

# SCIENTIFIC REPORTS



OPEN

## Cross-talk between AMPK and EGFR dependent Signaling in Non-Small Cell Lung Cancer

Paurush Praveen<sup>1,2,†</sup>, Helen Hülsmann<sup>3</sup>, Holger Sültmann<sup>3</sup>, Ruprecht Kuner<sup>3,‡</sup> & Holger Fröhlich<sup>1,§</sup>

Received: 24 July 2015

Accepted: 16 May 2016

Published: 09 June 2016

Lung cancers globally account for 12% of new cancer cases, 85% of these being Non Small Cell Lung Cancer (NSCLC). Therapies like erlotinib target the key player EGFR, which is mutated in about 10% of lung adenocarcinoma. However, drug insensitivity and resistance caused by second mutations in the EGFR or aberrant bypass signaling have evolved as a major challenge in controlling these tumors. Recently, AMPK activation was proposed to sensitize NSCLC cells against erlotinib treatment. However, the underlying mechanism is largely unknown. In this work we aim to unravel the interplay between 20 proteins that were previously associated with EGFR signaling and erlotinib drug sensitivity. The inferred network shows a high level of agreement with protein-protein interactions reported in STRING and HIPPIE databases. It is further experimentally validated with protein measurements. Moreover, predictions derived from our network model fairly agree with somatic mutations and gene expression data from primary lung adenocarcinoma. Altogether our results support the role of AMPK in EGFR signaling and drug sensitivity.

Lung cancer is the second most prevalent cancer in both men and women and accounts for 12% of all new cases of cancers reported worldwide<sup>1</sup>. It caused about 1.5 million deaths globally in 2010<sup>2</sup> and is the leading cause for cancer deaths<sup>3</sup>. Non small cell lung cancer (NSCLC) is a category of lung cancer in which malignant cells are formed in lung tissues. About 85% of lung cancers are NSCLC, including 40% lung adenocarcinoma (ADC)<sup>4</sup>. Like most types of cancer, lung ADCs are often perceived as a disease resulting from errant inter and intra-cellular communications manipulated by key signaling molecules. Being a highly heterogeneous malignancy, it is important to understand the etiology and pathogenesis of the disease in order to control and treat lung ADCs.

As a critical, disease relevant factor aberrant activation of EGFR dependent signaling has been implicated in lung ADCs<sup>5,6</sup>. In consequence several monoclonal antibodies against EGFR have been developed. These include gefitinib (Iressa) and erlotinib (Tarceva). Their efficacy is dependent on L858R/Deletion19 mutation<sup>7</sup>. Many of these therapies induce an initial tumor regression. However, in most cases tumors become insensitive to initial therapies and evolve into more aggressive and resistant phenotypes<sup>8,9</sup>. One explanation of the decreased therapeutic benefit is the acquisition of second EGFR mutations, which make cells drug resistant<sup>10</sup>. For example, a T790M mutation occurs in more than 50% of EGFR-mutant lung cancers<sup>11</sup>.

To overcome such treatment failures new targeted therapies need to be developed, possibly within a combinatorial or poly-pharmacological approach<sup>12,13</sup>. The reason is that most likely alternative cell signaling molecules are responsible for drug insensitivity and drug resistance. A recent study has shown that activation of AMPK sensitizes EGFR wildtype H1299 cells and xenografts to erlotinib treatment<sup>14</sup>. This synergistic effect was less obvious in EGFR mutated tumor models that may be due to endogenous AMPK activity and solely EGFR TKI (Tyrosine Kinase Inhibitor) sensitivity. This raises the question about possible molecular mechanisms.

<sup>1</sup>University of Bonn, Bonn-Aachen International Center for IT, Dahlmannstr. 2, Bonn Germany. <sup>2</sup>The Microsoft Research-University of Trento Center for Computational and Systems Biology, 38068 Rovereto, Italy. <sup>3</sup>Cancer Genome Research Group, German Cancer Consortium (DKTK), German Center for Lung Research (DZL), German Cancer Research Center (DKFZ), Im Neuenheimer Feld 460, D-69120 Heidelberg, Germany. <sup>†</sup>Present address: CluPoints SA, Mont-Saint-Guibert, 1435, Belgium. <sup>‡</sup>Present address: TRON—Translational Oncology at the University Medical Center of the Johannes Gutenberg University, Mainz Germany. <sup>§</sup>Present address: UCB Biosciences GmbH, 40789 Monheim Germany. Correspondence and requests for materials should be addressed to P.P. (email: praveen@bit.uni-bonn.de)

	Regulation of apoptic process	Cell death	Cell proliferation	Spindle organization	Cytokinesis	Meiosis	Mitosis	Cell differentiation	Cell cycle	Cell growth	Cell adhesion	Cell signaling	DNA binding
BCL1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>											
EGFR	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>						
ESPL1	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>						
GSK3B	<input checked="" type="checkbox"/>												
PIK3C3					<input checked="" type="checkbox"/>								
PRKAA1	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>					
PRKAB1								<input checked="" type="checkbox"/>					
RAF1	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>						
RPS6KA1	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>					
RPS6KB1	<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>							
SRC	<input checked="" type="checkbox"/>							<input checked="" type="checkbox"/>					
STK11	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>							
TSC1			<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>					
TSC2			<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>					
MTOR							<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>				
TRUB2										<input checked="" type="checkbox"/>			
ITGB4													
GSK3A												<input checked="" type="checkbox"/>	
LEPR												<input checked="" type="checkbox"/>	
WDR3													<input checked="" type="checkbox"/>

**Table 1. List of genes selected for siRNA based knockdown and annotated GO terms (biological processes).** Marked boxes indicate annotation with respective GO terms (full GO annotation in file S5).

