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

Mukesh Bansal[1,2,*#], Michael Peyton[3,*], Jing He[4,6,*], Manjunath Kaustagi[4], Archana Iyer[4], Michael Comb[10], Michael White[3], John Minna[11], Andrea Califano[4,5,6,7,8,9,#]

[1] Psychogenics Inc. Tarrytown, New York, USA
[2] Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York
[3] Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA.
[4] Department of Systems Biology, Columbia University, New York, NY
[5] Center for Computational Biology and Bioinformatics (C2B2), Columbia University, New York, NY, USA
[6] Department of Biomedical Informatics (DBMI), Columbia University, New York, NY, USA
[7] Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY
[8] Institute for Cancer Genetics, Columbia University, New York, NY
[9] Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY
[10] Cell Signaling Technology, 3 Trask Lane, Danvers, MA 01923, USA
[11] Hamon Center for Therapeutic Oncology Research, Simmons Comprehensive Cancer Center, Departments of Pharmacology, and Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas

* These authors contributed equally
[#] Correspondence should be address to M.B (mukesh.bansal@psychogenics.com); A.C. (ac2248@cumc.columbia.edu)

Summary

Signal transduction pathways are largely based on compilation of individual post-translational modification assays in 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 present systematic, computational reconstruction of tyrosine kinase (TK) signal transduction pathways, based on mass spectrometry-based proteomics profiling of phosphotyrosine-enriched

peptides from 250 samples representative of lung adenocarcinoma. The network represents 43 TKs and their predicted 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. Availability of the network allowed prediction of drug response in individual samples, including accurate prediction of synergistic EGFR and c-MET inhibitor activity in cells lacking mutations in either gene, which was experimentally validated, thus contributing to current precision oncology efforts.

Keywords: signaling network, lung cancer, tyrosine kinase, combination therapy, proteomics, EGFR, c-MET

41 # Introduction

42 Lung adenocarcinoma (LUAD) is a leading cause of cancer related deaths in United States,
43 representing 40% of the 225,500 new lung cancer cases every year and a 5-year survival rate of
44 only 16 % (1). Excluding immunotherapeutic agents, which have recently shown significant
45 success in a relatively small subset of patients (2), the most effective targeted therapies for this
46 diseases were designed to inhibit tyrosine kinases harboring genetic alterations inducing aberrant
47 activation of downstream signaling pathways (3-7). These events primarily include EGFR
48 mutations and ALK-EML4 fusion events, in ~15% and ~3–7% of LUAD patients, respectively (8,
49 9). Yet, while targeted therapy is initially effective in a significant fraction of patients harboring
50 these actionable alterations, the vast majority of them will either fail to respond or develop
51 resistance to mono-therapy (10, 11). In addition, most patient lack actionable alterations
52 altogether. This suggests that novel approaches are critically needed.

53 A possible alternative to minimize emergence of resistance is combination therapy, a strategy
54 that has been shown to be effective in many metastatic tumors, such as breast cancer, and acute
55 myeloid leukemia (12-14). However, systematic identification of effective drug combinations on a
56 genetic alteration basis is difficult, because the number of patients presenting multiple actionable
57 events is extremely low. As a result, combination therapy is generally hypothesized and tested on
58 an empirical basis or based on elucidation of complex mechanisms of tumor cell adaptation. To
59 address these limitations, we and other have proposed that rational design of combination therapy
60 will require a more mechanistic and tumor-context-specific understanding of the molecular
61 interactions that underlie their potential synergistic activity, starting with tyrosine kinases, which
62 represent a critical class of pharmacological targets in cancer (15). Such an approach requires
63 methodologies for the accurate and systematic elucidation of tumor-specific signaling
64 transduction pathways.

65 Dissection of signal transduction networks represents a complex endeavor, requiring elucidation
66 of hundreds of thousands of tissue-specific molecular interactions that mediate the post-
67 translational modification of protein substrates. *In vitro* approaches generally fail to capture the
68 tissue-specific nature of these interactions, thus providing "average" signal transduction pathways
69 that are both incomplete and inaccurate. In addition, experimental approaches that have been
70 successful in accelerating the analysis of molecular interactions in transcriptional regulation and
71 protein-protein interaction in stable-complexes, such as those based on co-expression or yeast-
72 2-hybrid assays, do not easily translate to elucidating signaling interactions. Similarly, approaches
73 based on the use of phospho-specific antibodies, while elegant and effective, are limited to only
74 a handful of proteins. Computationally, compared to the many algorithms that have been
75 developed for the reverse engineering of transcriptional and protein-complex interactions (16, 17),
76 only a handful of experimentally validated algorithms are available for the dissection of signaling
77 networks, none of which works at the proteome-wide level or is tumor-context specific (16, 18,
78 19).

79 Recent availability of proteome-wide molecular profile data, characterizing the abundance of
80 phospho-tyrosine-enriched peptides by liquid chromatography coupled to tandem mass
81 spectrometry (LC-MS/MS), suggests that additional methodologies may be developed to extend
82 approaches that have been successfully applied to the dissection of transcriptional networks from
83 gene expression profiling. In this manuscript, we propose extending the Algorithm for the
84 Reconstruction of Accurate Cellular Networks (ARACNe) (26) for the reverse engineering of
85 signal transduction networks from large-scale phosphoproteomic profiles. The new method,
86 *pARACNe*, addresses critical issues that prevented the direct application of the original ARACNe
87 algorithm on phosphoproteomic datasets. In particular, it has been modified to handle the highly
88 sparse nature of phosphopeptide abundance data, the large amount of noise and missing data,

89 and the degenerate peptides-to-protein mapping, as resulting from the use of LC-MS/MS-based
90 assays.

91 When used to analyze genome-wide phosphoproteomic data from 245 lung cancer samples,
92 including fresh-frozen biopsies and cell lines for non-small-cell lung carcinoma, as well as normal
93 lung tissue, pARACNe produced a dense network connecting 46 tyrosine kinases (TK) with 415
94 substrates (including 377 proteins without TK activity). This represents the first genome-wide,
95 tumor-context-specific model for a TK signal transduction network, capturing both protein-specific
96 and phospho-site specific events. We validated substrate predictions for two "hubs," whose
97 activity may play a key role in determining sensitivity to Erlotinib and Crizotinib, two FDA-approved
98 drugs for LUAD, including the EGFR and c-MET tyrosine kinases by independent SILAC assays
99 and database analysis, with an accuracy of over 60%. Of particular note, the pARACNe-predicted
100 TK-substrate network provided unique information about tyrosine kinase auto-phosphorylation,
101 either direct (cis) or via a second kinase (trans). Analysis of the resulting TK-network, via master
102 regulator analysis (27), recapitulated established genetic determinants of LUAD and identified
103 predictive biomarkers of response to Erlotinib and Crizotinib combination therapy, which were
104 validated in an independent set of LUAD cell lines, the majority of which harbored no genetic
105 alterations in the corresponding genes. Furthermore, predictions based on the analysis of the
106 corresponding patient cohort were strongly supported by genomic information, suggesting
107 potential value in using these analyses for the identification of effective combination therapies in
108 precision oncology.

109

110 # Results

111 ***Overview of the pARACNe algorithm***

112 Enzymatic activity of tyrosine kinase (TK) proteins – as assessed by the ability to phosphorylate
113 their downstream substrates – is effectively determined by their phosphorylated isoform
114 abundance (**Fig. 1A, B**). Therefore, we reasoned that computational inference of TK-substrate
115 interactions (**TK→S**) could be effectively performed by measuring the dependency between their
116 respective phospho-states, as measured over a large compendium of samples **(Fig. 1C)**.
117 Unfortunately, due to signal transduction cascade complexity and pathway cross-talk, such
118 dependencies can manifest between protein pairs that are not involved in direct **TK→S**
119 interactions. The ARACNe algorithm, which we had previously designed for the reverse
120 engineering of transcriptional networks, effectively addresses this problem by leveraging the Data
121 Processing Inequality (28), a critical property of mutual information (29) , that effectively allows
122 disambiguating between direct and indirect interactions. This is accomplished by assessing
123 whether the information transfer on a candidate direct interaction (e.g., $\mathbf{TK_1 \rightarrow S}$) is greater than
124 on all other indirect paths (e.g., $\mathbf{TK_1 \rightarrow TK_2 \rightarrow S}$). ARACNe has been highly successful in the
125 experimentally validated dissection of transcriptional networks via analysis of large gene
126 expression profile compendia. ARACNe-inferred targets of transcription factors were validated in
127 multiple cellular contexts, with an accuracy of 70% to 80% (26, 30-33).

128 However, ARACNe relies on molecular profile data that is both continuous and non-sparse,
129 properties that are not always provided by quantitative proteomic data sets, which can be
130 generated by a variety of methods. Those based on LC-MS/MS represent the most popular
131 approaches [PMID: 27629641], but different implementations have specific performance profiles
132 in terms of analyte throughput, consistency of measurement of peptides across samples and
133 linear dynamic range [PMID: 27049628]. Depending on the data acquisition method, one or both

134 of these assumptions of ARACNe are violated in proteome-wide datasets generated by the most
135 popular methods based on data-dependent acquisition. Particularly when employing
136 quantification by spectral counting, as is typically conducted for global protein-protein interaction
137 studies [PMID: 26186194, 28514442], phosphoproteomic data is both discrete (i.e., generally
138 represented by spectral counts), very sparse, with a majority of peptides having zero spectral
139 counts, and presenting a significantly skewed distribution for low-abundance peptides.

140

141 To address these limitations, we developed an entirely novel algorithm, pARACNe (phospho-
142 ARACNe) (**Fig. 1C**), specifically designed to measure phospho-state dependencies between TKs
143 and their candidate substrates using large-scale LC-MS/MS phosphoproteomic profiles.
144 pARACNe thus extends the original ARACNe framework to allow systematic inference of **TK→S**
145 interactions. Specifically, to handle the highly discrete nature of the data, pARACNe replaces the
146 kernel-density and adaptive partitioning based mutual information estimators with the naive
147 method (**Fig 1C4**), using gold standard data to select the most effective number of bins [12] (see
148 Methods). Furthermore, to deal with the skewed spectral count distribution, we introduce an
149 iterative quantile discretization method, where samples are binned together, based on their
150 spectral counts, to produce a distribution as close to uniform as possible (**Fig 1C3,** Methods).

151

152 ***pARACNe-inferred LUAD-specific TK-phosphorylation network***

153 We used pARACNe to reconstruct a LUAD-specific TK-signaling network, by analyzing
154 phosphopeptide profiles obtained from 245 LUAD samples (34). These data represent the
155 abundance of peptides containing at least one phospho-tyrosine, as obtained by
156 phosphoproteomic analysis of 46 LUAD cell lines, 151 LUAD tumors, and 48 adjacent normal
157 samples. LC-MS/MS profiling identified 3920 phospho-tyrosine residues on approximately 2600

158 different proteins. Based on these data, pARACNE identified a total of 2611 peptide-peptide
159 interactions, which could be further mapped to 2064 unique **TK→S** interactions (**Suppl. Table 1)**.
160 The latter represent interactions between 46 distinct TKs and their candidate substrates, including
161 174 $\mathbf{TK_1 \rightarrow TK_2}$ interactions between two TKs (**Fig 2A**), which represents a statistically significant
162 bias toward TK-TK interactions in the network ($p = 10^{-62}$). This suggests that TKs may form a more
163 densely inter-connected subnetwork within the complete TK signaling network than previously
164 assessed, thus potentially providing critical novel information about adaptive response, pathway
165 cross-talk, and auto-regulatory loops. Indeed, such highly interconnected structure provides
166 potential functional advantage compared to less connected or "flat" architectures, including the
167 ability to provide more fine-grain response to a highly heterogeneous variety of exogenous signals
168 and conditions, the ability to provide rapid adaptive response to changing stimuli, and the ability
169 to preserve cell state via autoregulatory feedback. Interestingly, in contrast to transcriptional
170 networks, the vast majority of pARACNe-inferred interactions have a positive Spearman
171 correlation, with higher counts of TK-mapped phosphopeptides corresponding to higher counts of
172 candidate substrate-mapped ones. This is consistent with the fact that, in contrast to
173 transcriptional networks where interactions may activate or suppress expression of a target gene,
174 TKs only phosphorylate their substrates, thus inducing positive phospho-state correlation. Only a
175 negligible number of interactions (0.5%) were associated with a negative Spearman correlation
176 ($N = 11$, $p \leqslant 0.05$). These may represent either indirect interactions where the TK activates a
177 substrate-specific phosphatase or direct interactions where phosphorylation of one phosphosite
178 may prevent phosphorylation of another site on the same protein.

