH participation in biological pathways and cancer, we generated two NSCLC cell lines overexpressing wild-type PTPRH (H23 PTPRH overexpression and H2023 PTPRH overexpression), followed by RNA-sequencing. We also included the parental cells (endogenous PTPRH expression) and H23 KO in our analysis. We followed similar conditions to the BioID2 experiment: cells were serum starved for 24 hours and stimulated with FBS in complete medium for 16 hours before RNA extraction. Principal component analysis (PCA) of the normalized reads shows that cells cluster primarily based on the PTPRH status (PC1- 82% variance): parental cells cluster on the right, while PTPRH overexpressing cell lines on the left (Fig. 6A). PC2 reflects 9% of the variance and further separate the clusters based on the transcriptome of the original cell type (H2023 vs H23 derived cells). Surprisingly, the H23 KO cells cluster closely to the H23 PTPRH overexpression cells, which is also evidenced by our clustering analysis (Supplemental Fig. 5A and

Supplemental Fig. 6). We believe that this reflects a drift of phenotype caused by the ablation of PTPRH after a few passages.

Differential expression analysis demonstrated that 1,360 genes present over a 2-fold increase (upregulated genes) and 2,060 genes present over a 2-fold decrease (downregulated genes) in a significant manner in PTPRH overexpressing cells compared to parental cells (Fig. 6B, Supplemental Fig. 5C, and Supplemental Table 2). These represent PTPRH signature genes and the top 30 genes up and downregulated are shown in Supplemental Fig. 6. The most upregulated genes are SERPINB2 (13.76 fold increase), a protein shown to inhibit migration and to increase the resolution of inflammation[50], the retrotransposon-derived gene RTL1 (13.25 fold increase), the transcription factor ERG (12.87 fold increase), and cyclin D2 (12.21 fold increase). Among the most downregulated genes are the cadherin CDH17 (19.9 fold decrease), AGR2 (19.38 fold decrease), which participates in protein folding and has been shown to be involved in cancer progression[51], and CYP4F11, a member of the cytochrome P450 family.

To understand how the differentially expressed genes (DEG) modulated by PTPRH are involved in biological pathways and cancer, we performed pathway enrichment analyses using GO and Molecular Signature (MSigDB) databases. Our gene set enrichment analysis (GSEA) using the oncogenic signature C6 revealed multiple significantly downregulated pathways in the cell lines overexpressing PTPRH, including NFE2L2, MEK, LEF1, E2F1, KRAS, RAF, P53, ERBB2, and others (Fig. 7Aand 7C). We have also observed significant downregulation of genes involved in the EGFR (NES = - 1.55) and AKT (NES = -1.53) pathways, in agreement with our hypothesis and previous findings that PTPRH negatively regulates EGFR and downstream signaling pathways (Fig. 6D and Supplemental Fig. 5B). The hallmark MSigDB showed downregulation of gene sets involved in estrogen response, interferon alpha and gamma responses, apical junction, reactive oxygen species, IL6-JAK-STAT3 signaling, and metabolism (adipogenesis, xenobiotic, and fatty acid metabolism), and upregulation of proliferation pathways (E2F targets, G2M checkpoint) and the WNT-Beta-catenin signaling (Fig. 7B and 7D). Concordant with or BioID results, pathway enrichment analysis using GO biological processes gene sets showed significant upregulation of RNA-associated pathways and ribonucleoprotein biogenesis (Table 2). Furthermore, DNA repair, chromosome organization, and chromatin remodeling pathways

were also positively enriched, whereas fatty acid and monocarboxylic acid metabolism were shown to be downregulated.

We next asked whether PTPRH overexpression would modulate the expression of other protein tyrosine phosphatases (PTPs) and tyrosine kinases (TKs). We interrogated our RNA-seq data based on the genes that were present in our DEG analysis and found that 34 PTPs (Fig. 6C) and 45 TKs (Fig. 6D) were differentially expressed in the PTPRH overexpressing cell lines. Overall, 24 PTPs were downregulated, suggesting redundancy in function among the PTPs. The PTPs showing the strongest downregulation were PTPN20, PTPRQ, PTPRVP, PTPRJ, PTPRM, PTPRN, and PTPRU, whereas PTPRD, PTPRR, PTPRH itself, PTPN2, and PTPRB were the most upregulated ones. Interestingly, the majority of TKs presented increased gene expression levels (n = 28) and some were found to be downregulated (n= 17), including EGFR. Among the TKs with the strongest upregulation are FLT3, EPHB1, KIT, ROR2, NTRK1, PDGFRA, EPHB6, and EPHA7. On the other hand, MST1R, EPHA8, ERBB3, LMTK3, TEK, ROR1, FGFR1, and PDGFRB were the most downregulated TKs. Collectively, these results reinforce that PTPRH modulates not only TKs but also PTPs, and participates in multiple oncogenic pathways and biological processes which were previously unknown.

Discussion

Little is known about the role of the receptor-like protein tyrosine phosphatase PTPRH in the regulation of various cell signaling pathways and how alterations may contribute to tumorigenesis. To address this gap of knowledge, we investigated PTPRH's candidate interactors and the participation of this phosphatase in biological and cellular processes in two NSCLC cell lines, in addition to its influence on the EGFR pathway. We were able to find novel protein-protein interactions and cell signaling pathways in which PTPRH is involved and a connection between EGFR signaling and PTPRH was demonstrated in our RNA-sequencing results and pEGFR screening. Moreover, we report for the first time that this protein tyrosine phosphatase is involved in RNA binding and processing, ribosome biogenesis, metabolism, and cell division-related pathways.

By assessing pEGFR sites in the absence of PTPRH or overexpression of various PTPRH constructs, our findings suggest that PTPRH may regulate EGFR phosphorylation status at tyrosine 1173 (1197) position but this regulation is not direct, as demonstrated by our Co-IP and BioID experiments. This regulation also seems to vary between lung adenocarcinoma cell lines. We observed that, although not significant, H23 KO cells tend to have higher pEGFR 1173 (1197), whereas rescuing wild-type PTPRH expression in H23 KO cells significantly decreases the phosphorylation of EGFR tyrosine 1173 (1197). Concordantly, MS/MS revealed no consistent changes in phosphorylation in EGFR tyrosine sites other than 1173 (1197) across PTPRH knockout replicates. Moreover, no significant changes in phospho-S473 AKT levels were observed in the knockouts; however, our GSEA analysis suggests downregulation of this pathway when PTPRH levels are increased. In our previous published results, we demonstrated that MMTV-*PyMT* mice harboring the conserved mutation V483M in the *Ptprh* gene present high pEGFR 1173 (1197) levels and consequent activation of the AKT pathway[24]. Similarly, PTPRH knockout in NSCLC NCI-H23 cells significantly increases pEGFR 1173 (1197) levels with a consequent increase in growth and proliferation. Nonetheless, the phosphorylation status of S473 AKT varied across the clones[23]. These conflicting findings are possibly explained by variations in research models (mouse vs human)[52], differences in mutation types (point mutations vs frameshift and indels)[53], and compensation mechanisms by other PTPs[54]. As we have used PTPRH knockout cell lines previously generated by our group, we believe that after several

passages, the knockout lines quickly drifted from the original phenotype and other phosphatases started to compensate for the ablation of PTPRH activity. Our RNA-sequencing results corroborate this theory, as we have observed that the H23 PTPRH knockout cells cluster surprisingly closely to the H23 PTPRH overexpressed cells, suggesting similar transcriptome profiles. In fact, studies in mice and *Drosophila* revealed that ablation of phosphatase genes leads to mild or no change in phenotype, likely due to redundancy in tyrosine phosphatase functions[38,54,55]. Additionally, we have also observed the up and downregulation of several other PTPs when PTPRH is overexpressed. Some of them have been described as EGFR regulators, including *PTPN2*[56], *PTPN6* (SHP-1)[57,58], *PTPN11* (SHP-2)[59], *PTPN12*[60], *PTPRK*[61,62], and *PTPRJ*[63,64]. Interestingly, PTPRJ dephosphorylates EGFR at the same tyrosine residue as PTPRH in a pH-dependent manner[63]. As PTPRJ is strongly downregulated in our PTPRH overexpressed cell lines, we believe that these two phosphatases have redundant roles.

In addition to our group, EGFR regulation by PTPRH has been reported by Yao and collaborators[25] and Nishiyama *et al*[65]. EGFR was first identified as a novel PTPRH interactor through a wide mammalian-membrane two-hybrid assay using 48 cloned mammalian receptor tyrosine kinases and 108 protein tyrosine phosphatases[25]. This interaction was further validated by PTPRH and EGFR co-immunoprecipitation in HEK293 cells, suggesting that, in embryonic kidney cells, the interaction between the two proteins is stable. Furthermore, in the HER2-positive breast cancer cell line SKBR-3, PTPRH was shown to be only partly involved in EGFR dephosphorylation, raising the hypothesis that more than one phosphatase is in charge of controlling inactivation of the EGFR pathway[65]. PTPRH sequestering by the connector enhancer of kinase suppressor of Ras1 (CNKSR1) prevents EGFR dephosphorylation in the SKBR-3 cells and the association between these two proteins was shown to be mediated by the intracellular phosphatase domain. Considering the mechanistic differences in the various cell lines, we cannot rule out the possibility that PTPRH interacts with different proteins depending on the cellular context. This is particularly true in this case, as we did not see a stable or transient interaction between PTPRH and EGFR in any of the NSCLC cell lines tested herein. Of note, the labeling radius of the BioID2 molecule is ~10 nm[42]; thus if EGFR is slightly outside our PTPRH-BioID2 construct range, no signal will be detected. This finding in combination with our observation that deletion of PTPRH's phosphatase domain decreases pEGFR 1173 (1197) levels at comparable levels as wild-type PTPRH overexpression leads us to hypothesize that PTPRH acts in an

intermediate pathway rather than being the primary responsible for EGFR dephosphorylation. This is the case of PTPRO, for example, which negatively regulates EGFR through direct interaction with SRC[66]. Since we observed a strong interaction between PTPRH and NFKB regulators in our BioID results (CDK5RAP3 and EIF2AK2) and NF-κB is activated downstream of the EGFR pathway through the IKK complex[67,68], this raises the possibility that PTPRH modulates EGFR through NF-κB associated proteins. This agrees with the finding that there is a crosstalk between NF-κB and EGFR signaling and that transcription of NF-κB target genes influences activation of EGFR-dependent pathways[69]. Importantly, gene expression of EGFR and EGFR-related pathways were also downregulated significantly in our RNA-seq results, suggesting that PTPRH also negatively regulates EGFR at the RNA level. Further studies are needed to clarify where in time and space PTPRH fits into the proposed mechanism.

To shed light on protein regulation by PTPRH and its interaction partners, we built a comprehensive protein-protein interaction network based on stable and transient interactions revealed by BioID. In agreement with previous studies, the two NSCLC cell lines revealed some level of variation between them, suggesting that the phosphatase behaves differently according to cellular context. In total, we found55 candidate interactors and the strongest hit wasPTPRH itselfPTPRH and other receptor protein tyrosine phosphatases form homodimers and it is speculated that this is a mechanism of inactivation of the phosphatase activity[34,44]. Hence, PTPRH dimers may be present in our cell lines or PTPRH monomers are biotinylating themselves. Even though some of the interacting proteins herein reported are mostly found in the nucleus, there is also evidence of the location of these proteins in the cytosol or the mebrane[70–72]. Since we did not observe the presence of the phosphatase in the nucleus of our transduced cells but rather in the cytosol and plasma membrane, we anticipate that the interaction is mediated in these cellular compartments. However, the Human Protein Atlas reports the presence of PTPRH nuclear speckles, hence, we cannot rule out PTPRH trafficking to the nucleus[73].

