Title
Dissecting AXL-mediated resistance to EGFR-targeted therapies in lung cancer

Author
Creixell, Marc

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Dissecting AXL-mediated resistance to EGFR-targeted therapies in lung cancer

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Bioengineering

by

Marc Creixell Santa Olalla

2023

ABSTRACT OF THE DISSERTATION

Dissecting AXL-mediated resistance to

EGFR-targeted therapies in lung cancer

by

Marc Creixell Santa Olalla

Doctor of Philosophy in Bioengineering

University of California, Los Angeles, 2023

Professor Aaron S Meyer, Chair

Cancer hijacks how cells sense, integrate, and respond to cues present in their environment. For instance, cancer cells can dysregulate their receptor tyrosine kinase (RTK) to drive proliferation and migration to distant tissues. RTK inhibitors, such as those targeting the epidermal growth factor receptor (EGFR), are therefore often effective but are invariably limited by the development of resistance. A well-appreciated means of resistance to RTK inhibition is so-called “bypass” resistance, wherein an RTK not targeted by therapy, such as AXL, activates alternative oncogenic pathways. In this work, we developed a paired experimental and computational strategy to comprehensively characterize the signaling changes that the RTK AXL regulates in EGFR-dependent lung cancer cells when driving resistance. To do so, we generated a panel of lung cancer cell lines with different AXL phosphosite mutations and then measured both the proteomic and phenotypic changes during bypass resistance. To model such data, we developed and applied an algorithm, Dual Data-Motif Clustering, that identified the most prominent signaling pathways that

AXL activates to mediate resistance to EGFR-targeted therapies, which we then experimentally validated. This work demonstrates a methodology for dissecting complex signaling networks and identifies several mechanisms by which AXL drives resistance-associated phenotypic changes.

The dissertation of Marc Creixell Santa Olalla is approved.

Thomas G Graeber

Stephanie Kristin Seidlits

Antoni Ribas

Aaron S Meyer, Committee Chair

University of California, Los Angeles

2023

This work is dedicated to my parents, Jordi and Cristina,

as well as my siblings Eva and Jordi.

TABLE OF CONTENTS

List of Figures

Chapter 1: Elucidating phosphorylation-mediated signaling networks of receptor tyrosine kinases through Y-to-F mutational studies

Chapter 2: Dual Data-Motif Clustering improves modeling of phosphoproteomic data

Chapter 3: Dissecting AXL-mediated bypass resistance and collateral phenotypes by phosphosite perturbations

Chapter 4: **Systems approaches to uncovering the contribution of environment-mediated drug**

Acknowledgments

For these past 4 years and a half I have been given the chance to develop myself as a scientist and as a person, and for that I would like to express my immense gratitude to my PhD advisor Dr. Meyer, my family, and my cousin, Pau Creixell.

When I was applying for graduate schools, my goal was pursuing a doctoral project within the field of cancer systems biology, specifically in kinase signaling transduction, in which I was able to learn and develop both computational and experimental skills. Therefore, I feel extremely grateful for the opportunity that Dr. Meyer provided me by welcoming me in his lab. I joined the lab as an experimental biologist with a little bit of practical, rather than theoretical, computational knowledge and the learning curve was very steep initially. However, under your mentorship, I have always had the resources and learning opportunities to overcome (the many) challenges that we have encountered along the way. One of the qualities I appreciate the most, was receiving the perfect balance between freedom to explore, make mistakes, improve, and repeat, and guidance when was really needed from you. There is not enough space in this thesis to enumerate all the things I have learned while working with you. Therefore, I will always feel grateful of your support, dedication, and mentorship when I think back about my years as a graduate student.

I would not be able to be here without the support of my family. Thank you for being supportive with my unconventional career plans. It is not easy to be so far from Barcelona every day, but I have always felt your support and care throughout this journey, so thank you.

Finally, I would like to deeply acknowledge the impact that my cousin, Pau Creixell, has had in my scientific career. Working with you during my master thesis allowed me to find a vocation in research and in this field. You taught me the importance of being true to oneself and tracing your own path, but also while having fun and celebrating every victory, regardless how small it is. Thank you for always offering advice and support when I need it.

Chapter 1 was adapted from a review article soon to be submitted from work that was led by Marc Creixell and Jacqueline Gerritsen:

Creixell, M., Gerritsen, J., White F., Meyer A (2022) 'Elucidating phosphorylation-mediated signaling networks in Receptor Tyrosine Kinases through Y-to-F mutational studies'. Manuscript in prep

Chapter 2 is a reproduction of the publication led by Marc Creixell and supervised by Aaron S Meyer:

Creixell, M and Meyer S. (2022) 'Dual Data and Motif Clustering improves the modeling and interpretation of phosphoproteomic data', *Cell Reports Methods,*

Chapter 3 was adapted from an article soon to be submitted from work that was led by Marc Creixell and supervised by Aaron S Meyer:

Creixell, M., Taylor, S., Gerritsen, J., Boixo C., Loui M., White F., Meyer A., 'Mapping of AXL-induced signaling changes driving bypass resistance to EGFR-targeted therapies in lung cancer'. Manuscript in prep.

Chapter 4 is a reproduction of the publication led by Marc Creixell and supervised by Aaron S Meyer:

Creixell, M. et al (2022) 'Systems approaches to uncovering the contribution of environment-mediated drug resistance', *Current Opinion in Solid State & Materials Science*

Marc Creixell was partially funded by the Johnson Comprehensive Cancer Research (JCCC) Fellowship Award

VITA

Research Experience

PhD Researcher with Aaron S Meyer, PhD | UCLA, Bioengineering department (2018-2023)

MSc Visiting MSc Researcher with Michael Yaffe | Koch Institute at MIT (2017—2018)

Industry Experience

Immuno-oncology Intern | Calico Life Sciences (June 2022)

Life Sciences Technology Fellow | Technology Development Group (TDG), UCLA (2022)

Analyst-Real World Evidence | IQVIA (2018)

Education

University of California, Los Angeles (UCLA)	PhD in Bioengineering	2018-Present
Imperial College London, UK	MSc in Novel Therapies	2016-2017
Technical University of Munich, Germany	ERASMUS program-Biomedicine	2015-2016
University of Barcelona, Spain	BSc in Biotechnology	2012-2016

Refereed Publications (In reverse chronological order)

Creixell, M., Taylor, S., Gerritsen, J., Boixo C., Loui M., White F., Meyer A., 'Mapping of AXL-induced signaling changes driving bypass resistance to EGFR-targeted therapies in lung cancer'. Manuscript in prep.

Creixell, M., Gerritsen, J., White F., Meyer A (2022) 'Elucidating phosphorylation-mediated signaling networks in Receptor Tyrosine Kinases through Y-to-F mutational studies'. Manuscript in prep

Creixell, M. et al (2022) 'Systems approaches to uncovering the contribution of environment-

mediated drug resistance', *Current Opinion in Solid State & Materials Science*, In Press.

Creixell, M and Meyer S. (2022) 'Dual Data and Motif Clustering improves the modeling and interpretation of phosphoproteomic data', *Cell Reports Methods,* In Press.

Creixell, P., Millán, L., **Creixell, M**. et al (2022) 'Mechanisms of Pro- and Anti-Oncogenic Signaling Downstream of Dual Tyrosine Protein Kinases'**.** Manuscript in pep.

Majumder, A., Hosseinian, S., Stroud, M., Adhikari E., Saller J., Smith M., Zhang G., Agarwal S., **Creixell, M** 'Integrated proteomics-based physical and functional mapping of AXL kinase signaling pathways and inhibitors define its role in cell migration', *Molecular Cancer Research,* In Press.

Magariños, A., Pedron, S., **Creixell, M**., et al (2018) 'The Feasibility of Encapsulated Embryonic Medullary Reticular Cells to Grow and Differentiate Into Neurons in Functionalized Gelatin-Based Hydrogels', Frontiers in Materials, In Press.

Creixell, P., Pandey, J., Palmeri, A., Bhattacharyya M., **Creixell M**., Ranganathan, R., Pincus, D., Yaffe M. 'Hierarchical Organization Endows the Kinase Domain with Regulatory Plasticity', Cell Systems, In Press.

Teaching & Mentorship Experience

Head Teaching Assistant (TA) | Bioengineering Laboratory, UCLA, 2022

Bioengineering B167L—Introduction to basic experimental techniques for undergraduate students

Mentorship | Aaron S Meyer Lab, UCLA

- Scott Taylor: Assistant Development Engineer
- Luka Karginov, NCI CSBC Undergraduate Summer Scholar, University of Illinois
- Carmen Boixo: Undergraduate Researcher
- Michelle Loui: PhD Researcher

Chapter 1:

Elucidating phosphorylation-mediated signaling networks of receptor tyrosine kinases through Y-to-F mutational studies

Creixell, M.[1†], Gerritsen, J.[2†], White F.[2], Meyer A[1]

[1]Department of Bioengineering, University of California Los Angeles

[2]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge MA, USA; Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge MA, USA; Center for Precision Cancer Medicine, Massachusetts Institute of Technology, Cambridge MA, USA

[†]These authors contributed equally to this work

Abstract

Receptor tyrosine kinases (RTKs) play a crucial role in cellular processes through complex phosphorylation-mediated signaling networks. Their aberrant activity in disease has resulted in many efforts to understand the underlying mechanisms that explain their contribution. However, RTK crosstalk and downstream signaling redundancy hinder the characterization of individual RTK members. One powerful approach that allows for mechanistic profiling of RTKs is the use of tyrosine-to-phenylalanine (Y-to-F) mutational studies, in which intracellular tyrosine residues are systematically mutated to non-phosphorylatable residues. Different Y-to-F mutations induce subtle but distinct RTK-specific signaling and phenotypic perturbations in a cellular environment. Both signaling and cellular responses driven by an RTK can then be functionally linked by coupling proteomic experiments with computational tools. Here, we review the mechanistic and translatable insights gained from these types of studies, as well as current approaches to collect and analyze this type of data. Finally, we offer a future perspective discussing the potential value that current advances in our tools and knowledge can add to the level of biological insight we can gain from this mechanistic research, especially as throughput and sensitivity of methods continues to improve.

2. Receptor Tyrosine Kinases in Disease

Receptor tyrosine kinases (RTKs) are key regulators of a plethora of cellular processes including cell proliferation, migration, differentiation, and metabolism[1,2]. Regulation occurs as RTKs propagate signals from the extracellular environment to an intracellular response. Given their critical cellular function, it is perhaps not surprising that aberrant RTK signaling has been associated with a variety of human diseases, including many cancers[3]. Large-scale omics studies have revealed various mechanisms through which abnormal RTK activation can occur, including

gene amplification, autocrine activation, chromosomal rearrangements, and gain-of-function or loss-of-function mutations[4–7]. As a result, many tyrosine kinase inhibitors (TKIs) and monoclonal antibodies have been developed and approved over the past few decades that target RTKs in various pathologies such as erlotinib and cetuximab in lung and head and neck cancer, panitumumab in colon cancer and trastuzumab and pertuzumab in breast cancer[8].

Despite these advancements, acquired resistance to these targeted therapies has often dampened success in the clinic. One well-known example of acquired resistance to EGFR-TKIs is the T790M mutation in non-small cell lung cancer (NSCLC). The threonine in position 790 regulates ATP and inhibitor specificity in the ATP binding pocket; mutation of this residue to methionine enhances ATP affinity and thereby confers resistance to ATP-competitive inhibitors[9,10]. Besides interfering with the drug-target interaction, other mechanisms of resistance can arise due to cellular adaptive responses. Cancer cells can modify their drug metabolism by changes in drug efflux and uptake or undergo bypass resistance, involving the hyperactivation of a receptor not targeted by therapy. For instance, MET and AXL have been shown to provide bypass resistance to EGFR-targeted therapies in lung carcinoma while HER3 drives resistance to HER2-inhibition in breast cancer[11–15]. Lack of treatment response or cancer growth after initial remission due to acquired resistance emphasizes the complexity of the signaling networks employed by RTKs and has fueled efforts to better understand receptor physiology.

3. RTK Structure and Signaling Network Models

RTKs (58 known in humans) can be grouped into 20 subfamilies based on their domain architecture. RTKs share a similar protein structure that includes an extracellular ligand binding domain, a single transmembrane helix, and an intracellular region that contains a juxtamembrane

regulatory region, a tyrosine kinase domain (TKD) and a carboxyl (C-) terminal tail[16]. In general, ligand-mediated receptor stimulation results in conformational changes and dimerization, allowing the TKD to assume an active conformation. Trans-autophosphorylation of tyrosine residues located on the TKD and C-terminal tail can then occur, initiating intracellular signaling[17]. Autophosphorylation leads to recruitment, phosphorylation, and activation of a wide variety of downstream adapter proteins that contain Src homology-2 (SH2) or phosphotyrosine-binding (PTB) domains. Phosphorylation of these adapter proteins and recognition by second messengers and downstream signaling molecules then trigger intracellular changes that aid in the receptor's biological signal relay and activate major signaling pathways including RAS/MAPK, PI3K/AKT and JAK/STAT[18,19](**Figure 1**).

Numerous computational methods have been generated over the past few decades to model RTK signaling response[20–22]. However, these models have generally ignored specific phosphorylation sites on the receptor due to the combinatorial complexity resulting from multiple phosphorylation sites; phosphorylation of each site results in a new species to be considered in the model. Moreover, interaction of a given scaffold or adaptor protein with different phosphorylation sites on the receptor can lead to a plethora of additional species, with associated interaction parameters for each phosphorylated residue and different adaptor proteins. Therefore, most computational models of RTK signaling consider the receptor as either non-phosphorylated and inactive or phosphorylated and active. Despite the challenges of modeling the full complexity of RTK phosphorylation and signaling, there have been extensive experimental efforts to identify specific phosphorylation sites on RTKs and define their respective interacting proteins. For instance, some experiments have used phosphorylated regions of an RTK to enrich for interacting proteins from cell lysate, while others have used protein microarrays to identify and quantify the interactions

between RTK phosphorylation sites and SH2 or PTB domains[23–25]. Although these studies, along with many others, have provided critical information of potential RTK phosphorylation-site specific interactions, most do not account for signaling dynamics, competition or recruitment of other kinases and phosphatases. Generating new insights describing RTK phosphorylation site-specific contributions will allow computational models to better explore their signaling dynamics. To access this additional level of information, phosphorylation site-specific studies need to be conducted in cells, within the full complexity of the intracellular environment.

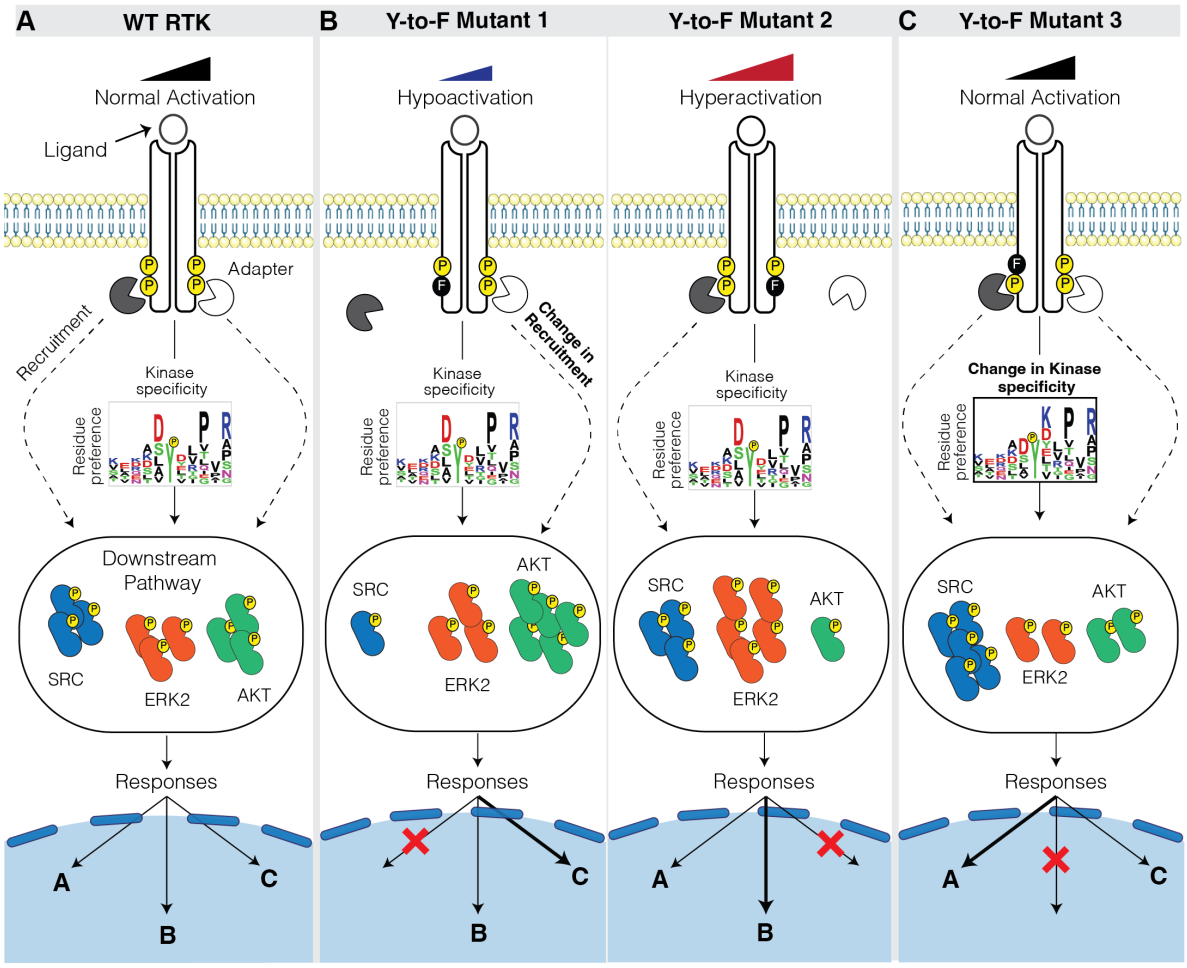


Figure 1. Signaling consequences lead to phenotypic responses in RTK Y-to-F mutant cells. (A) Upon RTK autophosphorylation, a wildtype (WT) two PTD-containing adapters binding different pTyr sites recruit further signaling molecules. According to its specificity, the RTK phosphorylates the recruited substrates to activate signaling

programs which will eventually lead to cell responses. RTKs harboring Y-to-F mutations can undergo specific signaling and phenotypic changes due to (B) either a decreased (left) or increased (red) RTK kinase activity and a change in adaptor recruitment to the unavailability of the mutated pTyr site or (C) a change in the RTK's kinase specificity towards its recruited substrates. Each of these signaling consequences lead to downstream signaling dysregulation that ultimately affect cell responses.

4. RTK Mutational Studies

Since the interaction between activated growth factor receptors and signal transduction molecules heavily relies on phosphorylation of the tyrosine residues in the TKD and C-terminal tail, many have used mutational strategies to identify specific residue functionalities in relation to downstream signaling responses.[26,27] In 'loss of function' studies, mutation of a given tyrosine to phenylalanine removes the hydroxyl group and eliminates the capacity for phosphorylation at that site while still maintaining many structural characteristics such as π-π and hydrophobic interactions[28,29]. In some studies, tyrosine sites have been replaced with alanine, however, alanine is structurally quite different from tyrosine and this substitution may confound signaling data due to changes in protein folding[30]. In many cases a 'gain of function' mutation is also tested, usually by substituting the tyrosine with a glutamic acid (e.g., Y-to-E) to mimic the negative charge of the phosphorylated residue. While this approach can work well on kinase activation loops, where the negative charge leads to altered protein structure and increased kinase activity, glutamic acid is a poor structural match for phosphotyrosine (pTyr)[31]. Therefore, Y-to-E mutations can lead to counterintuitive loss-of-function due to their poor binding affinity for SH2 and PTB domains. At the present time there are few options for a true gain-of-function experiment, as genetic incorporation of a constitutively phosphorylated tyrosine is highly challenging. Synthetic incorporation of constitutively phosphorylated tyrosines through intein chemistry may be feasible for relatively short proteins or for cases where the tyrosine site is at either end of the protein.

However, these synthetic proteins would then have to be inserted into cells to test their function. As an alternative to site-specific mutations, a more drastic option to assess loss of function is to delete the C-terminal tail, or sections of the C-terminal tail. While this approach can provide insights as to the function of sites in the deleted region, caution must be exerted given the risk of more extensive effects to protein folding and dimerization[32,33].

Given these options for site-specific interrogation, generating an isogenic panel of cell lines, with each line expressing an RTK with a selected Y-to-F mutation, represents a powerful approach to distinctly perturb RTK-driven signaling pathways while measuring the cell biological consequences. (**Figure 1B/C**). In studying the effects of selected Y-to-F mutations on RTK activity, protein interactions, and activation of downstream pathways, it is worth noting that selected Y-to-F mutations can also influence the overall activation state of the RTK. For instance, ligand induced RTK activation results in phosphorylation of several tyrosines within the TKD, including sites in the activation loop. On EGFR, the Y845 site lies in the kinase activation loop; phosphorylation of this site is thought to induce additional EGFR kinase activity. For FGFR1, Y677 was found to function as a stabilizer of the active conformation[34]. While the absolute amount of intracellular tyrosine phosphorylation is typically used as a proxy for RTK activation, not all tyrosine phosphorylation sites contribute to kinase activity, and sites may be phosphorylated at different stoichiometries[35]. On the C-terminal tail, EGFR Y1148 is phosphorylated 4-5 times more compared to other sites such as Y1173 or Y1068.

In addition to altering the RTK activation state, there is evidence of Y-to-F mutations shifting the competitive landscape of adaptor proteins binding to a given RTK[36,37]. Loss of a given phosphorylation site due to a Y-to-F mutation can strongly decrease the binding affinity for PTB

domain-containing proteins, leading to differences in RTK complex formation, thereby affecting the signaling response[23,38]. Interestingly, Cullati et al recently discovered that the kinase domain autophosphorylation of a particular threonine (T220) in the human non-receptor tyrosine kinase CK1δ significantly altered the conformation of the substrate binding cleft, affecting substrate specificity. Thus, while the autophosphorylation effects in RTK specificity remain uncharacterized, they cannot be ruled out[39] (**Figure 1C**).

5. Y-to-F interactome studies

Although there are several complicating factors when it comes to elucidation of signaling networks and individual tyrosine function, in many cases Y-to-F studies have successfully illuminated functional relevance of given RTK phosphorylation sites. One complicating factor in elucidating pTyr site function is redundancy, such as through the ability of adaptor molecules to bind multiple phosphosites[40]. With EGFR, for example, this built-in redundancy helps to maintain functional robustness in the event of low EGF stimulus or partial system failure[41]. A study by Gill *et al* found that EGFR containing just one functional tyrosine on the C-terminal tail was still signaling-competent[41]. **Figure 2** summarizes the efforts that involved mutation of selected C-terminal phosphorylation sites that revealed loss of a given adapter binding or confirmed the multiplicity of a binding interaction[27,42–53]. These built-in compensatory mechanisms make it more challenging to ascribe functions to individual residues, as RTKs can enlist other sites to adapt to loss of a given tyrosine phosphorylation site. Even with this redundancy, Schlesssinger determined a hierarchy of binding sites for Grb2 and Shc on EGFR, by using Y-to-F mutant cell lines and co-immunoprecipitation experiments. Although at least 5 tyrosines have been reported to bind Grb2 and Shc1 on EGFR, their data demonstrated a preference for Grb2 binding to Y1068 and Y1173,

and a Shc1 binding preference for Y1173[54]. In similar studies, preferred binding sites for PLCγ1 and Abl were EGFR-Y992 and EGFR-Y1173, respectively[41,55,56].

Beyond EGFR, Y-to-F mutant studies have also been used to determine site function and phosphorylation stoichiometry on multiple other RTKs. For instance, in IGF1R, Y-to-F mutations revealed Y1150 and Y1151 to be responsible for ~80% of receptor autophosphorylation[57]. On the other hand, IGFR1 Y1136F mutation only minimally affected its overall kinase activity yet was found to substantially reduce the phosphorylation of various substrates[58], providing a curious disconnect between site stoichiometry and importance for downstream signaling activity. In HER3, Y-to-F mutation of Y1325 revealed this site as the major binding site for Shc1 and the main site responsible for inducing heregulin (HRG)-dependent ERK activation[59].

Y-to-F mutations have also revealed intriguing crosstalk between RTKs. For instance, Y857 on PDGFRβ was identified as a critical site in transactivation of EGFR[60]. Interestingly, another study found this same tyrosine to affect cell proliferation, but not migration, while EGFR is known to affect both phenotypes[46,61]. Y-to-F studies have demonstrated cooperativity between residues on HER3 as well as requirement co-activation of HER2 and HER3 to induce PI3K activity[62]. In terms of transphosphorylation, Bae *et al.* showed that asymmetric receptor contact is required for autophosphorylation of FGFR1 and transphosphorylation of Y583[63].

6. Y-to-F studies in disease models

Although these examples clearly highlight the ability of Y-to-F RTK mutants to extract insight about given phosphorylation sites' function, it is worth noting that many of these experiments were performed in a single cell line. Depending on the cell context, the concentration of different adapter

proteins may vary and may represent limiting factors in one cell line relative to another, thus yielding different functional outcomes. Even for the same cancer types, the signaling pathways and targets identified across studies are largely context-dependent and thus cross-referencing insights generated in different systems can often be challenging and misleading. For example, AXL has been shown to promote cell proliferation, epithelial-to-mesenchymal transition, metastasis, macropinocytosis, metabolic oxidation, DNA damage response (DDR), and immunosuppression in various types of cancer[64–67]. It is easy to see how altering specific phosphorylation sites on the receptor may affect different pathways depending on the biological context of the RTK.

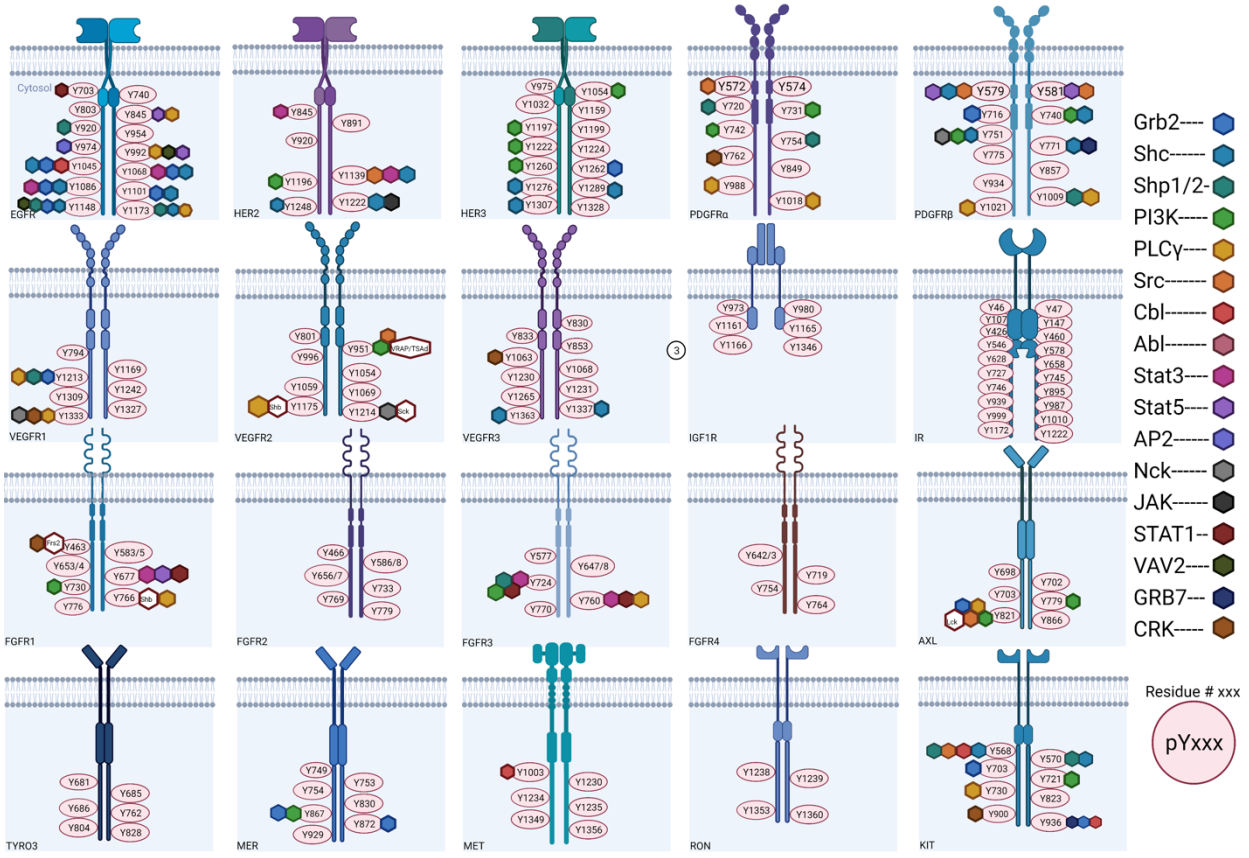


Figure 2: Known binding interactions between tyrosine residues of a subset of RTKs and secondary adaptor molecules. Numbers indicating residue location on protein. Proteins shown: EGFR[96,97], HER2[98], HER3[99],

PDGFRa/b[100], VEGFR1-3[101,102], IGF1R[103], IR[104], FGFR1-4[105], AXL/TYRO3/MER[106,107], MET[108], RON[109], KIT[100]. Several are still poorly characterized in terms of binding partners and function.