179

180 ***LUAD Network Accuracy and Sensitivity Analysis***

181 To estimate the accuracy of the inferred TK-signaling network, we investigated the substrates of
182 two TK-proteins, EGFR and c-MET, representing high-affinity binding targets of existing FDA-
183 approved TK inhibitors for LUAD. Specifically, we compared their pARACNe-inferred substrates
184 to those reported in the phosphoDB database (35) and those supported by experimental
185 evidence, based on previously published SILAC assays, following cell line treatment with
186 associated, selective TK inhibitors. pARACNe inferred 123 EGFR substrates (**Fig 2B**). Of these,
187 5 (blue and cyan) were included as high-confidence EGFR substrates in phosphoDB, out of 13 in
188 total (38%), including the established EGFR auto-phosphorylation site. Moreover, 50 additional
189 proteins (45%, green) showed significant decrease (alteast 2 fold) in the abundance of their
190 phosphorylated isoforms in SILAC assays (34), following treatment of H3255 cells with the EGFR
191 inhibitor Gefitinib. Similarly, pARACNe predicted 179 c-MET substrates (**Fig 2C)**. Notably, both
192 of the established substrates reported in PhosphoDB were identified by pARACNE (100%, blue).
193 Moreover, 126 additional proteins (71.5%, blue) showed significant decrease in the abundance
194 of their phosphorylated isoforms in SILAC assays(34), following treatment of MKN45 cells with
195 the first-generation c-MET-specific inhibitor Su11274.

196 We used MKN45, even though it represents a gastric cancer cell line, to assess the overall
197 conservation of predictions in different tissues. Indeed, while specific substrates are likely to be
198 highly tumor-context specific, signaling networks should be significantly more conserved than
199 transcriptional ones, since lineage-specific chromatin states represent a major determinant of
200 transcriptional regulation, while signal transduction is only affected indirectly by the availability of
201 specific partner proteins. As a result, it is reasonable to expect that an even greater overlap of
202 inferred vs. SILAC positive substrates may be achieved in native LUAD cells.

203

204 Taken together, these data suggest that pARACNe can identify a much larger subset of candidate
205 substrates, while both identifying a significant proportion of established substrates (46% on
206 average, based on phosphoDB) and maintaining high accuracy (~60% on average, by SILAC
207 assays). This also suggest that, similar to transcription factor targets reported in the literature, TK
208 substrates are poorly characterized, even for highly relevant and exceedingly well-studied kinases
209 such as EGFR and c-MET. As a result, pARACNe could provide significant novel hypotheses for
210 TK→S interactions that can be validated as required. We should also note that the reported
211 accuracy for pARACNe is estimated using SILAC data on a single cell line. SILAC assays have
212 significant false negatives and it would be reasonable to expect that, once tested in additional cell
213 lines, the accuracy of pARACNe could further increase. As a further performance benchmark, we
214 used the same SILAC benchmark to test predictions by NetworkIN, a reverse engineering method
215 based on protein sequence motif analysis and protein association networks (16). The analysis
216 found almost no consensus with SILAC assays, with only one out of 33 NetworkIN-predicted
217 EGFR substrate identified as significantly dephosphorylated following treatment with TK-specific
218 inhibitors.

219

220 ***Systematic, network-based inference of pharmacological dependencies***

221 Once an accurate model of signal transduction in LUAD cells was established by pARACNe
222 analysis, we interrogated the corresponding TK→S network using phosphoproteomic signatures
223 from 46 LUAD cell lines to identify key dependencies for experimental validation. For this purpose,
224 we extended the Virtual Proteomics by Enriched Regulon analysis (VIPER)(37), which had been
225 originally developed to identify key proteins controlling the transcriptional state of tumor-related
226 cells, on an individual sample basis. Specifically, VIPER was designed to identify these proteins
227 by assessing the enrichment of their transcriptional targets in differentially expressed genes, in a
228 tumor-specific gene expression signature. VIPER, which extends the Master Regulator Inference

229 algorithm (MARINa) to support analysis of single samples, has been instrumental in identifying
230 master regulator proteins representing key functional determinants of tumor-related phenotypes
231 in many cancer types, from glioblastoma (32, 33, 38), lymphoma (39, 40), and leukemia (41) to
232 prostate (42-44) and breast adenocarcinoma(45-47), among others. We thus reasoned that
233 VIPER could be modified to identify master regulator TKs, most likely to induce the differential
234 phosphorylation pattern observed in a specific tumor sample (**Fig. 3A**). A specific additional value
235 of the algorithm is that it could not only identify individual MR TK proteins, representing individual,
236 pharmacologically accessible dependencies of the tumor, but also TKs representing potential
237 synergistic MR-pair as candidate dependencies for combination therapy.

238 VIPER can be easily modified to analyze phosphoproteomic signatures. Specifically, rather than
239 assessing the enrichment of a protein's transcriptional targets (regulon) in differentially expressed
240 genes, pVIPER is designed to measure the enrichment of a TK's substrates (signalon) in
241 differentially phosphorylated proteins. We performed pVIPER analysis at phosphopeptide specific
242 level, where dependencies were assessed based on individual phosphopeptide ather than whole
243 protein state, and were finally combined the result of all phosphopeptide mapping to the same
244 protein. Consistent with VIPER's experimentally validated ability to identify synergistic master
245 regulators proteins by transcriptional network analysis (32, 40, 42), pVIPER can identify candidate
246 synergistic TKs by assessing the statistical significance of the enrichment of the common
247 substrates of both TKs over the enrichment of unique substrates of each individual TK (see
248 Method section). Systematic VIPER analysis of phosphoproteomic profiles from 46 LUAD cell
249 lines generated between 2 and 13 master regulator TKs or synergistic TK-pairs, as candidate

250 pharmacologically actionable dependencies, for each cell line, thus generating a plausible number
251 of hypothesis for each line (**Fig. 3B** and **Fig 3C**).

252

253 **VIPER identifies LUAD-specific dependencies**

254 VIPER analysis inferred several TK proteins as highly conserved individual dependencies across
255 multiple cell lines, including the Ephrin type-A receptor 2 (*EPHA2*), epidermal growth factor
256 receptor (*EGFR*), c-Met proto-oncogene (*MET*), and HER2 receptor tyrosine kinase 2 (*ERBB2*),
257 suggesting a critical role of these proteins in the maintenance of LUAD cell line state. This is also
258 in agreement with the functional role of these genes and the use of inhibitors of these kinases
259 across a large panel of patients in multiple cancer types (48-52).

260 In contrast to these established LUAD cell line dependencies, we also identified several TKs as
261 dependencies of specific cell lines. This can either be the result of associated genetic or
262 epigenetic alterations in these cell lines or the result of field effects, where multiple genetic
263 alterations or alterations in upstream pathway contribute to the cell line dependency on a specific
264 TK activity. For instance, we identified ALK (Anaplastic Lymphoma Receptor Tyrosine Kinase) to
265 be addiction point only in H2228 cell line. ALK is a conserved trans-membrane receptor tyrosine
266 kinase (RTK) protein in the insulin-receptor super family. Chromosomal alterations involving *ALK*
267 translocations and fusion events have been identified in several cancer types including LUAD (53,
268 54), diffuse large B-cell lymphomas (55), neuroblastoma (56), and inflammatory myofibroblastic
269 tumors (57), among others. Additionally, ALK fusion events with other genes, including EML4
270 (Echinoderm Microtubule-associated protein Like 4) in LUAD lead to aberrant protein activity
271 eliciting “oncogene addiction” (54). Presence of ALK-EML4 fusion transcripts, in ~3–7% of LUAD
272 patients (58-60), is a strong predictor of response to ALK inhibitors, such as Crizotinib, among
273 others (61, 62). Interestingly, among all available LUAD cell lines for which a phosphoproteomic

274 profile was available, H2228 was the only one with an established ALK-EML4 fusion event and
275 with established sensitivity to ALK inhibitor (63, 64). This further reflects the specificity of our
276 analysis as this was the only cell line predicted to depend on ALK activity. Interestingly, we
277 identified 4 additional H2228 dependencies, namely EGFR, Epha2, c-MET, and PTK2. H2228
278 sensitivity to EGFR inhibitors, in combination with ALK inhibitors, was already established (64).
279 In addition, we confirmed dependency of H2228 cells on the other 3 TKs (Epha2, c-MET and
280 PTK2) by cell viability assay following siRNA-mediated silencing of the corresponding gene,
281 (**Suppl. Fig. 2A**). BrdU assays confirmed H2228 sensitivity to c-MET and PTK2 silencing (**Suppl.
282 Fig. 2B).** Surprisingly, in contrast to inhibition by small molecule inhibitor, siRNA-mediated EGFR
283 silencing did not show decrease in cell viability, while ALK silencing showed little effect. This may
284 be due to the lower degree of siRNA-mediated inhibition, compared to specific inhibitors, or to the
285 multi-kinase effect of the latter. For instance, consistent with their identification as cell line specific
286 dependencies, Crizotinib is both an ALK and a c-MET inhibitor. Indeed, although siRNA-mediated
287 c-MET inhibition was modest, BrdU incorporation assays showed significant effect on cell viability.
288

289 ***EGFR and c-MET are predicted as dependencies in multiple LUAD cell lines***

290 As discussed, pVIPER analysis revealed several TK-pairs as candidate synergistic dependencies
291 across several cell lines, such as Epha2/c-MET, EGFR/PTK2, EGFR/Epha2, Epha2/c-MET, and
292 EGFR/c-MET. Among these the EGFR/c-MET pair emerged as the most conserved synergistic
293 TK-pair across the available cell lines. In addition, EGFR and c-MET were also identified as
294 candidate TK MRs in several of these cell lines, suggesting either a complementary or synergistic
295 role for these proteins and a potential therapeutic opportunity for combination therapy in LUAD
296 (65, 66).
297

298 To validate pVIPER-predicted, cell line specific EGFR/c-MET synthetic lethality, we selected a
299 panel of 14 cell lines, 11 of which were predicted to be synergistically dependent on EGFR/c-MET