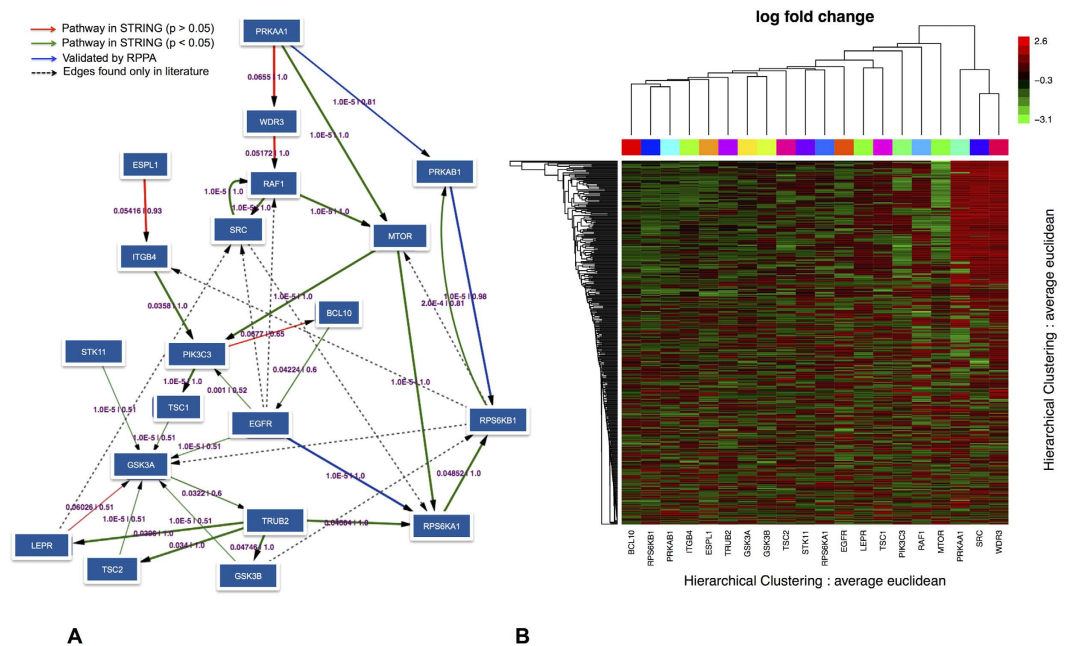
In this study we thus focused on the interplay between the AMPK and EGFR dependent signaling cascades in lung ADC. Accordingly, we reconstructed and subsequently validated a network of 20 genes that were associated with erlotinib response or harbor mutations in lung cancer xenograft models<sup>14</sup>. The approach for network reconstruction is based on single siRNA based knockdowns of each gene in the H1650 cell line (EGFR, delE746-A750) and subsequent gene expression profiling. Based on these data we employed Nested Effects Models (NEMs) as a statistical learning approach to unravel key elements of the interplay between AMPK and EGFR dependent signaling<sup>15</sup>. The resulting network is then validated using protein expression data in cell lines and lung ADC patient data (RNAseq plus somatic mutations) from The Cancer Genome Atlas (TCGA), demonstrating the relevance of our findings.

## Results

**Experimental Data.** A list of 20 genes was compiled, for which protein expression was associated with EGFR TKI response in patient-derived lung cancer xenografts, or mutations have been observed in these models<sup>14</sup>. In more detail, the proteins coding for mTOR, PRKAA1, RAF1, RPS6KA1, and RPS6KB1 were differentially expressed in RPPA analysis of NSCLC xenograft models<sup>14</sup>. For the genes BCL10, ESPL1, ITGB4, LEPR, TRUB2, and WDR3 mutations were identified across the same NSCLC xenografts (unpublished data). The remaining genes EGFR, GSK3A, GSK3B, PIK3C3, PRKAB1, SRC, STK11, TSC1, and TSC2 were individually selected from EGFR/ERBB/AMPK signaling pathways (KEGG<sup>16</sup>).

Table 1 shows a characterization of our 20 genes in terms of their Gene Ontology annotation (biological processes), demonstrating an involvement of all proteins into biological processes that are associated with tumor development and cell growth (see full Gene Ontology annotation in file S5).

Targeted siRNA-based knockdown experiments were performed in triplicates for each of these 20 genes in H1650 cells (EGFR mutated). The efficiency of each gene knockdown with respect to transcript reduction was verified by qPCR to be >70% compared to non-template controls (Figure 1 in S2 and data in S3 and S4). 72 h after siRNA transfection genome-wide expression profiling was performed. The expression measurements were done in two batches, and all experiments were done in triplicates using the Illumina Whole-Genome Gene Expression Bead Array Chips<sup>17</sup>. The data, exported from the Illumina BeadStudio, is available via the GEO (Gene Expression Omnibus) database (GSE69747). The lumi package<sup>18</sup> was used to read in and preprocess the data for further analysis, including quantile normalization. To get rid of the batch effect we used the ComBat function from sva package<sup>19</sup> (Figure 2 in S2).