One major application of Y-to-F mechanistic studies has been the identification of novel therapeutic targets with increased selectivity against oncogenic downstream signaling of an RTK. While RTKs are druggable themselves, the clinical efficacy of these therapies is limited, in part, by expression of the targeted receptor in healthy tissues leading to off-target toxic effects and the ability of cancer cells to overcome anti-RTK blockade by the overexpression of an RTK not targeted by therapy. Therefore, a major undertaking in the field has been the identification of specific downstream signaling pathways driving disease progression. For instance, motivated by a strong correlation between HER2 activation and constitutively active Stat3a in a variety of human tumors, Ren *et al* found HER2 Y1139 to be the main residue responsible for Stat3a activation through JAK2 and SRC-dependent mechanisms, and provided multiple novel targets for HER2-overexpressing tumors[48]. In glioblastoma cell lines expressing EGFR-VIII with Y-to-F mutations on different C-terminal phosphorylation sites, signaling network analysis combined with phenotypic data identified a set of phosphoproteins as potential candidates for future drug development[50].

These studies have also helped to validate model systems to study certain diseases. For instance, PDGFRα was identified to be important in development of cardiac and brain cells, providing a model to study aberrant neural crest cell (NCC) development[49]. Finally, a Y-to-F study in AXL determined that Y821 was found to mediate resistance to cetuximab through c-ABL in a head and neck squamous carcinoma (HNSCC) model, proposing ABL inhibition as a potential therapeutic strategy to resensitize tumor cells to cetuximab treatment[53].

7. Omics-based Y-to-F studies

So far, most Y-to-F studies have evaluated RTK-protein interactions via pTyr on the TKD or measured phenotypic variation in response to single Y-to-F mutations. However, to fully understand the mechanisms underlying RTK function, it is critical to be able to integrate signaling and phenotypic responses measured in the same biological systems. To integrate these data, a systems biology approach should be used that considers the entire network, ideally in a system that recapitulates the full complexity and competition of physiological cellular signaling networks. Phosphoproteomics-based approaches enable the acquisition of quantitative, network-level, dynamic signaling data, making it well-suited to interrogate the network-wide effects of site-specific tyrosine functions.

As reviewed by Baselga *et al*, drug resistance should be viewed and investigated as a multifaceted problem driven by the development of simultaneous "collateral" malignant phenotypes that coordinately promote tumor growth and metastasis in the presence of therapy[68]. Applying quantitative phosphoproteomics and phenotypic endpoint assays to a panel of isogenic mutant cell lines (e.g., with Y-to-F mutations in the TKD or C-terminal of a given RTK) can link specific RTK phosphorylation sites with downstream components that collectively drive a complex multifactorial response, e.g., cell proliferation, cell migration, or drug resistance[68] (**Figure 3A)**. These complex phenotypic responses require a systems biology approach that utilizes a variety of computational tools and multivariate analysis approaches to separate the underlying signaling drivers.

However, despite many advancements in phosphoproteomics over the past two decades, challenges inherent to phosphoproteomic data hinder the identification of relevant proteomic

alterations; namely, this technology measures substrates rather than kinases, provides incomplete and stochastic coverage of the phosphoproteome across experiments, and has high content but low sample throughput. The next section of this review will focus on how Y-to-F mutational model systems and omics-tools can be combined to measure, model, and integrate the signaling and phenotypic responses of RTKs. We describe ways in which the resulting model predictions of downstream drivers can be validated and provide a forward-looking perspective of the utility of this system.

8. Computational tools to analyze and interpret phosphoproteomic RTK-driven signaling networks

Quantitative phosphoproteomics enables the acquisition of large-scale data describing quantitative changes in thousands of protein phosphorylation sites across different biological samples. Such data allows one to extract signaling network information, including putative activated kinases and phosphatases. Although meaningful, the signaling consequences of RTK Y-to-F mutations in a panel of isogenic cell lines can be subtle and hard to discern, underscoring the need for computational methods capable of elucidating signaling behaviors specific to and across mutant cell lines. In the last decade there has been substantial advancements in the development of computational tools tailored to disentangle and reconstruct signaling networks from quantitative phosphoproteomic data. Here, we categorize these computational tools into two main groups: "Single-Sample" and "Multi-Sample" methods (**Figure 3B**).

Single-Sample methods are those that allow one to profile the phosphoproteome of an individual sample or RTK Y-to-F mutant cell line by comparing it to a control (e.g., wild-type RTK cell line). Kinome enrichment tools such as kinase-substrate enrichment analysis (KSEA) and Integrative

Inferred Kinase Activity (INKA) aim to identify upstream kinases using the quantitative phosphoproteomic data; KSEA assembles phosphoproteomic data into groups of known or inferred kinase substrates, then averages the signals of the group to identify putative activated upstream kinases[69], and INKA ranks kinases based on their inferred activity by integrating kinases' overall and activation loop phosphorylation with the phosphorylation abundance of its known substrates[70]. These methods require making assumptions about kinase-substrate relationships for which there is experimental evidence in the literature or are predicted by kinase prediction algorithms such as Scansite or KinomeXplorer. Scansite makes predictions using kinase specificity profiles generated through position-specific peptide library (PSPL, further described below) experiments while KinomeXplorer uses sequence motif and protein-protein network information[71,72]. Furthermore, ranked gene identifiers based on the signaling data of an RTK Y-to-F mutant sample can be used to run Gene Set Enrichment Analysis (GSEA) or STRING analysis to identify and visualize overrepresented genes and associated biological processes, as well as protein interaction network maps, respectively. Thus, Single-Sample methods can be used to find enriched signaling components and their associated biological processes in specific Y-to-F RTK perturbations. These tools are very useful resources to ascertain the mechanistic understanding of a particular Y-to-F mutation, especially if the generated signaling insights can be associated with prior knowledge such as a mutant-specific phenotype or dysregulated signaling pathway or kinase. As introduced earlier, McDaniel et al found that AXL pTyr 821 conferred cetuximab resistance via interaction with ABL[73]. These methods could help identify other kinases and signaling components acting downstream of AXL to drive resistance. For instance, KSEA or INKA could identify additional putative hyper- or hypo-active kinases in the Y-to-F mutant cell lines compared with wild-type. This inference, in combination with a visual representation of phospho-modulated protein-protein interactions such as STRING, can help reconstruct the signaling network changes

induced by the point mutation, beyond ABL1. Nevertheless, as argued above, pTyr-mediated RTK function is a coordinated, multifactorial process that occurs via the simultaneous regulation of multiple intracellular tyrosines and downstream components (**Figures 1 and 3A**). While "Single-Sample" methods help explore the role and associated downstream effects of a particular pTyr in isolation, they ignore the concomitant regulation of other pTyr sites. Hence, applying Multi-Sample analysis, to model the coordinated signaling and phenotypic consequences of a panel of Y-to-F mutant cell lines, should provide a better representation of the various regulatory mechanisms of an RTK.

Multi-Sample methods to analyze phosphoproteomic data, such as clustering or dimensionality reduction, are extremely useful to enhance our interpretation of a dramatically underdetermined system—i.e., thousands of peptides across a relatively small set of samples. As phosphoproteomic measurements contain the phosphorylation abundance of kinase substrates, the peptides displaying similar signaling behavior across samples clustered in the same group can be viewed as a single unit that acts downstream of the same set of upstream kinases. Therefore, clustering methods ideally organize phosphosites into groups of peptides that represent common signaling changes. OpenEnsembles is a clustering algorithm that generates and incorporates multiple unique clustering solutions utilizing different combinations of data transformations (e.g., untransformed, z-score, mean-centering), clustering algorithms (e.g., k-means, hierarchical clustering, GMM), and parameters (number of clusters, distance metrics, linkage methods, etc.). Phosphosites are then grouped based on how often they cluster together[74]. Via an iterative process of modifying the algorithm's parameters, OpenEnsembles can find consistently dysregulated signaling pathways across Y-to-F mutants. Recently, the Meyer lab constructed an expectation-maximization (EM) algorithm, Dual Data and Motif Clustering (DDMC) that can be viewed as a Multi-Sample form

of KSEA and INKA. DDMC generates clusters displaying similar phosphorylation patterns and sequence motif features, leveraging the prior knowledge that protein kinases possess extraordinary specificity toward the phosphorylation site. Incorporating the sequence information into the clustering criterion allows one to make predictions about the upstream kinases regulating clusters by comparing the clusters' sequence features with previously reported kinase specificity profiling data[75]. Therefore, DDMC can generate protein clusters that are phospho-modulated by the RTK under study, which in turn might be regulated by the predicted upstream kinases. This allows the reconstruction of RTK-specific signaling network comprised by the RTK, downstream kinase drivers, and clusters of phosphopeptides[74]. Thus, Multi-Sample methods such as the described clustering algorithms or other common dimensionality reduction techniques such as principal component analysis (PCA) or non-negative matrix factorization (NMF) can be applied to find behavioral patterns across RTK mutant samples which allow the systematic identification of RTK-driven signaling pathways and processes.

9. Establishing associations between cell phenotypes and RTK downstream signaling

RTKs are involved in a myriad of cellular processes, and data-driven multivariate regression modeling of a panel of Y-to-F mutants can help map the downstream signaling pathways that correspond to different phenotypes. In a regression model, the phosphoproteomic data—matrix X—is used to explain the cellular responses—matrix Y, and the observations in both matrices consist of the RTK mutant cell lines treated under the same conditions—such as in the presence of the cognate ligand for RTK activation—and the variables are phosphopeptides and phenotypes, respectively (**Figure 3A/C**). Measuring the signaling and phenotypic consequences under an identical cell culture environment ensures that the patterns identified by the model are exclusively driven by the chosen inputs (e.g., RTK activation). Practically any cell phenotype data of interest

such as cell proliferation, apoptosis, migration, or genetic signature scores can be incorporated in the Y matrix if the cell culture conditions match the ones utilized to generate the signaling data. Importantly, the signaling and phenotypic consequences that Y-to-F mutations might induce could be influenced by an impairment of RTK regulation at different levels. Thus, trafficking and degradation rates, gene and protein expression, as well as dynamic measurements of downstream signaling of the different RTKs are crucial factors that should be measured to obtain a holistic view of how these mutations induce the observed responses.

Y-to-F mutational studies can provide a unique perspective of the extent to which RTK point mutations affect distinct phenotypes through specific downstream signaling components. For instance, RTK mutant 1 might largely activate ERK to promote proliferation whereas mutant 2 mainly induces JNK activity and apoptosis (**Figure 3A**). However, as mentioned above, MS-based phosphoproteomic data sets tend to be hugely underdetermined given its high content but low sample throughput. One approach has been to use partial least squares (PLSR) models to robustly handle predictions in the presence of high-dimensional and correlated data[76]. However, as the number of phosphosites—or variables—increases, the system becomes increasingly underdetermined, and the prediction power diminishes. In addition, while these models can generally be predictive with such data, they are not easily interpretable. The lack of interpretability also arises due to the high number of variables. Like PCA, the output of PLSR models are scores and loadings. However, PLSR generates loadings for both the predictor (phosphosites) and response (phenotype) variables. In this scenario, the resulting loadings plot includes the few phenotype data points surrounded by thousands of phosphosites. Instead, first clustering phosphosites based on biologically meaningful features organizes such them into groups that constitute signaling nodes. The cluster averages are then used to fit PLSR to enhance the predictive

performance of the regression model while providing highly interpretable results wherein clusters distinctly correlate with cell phenotypes (**Figure 3C**). **Figure 3D** shows putative results of this strategy wherein cluster 2 tightly correlated with phenotype 2 whereas cluster 4 is associated with phenotype 3 in the loadings plot, whereas the scores display the correlation between the different RTK Y-to-F mutations and in comparison with the signaling and phenotypic effects across principal components (PC) 1 and 2. In this theoretical case, PC1 separates the RTK kinase activity with positive scores and loadings correlating with an increased RTK activation as well as cluster 4 and phenotype 3. Moreover, as introduced earlier, KSEA, INKA, DDMC, or KinomeXplorer can interrogate different features of the signaling clusters to predict upstream kinases regulating them[77].

Hence, the combination of phosphoproteomic with phenotypic measurements of isogenic RTK Y-to-F mutant cell lines represents a unique opportunity to associate key signaling events with biological consequences through Multi-Sample analysis. The predictions resulting from these models facilitate the generation of hypothesis about key signaling pathway components driving cell responses which in turn informs the design of downstream of validation experiments.

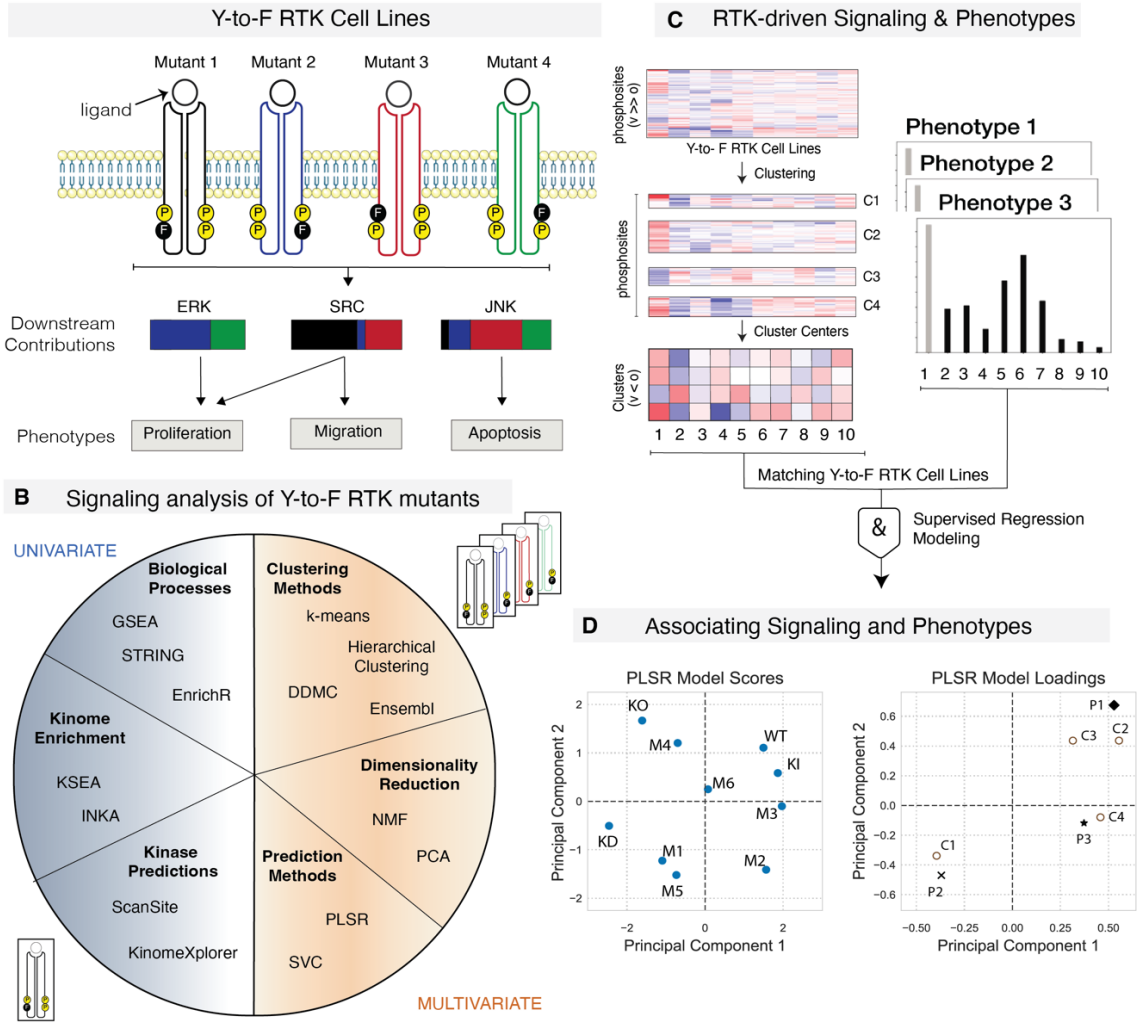


Figure 3: A combined experimental and computational strategy to associate RTK-specific signaling changes to cell phenotypes. (A) Collecting signaling and phenotypic data from a panel of Y-to-F mutants. (B) Categories of Single-Sample or Multi-Sample computational methods for signaling analysis of either individual or multiple Y-to-F perturbations, respectively. (C-D) A computational strategy to map signaling responses to Y-to-F mutations and their functional association with cell phenotypes. (C) First, phosphopeptides that behave similarly across Y-to-F mutant cell lines are clustered together to assemble peptides into a group that presumably acts downstream of the same upstream kinases. The resulting clusters are therefore summaries of the RTK-driven signaling changes that can be used to identify signaling nodes affecting cell phenotypes. (D) For instance, the generated cluster centers can be used to simultaneously predict several cell phenotypes through PLSR which allows the association between specific Y-to-F mutations, clusters, and phenotypes. V = variables, O = observations, C = Cluster, P = Phenotype.

10. Validation experiments to demonstrate model predictions

While a regression model can implicate measured RTK signaling changes with phenotypic responses, the causative significance of specific pathways must be experimentally validated. The predicted targets can be effectively validated through several experimental methods. Predicted changes in protein expression or protein post-translational modification can be validated by typical biochemical assays such as western blot; while predicted alteration in protein-protein interactions can be validated by proximity ligation assay (PLA), immunoprecipitation (IP), or 2-hybrid screening. Although these assays typically measure a single interaction and are not multiplexed, they provide evidence of direct protein-protein interactions. To validate RTK-protein or downstream protein-protein interactions with high throughput, BioID interaction mapping facilitates a global view of the interactome of any protein, including RTKs. This system utilizes a proximity-dependent labeling strategy wherein a bait protein is fused to a mutant form of biotin ligase (BirA*) that biotinylates interacting proteins within a 10-nm radius of the protein of interest[77]. Such data can serve as indirect evidence of adapter molecules, kinases, phosphatases and other signaling molecules distinctly interacting with RTK Y-to-F mutants which can then be subsequently associated with downstream signaling consequences. Many studies have used BioID-based strategies to elucidate the interactome of RTKs. For instance, Bareja et al used BioID to identify nexin6 as a novel interactor of IGFR that suppresses the recruitment of the adaptors IRS1 and SHC1 and ERK1/2 signaling[78]. Similarly, another study performed BioID on ALK and identified PEAK1 and SHP2 as interactors in neuroblastoma cells[79]. The concomitant inhibition of ALK and SHP2 lead to a synergistic drug response which established the functional role of the discovered interactions[79]. Therefore, by mapping the RTK interactome, multiplex BioID provides the opportunity of identifying changes in the recruitment of signaling molecules across Y-to-F

RTK mutants, thereby providing orthogonal evidence of the mechanistic insights generated with the phosphoproteomic analysis described above (**Figure 3**).

As mentioned above, kinase specificity profiling via PSPL experiments identifies the most favored or disfavored amino acids surrounding the target motif of a kinase, thereby providing the "optimal" substrate motif of a particular kinase. For instance, ABL1 has a very strong preference for proline at +3 and isoleucine at -1—relative to the phosphorylation site, whereas CK2 clearly favors acidic residues in the C-terminal part of the motif[80–83]. One can interrogate the phosphopeptides present in an RTK-specific signaling cluster of interest (see section above) and generate a position-specific scoring matrix (PSSM) that summarizes the frequencies of observed residues across phosphopeptides positions[84]. If the phosphoproteomic analysis causally identifies a particular kinase (e.g., cluster 4 and phenotype 3 in **Figure 3D**)**,** the PSSM of cluster 3 can be compared with the specificity profile of the kinase to explore its association. Thus, PSPL can help identify downstream kinases linking the activated RTK with groups of downstream phospho-modulated proteins. In this technique, a kinase of interest is individually incubated with each of 180 different peptide libraries in which each library contains a central phosphoacceptor residue (S/T or Y), a second fixed amino acid located at any of the 10-amino acid peptide spanning positions -5 to +4 relative to the phosphorylation site, and a degenerate mixture containing all natural amino acids at all other positions. The kinase and peptide libraries are incubated in the presence of radioactive ATP, which allows the quantification of phosphorylation abundance per residue and position and the identification of the kinase's ''optimal'' substrate motif. The accumulation of PSPL kinase profiles facilitates the association of putative upstream kinases regulating specific substrates or groups of phosphosites displaying similar kinase motifs[85–88]. However, while PSPL identifies the most favored amino acids throughout all peptide positions, it does not test for the interdependence

between residues. This means that the combination of the most preferred residues across peptide positions could still result in a highly disfavored substrate motif for the kinase due to the untested docking between the profiled kinase and the identified "optimal motif". This limitation could compromise kinase-substrate or kinase-cluster associations. Shah *et al* devised a kinase specificity screening platform combining bacterial surface-display of a genetically-encoded peptide library—including 2600 sequences, each spanning 15 residues surrounding known pTyr sites in the human proteome—with deep sequencing[89]. With this method, they incubated different kinases in the presence of bacteria displaying their human pTyr peptide library, with one peptide of defined sequence per cell. After kinase phosphorylation, the cells are sorted based on phosphorylation level measured by a pan-pTyr antibody. Finally, deep sequencing of the peptide-coding DNA before and after sorting provides an "enrichment score" describing the efficiency with which each substrate is phosphorylated by the kinase[89]. This technique quantifies the phosphorylation abundance of each motif, allowing a superior specificity profiling than PSPL. The increasing number of kinases profiled with this method will enable more accurate kinase-substrate or kinase-cluster predictions.

In the case of models integrating signaling networks with downstream phenotypic response, model predictions can be validated through pharmacological or genetic inhibition of the predicted signaling network nodes associated with a given phenotype. As one example, a previous PLSR model of phosphoproteomic data and cell proliferation data obtained from a set of Y-to-F mutant EGFRvIII isogenic cell lines yielded a counter-intuitive prediction that the ERK MAPKs were negatively associated with cell proliferation. This prediction was validated chemically through use of MEK inhibitors and genetically through use of constitutively active MEK. In summary, we

propose an experimental and computational framework to investigate, associate, and validate RTK-driven signaling pathways and phenotypes.

11. Forward-looking perspective

In this review, we highlighted the value of using Y-to-F studies to research the underlying mechanisms of signal propagation in RTKs. Linking signaling data to functional response in a cellular environment is crucial, whether the goal is improved mechanistic understanding or identification of novel biomarkers or drug targets. Recent advances in experimental and computational tools and methods have great potential to add additional insight we can gather with this type of data.

Advances in CRISPR methods have made it more feasible to make point mutations in endogenous RTKs, preserving expression levels and cellular environment. Viral-expressing model systems, although easy to use and manipulate, are challenging when it comes to controlling expression levels between cell lines and replicates, due to random insertion, even in an inducible system.[41] Furthermore, the use of a viral system limits the choice of model system, due to potential confounding signals from endogenous expressing genes. CRISPR point mutations are now also shown to be successfully made in *in vivo* settings, opening the door for potential Y-to-F studies in animal studies (e.g., PDX models).[90]

Increased sensitivity and accuracy of methods and instruments have yielded promising results for single-cell and single-molecule phosphoproteomics.[91,92] Automated microfluidic chip LC systems are an important component to allow quantitative identification of already low-abundant phosphorylated tyrosine.[93] These analyses are especially beneficial as they will reveal information about the stoichiometry of phosphorylation of the cell or receptor at a given time, and they could

allow profiling of a large panel of Y-to-F perturbations. One potential translational application would be in understanding tumor heterogeneity, which continues to be one of the main obstacles when it comes to cancer treatments.[94] Before this method can gain more traction, though, issues regarding sensitivity and low throughput will need to be resolved.

Finally, combining targeted phosphoproteomic strategies with absolute quantification methods, such as Surequant, will allow for reproducible pTyr profiling of commonly dysregulated oncogenic signaling proteins.[95] These advancements allow us to build on previously gathered mechanistic insight regarding RTK signal and function and to continue to explore the underlying mechanisms that explain the complex signaling network and responses that these RTKs regulate in health and disease.

Declaration of Competing Interest

The authors declare no competing financial interests.

Acknowledgments

This work was supported by NIH U01-CA215709 to A.S.M. and F.M.W., a grant from the Jayne Koskinas Ted Giovanis Foundation for Health and Policy to A.S.M., and in part by the UCLA Jonsson Comprehensive Cancer Center (JCCC) grant NIH P30-CA016042.

References

1. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).

2. Robinson, D. R., Wu, Y. M. & Lin, S. F. The protein tyrosine kinase family of the human genome. *Oncogene* **19**, 5548–57 (2000).

3. Sharma, S. V., Bell, D. W., Settleman, J. & Haber, D. A. Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* **7**, 169–181 (2007).

4. Bardelli, A. *et al.* Mutational analysis of the tyrosine kinome in colorectal cancers. *Science (1979)* **300**, 949 (2003).

5. Brewer, M. R. *et al.* Mechanism for activation of mutated epidermal growth factor receptors in lung cancer. *Proc Natl Acad Sci U S A* **110**, (2013).

6. Wang, T. H. *et al.* Integrated Omics Analysis of Non-Small-Cell Lung Cancer Cells Harboring the EGFR C797S Mutation Reveals the Potential of AXL as a Novel Therapeutic Target in TKI-Resistant Lung Cancer. *Cancers 2021, Vol. 13, Page 111* **13**, 111 (2020).

7. Friedlaender, A. *et al.* EGFR and HER2 exon 20 insertions in solid tumours: from biology to treatment. *Nature Reviews Clinical Oncology 2021 19:1* **19**, 51–69 (2021).

8. Jänne, P. A., Engelman, J. A. & Johnson, B. E. Epidermal growth factor receptor mutations in non-small-cell lung cancer: implications for treatment and tumor biology. *J Clin Oncol* **23**, 3227–3234 (2005).

9. Kobayashi, S. *et al.* EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* **352**, 786–792 (2005).

10. Remon, J., Steuer, C. E., Ramalingam, S. S. & Felip, E. Osimertinib and other third-generation EGFR TKI in EGFR-mutant NSCLC patients. *Annals of Oncology* **29**, i20–i27 (2018).

11. Wilson, T. R. *et al.* Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors. *Nature* **487**, 505–509 (2012).

12. Tao, J. J. *et al.* Antagonism of EGFR and HER3 enhances the response to inhibitors of the PI3K-Akt pathway in triple-negative breast cancer. *Sci Signal* **7**, (2014).

13. Turke, A. B. *et al.* Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC. *Cancer Cell* **17**, 77–88 (2010).

14. Wilson, T. R., Lee, D. Y., Berry, L., Shames, D. S. & Settleman, J. Neuregulin-1-mediated autocrine signaling underlies sensitivity to HER2 kinase inhibitors in a subset of human cancers. *Cancer Cell* **20**, 158–172 (2011).

15. Zhang, Z. *et al.* Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. *Nature Genetics* **44**, 852–860 (2012).

16. Hubbard, S. R. Structural analysis of receptor tyrosine kinases. *Prog Biophys Mol Biol* **71**, 343–358 (1999).

17. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–34 (2010).

18. Manning, B. D. & Cantley, L. C. AKT/PKB signaling: navigating downstream. *Cell* **129**, 1261–1274 (2007).

19. Patterson, R. L., Van Rossum, D. B., Nikolaidis, N., Gill, D. L. & Snyder, S. H. Phospholipase C-gamma: diverse roles in receptor-mediated calcium signaling. *Trends Biochem Sci* **30**, 688–697 (2005).

20. Oda, K., Matsuoka, Y., Funahashi, A. & Kitano, H. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology* **1**, 2005.0010 (2005).

21. Resat, H., Ewald, J. A., Dixon, D. A. & Wiley, H. S. An Integrated Model of Epidermal Growth Factor Receptor Trafficking and Signal Transduction. *Biophysical Journal* **85**, 730 (2003).

22. Wang, D. Y. Q., Cardelli, L., Phillips, A., Piterman, N. & Fisher, J. Computational modeling of the EGFR network elucidates control mechanisms regulating signal dynamics. *BMC Systems Biology* **3**, 1–17 (2009).

23. Hinsby, A. M., Olsen, J. V. & Mann, M. Tyrosine phosphoproteomics of fibroblast growth factor signaling: a role for insulin receptor substrate-4. *J Biol Chem* **279**, 46438–46447 (2004).

24. Schulze, W. X., Deng, L. & Mann, M. Phosphotyrosine interactome of the ErbB-receptor kinase family. *Molecular Systems Biology* (2005) doi:10.1038/msb4100012.

25. Jones, R. B., Gordus, A., Krall, J. A. & MacBeath, G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **439**, 168–174 (2006).

26. Westermark, B., Claesson-Welsh, L. & Heldin, C.-H. Structural and functional aspects of platelet-derived growth factor and its receptors. *Ciba Found Symp* **150**, 6–22 (1990).

27. Yokote, K. *et al.* Identification of Tyr-762 in the platelet-derived growth factor alpha-receptor as the binding site for Crk proteins. *Oncogene* **16**, 1229–1239 (1998).

28. Gallivan, J. P. & Dougherty, D. A. Cation-π interactions in structural biology. *Proc Natl Acad Sci U S A* **96**, 9459–9464 (1999).

29. Colette Daubner, S. & Fitzpatrick, P. F. Mutation to phenylalanine of tyrosine 371 in tyrosine hydroxylase increases the affinity for phenylalanine. *Biochemistry* **37**, 16440–16444 (1998).