300 (H226, H2122, H1666, H2172, Cal-12T, H2023, H1568, Calu-3, H1650, HCC78, and A549), as
301 well as 3 negative controls with no predicted synergistic or individual dependencies on the two
302 TKs (H2170, H460, and H520). To measure sensitivity to these agents, we used two different and
303 complementary assays, including: (a) colony formation assay to assess long term sensitivity (**Fig.**
304 **4A and Methods**) and (b) 3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide (MTT)
305 assay for short term sensitivity analysis (**Fig. 5A and Methods**). For colony formation assays,
306 cells were treated with either an EGFR inhibitor (Erlotinib, 1uM) or a c-MET/ALK inhibitor
307 (Crizotinib, 0.1uM), either individually or in combination (see Methods). To evaluate synergistic
308 dependency on EGFR/c-MET we used Excess Over Bliss (67), which measures the difference
309 between the observed effect on colony formation and the effect expected from a purely additive
310 model. For MTT assay, first, cells were treated with EGFR inhibitor (Erlotinib) or MET inhibitor
311 (Crizotinib) individually at various concentrations to identify IC50 (concentration resulting in 50%
312 cell death). Next, cells were treated with 1 uM of Erlotinib and varying concentrations of Crizotinib
313 to identify combinations resulting in IC50 and used the combination index (CI) statistic (68) to
314 measure interaction between the two drugs.
315
316 Across all 11 cell lines tested in colony formation assay, 8 showed significant sensitivity to either
317 individual inhibitors ($H226_{E,C}$, $H2122_{E}$, $H1666_{E,C}$, $Cal\text{-}12T_{E}$, $Calu\text{-}3_{E,C}$, $H1650_{E}$) or synergistic
318 sensitivity to the combination ($HCC78_{E+C}$ and $H2023_{E+C}$) (**Fig 4B-C and Fig S3)**. Surprisingly, all
319 of these cell lines were $EGFR^{WT}$, ALK^{WT}, and $c\text{-}MET^{WT}$, except for H1650, which was $EGFR^{Mut}$.
320 Thus, based on standard of care criteria, 7 out of 8 cell lines would not have been considered as
321 sensitive to either EGFR or ALK/c-MET inhibitors. Several cell lines presented striking sensitivity
322 to either one ($H2122_{E}$, $Cal\text{-}12T_{E}$, $H1650_{E}$) or both inhibitors ($H226_{E,C}$, $H1666_{E,C}$, $Calu\text{-}3_{E,C}$) in
323 isolation, thus making the assessment of synergistic drug sensitivity difficult. In addition, three
324 $EGFR^{WT}$ cell lines harboring BRAF (Cal-12T and H1666) or KRAS (H2122) mutations were also
325 highly sensitive to Erlotinib as a single agent, as predicted by pVIPER, despite the fact that KRAS

326 pathway mutations are mutually exclusive with EGFR mutations and predictive of Erlotinib
327 resistance (**Fig. 4C**). Finally, none of these cell lines was predicted to be sensitive to ALK
328 inhibitors, suggesting that Crizotinib sensitivity is mediated by c-MET specific dependencies. Of
329 the negative controls, only one (H2170) showed high sensitivity to Erlotinib. Taken together, 8/11
330 cell lines (73%) predicted as sensitive to the inhibitors were validated long term colony formation
331 assays, while only 1/3 negative controls showed sensitivity to these agents (33%).

332

333 To evaluate the short-term interaction between EGFR and c-MET, we performed MTT assay
334 across 11 cell lines (HCC78, H2023, H1650, Calu-3, H2172, H2122, H1568, A549, H1666, H520
335 and H2170) including 2 negative control cell lines (H520 and H2170). Similar to colony formation
336 assay, we found synergistic sensitivity to EGFR and c-MET inhibitors in 6/9 cell lines (67%), with
337 5 cell lines showing strong synergy (CI ≤ 0.8) (**Fig. 5B-C**) and 1 borderline synergy (CI = 0.82),
338 showing the consistency between two assays. However, for two cell lines, H1666 and H2170 (a
339 negative control), results were inconsistent between long term colony formation and MTT assays.
340 For both H1666 and H2170 cell lines, colony formation and MTT assay to showed sensitivity to
341 Erlotinib alone, where colony formation assay has complete abrogation of colonies at 1 μM of
342 Erlotinib, and later had IC_{50} =1.25 μM and 3.7 μM for H1666 and H2170 respectively. However,
343 in combination therapy, MTT assay showed antagonism (CI >1), despite the fact that colony
344 formation assay still showed complete abrogation which could be associated to the accumulation
345 of new mutations in these cell lines. However, this is just hypotheses and needs to be verified by
346 further experiments such as sequencing of these cell lines pre-and post-treatment.

347

348 ***Phosphosite specific phosphorylation predicts EGFR/c-MET inhibitor synergy.***

349 In previous section, we assessed the pVIPER predictions after consolidating the result at protein
350 level. Following the results from MTT and colony formation assays, we reanalyzed the pVIPER

351 predictions at the phosphopeptide level. Interestingly, this revealed that whenever synergistic
352 EGFR/c-MET dependencies were predicted from phosphosite $EGFR_{1197}$ and phosphosites other
353 than $c\text{-}MET_{1003}$ (H1666, Cal-12T, H1650), cell lines responded to Erlotinib in isolation, while when
354 predictions were based on phosphosites $EGFR_{1197}$ and $c\text{-}MET_{1003}$ (HCC78, H2023, and Calu-3),
355 cells exhibited bona fide synergistic sensitivity to the two inhibitors, with the only possible
356 exception of Calu-3, which showed synergistic sensitivity in MTT assays and additive sensitivity
357 to both inhibitors in colony formation assays. Conversely, when predictions were not based on
358 either phosphotyrosine, cells exhibited no sensitivity to the individual inhibitors or the combination
359 (H2172, H226, A549, H460, H520, H1568). Thus, predictions based on these two phosphosites
360 produced no false positives (6 out of 6 predicted and validated as non-sensitive) and only 2 false
361 negatives (H2170 and H2122), resulting in an error rate of only 2 out of 14 cell lines (14%, $p =$
362 0.0093 using fisher exact test).
363
364 This finding is in agreement with the established role of $EGFR_{1197}$ as a predictor of EGFR inhibitor
365 sensitivity (69). Intriguingly, when sensitivity was predicted using phosphosites other than
366 $EGFR_{1197}$ and $c\text{-}MET_{1003}$, cell lines did not respond to the inhibitors, either individually or in
367 combination. For these two peptides, we found their common substrates to be hyper
368 phosphorylated in the sensitive cell lines **(Fig. 6A**) compared to the specific substrates of each of
369 them, whereas cell line responding only to EGFR inhibitors showed more hyper phosphorylation
370 of EGFR only substrates (**Fig. 6B**). Cell lines resistant to both EGFR and c-MET inhibitors either
371 showed no change in the phosphorylation status or hypo-phosphorylation compared to the normal
372 samples (**Fig. 6C**). Hence, either the common substrates of EGFR and c-MET, or the
373 phosphorylation status of $EGFR_{1197}$ and $c\text{-}MET_{1003}$ could potentially be used as biomarkers for
374 predicting therapy with the dual inhibitors. However, this conclusion is based on a very limited
375 number of observations and lacks the statistical power. This finding needs a re-

evaluation/validation using larger cohort of samples to establish an appropriate biomarker for combination therapy.

Addictions points in patient samples

Similar to cell lines, when applied on patient data, pVIPER identified EGFR as one of the most common addiction points (**Fig. 6**). We inferred EGFR dependency in 12 patients. Of these, 5 harbored EGFR mutation, while the remaining 7 patients had not been tested for this mutation, showing a high consistency between our predictions and the genetic predisposition for sensitivity to EGFR inhibitors. In the entire cohort, there were only 3 patients with EGFR mutations that were not identified as EGFR dependent by pVIPER, resulting in an overall sensitivity of 62.5% (5/8). However, it is well known that >50% of patients harboring EGFR mutations do not respond to Erlotinib, suggesting that these may not be false negatives but rather patients with low activation of downstream EGFR pathways, despite their $EGFR^{Mut}$ state. Similarly, our analysis identified candidate ALK dependencies in 4 patients. Of these one had an established TFG-ALK fusion, whereas the others had not been tested for ALK fusion events.

Across all patient samples, we observed Discoidin Domain Receptor-1 (DDR1) to be the most frequent addiction point, which was not predicted for any of the 46 cell lines. One reason for the difference is that DDR1 is collagen dependent and there may be differences in the 3D structure of the tumor and the cell lines growing on the plate. An independent study (70) in a cohort of 83 lung cancer specimens found that silencing of DDR1 in these samples leads to the hampering of cell survival, reduced invasiveness in collagen matrices, increased apoptosis in basal condition and decreased metastatic activity in model of tumor metastasis to bone, signifying it as a potential novel therapeutic target. CONCLUDING SENTENCE.

399 ## Discussion

400 In this paper, we developed pARACNe to infer Tyrosine Kinase (TK) signaling network using
401 genome-wide phosphoproteomic data from lung cancer. The network prediction was validated
402 using SILAC experiments, with high accuracy. Interrogation of the predicted TK-substrate network
403 generated biologically meaningful hypotheses, followed by experimental validations illustrating
404 the effectiveness of predicted kinase inhibitor combination, EGFR and c-MET combination
405 inhibitors, in treating lung cancer cell lines. Furthermore, Master Regulator Analysis using patient
406 proteomics data provides implications for using targeted agent combinations to treat patients
407 based on patient proteomic data.
408
409 Notably, pARACNe is significant and powerful as of its genome-wide scale and context-specificity,
410 which were missing by previous methods. For example, methods proposed by Linding et al. (16)
411 combine motif-based phosphosite predictions with information of physical association, co-
412 occurrence, and co-expression to identify substrates with high specificity and accuracy, but with
413 low coverage and lack of contextual specificity. Bender et al (17) used reverse phase protein
414 assay data after various stimulations to cells and inferred signaling network using hidden Markov
415 models and genetic algorithms. Even though the resulting networks are context specific, they lack
416 genomic-scale coverage. There have been methods which used existing large-scale protein
417 networks and prune them using transcriptomic information to identify signaling pathways (20-22).
418 In addition, attempts have been made to reconstruct signaling network using gene expression
419 data (23, 24). However, as signaling complexity lies mostly in upper level of cellular processes,
420 inferring the cascades from downstream gene expression fails to capture all the dynamics. Also,
421 PrePPI proposed by Zhang et al. (25) used protein structure-based methods to infer global
422 protein-protein interaction, but this approach fails to address phosphorylation context specificity.
423 Innovative uses of multiplex and microarray-based approaches, where multiple antibodies can be

424 used to probe an ensemble of phosphoproteins, are finally becoming sufficiently mature to allow
425 characterization of small pathways. Yet, these methods are still far from providing an unbiased,
426 genome-wide view of signal-transduction processes and continue to be completely dependent on
427 antibody specificity and availability. Similarly, assays developed specifically to monitor
428 phosphorylation pathways, such as Stable Isotope Labeling with Amino acids in Cell culture
429 (SILAC), provides a simple and straightforward approach to detect differential protein abundance.
430 Coupled with phosphorylation enriched assays, it can provide high quality quantification for post-
431 translation phosphorylation changes in cell lines. However, these methods are 1) laborious and
432 costly; 2) can only be performed to dissect the substrates of a single enzyme at a time and 3) do
433 not differentiate between direct and indirect targets.
434
435 To be noted, as the LC-MS/MS experiments used here was generated based on Tyrosine-kinase
436 enrichment, which is only about ~2% of whole phosphoproteome. pARACNe is shown only on
437 TK-substrates network. The current methodology could be extended to whole phosphoproteomic
438 data based signaling network reconstruction where the data is available. In addition to label free
439 based LC-MS/MS proteomics data used in this work, label based approaches, such as ITRAQ or
440 TMT, could generate higher throughput whole proteomic profiles which might require future
441 redesign of ARACNe to incorporate both kinases and phosphatases in regulating their
442 downstream substrates. It is reasonable to expect that a version of ARACNe developed
443 specifically to dissect signaling networks should work at least as well as its transcriptional
444 counterpart. Since the relationship between the mRNA abundance of a gene encoding a
445 transcription factor (TF) and the activity of the corresponding protein is much looser than that
446 between the abundance of a phospho-isoform of a kinase and its enzymatic activity.
447
448

Even though research has attempted to identify addiction points based on gene expression data (71), predictions based on phosphoproteomic data appear superior in a way that they can directly reflect contextual specific signaling activity and are able to be directly targeted by kinase inhibitors. It is important to note that clinically, only patients with base-pair deletion at exon 19 (del746_A750) or a point mutation at exon 21 mutation (L858R) in EGFR shows sensitivity to EGFR inhibitor such as Cedirinib or Erlotinib (72).