**Figure 1.** (A) High confidence edges (50% probability cutoff) showing the interplay between PRKAA1/PRKAB1 and EGFR dependent signaling. Edges indicate upstream/downstream relations. They may not correspond to direct, physical protein interactions. Edge thickness is drawn according to the level of edge confidence. Edge labels indicate the p-value according to the comparison with the STRING<sup>22</sup> database (first number) as well as the level of edge stability (from 0 to 1, second number). Green edges indicate a nominally significant ( $p < 0.05$ ) overlap with pathways from STRING (using only edges with confidence  $> 0.5$ ), while red ones do not. Blue edges are validated by protein expression data. Dashed black edges can only be found as protein-protein interaction in STRING, but are not present in our network. An additional complete validation of our network with protein-protein interactions from the HIPPIE database<sup>23</sup> can be found in Figure 5 of supplemental file S2. (B) The heatmap shows the perturbation effects for individual gene silencings: rows = effects observed, columns = genes perturbed. The colors depicts log fold change for corresponding perturbations (green = down-regulation, red = up-regulation).

**Differential Expression Analysis.** Limma<sup>20</sup> was used to analyze expression profiles for each of the 20 gene knockdowns (see details in Materials and Methods). Genes with adjusted p-values  $< 0.05$  and absolute log fold change  $> 1.5$  were declared as differentially expressed. For 13 of 20 genes specific microarray probes reflected a transcript reduction of more than 50%. In total 388 genes (see supplementary file S1) from all 20 gene silencing experiments were retrieved. A heatmap of relative expression changes (log fold changes) of these genes is shown in Fig. 1B. For the following our focus was on the reconstruction of the network between the 20 perturbed genes, given the observed transcriptional responses.

**Network Reconstruction.** After the previously described preprocessing and filtering of gene expression data raw p-values were summarized in a  $20 \times 388$  matrix (Figure 3 in S2) and their densities estimated by fitting a mixture model (see Materials and Methods). These data were used to infer a causal network among the 20 silenced genes (henceforth referred as S-(silenced) genes) via Nested Effects Models (NEMs, see Materials and Methods). NEMs are a statistical learning method that aim at finding the most probable, unobserved signaling network given transcriptional response profiles of individual S-gene knockdowns. Edges in a NEM model indicate upstream/downstream relationships. Prior knowledge from several information sources (Gene Ontology, KEGG pathways, protein-protein interactions, protein domains, protein domain interactions) were used and probabilistically integrated at this step via our previously proposed Noisy-OR model<sup>21</sup>.

The whole NEM network reconstruction was repeated 1000 times while re-sampling from the set of 388 effect reporters (called E-genes in the following) with replacement, yielding a non-parametric bootstrap. The aim of such an approach is to assess the confidence by which individual edges can be learned from data. For the following we focused on edges that could be observed stably with more than 50% probability. The histogram of bootstrap probabilities for individual edges can be found in the supplements (Figure 4 in S2).

As consequence of our analysis we obtained a network (a graph of high confidence edges), which highlighted the possible interplay between EGFR and AMPK dependent signaling [Fig. 1A]. It is worth mentioning that the network reconstruction remained the same, even if different initial filtering criteria to our expression data were applied, indicating the robustness of our network.

**Network is in Agreement with Protein-Protein Interactions Reported in Databases.** We checked the agreement of our inferred network with literature reported and high confidence predictions (confidence value

larger 0.5) of protein-protein interactions taken from the STRING 10.0 database<sup>22</sup>. For this purpose we then mapped our silenced genes to the genome-wide protein-protein interaction networks and calculated shortest paths between them. This way we could find a mapping of each of our inferred edges to a pathway in the STRING database. Remarkably none of these pathways had more than two edges, i.e. there was at most one intermediate protein (Table 1 in S2).

One may ask, in how far such a result could have been expected by chance. In order to address this question we conducted a statistical permutation test: We randomly shuffled the node labels of the protein-protein interaction network 50,000 times. For each permuted network we then re-mapped our silenced genes and re-calculated shortest path distances. We then counted, in how many cases a shortest path between two mapped silenced genes in the permuted network was shorter than in the original one. This resulted into a p-value for each edge (Fig. 1A). The vast majority of the edges in our inferred network showed a significant p-value, thus demonstrating a significant agreement with known protein-protein interactions.

It is worth mentioning that several of the non-significant edges showed a high bootstrap probability (e.g. *WDR3* → *RAF1*). That means they appear as statistically stable patterns in our data, although the underlying biological mechanism cannot be uniquely identified.

We wondered, in how far the observed agreement of our network with STRING was reproducible when using an alternative database, namely HIPPIE<sup>23</sup>. HIPPIE comprises several other databases of literature reported protein-protein interactions, but does not contain any predictions. Only 14 out of our 20 silenced genes could be mapped to HIPPIE. All but two inferred edges between mapable genes showed a significant agreement with pathways in HIPPIE (Figure 5 in S2). Once again these pathways had all a short length (Table 2 in S2). Altogether our analysis thus demonstrates a good agreement of the inferred network with protein-protein interactions reported in databases.

Figure 1A and Figure 5 in S2, also depict direct protein-protein interactions that were found in STRING and HIPPIE, respectively, but show a bootstrap probability below 50% according to our data driven model. The number of these edges is comparably low, indicating a high coverage of reported protein-protein interactions by our network. It should be noted that a possible (but not necessary) reason for not finding a certain protein-protein interaction with high probability by our model is that indeed that interaction does not take place in our studied cell system.

**Interpretation of the Network.** The inferred network presented some interesting relations among genes that have been discussed in literature: PRKAA1 (AMPK Subunit  $\alpha$ ) was found as one of the major upstream regulators of the inferred network. One of the inferred edges is PRKAA1 → WDR3. The WDR gene family has been found to play critical role in cell cycle and apoptosis and gene regulation<sup>24</sup>. WDR3 has been shown to be controlled via P53, and P53 in turn by PRKAA1 in cancer<sup>25</sup>. The inferred edge is thus in agreement with these findings.