30. Teng, S., Srivastava, A. K., Schwartz, C. E., Alexov, E. & Wang, L. Structural Assessment of the Effects of Amino Acid Substitutions on Protein Stability and Protein-Protein Interaction. *International Journal of Computational Biology and Drug Design* **3**, 334 (2010).

31. Zisch, A. H. *et al.* Replacing two conserved tyrosines of the EphB2 receptor with glutamic acid prevents binding of SH2 domains without abrogating kinase activity and biological responses. *Oncogene 2000 19:2* **19**, 177–187 (2000).

32. G M Walton, W S Chen, M G Rosenfeld, G. N. G. Analysis of deletions of the carboxyl terminus of the epidermal growth factor receptor reveals self-phosphorylation at tyrosine 992 and enhanced in vivo tyrosine phosphorylation of cell substrates. *J Biol Chem.* **265**, 1750–4 (1990).

33. Roskoski, R. The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacological Research* **79**, 34–74 (2014).

34. Foehr, E. D., Raffioni, S., Murray-Rust, J. & Bradshaw, R. A. The Role of Tyrosine Residues in Fibroblast Growth Factor Receptor 1 Signaling in PC12 Cells: SYSTEMATIC SITE-DIRECTED MUTAGENESIS IN THE ENDODOMAIN *. *Journal of Biological Chemistry* **276**, 37529–37536 (2001).

35. Lemmon, M. A., Freed, D. M., Schlessinger, J. & Kiyatkin, A. The Dark Side of Cell Signaling: Positive Roles for Negative Regulators. *Cell* **164**, 1172–1184 (2016).

36. Songyang, Z., Gish, G., Mbamalu, G., Pawson, T. & Cantley, L. C. A single point mutation switches the specificity of group III Src homology (SH) 2 domains to that of group I SH2 domains. *J Biol Chem* **270**, 26029–26032 (1995).

37. Jin, L. L. *et al.* Tyrosine phosphorylation of the Lyn Src homology 2 (SH2) domain modulates its binding affinity and specificity. *Mol Cell Proteomics* **14**, 695–706 (2015).

38. Zhang, Y., McKeehan, K., Lin, Y., Zhang, J. & Wang, F. Fibroblast Growth Factor Receptor 1 (FGFR1) Tyrosine Phosphorylation Regulates Binding of FGFR Substrate 2α (FRS2α) But Not FRS2β to the Receptor. *Molecular Endocrinology* **22**, 167–175 (2008).

39. Cullati, S. N. *et al.* Kinase domain autophosphorylation rewires the activity and substrate specificity of CK1 enzymes. *Molecular Cell* (2022) doi:10.1016/j.molcel.2022.03.005.

40. Wagner, J. P. *et al.* Receptor tyrosine kinases fall into distinct classes based on their inferred signaling networks. *Sci Signal* **6**, ra58 (2013).

41. Gill, K., Macdonald-Obermann, J. L. & Pike, L. J. Epidermal growth factor receptors containing a single tyrosine in their C-terminal tail bind different effector molecules and are signaling-competent. *J Biol Chem* **292**, 20744–20755 (2017).

42. N, G., M, T. & M, S. Tyrosine phosphorylation sites at amino acids 239 and 240 of Shc are involved in epidermal growth factor-induced mitogenic signaling that is distinct from Ras/mitogen-activated protein kinase activation. *Mol Cell Biol* **17**, 1824–1831 (1997).

43. Salas, V. *et al.* Characterisation of tyrosine-phosphorylation-defective calmodulin mutants. *Protein Expr Purif* **41**, 384–392 (2005).

44. Sakaguchi, K. *et al.* Shc phosphotyrosine-binding domain dominantly interacts with epidermal growth factor receptors and mediates Ras activation in intact cells. *Mol Endocrinol* **12**, 536–543 (1998).

45. van der Geer, P. *et al.* A conserved amino-terminal Shc domain binds to phosphotyrosine motifs in activated receptors and phosphopeptides. *Curr Biol* **5**, 404–412 (1995).

46. Blanes, M. G., Oubaha, M., Rautureau, Y. & Gratton, J. P. Phosphorylation of tyrosine 801 of vascular endothelial growth factor receptor-2 is necessary for Akt-dependent endothelial nitric-oxide synthase activation and nitric oxide release from endothelial cells. *J Biol Chem* **282**, 10660–10669 (2007).

47. Meyer, R. D., Latz, C. & Rahimi, N. Recruitment and activation of phospholipase Cgamma1 by vascular endothelial growth factor receptor-2 are required for tubulogenesis and differentiation of endothelial cells. *J Biol Chem* **278**, 16347–16355 (2003).

48. Ren, Z. & Schaefer, T. S. ErbB-2 activates Stat3 alpha in a Src- and JAK2-dependent manner. *J Biol Chem* **277**, 38486–38493 (2002).

49. Tallquist, M. D. & Soriano, P. Cell autonomous requirement for PDGFRalpha in populations of cranial and cardiac neural crest cells. *Development* **130**, 507–518 (2003).

50. Huang, P. H. *et al.* ANALYSIS OF SITE-SPECIFIC MUTATIONS ON EGFRVIII PHOSPHOTYROSINE SIGNALING NETWORK IDENTIFIES DETERMINANTS GOVERNING U87MG GLIOBLASTOMA CELL GROWTH. *Molecular Biosystems* **6**, 1227 (2010).

51. Porta, R. *et al.* FGFR a promising druggable target in cancer: Molecular biology and new drugs. *Crit Rev Oncol Hematol* **113**, 256–267 (2017).

52. Hart, K. C., Robertson, S. C. & Donoghue, D. J. Identification of Tyrosine Residues in Constitutively Activated Fibroblast Growth Factor Receptor 3 Involved in Mitogenesis, Stat Activation, and Phosphatidylinositol 3-Kinase Activation. *Molecular Biology of the Cell* **12**, 931 (2001).

53. McDaniel, N. K. *et al.* AXL Mediates Cetuximab and Radiation Resistance Through Tyrosine 821 and the c-ABL Kinase Pathway in Head and Neck Cancer. *Clinical Cancer Research* **26**, 4349–4359 (2020).

54. Batzer, A. G., Rotin, D., Ureña, J. M., Skolnik, E. Y. & Schlessinger, J. Hierarchy of binding sites for Grb2 and Shc on the epidermal growth factor receptor. *Mol Cell Biol* **14**, 5192–5201 (1994).

55. Nogami, M. *et al.* Requirement of autophosphorylated tyrosine 992 of EGF receptor and its docking protein phospholipase Cγ1 for membrane ruffle formation. *FEBS Letters* **536**, 71–76 (2003).

56. Tanos, B. & Pendergast, A. M. Abl tyrosine kinase regulates endocytosis of the epidermal growth factor receptor. *J Biol Chem* **281**, 32714–32723 (2006).

57. Smith, J. E., Sheng, zu F. & Kallen, R. G. Effects of tyrosine-->phenylalanine mutations on auto- and trans-phosphorylation reactions catalyzed by the insulin receptor beta-subunit cytoplasmic domain. *DNA Cell Biol* **13**, 593–604 (1994).

58. Jiang, Y., Chan, J. L. K., Zong, C. S. & Wang, L. H. Effect of Tyrosine Mutations on the Kinase Activity and Transforming Potential of an Oncogenic Human Insulin-like Growth Factor I Receptor (∗). *Journal of Biological Chemistry* **271**, 160–167 (1996).

59. Vijapurkar, U., Cheng, K. & Koland, J. G. Mutation of a Shc binding site tyrosine residue in ErbB3/HER3 blocks heregulin-dependent activation of mitogen-activated protein kinase. *J Biol Chem* **273**, 20996–21002 (1998).

60. Kazlauskas, A., Durden, D. L. & Cooper, J. A. Functions of the major tyrosine phosphorylation site of the PDGF receptor beta subunit. *Cell Regul* **2**, 413–425 (1991).

61. Wardega, P., Heldin, C. H. & Lennartsson, J. Mutation of tyrosine residue 857 in the PDGF beta-receptor affects cell proliferation but not migration. *Cell Signal* **22**, 1363–1368 (2010).

62. Hellyer, N. J., Kim, M. S. & Koland, J. G. Heregulin-dependent activation of phosphoinositide 3-kinase and Akt via the ErbB2/ErbB3 co-receptor. *J Biol Chem* **276**, 42153–42161 (2001).

63. Bae, J. H. *et al.* Asymmetric receptor contact is required for tyrosine autophosphorylation of fibroblast growth factor receptor in living cells. *Proceedings of the National Academy of Sciences* **107**, 2866–2871 (2010).

64. Balaji, K. *et al.* AXL inhibition suppresses the DNA damage response and sensitizes cells to PARP inhibition in multiple cancers. *Molecular Cancer Research* **15**, 45–58 (2017).

65. Revach, O. Y., Sandler, O., Samuels, Y. & Geiger, B. Cross-talk between receptor tyrosine kinases AXL and ERBB3 regulates invadopodia formation in melanoma cells. *Cancer Research* **79**, 2634–2648 (2019).

66. Hoj, J. P., Mayro, B. & Pendergast, A. M. A TAZ-AXL-ABL2 Feed-Forward Signaling Axis Promotes Lung Adenocarcinoma Brain Metastasis. *Cell Rep* **29**, 3421-3434.e8 (2019).

67. Zhang, Z. *et al.* Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. *Nat Genet* **44**, 852–860 (2012).

68. Vasan, N., Baselga, J. & Hyman, D. M. A view on drug resistance in cancer. *Nature* **575**, 299–309 (2019).

69. Casado, P. *et al.* Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci Signal* **6**, (2013).

70. Beekhof, R. *et al.* INKA, an integrative data analysis pipeline for phosphoproteomic inference of active kinases. *Molecular Systems Biology* **15**, e8250 (2019).

71. Obenauer, J. C., Cantley, L. C. & Yaffe, M. B. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res* **31**, 3635–3641 (2003).

72. Horn, H. *et al.* KinomeXplorer: an integrated platform for kinome biology studies. *Nat Methods* **11**, 603–604 (2014).

73. McDaniel, N. K. *et al.* AXL Mediates Cetuximab and Radiation Resistance Through Tyrosine 821 and the c-ABL Kinase Pathway in Head and Neck Cancer. *Clin Cancer Res* **26**, 4349–4359 (2020).

74. KE, S., VS, S., EA, W., SC, G. & KM, N. Ensemble clustering of phosphoproteomic data identifies differences in protein interactions and cell-cell junction integrity of HER2-overexpressing cells. *Integr Biol (Camb)* **9**, 539–547 (2017).

75. Creixell, M. & Meyer, A. S. Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data. *Cell Reports Methods* **2**, 100167 (2022).

76. Morshed, N. *et al.* Quantitative phosphoproteomics uncovers dysregulated kinase networks in Alzheimer's disease. *Nature Aging 2021 1:6* **1**, 550–565 (2021).

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

78. Bareja, A., Hodgkinson, C. P., Soderblom, E., Waitt, G. & Dzau, V. J. The proximity-labeling technique BioID identifies sorting nexin 6 as a member of the insulin-like growth factor 1 (IGF1)–IGF1 receptor pathway. *Journal of Biological Chemistry* **293**, 6449–6459 (2018).

79. Uçkun, E. *et al.* BioID-Screening Identifies PEAK1 and SHP2 as Components of the ALK Proximitome in Neuroblastoma Cells. *Journal of Molecular Biology* **433**, 167158 (2021).

80. Songyang, Z. *et al.* Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. *Nature* **373**, 536–539 (1995).

81. Songyang, Z. *et al.* A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1. *Molecular and Cellular Biology* **16**, 6486–6493 (1996).

82. Wang, C. *et al.* Determination of CK2 Specificity and Substrates by Proteome-Derived Peptide Libraries. *Journal of Proteome Research* **12**, 3813–3821 (2013).

83. Meggio, F., Marin, O. & Pinna, L. A. Substrate specificity of protein kinase CK2. *Cell Mol Biol Res* **40**, 401–9 (1994).

84. Majumder, A. *et al.* Integrated Proteomics-Based Physical and Functional Mapping of AXL Kinase Signaling Pathways and Inhibitors Define Its Role in Cell Migration. *Molecular Cancer Research* 1–14 (2022) doi:10.1158/1541-7786.mcr-21-0275.

85. Alexander, J. *et al.* Spatial Exclusivity Combined with Positive and Negative Selection of Phosphorylation Motifs Is the Basis for Context-Dependent Mitotic Signaling. *Sci Signal* **4**, ra42 (2011).

86. Begley, M. J. *et al.* EGF-receptor specificity for phosphotyrosine-primed substrates provides signal integration with Src. *Nat Struct Mol Biol* **22**, 983–990 (2015).

87. Kooij, B. van de *et al.* Comprehensive substrate specificity profiling of the human nek kinome reveals unexpected signaling outputs. *Elife* **8**, (2019).

88. Obata, T. *et al.* Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. *J Biol Chem* **275**, 36108–36115 (2000).

89. Shah, N. H., Löbel, M., Weiss, A. & Kuriyan, J. Fine-tuning of substrate preferences of the Src-family kinase Lck revealed through a high-throughput specificity screen. *Elife* **7**, 1–24 (2018).

90. Levi, T., Sloutskin, A., Kalifa, R., Juven-Gershon, T. & Gerlitz, O. Efficient In Vivo Introduction of Point Mutations Using ssODN and a Co-CRISPR Approach. *Biol Proced Online* **22**, (2020).

91. Polat, A. N. & Özlü, N. Towards single-cell LC-MS phosphoproteomics. *Analyst* **139**, 4733–4749 (2014).

92. Gerritsen, J. S. & White, F. M. Phosphoproteomics: a valuable tool for uncovering molecular signaling in cancer cells. *Expert Rev Proteomics* **18**, 661–674 (2021).

93. Wei, W. *et al.* Single-Cell Phosphoproteomics Resolves Adaptive Signaling Dynamics and Informs Targeted Combination Therapy in Glioblastoma. *Cancer Cell* **29**, 563–573 (2016).

94. Lun, X. K. & Bodenmiller, B. Profiling Cell Signaling Networks at Single-cell Resolution. *Mol Cell Proteomics* **19**, 744–756 (2020).

95. Stopfer, L. E. *et al.* High-density, targeted monitoring of tyrosine phosphorylation reveals activated signaling networks in human tumors. *Cancer Research* **81**, canres.3804.2020 (2021).

96. Wee, P. & Wang, Z. Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways. *Cancers (Basel)* **9**, (2017).

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

98. Vu, T. & Claret, F. X. Trastuzumab: Updated mechanisms of action and resistance in breast cancer. *Frontiers in Oncology* **2 JUN**, 62 (2012).

99. Gaborit, N., Lindzen, M. & Yarden, Y. Emerging anti-cancer antibodies and combination therapies targeting HER3/ERBB3. *Human Vaccines & Immunotherapeutics* **12**, 576–592 (2015).

100. Heldin, C. H. & Lennartsson, J. Structural and functional properties of platelet-derived growth factor and stem cell factor receptors. *Cold Spring Harb Perspect Biol* **5**, (2013).

101. Álvarez-Aznar, A., Muhl, L. & Gaengel, K. VEGF Receptor Tyrosine Kinases: Key Regulators of Vascular Function. *Current Topics in Developmental Biology* **123**, 433–482 (2017).

102. Cross, M. J. & Claesson-Welsh, L. FGF and VEGF function in angiogenesis: signalling pathways, biological responses and therapeutic inhibition. *Trends Pharmacol Sci* **22**, 201–207 (2001).

103. Erickson, K. E. *et al.* Modeling cell line-specific recruitment of signaling proteins to the insulin-like growth factor 1 receptor. *PLOS Computational Biology* **15**, e1006706 (2019).

104. Myers, M. G. *et al.* The COOH-terminal tyrosine phosphorylation sites on IRS-1 bind SHP-2 and negatively regulate insulin signaling. *J Biol Chem* **273**, 26908–26914 (1998).

105. Ferguson, H. R., Smith, M. P. & Francavilla, C. Fibroblast Growth Factor Receptors (FGFRs) and Noncanonical Partners in Cancer Signaling. *Cells 2021, Vol. 10, Page 1201* **10**, 1201 (2021).

106. Auyez, A., Sayan, A. E., Kriajevska, M. & Tulchinsky, E. AXL Receptor in Cancer Metastasis and Drug Resistance: When Normal Functions Go Askew. *Cancers 2021, Vol. 13, Page 4864* **13**, 4864 (2021).

107. Linger, R. M. A., Keating, A. K., Earp, H. S. & Graham, D. K. TAM receptor tyrosine kinases: biologic functions, signaling, and potential therapeutic targeting in human cancer. *Adv Cancer Res* **100**, 35–83 (2008).

108. Gao, J., Inagaki, Y., Xue, X., Qu, X. & Tang, W. c-Met: A potential therapeutic target for hepatocellular carcinoma. *Drug Discov Ther* **5**, 2–11 (2011).

109. Penengo, L., Rubin, C., Yarden, Y. & Gaudino, G. c-Cbl is a critical modulator of the Ron tyrosine kinase receptor. *Oncogene 2003 22:24* **22**, 3669–3679 (2003).

Chapter 2:

Dual Data-Motif Clustering improves modeling of phosphoproteomic data

Creixell, M. and Meyer A

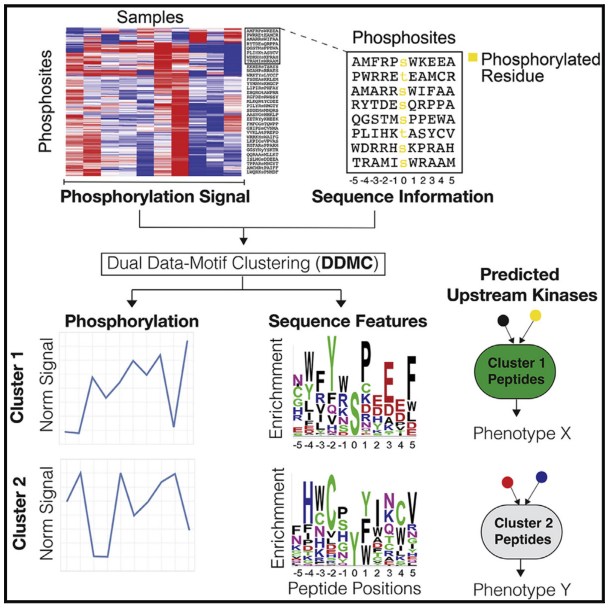
Department of Bioengineering, University of California Los Angeles

Cell Reports
Methods

Article

Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data

Graphical abstract



Highlights

- DDMC clusters phosphoproteomic data using the peptide's sequence and abundance
- DDMC accurately predicts upstream kinases regulating clusters
- Clusters identify signaling specific to tumors, mutations, and immune infiltration
- DDMC provides a general and flexible strategy for phosphoproteomic analysis

Authors

Marc Creixell, Aaron S. Meyer

Correspondence

ameyer@asmlab.org

In brief

Measuring cell signaling by mass spectrometry–based phosphoproteomics provides a promising opportunity to direct cancer therapy development. Here, we present dual data motif clustering (DDMC), a clustering and kinase prediction strategy that identifies signaling nodes by grouping phosphosites according to their phosphorylation signal and amino acid sequence.

Creixell & Meyer, 2022, Cell Reports Methods 2, 100167
February 28, 2022 a 2022 The Author(s).
https://doi.org/10.1016/j.crmeth.2022.100167

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

Cell Reports Methods

CellPress
OPEN ACCESS

Article

Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data

Marc Creixell[1] and Aaron S. Meyer[1,2,3,4,5,*]
[1]Department of Bioengineering, University of California, Los Angeles, Los Angeles, CA 90024, USA
[2]Department of Bioinformatics, University of California, Los Angeles, Los Angeles, CA 90024, USA
[3]Jonsson Comprehensive Cancer Center, University of California, Los Angeles, Los Angeles, CA 90024, USA
[4]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, Los Angeles, CA 90024, USA
[5]Lead contact
*Correspondence: ameyer@asmlab.org
https://doi.org/10.1016/j.crmeth.2022.100167

MOTIVATION Measuring cell signaling by mass spectrometry–based phosphoproteomics provides a promising opportunity to direct cancer therapy development. Despite continued progress in profiling the phosphoproteomes of patients across different cancer types, challenges inherent to these types of data hinder the identification of clinically relevant proteomic alterations. Here, we present DDMC: a clustering and kinase prediction strategy that identifies signaling nodes by grouping phosphosites according to their phosphorylation signal and amino acid sequence. The cluster centers, by virtue of being summaries of the phosphorylation changes of those phosphosites, can be used to establish associations with signaling responses. The sequence features of clusters can be used to identify the upstream kinases regulating them. In doing so, this method reconstructs signaling networks into biologically meaningful clusters that can be associated with cell responses and upstream kinase drivers.

SUMMARY

Cell signaling is orchestrated in part through a network of protein kinases and phosphatases. Dysregulation of kinase signaling is widespread in diseases such as cancer and is readily targetable through inhibitors. Mass spectrometry–based analysis can provide a global view of kinase regulation, but mining these data is complicated by its stochastic coverage of the proteome, measurement of substrates rather than kinases, and the scale of the data. Here, we implement a dual data and motif clustering (DDMC) strategy that simultaneously clusters peptides into similarly regulated groups based on their variation and their sequence profile. We show that this can help to identify putative upstream kinases and supply more robust clustering. We apply this clustering to clinical proteomic profiling of lung cancer and identify conserved proteomic signatures of tumorigenicity, genetic mutations, and immune infiltration. We propose that DDMC provides a general and flexible clustering strategy for the analysis of phosphoproteomic data.

INTRODUCTION

Cell signaling networks formed by protein kinases dictate cell fate and behavior through protein phosphorylation, including in diseases such as cancer (Hunter, 1995). Measuring cell signaling by mass spectrometry (MS)–based global phosphoproteomics provides a promising opportunity to direct therapy development (Yaffe, 2019), particularly given the accessibility of these signaling changes to drug targeting. Nevertheless, despite the rapid accumulation of large-scale phosphoproteomic clinical data, it is still difficult to link signaling events leading to observed proteomic alterations and phenotypic outcomes.

One approach to analyze phosphoproteomic measurements has been to infer the activity of upstream kinases. For instance, kinase–substrate enrichment analysis averages the signals of groups of known kinase substrates to infer enriched pathways in biological samples (Casado et al., 2013). Another method, integrative inferred kinase activity (INKA), infers kinase activity by integrating the overall and activation loop phosphorylation of kinases alongside the phosphorylation abundance of known substrates. Kinase–substrate relationships are either experimentally

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

determined or predicted by NetworKIN, an algorithm that uses sequence motif and protein–protein network information (Linding et al., 2007; Beekhof et al., 2019; Hornbeck et al., 2019). Finally, Scansite predicts kinase–substrate interactions using sequence motifs generated from oriented peptide library scanning experiments (Obenauer et al., 2003). These methods, sometimes in combination, help to reconstruct signaling pathway activities from individual samples.

However, due to several limitations, kinase–substrate inference still provides a limited view of signaling network changes. Kinase prediction methods are necessarily dependent on having well–characterized kinase–substrate interactions, but most of the phosphoproteome remains largely uncharacterized (Needham et al., 2019). Just 20% of kinases have been shown to phosphorylate 87% of currently annotated substrates, and around 80% of kinases have fewer than 20 substrates, with 30% yet to be assigned a single substrate (Needham et al., 2019). Insights dependent on this unequal knowledge distribution are less likely to identify understudied protein kinases. An additional major challenge, particularly with discovery–mode multiplexed tandem mass tag (TMT) MS, is missing values. The technique processes batches of samples with stochastic coverage in each experiment. This means that the portion of the phosphoproteome quantified in the samples of different TMT experiments varies (Tabb et al., 2010). Computational tools usually require complete datasets, and so data are frequently preprocessed by imputing missing values—inflating the effect of certain measurements or throwing out any peptides displaying missing values—at the expense of losing critical information (Chen et al., 2020; Gillette et al., 2020). Kinase enrichment and prediction methods are further compromised by this problem.

Clustering methods, such as hierarchical clustering or k–means, can be used to cluster phosphopeptides based on similarities in the patterns of their abundance (Mertins et al., 2016; Chen et al., 2020; Deb et al., 2020). This clustering criterion results in groups of peptides that display similar phosphorylation patterns across conditions, but that may be targeted by sets of different upstream kinases that are not directly inferred by these methods. The residues surrounding phosphorylation sites have evolved to become fine–tuned motifs that confer signaling specificity and fidelity (Zarrinpar et al., 2003; Tan et al., 2009). Clustering based on motif similarity might, therefore, improve model interpretation by facilitating the identification of upstream kinases modulating clusters that display conserved sequence motifs. On the other hand, clustering peptides based on sequence alone may result in groups of proteins that, while sharing the same set of upstream kinases, are differently regulated due to context. We therefore hypothesized that combining phosphorylation status and sequence similarity may enable a balanced characterization of the cell signaling state.

Here, we present an algorithm known as dual data and motif clustering (DDMC) that probabilistically and simultaneously models both the peptide phosphorylation variation and peptide sequence motifs of peptide clusters to reconstitute cell signaling networks (Figure 1). A key distinction of DDMC is that it analyzes multidimensional data, whereas kinase enrichment tools operate on individual samples, relying on prior knowledge. Importantly, DDMC clusters are not limited to pre–existent kinase motifs and therefore do not rely on previous kinase–substrate characterization. Thus, DDMC kinase predictions can lead to the association of understudied kinases and phenotypic responses. We propose that DDMC represents a unified alternative that overcomes fundamental methodologic issues of current tools. To test the utility of our method, we analyzed the phosphoproteomes of 110 treatment–naïve lung adenocarcinoma (LUAD) tumors and 101 paired normal adjacent tissues (NATs) from the National Cancer Institute (NCI)'s Clinical Proteomic Tumor Analysis Consortium (CPTAC) LUAD study (Gillette et al., 2020). We characterized the phosphoproteome of patients by identifying those signaling signatures associated with tumorigenesis, the presence of specific mutations, and tumor immune infiltration. In total, we demonstrated DDMC as a general strategy for improving the analysis of phosphoproteomic surveys.

RESULTS

Constructing an expectation–maximization algorithm tailored for clustering phosphoproteomic data

In seeking to cluster phosphoproteomic measurements, we recognized that these data provide two pieces of information: the exact site of phosphorylation on a peptide sequence and some measure of abundance within the measured samples. Both pieces of information are critical to the overall interpretation of the data. Based on this observation, we built a mixture model that probabilistically clusters phosphosites based on both their peptide sequence and abundance across samples (Figure S1). In each iteration, DDMC applies an expectation–maximization algorithm to optimize clusters that capture the average features of member sequences and their abundance variation (Figures 1A and S1). Both information sources—the peptide abundance and sequence—can be prioritized during cluster fitting by a weight parameter. With a weight of 0, DDMC becomes a Gaussian mixture model (GMM) that exclusively clusters peptides according to their phosphorylation signal. With a very large weight, DDMC primarily clusters peptides according to their peptide sequences. Clustering both the sequence and abundance measurements ensures that the resulting clusters are a function of both features, which we hypothesized would provide both more meaningful and robust clusters.

The resulting clustering provides coordinated outputs that can be used in a few different ways. The cluster centers, by virtue of being a summary for the abundance changes of these peptides, can be regressed against phenotypic responses (e.g., cell phenotypes or clinical outcomes) to establish associations between clusters and response (Figure 1B). Regression using the clusters instead of each peptide ensures that the model can be developed despite relatively few samples, with minimal loss of information since each peptide within a cluster varies in a similar manner. One can also interrogate the position–specific scoring matrices (PSSMs) from the resulting cluster sequence motifs. Given a set of peptide sequences, PSSMs quantify the amino acid frequencies across peptide positions and show to what extent each residue is enriched or depleted per position (Figure 1A). Thus, a cluster PSSM provides a general representation of the cluster sequence features and can be readily compared with other information, such as experimentally generated profiles

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

Cell Reports Methods
Article

CellPress
OPEN ACCESS

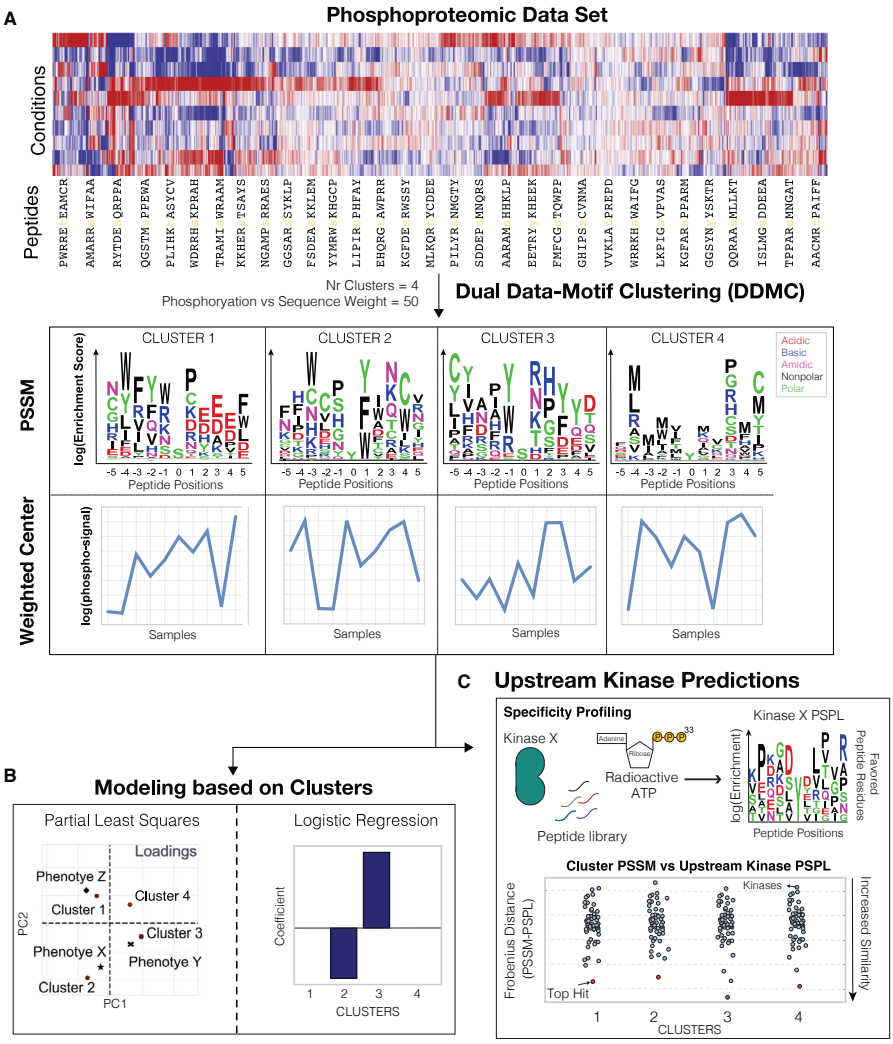


Figure 1. Schematic of the DDMC approach to cluster global signaling data and infer upstream kinases driving phenotypes

(A) DDMC is run to cluster an input phosphoproteomic dataset to generate four clusters of peptides that show similar sequence motifs and phosphorylation behavior.

