1 **Elucidating synergistic dependencies in lung adenocarcinoma by**
2 **proteome-wide signaling-network analysis**

3

4 **Mukesh Bansal[1,2,*#], Jing He[3,4,5, *], Michael Peyton[6*], Manjunath Kaustagi[3], Archana Iyer[3],**

5 **Michael Comb[10], Michael White[6], John Minna[11], Andrea Califano[3,4,5,7,8,9,#]**

6 [1] Psychogenics Inc. Tarrytown, New York, USA

7 [2] Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York

8 [3] Department of Systems Biology, Columbia University, New York, NY

9 [4] Center for Computational Biology and Bioinformatics (C2B2), Columbia University, New York, NY, USA

10 [5] Department of Biomedical Informatics (DBMI), Columbia University, New York, NY, USA

11 [6] Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

12 [7] Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY

13 [8] Institute for Cancer Genetics, Columbia University, New York, NY

14 [9] Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY

15 [10] Cell Signaling Technology, 3 Trask Lane, Danvers, MA 01923, USA

16 [11] Hamon Center for Therapeutic Oncology Research, Simmons Comprehensive Cancer Center,

17 Departments of Pharmacology, and Internal Medicine, University of Texas Southwestern Medical Center,

18 Dallas, Texas

19

20 * These authors contributed equally

21 # Correspondence should be address to M.B (mukesh.bansal@psychogenics.com); A.C.

22 (ac2248@cumc.columbia.edu)

Summary

Signaling pathway models are largely based on the compilation of literature data from heterogeneous cellular contexts. Indeed, *de novo* reconstruction of signaling interactions from large-scale molecular profiling is still lagging, compared to similar efforts in transcriptional and protein-protein interaction networks. To address this challenge, we introduce a novel algorithm for the systematic inference of protein kinase pathways, and applied it to published mass spectrometry-based phosphotyrosine profile data from 250 lung adenocarcinoma (LUAD) samples. The resulting network includes 43 TKs and 415 inferred, LUAD-specific substrates, which were validated at >60% accuracy by SILAC assays, including "novel' substrates of the EGFR and c-MET TKs, which play a critical oncogenic role in lung cancer. This systematic, data-driven model supported drug response prediction on an individual sample basis, including accurate prediction and validation of synergistic EGFR and c-MET inhibitor activity in cells lacking mutations in either gene, thus contributing to current precision oncology efforts.

Introduction

40 Lung adenocarcinoma (LUAD) is a leading cause of cancer related deaths in United States,

41 representing 40% of 225,500 new lung cancer cases every year, and has a 5-year survival rate

42 of only 16 % (1). Excluding immunotherapeutic agents, which have recently shown significant

43 success in a relatively small subset of patients (2), the most effective targeted therapies for this

44 diseases were designed to inhibit tyrosine kinase proteins harboring genetic alterations that

45 induce aberrant activation of downstream pathways (3-7). These the most frequent such

46 actionable alterations include EGFR mutations and ALK-EML4 fusion events, in ~15% and ~3-

47 7% of LUAD patients, respectively (8, 9). Yet, while targeted therapy is initially effective in a

48 significant fraction of tumors harboring these genetic alterations, the vast majority of treated

49 patients will either fail to respond or will develop resistance to mono-therapy (10, 11). In addition,

50 most patient lack actionable alterations altogether. This suggests that novel approaches are

51 critically needed.

52 A possible alternative to minimize emergence of resistance is combination therapy, a strategy

53 that has been shown to be effective in many metastatic tumors, such as breast cancer and acute

54 myeloid leukemia (12-14). However, systematic identification of effective drug combinations on a

55 genetic alteration basis is difficult, because the number of patients presenting multiple actionable

56 events is extremely low. As a result, combination therapy is generally hypothesized and tested on

57 an empirical basis or based on elucidation of complex mechanisms of tumor cell adaptation. In

58 addition, accurate prediction of response to available mono-therapy – including to EGFR inhibitors

59 – in patients lacking any genetic alteration represents an equally relevant challenge, especially

60 since a small fraction of EGFR^{WT} patients have been shown to respond to Afatinib, even though

61 a predictive biomarker is not available. To address these limitations, we and other have proposed

62 that rational design of combination therapy and the identification of critical targetable

3

63 dependencies may require a more mechanistic and tumor-context-specific understanding of the

64 molecular interactions that underlie their potential synergistic activity, starting with tyrosine

65 kinases, which represent a critical class of pharmacological targets in cancer (15). Such an

66 approach requires methodologies for the accurate and systematic elucidation of tumor-specific

67 signaling transduction pathways.

68 Dissection of signal transduction networks represents a complex endeavor, requiring elucidation

69 of hundreds of thousands of tissue-specific molecular interactions that mediate the post-

70 translational modification of protein substrates. *In vitro* approaches generally fail to capture the

71 tissue-specific nature of these interactions, thus providing "average" signal transduction pathways

72 that are both incomplete and inaccurate. In addition, experimental approaches that have been

73 successful in accelerating the analysis of molecular interactions in transcriptional regulation and

74 protein-protein interaction in stable-complexes, such as those based on co-expression or yeast-

75 2-hybrid assays, do not easily translate to elucidating signaling interactions. Similarly, approaches

76 based on the use of phospho-specific antibodies, while elegant and effective, are limited to only

77 a handful of proteins. Computationally, compared to the many algorithms that have been

78 developed for the reverse engineering of transcriptional and protein-complex interactions (16, 17),

79 only a handful of experimentally validated algorithms are available for the dissection of signaling

80 networks, none of which works at the proteome-wide level or is tumor-context specific (16, 18,

81 19).

82 Recent availability of proteome-wide molecular profile data, characterizing the abundance of

83 phospho-tyrosine-enriched peptides by liquid chromatography coupled to tandem mass

84 spectrometry (LC-MS/MS), suggests that additional methodologies may be developed to extend

85 approaches that have been successfully applied to the dissection of transcriptional networks from

86 gene expression profiling. In this manuscript, we propose extending the Algorithm for the

87 Reconstruction of Accurate Cellular Networks (ARACNe) (20) for the reverse engineering of

88 signal transduction networks from large-scale phosphoproteomic profiles. The new method,

89 *pARACNe*, addresses critical issues that prevented the direct application of the original ARACNe

90 algorithm on phosphoproteomic profile data. Briefly, the new algorithm addresses critical

91 computational challenges presented by LC-MS/MS and spectral counting data, while

92 incorporating enzymatic signaling characteristics into the algorithm design. In particular,

93 pARACNE is designed to handle three critical issues resulting from the use of LC-MS/MS assays,

94 including the highly sparse nature of phosphopeptide abundance data, the large amount of noise

95 and missing data, and the degenerate peptides-to-protein mapping.

96 We applied pARACNE to the analysis of previously published, genome-wide phosphoproteomic

97 data from 245 lung adenocarcinoma (LUAD) samples, including 151 fresh-frozen biopsies, 46 cell

98 lines, as well as 48 normal lung tissues. The resulting network comprised 46 tyrosine kinases

99 (TK) densely connected with 415 candidate substrates (including 377 proteins lacking any TK

100 activity), representing the first genome-wide, tumor-context-specific model for a TK signal

101 transduction network, capturing both protein-specific and phospho-site specific events. We

102 validated substrate predictions for two "hubs," whose activity may play a key role in determining

103 sensitivity to Erlotinib and Crizotinib, two FDA-approved drugs for LUAD, including the EGFR and

104 c-MET tyrosine kinases by independent SILAC assays and database analysis, with >60%

105 accuracy. Of particular note, the inferred TK-substrate network provided unique information about

106 tyrosine kinase auto-phosphorylation events, either direct (cis) or via a second kinase (trans).

107 Analysis of the resulting TK-network – by extending the VIPER (Virtual Proteomics by Enriched

108 Regulon analysis) algorithm (21), an established method for the inference of Master Regulator

109 proteins – recapitulated established genetic determinants of LUAD and was effective in predicting

110 sensitivity to Erlotinib and Crizotinib combination therapy. Predicted sensitivities were validated

111 in an independent set of LUAD cell lines, the majority of which harbored no genetic alterations in

112 the corresponding genes. Furthermore, predictions based on the analysis of the corresponding

113   patient cohort were strongly supported by genomic information, suggesting potential value in

114   using these analyses for the identification of effective combination therapies in precision oncology.

115

Results

117   ***Overview of the pARACNe Algorithm***

118   Enzymatic activity of tyrosine kinase (TK) proteins – as assessed by the ability to phosphorylate

119   their downstream substrates – is effectively determined by their phosphorylated isoform

120   abundance (**Fig. 1A, B**). Therefore, we reasoned that computational inference of TK-substrate

121   interactions (**TK→S**) could be effectively performed by measuring dependencies between their

122   respective phospho-states by mutual information analysis (22) over a large sample compendium

123   (**Fig. 1C**). Unfortunately, due to signal transduction cascade complexity and pathway cross-talk,

124   such dependencies can manifest between protein pairs that are not involved in direct **TK→S**

125   interactions. The ARACNe algorithm – previously designed for the reverse engineering of

126   transcriptional networks – effectively addresses this problem by leveraging the Data Processing

127   Inequality (23). This is a critical property of the mutual information that effectively allows

128   disambiguating between direct and indirect interactions by assessing whether information transfer

129   on any candidate direct interaction (e.g., $TK_1→S$) is greater than transfer on every other indirect

130   path (e.g., $TK_1→TK_2→S$). ARACNe has been highly successful in the experimentally validated

131   dissection of transcriptional networks via analysis of large gene expression profile compendia.

132   ARACNe-inferred targets of transcription factors were validated in multiple cellular contexts, with

133   an accuracy of 70% to 80% (20, 24-27).