Acknowledgments

We thank George Rosenberger for manuscript advice. This work was supported by the National Cancer Institute (NCI) Cancer Target Discovery and Development program (1U01CA168426 to A.C., and 1U01CA176284 to JDM), NCI Research Centers for Cancer Systems Biology Consortium (1U54CA209997 to A.C.), NCI Outstanding Investigator Award (R35CA197745-02 to A.C.), NCI SPORE in Lung Cancer (P50CA70907 to JDM), and CPRIT (Award RP110708 to JDM).

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.

471 # References

472 1. Ettinger DS, Akerley W, Borghaei H, Chang AC, Cheney RT, Chirieac LR, et al. Non-small cell lung
473 cancer. Journal of the National Comprehensive Cancer Network : JNCCN. 2012;10(10):1236-71.
474 2. Karasaki T, Nagayama K, Kawashima M, Hiyama N, Murayama T, Kuwano H, et al. Identification of
475 Individual Cancer-Specific Somatic Mutations for Neoantigen-Based Immunotherapy of Lung Cancer. J
476 Thorac Oncol. 2016;11(3):324-33.
477 3. Pasqualucci L, Kitaura Y, Gu H, Dalla-Favera R. PKA-mediated phosphorylation regulates the
478 function of activation-induced deaminase (AID) in B cells. Proc Natl Acad Sci U S A. 2006;103(2):395-400.
479 4. Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, Mukherjee P, et al. Functional organization
480 of the S. cerevisiae phosphorylation network. Cell. 2009;136(5):952-63.
481 5. Gainor JF, Varghese AM, Ou SH, Kabraji S, Awad MM, Katayama R, et al. ALK rearrangements are
482 mutually exclusive with mutations in EGFR or KRAS: an analysis of 1,683 patients with non-small cell lung
483 cancer. Clin Cancer Res. 2013;19(15):4273-81.
484 6. Janne PA, Shaw AT, Pereira JR, Jeannin G, Vansteenkiste J, Barrios C, et al. Selumetinib plus
485 docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-
486 controlled, phase 2 study. Lancet Oncol. 2013;14(1):38-47.
487 7. Mak KS, Gainor JF, Niemierko A, Oh KS, Willers H, Choi NC, et al. Significance of targeted therapy
488 and genetic alterations in EGFR, ALK, or KRAS on survival in patients with non-small cell lung cancer treated
489 with radiotherapy for brain metastases. Neuro Oncol. 2014.
490 8. Vendrell JA, Taviaux S, Beganton B, Godreuil S, Audran P, Grand D, et al. Detection of known and
491 novel ALK fusion transcripts in lung cancer patients using next-generation sequencing approaches. Sci Rep.
492 2017;7(1):12510.

493 9. Peters S, Camidge DR, Shaw AT, Gadgeel S, Ahn JS, Kim DW, et al. Alectinib versus Crizotinib in
494 Untreated ALK-Positive Non-Small-Cell Lung Cancer. N Engl J Med. 2017;377(9):829-38.
495 10. Tsvetkova E, Goss GD. Drug resistance and its significance for treatment decisions in non-small-
496 cell lung cancer. Curr Oncol. 2012;19(Suppl 1):S45-51.
497 11. Wilson FH, Johannessen CM, Piccioni F, Tamayo P, Kim JW, Van Allen EM, et al. A functional
498 landscape of resistance to ALK inhibition in lung cancer. Cancer Cell. 2015;27(3):397-408.
499 12. O'Shaughnessy J, Miles D, Vukelja S, Moiseyenko V, Ayoub JP, Cervantes G, et al. Superior survival
500 with capecitabine plus docetaxel combination therapy in anthracycline-pretreated patients with advanced
501 breast cancer: phase III trial results. J Clin Oncol. 2002;20(12):2812-23.
502 13. Pegram MD, Slamon DJ. Combination therapy with trastuzumab (Herceptin) and cisplatin for
503 chemoresistant metastatic breast cancer: evidence for receptor-enhanced chemosensitivity. Semin Oncol.
504 1999;26(4 Suppl 12):89-95.
505 14. Ravandi F, Cortes JE, Jones D, Faderl S, Garcia-Manero G, Konopleva MY, et al. Phase I/II study of
506 combination therapy with sorafenib, idarubicin, and cytarabine in younger patients with acute myeloid
507 leukemia. J Clin Oncol. 2010;28(11):1856-62.
508 15. Wu P, Nielsen TE, Clausen MH. Small-molecule kinase inhibitors: an analysis of FDA-approved
509 drugs. Drug Discov Today. 2016;21(1):5-10.
510 16. Linding R, Jensen LJ, Ostheimer GJ, van Vugt MA, Jorgensen C, Miron IM, et al. Systematic
511 discovery of in vivo phosphorylation networks. Cell. 2007;129(7):1415-26.
512 17. Bender C, Henjes F, Frohlich H, Wiemann S, Korf U, Beissbarth T. Dynamic deterministic effects
513 propagation networks: learning signalling pathways from longitudinal protein array data. Bioinformatics.
514 2010;26(18):i596-602.
515 18. Papin JA, Hunter T, Palsson BO, Subramaniam S. Reconstruction of cellular signalling networks and
516 analysis of their properties. Nature reviews. 2005;6(2):99-111.

517 19. Hashemikhabir S, Ayaz ES, Kavurucu Y, Can T, Kahveci T. Large-scale signaling network
518 reconstruction. IEEE/ACM Trans Comput Biol Bioinform. 2012;9(6):1696-708.
519 20. Scott J, Ideker T, Karp RM, Sharan R. Efficient algorithms for detecting signaling pathways in
520 protein interaction networks. Journal of computational biology : a journal of computational molecular cell
521 biology. 2006;13(2):133-44.
522 21. Steffen M, Petti A, Aach J, D'Haeseleer P, Church G. Automated modelling of signal transduction
523 networks. BMC bioinformatics. 2002;3:34.
524 22. Kelley BP, Sharan R, Karp RM, Sittler T, Root DE, Stockwell BR, et al. Conserved pathways within
525 bacteria and yeast as revealed by global protein network alignment. Proceedings of the National Academy
526 of Sciences of the United States of America. 2003;100(20):11394-9.
527 23. Shimoni Y, Fink MY, Choi SG, Sealfon SC. Plato's cave algorithm: inferring functional signaling
528 networks from early gene expression shadows. PLoS Comput Biol. 2010;6(6):e1000828.
529 24. Budak G, Eren Ozsoy O, Aydin Son Y, Can T, Tuncbag N. Reconstruction of the temporal signaling
530 network in Salmonella-infected human cells. Front Microbiol. 2015;6:730.
531 25. Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, Thu CA, et al. Structure-based prediction of protein-
532 protein interactions on a genome-wide scale. Nature. 2012;490(7421):556-60.
533 26. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of
534 regulatory networks in human B cells. Nat Genet. 2005;37(4):382-90.
535 27. Lefebvre C, Rajbhandari P, Alvarez MJ, Bandaru P, Lim WK, Sato M, et al. A human B-cell
536 interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol Syst
537 Biol. 2010;6:377.
538 28. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, et al. ARACNE: an
539 algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC
540 Bioinformatics. 2006;7 Suppl 1:S7.

541 29. MacKay DJC. Information theory, inference, and learning algorithms. Cambridge, UK ; New York:
542 Cambridge University Press; 2003. xii, 628 p. p.
543 30. Basso K, Saito M, Sumazin P, Margolin AA, Wang K, Lim WK, et al. Integrated biochemical and
544 computational approach identifies BCL6 direct target genes controlling multiple pathways in normal
545 germinal center B cells. Blood. 2010;115(5):975-84.
546 31. Della Gatta G, Palomero T, Perez-Garcia A, Ambesi-Impiombato A, Bansal M, Carpenter ZW, et al.
547 Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in
548 T-ALL. Nat Med. 2012;18(3):436-40.
549 32. Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, et al. The transcriptional network for
550 mesenchymal transformation of brain tumours. Nature. 2010;463(7279):318-25.
551 33. Sonabend AM, Bansal M, Guarnieri P, Lei L, Amendolara B, Soderquist C, et al. The transcriptional
552 regulatory network of proneural glioma determines the genetic alterations selected during tumor
553 progression. Cancer Res. 2014;74(5):1440-51.
554 34. Guo A, Villen J, Kornhauser J, Lee KA, Stokes MP, Rikova K, et al. Signaling networks assembled by
555 oncogenic EGFR and c-Met. Proc Natl Acad Sci U S A. 2008;105(2):692-7.
556 35. Giansanti P, Aye TT, van den Toorn H, Peng M, van Breukelen B, Heck AJ. An Augmented Multiple-
557 Protease-Based Human Phosphopeptide Atlas. Cell Rep. 2015;11(11):1834-43.
558 36. Awad MM, Shaw AT. ALK inhibitors in non-small cell lung cancer: crizotinib and beyond. Clin Adv
559 Hematol Oncol. 2014;12(7):429-39.
560 37. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Functional characterization of
561 somatic mutations in cancer using network-based inference of protein activity. Nat Genet.
562 2016;48(8):838-47.

563 38. Chudnovsky Y, Kim D, Zheng SY, Whyte WA, Bansal M, Bray MA, et al. ZFHX4 Interacts with the
564 NuRD Core Member CHD4 and Regulates the Glioblastoma Tumor-Initiating Cell State. Cell Rep.
565 2014;6(2):313-24.
566 39. Ying CY, Dominguez-Sola D, Fabi M, Lorenz IC, Hussein S, Bansal M, et al. MEF2B mutations lead
567 to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. Nat Immunol.
568 2013;14(10):1084-+.
569 40. Bisikirska B, Bansal M, Shen Y, Teruya-Feldstein J, Chaganti R, Califano A. Elucidation and
570 Pharmacological Targeting of Novel Molecular Drivers of Follicular Lymphoma Progression. Cancer Res.
571 2016;76(3):664-74.
572 41. Piovan E, Yu J, Tosello V, Herranz D, Ambesi-Impiombato A, Da Silva AC, et al. Direct reversal of
573 glucocorticoid resistance by AKT inhibition in acute lymphoblastic leukemia. Cancer Cell. 2013;24(6):766-
574 76.
575 42. Aytes A, Mitrofanova A, Lefebvre C, Alvarez MJ, Castillo-Martin M, Zheng T, et al. Cross-species
576 regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives
577 prostate cancer malignancy. Cancer Cell. 2014;25(5):638-51.
578 43. Mitrofanova A, Aytes A, Shen C, Abate-Shen C, Califano A. A systems biology approach to predict
579 drug response for human prostate cancer based on in vivo preclinical analyses of mouse models. Cell
580 Reports. 2015;12:1–12.
581 44. Talos F, Mitrofanova A, Bergren SK, Califano A, Shen MM. A computational systems approach
582 identifies synergistic specification genes that facilitate lineage conversion to prostate tissue. Nature
583 communications. 2017;8:14662.
584 45. Walsh LA, Alvarez MJ, Sabio EY, Reyngold M, Makarov V, Mukherjee S, et al. An Integrated Systems
585 Biology Approach Identifies TRIM25 as a Key Determinant of Breast Cancer Metastasis. Cell Rep.
586 2017;20(7):1623-40.