(B) Predictive modeling using clusters allows one to establish associations between specific clusters and features of interest.

(C) Putative upstream kinases regulating clusters can be predicted by comparing the experimentally generated specificity profiles of upstream kinases (kinase PSPL) and the cluster PSSMs PSSM; Position-specific scoring matrix, PSPL; Position scanning peptide library (Hutti et al., 2004; Begley et al., 2015).

See also Figures S1 and S2.

of putative upstream kinases via position-specific scanning libraries (PSPLs) (Obata et al., 2000; Snyder et al., 2010). In this technique, a kinase of interest is individually incubated with each of 180 different peptide libraries in which each library contains a central phosphoacceptor residue (S/T or Y), a second fixed amino acid located any of the peptide residues spanning positions 5 throughout +4 relative to the phosphorylation site, and a degenerate mixture containing all natural amino acids at

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

all other positions. The kinase and peptide libraries are incubated in the presence of radioactive ATP, which allows the quantification of phosphorylation abundance per residue and position and the identification of the kinase's ''optimal'' substrate motif. We extracted a collection of 42 kinase specificity profiles to identify which cluster motifs most resemble the optimal motif of putative upstream kinases (Figure 1C) (Hutti et al., 2004; Miller et al., 2008; Begley et al., 2015; van de Kooij et al., 2019). However, as kinase–substrate specificity is also dictated by features outside of the immediate substrate region, we also note that our approach is more general than strictly assembling kinase–substrate predictions, as non–enzymatic specificity information may be present in the DDMC sequence motifs. This overview demonstrates how DDMC can take complex, coordinated signaling measurements and find patterns in the phosphorylation signals to reconstruct signaling networks and associate clusters and phenotypes.

DDMC robustly imputes missing values

A major limitation of discovery–mode MS–based phosphoproteomic data is the presence of missing values due to the stochastic signaling coverage in each run. In the resulting dataset, many phosphosites are observed in groups of samples and missed in others (Figure 2A). To evaluate the robustness of DDMC in analyzing incomplete datasets, we designed a computational experiment wherein we synthetically removed random TMT experiments from the dataset and predicted them using the peptide–assigned cluster centers. The mean squared error of imputation was compared with other commonly used strategies, such as the peptides' mean, filling in zeros, or matrix completion by principal–component analysis (PCA) (Figure 2A). We applied this experiment across different numbers of clusters and sequence weighting to explore the imputation performance. We observed that increasing the number of clusters consistently improved performance (Figures 2B and 2C), whereas primarily prioritizing the sequence information yielded worse imputation estimates (Figures 2D–2G). However, a weight of 100 still allowed DDMC to accurately predict missing values while incorporating the sequence information into the clustering criterion (Figures 2C and 2E–2G). We concluded that DDMC clearly outperforms many common imputation strategies and imputes missing values with similar accuracy to matrix completion by PCA.

DDMC correctly identifies AKT1 and ERK2 as upstream kinases of signaling clusters containing their substrates

A major benefit of directly modeling the phosphopeptide sequence information is the construction of cluster motifs to infer which putative upstream kinases might preferentially target a specific cluster. To validate this ability, we used DDMC to cluster the phosphoproteomic measurements of MCF7 cells treated with a panel of 61 drug inhibitors reported by Hijazi et al. (2020). We hypothesized that the phosphoproteomic clusters align to specific and identifiable targeted kinases. Examining the clusters by PCA, the scores of AKT/PI3K/mTOR targeted inhibitors (shown in orange in Figure 3A) and the loading of cluster 16 were clearly opposed (Figures 3A and 3B). The additional inhibitors GSK2334470 and LY2584702 were also negatively associated with cluster 1; both inhibit kinases PDK1 and S6K1, respectively, expected to modulate the AKT/PI3K/mTORC pathway. A heatmap displaying cluster 1's phosphorylation signal across treatments corroborates that the abundance of these peptides is substantially decreased when treated with AKT/mTOR/PI3K inhibitors (Figure 3C). Encouragingly, the AKT profile was most closely matched to the PSSM of cluster 1 within a collection of 42 different kinase PSPL matrices (Figure 3D). In addition, NetPhorest identified AKT as the eighth top scoring upstream kinase of cluster 1, further corroborating DDMC's prediction (Figure 3E).

As a second test, we extracted the sequences of experimentally validated substrates of ERK2 to create an ''artificial'' ERK2–specific PSSM positive control (ERK2+ motif) (Carlson et al., 2011)(Figure 3F). As expected, ERK2 was predicted to be the upstream kinase with the highest preference for the cluster's motif(Figure 3G). Given the consistent enrichment of hydrophobic and polar residues throughout the entire ERK2 target motif (Figure 3F), we asked whether randomly shuffling the cluster PSSM positions surrounding the phosphoacceptor residue would affect the upstream kinase prediction. Randomization led to a marked increase in the distance between the ERK2 specificity profile and the ERK2+ motif (Figure 3H). Clusters from the CPTAC dataset that were preferentially favored by ERK2 showed a similar decline in specificity between the clusters PSSMs and ERK2 PSPL matrix on randomization (Figure 3H). This experiment shows that position–specific matching information is contained within the ERK2 target motif despite the uniform biophysical properties (Figures 3G and 3H). Altogether, these results illustrate two different validation scenarios in which DDMC successfully identifies the upstream kinases regulating clusters.

DDMC improves prediction of different phenotypes and provides more robust clustering

As detailed later (Figures 5, 6, and 7), we used DDMC to analyze the phosphoproteomes of 110 treatment–naïve LUAD tumors and 101 paired NATs from the NCI's CPTAC LUAD study. We used DDMC with the binomial sequence distance method and 30 clusters (Figures 1 and 2B–2D). We were able to include 30,561 peptides that were not observed in every sample through our ability to handle missing data but still filtered out 11,822 peptides that were only captured in one 10–plex TMT run. We used this fitting result throughout the rest of this study. The resulting cluster motifs can be found in Figure S2.

To evaluate the benefit of including peptide sequence information during clustering, we investigated whether different sequence weights would affect the performance of a regularized logistic regression model that predicts the mutational status of STK11, whether a patient harbors a mutation in the epidermal growth factor receptor (EGFRm), and the level of tumor infiltration (''hot'' versus ''cold''). Three independent DDMC runs were performed to observe the reproducibility of the prediction results. We found that for all three phenotypes, optimal predictions were derived when clustering was partly based on the peptide sequence—as highlighted in red circles. In the case of STK11, the use of the maximum performance is achieved with a weight of 250. Likewise, EGFRm samples were best classified

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

CellPress
OPEN ACCESS

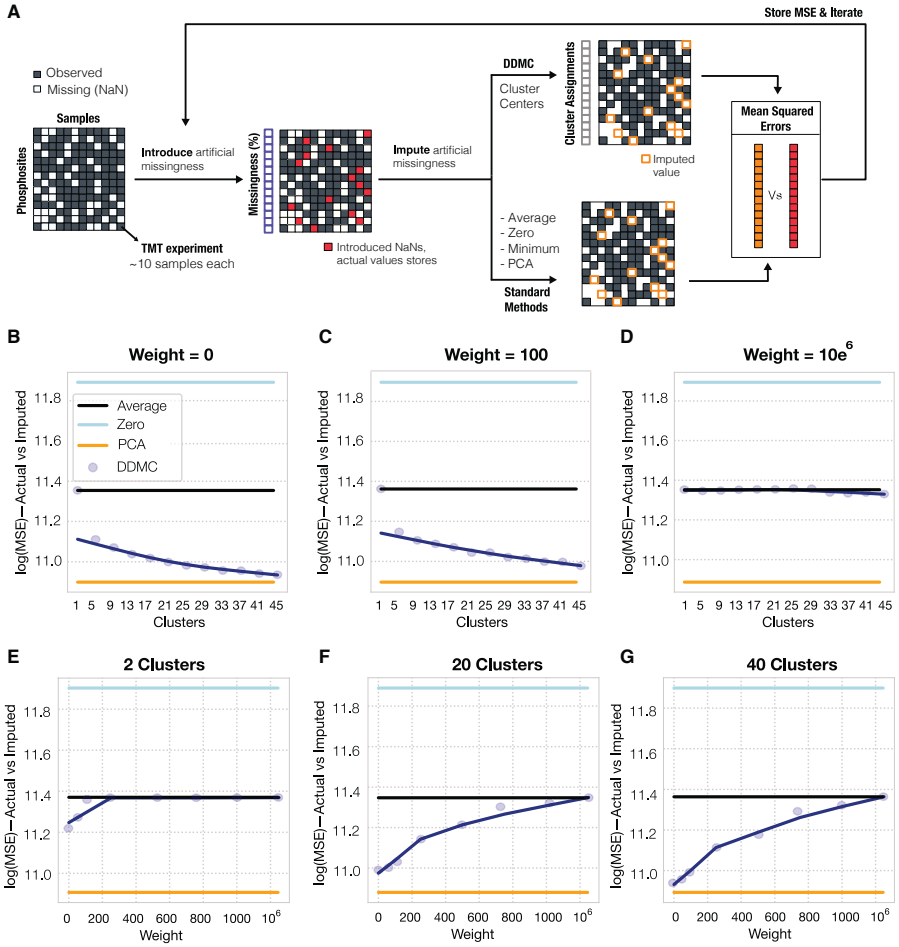


Figure 2. Benchmarking the robustness of motif clustering to missing measurements

(A) A schematic of the process for quantifying robustness to missing values. Any peptides containing fewer than seven TMT experiments were discarded. For the remaining 15,904 peptides, an entire random TMT experiment was removed per peptide and these values were stored for later comparison. Next, these artificial missing values were imputed using either a baseline strategy (peptide mean signal, constant zero, or matrix completion by PCA with five components) or the corresponding cluster center. Once a MSE was computed for each peptide, the second iteration repeats this process by removing a second TMT experiment. (B–G) A total of five random TMT experiments per peptide were imputed by clustering using a different number of clusters (B–D) or different weights (E–G). Note that the minimum signal imputation is not shown for clarity since its prediction performance was dramatically worse.

with a mix weight of 1,000. Finally, the regression model classifying whether a sample is "hot-tumor-enriched" (HTE) or "cold-tumor-enriched" (CTE) showed the best fitness with weights spanning from 100 to 750. Together, these results indicate that observing the motif information during clustering leads to final clusters that enhance the performance of downstream phenotype prediction models (Figures 4A and S3). Note that random chance is equal to 0.5 and perfect predictions 1.0, so an improvement of 0.1 (STK11 prediction) is a movement across 20% of this range.

Next, we explored how using different weights affects the overall phosphorylation signal and sequence information of the resulting clusters. To do so, we compared the model behavior after clustering the CPTAC data with a weight of 0 (peptide abundance only), 100 (mix), and 1,000,000 (mainly sequence). First, we hypothesized that the abundance-only model would generate

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

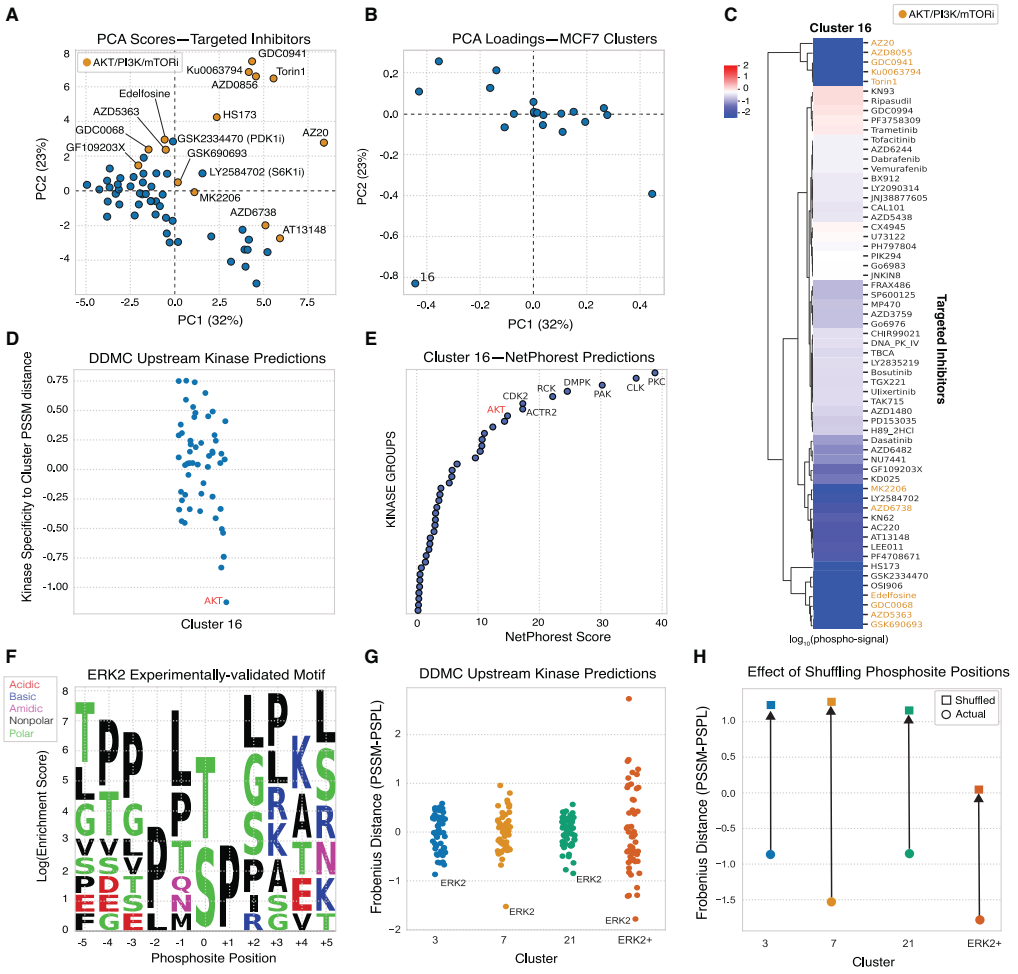


Figure 3. Validation of upstream kinase predictions

(A and B) PCA analysis of the DDMC phosphoproteome clusters of MCF7 cells subjected to a drug screen (Hijazi et al., 2020).

(C) Heatmap showing the effect of inhibitors on the phosphorylation signal of cluster 16.

(D) DDMC upstream kinase prediction of cluster 16.

(E) NetPhorest upstream kinase prediction of cluster 16.

(F) Resulting PSSM generated using reported ERK2 substrates (Carlson et al., 2011).

(G) Upstream kinase predictions of CPTAC clusters 3, 7, and 21 in addition to the ERK2 motif shown in (F).

(H) Upstream kinase predictions of the same PSSMs after randomly shuffling the motif positions.

clusters wherein its members would show less intra-cluster variation in phosphorylation signal and thus a lower mean squared error (MSE). To test this, we computed the average peptide-to-cluster MSE of 2,000 randomly selected peptides for each model across all clusters. We observed a direct correlation between weight and MSE (Figure 4B). Next, we calculated the cumulative PSSM enrichment by summing the sequence information (bits) of all cluster PSSMs per model. As expected, increasing the weight led to a corresponding increase in the cumulative sequence information (Figure 4C). We additionally observed that the clustering results generated by DDMC are noticeably different from those of eight standard clustering methods (Figure S4).

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

CellPress
OPEN ACCESS

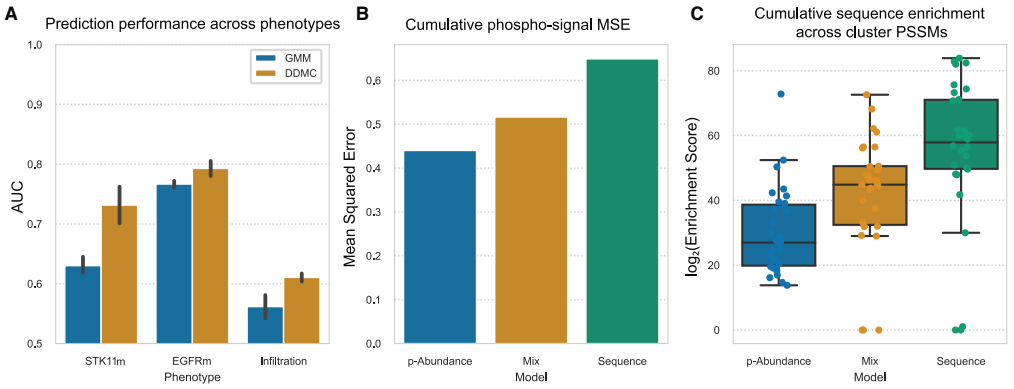


Figure 4. Sequence information enhances model prediction and provides more robust clustering

(A) Performance of a regression model predicting the mutational status of STK11 and EGFR, and the level of tumor infiltration in LUAD patients using either only phosphorylation abundance (weight =0), mainly sequence information (10^6), or both (0 <w % 10^6). Error bars indicate the standard error of the mean.

(B) MSE between the phosphorylation signal of 2,000 randomly selected peptides and the center of its assigned clusters using a weight of 0 (p-Abundance), 250 (Mix), or 10^6 (Sequence).

(C) Cumulative PSSM enrichment across positions comparing the p-Abundance, Mix, and Sequence clustering strategies. Error bars indicate the 95% confidence interval. The bottom and top of the box indicate the 25th and 75th percentiles. The line inside the box is the median.

See also Figure S3

We compared the classification performance of four regularized logistic regression models fit either the DDMC clusters, clusters generated by the standard methods GMM and k-means, or the raw phosphoproteomic data directly. It is worth noting that unlike DDMC, methods such as GMM, k-means, or direct regression cannot handle missing values. and thus for these strategies we used the 1,311 peptides that were observed in all samples, whereas DDMC was fit to the entire dataset comprising 30,561 phosphosites. In predicting STK11 mutational status, we found that DDMC fit to the fully observed 1,311 peptides yielded a moderately higher prediction performance than k-means, GMM, and DDMC fit to the entire dataset with missingness (Figure S3A). EGFR mutational status was noticeably better classified with both DDMC fittings (with and without missingness) than with k-means and GMM. Direct regression to the raw signaling data yielded excellent performance; however, this strategy assigns thousands of coefficients to different peptides that vary every time the model is run, rendering this approach unable to establish a consistent link between phenotypes and signaling (Figure S5D). These results show that using DDMC with a mixed weight that similarly prioritizes both information sources—peptide abundance and sequence—leads to more robust clustering of phosphosites through a tradeoff between phosphorylation abundance and sequence motifs.

Widespread, dramatic signaling differences exist between tumor and normal adjacent tissue

We explored whether DDMC could recognize conserved signaling patterns in tumors compared with NAT. The signaling difference between tumors and NAT samples was substantial, highlighting the significant signaling rewiring in tumor cells (Figure 5A). Using PCA, we could observe that NAT samples were more like one another than to each tumor sample (Figures 5B and 5C). Nearly every cluster was significantly different in its average abundance between tumor and NAT (Figures 5C and 5D). Not surprisingly given these enormous differences, samples could be almost perfectly classified using their phosphopeptide signatures, with or without DDMC (Figures 5E and S5A–S5C). However, directly classifying samples using the unclustered phosphoproteomic data and a regularized logistic regression model generates phosphosite weights that vary across runs. For instance, we saw that the associations of peptides MYH9: S1943-p, IFT140: S1443-p, and NCK1: Y105-p were selected in two runs but had an opposite association with sample status (Figure S5D). Using the DDMC clusters, a logistic regression model identified consistent associations between NAT versus tumor status and clusters 6, 15, and 20 (Figures 5E and 5F). With the abundance changes and regression results we observed, we further explored these three clusters.

Our DDMC results suggest that downregulation of NEKs and CLK2 promote cilia disassembly and migration in cancer cells, respectively, while CK1 activity correlates with tumor-specific signaling regulating cell cycle. Peptides in cluster 6, presumably targeted by NEK1&4, associate with hepatocyte growth factor (HGF) receptor signaling as well as cytoskeletal remodeling phenotypes (Figures 5G and S6A). Even though NEKs are fairly understudied, NEK1 has an established role in ciliagenesis and NEK4 is involved in regulating microtubule dynamics (Moniz et al., 2011; Meirelles et al., 2014). The absence of cilia in cancer cells promotes malignancy (Plotnikova et al., 2008; Fabbri et al., 2019), and NEK-regulated cluster 6 displays a

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

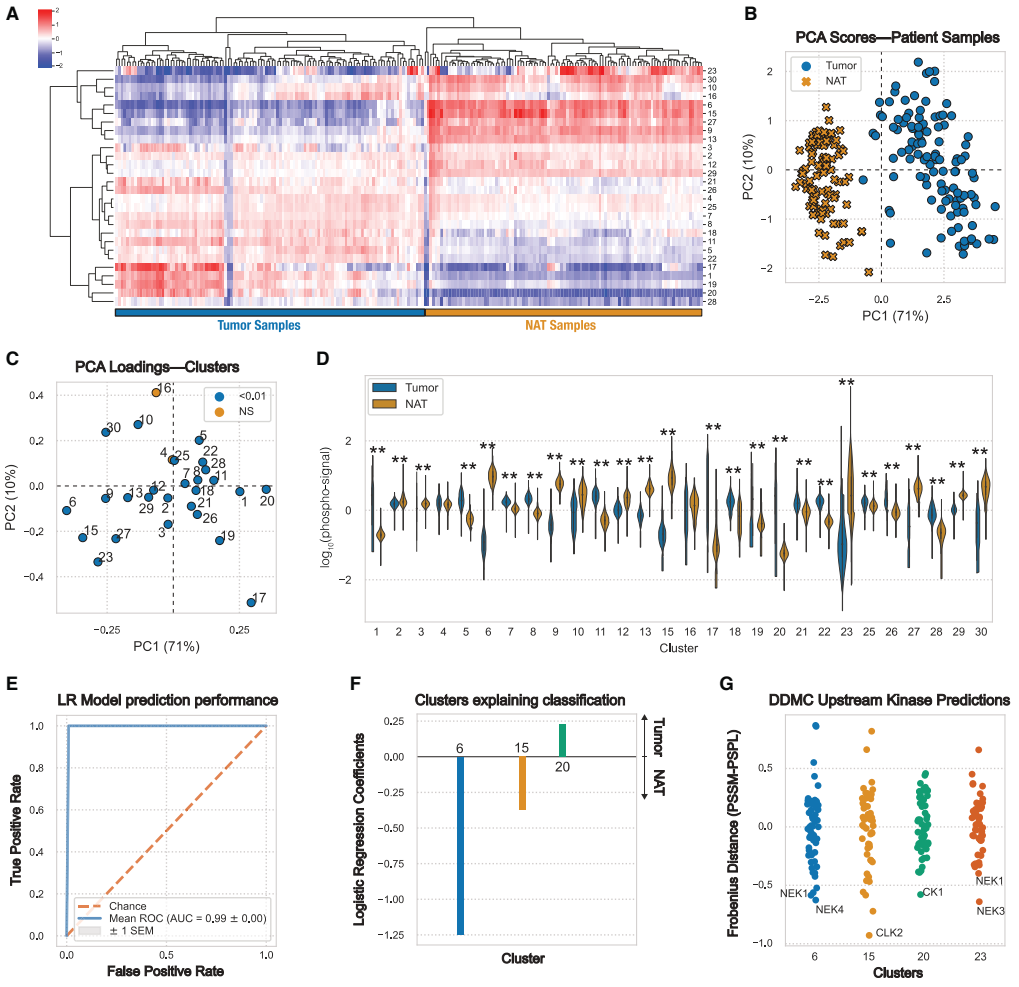


Figure 5. Conserved tumor differences compared with normal adjacent tissue

(A) Hierarchical clustering of the DDMC cluster centers.

(B and C) PCA scores (B) and loadings (C) of the samples and phosphopeptide clusters, respectively.

(D) Phosphorylation signal of tumor and NAT samples per cluster and statistical significance according to a Mann–Whitney U rank test (*p <0.05; **p <0.001).

(E) Receiver operating characteristic curve (ROC) of a regularizedlogistic regression model.

(F) Logistic regression weights per cluster.

(G) Upstream kinase predictions of clusters 6, 15, 20, and 23.

See also Figure S5

striking phosphorylation decrease in tumor samples compared with NATs, which might result in cilia disassembly. Interestingly, cluster 23, also downregulated in tumors, presents a motif favored by NEK1&3 and shows a marked enrichment of cilia–related processes (Figures 5D and S6A).

Similarly, cluster 15 is dramatically upregulated in NAT versus tumor samples, contributes toward correctly classifying NAT samples, and DDMC predicts CLK2 to be the most promising candidate for regulating its activation. CLK2 is a largely under–studied dual specificity kinase known to act as an RNA splicing

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

CellPress
OPEN ACCESS

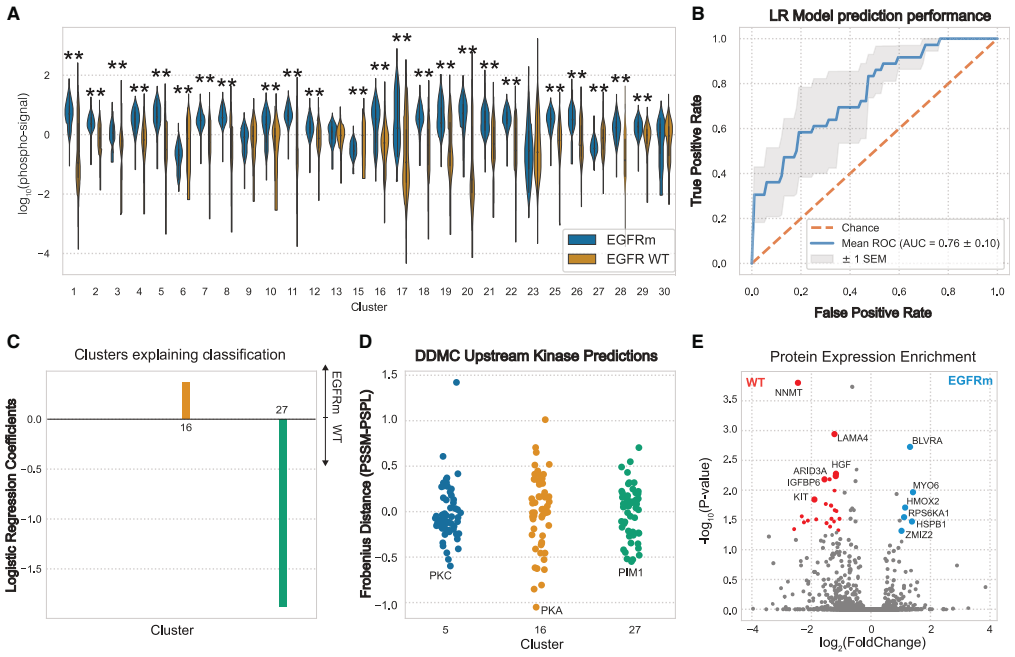


Figure 6. Phosphoproteomic aberrations associated with EGFR mutational status

(A) Phosphorylation signal of EGFR WT and mutant samples per cluster and statistical significance according to a Mann–Whitney U rank test (*p < 0.05; **p < 0.001).

(B and C) ROC of a logistic regression model predicting the EGFR mutational status and (C) its corresponding weights per sample type.

(D) Putative upstream kinases of clusters 5, 16, and 27.

(E) Volcano plot showing the differential protein expression between EGFR WT and mutant samples. Colored dots are statistically significant according to a Mann–Whitney U rank test (p values <0.05).

regulator. Gene set enrichment analysis (GSEA) indicates that integrin-mediated cell adhesion, cell junction assembly, and organization are the biological processes with highest enrichment scores (Figure S6A and S6B). These data are consistent with the observation that CLK2 downregulation enhanced cell migration and invasion and upregulated epithelial–to–mesenchymal transition (EMT)–related genes (Yoshida et al., 2015).