134   However, ARACNe relies on molecular profile data that is both continuous and non-sparse,

135   properties that are not always provided by quantitative proteomic data sets, which can be

136   generated by a variety of methods. Those based on LC-MS/MS represent the most popular

137   approaches (28), but different implementations have specific performance profiles in terms of

138   analyte throughput, consistency of measurement of peptides across samples and linear dynamic

139   range (29). Depending on the data acquisition method, one or both of these assumptions of

140   ARACNe are violated in proteome-wide datasets generated by the most popular methods based

141   on data-dependent acquisition. Particularly when employing quantification by spectral counting,

142   as is typically conducted for global protein-protein interaction studies (30, 31), phosphoproteomic

143   data is both discrete (i.e., generally represented by spectral counts) and very sparse, with a

144   majority of peptides having zero spectral counts and presenting a significantly skewed distribution

145   for low-abundance peptides.

146

147   To address these limitations, we developed a phospho-proteomic specific algorithm, pARACNe

148   (phospho-ARACNe) (**Fig. 1C**), specifically designed to measure phospho-state dependencies

149   between TKs and their candidate substrates from large-scale LC-MS/MS phosphoproteomic

150   profiles. pARACNe thus extends the original ARACNe framework to allow systematic inference

151   of **TK→S** interactions. Specifically, to handle the highly discrete nature of the data, we replaced

152   the kernel-density and adaptive partitioning based mutual information estimators in the original

153   algorithm with a bin-count based method (**Fig. 1C4**), using gold standard data to select the most

154   effective number of bins [12] (see Methods). Furthermore, to deal with the skewed spectral count

155   distribution, we introduce an iterative quantile discretization method, where samples are binned

156   together, based on their spectral counts, to produce a distribution as close to uniform as possible

157   (**Fig. 1C3,** Methods).

158

pARACNe-inferred LUAD-specific TK-phosphorylation Network

We used pARACNe to reconstruct a LUAD-specific TK-signaling network, by analyzing phosphopeptide profiles obtained from 245 LUAD samples from Guo et al. (32). These data represent the abundance of peptides containing at least one phospho-tyrosine, as obtained by phosphoproteomic analysis of 46 LUAD cell lines, 151 LUAD tumors, and 48 adjacent normal samples. LC-MS/MS profiling produced spectral counts for 3,920 phospho-tyrosine containing peptides mapping to ~2,600 different proteins. Based on these data, pARACNE identified 2,611 candidate phospho-peptide/phospho-peptide dependencies, which could be further mapped to 2,064 unique **TK→S** interactions (**Table S1, S2**). These represent interactions between 46 unique TKs and their candidate substrates. These include 174 **TK$_1$→TK$_2$** interactions between two TKs (**Fig. 2A**), representing a statistically significant bias toward TK-TK interactions in the network ($p = 10^{-62}$). This suggests that, within the complete TK signaling network, TKs themselves may form a more densely inter-connected subnetwork than previously assessed, providing potentially valuable novel information about adaptive response, pathway cross-talk, and auto-regulatory loops.

Such highly interconnected structure provides potential functional advantage compared to less interconnected (i.e., "flat") architectures, including the ability to provide more fine-grain response to a highly heterogeneous variety of exogenous signals and conditions, the ability to provide rapid adaptive response to changing stimuli, and the ability to preserve cell state via autoregulatory feedback. Consistent with the underlying biology, and in contrast to transcriptional networks, the vast majority of pARACNe-inferred interactions have a positive Spearman correlation, with higher counts of TK-mapped phosphopeptides corresponding to higher counts of candidate substrate-mapped ones. This is consistent with the fact that TKs only phosphorylate their substrates, thus inducing positive phospho-state correlation. Only a negligible number of inferred interactions (0.5%) were associated with a negative Spearman correlation ($N = 11$, $p \leqslant 0.05$). These may

184 represent either indirect interactions where the TK activates a substrate-specific phosphatase or

185 direct interactions where phosphorylation of one phosphosite may prevent phosphorylation of

186 another site on the same protein.

187

188 ***LUAD Network Accuracy and Sensitivity Analysis***

189 To estimate the accuracy of the inferred TK-signaling network, we investigated the substrates of

190 two TK-proteins, EGFR and c-MET, representing high-affinity binding targets of existing FDA-

191 approved TK inhibitors for LUAD. Specifically, we compared their pARACNe-inferred substrates

192 to those reported in the phosphoDB database (33) and those supported by experimental

193 evidence, based on previously published SILAC assays, following cell line treatment with

194 associated, selective TK inhibitors. pARACNe inferred 123 EGFR substrates (**Fig. 2B**). Of these,

195 5 (blue and cyan) were included as high-confidence EGFR substrates in phosphoDB, out of 13 in

196 total (38%), including the established EGFR auto-phosphorylation site. Moreover, 50 additional

197 proteins (45%, green) showed significant decrease (at least 2 fold) in the abundance of their

198 phosphorylated isoforms in SILAC assays (32), following treatment of H3255 cells with the EGFR

199 inhibitor Gefitinib. Similarly, pARACNe predicted 179 c-MET substrates (**Fig. 2C**). Notably, both

200 of the established substrates reported in PhosphoDB were identified by pARACNE (100%, blue).

201 Moreover, 126 additional proteins (71.5%, blue) showed significant decrease in the abundance

202 of their phosphorylated isoforms in SILAC assays(32), following treatment of MKN45 cells with

203 the first-generation c-MET-specific inhibitor Su11274.

204 We used MKN45 to assess overall prediction accuracy, even though it represents a gastric cancer

205 cell line, because signaling networks should are much more conserved across tissue contexts

206 than transcriptional ones. Indeed, while lineage-specific chromatin state represents a major

207 determinant of transcriptional regulation, it only affects signal transduction in terms of overall

208   protein availability. As a result, it is reasonable to expect that an even greater overlap of inferred

209   vs. SILAC positive substrates may be achieved in native LUAD cells.

210

211   Taken together, these data suggest that pARACNe can identify a much larger subset of candidate

212   substrates, while both identifying a significant proportion of established substrates (46% on

213   average, based on phosphoDB) and maintaining high accuracy (~60% on average, by SILAC

214   assays). This also suggest that, similar to transcription factor targets reported in the literature, TK

215   substrates are still poorly characterized in existing repositories, even for highly relevant and

216   exceedingly well-studied kinases such as EGFR and c-MET. As a result, pARACNe could provide

217   significant novel hypotheses for TK→S interactions that can be validated as required. We should

218   also note that the reported accuracy for pARACNe is estimated using SILAC data on a single cell

219   line. SILAC assays have significant false negatives and it would be reasonable to expect that,

220   once tested in additional cell lines, the accuracy of pARACNe could further increase. As a further

221   performance benchmark, we used the same SILAC benchmark to test predictions by NetworkIN,

222   a reverse engineering method based on protein sequence motif analysis and protein association

223   networks (16). The analysis found almost no consensus with SILAC assays, with only one out of

224   33 NetworkIN-predicted EGFR substrate identified as significantly dephosphorylated following

225   treatment with TK-specific inhibitors.

226

227   ***Systematic, Network-based Inference of Pharmacological Dependencies***

228   Once an accurate model of signal transduction in LUAD cells was established by pARACNe

229   analysis, we interrogated the corresponding TK→S network using phosphoproteomic signatures

230   from 46 LUAD cell lines to identify key dependencies for experimental validation. For this purpose,

231   we extended the VIPER algorithm (Virtual Proteomics by Enriched Regulon analysis)(21), which

232   was originally developed to identify the MR proteins that mechanistically regulate the

233   transcriptional state of a tumor by assessing the enrichment of their transcriptional targets in

234   differentially expressed genes in the tumor signature. VIPER and its predecessor MARINa

235   (Master Regulator Inference algorithm) (34) have been instrumental in inferring MR proteins

236   representing key functional determinants of tumor-related phenotypes in many cancer types, from

237   glioblastoma (26, 27, 35), neuroblastoma (34), lymphoma (36, 37), and leukemia (38) to prostate

238   (39-41) and breast adenocarcinoma(42-44), among others. We thus reasoned that VIPER could

239   be modified to identify master regulator TKs, most likely to mechanistically regulate the differential

240   phosphorylation pattern observed in a specific tumor sample (**Fig. 3A**). A specific additional value

241   of the algorithm is that, as previously shown {Lefebvre, 2010 #1763}{Aytes, 2014 #2520}{Carro,

242   2010 #4021}, it could not only identify MR TK proteins, representing individual, pharmacologically

243   accessible dependencies of the tumor, but also TKs representing potential synergistic MR-pair as

244   candidate dependencies for combination therapy.

245   VIPER can be easily modified to analyze phosphoproteomic signatures (pVIPER). Specifically,

246   rather than assessing the enrichment of a protein's transcriptional targets (regulon) in differentially

247   expressed genes, pVIPER is designed to measure the enrichment of a TK's substrates (signalon)

248   in differentially phosphorylated proteins. Since inferred TK-substrate interactions are virtually all

249   positive, this further increases the accuracy of the algorithm by supporting use of a single-tail

250   enrichment analysis as opposed to the three-tail analysis of the original implementation. We first

251   performed pVIPER analysis at the individual phosphopeptide level, rather than by averaging

252   phosphopeptide state on a whole protein level. We then combined the result of the analysis across

253   all phosphopeptides mapping to the same protein. Consistent with VIPER's experimentally

254   validated ability to identify synergistic master regulators proteins by transcriptional network

255   analysis, pVIPER inferred several candidate synergistic TK interactions based on the statistical

256   significance of the signature-enrichment of substrates shared by both TKs compared to that of

11

257   substrates uniquely mapped to either one or the other TK (see Method section). Systematic

258   VIPER analysis of phosphoproteomic profiles from 46 LUAD cell lines generated between 2 and

259   13 master regulator TKs or synergistic TK-pairs, as candidate pharmacologically actionable

260   dependencies, for each cell line, thus generating a plausible number of hypothesis for each cell

261   line (**Fig. 3B** and **Fig. 3C**).