587 46. Rodriguez-Barrueco R, Yu J, Saucedo-Cuevas LP, Olivan M, Llobet-Navas D, Putcha P, et al.
588 Inhibition of the autocrine IL-6-JAK2-STAT3-calprotectin axis as targeted therapy for HR-/HER2+ breast
589 cancers. Genes Dev. 2015;29(15):1631-48.
590 47. Putcha P, Yu J, Rodriguez-Barrueco R, Saucedo-Cuevas L, Villagrasa P, Murga-Penas E, et al. HDAC6
591 activity is a non-oncogene addiction hub for inflammatory breast cancers. Breast Cancer Res.
592 2015;17(1):149.
593 48. Davoli A, Hocevar BA, Brown TL. Progression and treatment of HER2-positive breast cancer.
594 Cancer Chemother Pharmacol.65(4):611-23.
595 49. Guo L, Kozlosky CJ, Ericsson LH, Daniel TO, Cerretti DP, Johnson RS. Studies of ligand-induced site-
596 specific phosphorylation of epidermal growth factor receptor. Journal of the American Society for Mass
597 Spectrometry. 2003;14(9):1022-31.
598 50. Okamoto I. Epidermal growth factor receptor in relation to tumor development: EGFR-targeted
599 anticancer therapy. FEBS J.277(2):309-15.
600 51. Wang S, Placzek WJ, Stebbins JL, Mitra S, Noberini R, Koolpe M, et al. Novel targeted system to
601 deliver chemotherapeutic drugs to EphA2-expressing cancer cells. J Med Chem. 2012;55(5):2427-36.
602 52. Birtwistle MR, Hatakeyama M, Yumoto N, Ogunnaike BA, Hoek JB, Kholodenko BN. Ligand-
603 dependent responses of the ErbB signaling network: experimental and modeling analyses. Mol Syst Biol.
604 2007;3:144.
605 53. Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, et al. Global survey of phosphotyrosine
606 signaling identifies oncogenic kinases in lung cancer. Cell. 2007;131(6):1190-203.
607 54. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the
608 transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448(7153):561-6.

609 55. Laurent C, Do C, Gascoyne RD, Lamant L, Ysebaert L, Laurent G, et al. Anaplastic lymphoma kinase-
610 positive diffuse large B-cell lymphoma: a rare clinicopathologic entity with poor prognosis. J Clin Oncol.
611 2009;27(25):4211-6.
612 56. Janoueix-Lerosey I, Lequin D, Brugieres L, Ribeiro A, de Pontual L, Combaret V, et al. Somatic and
613 germline activating mutations of the ALK kinase receptor in neuroblastoma. Nature. 2008;455(7215):967-
614 70.
615 57. Sokai A, Enaka M, Sokai R, Mori S, Mori S, Gunji M, et al. Pulmonary Inflammatory Myofibroblastic
616 Tumor Harboring EML4-ALK Fusion Gene. Jpn J Clin Oncol. 2014;44(1):93-6.
617 58. Bender C, Henjes F, Frohlich H, Wiemann S, Korf U, Beissbarth T. Dynamic deterministic effects
618 propagation networks: learning signalling pathways from longitudinal protein array data. Bioinformatics.
619 2010;26(18):i596-i602.
620 59. Koivunen JP, Mermel C, Zejnullahu K, Murphy C, Lifshits E, Holmes AJ, et al. EML4-ALK fusion gene
621 and efficacy of an ALK kinase inhibitor in lung cancer. Clin Cancer Res. 2008;14(13):4275-83.
622 60. TCGA-Consortium. Comprehensive genomic characterization defines human glioblastoma genes
623 and core pathways. Nature. 2008;455(7216):1061-8.
624 61. Crystal AS, Shaw AT. Variants on a theme: a biomarker of crizotinib response in ALK-positive non-
625 small cell lung cancer? Clin Cancer Res. 2012;18(17):4479-81.
626 62. Shaw AT, Solomon B. Targeting anaplastic lymphoma kinase in lung cancer. Clin Cancer Res.
627 2011;17(8):2081-6.
628 63. . !!! INVALID CITATION !!! [24].
629 64. Voena C, Di Giacomo F, Panizza E, D'Amico L, Boccalatte FE, Pellegrino E, et al. The EGFR family
630 members sustain the neoplastic phenotype of ALK+ lung adenocarcinoma via EGR1. Oncogenesis.
631 2013;2:e43.

632 65. Chan BA, Hughes BG. Targeted therapy for non-small cell lung cancer: current standards and the
633 promise of the future. Transl Lung Cancer Res. 2015;4(1):36-54.
634 66. Landi L, Minuti G, D'Incecco A, Cappuzzo F. Targeting c-MET in the battle against advanced
635 nonsmall-cell lung cancer. Curr Opin Oncol. 2013;25(2):130-6.
636 67. Borisy AA, Elliott PJ, Hurst NW, Lee MS, Lehar J, Price ER, et al. Systematic discovery of
637 multicomponent therapeutics. Proc Natl Acad Sci U S A. 2003;100(13):7977-82.
638 68. Chou TC. Drug Combination Studies and Their Synergy Quantification Using the Chou-Talalay
639 Method. Cancer research. 2010;70(2):440-6.
640 69. Zhang G, Fang B, Liu RZ, Lin H, Kinose F, Bai Y, et al. Mass spectrometry mapping of epidermal
641 growth factor receptor phosphorylation related to oncogenic mutations and tyrosine kinase inhibitor
642 sensitivity. J Proteome Res. 2011;10(1):305-19.
643 70. Valencia K, Ormazabal C, Zandueta C, Luis-Ravelo D, Anton I, Pajares MJ, et al. Inhibition of
644 collagen receptor discoidin domain receptor-1 (DDR1) reduces cell survival, homing, and colonization in
645 lung cancer bone metastasis. Clin Cancer Res. 2012;18(4):969-80.
646 71. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using
647 gene-expression signatures to connect small molecules, genes, and disease. Science.
648 2006;313(5795):1929-35.
649 72. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel
650 in pulmonary adenocarcinoma. The New England journal of medicine. 2009;361(10):947-57.
651 73. Steuer R, Kurths J, Daub CO, Weise J, Selbig J. The mutual information: detecting and evaluating
652 dependencies between variables. Bioinformatics. 2002;18 Suppl 2:S231-40.
653 74. Gandhi J, Zhang JL, Xie Y, Soh J, Shigematsu H, Zhang W, et al. Alterations in Genes of the EGFR
654 Signaling Pathway and Their Relationship to EGFR Tyrosine Kinase Inhibitor Sensitivity in Lung Cancer Cell
655 Lines. PloS one. 2009;4(2).

Figure Legends

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

(A) Schematic diagram of a **TK→S** interaction. The non-phosphorylated kinase is inactive in terms of phosphorylating a substrate, while the active isoform successfully phosphorylates the substrate. **(B)** Schematic diagram showing the correlation between TK phosphorylation and that of its potential substrates. The first two rows in the heatmap show proteins representing candidate TK substrates **(C)** Illustration of the pARACNe framework including 6 steps. Step-1 depicts peptides collection from primary lung cancer tissue and cell lines for whole phosphrtyrosine proteomics quantification. Step-2 depicts inferences of TK→S interactions using Mutual Information by Step-3 Naïve-Bayes estimator and Step-4 of the iterative quantile discretization methods. Step-5 and 6 depict network pruning and bootstrapping to construct final network. **(C)**. Workflow of pARACNe from LC-MS/MS data normalization, IQD process, MI calculation, DPI process, bootstrapping to network consolidation.

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

(A) pARACNe-inferred densely inter-connected TK-TK network, with red nodes representing candidate TKs involved in auto-phosphorylation, where the phospho-state of a tyrosine is correlated with the phospho-state of a different tyrosine on the same TK protein. **(B)** pARACNe-inferred EGFR and **(C)** c-MET substrate overlap with SILAC-based and Database reported substrates, respectively.

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

679 **(A)**. Schema of Master Regulator analysis in lung cancer using pVIPER. Prioritized Master
680 Regulators **(B)** and Prioritized Master Regulator Pairs **(C)** as significantly activated (red circle) or
681 de-activated (blue) in different lung cancer cell lines (column). Red color represents an enrichment
682 of substrates hyper-phosphorylation by a Master Regulator or Master Regulator Pairs. Blue color
683 represents that of hypo-phosphorylation.

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

686 **(A)** Colony formation assay schema. **(B)** shows the image of long-term EGFR and c-MET double
687 inhibition effects in HCC78 cell line with different treatments. **(C)** shows long-term colony
688 formation data for 14 cell lines with different EGFR, BRAF and KRAS genomic mutation status.

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

690 **(A)**. MTT assay experimental schema. **(B)** MTT assay of HCC78 cell line shows synergistic effects
691 of Crizotinib and Erlotinib treatment. **(C)** shows short-term effects of EGFR and c-MET inhibitors'
692 combination index in 11 cell lines include 2 control cell lines (red).

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

694 EGFR and c-MET co-regulate in three scenarios **(A)** when their common substrates are
695 hyperphosphrylated, the patient responds to combination treatment well; **(B)** when most EGFR
696 substrates are hyper-phosphorylated, the patient responds to EGFR inhibitor; **(C)** when
697 substrates of both EGFR and c-MET are mostly hypophosphorylated, the patient does not
698 respond. **(D)** show the Master Regulator and Master Regulator Pairs regulating hyper/hypo-
699 phosphorylation of their network substrates in each primary samples.

700

701 # Methods and Data

702 ***Phosphoproteomic data***

703 The phosphoproteomic data, downloaded from Rikova K. et al. (53), representing the abundance
704 of phospho-tyrosine containing peptides, was obtained by tandem mass spectrometry analysis of
705 46 non-small cell lung cancer (NSCLC) cell lines, 151 NSCLC tumors, and 48 normal lung tissue
706 samples. Immunohistochemistry and a phospho-tyrosine specific antibody were used to screen
707 96 paraffin-embedded, formalin fixed tissue samples from NSCLC patients. About 30% of tumors
708 showed high-levels of phospho-tyrosine expression. Immunoblotting of 46 NSCLC cell lines with
709 a phospho-tyrosine specific antibody also showed heterogeneous reactivity especially in the
710 molecular weight range characteristic of receptor tyrosine kinases.

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

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

721 ***pARACNe algorithm***

722 ARACNe is originally designed for gene expression data, where expression of genes is usually
723 continuous and non-sparse. Quantitative data obtained from label-free LC-MS/MS by data-
724 dependent acquisition via spectral counting is discrete and very sparse, with many
725 phosphopeptides counts not observed for multiple peptides in each sample causing the current
726 version of ARACNe to be not suitable for this data, which thus required major modifications to
727 handle discrete data. To handle discrete abundances, we modified the mutual information
728 computation approach from a kernel density estimation based method to a Naïve based
729 estimation of mutual information, which is a histogram based technique(73). Briefly, consider a
730 collection of *N* simultaneous measurements of two genes X and Y. Data is partitioned into M
731 discrete bins a_i, and k_i denotes the number of measurements that lie within the bin a_i. The
732 probabilities $p(a_i)$ are then approximated by the corresponding relative frequencies of
733 occurrence $p(a_i) \rightarrow \frac{k_i}{N}$ and the mutual information $I(X,Y)$ between datasets X and Y is
734 expressed as

735 $$I(X,Y) = \log N + \frac{1}{N}\sum_{ij} k_{ij} \log \frac{k_{ij}}{k_i k_j}$$

736 Here k_{ij} denotes the number of measurements where X lies in a_i and Y in a_j and *N* total number
737 of samples.

738 Accuracy of mutual information is dependent on correct numbers of bins, M. To find the optimal
739 number of bins we applied ARACNe on the whole dataset by varying M from 1 to 20 and testing
740 the connections in predicted sub-network against the set of known connections (gold standard)
741 from databases (phosphoDB) (35).