Conversely, the phosphorylation signal in cluster 20 is significantly higher in tumors compared with NATs and explains tumor–specific signaling that could be driven by CK1 (Figures 5D and 5F). CK1 has been identified to induce acquired resistance to the EGFR inhibitor erlotinib in several EGFR–mutant non–small cell lung cancer (NSCLC) cell lines (Lantermann et al., 2015). Taken together, DDMC builds phosphoproteomic clusters that present signaling dysregulation common to tumors compared with NATs and identifies putative upstream kinases modulating them.

Genetic driver mutations are associated with more targeted phosphoproteomic rewiring

Tyrosine kinase inhibitors (TKIs) targeting the receptor tyrosine kinase (RTK) EGFR are effective treatments in cancer patients with EGFRm. However, these treatments are limited by drug resistance, which in some cases is mediated by cell signaling rewiring that bypasses EGFR inhibition. Thus, we aimed to identify the phosphoproteomic aberrations triggered by mutant EGFR.

Most clusters were significantly altered on average, generally toward higher abundances with an EGFR mutation (Figure 6A). The cluster centers corresponding to each patient's tumor samples, excluding NATs, could successfully predict the EGFR mutational status by regularized logistic regression. We observed the largest statistically significant phosphorylation abundance increase in EGFRm samples with cluster 5 (Figure 6B). Moreover, the regression model identified clusters 16 and 27 to explain the signaling differences between EGFRm and wild–type (WT) samples, respectively (Figure 6C). DDMC identified PKC, PKA, and PIM1, respectively, as putative upstream kinases of clusters 5, 16, and 27 (Figure 6D). As elaborated below, our data suggest that EGFRm tumors might be regulated by two groups of proteins acting downstream of PKC and PKA, whereas PIM1 might support the signaling of EGFR WT tumors that are possibly driven by further RTKs.

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

In different EGFR–dependent tumors, PKC—putative regulator of cluster 5—has been shown to mediate receptor transactivation, induce mTOR signaling, and confer acquired resistance to EGFR inhibitors (Stewart and O'Brian, 2005; Fan et al., 2009; Salama et al., 2016; Chen et al., 2021). Enrichment analysis of the global protein expression data across all tumor samples showed that the heme degradation pathway enzymes BLVRA and HMOX2, as well as the mitogenic kinase RPS6KA1, among others, are significantly upregulated in EGFRm samples (Figure 6E). Consistent with the DDMC prediction, the kinase domains of RPS6KA1 and BVLRA are phosphorylated by PKC (Meshki et al., 2010; Miralem et al., 2012). GSEA shows an overrepresentation of the EGFR, human epidermal growth factor receptor (HER), and vascular endothelial growth factor receptor (VEGFR) signaling pathways in cluster 5, which might suggest crosstalk among the three RTKs' signaling (Figure S6A).

PKA, which might regulate cluster 16, is crucial for EMT, migration and invasion, and tumorigenesis (Shaikh et al., 2012; Coles et al., 2020). This kinase induces the activation of EGFR and its inhibition leads to a ligand–independent degradation of the receptor (Chen et al., 2002; Piiper et al., 2003; Oksvold et al., 2008; Feng et al., 2014). The EGFR and VEGFR signaling pathways are also enriched in cluster 16 alongside the ATM pathway (Figure S6A).

PIM1 might act upstream of cluster 27, which in turn is upregulated in EGFR WT tumor samples (Figures 6A, 6C, and 6D). PIM1 is an established oncogenic driver, and its inhibition was shown to re–sensitize cancer cells to radiotherapy as well as c–MET and ALK inhibition in NSCLC tumors (Kim et al., 2013; Cao et al., 2019; Trigg et al., 2019; Attili et al., 2020). Interestingly, the c–MET ligand HGF is more abundant in EGFR WT samples (Figure 6E). Moreover, activation of the KIT receptor, which can also mediate bypass resistance to targeted therapies and is enriched in EGFR WT samples, is reportedly regulated, at least in part, by PIM1 (An et al., 2016; Dziadziuszko et al., 2016; Ebeid et al., 2020) (Figures 6D and 6E). In total, our analysis identifies a consistent association between EGFR activity with established and previously unknown signaling mechanisms.

Finally, to show that DDMC can accurately predict other genotypes, we again used the signaling cluster centers with regularized logistic regression to classify the mutational status of STK11. Inactivating somatic mutations in the tumor suppressor STK11 leads to increased tumorigenesis and metastasis (Ji et al., 2007). This context is consistent with our results that clusters 9 (TLK1) and 11 (CK2) are associated with STK11m signaling, whereas clusters 16 (PKA) and 18 (CK1) are associated with WT samples (Figure S7).

Exploration of immune infiltration–associated signaling patterns in tumors

Immune checkpoint inhibitors (ICIs) have emerged as effective treatment options for NSCLC patients. However, there still is a need to identify or influence which patients will respond to these therapies. Patients who do not respond to ICIs often have tumors with poor immune infiltration either inherently or via an adaptive process after long exposure to the drug. However, the signaling mechanism by which malignant cells prevent tumor infiltration remains elusive. We used our DDMC clusters to explore the shared signaling patterns that differentiate HTE from CTE LUAD patients. HTE and CTE status per patient was determined using xCell (Aran et al., 2017; Gillette et al., 2020).

Only cluster 21 had a significantly different abundance between CTE and HTE samples (Figure 7A); however, infiltration status could still be accurately classified using combinations of the DDMC clusters. This predictive performance was mainly explained by a positive association of cluster 17 with HTE status and clusters 20 and 21 with CTE samples. Other clusters contributed to explain the signaling differences between both groups but to a lesser extent (Figure 7B). These results prompted us to further investigate clusters 17, 20, and 21, which our model inferred were regulated by CK2/TGFBR2, CK1, and ERK2, respectively (Figure 7C). We found that CK2 and TFGBR2 associate with the regulation of B cell homeostasis in HTE samples, whereas CK1 and ERK2 correlate with the activity of immunosuppressive regulatory T cells (Tregs) in CTE samples.

We performed GSEA on these three clusters using a compendium of gene sets associated with immunological signatures (Godec et al., 2016). Cluster 17 presents a marked enrichment of downregulated genes upon lipopolysaccharide stimulation, an upregulation of B cell– over eosinophil–specific genes, the enrichment of genes upregulated by an influenza vaccine, and genes upregulated in immunoglobulin(Ig)D+B cells. Thus, these might suggest that CK2 and TGFBR2 could regulate cluster 17 to direct B cell homeostasis. In line with this interpretation, a recent study found that CK2 knockout in B cells resulted in lower B cell receptor signaling, which perturbed B cell differentiation (Wei et al., 2021). Transforming growth factor (TGF)–b signaling is involved in several processes regulating B cell maturation. For instance, a study showed that IgD+ B cells were observed in the presence of TGF–b signaling, whereas genetic deletion of the receptor led to complete loss of IgD (Albright et al., 2019).

Consistent with their higher abundance in CTE samples and negative logistic regression coefficients, both cluster 20 and 21 showed enrichment of several phenotypes describing the induction of Tregs. ERK2 is known to modulate PD–L1 expression and its inhibition has been shown to improve anti–PD–L1 blockade in several cancer types, including NSCLC (Ng et al., 2018; Kumar et al., 2020; Henry et al., 2021; Luo et al., 2021). Conversely, while CK1 is associated with tumorigenesis, tumor growth, and drug resistance in cancer cells, its role in different immune cells and its ability to promote immune evasion has not been addressed. Overall, these data demonstrate that the presence or lack of tumor immune infiltration can be accurately predicted by the DDMC clusters, which in turn help identify putative upstream kinases modulating immune evasion.

DISCUSSION

Phosphorylation–based cell signaling through the coordinated activity of protein kinases allows cells to swiftly integrate environmental cues and orchestrate a myriad of biological processes. MS–based global phosphoproteomic data provide the unique opportunity to globally interrogate signaling networks to better understand cellular decision–making and its therapeutic implications. However, these data also present challenging issues because of their incomplete and stochastic coverage,

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

CellPress
OPEN ACCESS

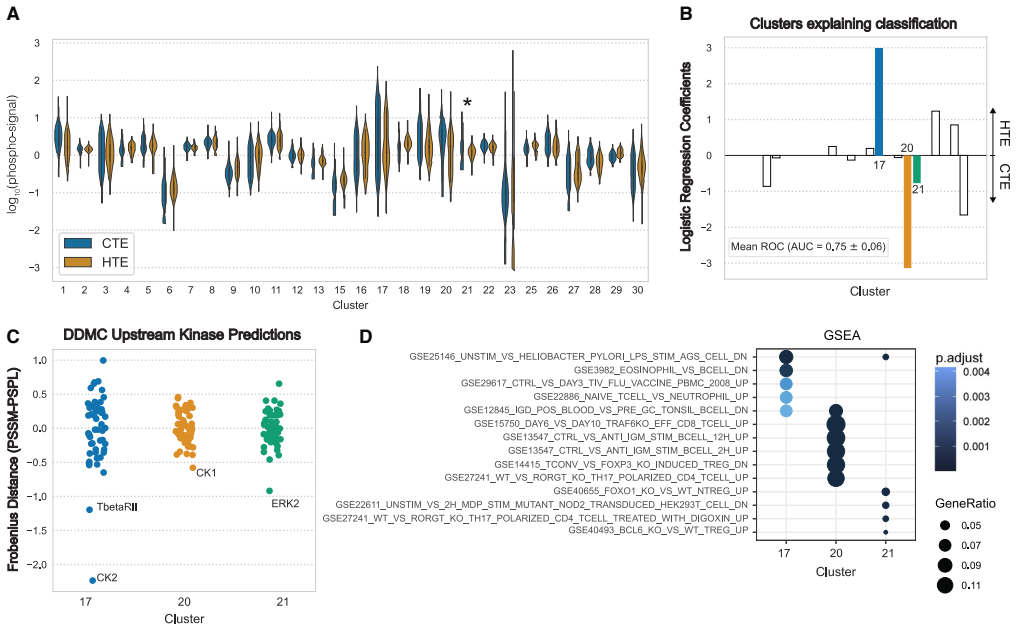


Figure 7. Phosphoproteomic signatures correlating with tumor immune infiltration

(A) Phosphorylation abundance of CTE and HTE samples per cluster and statistical significance according to a Mann–Whitney U rank test (*p <0.05; **p <0.001).

(B) Mean ROC and coefficients of a logistic regression model predicting infiltration status—cold-tumor enriched (CTE) versus hot-tumor enriched (HTE).

(C) Putative upstream kinases of clusters 17, 20, and 21.

(D) GSEA of immunological processes.

high-content but low-sample throughput, and variation in coverage across experiments. Here, we propose a clustering method, DDMC, that untangles the coordinated signaling changes by grouping phosphopeptides based on their phosphorylation behavior and sequence similarity (Figure 1). To test the utility of DDMC, we clustered the phosphoproteomes of LUAD patients and used the resulting groups of peptides to decipher signaling dysregulation associated with tumors, genetic backgrounds, and tumor infiltration status (Figures 5, 6, and 7).

Previous efforts in regressing MS-based phosphorylation measurements against phenotypic or clinical data have been based on the ability of certain regression models such as PLSR or LASSO to robustly predict using high-dimensional and correlated data (Kourou et al., 2015). While these models can generally be predictive with such data, they are not easily interpretable (Figure S5D). We hypothesized that clustering large-scale MS measurements based on biologically meaningful features and using the cluster centers could enhance the predictive performance of the model while providing highly interpretable results, wherein clusters constitute signaling nodes distinctly correlated with patient phenotypes. Here, we demonstrate that DDMC enhances model prediction and interpretation (Figures 3, 4A, and S3).

A key benefit of DDMC is that the identified clusters are not limited to pre-existing motifs and are therefore not dependent on prior experimentally validated kinase-substrate interactions. This method could therefore likely improve our understanding of the signaling effects of understudied kinases. For instance, our model predicts that NEKs promote, at least in part, a cluster with strikingly increased signaling in NATs compared with tumors. Further exploration of this cluster led us to hypothesize that the lack of NEK signaling in tumor samples might be associated with the absence of cilia in lung tumors (Figures 5G and S6A). In addition, we show that cluster 20 greatly contributes to explain a low immune infiltration status and might be regulated by the kinase CK1, which to our knowledge has not been studied in this context. While DDMC models the peptide sequence information without any constraints or assumptions defined by prior knowledge, the method could be easily adapted to populate clusters with the substrate motif information of specific upstream kinases. This "fixed" method could help improve granularity within a specific kinase signaling pathways.

An additional major challenge during the analysis of large-scale signaling data is missingness. Statistical tools often require complete datasets and, while researchers can use standard methods to impute missing values such as the peptides' mean

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

signal, imputation strategies generally work best when missing values only comprise a small fraction of the dataset (Chen et al., 2020; Deb et al., 2020; Gillette et al., 2020). In this study we show that DDMC can model a dataset of 30,561 peptides after filtering out any phosphosites that were not captured in more than one TMT run (up to 80% of missingness) by imputation during the expectation–maximization (EM) fitting process (see STAR Methods). Furthermore, DDMC clearly outperforms the imputation performance of using the peptides' mean or constant zero and provides similar results to PCA imputation (Figure 3). This important feature could offer the possibility of conducting pan-cancer phosphoproteomics studies using readily available large-scale clinical phosphoproteomic data by overcoming the fractional overlap in peptide coverage.

More generally, DDMC is tailored to model any biological datasets that combine a given signal with sequence information. In addition to TMT multiplex liquid chromatography–tandem MS datasets (as used here), this method may be equally useful with other techniques such as targeted MS via data-independent acquisition (Venable et al., 2004; Gillet et al., 2012). Beyond phosphoproteomics, DDMC can also be used to cluster transcription factor motifs or neoantigen sequences coupled with their gene or protein expression data. The benefit of building algorithms combining different information sources is evident in previously published approaches. For instance, INKA predicts active kinases by integrating scores reflecting both phosphorylation status and substrate abundance (Beekhof et al., 2019). A similar approach to that taken here could be applied with other generative algorithms, such as probabilistic PCA or probabilistic generative adversarial networks, with similar benefits. Integrating yet other information may reveal further improvements in the dimensionality reduction and interpretation of other high-throughput molecular measurements.

In total, we show that combining the information about the sequence features and phosphorylation abundance leads to more robust clustering of global signaling measurements. Use of the DDMC clusters to regress against cell phenotypes led to enhanced model predictions and interpretation. Thus, we propose DDMC as a general and flexible strategy for phosphoproteomic analysis.

Limitations of this study

Our present analysis is limited to a single clinical phosphoproteomics dataset. Examining other datasets, and integrating phosphoproteomics measurements with other molecular measurement modalities, will reveal new insights and other ways to improve the model. For instance, it remains unclear how DDMC might perform with smaller cohorts or with measurements across different cancer types.

DDMC interpretation is enhanced by comparing the resulting cluster PSSMs with kinase specificity data such as PSPL to identify putative upstream kinases for each cluster. Validation experiments showed that DDMC was able to correctly associate AKT1 and ERK2 with clusters containing their respective substrates (Figure 3). Kinase specificity is defined by additional features beyond the phosphosite motif, however, such as kinase–substrate co-localization, regulation by phosphosite-binding domains (e.g., SH2, PTB domains), or docking. These other kinase regulatory processes could compromise kinase–cluster associations established by DDMC. Refined methods of quantifying kinase specificity, alongside adjustments to DDMC to account for these other regulatory processes, could improve both upstream kinase predictions and the resulting peptide clustering (Shah et al., 2018).

STAR+METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Expectation–maximization (EM) algorithm architecture
  - Phosphorylation site abundance clustering in the presence of missing values
  - Sequence–cluster comparison
  - Quantifying the influence of sequence versus data
  - Generating cluster motifs and upstream kinase predictions
  - Evaluate clustering by imputation of values
  - Associating clusters with molecular and clinical features
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2022.100167.

ACKNOWLEDGMENTS

This work was supported by NIH U01-CA215709 to A.S.M. and in part by the UCLA Jonsson Comprehensive Cancer Center (JCCC) grant NIH P30-CA016042.

AUTHOR CONTRIBUTIONS

Conceptualization, A.S.M.; methodology, M.C. and A.S.M.; investigation, M.C. and A.S.M.; writing, M.C. and A.S.M.; funding acquisition, A.S.M.; resources; A.S.M.; supervision; A.S.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 15, 2021
Revised: November 28, 2021
Accepted: January 24, 2022
Published: February 14, 2022

SUPPORTING CITATIONS

The following references appear in the supplemental information: Comaniciu and Meer (2002), Frey and Dueck (2007), Knyazev (2001), and Zhang et al. (1996).

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

CellPress
OPEN ACCESS

REFERENCES

Albright, A.R., Kabat, J., Li, M., Raso, F., Reboldi, A., and Muppidi, J.R. (2019). TGFb signaling in germinal center B cells promotes the transition from light zone to dark zone. J. Exp. Med. 216, 2531–2545. https://doi.org/10.1084/jem.20181868.

Alexander, J., Lim, D., Joughin, B.A., Hegemann, B., Hutchins, J.R., Ehrenberger, T., Ivins, F., Sessa, F., Hudecz, O., Nigg, E.A., et al. (2011). Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling. Sci. Signal. 4, ra42. https://doi.org/10.1126/scisignal.2001796.

An, N., Cen, B., Cai, H., Song, J.H., Kraft, A., and Kang, Y. (2016). Pim1 kinase regulates c-Kit gene translation. Exp. Hematol. Oncol. 5, 31–38. https://doi.org/10.1186/s40164-016-0060-3.

Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 18, 220. https://doi.org/10.1186/s13059-017-1349-1.

Attili, I., Bonanno, L., Karachaliou, N., Bracht, J.W.P., Berenguer, J., Codony-Servat, C., Codony-Servat, J., Aldeguer, E., Gimenez-Capitan, A., Dal Maso, A., et al. (2020). SRC and PIM1 as potential co-targets to overcome resistance in MET deregulated non-small cell lung cancer. Transl. Lung Cancer Res. 9, 1810–1821. https://doi.org/10.21037/tlcr-20-681.

Beekhof, R., van Alphen, C., Henneman, A.A., Knol, J.C., Pham, T.V., Rolfs, F., Labots, M., Henneberry, E., Le Large, T.Y., de Haas, R.R., et al. (2019). INKA, an integrative data analysis pipeline for phosphoproteomic inference of active kinases. Mol. Syst. Biol. 15, e8250. https://doi.org/10.15252/msb.20188250.

Begley, M.J., Yun, C.H., Gewinner, C.A., Asara, J.M., Johnson, J.L., Coyle, A.J., Eck, M.J., Apostolou, I., and Cantley, L.C. (2015). EGF-receptor specificity for phosphotyrosine-primed substrates provides signal integration with Src. Nat. Struct. Mol. Biol. 22, 983–990. https://doi.org/10.1038/nsmb.3117.

Cao, L., Wang, F., Li, S., Wang, X., Huang, D., and Jiang, R. (2019). PIM1 kinase promotes cell proliferation, metastasis and tumor growth of lung adenocarcinoma by potentiating the c-MET signaling pathway. Cancer Lett. 444, 116–126. https://doi.org/10.1016/j.canlet.2018.12.015.

Carlson, S.M., Chouinard, C.R., Labadorf, A., Lam, C.J., Schmelzle, K., Fraenkel, E., and White, F.M. (2011). Large-scale discovery of ERK2 substrates identifies ERK-mediated transcriptional regulation by ETV3. www.SCIENCESIGNALING.org.

Casado, P., Rodriguez-Prados, J.C., Cosulich, S.C., Guichard, S., Vanhaesebroeck, B., Joel, S., and Cutillas, P.R. (2013). Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. www.SCIENCESIGNALING.org.

Chen, C.-H., Wang, B.-W., Hsiao, Y.-C., Wu, C.-Y., Cheng, F.-J., Hsia, T.-C., Chen, C.-Y., Wang, Y., Weihua, Z., Chou, R.-H., et al. (2021). PKCd-mediated SGLT1 upregulation confers the acquired resistance of NSCLC to EGFR TKIs. Oncogene 40, 4796–4808. https://doi.org/10.1038/s41388-021-01889-0.

Chen, Y., and Gonzá lez, A. (2002). Novel mechanism for regulation of epidermal growth factor receptor endocytosis revealed by protein kinase A inhibition. Mol. Biol. Cell 13, 1227–1237. https://doi.org/10.1091/mbc.01.

Chen, Y.J., Roumeliotis, T.I., Chang, Y.H., Chen, C.T., Han, C.L., Lin, M.H., Chen, H.W., Chang, G.C., Chang, Y.L., Wu, C.T., et al. (2020). Proteogenomics of non-smoking lung cancer in East Asia delineates molecular signatures of pathogenesis and progression. Cell 182, 226–244.e17. https://doi.org/10.1016/j.cell.2020.06.012.

Coles, G.L., Cristea, S., Webber, J.T., Levin, R.S., Moss, S.M., He, A., Sangodkar, J., Hwang, Y.C., Arand, J., Drainas, A.P., et al. (2020). Unbiased proteomic profiling uncovers a targetable GNAS/PKA/PP2A Axis in small cell lung cancer stem cells. Cancer Cell 38, 129–143.e7. https://doi.org/10.1016/j.ccell.2020.05.003.

Comaniciu, D., and Meer, P. (2002). Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24, 603–619. https://doi.org/10.1109/34.1000236.

Deb, B., Sengupta, P., Sambath, J., and Kumar, P. (2020). Bioinformatics analysis of global proteomic and phosphoproteomic data sets revealed activation of NEK2 and AURKA in cancers. Biomolecules 10, 237. https://doi.org/10.3390/biom10020237.

Dziadziuszko, R., Le, A.T., Wrona, A., Jassem, J., Camidge, D.R., Varella-Garcia, M., Aisner, D.L., and Doebele, R.C. (2016). An activating KIT mutation induces crizotinib resistance in ROS1-positive lung cancer. J. Thorac. Oncol. 11, 1273–1281. https://doi.org/10.1016/j.jtho.2016.04.001.

Ebeid, D.E., Firouzi, F., Esquer, C.Y., Navarrete, J.M., Wang, B.J., Gude, N.A., and Sussman, M.A. (2020). PIM1 promotes survival of cardiomyocytes by upregulating c-Kit protein expression. Cells 9, 11–13. https://doi.org/10.3390/cells9092001.

Fabbri, L., Bost, F., and Mazure, N.M. (2019). Primary cilium in cancer hallmarks. Int. J. Mol. Sci. 20, 1336. https://doi.org/10.3390/ijms20061336.

Fan, Q.W., Cheng, C., Knight, Z.A., Haas-Kogan, D., Stokoe, D., James, C.D., McCormick, F., Shokat, K.M., and Weiss, W.A. (2009). EGFR signals to mTOR through PKC and independently of Akt in glioma. Sci. Signal. 2, ra4. https://doi.org/10.1126/scisignal.260er4.

Feng, H., Hu, B., Vuori, K., Sarkaria, J.N., Furnari, F.B., Cavenee, W.K., and Cheng, S.Y. (2014). EGFRvIII stimulates glioma growth and invasion through PKA-dependent serine phosphorylation of Dock180. Oncogene 33, 2504–2512. https://doi.org/10.1038/onc.2013.198.

Frey, B.J., and Dueck, D. (2007). Clustering by passing messages between data points. Science 315, 972–976. https://doi.org/10.1126/science.1136800.

Gillet, L.C., Navarro, P., Tate, S., Rö st, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell Proteomics 11, O111.O016717. https://doi.org/10.1074/mcp.O111.016717.

Gillette, M.A., Satpathy, S., Cao, S., Dhanasekaran, S.M., Vasaikar, S.V., Krug, K., Petralia, F., Li, Y., Liang, W.W., Reva, B., et al. (2020). Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. Cell 182, 200–225.e35. https://doi.org/10.1016/j.cell.2020.06.013.

Godec, J., Tan, Y., Liberzon, A., Tamayo, P., Bhattacharya, S., Butte, A.J., Mesirov, J.P., and Haining, W.N. (2016). Compendium of immune signatures identifies conserved and species-specific biology in response to inflammation. Immunity 44, 194–206. https://doi.org/10.1016/j.immuni.2015.12.006.

Henry, K.E., Mack, K.N., Nagle, V.L., Cornejo, M., Michel, A.O., Fox, I.L., Davydova, M., Dilling, T.R., Pillarsetty, N., and Lewis, J.S. (2021). ERK inhibition improves anti-PD-L1 immune checkpoint blockade in preclinical pancreatic ductal adenocarcinoma. Mol. Cancer Ther. 20, 2026–2034. https://doi.org/10.1158/1535-7163.mct-20-1112.

Hijazi, M., Smith, R., Rajeeve, V., Bessant, C., and Cutillas, P.R. (2020). Reconstructing kinase network topologies from phosphoproteomics data reveals cancer-associated rewiring. Nat. Biotechnol. 38, 493–502. https://doi.org/10.1038/s41587-019-0391-9.

Horn, H., Schoof, E.M., Kim, J., Robin, X., Miller, M.L., Diella, F., Palma, A., Cesareni, G., Jensen, L.J., and Linding, R. (2014). KinomeXplorer: an integrated platform for kinome biology studies. Nature Methods 11, 603–604.

Hornbeck, P.V., Kornhauser, J.M., Latham, V., Murray, B., Nandhikonda, V., Nord, A., Skrzypek, E., Wheeler, T., Zhang, B., and Gnad, F. (2019). 15 years of PhosphoSitePlus : integrating post-translationally modified sites, disease variants and isoforms. Nucleic Acids Res. 47, D433–D441. https://doi.org/10.1093/nar/gky1159.

Hunter, T. (1995). Protein kinases and phosphatases: the Yin and Yang of protein phosphorylation and signaling. Cell 80, 225–236. https://doi.org/10.1016/0092-8674(95)90405-0.

Hutti, J.E., Jarrell, E.T., Chang, J.D., Abbott, D.W., Storz, P., Toker, A., Cantley, L.C., and Turk, B.E. (2004). A rapid method for determining protein kinase phosphorylation specificity. Nat. Methods 1, 27–29. https://doi.org/10.1038/nmeth708.

Ji, H., Ramsey, M.R., Hayes, D.N., Fan, C., McNamara, K., Kozlowski, P., Torrice, C., Wu, M.C., Shimamura, T., Perera, S.A., et al. (2007). LKB1 modulates

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

lung cancer differentiation and metastasis. Nature 448, 807–810. https://doi.org/10.1038/nature06030.

Kim, W., Youn, H., Kwon, T., Kang, J., Kim, E., Son, B., Yang, H.J., Jung, Y., and Youn, B. (2013). PIM1 kinase inhibitors induce radiosensitization in non-small cell lung cancer cells. Pharmacol. Res. 70, 90–101. https://doi.org/10.1016/j.phrs.2013.01.005.

Kinney, J.B. (2019).Logomaker: beautifulsequence logos in Python. Bioinformatics 36, 4–6.

Knyazev, A.V. (2001). Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method. SIAM J. Sci. Comput. 2, 158–178.

Kourou, K., Exarchos, T.P., Exarchos, K.P., Karamouzis, M.V., and Fotiadis, D.I. (2015).Machine learning applications in cancer prognosis and prediction. Comput. Struct. Biotechnol. J. 13, 8–17. https://doi.org/10.1016/j.csbj.2014.11.005.

Kumar, S., Principe, D.R., Singh, S.K., Viswakarma, N., Sondarva, G., Rana, B., and Rana, A. (2020). Mitogen-activated protein kinase inhibitors and T-cell-dependent immunotherapy in cancer. Pharmaceuticals 13, 1–11. https://doi.org/10.3390/ph13010009.

Lantermann, A.B., Chen, D., McCutcheon, K., Hoffman, G., Frias, E., Ruddy, D., Rakiec, D., Korn, J., McAllister, G., Stegmeier, F., et al. (2015). Inhibition of casein kinase 1 alpha prevents acquired drug resistance to erlotinib in EGFR-mutant non-small cell lung cancer. Cancer Res. 75, 4937–4948. https://doi.org/10.1158/0008-5472.CAN-15-1113.

Linding, R., Jensen, L.J., Ostheimer, G.J., van Vugt, M.A., Jørgensen, C., Miron, I.M., Diella, F., Colwill, K., Taylor, L., Elder, K., et al. (2007).Systematic discovery of in vivo phosphorylation networks. Cell 129, 1415–1426. https://doi.org/10.1016/j.cell.2007.05.052.

Luo, M., Xia, Y., Wang, F., Zhang, H., Su, D., Su, C., Yang, C., Wu, S., An, S., Lin, S., and Fu, L. (2021).PD0325901,an ERK inhibitor,enhances the efficacy of PD-1 inhibitorin non-smallcell lungcarcinoma. Acta Pharm. Sin B 11, 3120–3133. https://doi.org/10.1016/j.apsb.2021.03.010.

Meirelles,G.V., Perez, A.M., de Souza, E.E., Basei, F.L., Papa, P.F., Melo Hanchuk, T.D., Cardoso, V.B., and Kobarg, J. (2014).Stop Ne(c)king around'': how interactomics contributes to functionally characterize Nek family kinases. World J. Biol. Chem. 5, 141–160. https://doi.org/10.4331/wjbc.v5.i2.141.

Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., et al. (2016).Proteogenomics connects somatic mutations to signalling in breast cancer. Nature 534, 55–62. https://doi.org/10.1038/nature18003.

Meshki, J., Caino, M.C., von Burstin, V.A., Griner, E., and Kazanietz, M.G. (2010).Regulation of prostate cancer cell survival by proteinkinase Cinvolves bad phosphorylationandmodulationofthe TNFa/JNK pathway. J. Biol. Chem. 285, 26033–26040. https://doi.org/10.1074/jbc.M110.128371.