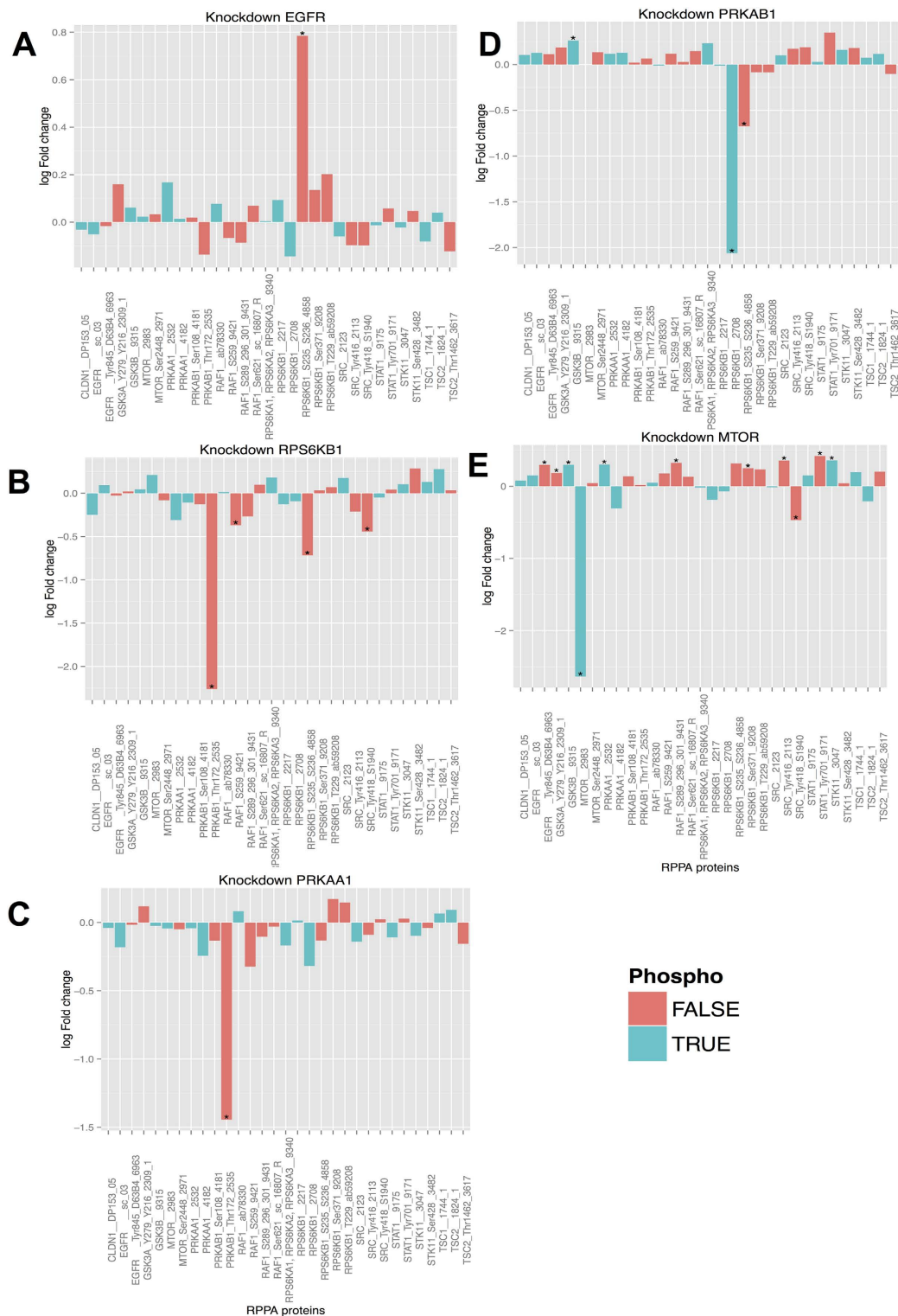
PRKAB1 (AMPK Subunit  $\beta$ ) is a regulatory subunit of the AMP-activated protein kinase, which was inferred to yield identical perturbation effects than PRKAA1 according to our data. These two subunits together influence the ribosomal protein kinase RPS6, which plays an important role in protein synthesis. This is reflected by the path PRKAA1 → PRKAB1 ↔ RPS6KB1 in our network. It has been shown that AMPK can further induce an effect on the RPS6K complex via mTOR in NSCLC<sup>26–28</sup>. Our network contains the path PRKAA1 → mTOR → RPS6KA1 in agreement with these findings.

Subunits of Glycogen synthase kinase-3 (GSK3A and GSK3B) are known for their role in cell division, proliferation, motility and survival together with a number of pathological conditions including cancer<sup>29</sup>. Our network shows GSK3A and GSK3B subunits downstream of EGFR. This is in agreement with the understanding that EGFR stimulation leads to downstream activation of AKT, which in turn influences GSK3<sup>30,31</sup> in normal and cancer cells. The GSK3 complex is activated via phosphorylation by EGFR altering the *Wnt* pathway components in NSCLC<sup>32</sup>. The inferred network maps this in terms of the relationship between the GSK3 complex and EGFR (EGFR → GSK3A ← GSK3B).

EGFR is an upstream activator of PI3K that is frequently altered in cancer<sup>33</sup>. In lung cancers harboring somatic activating mutations in EGFR, PI3-Kinase has been found to be activated by EGFR via a direct binding<sup>34,35</sup>. Our network depicted this in terms of a direct edge between EGFR → PIK3C3.