742 In case of continuous data, partitioning can be achieved by dividing the range of data into M
743 equally spaced distance bins. Our data being discrete, equally spaced partitioning was not
744 possible. So, to overcome this problem, we used an iterative approach of partitioning (Fig 1C, S
745 Fig 1A). The basic idea is to divide the number of N data points into M with each bin containing
746 equal number of data point. If the data point(s) with the same value falls into consecutive bin(s),
747 we put those data point(s) into current bin and repartition the remaining points into remaining
748 number of bins. We keep iterating this till we finish either the bins or there are no more data points
749 to bin. For example, in Supplementary Fig. 1A, we initially partition N points into 4 bins. The data
750 points with 0 value does not fit into first bin and falls into subsequent bins, so we assign all data
751 points with 0 value into first bin and repartition the remaining points into 3 bins. We keep on doing
752 this process till we achieve 4 bins.

753 To evaluate initial performance and decide number of bins, we computed the network among all
754 tyrosine kinases and substrates, parsed the sub-network between 49 tyrosine kinases and 114
755 substrates which were present in PhosphoSite database and compared the results with the
756 connections present in database. From our analysis, we found that M=10 to be an optimal number
757 (S Fig.1B) which gave us precision of 14% and sensitivity of 24%. This precision is an
758 underestimate of real precision as in the gold standard many interactions are not present.

759 ***Master Regulator analysis***

760 To discover the master regulator in various cell lines, we interrogated the network obtained from
761 pARACNe using a novel algorithm, MARINa (27) (Master Regulator Inference Analysis), designed
762 to infer kinases that are key players in a particular cell line. Protein activity is a good indicator of
763 key kinases in a particular phenotype but often phosphorylated state of a protein is not sufficient
764 to determine its activity both due to measurement noise in phosphorylated state as well as LC-
765 MS/MS technique noise. To overcome this problem, MARINa infers kinase activity from the global

766 kinase substrate relationship and its biological relevance by overlapping this information in a
767 particular phenotype-specific program.

768 MARINa requires a network model and signature of the phenotype transition (i.e., all genes
769 ranked by their differential phosphorylation in two phenotypes). Here, the signature, S_{kin}, was
770 obtained by t-test analysis by comparing each cell line against all normal samples. First, we
771 associate each kinase with positive and negative activity targets, by computing the correlation
772 between each kinase and its predicted substrates and selecting only those substrates which had
773 a significant correlation (p-value ≤ 0.05, Bonferroni corrected). Second, for each kinase we
774 computed an *activity* by measuring the enrichment of the S_{kin} signature in predicted substrates
775 list, separately for both positive and negative correlated, (S_{kin}-*enrichment*). Enrichment was
776 computed by Gene Set Enrichment Analysis (GSEA). Since very small percentage of kinases are
777 found to have negative correlation, we did not use those interactions to evaluate enrichment.

778 ***Cell Culture***

779 All cell lines were grown in RPMI-1640 with 5% fetal bovine serum and incubated at 37°C in a
780 humidified atmosphere containing 5% CO_2. Cell lines were fingerprinted using the Perplex 1.2
781 system (Promega, Madison, WI). Fingerprints were compared to those generated at ATCC and/or
782 our internal database.

783 ***MTS Assay***

784 Short term MTS assays (CellTiter 96® AQ_{ueous} One Solution Cell Proliferation Assay, Promega,
785 Madison WI) were performed as previously described in (74). Specifically, each drug
786 concentration is octuplicated and the mean with standard deviation of all replicates were used to
787 generate a curve to allow calculation of the drug IC50 (Inhibitory Concentration of 50%) value.
788 The assays were repeated at least 3 times and the IC50s are the average of all replicates.

789

790 ***Colony formation assay***

791 Long term colony formation assays were performed in triplicate in 6-well plates. Cells were added
792 to media containing drug and incubated for 1-2 weeks such that control wells (no drug) contained
793 colonies of 50-70 cells each. At such time media was removed and all wells stained with a solution
794 containing 0.5% crystal violet and 6% glutaraldehyde for 1 hour. The plates were then rinsed,
795 dried, and colonies were manually counted.

796 ***SILAC Experiment***

797 EGFR SILAC experiment was performed in H3255 cell line by treating samples with Gefitinib. c-
798 MET SILAC experiment was performed in c-MET-driven gastric cancer cell line, MKN45, by using
799 c-MET inhibitor Su11274. For both genes, cells were treated with inhibitors for 3 and 24hr. For
800 control, cells were grown in same conditions but were not treated with the drug. For our
801 comparison we combined the peptides, which were differentially obtained between treated and
802 untreated samples, for 3 and 24 hr. More details about the experiment can be obtained from Guo
803 et al (34).

804 **Gold standard**: In PhosphoSite database, there were 282 connections between 49 tyrosine
805 kinases and 114 substrates.

806

807 # Supplemental Information

808 ***S1 Fig. Performance of pARACNe algorithm.***

809 **(A).** To select optimal bin number in pARACNe algorithm, precision and recall curves for various
810 number of bins were computed. Black curve is when no binning of data is done. When using 10
811 bins, the algorithm achieved the best performance.

812 ***S2 Fig. Validation of Master Regulators and Master Regulator combination.***

813 **(A)** Cell viability assay for top Master Regulators including ALK, EGFR, EPHA2, MET, PTK2, and

814 UBB in H2228 cell line. **(B)** BrdU assay for MET and PTK2 and its combination.

815 ***S3 Table. pARACNe inferred TK peptides – substrate peptides regpresented signaling***
816 ***network.***

817 ***S4 Table. pARACNe inferred TK protein – substrate protein regpresented signaling***
818 ***network.***

819 ***S4 Table. Colony formation assay and MTS assay results.***

820

Figure 1

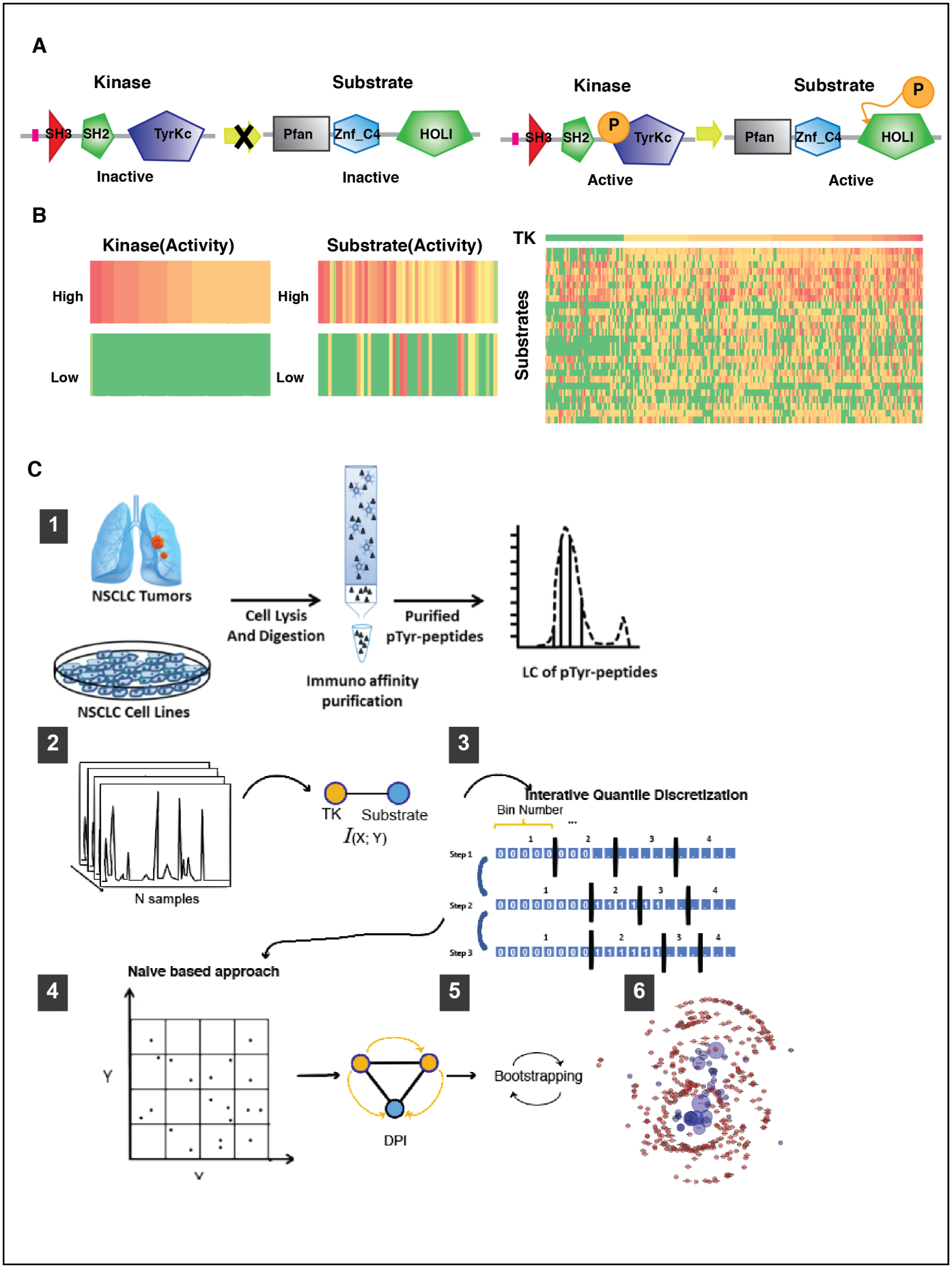

A
Kinase
Substrate
SH3
SH2
TyrKc
Pfan
Znf_C4
HOLI
Inactive
Inactive
Kinase
Substrate
P
Active
Active
B
Kinase(Activity)
Substrate(Activity)
High
Low
High
Low
TK
Substrates
C
1
NSCLC Tumors
NSCLC Cell Lines
Cell Lysis
And Digestion
Immuno affinity
purification
Purified
pTyr-peptides
LC of pTyr-peptides
2
N samples
TK
Substrate
I(X; Y)
3
Interative Quantile Discretization
Bin Number
Step 1
Step 2
Step 3
4
Naive based approach
Y
X
DPI
5
Bootstrapping
6