Miller, M.L., Jensen, L.J., Diella, F., Jørgensen, C., Tinti, M., Li, L., Hsiung, M., Parker, S.A., Bordeaux, J., Sicheritz-Ponten, T., et al. (2008).Linear motifatlas for phosphorylation-dependent signaling. Sci. Signal. 1, ra2–12. https://doi.org/10.1126/scisignal.1159433.

Miralem, T., Lerner-Marmarosh,N., Gibbs, P.E., Tudor, C., Hagen, F.K., and Maines, M.D. (2012). The human biliverdin reductase-based peptide fragments and biliverdin regulate proteinkinase Cdactivity: the peptides are inhibitors or substrate for the protein kinase C. J. Biol. Chem. 287, 24698–24712. https://doi.org/10.1074/jbc.M111.326504.

Moniz, L., Dutt, P., Haider, N., and Stambolic, V. (2011).Nek family of kinases in cell cycle, checkpoint control and cancer. Cell Div. 6, 18. https://doi.org/10.1186/1747-1028-6-18.

Needham, E.J., Parker, B.L., Burykin, T., James, D.E., and Humphrey, S.J. (2019). Illuminating the dark phosphoproteome. Sci. Signal. 12, eaau8645. https://doi.org/10.1126/scisignal.aau8645.

Ng, H.Y., Li, J., Tao, L., Lam, A.K., Chan, K.W., Ko, J.M.Y., Yu, V.Z., Wong, M., Li, B., and Lung, M.L. (2018).Chemotherapeutic treatments increase PD-L1 expression in esophageal squamous cell carcinoma throughEGFR/ERK activation. Transl. Oncol. 11, 1323–1333. https://doi.org/10.1016/j.tranon.2018.08.005.

Obata, T., Yaffe, M.B., Leparc, G.G., Piro, E.T., Maegawa, H., Kashiwagi, A., Kikkawa, R., and Cantley, L.C. (2000). Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. J. Biol. Chem. 275, 36108–36115. https://doi.org/10.1074/jbc.M005497200.

Obenauer, J.C., Cantley, L.C., and Yaffe, M.B. (2003).Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res. 31, 3635–3641. https://doi.org/10.1093/nar/gkg584.

Oksvold, M.P., Funderud, A., Kvissel, A.K., Skarpen, E., Henanger, H., Huitfeldt,H.S., Skå lhegg,B.S., and Ørstavik, S. (2008).Epidermal growth factor receptor levels are reduced in mice with targeted disruption of the protein kinase A catalytic subunit. BMC Cell Biol. 9, 16–19. https://doi.org/10.1186/1471-2121-9-16.

Piiper, A., Lutz, M.P., Cramer, H., Elez, R., Kronenberger, B., Dikic, I., Müller-Esterl, W., and Zeuzem, S. (2003). Protein kinase A mediates cAMP-induced tyrosine phosphorylation of the epidermal growth factor receptor. Biochem. Biophys. Res. Commun. 301, 848–854. https://doi.org/10.1016/S0006-291X(03)00055-X.

Plotnikova, O.V., Golemis, E.A., and Pugacheva, E.N. (2008). Cell cycle-dependent ciliogenesis and cancer. Cancer Res. 68, 2058–2061.https://doi.org/10.1158/0008-5472.CAN-07-5838.

Salama, M.F., Liu, M., Clarke, C.J., Espaillat, M.P., Haley, J.D., Jin, T., Wang, D., Obeid, L.M., and Hannun, Y.A. (2016). PKCa is required for Akt-mTORC1 activation in non-small cell lung carcinoma (NSCLC) with EGFR mutation. Physiol. Behav. 176, 100–106. https://doi.org/10.1038/s41388-019-0950-z. PKC.

Schwartz, D., and Gygi, S.P. (2005). An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat. Biotechnol. 23, 1391–1398. https://doi.org/10.1038/nbt1146.

Shah, N.H., Löbel, M., Weiss, A., and Kuriyan, J. (2018). Fine-tuning of substrate preferences of the Src-family kinase Lck revealed through a high-throughput specificity screen. eLife 7, 1–24. https://doi.org/10.7554/elife.35190.

Shaikh, D., Zhou, Q., Chen, T., Ibe, J.C., Raj, J.U., and Zhou, G. (2012).CAMP-dependent protein kinase is essential for hypoxia-mediated epithelial-mesenchymal transition, migration, and invasion in lung cancer cells. Cell Signal 24, 2396–2406. https://doi.org/10.1016/j.cellsig.2012.08.007.

Mok, J., Kim, P.M., Lam, H.Y.K., Piccirillo, S., Zhou, X., Jeschke, G.R., Sheridan, D.L., Parker, S.A., Desai, V., Jwa, M., et al. (2010). Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. Sci. Signal. 3, ra12. https://doi.org/10.1126/scisignal.2000482.

Stewart, J.R., and O'Brian, C.A. (2005). Protein kinase C-{alpha} mediates epidermal growth factor receptor transactivation in human prostate cancer cells. Mol. Cancer Ther. 4, 726–732. https://doi.org/10.1158/1535-7163.MCT-05-0013.

Tabb, D.L., Vega-Montoto,L., Rudnick, P.A., Variyath, A.M., Ham, A.J., Bunk, D.M., Kilpatrick, L.E., Billheimer, D.D., Blackman, R.K., Cardasis, H.L., et al. (2010). Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. J. Proteome Res. 9, 761–776. https://doi.org/10.1021/pr9006365.

Tan, C.S., Pasculescu, A., Lim, W.A., Pawson, T., Bader, G.D., and Linding,R. (2009). Positive selection of tyrosine loss in metazoan evolution. Science 325, 1686–1688. https://doi.org/10.1126/science.1174301.

Trigg,R.M., Lee, L.C., Prokoph, N., Jahangiri, L., Reynolds, C.P., Amos Burke, G.A., Probst, N.A., Han, M., Matthews, J.D., Lim, H.K., et al. (2019).The targetable kinase PIM1 drives ALK inhibitorresistance in high-riskneuroblastomain-dependent of MYCN status. Nat. Commun. 10, 5428. https://doi.org/10.1038/s41467-019-13315-x.

van de Kooij, B., Creixell, P., van Vlimmeren, A., Joughin, B.A., Miller, C.J., Haider, N., Simpson, C.D., Linding, R., Stambolic, V., Turk, B.E., and Yaffe, M.B. (2019). Comprehensive substrate specificity profiling of the human nek

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

kinome reveals unexpected signaling outputs. eLife 8, e44635.https://doi.org/10.7554/eLife.44635.

Venable, J.D., Dong, M.Q., Wohlschlegel, J., Dillin, A., and Yates, J.R. (2004). Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nat. Methods 1, 39–45. https://doi.org/10.1038/nmeth705.

Wei, H., Yang, W., Hong, H., Yan, Z., Qin, H., and Benveniste, E.N. (2021). Protein kinase CK2 regulates B cell development and differentiation. J. Immunol. 207, 799–808. https://doi.org/10.4049/jimmunol.2100059.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation 2, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

Yaffe, M.B. (2019). Why geneticists stole cancer research even though cancer is primarily a signaling disease. Sci. Signal. 12, eaaw3483. https://doi.org/10.1126/scisignal.aaw3483.

Yoshida, T., Kim, J.H., Carver, K., Su, Y., Weremowicz, S., Mulvey, L., Yamamoto, S., Brennan, C., Mei, S., Long, H., et al. (2015). CLK2 is an oncogenic kinase and splicing regulator in breast cancer. Cancer Res. 75, 1516–1526. https://doi.org/10.1158/0008-5472.CAN-14-2443.

Zarrinpar, A., Park, S.H., and Lim, W.A. (2003). Optimization of specificity in a cellular protein interaction network by negative selection. Nature 426, 676–680. https://doi.org/10.1038/nature02178.

Zhang, T., Ramakrishnan, R., and Livny, M. (1996). Birch. SIGMOD Rec. 25, 103–114. https://doi.org/10.1145/235968.233324.

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

STAR+METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
LUAD phosphoproteomics, proteomics, and clinical data	Gillette et al., 2020	https://cptac-data-portal.georgetown.edu/study-summary/S056
Upstream kinase PSPLs	Begley et al., 2015; Horn et al., 2014; Miller et al., 2008; Obata et al., 2000; van de Kooij et al., 2019	https://netphorest.info/download.shtml
Software and algorithms		
Python v3.9	Python Software Foundation	https://python.org/
R	The R Foundation	https://r-project.org/
NetPhorest	Horn et al., 2014	https://netphorest.info/download.shtml
Bioinfokit 0.3	NA	https://pypi.org/project/bioinfokit/0.3/
clusterProfiler 4.2	Wu et al., 2021	https://guangchuangyu.github.io/software/clusterProfiler/
DDMC	This paper	https://doi.org/10.5281/zenodo.5856274
fancyimpute v0.5.5	NA	https://github.com/iskandr/fancyimpute

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Aaron Meyer (ameyer@asmlab.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- No new standardized datasets were generated by this study.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. The DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Expectation–maximization (EM) algorithm architecture

We constructed a modified mixture model that clusters peptides based on both their abundance across conditions and sequence. The model is defined by a given number of clusters and weighting factor to prioritize either the data or the sequence information. Fitting was performed using expectation–maximization, initialized at a starting point. The starting point was derived from k–means clustering the abundance data after missing values were imputed. During the expectation (E) step, the algorithm calculates the probability of each peptide being assigned to each cluster. In the maximization (M) step, each cluster's distributions are fit using the weighted cluster assignments. The peptide sequence and abundance assignments within the E step are combined by taking the sum of the log-likelihood of both assignments. The peptide log-likelihood is multiplied by the user-defined weighting factor immediately before to influence its importance. Both steps repeat until convergence as defined by the increase in model log-likelihood between iterations falling below a user-defined threshold.

Phosphorylation site abundance clustering in the presence of missing values

We modeled the log-transformed abundance of each phosphopeptide as following a multivariate Gaussian distribution with diagonal co-variance matrix. Each dimension of this distribution represents the abundance of that peptide within a given sample. For example, within a data set of 100 patients and 1000 peptides, using 10 clusters, the data is represented by 10 Gaussian distributions of 100

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

dimensions. Unobserved/missing values were initially indicated as NaN and subsequently imputed using the method SoftImpute (using the package fancyimpute) upon model initialization. During every iteration of the EM algorithm, the missing values are then updated according to the current model. Any peptides that were detected in only one TMT experiment were discarded.

Sequence–cluster comparison

PAM250

During model initialization, the pairwise distance between all peptides in the dataset was calculated using the PAM250 matrix. The mean distance from each peptide to a given cluster could then be calculated by:

$$w = \frac{1}{n}\delta P, v\flat$$

where P is the n 3 n distance matrix, n is the number of peptides in the dataset, v is the probability of each peptide being assigned to the cluster of interest, and w is the log–probabilities of cluster assignment.

Binomial enrichment

We alternatively used a binomial enrichment model for the sequence representation of a cluster based on earlier work (Schwartz and Gygi, 2005). Upon model initialization, a background matrix i 3 j 3 k was created with a position–specific scoring matrix of all the sequences together. Next, a data tensor TT was created where i is the number of peptides, j is the number of amino acid possibilities, and k is the position relative to the phosphorylation site. This tensor contained 1 where an amino acid was present for that position and peptide, and 0 elsewhere.

Within each iteration, the cluster motif would be updated using v, the probability of each peptide being assigned to the cluster of interest. First, a weighted count for each amino acid and position would be assembled:

$$k = \delta T^{u}, v\flat^{u}$$

Because peptides can be partially assigned to a cluster, the counts of each amino acid and position can take continuous values. We therefore generalized the binomial distribution to allow continuous values using the regularized incomplete Beta function:

$$M = B\delta k\acute{v}\, k_1\, k;\, k + 1;1\, G\flat$$

Finally, the log–probability of membership for each peptide was calculated based on the product of each amino acid–position probability.

$$w = \log\delta T\ 3\ M\flat$$

We confirmed that this provided identical results to a binomial enrichment model for integer counts of amino acids but allowed for partial assignment of peptides to clusters.

Quantifying the influence of sequence versus data

The magnitude of the weight used to scale the sequence and data scores is arbitrary. We do know that with a weight of 0 the model only uses the phosphorylation measurements. Alternatively, with an enormously large weight the motif information is prioritized. However, we do not know to what extent each information source is prioritized in general. Therefore, to quantify the relative importance of each type of data, we calculated our clustering results at each weighting extreme, and then calculated the Frobenius norm of the resulting peptide assignments between those and the clustering of interest.

Generating cluster motifs and upstream kinase predictions

For each cluster we computed a position–specific–scoring matrix (PSSM). To do so, we populated a residue/position matrix with the sum of the corresponding cluster probabilities for every peptide. Once all peptides were accounted for, the resulting matrix was normalized by averaging the mean probability across amino acids and log2–transforming to generate a PSSM. In parallel, we computed a PSSM including all sequences that served as background to account for the different amino acid occurrences within the data set. Then, we subtracted each cluster PSSM with the background PSSM to generate the final enrichment scores. Positive scores represent enriched residues while negative scores represent depleted amino acids across positions. Next, we extracted several kinase specificity profiling results (PSPL) from the literature (Miller et al., 2008; Alexander et al., 2011; Begley et al., 2015; van de Kooij et al., 2019). The distance between each cluster PSSM and kinase PSPL motif was calculated using by the Frobenius norm of the difference between both matrices, considering only positive enrichment scores. Motif logo plots were generated using logomaker (Kinney, 2019).

Evaluate clustering by imputation of values

To evaluate the ability of our model to handle missing values, we removed random, individual TMT experiments for each peptide and used the model to impute these values. We then computed the mean squared error between the actual values and predictions made by each method. We calculated the reconstruction error across different combinations of cluster numbers and weights using the same process.

Please cite this article in press as: Creixell and Meyer, Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data, Cell Reports Methods (2022), https://doi.org/10.1016/j.crmeth.2022.100167

Associating clusters with molecular and clinical features

To find clusters that tracked with specific molecular or clinical features we implemented two different strategies: logistic regression and hypothesis testing. For binary problems such as tumor vs NAT samples or mutational status we used l1-regularized logistic regression and the Mann–Whitney U rank test. In the former, we tried to predict the feature of interest using the phosphorylation signal of the cluster centers, whereas in the latter, for each cluster we split all patients according to their specific feature and tested whether the difference in the median signal between both groups was statistically different. We performed Bonferroni correction on the p-values computed by the Mann–Whitney U rank test. GSEA analysis was performed using clusterProfiler (4.0.2) implemented in R. The enrichment method used was ''enrichWP'' or ''enrichGO'' (WikiPathway or GeneOntology gene sets) with the p-value adjustment method was set to Bonferroni (Wu et al., 2021).

QUANTIFICATION AND STATISTICAL ANALYSIS

All the statistical and quantification descriptions of each analysis can be found in the corresponding figure legends and results sections. The statistical enrichment of phosphorylation abundance between different binary phenotypes (tumor vs NAT, mutation vs WT, or HTE vs CTE) was calculated using the Mann–Whitney U rank test, with each subjects tumor treated as an independent observation ($N = 110$). The test results were adjusted for multiple tests via Bonferroni's correction method. ''*'' means that the p-value is lower than 0.05 but higher than 0.001 and ''**'' that it is lower than 0.001. The volcano plot showing up- and down-regulated proteins in EGFR mutant vs WT samples was generated after calculating their log2 fold-change and p-values according to a Mann–Whitney U rank test using Bonferroni's correction for multiple tests. Biokit v.2.0.8 was used to generate the volcano plot using the default log fold change and p-value cutoffs set to 1.0 and 0.05, respectively.

Cell Reports Methods, Volume 2

Supplemental information

Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data

Marc Creixell and Aaron S. Meyer

SUPPLEMENTAL INFORMATION

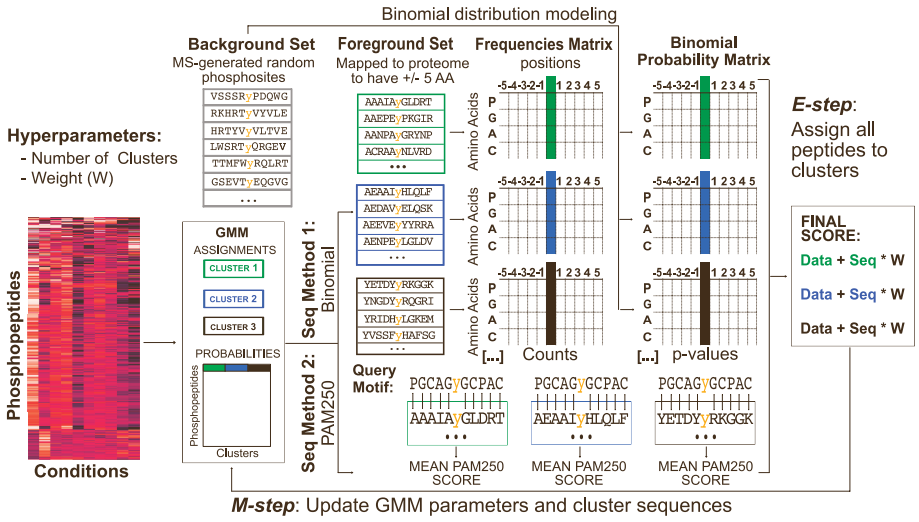


Figure S1: **Schematic of the DDMC simultaneous data and peptide sequence clustering approach.** Related to figure 1. Peptides are initialized into clusters at random. The measurements of abundance are represented by a multivariate Gaussian mixture model where each dimension of the distribution represents the abundance within a sample. Next, an expectation-maximization fitting scheme is used. During the expectation step, the distance of each peptide sequence to each cluster is calculated. This is done either through a binomial enrichment scheme (method 1, derived from (Schwartz and Gygi, 2005)) or using the average PAM250 distance (method 2). In parallel, the distance of each peptide abundance is compared to the cluster centers. These two distances are combined to update the assignments of each peptide to each cluster. During the maximization step, the cluster centers of the data are updated based on the weighted average of the peptide abundances in each condition. The peptide motifs are similarly updated through a weighted combination of the assigned peptides. Both steps continue sequentially until the change in peptide assignments between each iteration drops below a threshold.

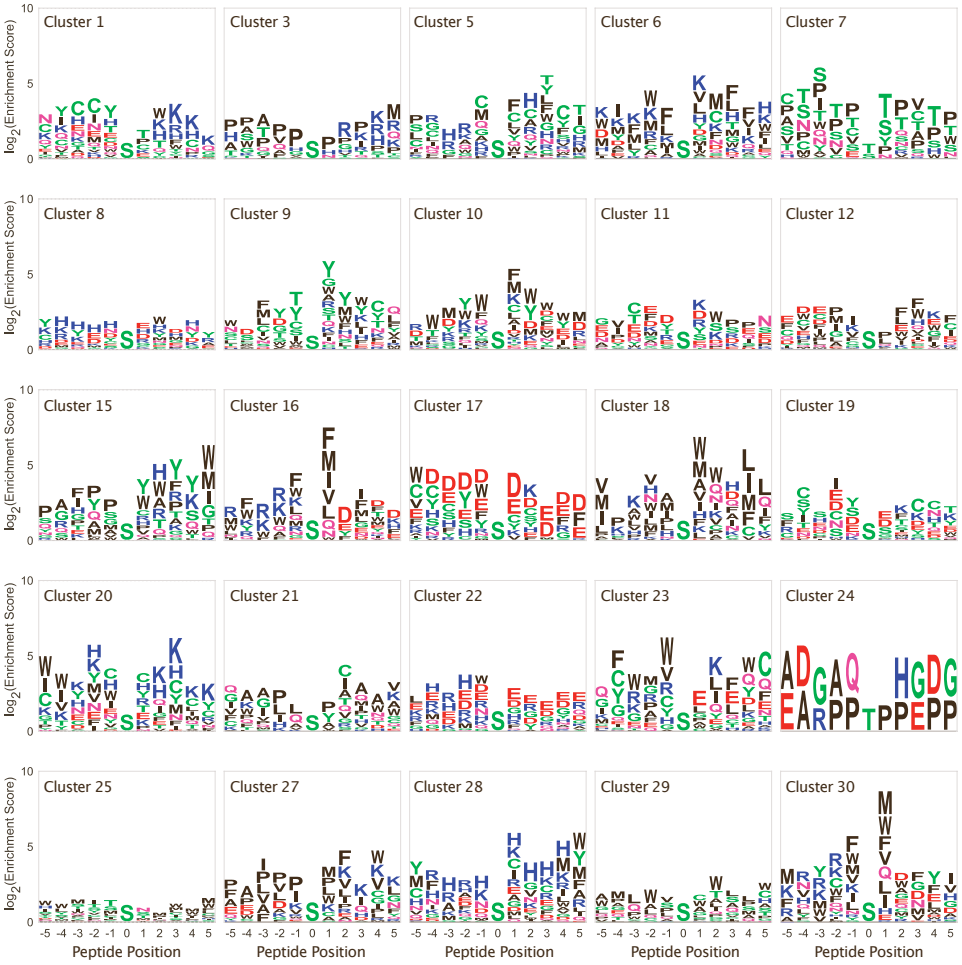


Figure S2: **Logo plots of all CPTAC Cluster PSSMs.** Related to figures 4-7. (A-U) Sequence motifs of clusters 1 through 30. Note that clusters that are not shown do not contain any peptides that are most likely assigned to them and that only 2 peptides are assigned to cluster 24. Amino acid types per color: Red=Acidic, Blue=Basic, Pink=Amidic, Black=Nonpolar, Green=polar.

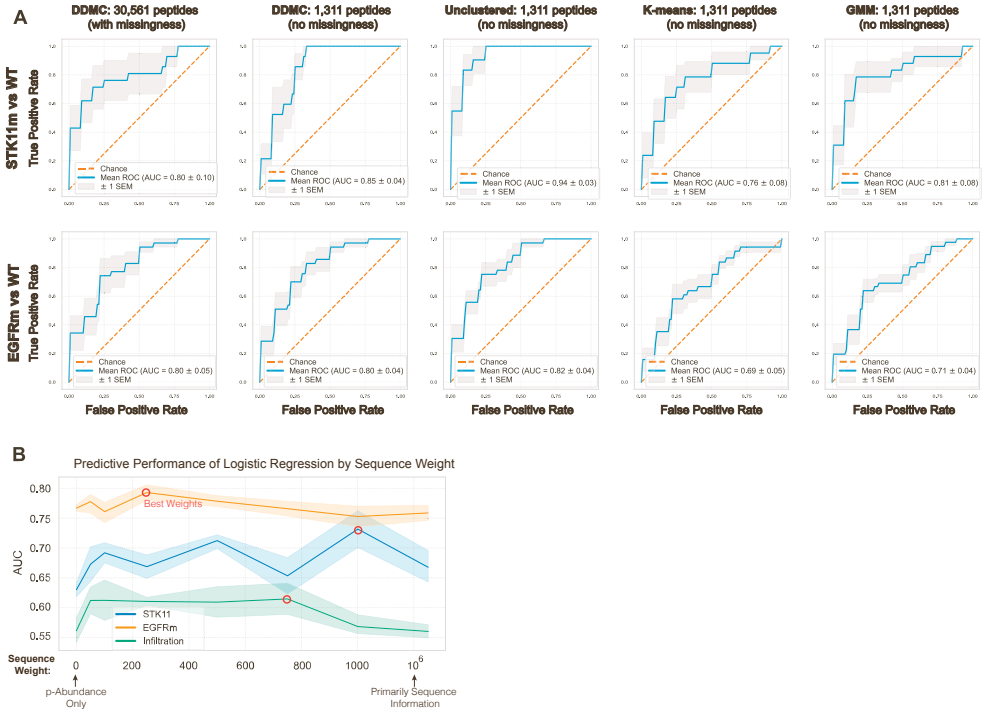


Figure S3: **Effect of sequence information on predictive performance.** Related to figure 4. A) Prediction of STK11m using either using DDMC fit to the complete data set with missingness (30,561 peptides) or the portion of the data without missingness (1,311 peptides). Using the data set without missing values, STK11m was predicted without clustering, with k-means, or GMM. The same approach was repeated to predict EGFRm. A regularized logistic regression model was used to predict the mutational statuses. B) Performance of a regression model predicting the mutational status of STK11 (blue) EGFR (yellow) and tumor infiltration level (hot versus cold) (green) in LUAD patients using either only phosphorylation data (1000,000), mainly peptide sequences (10^6), or a mix (50, 100, 250, 500, 750, 1000). Red circles denote the best predictive performance among weights.

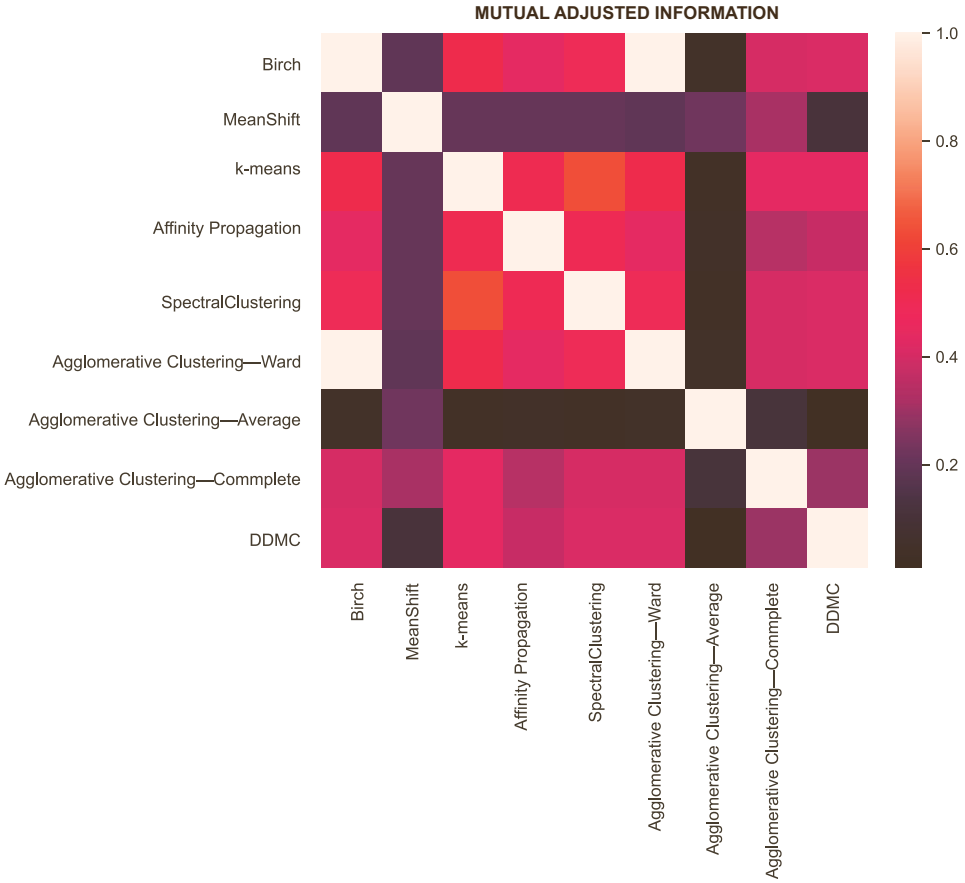


Figure S4: **Mutual adjustment information across clustering methods.** Related to figure 4. 8 different standard clustering methods and DDMC were fit to the phosphoproteomic data set generating 30 clusters. The peptide cluster assignments of each clustering method case were pairwise compared by calculating the adjusted mutual information score. Nearest neighbors was the affinity method used for spectral clustering, the remaining methods were run with the default parameters provided within scikit-learn (Zhang, Ramakrishnan and Livny, 1996; Knyazev, 2001; Comaniciu and Meer, 2002; Frey and Dueck, 2007) .

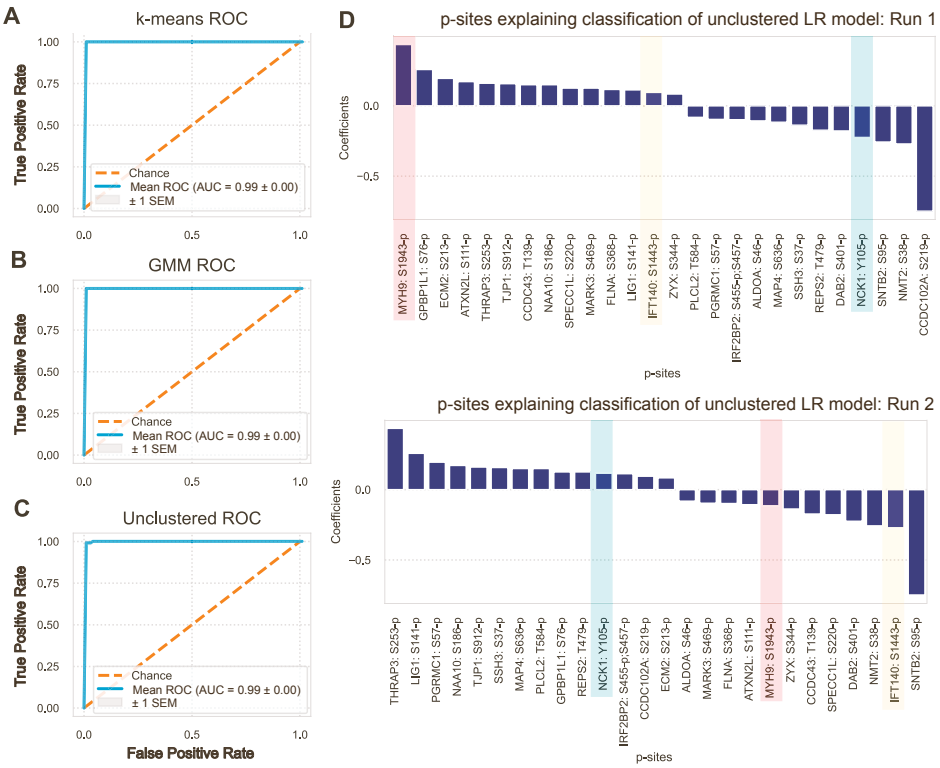


Figure S5: **Additional modeling strategies to find conserved tumor differences compared to NATs**. Related to figure 5. A-B) ROC plot of a logistic regression model fit to the complete portion of the signaling data set clustered by k-means and GMM. C-D) ROC plot of a logistic regression model fit to the complete portion of the phosphoproteomic CPTAC LUAD data set and (D) phospho-peptides with largest weights (w < 0.5 | w < -0.5) explaining the observed differences between tumors and NATs in two separate runs.