262

263   **pVIPER Identifies LUAD-specific Dependencies**

264   pVIPER analysis inferred several TK proteins as highly conserved individual dependencies across

265   multiple cell lines, including the Ephrin type-A receptor 2 (*EPHA2*), epidermal growth factor

266   receptor (*EGFR*), c-Met proto-oncogene (*MET*), and HER2 receptor tyrosine kinase 2 (*ERBB2*),

267   suggesting a critical role of these proteins in the maintenance of LUAD cell line state. This is also

268   in agreement with the functional role of these genes and the use of inhibitors of these kinases

269   across a large panel of patients in multiple cancer types (45-49).

270   In contrast to these established LUAD cell line dependencies, we also identified several TKs as

271   dependencies of specific cell lines. This can either be the result of associated genetic or

272   epigenetic alterations in these cell lines or the result of field effects, where multiple genetic

273   alterations or alterations in upstream pathway contribute to the cell line dependency on a specific

274   TK activity. For instance, we identified ALK (Anaplastic Lymphoma Receptor Tyrosine Kinase) to

275   be addiction point only in H2228 cell line. ALK is a conserved trans-membrane receptor tyrosine

276   kinase (RTK) protein in the insulin-receptor super family. Chromosomal alterations involving *ALK*

277   translocations and fusion events have been identified in several cancer types including LUAD (50,

278   51), diffuse large B-cell lymphomas (52), neuroblastoma (53), and inflammatory myofibroblastic

279   tumors (54), among others. Additionally, ALK fusion events with other genes, including EML4

280   (Echinoderm Microtubule-associated protein Like 4) in LUAD lead to aberrant protein activity

281   eliciting "oncogene addiction" (51). Presence of ALK-EML4 fusion transcripts, in ~3–7% of LUAD

282   patients (55-57), is a strong predictor of response to ALK inhibitors, such as Crizotinib, among

283   others (58, 59). Interestingly, among all available LUAD cell lines for which a phosphoproteomic

284   profile was available, H2228 was the only one with an established ALK-EML4 fusion event and

285   with established sensitivity to ALK inhibitor (60, 61). This further reflects the specificity of our

286   analysis as this was the only cell line predicted to depend on ALK activity. Interestingly, we

287   identified 4 additional H2228 dependencies, namely EGFR, Epha2, c-MET, and PTK2. H2228

288   sensitivity to EGFR inhibitors, in combination with ALK inhibitors, was already established (61).

289

290   ***EGFR and c-MET are Predicted Dependencies in Multiple LUAD Cell Lines***

291   As discussed, pVIPER analysis revealed several TK-pairs as candidate synergistic dependencies

292   across several cell lines, such as Epha2/c-MET, EGFR/PTK2, EGFR/Epha2, Epha2/c-MET, and

293   EGFR/c-MET. Among these the EGFR/c-MET pair emerged as the most conserved synergistic

294   TK-pair across the available cell lines. In addition, EGFR and c-MET were also identified as

295   candidate TK MRs in several of these cell lines, suggesting either a complementary or synergistic

296   role for these proteins and a potential therapeutic opportunity for combination therapy in LUAD

297   (62, 63).

298

299   To validate pVIPER-predicted, cell line specific EGFR/c-MET synthetic lethality, we selected a

300   panel of 14 cell lines, 11 of which were predicted to be synergistically dependent on EGFR/c-MET

301   (H226, H2122, H1666, H2172, Cal-12T, H2023, H1568, Calu-3, H1650, HCC78, and A549), as

302   well as 3 negative controls with no predicted synergistic or individual dependencies on the two

303   TKs (H2170, H460, and H520). To measure sensitivity to these agents, we used two different and

304   complementary assays, including: (a) colony formation assay to assess long term sensitivity (**Fig.**

305   **4A and Methods**) and (b) 3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide (MTS)

306   assay for short term sensitivity analysis (**Fig. 5A, Table S3 and Methods**). For colony formation

307 assays, cells were treated with either an EGFR inhibitor (Erlotinib, 1uM) or a c-MET/ALK inhibitor

308 (Crizotinib, 0.1uM), either individually or in combination (see Methods). To evaluate synergistic

309 dependency on EGFR/c-MET we used Excess Over Bliss (64), which measures the difference

310 between the observed effect on colony formation and the effect expected from a purely additive

311 model. For MTT assay, first, cells were treated with EGFR inhibitor (Erlotinib) or MET inhibitor

312 (Crizotinib) individually at various concentrations to identify IC50 (concentration resulting in 50%

313 cell death). Next, cells were treated with 1 uM of Erlotinib and varying concentrations of Crizotinib

314 to identify combinations resulting in IC50 and used the combination index (CI) statistic (65) to

315 measure interaction between the two drugs.

316

317 Across all 11 cell lines tested in colony formation assay, 8 showed significant sensitivity to either

318 individual inhibitors (H226$_{E,C}$, H2122$_E$, H1666$_{E,C}$, Cal-12T$_E$, Calu-3$_{E,C}$, H1650$_E$) or synergistic

319 sensitivity to the combination (HCC78$_{E+C}$ and H2023$_{E+C}$) (**Fig. 4B-C)**. Surprisingly, all of these cell

320 lines were EGFR^{WT}, ALK^{WT}, and c-MET^{WT}, except for H1650, which was EGFR^{Mut}. Thus, based

321 on standard of care criteria, 7 out of 8 cell lines would not have been considered as sensitive to

322 either EGFR or ALK/c-MET inhibitors. Several cell lines presented striking sensitivity to either one

323 (H2122$_E$, Cal-12T$_E$, H1650$_E$) or both inhibitors (H226$_{E,C}$, H1666$_{E,C}$, Calu-3$_{E,C}$) in isolation, thus

324 making the assessment of synergistic drug sensitivity difficult. In addition, three EGFR^{WT} cell lines

325 harboring BRAF (Cal-12T and H1666) or KRAS (H2122) mutations were also highly sensitive to

326 Erlotinib as a single agent, as predicted by pVIPER, despite the fact that KRAS pathway mutations

327 are mutually exclusive with EGFR mutations and predictive of Erlotinib resistance (**Fig. 4C**).

328 Finally, none of these cell lines was predicted to be sensitive to ALK inhibitors, suggesting that

329 Crizotinib sensitivity is mediated by c-MET specific dependencies. Of the negative controls, only

330 one (H2170) showed high sensitivity to Erlotinib. Taken together, 8/11 cell lines (73%) predicted

331 as sensitive to the inhibitors were validated long term colony formation assays, while only 1/3

332 negative controls showed sensitivity to these agents (33%).

333

334   To evaluate the short-term interaction between EGFR and c-MET, we performed MTT assay

335   across 11 cell lines (HCC78, H2023, H1650, Calu-3, H2172, H2122, H1568, A549, H1666, H520

336   and H2170) including 2 negative control cell lines (H520 and H2170). Similar to colony formation

337   assay, we found synergistic sensitivity to EGFR and c-MET inhibitors in 6/9 cell lines (67%), with

338   5 cell lines showing strong synergy (CI ≤ 0.8) (**Fig. 5B-C**) and 1 borderline synergy (CI = 0.82),

339   showing the consistency between two assays. However, for two cell lines, H1666 and H2170 (a

340   negative control), results were inconsistent between long term colony formation and MTT assays.

341   For both H1666 and H2170 cell lines, colony formation and MTT assay to showed sensitivity to

342   Erlotinib alone, where colony formation assay has complete abrogation of colonies at 1 µM of

343   Erlotinib, and later had IC_{50} =1.25 µM and 3.7 µM for H1666 and H2170 respectively. However,

344   in combination therapy, MTT assay showed antagonism (CI >1), despite the fact that colony

345   formation assay still showed complete abrogation which could be associated to the accumulation

346   of new mutations in these cell lines. However, this is just hypotheses and needs to be verified by

347   further experiments such as sequencing of these cell lines pre-and post-treatment.

348

349   ***Phosphosite-specific Phosphorylation Predicts EGFR/c-MET Inhibitor Synergy.***

350   In previous section, we assessed the pVIPER predictions after consolidating the result at protein

351   level. Following the results from MTT and colony formation assays, we reanalyzed the pVIPER

352   predictions at the phosphopeptide level. Interestingly, this revealed that whenever synergistic

353   EGFR/c-MET dependencies were predicted from phosphosite $EGFR_{1197}$ and phosphosites other

354   than c-MET_{1003} (H1666, Cal-12T, H1650), cell lines responded to Erlotinib in isolation, while when

355   predictions were based on phosphosites $EGFR_{1197}$ and c-MET_{1003} (HCC78, H2023, and Calu-3),

356   cells exhibited bona fide synergistic sensitivity to the two inhibitors, with the only possible

357   exception of Calu-3, which showed synergistic sensitivity in MTT assays and additive sensitivity

358   to both inhibitors in colony formation assays. Conversely, when predictions were not based on

359   either phosphotyrosine, cells exhibited no sensitivity to the individual inhibitors or the combination

360   (H2172, H226, A549, H460, H520, H1568). Thus, predictions based on these two phosphosites

361   produced no false positives (6 out of 6 predicted and validated as non-sensitive) and only 2 false

362   negatives (H2170 and H2122), resulting in an error rate of only 2 out of 14 cell lines (14%, p =

363   0.0093 using fisher exact test).