Functional and pathway enrichment analysis of interacting proteins and DEG demonstrated for the first time that PTPRH participates in novel cellular and biological processes in NSCLC cell lines. Pathways linked to RNA (binding, processing, and splicing) and ribosomal proteins (translation and biogenesis) were enriched at both the protein and RNA levels. In addition, GSEA analysis revealed the downregulation of several oncogenic signatures and hallmark pathways in

NSCLC cells overexpressing PTPRH, establishing a link for how this phosphatase may contribute to lung cancer progression. These include NFE2L2, KRAS-associated pathways (MEK, RAF), LEF1, p53, BMI1, ERBB2, IL6-JAK-STAT3 signaling axis, interferon response, IL-21, reactive oxygen species, estrogen response, and metabolism of fatty acids. Curiously, these cells show upregulation of DNA repair and chromatin remodeling pathways, Wnt signaling, besides processes linked to cellular proliferation, such as E2F1 targets, G2M checkpoint, and chromosome organization. A few studies were published in the early 2000s implying that PTPRH participates in some other cellular processes. Overexpression of the phosphatase in Chinese hamster ovary cells disrupted cellular responses mediated by integrins (cell growth and motility) by dephosphorylation of focal adhesion-associated proteins[31]. In Jukart T cells, PTPRH inhibits the T cell antigen receptor (TCR) activity by interacting with the kinase Lck[29]. Moreover, overexpression of PTPRH in NIH 3T3 fibroblasts was shown to induce apoptosis mediated by the inactivation of the PI3K-AKT-ILK pathway and activation of cell death pathways dependent on caspases[33]. Even though in our study we did not observe PTPRH interaction with integrins or caspases-associated molecules, we found interaction with proteins involved in the actin-cytoskeleton and adhesion, such as THBS1, MYO1B, and MYO1C. Furthermore, our pathway enrichment analysis revealed the downregulation of several pathways controlling proliferation, AKT, apical junction, and some immune-related pathways as mentioned above. All these findings highlight differences in functionality according to the cell type and condition

When gene expression of tyrosine kinases were analyzedafter PTPRH overexpression, , we observed that several of thempresented significant changes. There is strong upregulation of FLT3, EPHB1, KIT, ROR2, NTRK1, PDGFRA, EPHB6, and EPHA7, and downregulation of MST1R, EPHA8, ERBB3, LMTK3, TEK, ROR1, FGFR1, and PDGFRB, besides EGFR. This could also be a consequence of the variety of PTPs that are perturbed concomitantly. In addition, we had previously reported that the phosphorylated levels of FGFR1 increase upon ablation of PTPRH in H23 cells[23].

In summary, we found that PTPRH indirectly regulates the EGFR pathway at both the RNA and protein levels and is primarily involved in RNA and translation-associated pathways in NSCLC cells. Moreover, overexpression of the phosphatase negatively modulated several oncogenic

signature pathways and genes, shedding light on how this phosphatase may contribute to lung cancer progression.

Experimental Procedures

Cell culture

Cells were acquired from ATCC and H23 PTPRH knockout cells were generated as described previously[23]. NCI-H23 cells and derived lineages were cultured in RPMI medium (Corning #50-020-PC) supplemented with 10% fetal bovine serum, 1% Penicillin-Streptomycin, 23.8 mM sodium bicarbonate, 25 mM HEPES, and 1 mM sodium pyruvate. NCI-H2023 and derived lineages were cultured in DMEM-F12/HEPES medium (Gibco #11330032) supplemented with 5% fetal bovine serum (FBS), 1% Penicillin-Streptomycin, 30nM sodium selenite, 10 nM hydrocortisone, 10 nM beta-estradiol, 0.01 mg/mL transferrin, 0.005 mg/mL insulin, and 2 mM L-glutamine. HT-29 cells were cultured in McCoy's 5a (modified) medium (Gibco #16600082) supplemented with 10% FBS and 1% Penicillin-Streptomycin. 293T cells were cultured in DMEM high glucose medium (Sigma-Aldrich #56436C) supplemented with 10% FBS and 1% Penicillin-Streptomycin. Cells were incubated at 37°C and 5% CO_2 and split using 0.25% Trypsin (Gibco #15090046) when 80% confluency was reached.

EGF stimulation

Recombinant human EGF (Gibco #PHG0311) was added to the medium of NCI-H23, H23 KO, and H2023 cells at 80% confluency to reach a final concentration of 100 ng/mL. Cells were incubated at 37°C and 5% CO_2 for 2, 5, 10, 15, 30, 60, or 120 minutes. Cells were washed twice with ice-cold HEPES- Buffered Saline (HBS) to stop stimulation followed by cell lysis to collect proteins.

Cell lysis and protein collection

After washing twice with ice-cold HBS, cells were scrapped from the dish and resuspended in ice-cold HBS. The cell suspension was centrifuged at 1,000 x g for 5 minutes at 4°C, followed by lysis in ice with ice-cold Pierce RIPA buffer (immunoblot and LC/MS/MS experiments) (Thermo Scientific #89901) or IP lysis/wash buffer (co-immunoprecipitation experiments) (Pierce Classic Magnetic IP/Co-IP kit #88804) + Halt Protease and Phosphatase inhibitor (Thermo Scientific #78441) for 20 minutes. Cell extracts were centrifuged at 16,000 x g for 20 minutes at 4°C and the supernatant containing the protein fraction was collected and stored at -

80°C until use. Protein quantification was performed using Pierce BCA protein assay (Thermo Scientific #23225) according to the manufacturer's instructions.

Immunoblotting and Densitometry

Protein samples were prepared using Laemmli SDS reducing sample buffer (Thermo Scientific #J61337-AC), followed by boiling in the dry bath at 95°C for 10 minutes. Samples were loaded in an 8% SDS-PAGE gel and electrophoresis was conducted at 100 V in Tris/Glycine/SDS buffer (BioRad #1610732). Following, samples were transferred to a PVDF Immobilon-P membrane (Millipore #IPVH00010) at 100 V for 1,5 hours at 4°C in ice-cold Tris/Glycine buffer containing 20% methanol. Membranes were blocked in 5% non-fat dry milk (diluted in TBS-Tween 0.1%) for one hour at room temperature (RT) under agitation and washed 3 times for 5 minutes with TBS-Tween 0.1% in the shaker. Primary antibodies were diluted in TBS-Tween 0.1% and 5% bovine serum albumin (BSA) and incubated with the membrane overnight at 4°C under agitation. The following dilutions were used for the primary antibodies: PTPRH 1:1000 (Abcam #ab231767), pEGFR 1173 1:500 (Cell Signaling #4407), EGFR 1:1000 (Cell Signaling #4267), B-actin 1:1000 (Cell Signaling #4967), HA Tag 1:1000 (Cell Signaling #3724), anti-Biotin 1:1000 (Cell Signaling #5597), BioID2 1:20000 (BioFront Technologies #BID2-CP-100), pAKT S473 1:2000 (Cell Signaling #4060), AKT 1:1000 (Cell Signaling #4685), and Vinculin 1:1000 (Cell Signaling #13901). Following 3 washes of 5 minutes with TBS-Tween 0.1% at RT, membranes were incubated for one hour and under agitation with fluorescent IRDye® 800CW goat anti-rabbit (LICOR #926-32211) or donkey anti-chicken (LICOR #926-32218) secondary antibodies diluted 1:10000 in TBS-Tween 0.1% and 5% BSA. Unbound antibodies were removed by washing the membranes three times, 5 minutes/each, with TBS-Tween 0.1%. Images were taken using the LICOR Odyssey M machine.

Densitometry was performed using ImageJ 1.54d software. First, the area selection size was set based on the sample with the largest band for each target protein (e.g. pEGFR 1173 and EGFR) and the loading control (i.e. B-actin). The grey mean value (X) of the band and background values for all samples were then measured and recorded. After the measurements were inverted using the 255-X formula, the background of each sample was subtracted from their band values. The resulting values for the target proteins were then divided by B-actin for data normalization.

Finally, the normalized phospho-protein values were divided by the normalized total protein values of each sample to obtain the phospho-photein/total protein ratios.

Immunoprecipitation and Co-immunoprecipitation

Protein lysates containing the same amount of protein from EGF-stimulated cells were mixed with EGFR monoclonal antibody (Cell Signaling #4267, 1:50 dilution) and incubated overnight at 4°C using end-to-end rotation. 50 mg of Dynabeads Protein A (Thermo Scientific #10006D) was added to the antibody-antigen conjugate and incubated at RT for 30 min using rotation. Samples were placed in a magnetic rack and the supernatant was discarded. The magnetic bead-antibody-antigen complex was washed and eluted as recommended by the manufacturer. The immunoprecipitated fraction was then analyzed for co-immunoprecipitation of EGFR and PTPRH by immunoblotting as described previously or sent to LC/MS/MS for screening of phospho-tyrosine sites.

Mass Spectrometry

LC/MS/MS was performed at the MSU Mass Spectrometry Core. On-bead antibody-bound proteins were first washed 3 times with 50 mM ammonium bicarbonate and digested with trypsin (5 ng/uL, incubation at 37°C for 6 hours). The solution was acidified using 1% trifluoroacetic acid and centrifuged at 14000 x g. After supernatant collection, the peptides were concentrated by solid phase extraction using StageTips[74] and dried by vacuum centrifugation. An injection of 10 uL of dried peptides resuspended in 2% acetonitrile and 0.1% trifluoroacetic acid was conducted using the Thermo Scientific EASY-nLC 1200 system and 0.1mm x 20 mm Acclaim™ PepMap™ C18 RSLC trapping columns (Thermo Scientific), followed by a 5min wash with buffer A (99.9% Water/0.1% Formic Acid). Peptides were eluted for 35 minutes onto a 0.075 mm x 500 mm Acclaim™ PepMap™ RSLC resolving column (Thermo Scientific) using a liner gradient ranging from 5-28% of buffer B (80% Acetonitrile/0.1% Formic Acid/19.9% Water) for 24 minutes. Next, column wash was performed using 90% buffer B for 10 minutes (constant flow rate = 300nl/min). Column temperature was kept constant at 50°C with the aid of an integrated column oven- PRSO-V2 (Sonation GmbH). Using a FlexSpray ion source, the peptides were then sprayed into the Q-Exactive mass spectrometer (Thermo Scientific) and survey scans of 45000 resolution (based on m/z 200) were collected using Orbi trap. The predominant 15 ions from each survey underwent an automatic higher energy collision-induced

association (HCD) and the resulting fragment spectra were collected using a resolution of 7500. These spectra were then transformed into a peaks list by the software Mascot Distiller v2.8.3. Using the Mascot[2] searching algorithm v 2.8.0.1, the peaks list was compared to a human protein database containing all sequences from UniProt and supplemented with frequent laboratory contaminants (downloaded from www.thegpm.org, cRAP project). Parameters were set to allow up to 2 missed tryptic sites, fixed modification of carbamidomethyl cysteine, variable modification of oxidation of Methionine and/or phosphorylation of Tyrosine, MS/MS tolerance of 0.02 Da, and peptide tolerance +/- 10 ppm. Following, the output list was analyzed using Scaffold v5.3.0. Proteins identified with 99% protein threshold, 95% peptide threshold, and 2 minimum peptide number were considered true. Quantitative analysis was performed based on normalized total precursor intensity and missing values were replaced by imputation (1,000 or 10,000).

Plasmids

pCMV-VSV-G was a gift from Bob Weinberg (Addgene plasmid # 8454), pCMV delta R8.2 was a gift from Didier Trono (Addgene plasmid # 12263), and MCS-BioID2-HA pBabe-puro was a gift from Kyle Roux (Addgene plasmid #120308). The lentiviral PTPRH WT- HA plasmid was prebuilt by Addgene. Briefly, the lentiCas9-Blast plasmid (Addgene plasmid # 52962), a gift from Feng Zhang, was modified to replace the Cas9 sequence with the human *PTPRH* coding sequence (CCDS 33110.1) and a C-terminal HA tag. A map of the generated plasmid is provided in the Supplemental Figure 3A.

Bacterial Transformation and Plasmid Prep

DH5-α competent cells were transformed by heat shock. 100 uL of ice-thawed cells were added to a centrifuge tube containing ~5 ng of purified plasmid. The mixture was incubated for 30 minutes in ice, followed by 45 seconds in the 42°C water bath, and 2 minutes in ice. 500 uL of warm SOC outgrowth medium (New England Biolabs #B9020S) was added to the tubes and the mixture was incubated overnight at 37°C and 220 rpm agitation. On the next day, transformed bacteria were streaked onto LB agar plates containing ampicillin (100 ug/mL) and incubated at 37°C overnight. Following, single-grown colonies were diluted in warm LB broth + ampicillin (100 ug/mL) and incubated at 37°C and 220 rpm agitation overnight. Plasmids were recovered using Monarch Plasmid Miniprep Kit (New England Biolabs #T1010S) or E.Z.N.A. Plasmid

DNA Midi Kit (Omega Bio-tek #D6904). Full-length sequencing was carried out by Plasmidsaurus using Oxford Nanopore Technologies.