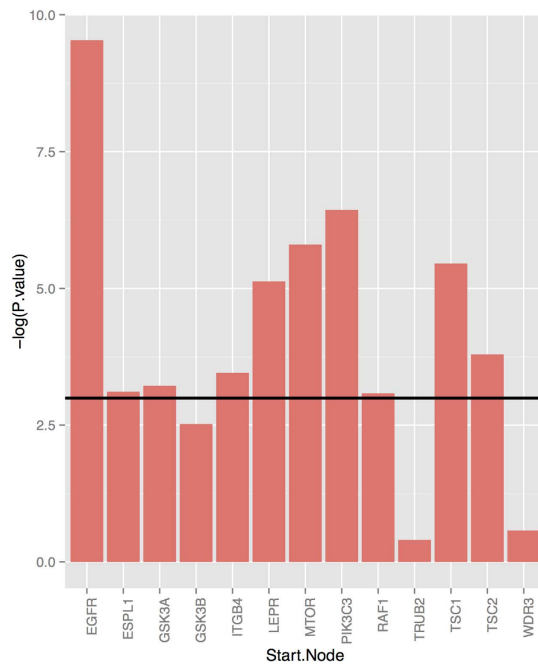
BCL10 is a gene that is known to induce apoptosis and to activate NF- $\kappa$ B<sup>36</sup>. The protein forms a complex with MALT1, which in turn has recently shown to be required for EGFR-induced NF- $\kappa$ B activation and to contribute to EGFR-driven lung cancer progression<sup>37</sup>. RAF1 is a member of the PI3K/AKT signaling cascade, which is known to activate the mTOR pathway<sup>38</sup>. Hence, this is in agreement with our inferred edge RAF1 → mTOR. mTOR itself yields downstream effects on GSK3<sup>39</sup>, which is also consistent with our network.

**Protein Measurements Confirm Inferred Network Features.** The response of 107 phospho and total protein concentrations to the knockdown of five selected network genes (EGFR, PRKAA1, PRKAB1, RPS6KB1, mTOR) was measured via Reverse Phase Protein Arrays (RPPA) (Materials and Methods, [Fig. 2]). The data were analyzed using limma. The knockdown for EGFR yielded an up-regulation of the protein kinase RPS6KB1 (Fig. 2A), which is located downstream of EGFR in our network (Fig. 1). The downstream relation EGFR → RPS6KB1 had been inferred with extremely high stability (100%) in our network and our observation at protein level validates this relation. The second knockdown was for protein kinase RPS6KB1 (p70S6K). RPS6KB1 and PRKAB1 form a cycle in our inferred network. Knockdown of RPS6KB1 yielded a ~four fold down-regulation of PRKAB1 (Fig. 2B). Knockdown of PRKAB1 caused a four fold reduction of RPS6KB1 expression (Fig. 2D). Together this verifies a positive feedback loop between both proteins. Likewise, PRKAA1 knockdown yielded a ~2.8 fold down regulation of PRKAB1 (Fig. 2C), which is in agreement with our inferred edge PRKAA1 → PRKAB1.



**Figure 2. Protein log fold changes after different knockdowns.** The RPPA technique is antibody based, and thus for one and the same protein different identifiers can appear on the x-axis. The employed antibody is indicated as part of the protein name. Bar colors indicate total (red) or phospho-protein (blue) concentration changes. A star (\*) indicates a statistically significant result (FDR < 5%). Effects are given for knockdown of: EGFR (A), RPS6KB (B), PRKAA1 (C), PRKAB1 (D) and mTOR (E).

The knockdown of mTOR yielded statistically significant (FDR < 5%) expression changes of several network proteins (Fig. 2E). However, only the fold change of mTOR itself (around eight fold) was in a range that would usually be regarded as biologically meaningful. Hence, we abstain from interpreting these data further.



**Figure 3. Agreement of inferred network with patient data.** The X-axis shows S-genes, for which somatic mutations in patients were observed. The Y-axis depicts the significances ( $-\log p$ -values) of the overlaps of expected effects on S and E-genes with observed effects (i.e. differentially expressed genes). The horizontal line shows a significance cutoff of 5%.

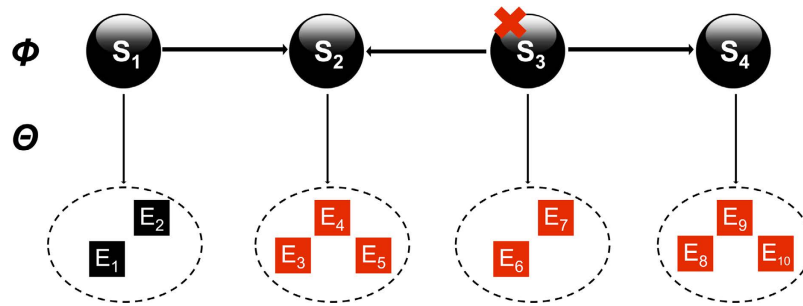
**Network Reconstruction is in Agreement with Patient Data.** In order to establish the further relevance of our reconstructed network The Cancer Genome Atlas (TCGA) database was explored to look for patient samples with somatic mutations in lung adenocarcinoma (file S6), resulting in hits for 13 out of the 20 network genes under study. We extracted RNAseq data for these patients as well as for normal samples ( $n = 5$ ). As expected, the number of patients carrying a specific somatic mutation varied rather significantly. We compared gene expression data of patients with a specific somatic mutation to normal samples using voom and limma<sup>40</sup>. This resulted in 97 to ~4300 differentially expressed genes per somatic mutation (FDR < 5%). The heatmap of the union of these genes is shown in Figure 9 in S2).