364

365   This finding is in agreement with the established role of $EGFR_{1197}$ as a predictor of EGFR inhibitor

366   sensitivity (66). Intriguingly, when sensitivity was predicted using phosphosites other than

367    $EGFR_{1197}$ and $c\text{-}MET_{1003}$, cell lines did not respond to the inhibitors, either individually or in

368   combination. For these two peptides, we found their common substrates to be hyper

369   phosphorylated in the sensitive cell lines **(Fig. 6A**) compared to the specific substrates of each of

370   them, whereas cell line responding only to EGFR inhibitors showed more hyper phosphorylation

371   of EGFR only substrates (**Fig. 6B**). Cell lines resistant to both EGFR and c-MET inhibitors either

372   showed no change in the phosphorylation status or hypo-phosphorylation compared to the normal

373   samples (**Fig. 6C**). Hence, either the common substrates of EGFR and c-MET, or the

374   phosphorylation status of $EGFR_{1197}$ and $c\text{-}MET_{1003}$ could potentially be used as biomarkers for

375   predicting therapy with the dual inhibitors. However, this conclusion is based on a very limited

376   number of observations and lacks the statistical power. This finding needs a re-

377   evaluation/validation using larger cohort of samples to establish an appropriate biomarker for

378   combination therapy.

379

380   ***Systematic Inference of Patient-specific Dependencies***

381   Similar to cell lines, when applied on patient data (32), pVIPER identified EGFR as one of the

382   most common addiction points (**Fig. 6**). We inferred EGFR dependency in 12 patients. Of these,

383   5 harbored EGFR mutations, while the remaining 7 patients had not been tested for this mutation,

384   showing a high consistency between our predictions and the genetic predisposition for sensitivity

385   to EGFR inhibitors. In the entire cohort, there were only 3 patients with EGFR mutations that were

386   not identified as EGFR dependent by pVIPER, resulting in an overall sensitivity of 62.5% (5/8).

387   However, it is well known that >50% of patients harboring EGFR mutations do not respond to

388   Erlotinib, suggesting that these may not be false negatives but rather patients with low activation

389   of downstream EGFR pathways, despite their EGFR^{Mut} state. Similarly, our analysis identified

390   candidate ALK dependencies in 4 patients. Of these one had an established TFG-ALK fusion,

391   whereas the others had not been tested for ALK fusion events.

392   Across all patient samples, we observed Discoidin Domain Receptor-1 (DDR1) to be the most

393   frequent addiction point, which was not predicted for any of the 46 cell lines. One reason for the

394   difference is that DDR1 is collagen dependent and there may be differences in the 3D structure

395   of the tumor and the cell lines growing on the plate. An independent study (67) in a cohort of 83

396   lung cancer specimens found that silencing of DDR1 in these samples leads to the hampering of

397   cell survival, reduced invasiveness in collagen matrices, increased apoptosis in basal condition

398   and decreased metastatic activity in model of tumor metastasis to bone, signifying it as a potential

399   novel therapeutic target.

400   ## Discussion

401   In this paper, we developed pARACNe to infer Tyrosine Kinase (TK) signaling network using

402   published genome-wide phosphoproteomic data from lung cancer. The network prediction was

403   validated using SILAC experiments, with high accuracy. Interrogation of the predicted TK-

404   substrate network generated biologically meaningful hypotheses, followed by experimental

405   validations illustrating the effectiveness of predicted kinase inhibitor combination, EGFR and c-

406   MET combination inhibitors, in treating lung cancer cell lines. Furthermore, Master Regulator

407   Analysis using patient proteomics data provides implications for using targeted agent

408   combinations to treat patients based on their proteomic profile data.

409

410   Notably, pARACNe is significant and powerful as of its genome-wide scale and context-specificity

411   in discovering global signaling cascading relationships, which were missing by previous methods.

412   For example, methods proposed by Linding et al. (16) combine motif-based phospho-site

413   predictions with information of physical association, co-occurrence, and co-expression to identify

414   substrates with high specificity and accuracy, but with low coverage and lack of contextual

415   specificity. Bender et al (17) used reverse phase protein assay data after various stimulations to

416   cells and inferred signaling network using hidden Markov models and genetic algorithms. Even

417   though the resulting networks are context specific, they lack genomic-scale coverage. There have

418   been methods which used existing large-scale protein networks and prune them using

419   transcriptomic information to identify signaling pathways (68-70). In addition, attempts have been

420   made to reconstruct signaling network using gene expression data (71, 72). However, as signaling

421   complexity lies mostly in upper level of cellular processes, inferring the cascades from

422   downstream gene expression fails to capture all the dynamics. Also, PrePPI proposed by Zhang

423   et al. (73) used protein structure-based methods to infer global protein-protein interaction, but this

424   approach fails to address phosphorylation context specificity. Innovative uses of multiplex and

425   microarray-based approaches, where multiple antibodies can be used to probe an ensemble of

426   phosphoproteins, are finally becoming sufficiently mature to allow characterization of small

427   pathways. Yet, these methods are still far from providing an unbiased, genome-wide view of

428   signal-transduction processes and continue to be completely dependent on antibody specificity

429   and availability. Similarly, assays developed specifically to monitor phosphorylation pathways,

430   such as Stable Isotope Labeling with Amino acids in Cell culture (SILAC), provides a simple and

431   straightforward approach to detect differential protein abundance. Coupled with phosphorylation

432   enriched assays, it can provide high quality quantification for post-translation phosphorylation

433   changes in cell lines. However, these methods are 1) laborious and costly; 2) can only be

434 performed to dissect the substrates of a single enzyme at a time and 3) do not differentiate

435 between direct and indirect targets.

436

437 To be noted, as the LC-MS/MS experiments used here was generated based on Tyrosine-kinase

438 enrichment, which is only about ~2% of whole phosphoproteome. pARACNe is shown only on

439 TK-substrates network. The current methodology could be extended to whole phosphoproteomic

440 data based signaling network reconstruction where the data is available. In addition to label free

441 based LC-MS/MS proteomics data used in this work, label based approaches, such as ITRAQ or

442 TMT, could generate higher throughput whole proteomic profiles which might require future

443 redesign of ARACNe to incorporate both kinases and phosphatases in regulating their

444 downstream substrates. It is reasonable to expect that a version of ARACNe developed

445 specifically to dissect signaling networks should work at least as well as its transcriptional

446 counterpart. Since the relationship between the mRNA abundance of a gene encoding a

447 transcription factor (TF) and the activity of the corresponding protein is much looser than that

448 between the abundance of a phospho-isoform of a kinase and its enzymatic activity.

449

450

451 Even though research has attempted to identify addiction points based on gene expression data

452 (74), predictions based on phosphoproteomic data appear superior in a way that they can directly

453 reflect contextual specific signaling activity and are able to be directly targeted by kinase

454 inhibitors. It is important to note that clinically, only patients with base-pair deletion at exon 19

455 (del746_A750) or a point mutation at exon 21 mutation (L858R) in EGFR shows sensitivity to

456 EGFR inhibitor such as Cedirinib or Erlotinib (75).

457

Acknowledgments

Author Contributions

M.B. and M.K. developed the algorithm. M.B. and J.H. performed the analysis. M.P. performed the experimental validation. A.I., M.B., M. W., J.M and A.C. guided the experiments. M.B. J.H., A.I. and A.C wrote the manuscript. M.B. and A.C. provided supervision and guidance. A.C. conceived, funded and administrated the project.

Declaration of Interests

A.C.is a founder and shareholder of DarwinHealth Inc. and a member of the Tempus Inc. SAB and shareholder. Columbia University is a shareholder of DarwinHealth Inc.

References

1.   Ettinger DS, Akerley W, Borghaei H, Chang AC, Cheney RT, Chirieac LR, et al. Non-small cell lung cancer. Journal of the National Comprehensive Cancer Network : JNCCN. 2012;10(10):1236-71.

2.   Karasaki T, Nagayama K, Kawashima M, Hiyama N, Murayama T, Kuwano H, et al. Identification of Individual Cancer-Specific Somatic Mutations for Neoantigen-Based Immunotherapy of Lung Cancer. J Thorac Oncol. 2016;11(3):324-33.

3.   Pasqualucci L, Kitaura Y, Gu H, Dalla-Favera R. PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. Proc Natl Acad Sci U S A. 2006;103(2):395-400.

4.   Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, Mukherjee P, et al. Functional organization of the S. cerevisiae phosphorylation network. Cell. 2009;136(5):952-63.

5.   Gainor JF, Varghese AM, Ou SH, Kabraji S, Awad MM, Katayama R, et al. ALK rearrangements are mutually exclusive with mutations in EGFR or KRAS: an analysis of 1,683 patients with non-small cell lung cancer. Clin Cancer Res. 2013;19(15):4273-81.

6.   Janne PA, Shaw AT, Pereira JR, Jeannin G, Vansteenkiste J, Barrios C, et al. Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study. Lancet Oncol. 2013;14(1):38-47.

7.   Mak KS, Gainor JF, Niemierko A, Oh KS, Willers H, Choi NC, et al. Significance of targeted therapy and genetic alterations in EGFR, ALK, or KRAS on survival in patients with non-small cell lung cancer treated with radiotherapy for brain metastases. Neuro Oncol. 2014.

8.   Vendrell JA, Taviaux S, Beganton B, Godreuil S, Audran P, Grand D, et al. Detection of known and novel ALK fusion transcripts in lung cancer patients using next-generation sequencing approaches. Sci Rep. 2017;7(1):12510.

9.   Peters S, Camidge DR, Shaw AT, Gadgeel S, Ahn JS, Kim DW, et al. Alectinib versus Crizotinib in Untreated ALK-Positive Non-Small-Cell Lung Cancer. N Engl J Med. 2017;377(9):829-38.

10.   Tsvetkova E, Goss GD. Drug resistance and its significance for treatment decisions in non-small-cell lung cancer. Curr Oncol. 2012;19(Suppl 1):S45-51.