Figure 2

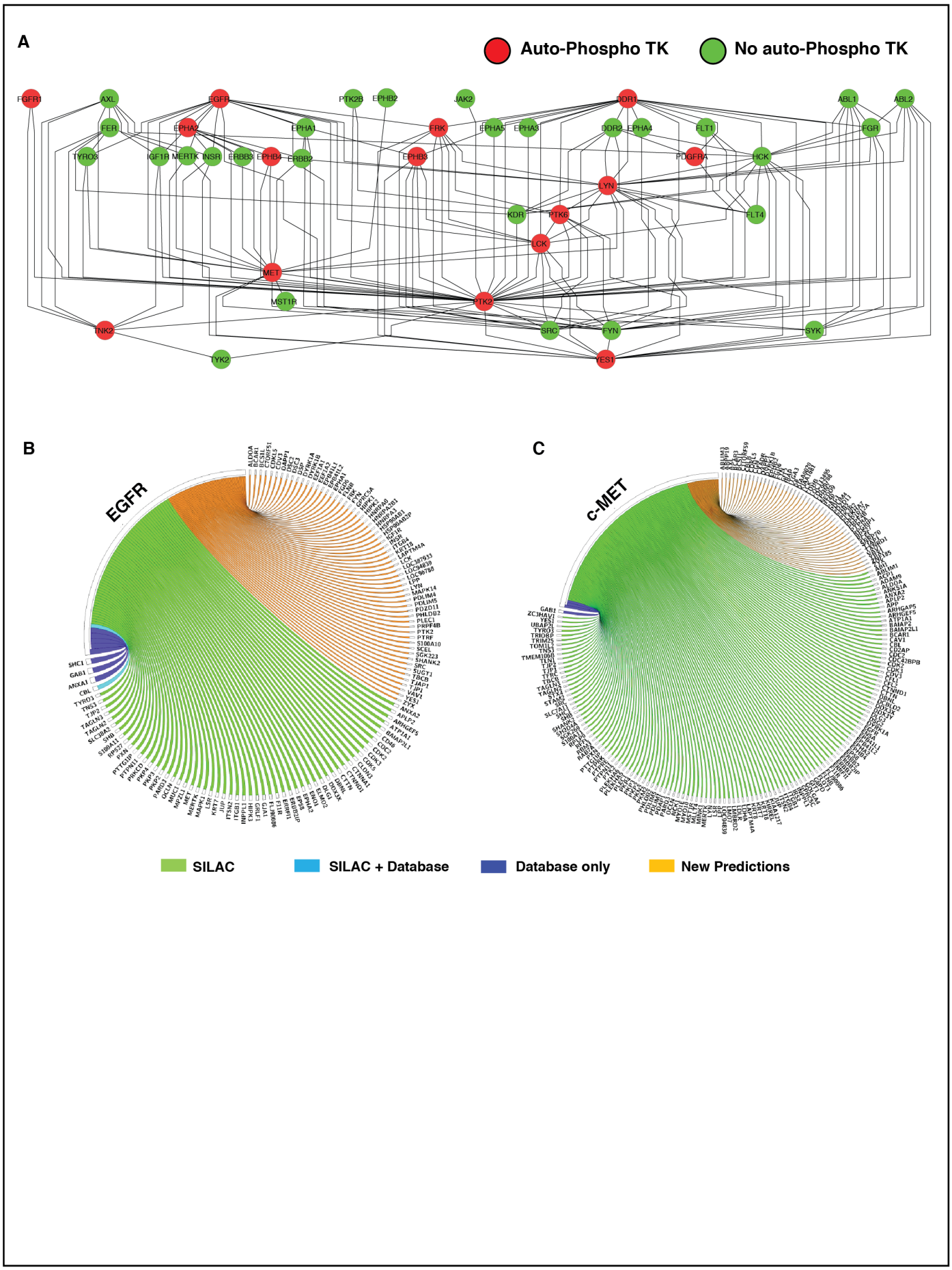

A
Auto-Phospho TK
No auto-Phospho TK
FGFR1
AXL
EGFR
PTK2B
EPHB2
JAK2
DDR1
ABL1
ABL2
FER
EPHA2
EPHA1
FRK
EPHA5
EPHA3
DDR2
EPHA4
FLT1
FGR
TYRO3
IGF1R
MERTK
INSR
ERBB3
EPHB4
ERBB2
EPHB3
PDGFRA
HCK
LYN
KDR
PTK6
FLT4
LCK
MET
MST1R
PTK2
TNK2
SRC
FYN
SYK
TYK2
YES1
B
EGFR
C
c-MET
SILAC
SILAC + Database
Database only
New Predictions

Figure 3

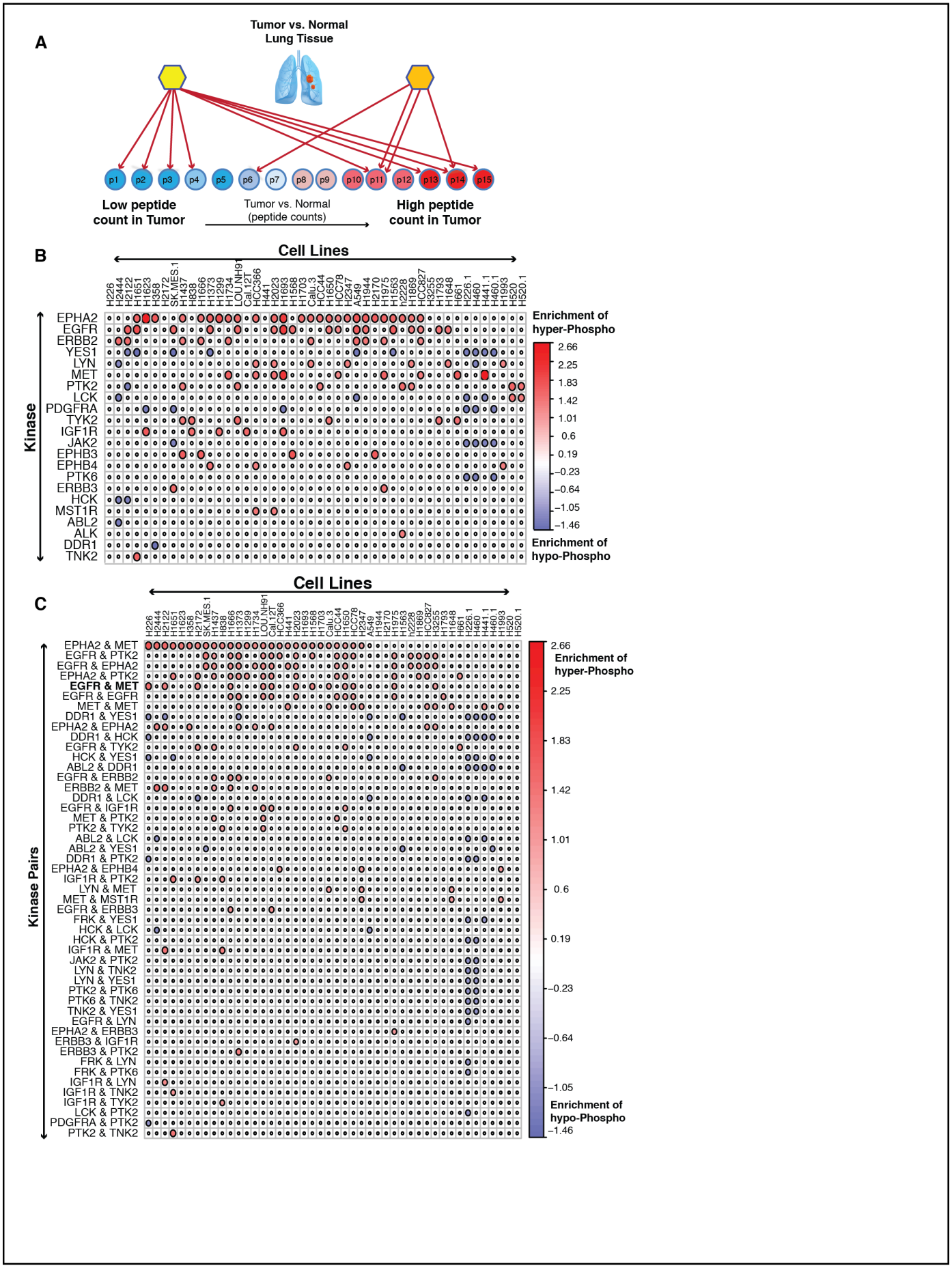

A
Tumor vs. Normal
Lung Tissue
p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12 p13 p14 p15
Low peptide count in Tumor
Tumor vs. Normal (peptide counts)
High peptide count in Tumor
B
Cell Lines
Kinase
EPHA2
EGFR
ERBB2
YES1
LYN
MET
PTK2
LCK
PDGFRA
TYK2
IGF1R
JAK2
EPHB3
EPHB4
PTK6
ERBB3
HCK
MST1R
ABL2
ALK
DDR1
TNK2
Enrichment of hyper-Phospho
2.66
2.25
1.83
1.42
1.01
0.6
0.19
−0.23
−0.64
−1.05
−1.46
Enrichment of hypo-Phospho
C
Cell Lines
Kinase Pairs
EPHA2 & MET
EGFR & PTK2
EGFR & EPHA2
EPHA2 & PTK2
EGFR & MET
EGFR & EGFR
MET & MET
DDR1 & YES1
EPHA2 & EPHA2
DDR1 & HCK
EGFR & TYK2
HCK & YES1
ABL2 & DDR1
EGFR & ERBB2
ERBB2 & MET
DDR1 & LCK
EGFR & IGF1R
MET & PTK2
PTK2 & TYK2
ABL2 & LCK
ABL2 & YES1
DDR1 & PTK2
EPHA2 & EPHB4
IGF1R & PTK2
LYN & MET
MET & MST1R
EGFR & ERBB3
FRK & YES1
HCK & LCK
HCK & PTK2
IGF1R & MET
JAK2 & PTK2
LYN & TNK2
LYN & YES1
PTK2 & PTK6
PTK6 & TNK2
TNK2 & YES1
EGFR & LYN
EPHA2 & ERBB3
ERBB3 & IGF1R
ERBB3 & PTK2
FRK & LYN
FRK & PTK6
IGF1R & LYN
IGF1R & TNK2
IGF1R & TYK2
LCK & PTK2
PDGFRA & PTK2
PTK2 & TNK2
Enrichment of hyper-Phospho
2.66
2.25
1.83
1.42
1.01
0.6
0.19
−0.23
−0.64
−1.05
−1.46
Enrichment of hypo-Phospho