Site-Directed Mutagenesis

To generate PTPRH constructs with deletions in various sites, primers were designed using the NEBaseChanger software (https://nebasechanger.neb.com/) using the human *PTPRH* coding sequence (CCDS 33110.1) as a template. The sequence of the primers is described in the table below. Mutagenesis was carried out in the PTPRH WT- HA plasmid using the Q5 Site-Direct Mutagenesis Kit (New England Biolabs #E0554) according to the manufacturer's instructions. The generated plasmids were sequenced as previously described and sequencing results were aligned against the human *PTPRH* coding sequence using local pairwise alignment (MAFFT v7) from Benchling. Constructs containing *in-frame* deletions in the intended site were used for lentiviruses generation and cell transduction.

Construct	Forward (5` - 3`)	Reverse (5` - 3`)
Δ1	ACTCGAAATGCCACCACAGCTCACAACCCAG	GGGGGCAGGCGCCCTGGC
Δ2	TCCGTGTGGGTGGGGAAG	GTTGTGAGCTGTGGTGGC
Δ3	ACCCTCACTGCAGCCACT	AGTCTCCATATGGAGGTTTCTCAC
Δ4	CAACAGTCAGCCCAGGCC	GTGGAATTGGCGCACAGAC

Lentivirus Generation and Cell Transduction

293T cells were seeded on 10 cm dishes (3 x 10^6 cells/dish) on the day before the transfection. On the next day, pCMV-VSV-G, pCMV delta R8.2, and transfer plasmid were mixed with PEI Prime (Sigma-Aldrich #919012) in Opti-MEM medium (Gibco #11058021) in a 1 DNA: 3 PEI ratio and added to the cells. Mock viruses were generated using pCMV-VSV-G and pCMV delta R8.2 only. After 6 hours of incubation, the medium was discarded, and cells were replenished with a complete DMEM high glucose medium. The supernatant containing the lentiviruses was harvested at 48 and 72 hours after transfection and filtered through a 0.45 µm Millex PES filter

(Millipore Sigma #SLHPR33RS). The lentiviruses were concentrated 10x using Lenti-X Concentrator (Takara Bio #631232) following the manufacturer's instructions and stored at -80°C.

For the transduction step, target cells were plated in 6-well plates using 10^5 cells/well on day 1. On the next day, when cells were ~40% confluent, a mixture of complete media, Viral Entry Transduction Enhancer (abm #G515), and an equal volume of lentiviruses were added. The transduction was stopped on day 3 by replenishing the cells with a fresh complete medium. On day 5, we started the selection of infected cells with blasticidin (Dot Scientific Inc. #3513-03-9). We defined the optimal antibiotic concentration using a killing curve, and NCI-H2023 and NCI-H23 (parental and transduced) were treated with 4, 6, and 8 ug/mL of blasticidin, respectively. The selection was stopped when all parental cells were dead.

BioID2 experiment

BioID2-HA was cloned downstream of the EF-1α core promoter into the C-terminal of *PTPRH* WT or an empty lentiviral backbone to generate the PTPRH WT-BioID2 or BioID2-only plasmids, respectively, using the golden gate method. To generate the PTPRH WT linearized vector, the PTPRH-HA WT plasmid was amplified to include all sequences but the HA tag using the primers F: 5' GACCTCCAACTTGTGGGC 3' and R: 5' GGATCCGGCGCAACAAAC 3'. To generate a linearized empty lentiviral backbone (without the PTPRH WT-HA sequence), the same plasmid was amplified using the primers F: 5' ACCGGTCCTGTGTTCTGG 3' and R: 5' GGATCCGGCGCAACAAAC3'. BioID2-HA sequence was extracted from MCS-BioID2-HA pBabe-puro using the primers F: 5' aagtttgttgcgccggatccTGCGTAATCCGGTACATC 3' and R: 5' aggcccacaagttggaggtcTTCAAGAACCTGATCTGG 3' or F: 5' aagtttgttgcgccggatccTGCGTAATCCGGTACATC 3' and R: 5' cgccagaacacaggaccggtgccaccatgTTCAAGAACCTGATCTGG 3' to include the start codon ATG. The lower caps letters in the primer sequences indicate the overlapping regions with the vector. Platinum Superfi II PCR MM (Invitrogen #12368010) was used for all amplification reactions following the manufacturer's instructions. The PCR products were purified using the Monarch DNA and PCR Cleaning kit (New England Biolabs #T1030S) and templates were digested using Dnp1 (New England Biolabs #R0176). Assembly reaction was performed at 1:2 vector: insert ratio using the NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs #E5520)

according to the manufacturer's protocol. After bacterial transformation and purification of the generated plasmid, lentiviruses were produced as previously described.

NCI-H2023 and H23 KO cells were transduced with BioID2-only, PTPRH WT-BioID2, or mock lentiviruses. Cells expressing the BioID2-only and PTPRH WT-BioID2 transgenes were selected by treatment with blasticidin and all cell lines were expanded to 100 mm cell culture dishes (n= 3 technical replicates/cell line). When ~ 60% confluency was reached, cells were serum-starved for 24 hours by adding fresh complete medium without FBS. After the incubation period, the old medium was removed, and cells were replenished with a complete medium containing FBS and 50 uM biotin (Thermo Scientific #29129). Cells were lysed 18 hours after treatment for protein collection and analyzed by immunoblotting. For immunoprecipitation of biotinylated proteins, 800 ug of protein lysates and 0.5 mg of MagReSyn Streptavidin MS beads (ReSyn Bio #MR-STV002) equilibrated in Pierce RIPA buffer + Halt Protease and Phosphatase inhibitor (Thermo Scientific #78441) were incubated for 24 hours at 4°C using end-to-end rotation. The washing of the antigen-bead complex was performed following the manufacturer's instructions. Next, the immunoprecipitated fraction was run in LC/MS/MS for protein identification. The analysis of PTPRH interactors adhered to the recommendations outlined by Roux *et al* [75,76]. Briefly, the identified proteins underwent initial filtration to exclude those commonly observed in biotinylation experiments, referencing the CRAPome database. Those with a CRAPome score > 300 were considered contaminants and removed from the analysis. Furthermore, histones, endogenous carboxylases, and other known BioID2 interactors (i.e. AHNAK, FLNA, PARP1, EEF1A1, PRKDC, TOP1, PKM) were manually curated. Candidate interactors were defined as having a minimum of a 2-fold increase in normalized total precursor intensity in the PTPRH WT-BioID2 group when compared to BioID2-only and$p < 0.05$. STRING analysis of these candidate interactors was done using Cytoscape 3.10.2.

Immunofluorescence

Mock, BioID2-only, and PTPRH WT-BioID2 cells were seeded in duplicate at 10^5 cells/well in 6-well plates containing coverslips on the day before. Cells were fixed in 4% Paraformaldehyde (PFA, pH 7.4) for 10 minutes at RT and rinsed with ice-cold PBS three times. Next, cells were permeabilized using PBS-Tween 0.4%, washed three times with PBS for 5 min/wash, and blocked with Image-iT FX signal enhancer (Invitrogen #I36933) for 30 minutes. The coverslips

were then incubated with primary antibodies against BioID2 (1:5000, BioFront #BID2-CP-100) or HA tag (1:800, Cell Signaling #3724) diluted in 1% BSA/ PBS-Tween 0.1% for one hour in a humidified chamber at RT. Following three washes with PBS of 5 minutes/each, Alexa Fluor 488 goat anti-chicken (1:400, Invitrogen #A-11039) or Alexa Fluor 594 goat anti-rabbit (1:400, Invitrogen #A-11012) diluted in 1% BSA/ PBS-Tween 0.1% was added to the coverslips and incubated for 1 hour at RT. Coverslips were mounted to glass slides with Prolong Gold Antifade Mountant with DAPI (Invitrogen #P36935). Images were acquired in the Nikkon Ti2 microscope using FITC (700 ms exposure), TRITC (500 ms exposure), and DAPI (autoexposure) filters. Images were processed with ImageJ 1.54g.

qPCR

NCI-H23, A427, and HT-29 cells were seeded as triplicates in 6-well plates (2 x 10^5 cells/well) in complete medium. On the following morning, EGF was added to each well (500 ng/mL) and cells were incubated for 18 hours. Total RNA was extracted using the RNeasy Mini Kit (Qiagen #74104) according to the manufacturer's instructions. cDNA generation and qPCR were performed in one-step using 100 ng of template, 100 nM of primers (PTPRH, F: 5' TGTGCGTCCACATACCCAGA 3', R: 5' CAGCCTCCCCACATGAAGAT 3'; GAPDH, F: 5' TCATGACCACAGTGGATGCC 3', R: 5' GGAGTTGCTGTTGAAGTCGC 3'), and Power SYBR™ Green RNA-to-CT™ 1-Step Kit (Applied Biosystems #4391178) following manufacturer's protocol. ΔΔ Ct analysis was performed as described by Livak & Schmittgen.[77]

RNA-sequencing

NCI-H2023 and H23 KO cells were transduced with *PTPRH* WT-HA lentivirus to stably overexpress PTPRH and protein levels were validated by western blotting as described above. The parental and transduced cell lines were seeded as triplicates in 6-well plates (10^5 cells/well) and incubated overnight. On the next day, cells were replenished with serum-free complete medium. After 24 hours of serum starvation, a complete medium containing FBS was added to the cells and incubated for an additional 16 hours. Total RNA was collected from the FBS-stimulated cells using RNeasy Mini Kit (Qiagen #74104) according to the manufacturer's protocol. Library preparation and paired-end RNA sequencing (Illumina PE150) were conducted by Novogene using NovaSeq X Plus Series (Illumina) using a minimum coverage of 30M reads per sample. The quality of the raw reads was assessed through *FastQC* 0.11.7 and trimmed using

Trimmomatic 0.39. Trimmed reads were aligned against the human genome (grch38) using *Hisat2* 2.1.0 and a count matrix of raw reads to be used as input for DESeq2 was generated by *StringTie* 2.1.3 as described in the *StringTie StringTie* documentation. Differential expression analysis was performed using DESeq2 1.42.1 and differentially expressed genes (DEG) with padj < 0.05 were considered significant. The *gage* package 2.52.0 was used for gene set enrichment for pathway analysis of GO biological processes in our list of DEG, as well as GSEA for hallmark and C6 oncogenic signature gene sets from MSigDB using the *fgsea* package 1.28.0. Pathways with padj < 0.05 were considered enriched and reported in our analysis.

Data Visualization and Statistical Analysis

Data visualization and statistical tests (i.e. two-tail t-test, ANOVA) were performed in R version 4.3.3. * $p < 0.05$ was considered as a significant result.

Data Availability

RNA-sequencing results are publicly available on the GEO repository under accession number GSE270167.

Supporting Information

This article contains supporting information.

Acknowledgments

We would like to thank Douglas Whitten from the MSU Proteomics Core for his guidance and expertise with the MS/MS data analysis. We also would like to acknowledge the members of the Andrechek lab – Carson Broeker, John Vusich, and Jesus Garcia-Lerena – as well as Dr. Amy Ralston, Dr. Sachi Horibata, Dr. Ana Vazquez, and Dr. Dohun Pyeon for providing insightful opinions about this research project.

Authors' contributions:

MO and DP performed all experiments, MS contributed to the experimental design, MO and ERA contributed to the experimental design and drafting of the manuscript.