11.   Wilson FH, Johannessen CM, Piccioni F, Tamayo P, Kim JW, Van Allen EM, et al. A functional landscape of resistance to ALK inhibition in lung cancer. Cancer Cell. 2015;27(3):397-408.

12.   O'Shaughnessy J, Miles D, Vukelja S, Moiseyenko V, Ayoub JP, Cervantes G, et al. Superior survival with capecitabine plus docetaxel combination therapy in anthracycline-pretreated patients with advanced breast cancer: phase III trial results. J Clin Oncol. 2002;20(12):2812-23.

13.   Pegram MD, Slamon DJ. Combination therapy with trastuzumab (Herceptin) and cisplatin for chemoresistant metastatic breast cancer: evidence for receptor-enhanced chemosensitivity. Semin Oncol. 1999;26(4 Suppl 12):89-95.

14.   Ravandi F, Cortes JE, Jones D, Faderl S, Garcia-Manero G, Konopleva MY, et al. Phase I/II study of combination therapy with sorafenib, idarubicin, and cytarabine in younger patients with acute myeloid leukemia. J Clin Oncol. 2010;28(11):1856-62.

15.   Wu P, Nielsen TE, Clausen MH. Small-molecule kinase inhibitors: an analysis of FDA-approved drugs. Drug Discov Today. 2016;21(1):5-10.

16.   Linding R, Jensen LJ, Ostheimer GJ, van Vugt MA, Jorgensen C, Miron IM, et al. Systematic discovery of in vivo phosphorylation networks. Cell. 2007;129(7):1415-26.

17.   Bender C, Henjes F, Frohlich H, Wiemann S, Korf U, Beissbarth T. Dynamic deterministic effects propagation networks: learning signalling pathways from longitudinal protein array data. Bioinformatics. 2010;26(18):i596-602.

517   18.   Papin JA, Hunter T, Palsson BO, Subramaniam S. Reconstruction of cellular signalling networks and
518   analysis of their properties. Nature reviews. 2005;6(2):99-111.
519   19.   Hashemikhabir S, Ayaz ES, Kavurucu Y, Can T, Kahveci T. Large-scale signaling network
520   reconstruction. IEEE/ACM Trans Comput Biol Bioinform. 2012;9(6):1696-708.
521   20.   Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of
522   regulatory networks in human B cells. Nat Genet. 2005;37(4):382-90.
523   21.   Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Functional characterization of
524   somatic mutations in cancer using network-based inference of protein activity. Nat Genet.
525   2016;48(8):838-47.
526   22.   MacKay DJC. Information theory, inference, and learning algorithms. Cambridge, UK ; New York:
527   Cambridge University Press; 2003. xii, 628 p. p.
528   23.   Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, et al. ARACNE: an
529   algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC
530   bioinformatics. 2006;7 Suppl 1:S7.
531   24.   Basso K, Saito M, Sumazin P, Margolin AA, Wang K, Lim WK, et al. Integrated biochemical and
532   computational approach identifies BCL6 direct target genes controlling multiple pathways in normal
533   germinal center B cells. Blood. 2010;115(5):975-84.
534   25.   Della Gatta G, Palomero T, Perez-Garcia A, Ambesi-Impiombato A, Bansal M, Carpenter ZW, et al.
535   Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in
536   T-ALL. Nat Med. 2012;18(3):436-40.
537   26.   Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, et al. The transcriptional network for
538   mesenchymal transformation of brain tumours. Nature. 2010;463(7279):318-25.
539   27.   Sonabend AM, Bansal M, Guarnieri P, Lei L, Amendolara B, Soderquist C, et al. The transcriptional
540   regulatory network of proneural glioma determines the genetic alterations selected during tumor
541   progression. Cancer Res. 2014;74(5):1440-51.
542   28.   Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function.
543   Nature. 2016;537(7620):347-55.
544   29.   Gillet LC, Leitner A, Aebersold R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering
545   the High-Throughput Era for Hypothesis Testing. Annu Rev Anal Chem (Palo Alto Calif). 2016;9(1):449-72.
546   30.   Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, et al. The BioPlex Network: A
547   Systematic Exploration of the Human Interactome. Cell. 2015;162(2):425-40.
548   31.   Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, et al. Architecture of the human
549   interactome defines protein communities and disease networks. Nature. 2017;545(7655):505-9.
550   32.   Guo A, Villen J, Kornhauser J, Lee KA, Stokes MP, Rikova K, et al. Signaling networks assembled by
551   oncogenic EGFR and c-Met. Proc Natl Acad Sci U S A. 2008;105(2):692-7.
552   33.   Giansanti P, Aye TT, van den Toorn H, Peng M, van Breukelen B, Heck AJ. An Augmented Multiple-
553   Protease-Based Human Phosphopeptide Atlas. Cell Rep. 2015;11(11):1834-43.
554   34.   Lefebvre C, Rajbhandari P, Alvarez MJ, Bandaru P, Lim WK, Sato M, et al. A human B-cell
555   interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol Syst
556   Biol. 2010;6:377.
557   35.   Chudnovsky Y, Kim D, Zheng SY, Whyte WA, Bansal M, Bray MA, et al. ZFHX4 Interacts with the
558   NuRD Core Member CHD4 and Regulates the Glioblastoma Tumor-Initiating Cell State. Cell Rep.
559   2014;6(2):313-24.
560   36.   Ying CY, Dominguez-Sola D, Fabi M, Lorenz IC, Hussein S, Bansal M, et al. MEF2B mutations lead
561   to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. Nat Immunol.
562   2013;14(10):1084-+.

563 37.   Bisikirska B, Bansal M, Shen Y, Teruya-Feldstein J, Chaganti R, Califano A. Elucidation and
564 Pharmacological Targeting of Novel Molecular Drivers of Follicular Lymphoma Progression. Cancer Res.
565 2016;76(3):664-74.
566 38.   Piovan E, Yu J, Tosello V, Herranz D, Ambesi-Impiombato A, Da Silva AC, et al. Direct reversal of
567 glucocorticoid resistance by AKT inhibition in acute lymphoblastic leukemia. Cancer Cell. 2013;24(6):766-
568 76.
569 39.   Aytes A, Mitrofanova A, Lefebvre C, Alvarez MJ, Castillo-Martin M, Zheng T, et al. Cross-species
570 regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives
571 prostate cancer malignancy. Cancer Cell. 2014;25(5):638-51.
572 40.   Mitrofanova A, Aytes A, Shen C, Abate-Shen C, Califano A. A systems biology approach to predict
573 drug response for human prostate cancer based on in vivo preclinical analyses of mouse models. Cell
574 Reports. 2015;12:1–12.
575 41.   Talos F, Mitrofanova A, Bergren SK, Califano A, Shen MM. A computational systems approach
576 identifies synergistic specification genes that facilitate lineage conversion to prostate tissue. Nature
577 communications. 2017;8:14662.
578 42.   Walsh LA, Alvarez MJ, Sabio EY, Reyngold M, Makarov V, Mukherjee S, et al. An Integrated Systems
579 Biology Approach Identifies TRIM25 as a Key Determinant of Breast Cancer Metastasis. Cell Rep.
580 2017;20(7):1623-40.
581 43.   Rodriguez-Barrueco R, Yu J, Saucedo-Cuevas LP, Olivan M, Llobet-Navas D, Putcha P, et al.
582 Inhibition of the autocrine IL-6-JAK2-STAT3-calprotectin axis as targeted therapy for HR-/HER2+ breast
583 cancers. Genes Dev. 2015;29(15):1631-48.
584 44.   Putcha P, Yu J, Rodriguez-Barrueco R, Saucedo-Cuevas L, Villagrasa P, Murga-Penas E, et al. HDAC6
585 activity is a non-oncogene addiction hub for inflammatory breast cancers. Breast Cancer Res.
586 2015;17(1):149.
587 45.   Davoli A, Hocevar BA, Brown TL. Progression and treatment of HER2-positive breast cancer.
588 Cancer Chemother Pharmacol.65(4):611-23.
589 46.   Guo L, Kozlosky CJ, Ericsson LH, Daniel TO, Cerretti DP, Johnson RS. Studies of ligand-induced site-
590 specific phosphorylation of epidermal growth factor receptor. Journal of the American Society for Mass
591 Spectrometry. 2003;14(9):1022-31.
592 47.   Okamoto I. Epidermal growth factor receptor in relation to tumor development: EGFR-targeted
593 anticancer therapy. FEBS J.277(2):309-15.
594 48.   Wang S, Placzek WJ, Stebbins JL, Mitra S, Noberini R, Koolpe M, et al. Novel targeted system to
595 deliver chemotherapeutic drugs to EphA2-expressing cancer cells. J Med Chem. 2012;55(5):2427-36.
596 49.   Birtwistle MR, Hatakeyama M, Yumoto N, Ogunnaike BA, Hoek JB, Kholodenko BN. Ligand-
597 dependent responses of the ErbB signaling network: experimental and modeling analyses. Mol Syst Biol.
598 2007;3:144.
599 50.   Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, et al. Global survey of phosphotyrosine
600 signaling identifies oncogenic kinases in lung cancer. Cell. 2007;131(6):1190-203.
601 51.   Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the
602 transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448(7153):561-6.
603 52.   Laurent C, Do C, Gascoyne RD, Lamant L, Ysebaert L, Laurent G, et al. Anaplastic lymphoma kinase-
604 positive diffuse large B-cell lymphoma: a rare clinicopathologic entity with poor prognosis. J Clin Oncol.
605 2009;27(25):4211-6.
606 53.   Janoueix-Lerosey I, Lequin D, Brugieres L, Ribeiro A, de Pontual L, Combaret V, et al. Somatic and
607 germline activating mutations of the ALK kinase receptor in neuroblastoma. Nature. 2008;455(7215):967-
608 70.
609 54.   Sokai A, Enaka M, Sokai R, Mori S, Mori S, Gunji M, et al. Pulmonary Inflammatory Myofibroblastic
610 Tumor Harboring EML4-ALK Fusion Gene. Jpn J Clin Oncol. 2014;44(1):93-6.