In order to check the agreement of these data with our network we collected for each somatic mutation of one of the network genes all downstream genes, including attached transcriptional effect reporters (see details in Material and Methods). These genes thus comprise the effects predicted by our network. We then tested the overlap with these genes to the genes declared as differentially expressed according to our RNAseq data. The significance of the overlap was assessed using a hyper-geometric test. For 10 out of 13 tested somatic mutations, we could in this way establish a significant overlap with  $p$ -value < 0.05 [Fig. 3]. This is a remarkable result given the large heterogeneity of patients, the different number of patients carrying a specific somatic mutation and the fact that combinations of somatic mutations may induce in a specific patient rather complex and non-linear changes of gene expression.

## Discussion

Targeted therapy of EGFR is well established for lung ADC, but frequently fails due to acquired drug resistances. Moreover, the existence of alternative signaling pathways could be responsible for an incomplete drug response or complete treatment failure, raising questions about additional treatment options. In that context, AMPK activation has been recently suggested as a possible way to sensitize cells to erlotinib<sup>14</sup>. Our present study aims to shed light on the possible underlying mechanism by focusing on the interplay between EGFR and AMPK dependent signaling cascades. We used siRNA based gene knockdowns together with gene expression profiling to gain insights into the network between 20 proteins, which according to previous experiments showed significant expression changes after AMPK activation and erlotinib treatment. Application of probabilistic network inference using Nested Effects Models yielded a number of high confidence edges, which are largely in agreement with the common literature and further experimental validation data, including protein measurements after siRNA perturbation. Remarkably, our network is also in agreement with patient gene expression and somatic mutation data.

A limitation of our network model is the restriction to 20 selected proteins. While the prior relevance of these proteins has been implicated in previous experiments in model systems<sup>14</sup>, their actual involvement into the human disease pathology is not guaranteed. Also other important proteins might play a role, which are not included into our network. In addition, all experimental data were generated in the context of the cell line H1650, its genome and its activation status of signaling molecules at a specific time point after perturbation. Further studies should focus on the dynamical aspect of signaling cascades by measuring data at several time points.



**Figure 4. Idea behind Nested Effects Models.** Perturbation of S-gene S3 is expected to yield measurable downstream effects on transcriptional reporters E6 and E7. Since S3 is at the same time upstream of S2 and S4, reporters attached to these S-genes are predicted to show an effect. On the other hand, perturbation of S4 does only effect reporters E8, E9, E10. NEM structure learning aims for reconstructing the wiring of the S-genes based on the observable effects on downstream reporters. The attachment of individual reporter genes is treated as an unknown nuisance parameter, which is integrated out, i.e. the network score averages over all possible attachment positions of individual reporters.

In summary, our approach allowed us to shed light on the interplay between EGFR and AMPK dependent signaling in a NSCLC model system.

## Materials and Methods

**Cell culture, Treatment and siRNA-based Gene Knockdowns.** H1650 cells ( $2 \times 10^5$ ) were cultured in 2 ml RPMI-1640 medium on 6-well plates and transfected after 24 hours growth using Lipofectamine RNAiMAX Reagent (Life Technologies). Gene-specific siRNAs pools (Dharmacon) and non-template controls (40 nM) were transfected and harvested after 48 h (triplicates). RNA was isolated using the RNeasy Plus 96 Kit according to the manufacturer's instructions. mRNA gene levels were determined by qPCR using Taqman assays (Life Technologies) and the Lightcycler system (Roche). Knockdown efficiency was determined by delta-delta Ct method using internal control genes (ESD and B2M) and the non-template control as reference sample. siRNAs and qPCR assays are shown in files S3 and S4.

**Reverse Phase Protein Array.** H1650 protein lysates upon 72 h gene knockdown (PRKAA1, PRKAB1, EGFR, MTOR and RPS6KB1) were analyzed via Reverse Phase Protein Array as previously described<sup>41</sup>. Quantitative data is given in supplementary data (file S2, Figure 8).

**Limma Analysis.** Limma is a linear modeling based technique, which employs an empirical Bayes estimate of the gene wise variance<sup>20</sup>. In this article gene expression was measured in triplicates in a control situation and after siRNA knockdown of a specific gene. Accordingly, a linear model with one factor describing the sample type (control, knockdown of gene 1, knockdown of gene 2, ...) was fitted. Afterwards the contrasts to the control group were extracted.

**Estimating Perturbation Effects.** P-values reflect the probability to obtain a test statistic at least as extreme as the one observed, given the null hypothesis is true. However, for our network reconstruction we are interested to quantify for each transcript the probability for its perturbation, which is a different quantity. In order to estimate these transcript specific perturbation effects a Beta-Uniform Mixture model (BUM) can be fitted to the observed p-value distributions<sup>42</sup>. The BUM model decomposes the observed p-value distribution into a uniform part (the null distribution) and a second part (the alternative distribution), represented itself by two Beta distributions. Based on the alternative distribution it is now possible to estimate the transcript specific perturbation effect. After applying this technique to each of our 20 gene silencing experiments, we arrived at a  $20 \times 388$  matrix of p-value densities (Figure 3 in S2), which served as the input for the NEM inference explained in the next section.

**Nested Effects Model (NEM).** NEMs (Nested Effects Models) are a class of probabilistic graphical models used to reconstruct a pathway based on measurable downstream response profiles of individual gene knockdowns<sup>15,43</sup>. Gene silencing typically alters steady-state expression/activity levels of downstream transcripts<sup>44,45</sup>. Accordingly, a NEM has two components: the unknown and not directly observable system of silenced proteins (called S-genes) and the connection of these proteins to downstream effect reporters (E-genes, Fig. 4). NEM network inference is based on a data matrix  $D$  where  $D_{sk}$  is the effect on E-gene  $k$  for the knockdown of S-gene  $s$ . In our case the probabilities for these effects are quantified via p-values densities (see above). Based on the given data for any candidate network structure  $\mathcal{M}$  it is possible to calculate a marginal log-likelihood in polynomial time complexity<sup>42,43</sup>. This marginal log-likelihood serves as a score for each network structure.