Conflict of Interest:

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Vitrac, H., Mallampalli, V. K. P. S. & Dowhan, W. Importance of phosphorylation/dephosphorylation cycles on lipid-dependent modulation of membrane protein topology by posttranslational phosphorylation. *J Biol Chem* **294**, 18853–18862 (2019).
2. Bononi, A. *et al*. Protein Kinases and Phosphatases in the Control of Cell Fate. *Enzyme Res* **2011**, 329098 (2011).
3. Qi, X.-M., Wang, F., Mortensen, M., Wertz, R. & Chen, G. Targeting an oncogenic kinase/phosphatase signaling network for cancer therapy. *Acta Pharmaceutica Sinica B* **8**, 511–517 (2018).
4. Ardito, F., Giuliani, M., Perrone, D., Troiano, G. & Lo Muzio, L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *International Journal of Molecular Medicine* **40**, 271–280 (2017).
5. Turdo, A. *et al*. Targeting Phosphatases and Kinases: How to Checkmate Cancer. *Front Cell Dev Biol* **9**, 690306 (2021).
6. Harrison, P. T., Vyse, S. & Huang, P. H. Rare epidermal growth factor receptor (EGFR) mutations in non-small cell lung cancer. *Seminars in Cancer Biology* **61**, 167–179 (2020).
7. Gijtenbeek, R. G. P. *et al.* Overall survival in advanced epidermal growth factor receptor mutated non-small cell lung cancer using different tyrosine kinase inhibitors in The Netherlands: a retrospective, nationwide registry study. *The Lancet Regional Health – Europe* **27**, (2023).
8. Amrhein, J. A. *et al*. Macrocyclization of Quinazoline-Based EGFR Inhibitors Leads to Exclusive Mutant Selectivity for EGFR L858R and Del19. *J. Med. Chem.* **65**, 15679–15697 (2022).

9. Reguart, N. & Remon, J. Common EGFR-mutated subgroups (Del19/L858R) in advanced non-small-cell lung cancer: chasing better outcomes with tyrosine kinase inhibitors. *Future Oncol* **11**, 1245–1257 (2015).

10. Galdadas, I. *et al.* Structural basis of the effect of activating mutations on the EGF receptor. *eLife* **10**, e65824.

11. Red Brewer, M. *et al.* Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. *Proceedings of the National Academy of Sciences* **110**, E3595–E3604 (2013).

12. Gazdar, A. Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene* **28**, S24–S31 (2009).

13. Liu, Q. *et al.* EGFR-TKIs resistance via EGFR-independent signaling pathways. *Molecular Cancer* **17**, 53 (2018).

14. Hsu, P.-C., Jablons, D. M., Yang, C.-T. & You, L. Epidermal Growth Factor Receptor (EGFR) Pathway, Yes-Associated Protein (YAP) and the Regulation of Programmed Death-Ligand 1 (PD-L1) in Non-Small Cell Lung Cancer (NSCLC). *Int J Mol Sci* **20**, 3821 (2019).

15. Grosse, A., Grosse, C., Rechsteiner, M. & Soltermann, A. Analysis of the frequency of oncogenic driver mutations and correlation with clinicopathological characteristics in patients with lung adenocarcinoma from Northeastern Switzerland. *Diagnostic Pathology* **14**, 18 (2019).

16. Dang, A.-T. H. *et al.* Actionable Mutation Profiles of Non-Small Cell Lung Cancer patients from Vietnamese population. *Sci Rep* **10**, 2707 (2020).

17. cBioPortal for Cancer Genomics. https://www.cbioportal.org/.

18. Du, Z. & Lovly, C. M. Mechanisms of receptor tyrosine kinase activation in cancer. *Mol Cancer* **17**, 58 (2018).

19. Bollu, L. R., Mazumdar, A., Savage, M. I. & Brown, P. H. Molecular Pathways: Targeting Protein Tyrosine Phosphatases in Cancer. *Clinical Cancer Research* **23**, 2136–2142 (2017).

20. Gkountakos, A. *et al.* PTEN in Lung Cancer: Dealing with the Problem, Building on New Knowledge and Turning the Game Around. *Cancers (Basel)* **11**, 1141 (2019).

21. Sirhan, Z., Alojair, R., Thyagarajan, A. & Sahu, R. P. Therapeutic Implications of PTEN in Non-Small Cell Lung Cancer. *Pharmaceutics* **15**, 2090 (2023).

22. Richards, C. E. *et al.* Protein Tyrosine Phosphatase Non-Receptor 11 (PTPN11/Shp2) as a Driver Oncogene and a Novel Therapeutic Target in Non-Small Cell Lung Cancer (NSCLC). *Int J Mol Sci* **24**, 10545 (2023).

23. Swiatnicki, M. R. *et al.* Elevated phosphorylation of EGFR in NSCLC due to mutations in PTPRH. *PLOS Genetics* **18**, e1010362 (2022).

24. Rennhack, J. P. *et al.* Integrated analyses of murine breast cancer models reveal critical parallels with human disease. *Nat Commun* **10**, 3261 (2019).

25. Yao, Z. *et al.* A Global Analysis of the Receptor Tyrosine Kinase-Protein Phosphatase Interactome. *Mol Cell* **65**, 347–360 (2017).

26. Protein tyrosine phosphatases: from genes, to function, to disease | Nature Reviews Molecular Cell Biology. https://www.nature.com/articles/nrm2039.

27. Matozaki, T. *et al.* Molecular cloning of a human transmembrane-type protein tyrosine phosphatase and its expression in gastrointestinal cancers. *J Biol Chem* **269**, 2075–2081 (1994).

28. Phosphotyrosine phosphatase R3 receptors: Origin, evolution and structural diversification | PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0172887.

29. Interaction of SAP-1, a Transmembrane-type Protein-tyrosine Phosphatase, with the Tyrosine Kinase Lck: ROLES IN REGULATION OF T CELL FUNCTION - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0021925820835364?via%3Dihub.

30. Protein tyrosine phosphatase SAP-1 protects against colitis through regulation of CEACAM20 in the intestinal epithelium | PNAS. https://www.pnas.org/doi/full/10.1073/pnas.1510167112.

31. Inhibition of Cell Growth and Spreading by Stomach Cancer-associated Protein-tyrosine Phosphatase-1 (SAP-1) through Dephosphorylation of p130cas - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0021925819568181?via%3Dihub.

32. Nishiyama, K. *et al*. CNKSR1 serves as a scaffold to activate an EGFR phosphatase via exclusive interaction with RhoB-GTP. *Life Science Alliance* **4**, (2021).

33. Induction of Apoptosis by Stomach Cancer-associated Protein-tyrosine Phosphatase-1 - ScienceDirect. https://www.sciencedirect.com/science/article/pii/S002192582074336X?via%3Dihub.

34. Wälchli, S., Espanel, X. & Hooft van Huijsduijnen, R. Sap-1/PTPRH activity is regulated by reversible dimerization. *Biochem Biophys Res Commun* **331**, 497–502 (2005).

35. Seo, Y. *et al*. Overexpression of SAP-1, a Transmembrane-Type Protein Tyrosine Phosphatase, in Human Colorectal Cancers. *Biochemical and Biophysical Research Communications* **231**, 705–711 (1997).

36. Chen, A., Ding, S., Shen, X. & Lin, X. The High Expression of PTPRH Is Associated with Poor Prognosis of Human Lung Adenocarcinoma. *Computational and Mathematical Methods in Medicine* **2021**, e9932088 (2021).

37. Bujko, M. *et al.* Downregulation of PTPRH (Sap-1) in colorectal tumors. *International Journal of Oncology* **51**, 841–850 (2017).

38. Sadakata, H. *et al.* SAP-1 is a microvillus-specific protein tyrosine phosphatase that modulates intestinal tumorigenesis. *Genes to Cells* **14**, 295–308 (2009).

39. Nagano, H. *et al.* Downregulation of stomach cancer-associated protein tyrosine phosphatase-1 (SAP-1) in advanced human hepatocellular carcinoma. *Oncogene* **22**, 4656–4663 (2003).

40. Capuani, F. *et al.* Quantitative analysis reveals how EGFR activation and downregulation are coupled in normal but not in cancer cells. *Nat Commun* **6**, 7999 (2015).

41. Salazar-Cavazos, E. *et al.* Multisite EGFR phosphorylation is regulated by adaptor protein abundances and dimer lifetimes. *MBoC* **31**, 695–708 (2020).

42. Kim, D. I. *et al.* An improved smaller biotin ligase for BioID proximity labeling. *MBoC* **27**, 1188–1196 (2016).

43. Bonham, C. A., Mandati, V., Singh, R. K., Pappin, D. J. & Tonks, N. K. Coupling substrate-trapping with proximity-labeling to identify protein tyrosine phosphatase PTP1B signaling networks. *Journal of Biological Chemistry* **299**, (2023).

44. Tonks, N. K. Protein tyrosine phosphatases: from genes, to function, to disease. *Nat Rev Mol Cell Biol* **7**, 833–846 (2006).

45. Wang, J., An, H., Mayo, M. W., Baldwin, A. S. & Yarbrough, W. G. LZAP, a putative tumor suppressor, selectively inhibits NF-kappaB. *Cancer Cell* **12**, 239–251 (2007).

46. Bonnet, M. C., Weil, R., Dam, E., Hovanessian, A. G. & Meurs, E. F. PKR stimulates NF-kappaB irrespective of its kinase function by interacting with the IkappaB kinase complex. *Mol Cell Biol* **20**, 4532–4542 (2000).

47. Gil, J. *et al.* TRAF family proteins link PKR with NF-kappa B activation. *Mol Cell Biol* **24**, 4502–4512 (2004).

48. Silva, A. M. *et al.* Protein kinase R (PKR) interacts with and activates mitogen-activated protein kinase kinase 6 (MKK6) in response to double-stranded RNA stimulation. *J Biol Chem* **279**, 37670–37676 (2004).

49. Matheny, S. A. *et al.* Ras regulates assembly of mitogenic signalling complexes through the effector protein IMP. *Nature* **427**, 256–260 (2004).

50. Schroder, W. A. *et al.* SerpinB2 inhibits migration and promotes a resolution phase signature in large peritoneal macrophages. *Sci Rep* **9**, 12421 (2019).

51. Moidu, N. A., A Rahman, N. S., Syafruddin, S. E., Low, T. Y. & Mohtar, M. A. Secretion of pro-oncogenic AGR2 protein in cancer. *Heliyon* **6**, e05000 (2020).

52. Ortiz, M. M. O. & Andrechek, E. R. Molecular Characterization and Landscape of Breast cancer Models from a multi-omics Perspective. *J Mammary Gland Biol Neoplasia* **28**, 12 (2023).

53. Genetic Mutation | Learn Science at Scitable. http://www.nature.com/scitable/topicpage/genetic-mutation-1127.

54. Sun, Q., Schindelholz, B., Knirr, M., Schmid, A. & Zinn, K. Complex genetic interactions among four receptor tyrosine phosphatases regulate axon guidance in Drosophila. *Mol Cell Neurosci* **17**, 274–291 (2001).

55. den Hertog, J. & van Eekelen, M. Chapter 101 - Zebrafish and Phosphatase Function. in *Handbook of Cell Signaling (Second Edition)* (eds. Bradshaw, R. A. & Dennis, E. A.) 827–833 (Academic Press, San Diego, 2010). doi:10.1016/B978-0-12-374145-5.00101-7.

56. Stanoev, A. *et al.* Interdependence between EGFR and Phosphatases Spatially Established by Vesicular Dynamics Generates a Growth Factor Sensing and Responding Network. *Cell Syst* **7**, 295-309.e11 (2018).

57. Wu, Y. *et al.* Protein tyrosine phosphatase SHP-1 sensitizes EGFR/HER-2 positive breast cancer cells to trastuzumab through modulating phosphorylation of EGFR and HER-2. *OTT* **8**, 2577–2587 (2015).

58. Forbes, K., Skinner, L., Aplin, J. D. & Westwood, M. The tyrosine phosphatase SHP-1 negatively regulates cytotrophoblast proliferation in first-trimester human placenta by modulating EGFR activation. *Cell. Mol. Life Sci.* **69**, 4029–4040 (2012).

59. Agazie, Y. M. & Hayman, M. J. Molecular Mechanism for a Role of SHP2 in Epidermal Growth Factor Receptor Signaling. *Mol Cell Biol* **23**, 7875–7886 (2003).

60. Sun, T. *et al.* Activation of multiple proto-oncogenic tyrosine kinases in breast cancer via loss of the PTPN12 phosphatase. *Cell* **144**, 703–718 (2011).

61. Xu, Y., Tan, L.-J., Grachtchouk, V., Voorhees, J. J. & Fisher, G. J. Receptor-type Protein-tyrosine Phosphatase-κ Regulates Epidermal Growth Factor Receptor Function*. *Journal of Biological Chemistry* **280**, 42694–42700 (2005).