611 55.   Bender C, Henjes F, Frohlich H, Wiemann S, Korf U, Beissbarth T. Dynamic deterministic effects
612 propagation networks: learning signalling pathways from longitudinal protein array data. Bioinformatics.
613 2010;26(18):i596-i602.
614 56.   Koivunen JP, Mermel C, Zejnullahu K, Murphy C, Lifshits E, Holmes AJ, et al. EML4-ALK fusion gene
615 and efficacy of an ALK kinase inhibitor in lung cancer. Clin Cancer Res. 2008;14(13):4275-83.
616 57.   TCGA-Consortium. Comprehensive genomic characterization defines human glioblastoma genes
617 and core pathways. Nature. 2008;455(7216):1061-8.
618 58.   Crystal AS, Shaw AT. Variants on a theme: a biomarker of crizotinib response in ALK-positive non-
619 small cell lung cancer? Clin Cancer Res. 2012;18(17):4479-81.
620 59.   Shaw AT, Solomon B. Targeting anaplastic lymphoma kinase in lung cancer. Clin Cancer Res.
621 2011;17(8):2081-6.
622 60.   . !!! INVALID CITATION !!! [24].
623 61.   Voena C, Di Giacomo F, Panizza E, D'Amico L, Boccalatte FE, Pellegrino E, et al. The EGFR family
624 members sustain the neoplastic phenotype of ALK+ lung adenocarcinoma via EGR1. Oncogenesis.
625 2013;2:e43.
626 62.   Chan BA, Hughes BG. Targeted therapy for non-small cell lung cancer: current standards and the
627 promise of the future. Transl Lung Cancer Res. 2015;4(1):36-54.
628 63.   Landi L, Minuti G, D'Incecco A, Cappuzzo F. Targeting c-MET in the battle against advanced
629 nonsmall-cell lung cancer. Curr Opin Oncol. 2013;25(2):130-6.
630 64.   Borisy AA, Elliott PJ, Hurst NW, Lee MS, Lehar J, Price ER, et al. Systematic discovery of
631 multicomponent therapeutics. Proc Natl Acad Sci U S A. 2003;100(13):7977-82.
632 65.   Chou TC. Drug Combination Studies and Their Synergy Quantification Using the Chou-Talalay
633 Method. Cancer research. 2010;70(2):440-6.
634 66.   Zhang G, Fang B, Liu RZ, Lin H, Kinose F, Bai Y, et al. Mass spectrometry mapping of epidermal
635 growth factor receptor phosphorylation related to oncogenic mutations and tyrosine kinase inhibitor
636 sensitivity. J Proteome Res. 2011;10(1):305-19.
637 67.   Valencia K, Ormazabal C, Zandueta C, Luis-Ravelo D, Anton I, Pajares MJ, et al. Inhibition of
638 collagen receptor discoidin domain receptor-1 (DDR1) reduces cell survival, homing, and colonization in
639 lung cancer bone metastasis. Clin Cancer Res. 2012;18(4):969-80.
640 68.   Scott J, Ideker T, Karp RM, Sharan R. Efficient algorithms for detecting signaling pathways in
641 protein interaction networks. Journal of computational biology : a journal of computational molecular cell
642 biology. 2006;13(2):133-44.
643 69.   Steffen M, Petti A, Aach J, D'Haeseleer P, Church G. Automated modelling of signal transduction
644 networks. BMC bioinformatics. 2002;3:34.
645 70.   Kelley BP, Sharan R, Karp RM, Sittler T, Root DE, Stockwell BR, et al. Conserved pathways within
646 bacteria and yeast as revealed by global protein network alignment. Proceedings of the National Academy
647 of Sciences of the United States of America. 2003;100(20):11394-9.
648 71.   Shimoni Y, Fink MY, Choi SG, Sealfon SC. Plato's cave algorithm: inferring functional signaling
649 networks from early gene expression shadows. PLoS Comput Biol. 2010;6(6):e1000828.
650 72.   Budak G, Eren Ozsoy O, Aydin Son Y, Can T, Tuncbag N. Reconstruction of the temporal signaling
651 network in Salmonella-infected human cells. Front Microbiol. 2015;6:730.
652 73.   Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, Thu CA, et al. Structure-based prediction of protein-
653 protein interactions on a genome-wide scale. Nature. 2012;490(7421):556-60.
654 74.   Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using
655 gene-expression signatures to connect small molecules, genes, and disease. Science.
656 2006;313(5795):1929-35.
657 75.   Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel
658 in pulmonary adenocarcinoma. The New England journal of medicine. 2009;361(10):947-57.

659   76.   Steuer R, Kurths J, Daub CO, Weise J, Selbig J. The mutual information: detecting and evaluating
660   dependencies between variables. Bioinformatics. 2002;18 Suppl 2:S231-40.
661   77.   Gandhi J, Zhang JL, Xie Y, Soh J, Shigematsu H, Zhang W, et al. Alterations in Genes of the EGFR
662   Signaling Pathway and Their Relationship to EGFR Tyrosine Kinase Inhibitor Sensitivity in Lung Cancer Cell
663   Lines. PloS one. 2009;4(2).

664

665 # Figure Legends

666 **Figure 1. Framework for the reverse engineering of TK signaling networks from**
667 **phosphoproteomic profiles.**

668 **(A)** Schematic diagram of a **TK→S** interaction. The non-phosphorylated kinase is inactive in terms

669 of phosphorylating a substrate, while the active isoform successfully phosphorylates the

670 substrate. **(B)** Schematic diagram showing the correlation between TK phosphorylation and that

671 of its potential substrates. The first two rows in the heatmap show proteins representing candidate

672 TK substrates **(C)** Illustration of the pARACNe framework including 6 steps. Step-1 depicts

673 peptides collection from primary lung cancer tissue and cell lines for whole phosphrtyrosine

674 proteomics quantification. Step-2 depicts inferences of TK→S interactions using Mutual

675 Information by Step-3 Naïve-Bayes estimator and Step-4 of the iterative quantile discretization

676 methods. Step-5 and 6 depict network pruning and bootstrapping to construct final network. **(C)**.

677 Workflow of pARACNe from LC-MS/MS data normalization, IQD process, MI calculation, DPI

678 process, bootstrapping to network consolidation.

679 **Figure 2. Predicted TK-TK network and validation of EGFR and c-MET prediction**

680 **(A)** pARACNe-inferred densely inter-connected TK-TK network, with red nodes representing

681 candidate TKs involved in auto-phosphorylation, where the phospho-state of a tyrosine is

682 correlated with the phospho-state of a different tyrosine on the same TK protein. **(B)** pARACNe-

683 inferred EGFR and **(C)** c-MET substrate overlap with SILAC-based and Database reported

684 substrates, respectively.

685

686 **Figure 3. Inference of Master Regulator and combination**

687 **(A)**. Schema of Master Regulator analysis in lung cancer using pVIPER. Prioritized Master

688 Regulators **(B)** and Prioritized Master Regulator Pairs **(C)** as significantly activated (red circle) or

689   de-activated (blue) in different lung cancer cell lines (column). Red color represents an enrichment

690   of substrates hyper-phosphorylation by a Master Regulator or Master Regulator Pairs. Blue color

691   represents that of hypo-phosphorylation.

692   **Figure 4. Experimental validation of EGFR and c-MET combination by colony**
693   **formation assay**

694   **(A)** Colony formation assay schema. **(B)** shows the image of long-term EGFR and c-MET double

695   inhibition effects in HCC78 cell line with different treatments. **(C)** shows long-term clony formation

696   data for 14 cell lines with different EGFR, BRAF and KRAS genomic mutation status.

697   **Figure 5. MTT Assay validation of EGFR and c-MET combination**

698   **(A)**. MTT assay experimental schema. **(B)** MTT assay of HCC78 cell line shows synergistic effects

699   of Crizotinib and Erlotinib treatment. **(C)** shows short-term effects of EGFR and c-MET inhibitors'

700   combination index in 11 cell lines include 2 control cell lines (red).

701   **Figure 6. Master Regulating peptides in primary lung cancer samples**

702   EGFR and c-MET co-regulate in three scenarios **(A)** when their common substrates are

703   hyperphosphrylated, the patient responds to combination treatment well; **(B)** when most EGFR

704   substrates are hyper-phosphorylated, the patient responds to EGFR inhibitor; **(C)** when

705   substrates of both EGFR and c-MET are mostly hypophosphorylated, the patient does not

706   respond. **(D)** show the Master Regulator and Master Regulator Pairs regulating hyper/hypo-

707   phosphorylation of their network substrates in each primary samples.

708

Methods and Data

Phosphoproteomic Data

The previously published phosphoproteomic data used to reconstruct signaling network was download from (50). This dataset, representing the abundance of phospho-tyrosine containing peptides, was obtained by tandem mass spectrometry analysis of 46 non-small cell lung cancer (NSCLC) cell lines, 151 NSCLC tumors, and 48 normal lung tissue samples. Immunohistochemistry and a phospho-tyrosine specific antibody were used to screen 96 paraffin-embedded, formalin fixed tissue samples from NSCLC patients as described by Rikova et al.(50). About 30% of tumors showed high-levels of phospho-tyrosine expression. Immunoblotting of 46 NSCLC cell lines with a phospho-tyrosine specific antibody also showed heterogeneous reactivity especially in the molecular weight range characteristic of receptor tyrosine kinases.