The network score may degrade, if reporter genes are uninformative. We therefore applied a trick suggested by Tresch and Markowitz<sup>46</sup>. More specifically, an isolated, virtual S-gene *null* is added to the network structure. It

has been demonstrated that this *null* node filters out irrelevant reporters effectively, because irrelevant reporters have a higher attachment likelihood to *null* than to any other S-gene.

Since the space of hypothetically possible network structures grows at least exponentially with the number of S-genes, heuristic algorithms are required to scan the space of possible candidate structures effectively. In our case we employed the Module Network learning algorithm<sup>42</sup>. The algorithm first splits the overall set of S-genes into smaller groups in a data driven manner. For these groups of S-genes the optimal network structure is found via exhaustive search. In a final step identified sub-networks (modules) are connected using a local search strategy.

The search in network space can be enhanced by incorporating prior knowledge. More specifically, the method proposed earlier by one of the authors<sup>42</sup> allows to specify for each edge an a priori probability. Here these prior probabilities were estimated from several information sources using the Noisy-OR approach suggested in Praveen and Föhlich<sup>21</sup>. Details about the individual employed information sources can be found in text S2.

The output of the NEM learning procedure consists of a network between S-genes as well as estimated attachments of downstream effect reporters (E-genes). For the data used in this article estimated attachment probabilities are shown as a heatmap in Figure 6 in S2. Figure 7 in S2 gives another view on the data by depicting perturbation effects for each knockdown grouped by their most likely S-gene attachment. The plot specifically shows a larger group of reporters, which are most likely uninformative, i.e. assigned to *null*.

**Predicting Perturbation Effects with NEMs.** The NEM models allows for qualitatively predicting the effect of a knockdown of each of the S-genes. According to the NEM model perturbation of a particular S-gene S is supposed to affect.

1. All S-genes reachable from S via directed edges, including S itself.
2. All E-genes attached to these S-genes.

Figure 4 depicts an example of the effects predicted by the silencing of S3 in a toy example with four S-genes.

## References

1. van Meerbeeck, J. P., Fennell, D. A. & De Ruyscher, D. K. Small-cell lung cancer. *Lancet* **378**, 1741–1755 (2011).
2. Report, L. Lung cancer: a global scourge. *Lancet* **382**, 659 (2013).
3. Jemal, A. *et al.* Global cancer statistics. *CA: A Cancer Journal for Clinicians* **61**, 69–90 (2011).
4. Reck, M., Heigener, D. F., Mok, T., Soria, J.-C. & Rabe, K. F. Management of non-small-cell lung cancer: recent developments. *The Lancet* **382**, 709–719 (2013).
5. Atlas, T. C. G. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
6. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107–1120 (2012).
7. D'Angelo, S. P. *et al.* Incidence of egfr exon 19 deletions and l858r in tumor specimens from men and cigarette smokers with lung adenocarcinomas. *Journal of Clinical Oncology* **29**, 2066–2070 (2011).
8. Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nature reviews. Cancer* **13**, 714–26 (2013).
9. Mumenthaler, S. M. *et al.* Evolutionary modeling of combination treatment strategies to overcome resistance to tyrosine kinase inhibitors in non-small cell lung cancer. *Molecular pharmacology* **8**, 2069–79 (2011).
10. Chong, C. R. & Jänne, P. a. The quest to overcome resistance to egfr-targeted therapies in cancer. *Nature medicine* **19**, 1389–400 (2013).
11. Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the egfr kinase domain. *PLoS Med* **2**, doi: 10.1371/journal.pmed.0020073 (2005).
12. Innocenti, F. Polypharmacology in drug discovery. *Clin Pharmacol Ther* **92**, 279–280 (2012).
13. Al-Lazikani, B., Banerji, U. & Workman, P. Combinatorial drug therapy for cancer in the post-genomic era. *Nat Biotech* **30**, 679–692 (2012).
14. Hülsmann, H. J. *et al.* Activation of amp-activated protein kinase sensitizes lung cancer cells and h1299 xenografts to erlotinib. *Lung Cancer* **86**, 151–157 (2015).
15. Markowitz, F. *Probabilistic Models for Gene Silencing Data*. Ph.D. thesis, Free University of Berlin (2005).
16. Kanehisa, M. *et al.* Kegg for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484 (2008).
17. Chen, J., April, C. S. & Fan, J.-B. Mirna expression profiling using illumina universal beadchips. *Next-Generation MicroRNA Expression Profiling Technology* **822**, doi: 10.1007/978-1-61779-427-8\_7 (2012).
18. Du, P., Kibbe, W. A. & Lin, S. M. Lumi: a pipeline for processing illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).
19. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics* **8**, 118–127 (2007).
20. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**, 1–25 (2004).
21. Praveen, P. & Fröhlich, H. Boosting probabilistic graphical model inference by incorporating prior knowledge from multiple sources. *PLoS ONE* **8**, doi: 10.1371/journal.pone.0067410 (2013).
22. Szklarczyk, D. *et al.* String v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, doi: 10.1093/nar/gku1003 (2014).
23. Schaefer, M. H. *et al.* Hippie: Integrating protein interaction networks with experiment based quality scores. *PLoS one.* **7**, doi: 10.1371/journal.pone.0031826 (2012).
24. Clemen, C., Eichinger, L., Rybakina, V. & Smith, T. *Subcellular Biochemistry*, vol. 48, 20–30 (Springer, New York, 2008).
25. McMahon, M., Ayll, V., Panov, K. I. & O'Connor, R. Ribosomal 18s rna processing by the igf-i-responsive wdr3 protein is integrated with p53 function in cancer cell proliferation. *Journal of Biological Chemistry* doi: 10.1074/jbc.M110.108555 (2010).
26. Hardie, D. & Alessi, D. Lkb1 and ampk and the cancer-metabolism link - ten years after. *BMC Biology* **11**, doi: 10.1186/1741-7007-11-36 (2013).
27. Shackelford, D. B. & Shaw, R. J. The lkb1-ampk pathway: metabolism and growth control in tumour suppression. *Nat Rev Cancer* **9**, 563–575 (2009).
28. Han, S., Khuri, F. R. & Roman, J. Fibronectin stimulates non-small cell lung carcinoma cell growth through activation of akt/mammalian target of rapamycin/s6 kinase and inactivation of lkb1/amp-activated protein kinase signal pathways. *Cancer Research* **66**, 315–323 (2006).
29. Forde, J. E. & Dale, T. C. *Glycogen synthase kinase 3: A key regulator of cellular fate* **64**, 1930–1944 (2007).