62. Xu, Y. *et al.* Receptor Type Protein Tyrosine Phosphatase-Kappa Mediates Cross-Talk between Transforming Growth Factor-Beta and Epidermal Growth Factor Receptor Signaling Pathways in Human Keratinocytes. *MBoC* **21**, 29–35 (2010).

63. Rizzo, S. *et al.* Promoting the activity of a receptor tyrosine phosphatase with a novel pH-responsive transmembrane agonist inhibits cancer-associated phenotypes. *Protein Science* **32**, e4742 (2023).

64. Schwarz, M. *et al*. Disrupting PTPRJ transmembrane-mediated oligomerization counteracts oncogenic receptor tyrosine kinase FLT3 ITD. *Front. Oncol*. **12**, (2022).

65. Nishiyama, K. *et al*. CNKSR1 serves as a scaffold to activate an EGFR phosphatase via exclusive interaction with RhoB-GTP. *Life Sci Alliance* **4**, e202101095 (2021).

66. Asbagh, L. A. *et al.* The tyrosine phosphatase PTPRO sensitizes colon cancer cells to anti-EGFR therapy through activation of SRC-mediated EGFR signaling. *Oncotarget* **5**, 10070–10083 (2014).

67. Blakely, C. M. *et al.* NF-κB activating complex engaged in response to EGFR oncogene inhibition drives tumor cell survival and residual disease in lung cancer. *Cell Rep* **11**, 98–110 (2015).

68. Biswas, D. K., Cruz, A. P., Gansberger, E. & Pardee, A. B. Epidermal growth factor-induced nuclear factor κB activation: A major pathway of cell-cycle progression in estrogen-receptor negative breast cancer cells. *Proceedings of the National Academy of Sciences* **97**, 8542–8547 (2000).

69. Shostak, K. & Chariot, A. EGFR and NF-κB: partners in cancer. *Trends in Molecular Medicine* **21**, 385–393 (2015).

70. Huntzinger, E., Sinteff, J., Morlet, B. & Séraphin, B. HELZ2: a new, interferon-regulated, human 3′-5′ exoribonuclease of the RNB family is expressed from a non-canonical initiation codon. *Nucleic Acids Res* **51**, 9279–9293 (2023).

71. Fagerlund, R., Kinnunen, L., Köhler, M., Julkunen, I. & Melén, K. NF-{kappa}B is transported into the nucleus by importin {alpha}3 and importin {alpha}4. *J Biol Chem* **280**, 15942–15951 (2005).

72. The Human Protein Atlas. https://www.proteinatlas.org/.

73. PTPRH protein expression summary - The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000080031-PTPRH.

74. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896–1906 (2007).

75. Roux, K. J., Kim, D. I., Burke, B. & May, D. G. BioID: A Screen for Protein-Protein Interactions. *Curr Protoc Protein Sci* **91**, 19.23.1-19.23.15 (2018).

76. Sears, R. M., May, D. G. & Roux, K. J. BioID as a Tool for Protein-Proximity Labeling in Living Cells. *Methods Mol Biol* **2012**, 299–313 (2019).

77. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).

Abbreviations

EGFR = Epidermal Growth Factor Receptor, EGF = Epidermal Growth Factor, PTPRH = protein tyrosine phosphatase receptor type H, NSCLC = non-small cell lung cancer, RTK = receptor tyrosine kinase, PI3K = phosphatidylinositol 3-kinases, mTOR = mammalian target of rapamycin, Akt = Protein kinase B, RAS = Rat sarcoma virus, RAF = Raf-1 proto-oncogene, MEK = MAPK/ERK kinase, ERK = Extracellular signal-regulated kinases, IL-6 = Interleukin 6, JAK = Janus kinase, STAT3 = Signal transducer and activator of transcription 3, EGFRi = EGFR inhibitors, PIK3CA = Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha, SKT11 = Serine/Threonine Kinase 11, MET = MET Proto-Oncogene, ALK = ALK Receptor Tyrosine Kinase, BRAF = B-Raf Proto-Oncogene, HER2/ Erbb2 = Erb-B2 Receptor Tyrosine Kinase 2, ROS = ROS Proto-Oncogene 1, RET = Ret Proto-Oncogene, MAP2K1 = Mitogen-

Activated Protein Kinase Kinase 1, PTEN = Phosphatase And Tensin Homolog, PTPN11 = Protein Tyrosine Phosphatase Non-Receptor Type 11, MAPK = Mitogen-activated protein kinase, RPTP = receptor-like protein tyrosine phosphatase, MMTV = mouse mammary tumor virus, PyMT = polyoma middle tumor-antigen, FN3 = fibronectin type 3, MS/MS = tandem mass spectrometry, Y = tyrosine, Co-IP = co-immunoprecipitation, WCL = whole cell lysate, PCA =principal component analysis, PTP1B = tyrosine-protein phosphatase non-receptor type 1, CDK5RAP3 = CDK5 Regulatory Subunit Associated Protein 3, EIF2AK2 = Eukaryotic Translation Initiation Factor 2 Alpha Kinase 2, NKAP = NFKB Activating Protein, BRAP = BRCA1 Associated Protein, NFKB2 = Nuclear Factor Kappa B Subunit 2, THBS-1 = thrombospondin 1, DAVID = Database for Annotation, Visualization and Integrated Discovery, SERPINB2 = Serpin Family B Member 2, RTL1 = Retrotransposon Gag Like 1, ERG = ETS Transcription Factor ERG, CDH17 = Cadherin 17, AGR2 = Anterior Gradient 2, CYP4F11 = Cytochrome P450 Family 4 Subfamily F Member 11, DEG = differentially expressed genes, GSEA = gene set enrichment analysis, MSigDB = Molecular Signature Database, GO =gene ontology, PTP =protein tyrosine phosphatase, TK = tyrosine kinase, PTPN20 = Protein Tyrosine Phosphatase Non-Receptor Type 20, PTPRQ = Protein Tyrosine Phosphatase Receptor Type Q, PTPRVP = Protein Tyrosine Phosphatase Receptor Type V, PTPRJ = Protein Tyrosine Phosphatase Receptor Type J, PTPRM = Protein Tyrosine Phosphatase Receptor Type M, PTPRN = Protein Tyrosine Phosphatase Receptor Type N, PTPRU = Protein Tyrosine Phosphatase Receptor Type U, PTPRD = Protein Tyrosine Phosphatase Receptor Type D, PTPRR = Protein Tyrosine Phosphatase Receptor Type R, PTPN2 = Protein Tyrosine Phosphatase Non-Receptor Type 2, PTPRB = Protein Tyrosine Phosphatase Receptor Type B, FLT3 = Fms Related Receptor Tyrosine Kinase 3, EPHB1 = EPH Receptor B1, KIT = KIT Proto-Oncogene, ROR2 = Receptor Tyrosine Kinase Like Orphan Receptor 2, NTRK1 = Neurotrophic Receptor Tyrosine Kinase 1, PDGFRA = Platelet Derived Growth Factor Receptor Alpha, EPHB6 = EPH Receptor B6, EPHA7 = EPH Receptor A7, MST1R = Macrophage Stimulating 1 Receptor, EPHA8 = EPH Receptor A8, ERBB3 = Erb-B2 Receptor Tyrosine Kinase 3, LMTK3 = Lemur Tyrosine Kinase 3, TEK = TEK Receptor Tyrosine Kinase, ROR1 = Receptor Tyrosine Kinase Like Orphan Receptor 1, FGFR1 = Fibroblast Growth Factor Receptor 1, PDGFRB = Platelet Derived Growth Factor Receptor Beta, LEF1 = Lymphoid Enhancer Binding Factor 1, BMI1 = BMI1 Proto-Oncogene, NFE2L2 = NFE2 Like BZIP Transcription

Factor 2, IL-21 = Interleukin 21, KRAS = KRAS Proto-Oncogene, MYO1B = Myosin IB, MYO1C = Myosin IC, NES =normalized enrichment score

Table 1: DAVID Functional Analysis (BioID)

Category	Term	PValue	Fold Enrichment	Bonferroni	Benjamini	FDR
GOTERM_MF_DIRECT	GO:0003723~RNA binding	4.58E-20	7.349316	7.6E-18	7.6E-18	7.41E-18
UP_KW_DOMAIN	KW-0809~Transit peptide	3.44E-07	8.200572	3.79E-06	3.79E-06	3.79E-06
UP_KW_MOLECULAR_FUNCTION	KW-0687~Ribonucleoprotein	1.44E-06	8.592936	4.47E-05	4.47E-05	4.33E-05
GOTERM_BP_DIRECT	GO:0032543~mitochondrial translation	2.2E-07	26.55477	7.34E-05	7.34E-05	7.34E-05
UP_KW_PTM	KW-0007~Acetylation	1.06E-05	2.238196	8.46E-05	9.52E-05	9.52E-05
UP_KW_MOLECULAR_FUNCTION	KW-0689~Ribosomal protein	5.85E-06	11.01299	0.000181	9.07E-05	8.78E-05
UP_KW_MOLECULAR_FUNCTION	KW-0694~RNA-binding	9.58E-06	4.658216	0.000297	9.9E-05	9.58E-05
INTERPRO	IPR011545:DEAD/DEAH_box_helicase_dom	1.17E-06	31.65634	0.000298	0.000298	0.000297
SMART	SM00490:HELICc	8.19E-06	20.29105	0.000369	0.000202	0.000202
SMART	SM00487:DEXDc	8.96E-06	19.92544	0.000403	0.000202	0.000202
UP_SEQ_FEATURE	TRANSIT:Mitochondrion	1.36E-06	7.506372	0.000477	0.000477	0.000476
UP_KW_CELLULAR_COMPONENT	KW-0496~Mitochondrion	8.26E-05	3.542927	0.001568	0.001569	0.001569
GOTERM_CC_DIRECT	GO:0005739~mitochondrion	1.34E-05	3.92196	0.001586	0.000838	0.000803
GOTERM_CC_DIRECT	GO:0005762~mitochondrial large ribosomal subunit	1.42E-05	33.31331	0.001675	0.000838	0.000803
INTERPRO	IPR001650:Helicase_C	9.72E-06	20.62018	0.002476	0.000943	0.00094
INTERPRO	IPR014001:Helicase_ATP-bd	1.11E-05	20.06786	0.002826	0.000943	0.00094
UP_SEQ_FEATURE	DOMAIN:Helicase C-terminal	9.35E-06	20.7697	0.003268	0.0016	0.001595
GOTERM_MF_DIRECT	GO:0003676~nucleic acid binding	2.42E-05	9.024134	0.004015	0.002011	0.001963
UP_SEQ_FEATURE	DOMAIN:Helicase ATP-binding	1.37E-05	19.19888	0.004788	0.0016	0.001595
UP_KW_BIOLOGICAL_PROCESS	KW-0698~rRNA processing	0.000142	17.92515	0.005082	0.005095	0.005095
INTERPRO	IPR027417:P-loop_NTPase	2.07E-05	4.912787	0.005263	0.001319	0.001314
KEGG_PATHWAY	hsa03010:Ribosome	0.000106	11.75837	0.00529	0.005304	0.005304
UP_KW_MOLECULAR_FUNCTION	KW-0347~Helicase	0.000179	11.01299	0.005539	0.001388	0.001344