Since phospho-tyrosine represents less than 1% of the cellular phosphoproteome, as determined by tandem mass spectrometry (MS/MS), and is difficult to analyze by conventional methods, immuno-affinity purification was performed with a phospho-tyrosine antibody to enrich for phospho-tyrosine containing peptides prior to tandem mass spectrometry. All tumors were identified as NSCLC based on standard pathology. Only those tumors with greater than 50% of cancer cells were considered for further analysis. NSCLC cell lines were grown overnight in low serum to reduce background phosphorylation from culture conditions.

Tandem MS profiling identified 3920 tyrosine phosphorylation sites on approximately 2600 different proteins. 85% of these sites appeared to be novel when compared against PhosphoSite (http://www.phosphosite.org), a comprehensive resource of known phosphorylation sites.

730   *pARACNe Algorithm*

731   ARACNe is originally designed for gene expression data, where expression of genes is usually

732   continuous and non-sparse. Quantitative data obtained from label-free LC-MS/MS by data-

733   dependent acquisition via spectral counting is discrete and very sparse, with many

734   phosphopeptides counts not observed for multiple peptides in each sample causing the current

735   version of ARACNe to be not suitable for this data, which thus required major modifications to

736   handle discrete data. To handle discrete abundances, we modified the mutual information

737   computation approach from a kernel density estimation based method to a Naïve based

738   estimation of mutual information, which is a histogram based technique(76). Briefly, consider a

739   collection of N simultaneous measurements of two genes X and Y. Data is partitioned into M

740   discrete bins a_i, and k_i denotes the number of measurements that lie within the bin a_i. The

741   probabilities $p(a_i)$ are then approximated by the corresponding relative frequencies of

742   occurrence $p(a_i) \rightarrow \dfrac{k_i}{N}$ and the mutual information $I(X,Y)$ between datasets X and Y is

743   expressed as

744
$$I(X,Y) = \log N + \frac{1}{N} \sum_{ij} k_{ij} \log \frac{k_{ij}}{k_i k_j}$$

745   Here k_{ij} denotes the number of measurements where X lies in a_i and Y in a_j and N total number

746   of samples.

747   Accuracy of mutual information is dependent on correct numbers of bins, M. To find the optimal

748   number of bins we applied ARACNe on the whole dataset by varying M from 1 to 20 and testing

749   the connections in predicted sub-network against the set of known connections (gold standard)

750   from databases (phosphoDB) (33).

751   In case of continuous data, partitioning can be achieved by dividing the range of data into M

752   equally spaced distance bins. Our data being discrete, equally spaced partitioning was not

753   possible. So, to overcome this problem, we used an iterative approach of partitioning **(Fig. 1C,**

754   **Fig. S1A)**. The basic idea is to divide the number of N data points into M with each bin containing

755   equal number of data point. If the data point(s) with the same value falls into consecutive bin(s),

756   we put those data point(s) into current bin and repartition the remaining points into remaining

757   number of bins. We keep iterating this till we finish either the bins or there are no more data points

758   to bin. For example, in **Fig. S1A**, we initially partition N points into 4 bins. The data points with 0

759   value does not fit into first bin and falls into subsequent bins, so we assign all data points with 0

760   value into first bin and repartition the remaining points into 3 bins. We keep on doing this process

761   till we achieve 4 bins.

762   To evaluate initial performance and decide number of bins, we computed the network among all

763   tyrosine kinases and substrates, parsed the sub-network between 49 tyrosine kinases and 114

764   substrates which were present in PhosphoSite database and compared the results with the

765   connections present in database. From our analysis, we found that M=10 to be an optimal number

766   **(Fig. S1B)** which gave us precision of 14% and sensitivity of 24%. This precision is an

767   underestimate of real precision as in the gold standard many interactions are not present.

768   ***Master Regulator Analysis***

769   To discover the master regulator in various cell lines, we interrogated the network obtained from

770   pARACNe using a novel algorithm, VIPER (Virtual Proteomics by Enriched Regulon analysis)

771   (21), designed to infer kinases that are key players in a particular cell line. Protein activity is a

772   good indicator of key kinases in a particular phenotype but often phosphorylated state of a protein

773   is not sufficient to determine its activity both due to measurement noise in phosphorylated state

774   as well as LC-MS/MS technique noise. To overcome this problem, VIPER infers kinase activity

775   from the global kinase substrate relationship and its biological relevance by overlapping this

776   information in a particular phenotype-specific program.

777   VIPER requires a network model and signature of the phenotype transition (i.e., all genes ranked

778   by their differential phosphorylation in two phenotypes). Here, the signature, S_{kin}, was obtained

779   by t-test analysis by comparing each cell line against all normal samples. First, we associate each

780   kinase with positive and negative activity targets, by computing the correlation between each

781   kinase and its predicted substrates and selecting only those substrates which had a significant

782   correlation (p-value ≤ 0.05, Bonferroni corrected). Second, for each kinase we computed an

783   *activity* by measuring the enrichment of the S_{kin} signature in predicted substrates list, separately

784   for both positive and negative correlated, (S_{kin}-*enrichment*). Enrichment was computed by Gene

785   Set Enrichment Analysis (GSEA). Since very small percentage of kinases are found to have

786   negative correlation, we did not use those interactions to evaluate enrichment.

787   ***Cell Culture***

788   All cell lines were grown in RPMI-1640 with 5% fetal bovine serum and incubated at 37°C in a

789   humidified atmosphere containing 5% CO_2. Cell lines were fingerprinted using the Perplex 1.2

790   system (Promega, Madison, WI). Fingerprints were compared to those generated at ATCC and/or

791   our internal database.

792   ***MTS Assays***

793   Short term MTS assays (CellTiter 96® AQ$_{ueous}$ One Solution Cell Proliferation Assay, Promega,

794   Madison WI) were performed as previously described in (77). Specifically, each drug

795   concentration is octuplicated and the mean with standard deviation of all replicates were used to

796   generate a curve to allow calculation of the drug IC50 (Inhibitory Concentration of 50%) value.

797   The assays were repeated at least 3 times and the IC50s are the average of all replicates.

798

799 ***Colony-formation Assays***

800 Long term colony formation assays were performed in triplicate in 6-well plates. Cells were added

801 to media containing drug and incubated for 1-2 weeks such that control wells (no drug) contained

802 colonies of 50-70 cells each. At such time media was removed and all wells stained with a solution

803 containing 0.5% crystal violet and 6% glutaraldehyde for 1 hour. The plates were then rinsed,

804 dried, and colonies were manually counted.

805 ***SILAC Experiments***

806 EGFR SILAC experiment was performed in H3255 cell line by treating samples with Gefitinib. c-

807 MET SILAC experiment was performed in c-MET-driven gastric cancer cell line, MKN45, by using

808 c-MET inhibitor Su11274. For both genes, cells were treated with inhibitors for 3 and 24hr. For

809 control, cells were grown in same conditions but were not treated with the drug. For our

810 comparison we combined the peptides, which were differentially obtained between treated and

811 untreated samples, for 3 and 24 hr. More details about the experiment can be obtained from Guo

812 et al (32).

813 **Gold standard**: In PhosphoSite database, there were 282 connections between 49 tyrosine

814 kinases and 114 substrates.

815

Supplemental Information

817 ***Fig. S1. Performance of the pARACNe Algorithm.***

818 **(A).** To select optimal bin number in pARACNe algorithm, precision and recall curves for various

819 number of bins were computed. Black curve is when no binning of data is done. When using 10

820 bins, the algorithm achieved the best performance.

821   ***Table S1. pARACNe-inferred TK-peptides/substrate-peptides Interaction Network.***