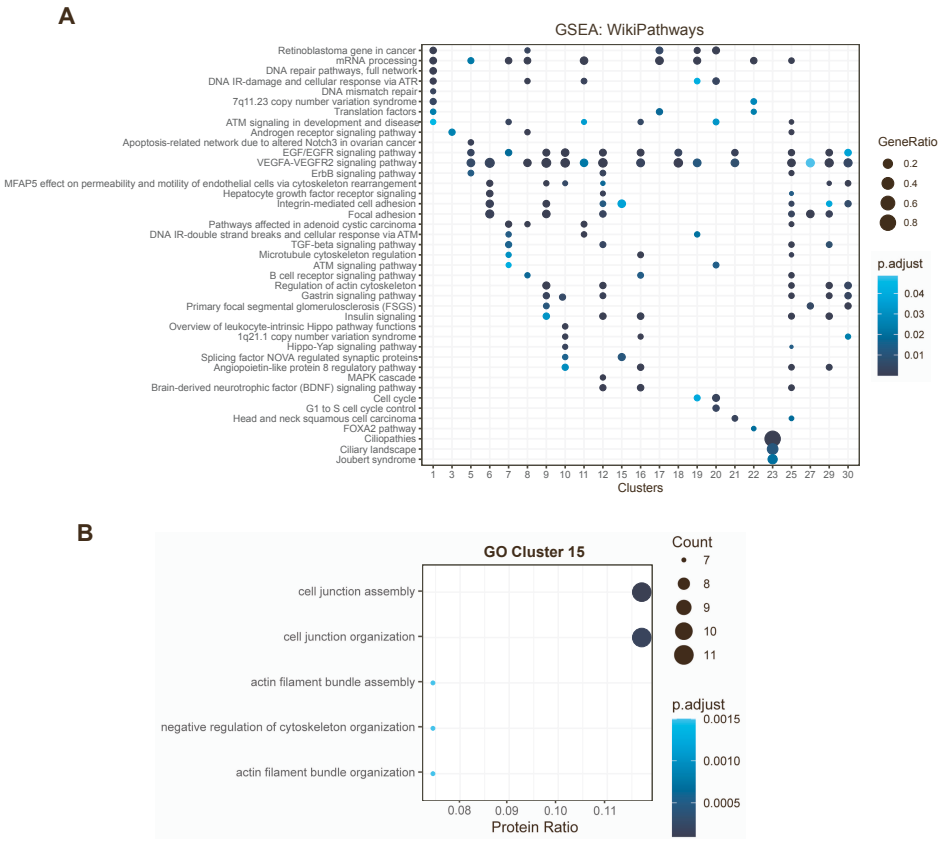


Figure S6**: Biological processes enriched in phosphoproteomic clusters**. Related to figures 5-7. A) GSEA per cluster using WikiPathway's gene set. B) Gene ontology analysis of cluster 15. GSEA was implemented using clusterProfiler (Wu et al., 2021).

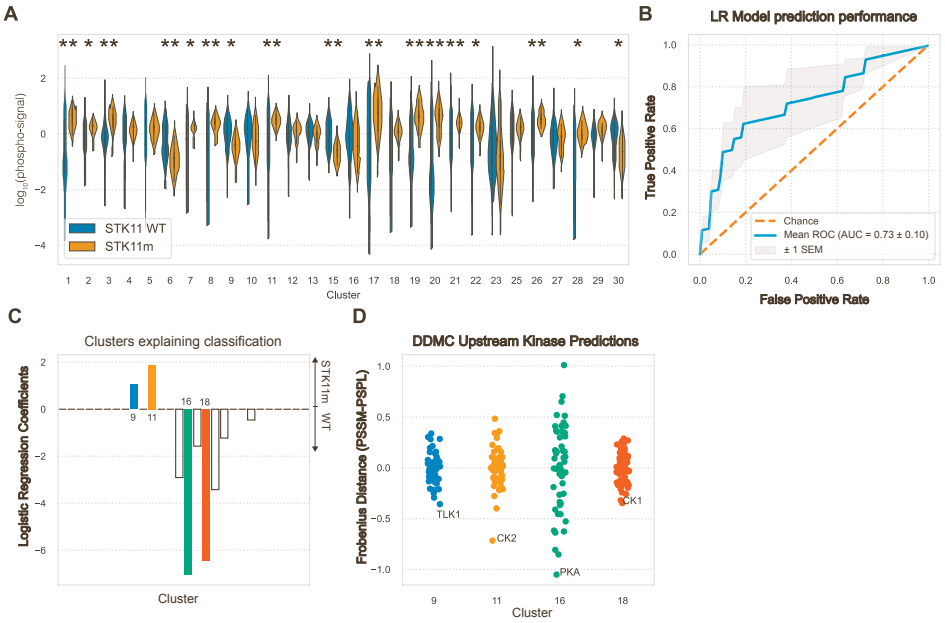


Figure S7: **Prediction of patient samples harboring STK11 mutations (STK11m)**. Related to figure 6. A) Phosphorylation signal of DDMC clusters grouped by STK11m and WT samples. Its statistical significance is indicated on the top part of the plot via a series of Mann Whitney rank test. B) ROC plot of the logistic regression model fit to the DDMC clusters. C) Logistic regression weights explaining classification of mutational status. D) Upstream kinase predictions of clusters 9, 11, 16, and 18.

Chapter 3:

Dissecting AXL-mediated bypass resistance and collateral phenotypes by phosphosite perturbations

Creixell M.[1], Taylor S., Gerritsen J.[2], Loui M.[1], Boixo C.[1], White F.[2], Meyer A[1]

[1]Department of Bioengineering, University of California Los Angeles

[2]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge MA, USA; Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge MA, USA; Center for Precision Cancer Medicine, Massachusetts Institute of Technology, Cambridge MA, USA

Abstract

Receptor tyrosine kinase (RTK)-targeted therapies are often effective but invariably limited by drug resistance. A major mechanism of acquired resistance involves "bypass" switching to alternative pathways driven by non-targeted RTKs that reactivate proliferation. Overexpression of the RTK AXL is frequently observed in bypass resistant tumors which, in addition to cell survival, drives further malignant phenotypes such as EMT and migration. However, the signaling molecules and pathways eliciting these responses remain elusive. To explore these coordinated effects, we generated a panel of mutant lung adenocarcinoma (LUAD) PC9 cell lines in which each AXL intracellular tyrosine is mutated to phenylalanine. Measuring phosphorylation signaling alongside several phenotypic changes associated with resistance, we related specific signaling and phenotypic changes through multivariate modeling. Our model identified an "AXL downstream signature" that correlates with progressive disease and poor clinical outcomes in patients. This signature comprises the activation of two signaling cores correlating with cell proliferation and migration; on one hand SFK, ABL1, and FAK acting upstream of the transcription factor YAP, and on the other hand CK2 protecting against E-induced DNA-damage. In total, this work demonstrates a methodology approach for dissecting pleiotropic signaling regulators and identifies several mechanisms by which AXL drives resistance-associated phenotypic changes.

Summary points

- AXL expression and abundance are associated with Yap activation in LUAD tumors.
- AXL phosphosite perturbations lead to varied phosphoproteomic and phenotypic effects.
- AXL downstream signaling correlates with poor clinical outcomes and progressive disease.
- AXL activates upstream kinases SFK, Abl1, and FAK of YAP to induce its activation.
- AXL protects cancer cells against erlotinib-induced DNA damage through CK2 activity.

Introduction

Lung cancer is the leading cause of cancer mortality, accounting for almost 25% of all cancer deaths in the United States for 2022[1]. Comprehensive genomic sequencing and expression profiling of non-small cell lung cancer (NSCLC) patients, the most common form of lung cancer, has helped to identify genetic alterations that drive disease progression and can be therapeutically targeted. These agents can have improved clinical efficacy and safety compared to conventional chemotherapy. One such therapy is the EGFR tyrosine kinase inhibitor (TKI) erlotinib which is effective in patients with advanced EGFR mutant (EGFRm) NSCLC[2]. However, despite being initially effective, targeted therapies invariably result in incomplete responses, with tumors regrowing after acquiring drug resistance. A major source of resistance to EGFR targeted therapy arises from secondary activating mutations occurring in the kinase domain of the receptor, such as the "gatekeeper" T790M alteration that increases the affinity of EGFR to ATP, relative to its TKI affinity. Second- and third-generation EGFR TKI therapies, such as afatinib and osimertinib respectively, have been developed to effectively target resistance derived from mutated forms of EGFR[3]. However, while agents targeting EGFR secondary mutations delay tumor relapse, the efficacy of these therapies are ultimately limited by resistance through still other mechanisms.

Another well-appreciated means of resistance to EGFR inhibition is receptor tyrosine kinase (RTK) "bypass" resistance, wherein alternative pathways are activated so that cells are no longer reliant on the drug-targeted pathway. Our lab and others have shown that, while individual RTKs are able to activate a common set of downstream pathways, they do so to varying extents and thus have varied capacity to confer bypass resistance[4–6]. Two well-studied RTK bypass resistance mechanisms are Her3 signaling providing resistance to Her2-targeted therapy in breast carcinoma and Met signaling driving resistance to EGFR-targeted therapies in lung carcinoma[5,7–9]. Bypass resistance-conferring RTKs may contribute to intrinsic or acquired resistance, can become

activated by ligand-mediated autocrine or paracrine induction, amplification, or mutation, and are effectively blocked by combination therapy[10].

A member of the Tyro3, AXL, and MerTK (TAM) RTK family, AXL, is frequently upregulated in tumors resistant to chemotherapy, targeted therapies, and immunotherapy, across cancer types, including EGFR-driven NSCLC[10–21]. AXL expression is additionally associated with collateral phenotypic changes in resistant cells, including epithelial-to-mesenchymal transition (EMT) and cell migration, indicative of increased metastatic capacity[22–28]. Furthermore, AXL has been shown to sustain the viability of osimertinib resistant cells in lung cancer *in vitro* and *in vivo* EGFRm lung cancer models[29,30]. A landmark study describing AXL-driven erlotinib resistance showed that tumor xenografts with acquired resistance harbored alteration in the expression of several EMT marker genes[15]. AXL additionally drives resistance to EGFR inhibition of cycling drug-tolerant persister (DTP) cells by protecting them against treatment-induced DNA damage through the activation of low-fidelity DNA polymerases, other endogenous hypermutators such as MYC, and a pyridine/pyrimidine metabolism imbalance[30]. Notably, the triple combination treatment of osimertinib, cetuximab, and an anti-AXL antibody led to cures in mice whose tumors had already acquired resistance to osimertinib[30]. Given the established role of AXL to promote tumor resistance, it is not surprising that ongoing phase I/II clinical trials are testing the clinical benefit of AXL and EGFR inhibitor combinations in EGFRm NSCLC (NCT02729298, phase I/II).

It is worth noting that although drug resistance is commonly quantified as a measure or proxy of cell number, tumor relapse is a multifaceted challenge driven by the development of simultaneous "collateral" malignant phenotypes that coordinately promote tumor growth and metastasis[31]. AXL has been involved in a myriad of biological processes that direct cancer progression, including EMT and metastasis[22,23,25,32], inhibition of apoptosis[33] and induction of cell proliferation[11,29,34],

DNA damage repair[30,35,36], endocytosis[14,37], and tumor immunosuppression[10,17,38,39]. However, despite evidence implicating AXL in mediating resistance and collateral phenotypes in response to EGFR targeted therapies, it is unclear exactly which pathways AXL activates to promote each of these processes. Identifying these pathways is hindered by RTK crosstalk and signaling pleiotropy: Each RTK regulates a set of downstream pathways that can be also regulated by a different RTK, or a combination of them. Moreover, downstream pathways can influence each other which complicates associating a specific signaling axis with a particular phenotype. Therefore, we require a methodology that specifically addresses RTK redundancy and signaling pleiotropy to mechanistically characterize AXL-mediated bypass resistance.

Upon interaction with its cognate ligand, two RTK monomers form a homodimer that autophosphorylates tyrosine (Y) residues located within its kinase domain, creating binding sites for interacting adapters and kinases that contain phospho-binding domains (PBD) such as SH2 domains[40]. The RTKs then phosphorylate these signaling proteins to initiate a downstream signaling cascade that will ultimately influence a myriad of cellular processes[6]. Since phenylalanine (F) is a non-phosphorylatable mimetic of Y, Y-to-F mutational studies provide a tool for dissecting the signaling role of individual phosphosites. For instance, AXL mediates resistance to cetuximab and radiation through the interaction of Abl1 and a Y residue on AXL kinase domain (Y821) in head and neck cancer and validated that editing Y821 to F abolished the ability of cells to develop resistance[41]. Other Y-to-F mutational studies in AXL have shown that the regulatory subunit of PI3K, p38, binds to Y779 to activate MAPK, whereas SRC, LYK, and PLCα engage with Y821 and Y866 to activate PKC and STAT[42–44]. As individual AXL Y-to-F mutations have different effects in the receptor's intracellular signaling, we hypothesized that these mutations would also distinctly mediate AXL's ability to confer erlotinib resistance which would

provide us with the opportunity of establishing AXL-specific signaling and phenotypic associations during AXL bypass resistance.

Here, we apply an integrated experimental and computational strategy to systematically elucidate the downstream signaling proteins and pathways driving cell survival and collateral phenotypes in response to AXL activation. To do so, we generated a cell line panel wherein each cell line carries a single Y-to-F mutation in AXL. We measured phosphorylation signaling alongside cell viability, apoptosis, migration, and an erlotinib-induced clustering effect in AXL-activated cells treated with erlotinib. Through statistical modeling to analyze the paired AXL-driven phosphoproteomic and phenotypic measurements, our results indicate that AXL activates SFK, Abl1, and FAK which serve as key upstream components of the YAP pathway to promote cancer cell growth. We correspondingly observed that AXL activation led to increased YAP nuclear translocation. Moreover, we observed a strong correlation between the identified AXL downstream signaling signature and AXL expression in EGFRm tumors, poor survival, and progressive disease in LUAD patients. Separately, we found that AXL protects lung cancer cells against DNA damage induced by erlotinib, at least in part, through CK2 kinase activity.

Results

AXL-high LUAD tumors display strong Yap activation and EMT marker enrichment

To explore AXL's association with tumor progression and identify the signaling pathways most strongly correlated with AXL expression and abundance, we examined the proteogenomic data of 110 treatment-naïve LUAD patients from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) [45]. We found that AXL abundance is modestly increased in EGFRm versus EGFR WT tumors across stages and that within EGFRm patients, AXL increases throughout disease

progression (**Figure 1A**). We then stratified LUAD tumors into the top and bottom 33% AXL abundance. Comparing these two patient groups, we found enrichment of a Yap signaling signature[46] in AXL-high LUAD tumors by GSEA (**Figure 1B**). Consistent with this observation, we observed a strong correlation between AXL expression and established Yap targets including *CYR61* and *CTGF,* alongside a decrease in the inhibitory phosphorylation site of Yap S382 in AXL-high tumors (**Figure 1C-D**). Yap S382-p displayed an inverse correlation with tumor stage (**Figure 1E**). Finally, we observed a significant increase in EMT markers *TWIST1*, *VIM*, and *CDH11* in AXL-high compared to AXL-low tumors (**Figure 1F**). Overall, these results indicated to us that the Yap signaling is associated with AXL expression and abundance in LUAD tumors.

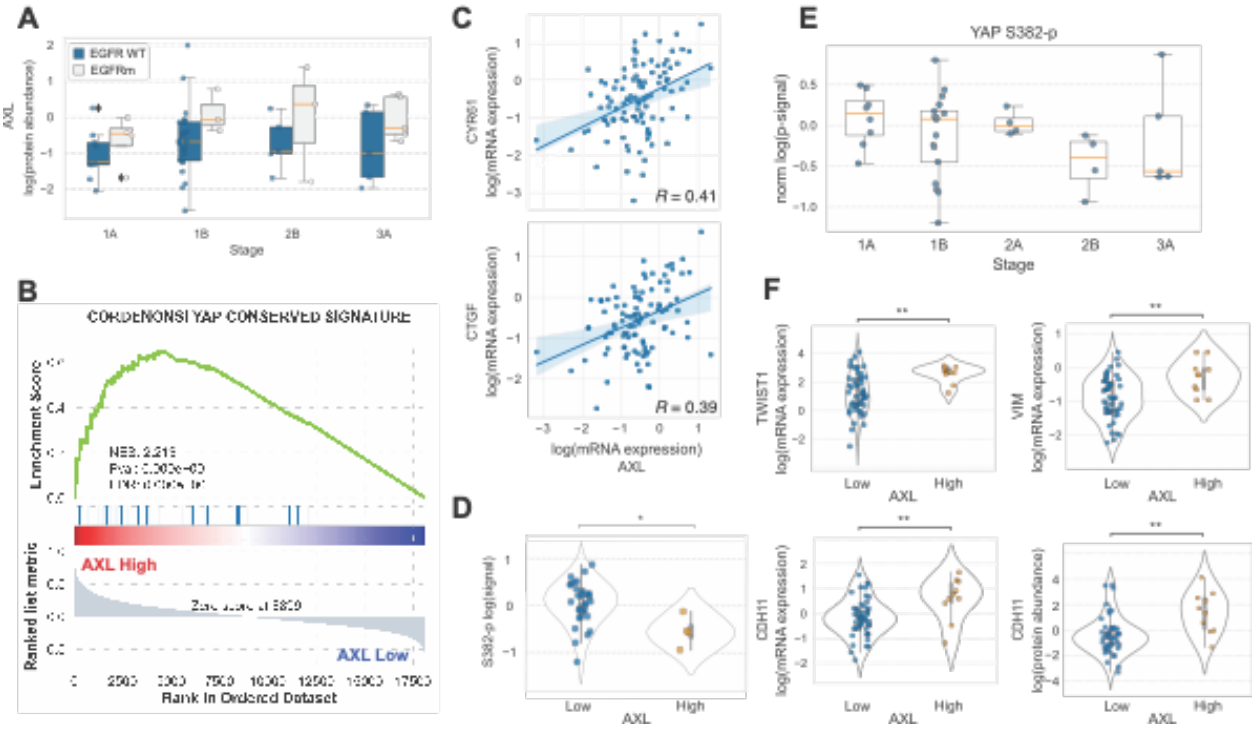


Figure 1. AXL-high LUAD tumors display increased YAP activation and EMT markers. (A) AXL protein levels grouped by EGFR mutational status and tumor stage. (B) Transcriptomic YAP signature in AXL-high vs AXL-low tumors (C-D) Abundance of YAP inhibitory phosphosite by AXL levels (C) and tumor stage (D). (E) Expression of mesenchymal markers by AXL levels. *p-value < 0.05 and **p-value < 0.001 according to a Student T-test.

AXL's capacity for erlotinib resistance varies across AXL Y-to-F mutants

To isolate the pathways driving specific phenotypic effects, we generated a panel of AXL Y-to-F mutant cell lines and collected phosphoproteomic and phenotypic measurements of each during AXL-driven bypass signaling. We knocked-out (KO) AXL in PC9 cells using CRISPR/Cas9 and then expressed either AXL wild-type (WT) (KI), AXL kinase-dead (K567R, KD), or one Y-to-F AXL mutant cassette—namely Y634F, Y643F, Y698F, Y726F, Y750F, or Y821F—using lentiviral transduction (**Figure 2A**). We confirmed AXL total abundance, cell-surface abundance, and activation in each cell line (**Supplementary Figure 1A**). AXL activation is challenging to faithfully manipulate given its regulation by phosphatidylserine (PS) present in apoptotic bodies and the spatially heterogeneous presentation of its ligand Gas6[47]. Therefore, we used the AXL-activating antibody AF154 (A) as a reliable means for potent receptor activation (**Supplementary Figure 1B**). AXL activity was confirmed by phosphorylation of Akt (**Supplementary Figure 1C**). As expected, we observed complete inhibition of p-EGFR upon erlotinib treatment (**Supplementary Figure 1D**). We measured the ability of the different AXL mutants to proliferate, survive, migrate, and form cell "islands" in the presence of erlotinib (E) or the combination of E and A (EA) (**Figure 2B**).

We simultaneously monitored cell proliferation and cell death using cell live imaging throughout 72 hours of treatment with either E or EA. We observed a significant increase in cell proliferation and decrease in apoptosis in EA-treated WT cells compared to the E condition. KI cells followed the same trend with lesser magnitude, likely explained by the fact that each PC9 AXL mutant cell line is comprised by a mixed population of PC9 AXL KO cells and cells expressing AXL. As expected, EA did not enhance proliferation or survival versus E in AXL KO and KD. While the mutants Y698F and Y726 promoted proliferation and prevented apoptosis upon AXL activation, Y643F and Y750F effectively behaved like AXL KO or KD. Interestingly, EA-treated Y821F and

Y634F cells failed to increase cell proliferation but significantly blocked apoptosis (**Figure 2A/B and Supplementary Figure 2A/B**).

We performed a scratch-wound assay to evaluate the migratory capacity each cell line after treatment (**Supplementary Figure 1E**). After making a wound, we used cell live imaging for 24 hours to quantify the ability of PC9 cells to migrate and re-occupy the space in the presence of E or EA. WT cells treated with E unexpectedly migrated similarly in the presence or absence of the AXL-activating antibody, whereas the migratory ability of KO and KD was blocked, regardless of treatment. Moreover, Y643F, Y698F, and Y726F became more motile after receptor activation, while the mutations Y634F, Y750F, Y821F did not respond (**Figure 2F** and **Supplementary Figure 2C**).

We observed that E induces a clustering effect wherein cells establish cell-to-cell adhesions that result in the formation of small "cell islands" (**Figure 2B** and **Supplementary Figure 1F**). Thus, we asked whether the different PC9 AXL mutant cell lines would differently affect this phenotype. To quantify the extent of cell clustering, we applied Ripley's K function, a spatial clustering estimator. This metric frequently used in astronomy to model whether an object, such as stars are found closer to one another than would be expected by chance[48,49]. This algorithm allowed us to test the spatial distribution of PC9 cells against the null hypothesis that the cells are distributed randomly. As expected, we found that PC9 cells were more clustered than chance in response to E, whereas AXL activation abolished this clustering. Each Y-to-F mutants displayed the trends counter to those observed in the scratch wound assay: Y643F, Y698F and Y726F effectively behaved like WT and KI, whereas Y634F, Y750F, and Y821F remained clustered upon EA treatment. (**Figure 2G** and **Supplementary Figure 2D**).

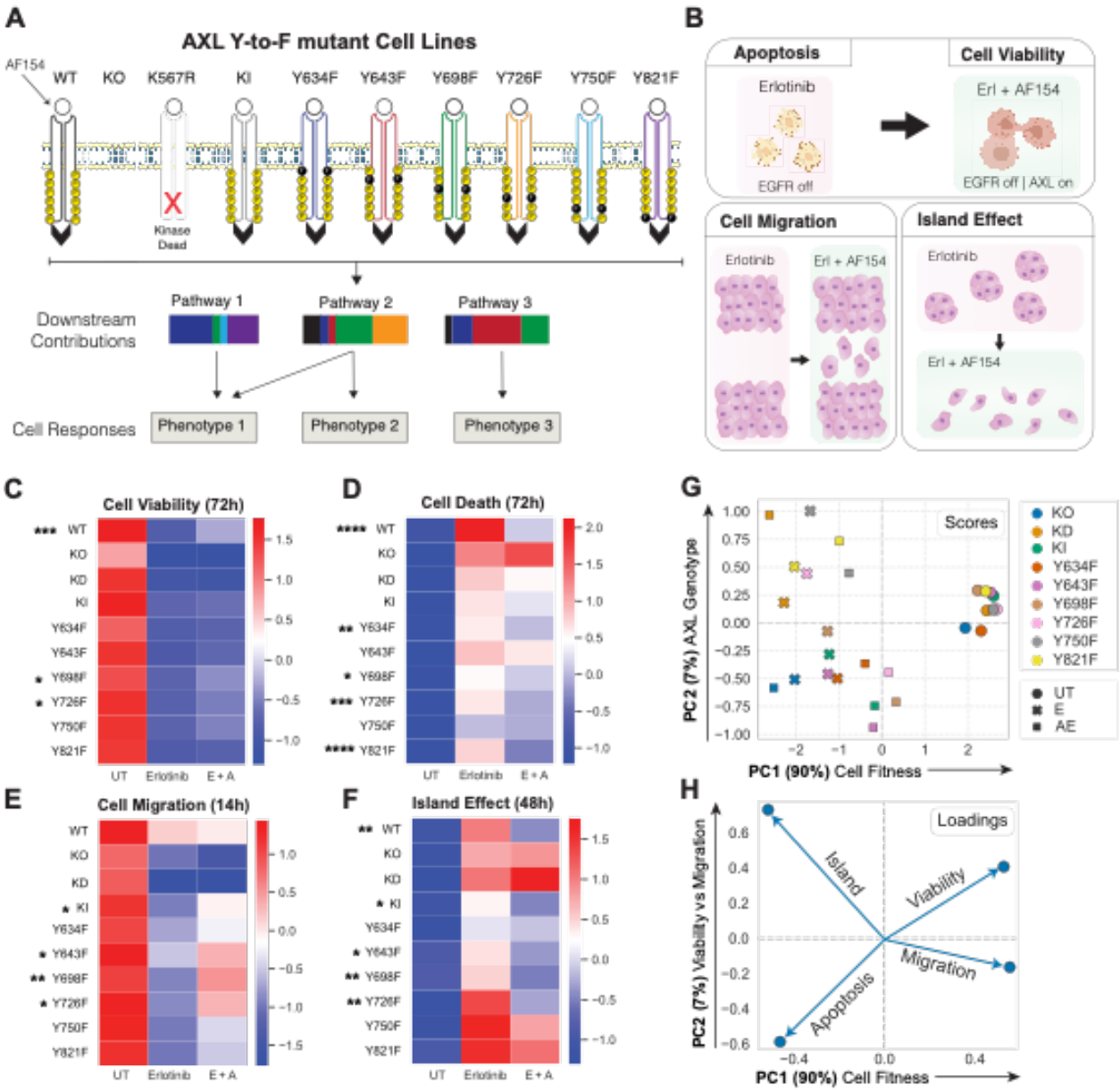


Figure 2. The ability of AXL to enhance cell fitness in the presence of erlotinib varies among PC9 AXL Y-to-F mutants. (A) Schematic of the AXL Y-to-F mutant cell lines each causing distinct signaling and phenotypic consequences upon treatment with erlotinib for 4h and an AXL-activating antibody AF154 for 10 minutes. (B) Cell proliferation and cell death were monitored for 72h, whereas cell migration was quantified via a scratch wound assay. Finally, AXL activation blocks the formation of cell islands induced by erlotinib. The extent of cell clustering was quantified using Ripley's K function. (C-F) Quantification of cell proliferation, cell death, cell migration, and island effect by PC9 cell line. (G-H) Scores (G) and loadings (H) of a PCA analysis using the phenotypic measurements shown in (C-F).

We used principal components analysis (PCA) to explore the relationships among the four phenotypes we measured (**Figure 2H/G**). Most of the variation, represented by principal component (PC) 1, prominently separated untreated cells from the other two treatment groups, and partly separated EA from E cells. We took this component to represent overall cell fitness, which is increased by moving positively along PC 1 for all the mutants except KO and KD, where A had no effect (**Figure 2G**). Phenotypes increased by AXL activation are positively associated with PC 1, while those decreased are negatively associated (**Figure 2H**). By contrast, PC 2 separated AXL-induced viability from migration effects; moving positively along PC 2 indicated an increase in viability/decrease in apoptosis, while migration decreased, and cells formed islands (**Figure 2C, loadings**). Certain cell lines showed varying effects with AXL activation: Y726F and Y750F shifted positively along PC1 and negatively along PC2 with A treatment, reflecting migration-selective effects; Y634F and Y821F moved towards positively along PC1 and PC2, reflecting cell viability/apoptosis effects.

Thus, overall, these results confirm our hypothesis that each Y-to-F mutant affects cell viability, death, migration, and clustering differently. Y634F and Y821F blocked AXL-induced cell viability, migration, and scattering, but still improved survival of PC9 cells in the presence of E. Y750F blocked all phenotypic effects, behaving like KO and KD. Finally, Y698F and Y726F behaved like WT and KI, with little impact on phenotypic outcomes.

AXL mutants selectively disrupt downstream pathway effects

As previously described, we applied a mass spectrometry approach to measure the signaling effect of AXL activation in the PC9 AXL mutation cells during EGFR inhibition. Each PC9 mutant cell

line was treated with E for 4 hours and activated with A for 10 minutes afterwards before performing cell lysis. Hierarchical clustering of the resulting phosphoproteomic data set showed AXL KO and KD clustered together, as expected, whereas the other cell lines did not display obvious clustering (**Figure 3A**).