GOTERM_BP_DIRECT	GO:0006412~translation	2.68E-05	11.60999	0.008904	0.004472	0.004472
GOTERM_MF_DIRECT	GO:0003724~RNA helicase activity	5.8E-05	23.31235	0.009582	0.003209	0.003132
UP_KW_LIGAND	KW-0547~Nucleotide-binding	0.001199	2.219455	0.011927	0.011992	0.010793
UP_KW_PTM	KW-1017~Isopeptide bond	0.002651	2.333118	0.021014	0.00797	0.00797
UP_KW_PTM	KW-0597~Phosphoprotein	0.002657	1.347864	0.021057	0.00797	0.00797
INTERPRO	IPR014014:RNA_helicase_DEAD_Q_motif	9.61E-05	44.07059	0.024221	0.004904	0.004884
UP_SEQ_FEATURE	DOMAIN:DEAD-box RNA helicase Q	7.61E-05	47.59722	0.026275	0.005334	0.005319
UP_SEQ_FEATURE	CROSSLNK:Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in SUMO2)	7.73E-05	3.87636	0.026709	0.005334	0.005319
GOTERM_MF_DIRECT	GO:0003735~structural constituent of ribosome	0.000181	11.16017	0.02955	0.007498	0.007317
UP_SEQ_FEATURE	MOTIF:DEAD box	9.14E-05	44.79739	0.0315	0.005334	0.005319
UP_KW_DISEASE	KW-1274~Primary mitochondrial disease	0.00739	8.506941	0.043531	0.044342	0.044342
UP_KW_LIGAND	KW-0949~S-adenosyl-L-methionine	0.004567	6.921	0.04474	0.017621	0.015859
UP_SEQ_FEATURE	MOTIF:Q motif	0.000138	39.05413	0.047286	0.00692	0.0069
UP_KW_LIGAND	KW-0067~ATP-binding	0.005286	2.299017	0.051622	0.017621	0.015859
GOTERM_CC_DIRECT	GO:0005730~nucleolus	0.000853	3.21877	0.09576	0.025861	0.024765
GOTERM_CC_DIRECT	GO:1990904~ribonucleoprotein complex	0.000877	7.91044	0.098315	0.025861	0.024765
GOTERM_CC_DIRECT	GO:0005840~ribosome	0.001135	10.72153	0.125399	0.026782	0.025647
UP_KW_CELLULAR_COMPONENT	KW-0539~Nucleus	0.010685	1.534441	0.184627	0.101509	0.101509
GOTERM_MF_DIRECT	GO:0016887~ATPase activity	0.001236	5.705819	0.185601	0.041036	0.040047
UP_KW_MOLECULAR_FUNCTION	KW-0489~Methyltransferase	0.006847	6.424242	0.191827	0.04245	0.04108
GOTERM_CC_DIRECT	GO:0005654~nucleoplasm	0.001835	1.972631	0.194829	0.031216	0.029894
GOTERM_CC_DIRECT	GO:0005743~mitochondrial inner membrane	0.001852	5.27629	0.196451	0.031216	0.029894

Table 2: GO Biological Processes (RNA-seq)

Upregulated GO	p.geomean	stat.mean	p.val	q.val	set.size	exp1
GO:0006281 DNA repair	9.24E-06	4.318077	9.24E-06	0.0398732	432	9.24E-06
GO:0008380 RNA splicing	2.22E-05	4.12196	2.22E-05	0.0398732	327	2.22E-05

GO:0051276 chromosome organization	2.32E-05	4.098638	2.32E-05	0.0398732	449	2.32E-05
GO:0006397 mRNA processing	3.50E-05	4.011524	3.50E-05	0.0430677	347	3.50E-05
GO:0022613 ribonucleoprotein complex biogenesis	4.18E-05	3.971604	4.18E-05	0.0430677	339	4.18E-05
GO:0051960 regulation of nervous system development	5.30E-05	3.901326	5.30E-05	0.0450794	319	5.30E-05
GO:0006338 chromatin remodeling	6.13E-05	3.863507	6.13E-05	0.0450794	413	6.13E-05
Downregulated GO						
GO:0032787 monocarboxylic acid metabolic process	8.11E-06	-4.339445	8.11E-06	0.0373209	386	8.11E-06
GO:0006631 fatty acid metabolic process	1.45E-05	-4.222542	1.45E-05	0.0373209	244	1.45E-05

Figure legends:

Figure 1. Investigation of pEGFR 1173 (1197) levels on H23 *PTPRH* knockout cell lines. *A)* H23 and H23 KO cells were stimulated with EGF (100 ng/mL) and cell lysates were collected after 5, 15, 30, 60, and 120 minutes. Timepoint 0 represents non-stimulated (untreated) cells. Differences in pEGFR 1173 (1197) (*n*=4/group in each time point) and pAKT S473 (*n*=3/group in each time point) levels were assessed by immunoblot and representative blots are shown (mean +/- s.d.). Results of densitometry for pEGFR 1173 (1197)/total EGFR and pAKT/Total AKT ratio for the replicates are demonstrated as individual dots in the bar plots. T-test was performed to compare the PTPRH KO group with the parental cell (H23) at each time point and *p<0.05 was considered a significant result. *B)* EGFR was immunoprecipitated from H23 (n=2) and H23 KO (n=3) cells stimulated with EGF (100 ng/mL) for 15 minutes and sent for mass spectrometry analysis. Quantitative values of normalized total precursor intensities for the identified phospho-Tyr sites in each cell line are represented in the table and the bar plot (mean +/- s.d.). Abbreviations: ns = non-significant; Tyr (Y) = tyrosine; IB = immunoblot

Figure 2. Investigating pEGFR 1173 (1197) levels in H23 cells overexpressing *PTPRH WT* or various *PTPRH* deletions . *A)* Representation of the PTPRH domains and their location in

the protein. The target region for deletion in each Δ construct is indicated by the blue bar. The table provides a complete description of the deletions and their position in the PTPRH protein. *B)* Lung adenocarcinoma H23 (parental), H23 KO, H23 KO mock (control for lentivirus transduction), or H23 KO cells in which *PTPRH wild-type* or various PTPRH deletions were added back to overexpress the protein were stimulated with EGF (100 ng/mL) for 15 min. The whole cell lysates (*n*=3/group) were used to immunoblot pEGFR 1173 (1197), EGFR, B-actin, PTPRH, and HA-tag, and the representative blot is shown. Densitometry results for pEGFR 1173 (1197)/total EGFR ratio for the replicates are demonstrated as individual dots in the bar plot, and the mean +/- s.d. is shown. T-test was performed to compare each group with the parental cell (H23) and *$p < 0.05$ was considered a significant result. *Abbreviations:* FN3 = fibronectin3*;* ns = non-significant; IB = immunoblot; nt = nucleotide IB = immunoblot

Figure 3. PTPRH does not coimmunoprecipitate with EGFR. Cell lysates were immunoprecipitated (IP) for EGFR and immunoblotted (IB) for EGFR and PTPRH. Rabbit IgG antibody was used as a control isotype. *A)* Lung adenocarcinoma cells H23 or H23 KO were stimulated with EGF (100 ng/mL) for 15 min. The whole cell lysate (WCL) of the colorectal cancer cell line HT29, which expresses high levels of PTPRH, was used as a positive control for IB, whereas H23 KO cells were used as a negative control. B) H23 KO cells in which *PTPRH* WT was added back were stimulated with EGF (100 ng/mL) at different time points (2, 5, 10, and 15 minutes) prior to cell lysis and IP. The same cell line was used for WCL positive control. *Abbreviations:* IB= immunoblot, IP= immunoprecipitation, m= minutes

Figure 4. Validation of lung adenocarcinoma cell lines expressing BioID2-only or PTPRH WT-BioID2. H2023 and H23 KO cell lines transduced with BioID2-HA-only, PTPRH WT-BioID2-HA, or mock (control) were validated for expression of the constructs and their subcellular localization using immunoblot and immunofluorescence, respectively. *A)* Immunoblot against the HA-tag for mock (n=1/cell line), BioID2-only (n=2/cell line), and PTPRH WT-BioID2-HA (n=2/cell line) H2023 and H23 KO cells. Bands matching the correct molecular size for BioID2-HA and PTPRH WT-BioID2-HA are indicated by the arrows. *B)* Immunofluorescence was performed in all groups using antibodies against the HA-tag which are shown in magenta. DAPI staining is shown in blue. Representative cells showing localization of PTPRH WT-BioID2-HA in the plasma membrane and/or cytosol are indicated by arrows. Images

are representative fields using 40x magnification, scale bar = 400 pixels. *Abbreviations*: ab = antibody; IB= immunoblot.

Figure 5. PTPRH interactors revealed by *in vitro* labeling technique (BioID) in lung adenocarcinoma cell lines. H2023 and H23 KO cell lines transduced with BioID2-HA-only, PTPRH WT-BioID2-HA, or mock (negative control) were serum starved for 24h, followed by stimulation with FBS and biotin for 18h. *A)* Immunoblot against biotin showing the presence of biotinylated proteins in the BioID2-only and PTPRH WT-BioID2-HA groups, with some bands (arrows) showing up exclusively in the PTPRH-WT-BioID2-HA groups. *B)* STRING analysis of candidate PTPRH interactors identified in H2023 and H23 KO cell lines transduced with PTPRH WT-BioID2-HA. *C)* Volcano plots representing PTPRH WT-BioID2-HA vs. BioID-HA-only analysis of individual cell lines (H2023 and H23 KO, n=3 replicates/group) or in combination (H23 KO + H2023). Proteins were considered a PTPRH candidate interactor if log2 fold-change>1 in normalized total precursor intensity compared to their respective BioID2-only transduced cell line and $p<0.05$. Candidate interactors are labeled in blue. *D)* Venn diagram of PTPRH candidate interactors identified in each cell line. *E)* Candidate PTPRH interactors were screened for the presence of reported phosphosites in the PhosphoSitePlus database. The percentage of proteins containing at least one pY or pS/pT sites are shown. *Abbreviations*: IB = immunoblot; pY = phospho-Tyrosine; pS/pT = phospho-Serine/phospho-Threonine.

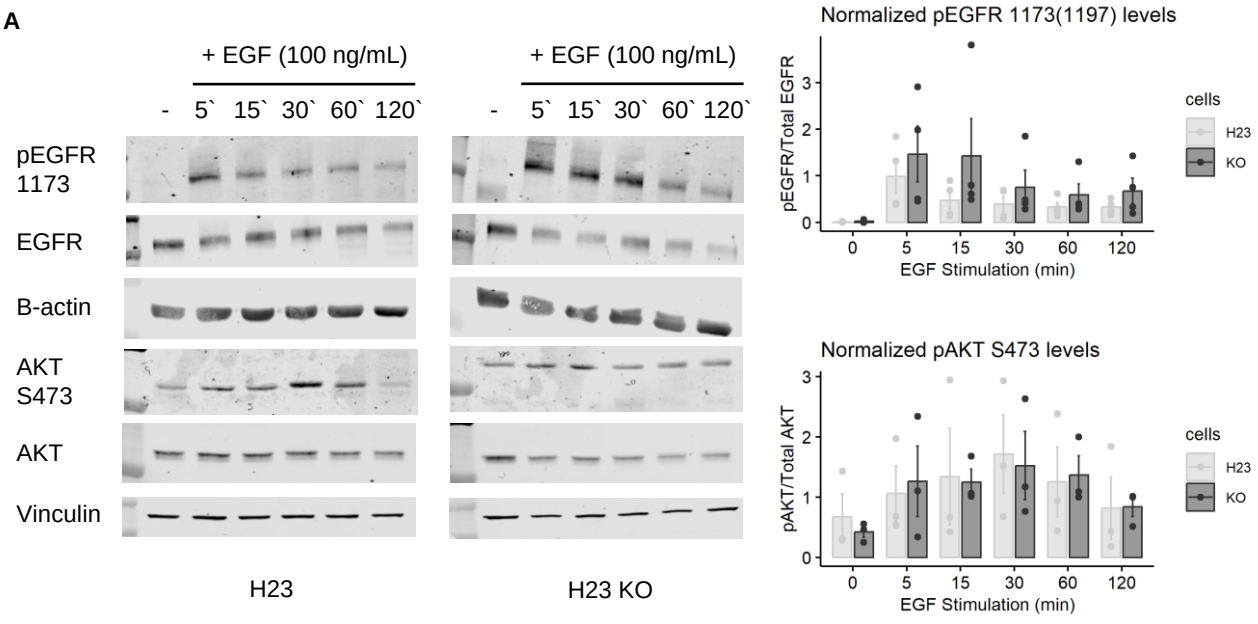
Figure 6. Transcriptome analysis of lung adenocarcinoma cell lines overexpressing PTPRH. H2023 and H23 cell lines expressing endogenous or overexpressed PTPRH levels and H23 knockout cells were serum starved for 24h, followed by stimulation with FBS for 16h prior to RNA extraction (n=3 replicates/group). *A)* PCA plot of counts normalized by variance stabilizing transformation (VST). *B)* Volcano plot of differentially expressed genes in the PTPRH overexpression cell lines compared to cell lines expressing endogenous PTPRH (parental cells) with padj < 0.05. Genes were considered upregulated (red) when log2 fold change > 2 and downregulated (blue) when log2 fold change < -2. *C and D)* Expression of differentially expressed protein tyrosine phosphatases *(C)* and tyrosine kinases *(D)* are represented as log10 normalized counts (padj <0.05).