822   ***Table S2. pARACNe-inferred TK-Protein/Substrate Interaction Network.***

823   ***Table S3. Colony Formation Assay and MTS Assay Results.***
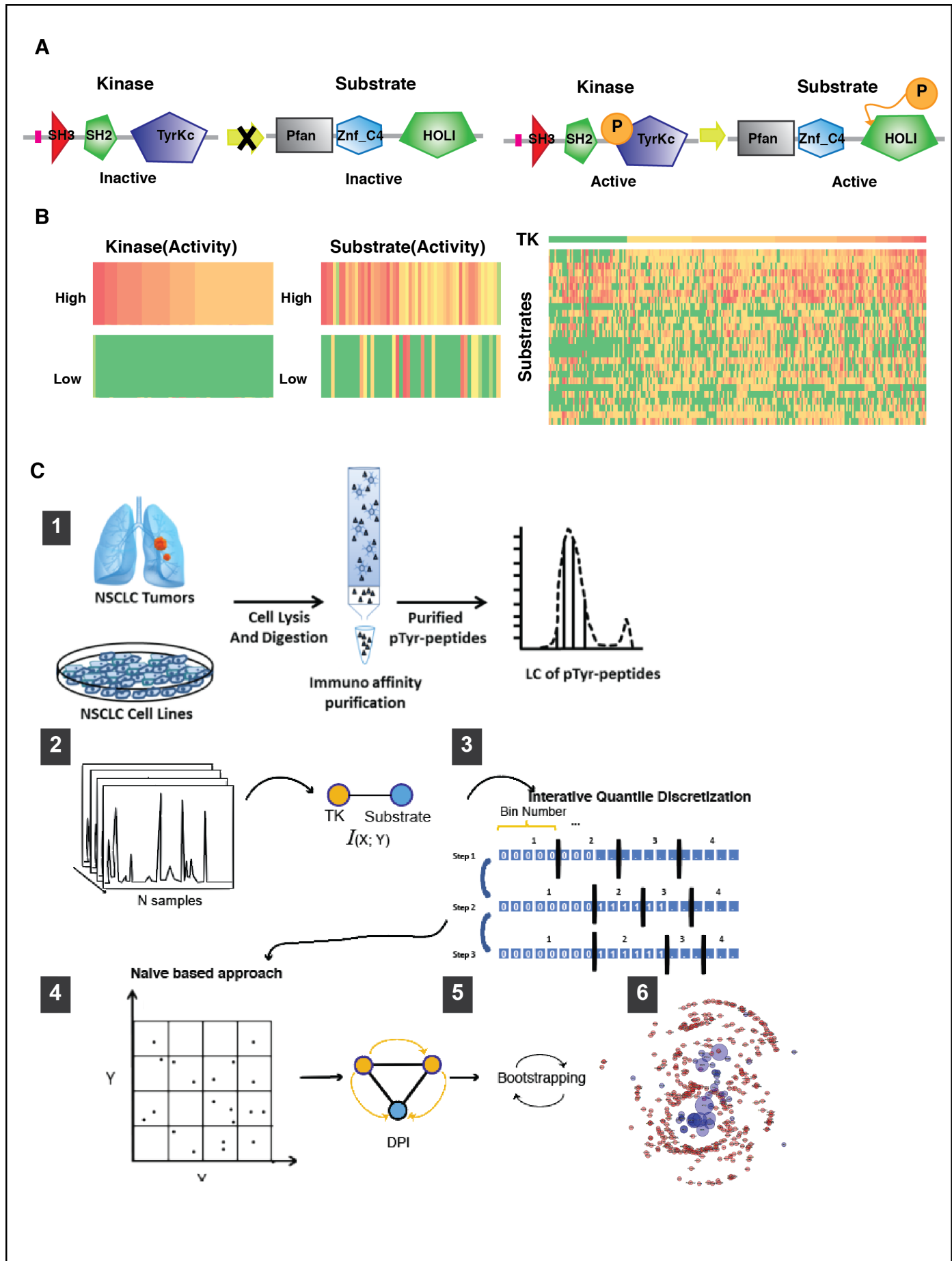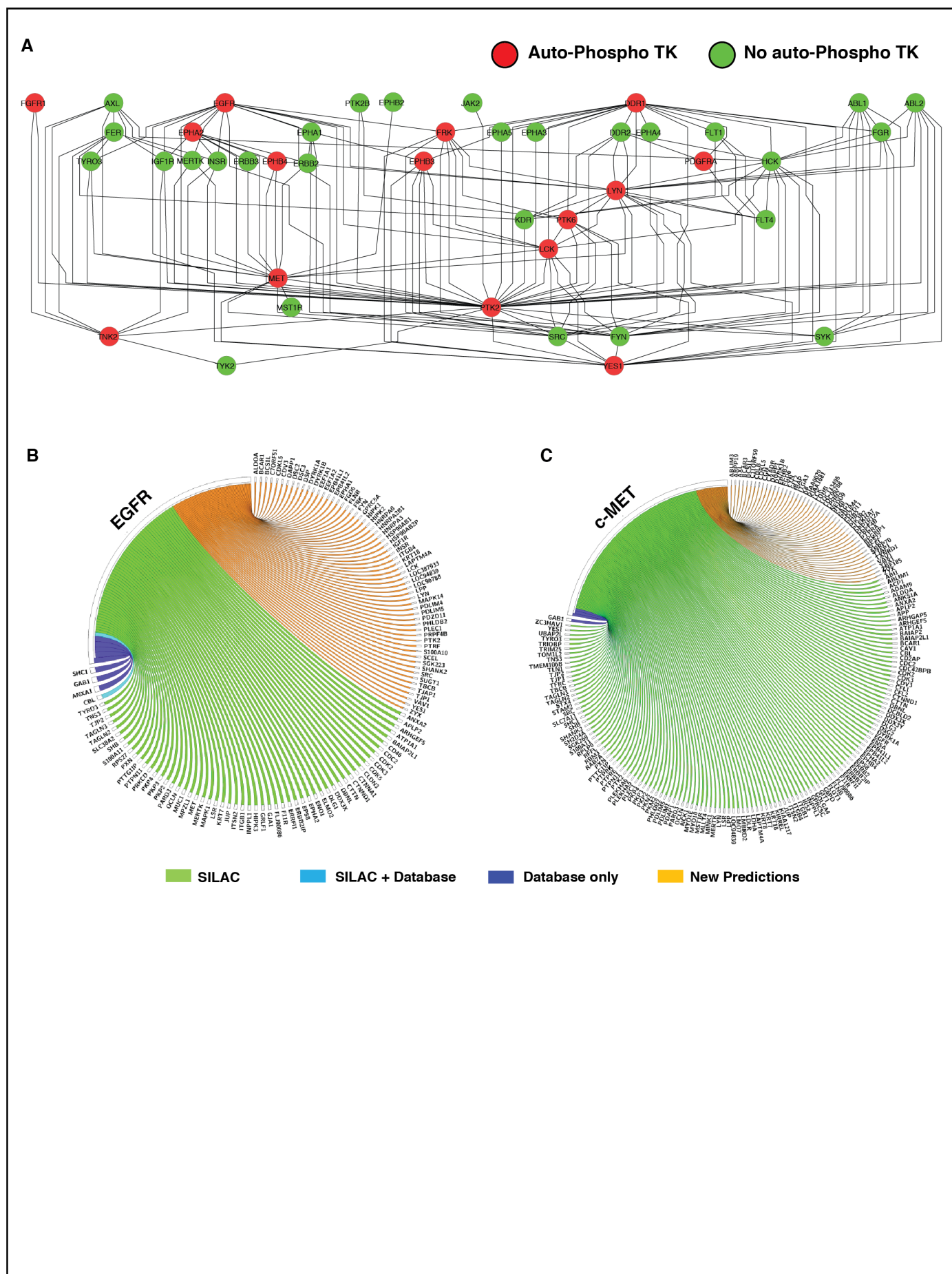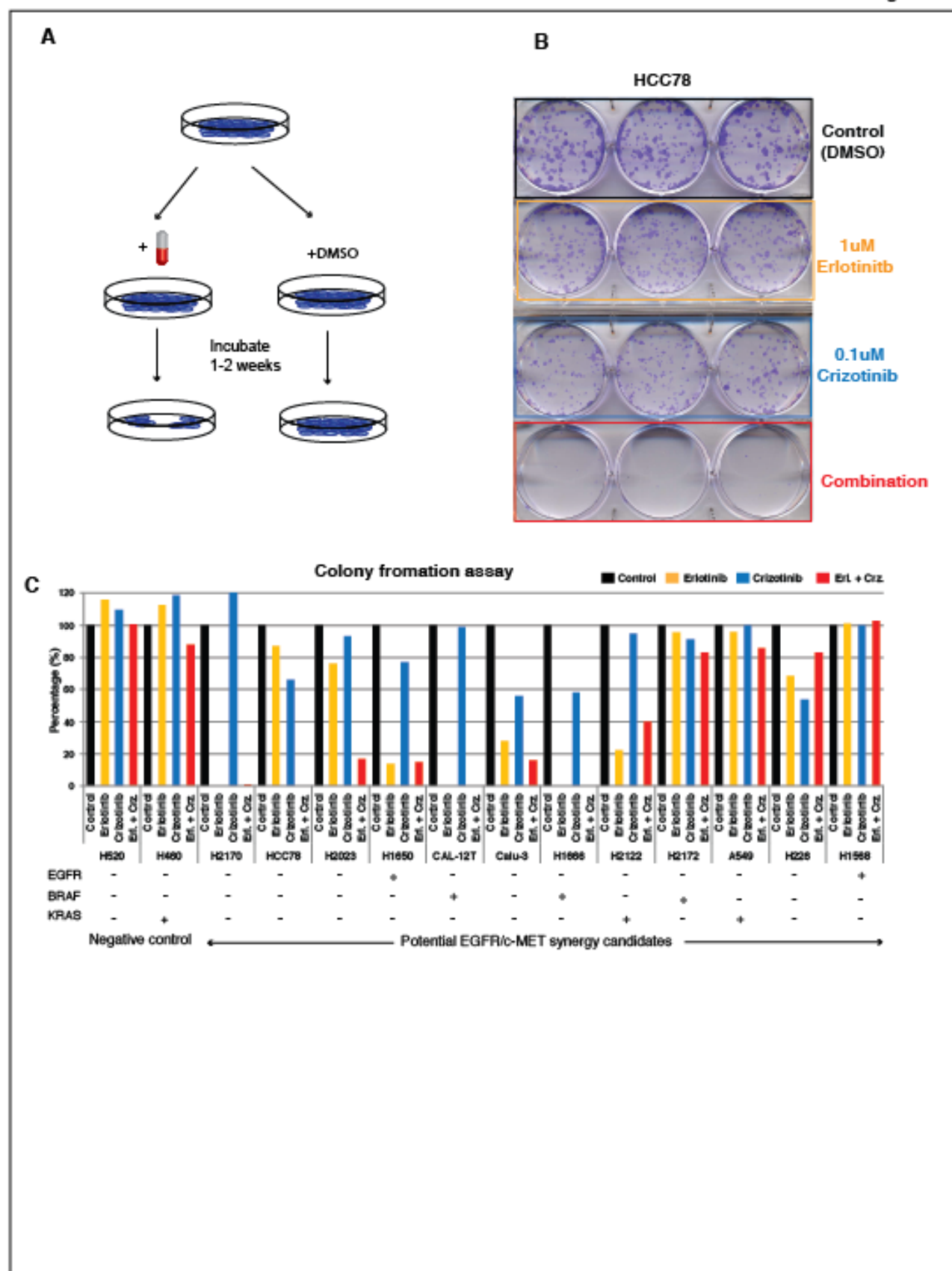
824

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6