Figure 4

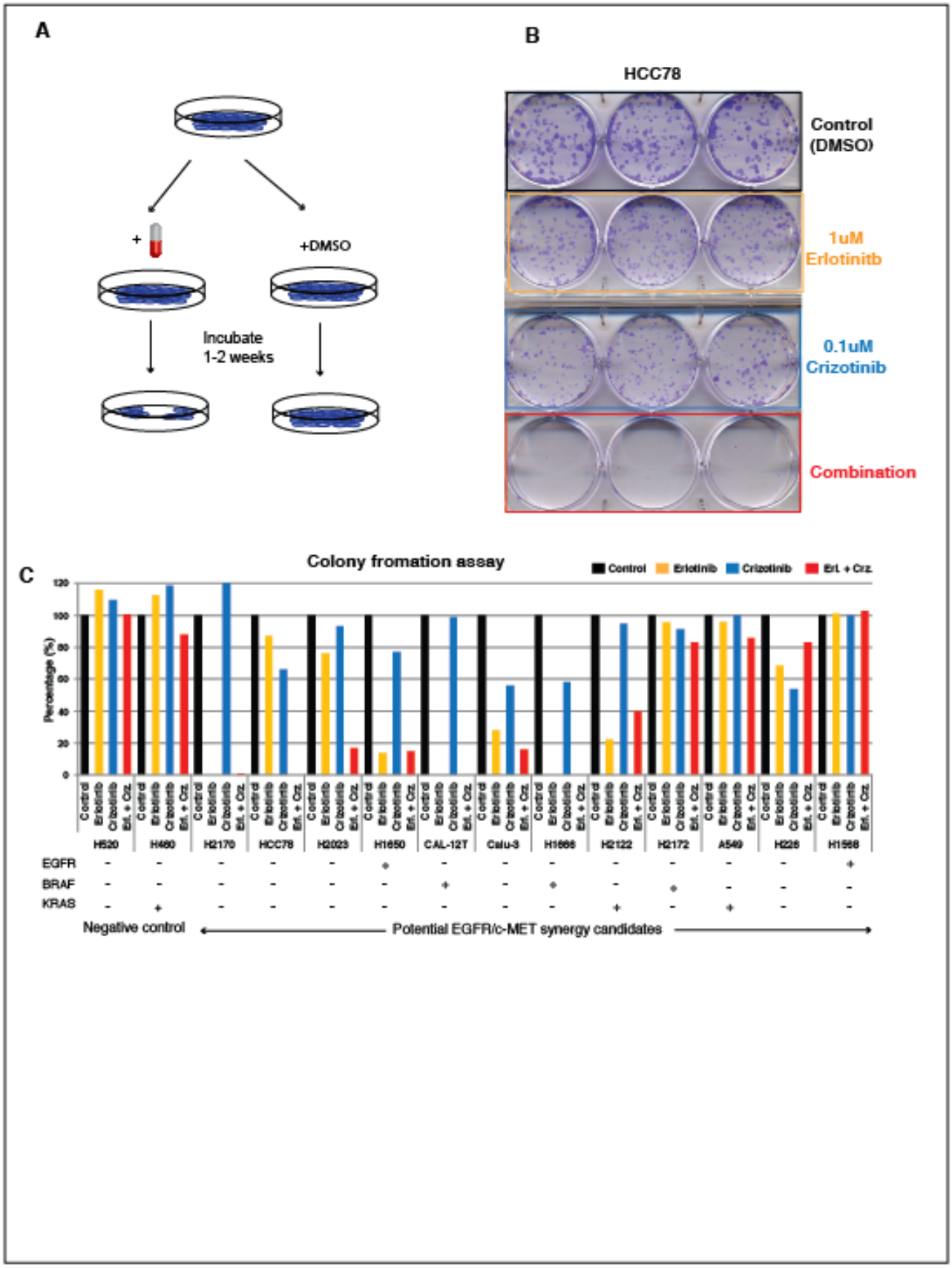

A
+
+DMSO
Incubate
1-2 weeks
B
HCC78
Control
(DMSO)
1uM
Erlotinitb
0.1uM
Crizotinib
Combination
C
Colony fromation assay
Control
Erlotinib
Crizotinib
Erl. + Crz.
Percentage (%)
120
100
80
60
40
20
0
H520
H460
H2170
HCC78
H2023
H1650
CAL-12T
Calu-3
H1666
H2122
H2172
A549
H226
H1568
EGFR
BRAF
KRAS
Negative control
Potential EGFR/c-MET synergy candidates

Figure 5

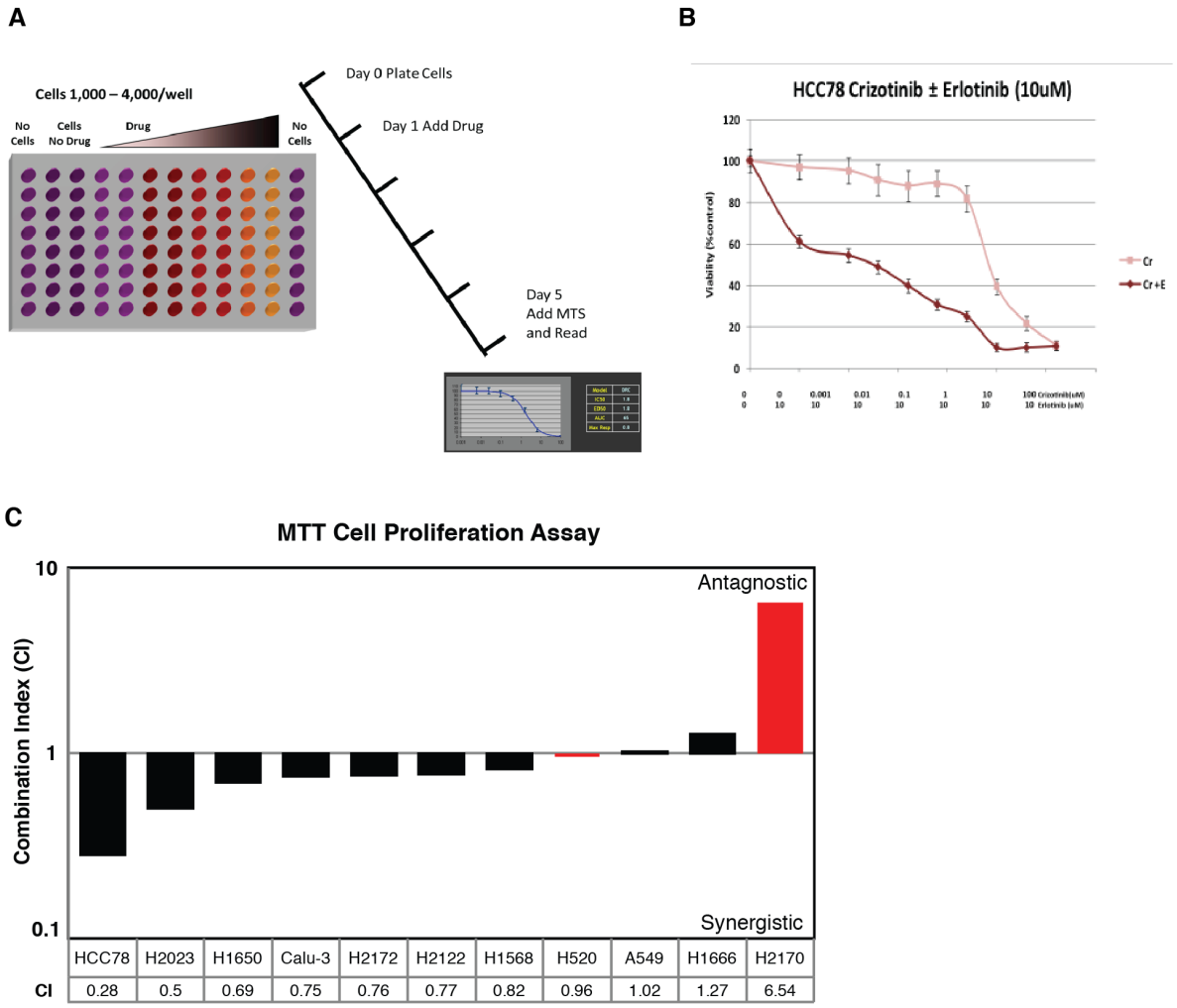
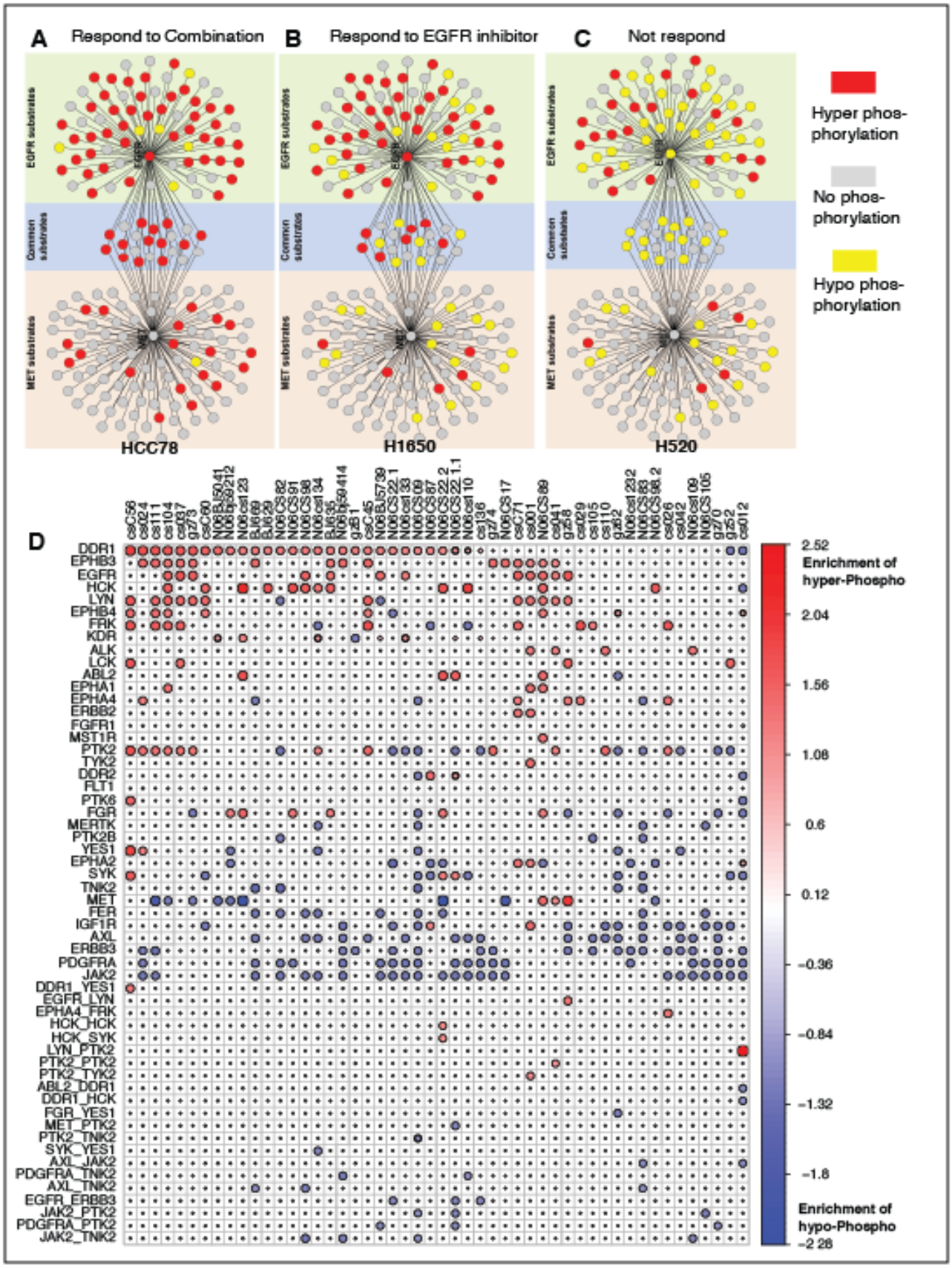

A
Cells 1,000 – 4,000/well
No Cells
Cells No Drug
Drug
No Cells
Day 0 Plate Cells
Day 1 Add Drug
Day 5 Add MTS and Read
B
HCC78 Crizotinib ± Erlotinib (10uM)
Viability (%control)
120
100
80
60
40
20
0
Cr
Cr+E
C
MTT Cell Proliferation Assay
Combination Index (CI)
10
1
0.1
Antagnostic
Synergistic
HCC78 H2023 H1650 Calu-3 H2172 H2122 H1568 H520 A549 H1666 H2170
CI 0.28 0.5 0.69 0.75 0.76 0.77 0.82 0.96 1.02 1.27 6.54

Figure 6


A
Respond to Combination
B
Respond to EGFR inhibitor
C
Not respond
EGFR substrates
Common substrates
MET substrates
HCC78
H1650
H520
Hyper phos-phorylation
No phos-phorylation
Hypo phos-phorylation
D
DDR1
EPHB3
EGFR
HCK
LYN
EPHB4
FRK
KDR
ALK
LCK
ABL2
EPHA1
EPHA4
ERBB2
FGFR1
MST1R
PTK2
TYK2
DDR2
FLT1
PTK6
FGR
MERTK
PTK2B
YES1
EPHA2
SYK
TNK2
MET
FER
IGF1R
AXL
ERBB3
PDGFRA
JAK2
DDR1_YES1
EGFR_LYN
EPHA4_FRK
HCK_HCK
HCK_SYK
LYN_PTK2
PTK2_PTK2
PTK2_TYK2
ABL2_DDR1
DDR1_HCK
FGR_YES1
MET_PTK2
PTK2_TNK2
SYK_YES1
AXL_JAK2
PDGFRA_TNK2
AXL_TNK2
EGFR_ERBB3
JAK2_PTK2
PDGFRA_PTK2
JAK2_TNK2
2.52
Enrichment of hyper-Phospho
2.04
1.56
1.08
0.6
0.12
−0.36
−0.84
−1.32
−1.8
Enrichment of hypo-Phospho
−2.28