Figure 7. GSEA analysis of lung adenocarcinoma cell lines overexpressing PTPRH. *A* and *B)* GSEA plots showing significantly enriched pathways in the PTPRH overexpression cell lines

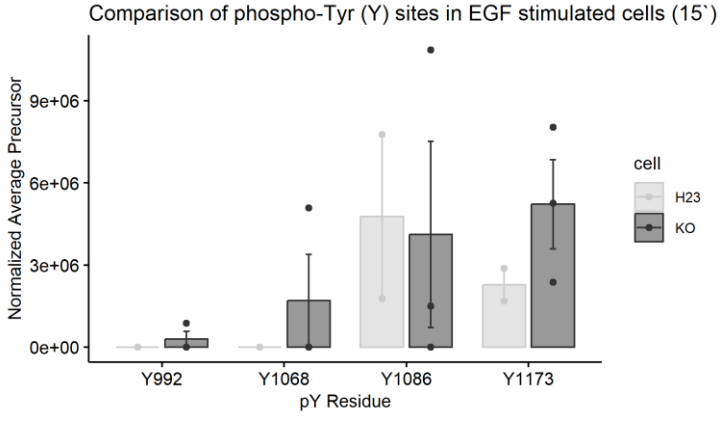
using the C6 oncogenic signature gene set *(A*) and the Hallmark gene set *(B). C* and *D*) The most enriched pathway in each gene set was extracted and the enrichment plot is shown.

Fig 1: Investigation of pEGFR levels on H23 PTPRH knockout cells

A



B



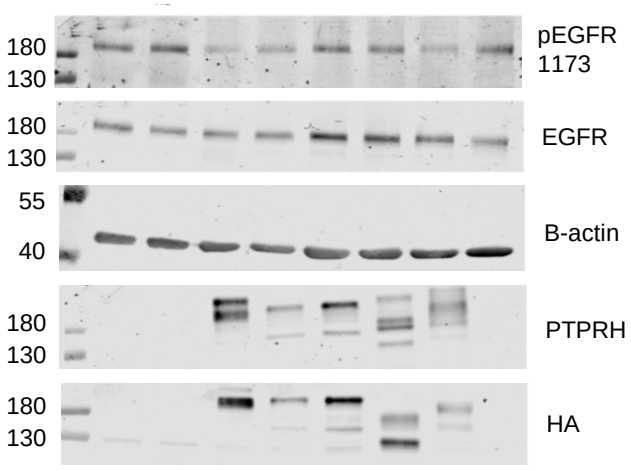
Peptide/Sample	KO #1	KO #2	KO #3	H23 #1	H23 #2
ALMDEEDMDDVVDADEYLIPQQG FFSSPSTSR (Y992)	871939.1	0	0	0	0
GSTAENAEYLR (Y1173)	2365493	5256077	8026998.2	1683354	2882294
RPAGSVQNPVYHNQPLNPAPSR (Y1086)	1490507	10848465	0	1768511	7758721
YSSDPTGALTEDSIDDTFLPVPEY INQSVPK (Y1068)	5076659	0	0	0	0

Fig 2: Investigation of pEGFR 1173 levels in H23 cells overexpressing PTPRH WT or various PTPRH deletions

A

Construct name	Description	Deletion Position
PTPRH-HA Δ1	Deletion of fibronectin3 domains 1 & 2	30 – 200 aa (90-600 nt)
PTPRH-HA Δ2	Deletion of fibronectin3 domains 3 & 4	209 – 364 aa (627-1092 nt)
PTPRH-HA Δ3	Deletion of fibronectin3 domains 5 & 6	396 – 556 aa (1188-1668 nt)
PTPRH-HA Δ4	Deletion of the end portion of the phosphatase domain	980 – 1080 aa (2940-3240 nt)

B



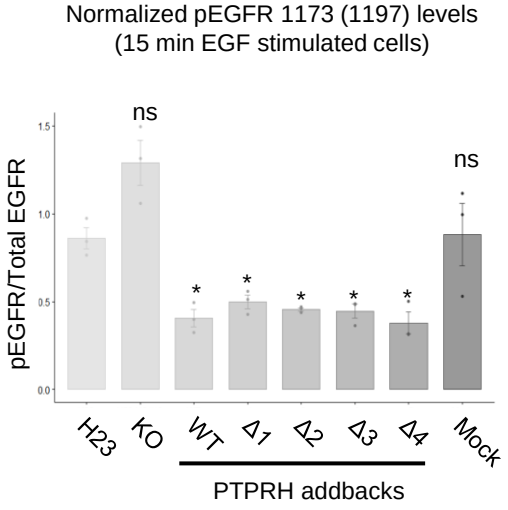


Fig 3: PTPRH does not co-immunoprecipitated with EGFR

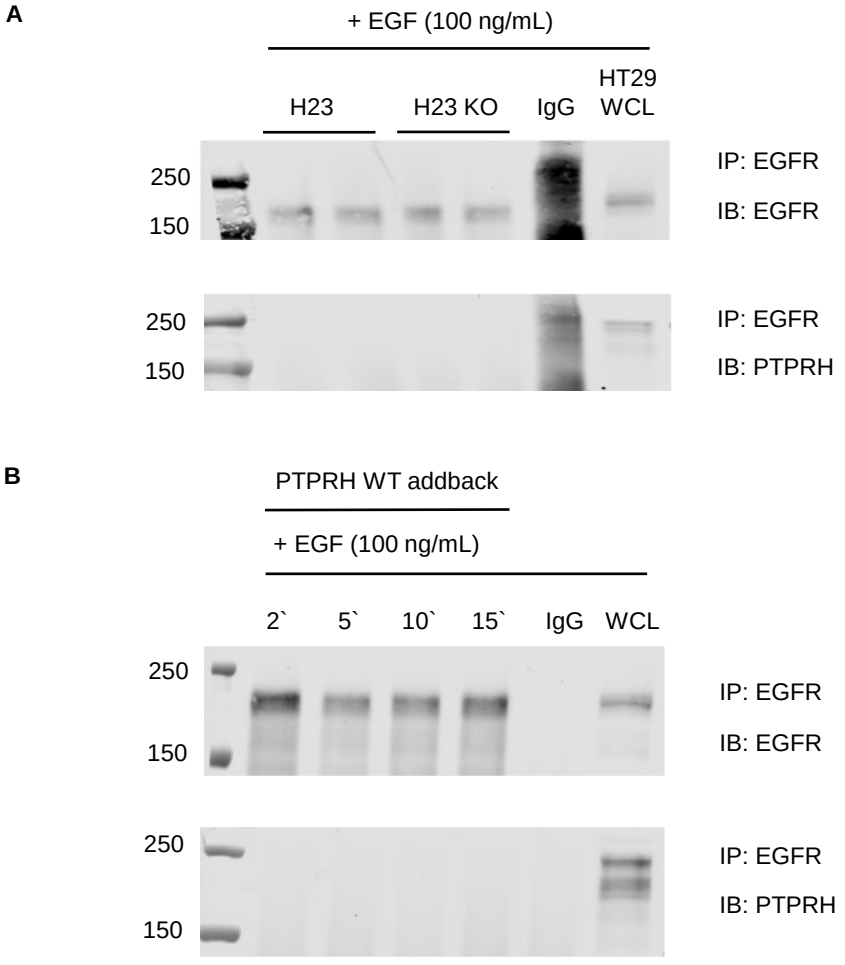


Fig 4: Validation of lung adenocarcinoma cell lines expressing BioID2-only or PTPRH WT-BioID2