30. Scaltriti, M. & Baselga, J. The epidermal growth factor receptor pathway: a model for targeted therapy. *Clinical Cancer Research* **12**, 5268–5272 (2006).
31. Matsuda, S., Nakanishi, A., Wada, Y. & Kitagishi, Y. Roles of pi3k/akt/pten pathway as a target for pharmaceutical therapy. *The open medicinal chemistry journal* **7**, 23–29 (2013).
32. David J. Stewart. Wnt Signaling Pathway in Non-Small Cell Lung Cancer. *Journal of the National Cancer Institute* **106**(1), doi: 10.1093/jnci/djt356 (2014).
33. Yuan, T. L. & Cantley, L. C. PI3K pathway alterations in cancer: variations on a theme. *Oncogene* **27**, 5497–5510 (2008).
34. Law, C. *et al.* Cloning of PI3Kinase-associated p85 utilizing a novel method for expression/cloning of target proteins for receptor tyrosine kinases. *Cell* **65**, 83–90 (1991).
35. Engelman *et al.* MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* **316**, 1039–1043 (2007).
36. Willis, T. G. *et al.* Bcl10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and mutated in multiple tumor types. *Cell* **96**, 35–45 (1999).
37. Pan, D. *et al.* MALT1 is required for EGFR-induced NF- $\kappa$ B activation and contributes to EGFR-driven lung cancer progression. *Oncogene* **35**, 919–928 (2016).
38. Rafalski, V. A. & Brunet, A. Energy metabolism in adult neural stem cell fate. *Progress in Neurobiology* **93**, 182–203 (2011).
39. Ojeda, L. *et al.* Critical role of PI3K/Akt/GSK3 $\beta$  in motoneuron specification from human neural stem cells in response to FGF2 and EGF. *PLoS ONE* **6**(8), e23414, doi: 10.1371/journal.pone.0023414 (2011).
40. Law, C., Chen, Y., Shi, W. & Smyth, G. Voom: precision weights unlock linear model analysis tools for rna-seq read counts. *Genome Biology* **15**, doi: 10.1186/gb-2014-15-2-r29 (2014).
41. Loebke, C. *et al.* Infrared-based protein detection arrays for quantitative proteomics. *Proteomics* **7**, 558–564 (2007).
42. Fröhlich, H., Fellmann, M., Sülthmann, H., Poustka, A. & Beissbarth, T. Estimating large-scale signaling networks through nested effect models with intervention effects from microarray data. *Bioinformatics* **24**, 2650–2656 (2008).
43. Fröhlich, H., Fellman, M., Sülthmann, H., Poustka, A. & Beissbarth, T. Large scale statistical inference of signaling pathways from rna and microarray data. *BMC Bioinformatics* **8**, doi: 10.1186/1471-2105-8-386 (2007).
44. Ideker, T. E., Thorsson, V. & Karp, R. M. Discovery of regulatory interactions through perturbation: Inference and experimental design. *Proceedings of the Pacific Symposium on Biocomputing* **5**, 302–313 (World Scientific Press, 2000).
45. Badaloni, S., Camillo, B. D. & Sambo, F. Qualitative reasoning for biological network inference from systematic perturbation experiments. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **9**, 1482–1491 (2012).
46. Tresch, A. & Markowitz, F. Structure learning in nested effects models. *Statistical Applications in Genetics and Molecular Biology* **7**(1), doi: 10.2202/1544-6115.1332 (2008).

## Acknowledgements

Parts of the presented results used data generated by the TCGA Research Network: <http://cancergenome.nih.gov/>. PP was partially supported in part by a the B-IT research school (NRW graduate school). HH, RK and HS were supported by the German Federal Ministry of Education and Research BMBF including the grants PREDICT (0315428C), and the BMBF grant German Center for Lung Research (82DZL00404).

## Author Contributions

Conceived and designed the experiments: P.P. and H.F. Performed the knockdown experiments: H.H. Analyzed the data: P.P., H.F. and R.K. Contributed reagents/materials/analysis tools: P.P., H.H., R.K., H.S. and H.F. Wrote the paper: P.P., H.F., R.K. and H.S.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Praveen, P. *et al.* Cross-talk between AMPK and EGFR dependent Signaling in Non-Small Cell Lung Cancer. *Sci. Rep.* **6**, 27514; doi: 10.1038/srep27514 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>