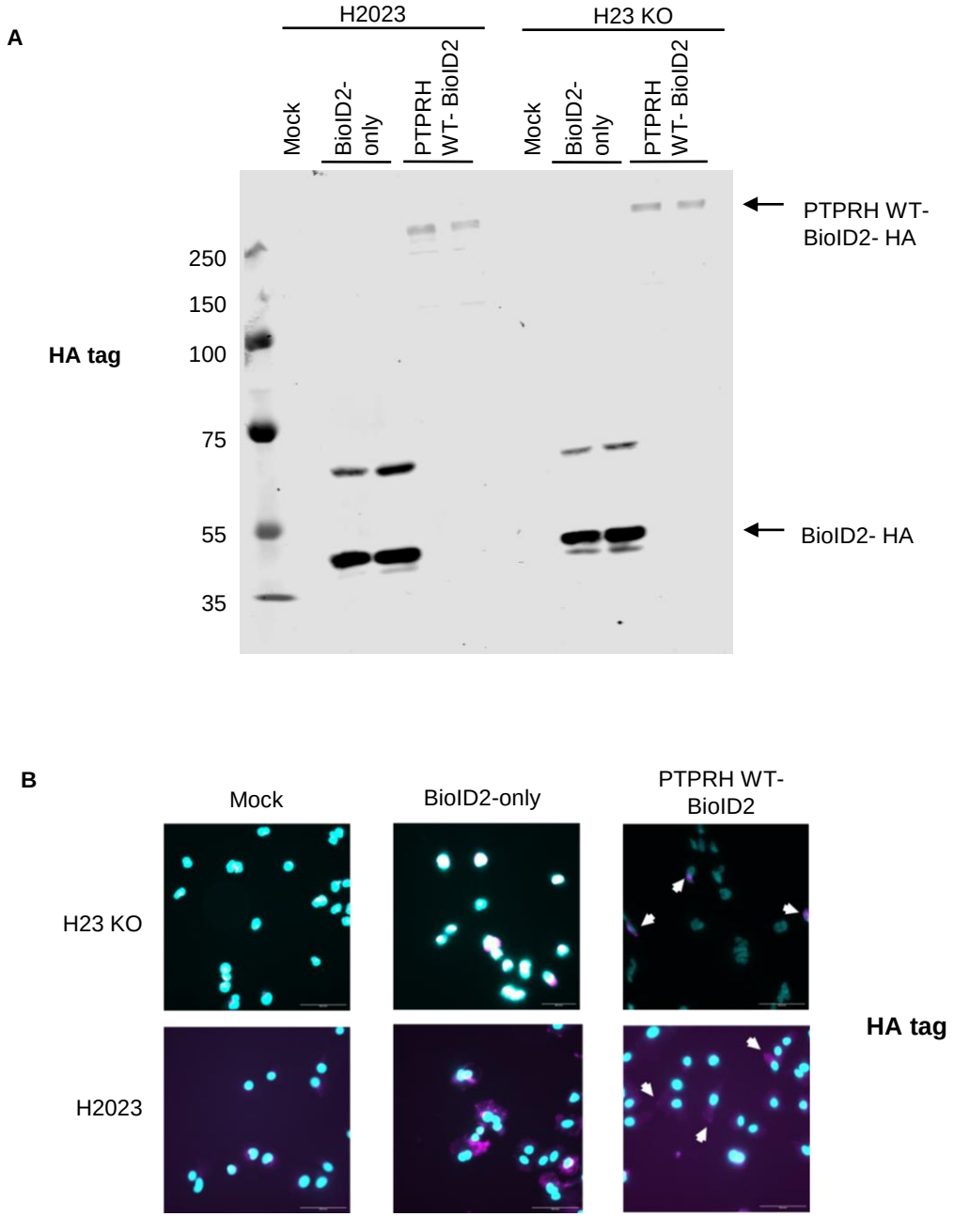


Fig 5: BioID reveals PTPRH interactors

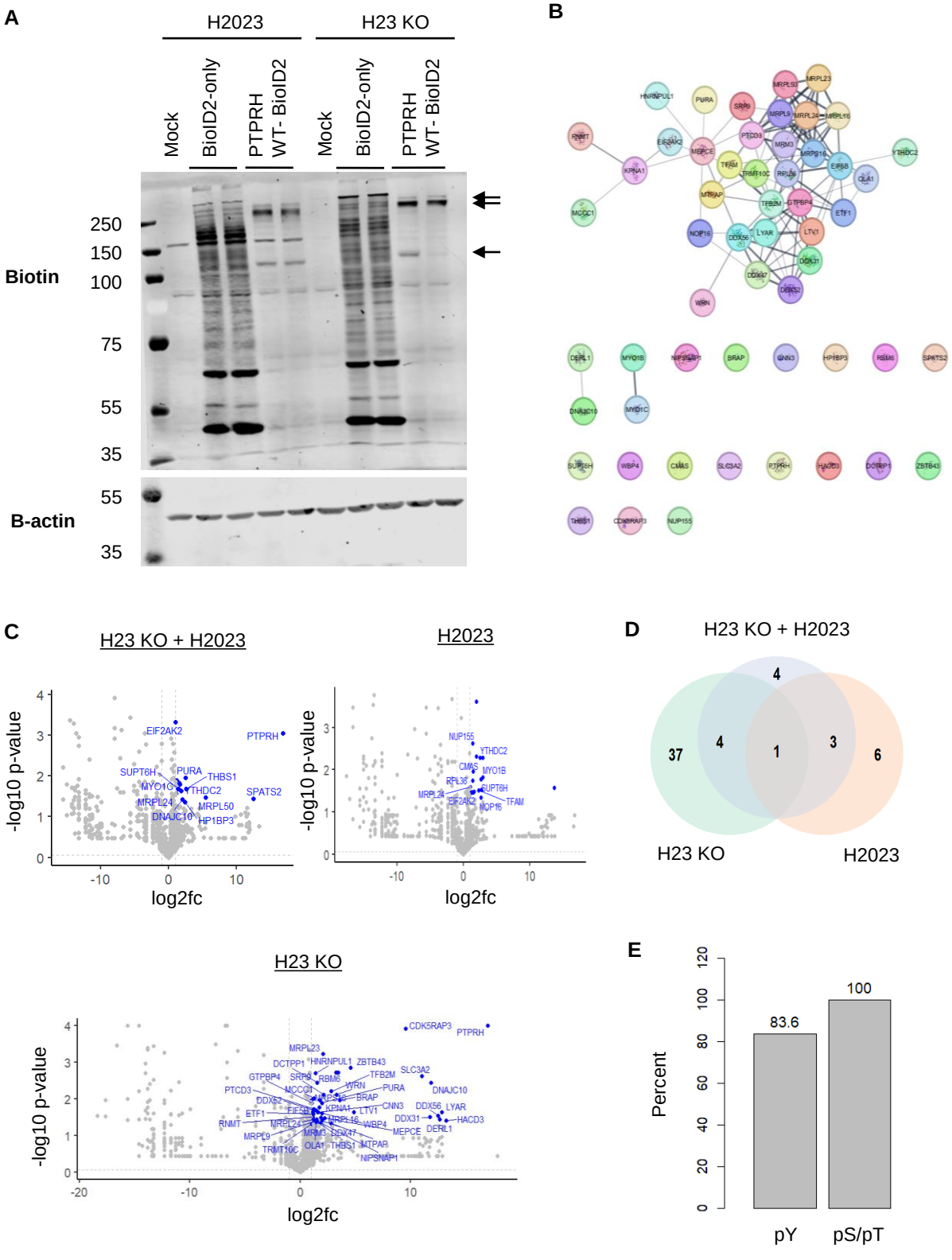
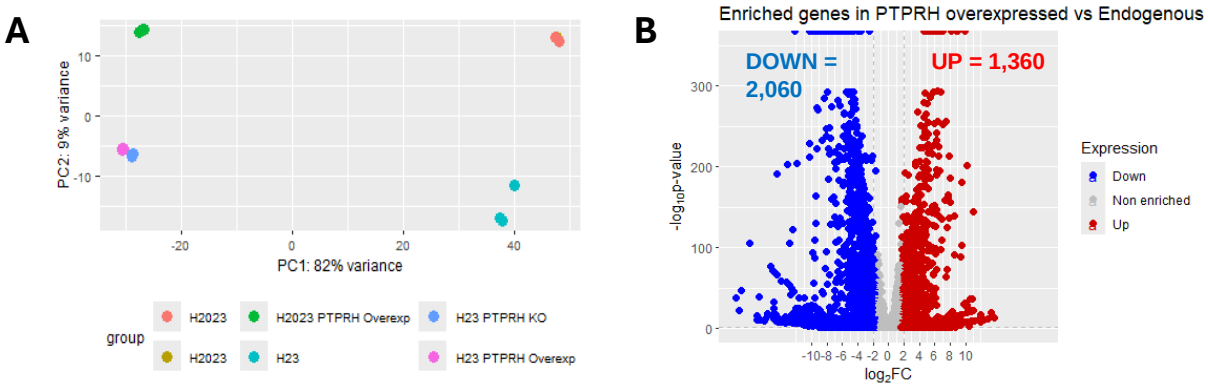
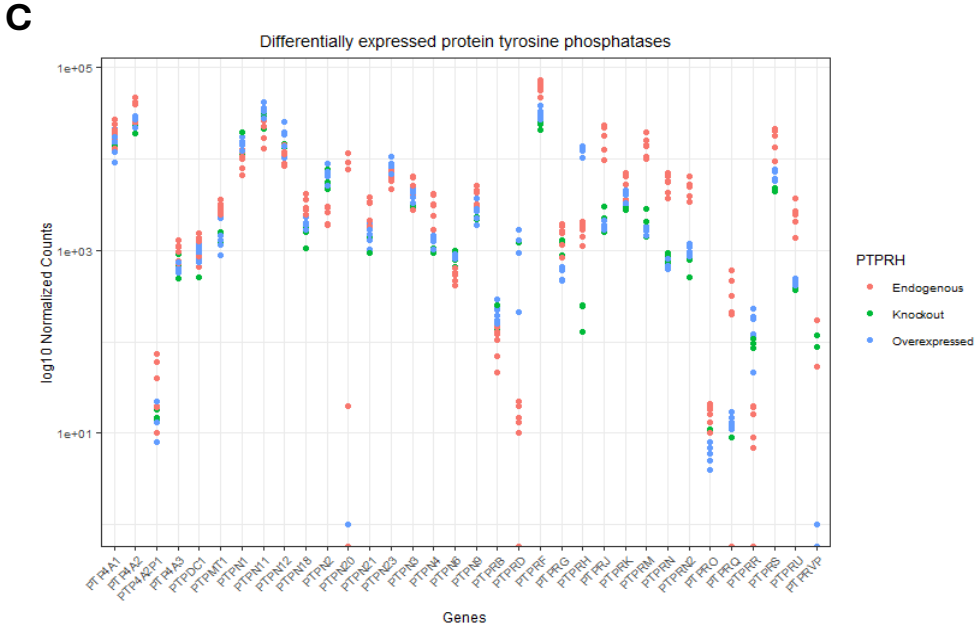


Fig 6: Transcriptome analysis of NSCLC cells overexpressing PTPRH





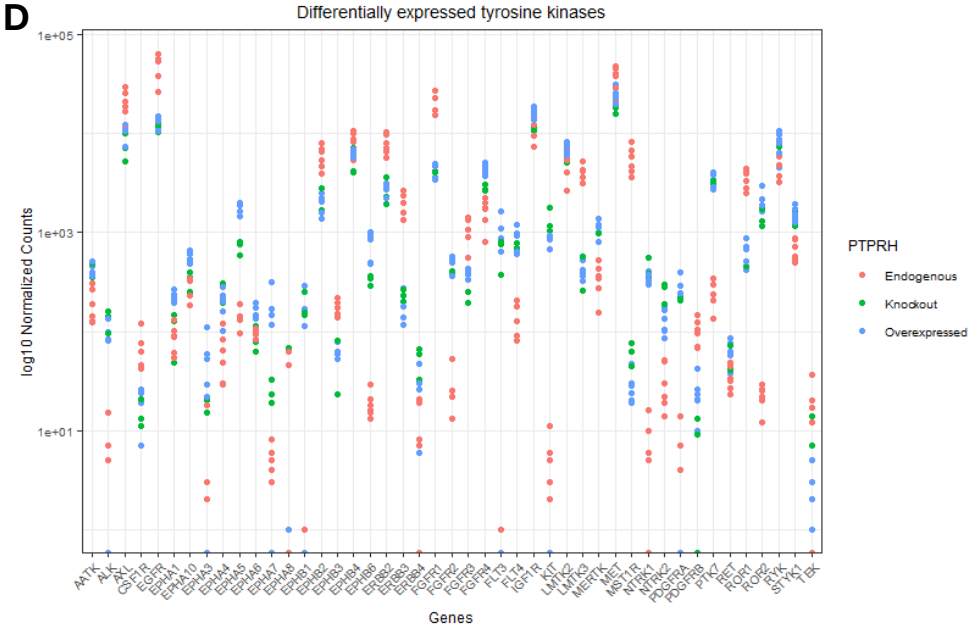
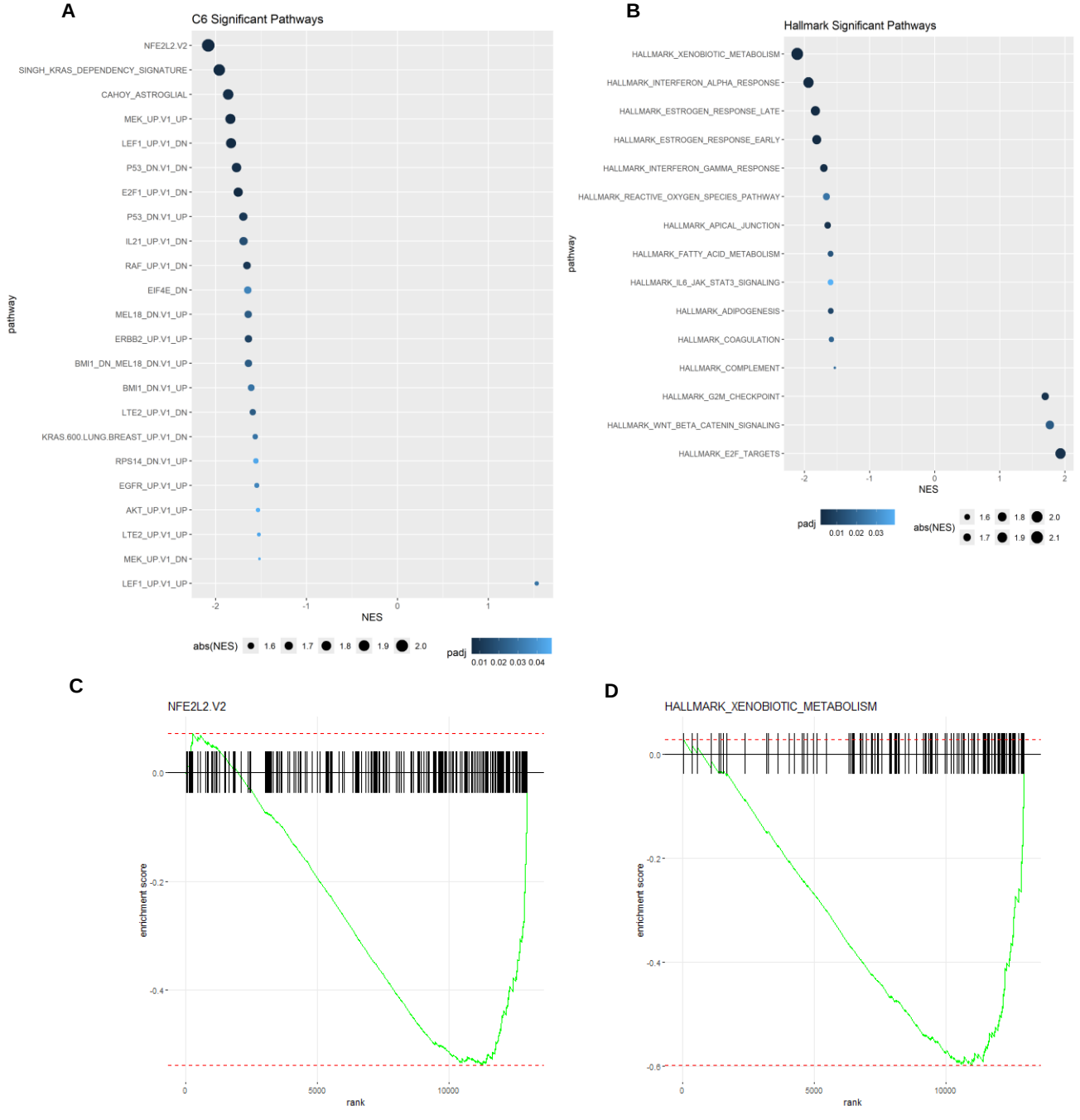



Fig 7: GSEA analysis of lung adenocarcinoma cell lines overexpressing PTPRH