To analyze these data, we applied a previously-developed approach, Dual Data-Motif Clustering (DDMC). DDMC clusters phosphoproteomic data based on both abundance variation and sequence information, both of which can be leveraged to improve modeling[50]. We applied DDMC with 5 clusters using the PAM250 method to model the peptide sequence information. The resulting cluster (C) centers summarized the AXL-responsive behaviors. C1, C2, and especially C3 were markedly decreased in PC9 AXL KO and KD cells, with varied phosphorylation signal across Y-to-F mutants. Interestingly, in addition to a dramatic phosphorylation decrease in AXL KO and KD, C3 peptides show an increased abundance in PC9 AXL Y698F—a mutant that phenotypically behaves like WT (**Figure 2**)—compared with the other clusters. Conversely, C4 shows an increase in PC9 KO—but not KD—with respect to WT, and a lower phosphorylation signal in Y634F compared to the rest of clusters (**Figure 3C**). Cluster 5 had biologically inexplicable signaling trends, and so was discarded. We searched for AXL phosphosites and found thatall of them were clustered together in C3 which is not surprising since C3 displays the most dramatic phosphorylation changes based on AXL activity (**Figure 3C/D**). None of the phosphorylation differences observed in the AXL phosphosites correlate with total or cell surface AXL levels per cell line which provides confidence for the absence of any experimental bias in our signaling data due to varying ectopic AXL expression (**Supplementary figure 1A**).

We ran GSEA on each cluster to explore their functional roles. We found that C1 is strongly enriched in a TGF-beta signaling pathway signature; C2 and C3 share the enrichment of several biological processes associated with the activation of RTK signaling (EGFR / VEGFR) and the regulation of focal adhesions; and C4 is uniquely enriched in a STAT3 signaling pathway signature (**Figure 3E**). The enrichment of an EGFR downstream signature in EA treated cells might underscore the capacity of AXL to reactivate EGFR-mediated pathways required for survival. Thus, these results illustrate the activation of known biological processes regulated by AXL such as focal adhesion dynamics to regulate cell migration[27] and RTK signaling activation, in addition to less established associations such as its relation to TGF-beta and STAT3 signaling.

To obtain a general view of the composition of the different clusters, we next explored the cluster assignments and phosphorylation of RTKs, receptor adapters, and canonical protein kinases in PC9 WT and AXL KO cells treated with EA (**Figure 3F**). Among RTKs, we found a decrease in the phosphorylation of Epha2 Y594-p and Her2 Y877-p in AXL KO compared to WT. This is consistent with previous studies that have observed signaling crosstalk between AXL and HER2 to drive resistance to anti-HER2 therapy in breast cancer, as well as between AXL and EPHA2 to confer resistance to EGFR inhibitors[19,51,52]. We found that the RTK adapters Gab1/2, Eps8, Sos1, or Dapp1 as well as the E3-ubiquitin ligase Cbl-b are markedly downregulated in AXL KO. Interestingly, the phosphorylation of these proteins is dependent on the availability of AXL Y821; these phosphosites were not phosphorylated in the AXL Y821F cell line (**Supplementary Figure 3A**). Moreover, we observed increased phosphorylation of SFKs Frk, Lyn, Lck, Yes1, as well as Abl1 and its substrate and regulator Abi2 in C2 and C3. In addition, while Frk, Lyn, and Lck phosphorylation was increased upon E treatment, monotherapy or concomitant addition of the AXL inhibitor R428 downregulated them. R428 also inhibited Abi2 and Abl1, suggesting an AXL-

specific activation of SFK and Abl1 (**Supplementary Figure 3B**). AXL activated the established drivers of resistance to TKI Erk1&2 which are both within C3[15]. Administration of E (in the absence of the AXL-activating antibody) and R428, as single agents or in combination, downregulated ERK1/2 which suggests an AXL-specific activation of ERK signaling as well. Conversely, JNK2 and JNK3 phosphosites are modestly more abundant in AXL KO than WT and are clustered in C4. Given the interesting signaling behavior and cluster composition of C3, we decided to use STRING to map the protein-protein interactions while observing the phosphorylation changes of PC9 WT compared AXL KO cells within C3. We found that in addition to Erk1/2, other kinases are highly regulated by AXL activation within C3 such as Cdk1 (as shown by the lack of phosphorylation of its inhibitory phosphosite Y15), Prkcd, Cilk1, Ack1, and the phosphatase Shp2 (**Supplementary Figure 3C**). As indicated by GSEA, this signaling cluster is highly enriched with regulators of focal adhesions and cytoskeletal remodeling such as Tjp1/2, p130cas (*BCAR1*), Actin1, and Afadin (*MLLT4*) (**Figure 3E/F**).

In conclusion, the phosphorylation-based signaling landscape of the different PC9 cells varies by AXL genotype, demonstrating that single Y-to-F mutations lead to distinct AXL-specific signaling perturbations in addition to different cell responses (**Figures 2&3**). DDMC summarized the network-level phosphorylation changes across PC9 Y-to-F mutants during switched AXL activation and found C1, C2, and especially C3 correlate with AXL activation levels, while peptides in C4 are slightly increased in AXL KO cells compared with WT. These results led us to posit C3 as a core downstream component during AXL bypass signaling.

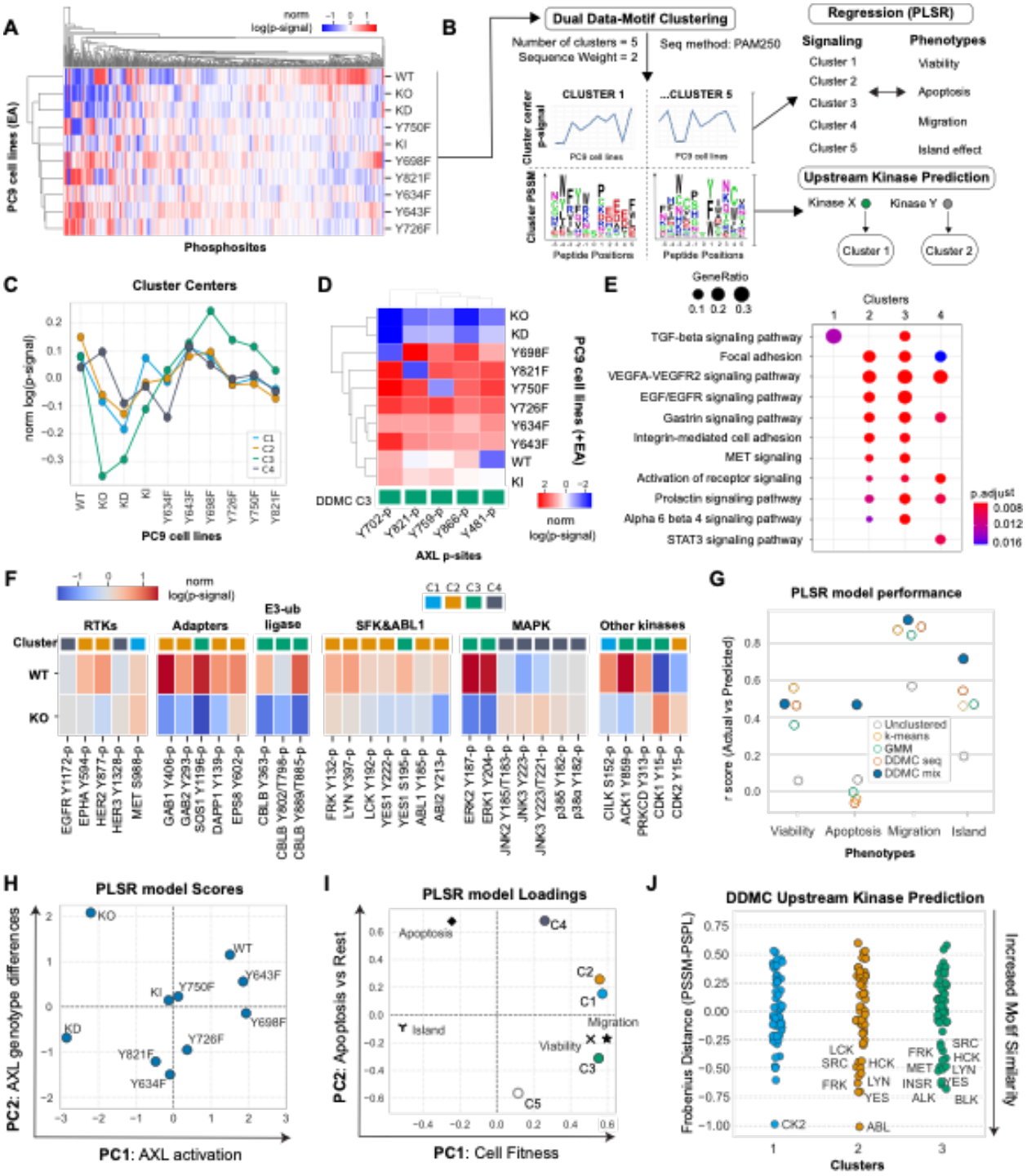


Figure 3. DDMC signaling clusters predict the AXL-mediated phenotypes and identifies CK2, ABL1, and SFK as putative bypass signaling drivers. (A) Global phosphoproteomic measurements of PC9 AXL Y-to-F cell lines. (B) Computational strategy to map the network-level phosphoproteomic changes driving AXL-mediated phenotypic responses. The signaling data was clustered using Dual Data-Motif clustering (DDMC) to generate 5 clusters of peptides displaying similar phosphorylation behavior and sequence features. The cluster centers were then fit to a

PLSR model to predict the phenotypic responses and find associations between clusters and phenotypes. DDMC was used to infer putative upstream kinases regulating clusters. (C) Phosphorylation signal of DDMC cluster centers. (D) Phosphorylation signal of AXL phosphosites per PC9 cell line and its cluster assignments. (E) Ranked GSEA analysis of DDMC clusters using ClusterProfiler. Gene lists per cluster were ranked based on the log phosphorylation abundance fold change of PC9 WT versus AXL KO cells. (F) Indicated phosphosites and its cluster assignments in PC9 WT and AXL KO cells. (G) PLSR model prediction performance using 5 different clustering strategies namely either no clustering: directly fitting the phosphoproteomic data—, or the 5 cluster centers generated by k-means, a Gaussian Mixture Model (GMM), DDMC using only the peptide sequence information to make the cluster assignments and ignoring the peptide phosphorylation abundance or using DDMC equally prioritizing the sequence and phosphorylation information. (H-I) PSLR scores and loadings. (J) DDMC upstream kinase predictions.

DDMC clusters predict AXL-mediated phenotypic responses and identifies SFK, ABL1 and CK2 as downstream drivers of erlotinib resistance

To associate the different AXL-mediated signaling clusters and phenotypes, we regressed the DDMC cluster centers against the phenotypic responses using a partial least squares regression (PLSR) model (**Figure 3B**). To verify the importance of DDMC-mediated clustering for prediction, we assessed the prediction performance of a PLSR model using different clustering strategies: no clustering (i.e. regressing the phosphoproteomic data set directly), k-means clustering, clustering with a Gaussian Mixture Model (GMM), DDMC using only the sequence information (DDMC seq), or DDMC equally prioritizing the phosphorylation abundance and peptide sequences (DDMC mix). We found that PLSR was only able to accurately predict all four phenotypes when using the cluster centers generated by DDMC mix. By contrast, a PLSR model fitting the MS data directly is not able to predict any of the phenotypes, likely due to the overwhelmingly higher number of peptides (498) versus cell lines (10) (**Figure 3F**).

The scores and loadings of the resulting PLSR model illustrate the varying capacity of the different PC9 cell lines to promote cell proliferation and migration along PC 1, which, similar to the PC 1

of the PCA analysis of the AXL phenotypes (**Figure 2G/H**), are indicative of overall cancer cell fitness (**Figures G/H**). PC9 WT, as well as the AXL mutants Y698F and Y643F correlate with increased cell fitness by moving positively along PC 1 whereas PC9 AXL KO and KD are heavily negatively weighted along the same axis, thus strongly correlating with erlotinib-induced apoptosis and the island effect. Moreover, PC 2 seems to capture variation that is specific to apoptosis which increases by moving positively along this axis. The PLSR model captures an inverse correlation between apoptosis and the PC9 AXL mutants Y634F and Y821F as they are negatively associated with PC 2, consistent with the exclusive phenotypic effect of Y634F and Y821F in reducing apoptosis (**Figure 2C/D**); their PLSR scores are not weighted along PC 1 and therefore not associated with proliferation, migration, nor the island effect. These results highlight that the phosphoproteomic variation of AXL Y634F and Y821F explains their ability to exclusively block cell death, without affecting any of the other phenotypes. Importantly, we found that only C3 strongly correlates with cell migration and proliferation along both PC1 and PC 2, whereas C1, C2, and C3 strongly correlate with cell migration and proliferation across PC1 but not PC 2 (**Figure 3G/H**).

A feature of DDMC is the construction of position-specific scoring matrices (PSSMs) for each cluster which represent the frequency of each residue across peptide positions. These computationally-derived kinase motifs cab then be compared against a compendium of 60 experimentally-determined kinase motifs to infer the upstream kinases regulating each cluster[50] (**Figure 3B**). DDMC found that C1 displays a kinase motif characterized by the strong enrichment of an acidophilic C-terminus, a known hallmark of CK2 specificity[53,54]. The GSEA results of C1 provides support to this DDMC upstream kinase prediction as CK2 has been reported to be activated by TGFβ treatment and required for TGFβ-induced EMT[55]. CK2 has also been shown to

phosphorylate the C1 members Mcm2, Ldlr, and Sptbn1[56]. With C2, the PSSM shows an Abl1 kinase motif since the kinase specificity of Abl1 features a proline at position +3 with respect the phosphorylation site, an isoleucine at -1 and an alanine at +1. Both C2 and C3 have sequence motif features associated with SFK: C2 has a strong enrichment of acidic residues at position -3 and to a lesser extent at position -2. The SFK Yes1 specifically favors glycine, threonine, and tryptophan at position +1 as well as serine, glycine, and methionine at -2 which are all included in the PSSM of C2. On the other hand, C3 displays a strong enrichment of hydrophobic residues leucine and phenylalanine at position +3, serine and threonine at +2, and basic residues at position +4 and +5 which are all specificity drivers of SFK. C3 was also inferred to regulated by the RTKs Alk, Met, and INSR. These receptors tend to target substrates that present highly hydrophobic C-terminus in addition to hydrophilic residues at -1 which are characteristics of the motif of C3 (**Figure 3I** and **Supplementary Figure 3D**). The association of C3 with RTKs is consistent with its enrichment and STRING analysis (**Supplementary Figure 3A/D**).

In summary, the results of our paired experimental and computational approach suggest that AXL activates Ck2, Abl1, and SFK which in turn regulate C1, C2, and C3, respectively, to promote cell proliferation and migration in the presence of E.

AXL downstream signaling correlates with poor patient survival and progressive disease in EGFRm LUAD patients

We sought to explore the clinical relevance of the AXL downstream signaling identified by our analysis and its association with resistance to EGFR therapy in vitro and in LUAD patients. To do so, we first selected all phosphosites in C1, C2, and C3 and ranked them based on the log2 fold-change in PC9 WT versus PC9 AXL KO. The resulting “AXL downstream signature” shows an

enrichment of an EGFR TKI resistance signature according to a ranked GSEA analysis (**Figure 4A**).

We then asked whether the AXL downstream signature is enriched in AXL-high LUAD tumors in patients and whether it correlates with poor clinical outcomes. To do so, we again stratified the CPTAC LUAD patient samples into AXL high or AXL low tumors and looked at the overall and phosphorylation abundance of the members of C1-3. We found a significant enrichment in the protein expression of C1, C2, and C3 in AXL-high and EGFRm compared to AXL-low and EGFR WT tumors (**Figures 4B/C**). We did not observe a difference in phosphorylation abundance in C1, C2, and C3 in AXL-high versus AXL-low tumors (data not shown). However, unlike AXL-low samples, AXL-high tumors display a remarkable phosphorylation enrichment of the AXL signature in EGFRm compared with EGFR WT (**Figure 4D** and **Supplementary Figure 4A**).

With single cell RNAseq (scRNAseq), we asked, whether in EGFRm tumors, the gene expression of the AXL downstream signature is higher in patients displaying progressive disease, and thus resistance to EGFR-targeted therapies. Our AXL signature was significantly upregulated in cancer cells compared with other cell types, as well as in resistant and metastatic tumors (**Figure 4G-J**). By plotting each cluster separately, we found that these differences are explained by the contribution of the genes present in C2 and C3, but not C1 (**Supplementary Figure 4C-N**).

Using TCGA data, we explored whether the gene expression of AXL downstream signaling correlates with poor overall survival of LUAD patients. Interestingly, we observed statistically significant decreases in the percent survival of patients with higher gene expression of the AXL signature in LUAD and PAAD tumors (**Figure 4J/K** and **Supplementary Figure 4R**). Enrichment of this signature in PAAD patients with poor clinical outcomes is consistent with the fact that many

PAAD tumors are driven by mutant EGFR and that AXL has indeed also been implicated in resistance to TKI in this context[57].

In all, these results support a strong association between the identified AXL downstream signaling and poor clinical outcomes of EGFRm LUAD patients.

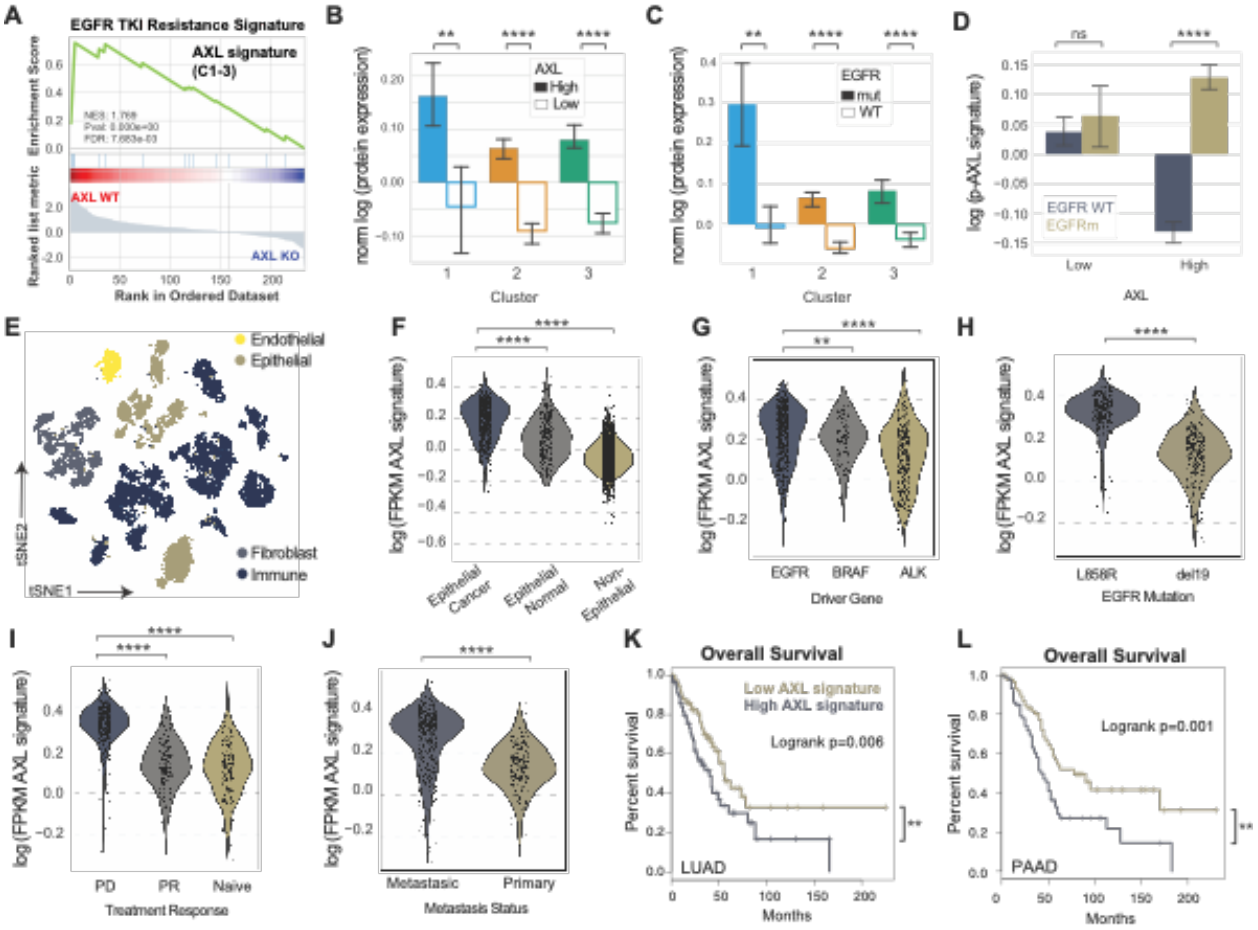


Figure 4. AXL downstream signature based on C2 and C3 is specific to AXL-high EGFRm LUAD tumors and correlate with progressive disease. (A) EGFR TKI resistance signature found by a ranked GSEA analysis using the list of gene names included in C1, C2, and C3 and ranked by their log fold-change phosphorylation between PC9 WT and AXL KO cells. (B-C) Protein expression of C1, C2, and C3 members in (B) AXL-low versus AXL-high tumors or (C) EGFRm versus EGFR WT tumors. (D) Phosphorylation signal of AXL signature by AXL levels and EGFR genotype. (E) tSNE plot of the different cell types defined by Louvain clustering. (F) AXL signature score as defined

by the mean gene expression of C1, C2, and C3 per cell in cancer cells, epithelial normal cells, or non-epithelial cells. (G-J) AXL signaling score of cancer cells by (G) Driver mutation, (H) EGFR mutation, (I) treatment response or by (J) Metastatic status. (K-L) Kaplan-Meier curve of (K) LUAD and (L) PAAD patients according to their AXL signature score. PD: Progressive disease, PR: Partial Response. PAAD: Pancreatic adenocarcinoma.

Dasatinib inhibition validates SFK and Abl1 as regulators of C2 and C3

To experimentally validate that SFK and Abl1 regulate C2 and C3, we measured cell proliferation and the global phosphoproteomic state of PC9 cells upon treatment with the targeted inhibitor dasatinib. We first asked whether dasatinib was able to block the cell proliferation increase induced by AXL activation. We treated PC9 cells with E, EA, or E with the AXL inhibitor R428. As expected, EA-treated and E+R428 cells were significantly more, and less proliferative, respectively, than cells treated with E, and we observed a dose-response decrease of cell proliferation with increasing concentrations of dasatinib (**Figure 5A**).

We performed pY-based mass spectrometry once again to ascertain whether there was an enrichment of phosphosites from C2 and C3 in those most prominently depleted by dasatinib treatment. We activated AXL in PC9 WT cells after pretreating with E and increasing concentration of dasatinib for 4 hours. In parallel, we also pretreated AXL KO cells. We then lysed cells and run mass spectrometry (**Figure 5B**). We ran hierarchical clustering and identified a cluster of peptides that displays clear decreasing phosphorylation levels with increasing dasatinib (**Supplementary Figure 5A** and **Figure 5C**). Asking which of the DDMC AXL clusters has a statistically significant enrichment of dasatinib-responsive peptides we saw extensive overlap with C2 and C3 (**Figure 5C/D**). Consistent with our AXL phosphoproteomics data, beyond ABL1- and SFK-related phosphosites, this cluster includes most of the AXL-regulated kinases previously described in **Figure 3** including Ack1, Prkcδ, FAK1/2, and Epha2, as well as proteins associated

with focal adhesion regulation and cytoskeletal remodeling such as p130cas, CasL, Git1, Tjp1, Paxilin, and Actin1 (**Figure 5C**).

A previous FAK pathway signature in EMT-mediated erlotinib resistant lung cancer cells was counter-acted by dasatinib[58]. These mesenchymal resistant cells overexpress many of the proteins that are our model identified to correlate with migration and proliferation when differentially phosphorylated upon AXL activation (**Figure 3E/H** and **Supplementary Figure 3C**). Thus, we asked whether AXL downstream signature overlaps with the FAK signature. We found that following AXL activation, PC9 cells display significantly higher FAK signaling with ACK1 and CDK1 being most dramatically activated (**Figure 5E**). We did not find transcriptional association with this pathway (not shown), suggesting that the regulation of the FAK signaling pathway by AXL is mediated through phosphorylation.

Finally, the dasatinib treatment data revealed an additional cluster of interest that displays an AXL-dependent increase of phosphosites (**Supplementary Figure 5B**). For example, ERK1 Y187-p and ERK2 Y204-p remain phosphorylated in PC9 WT cells but are strongly inhibited in AXL KO cells. This suggests that the AXL-mediated activation of ERK is independent of but affected by SFK and ABL1. Together, these results indicate that SFK and ABL1 inhibition impairs key downstream components of AXL signaling that correlate with cancer cell growth and migration.

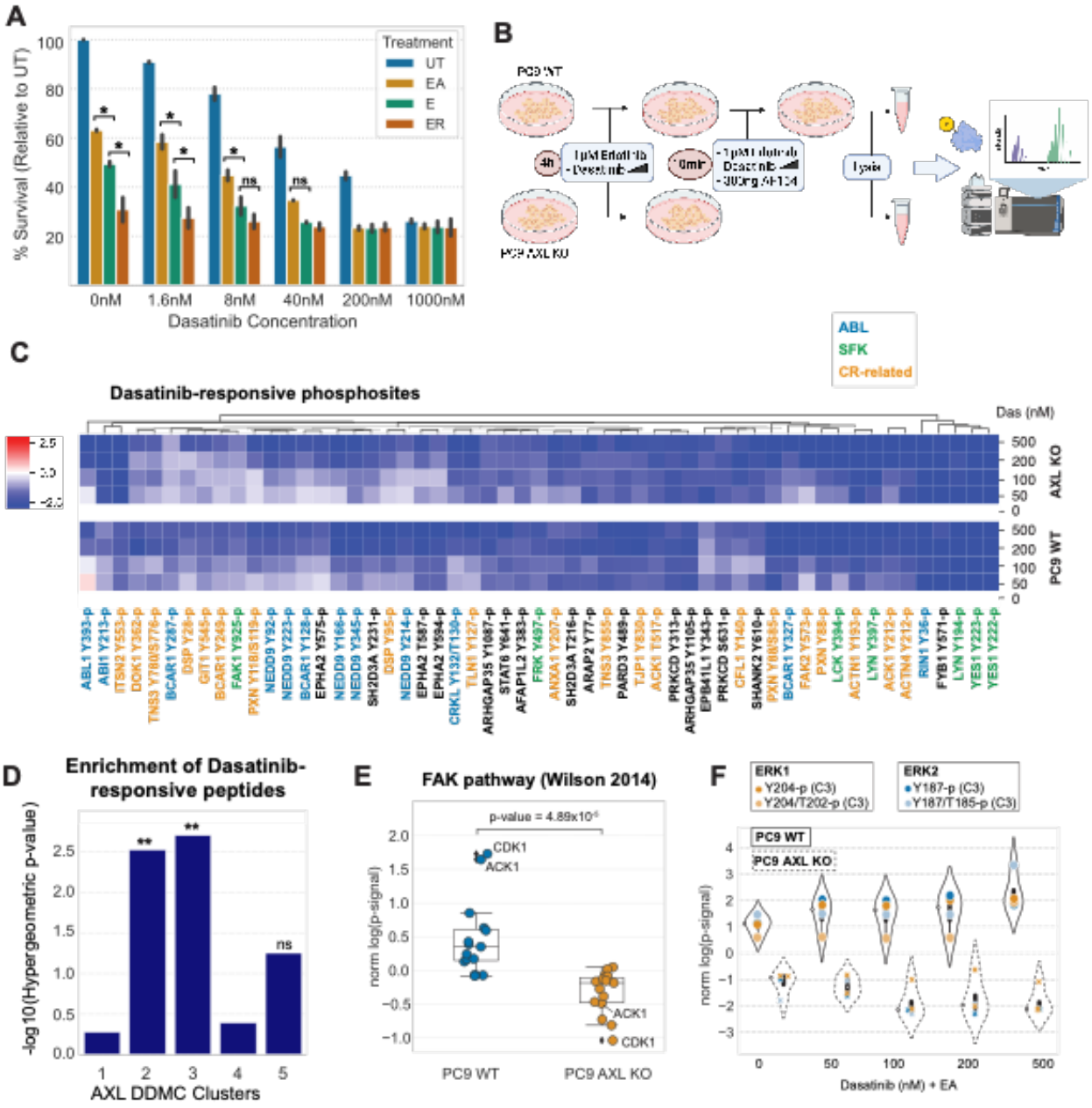


Figure 5. ABL1 and SFK regulate C2 and C3. (A) Cell confluency of PC9 WT cells exposed for 72 hours to the indicated treatments with increasing concentrations of dasatinib and relative to UT. (B) Diagram of the MS experiment; Cells were treated with E in addition to the indicated concentration of dasatinib for 4h and subsequently with AF154 for 10 minutes. Cells were then lysed and subjected to mass spectrometry (see Methods). (C) Heatmap showing the log phosphorylation signal of peptides normalized to 0nM of dasatinib per cell line. (D) Hypergeometric test to evaluate the enrichment of dasatinib-responsive peptides per DDMC cluster. (E) Phosphorylation signal of the FAK pathway signature in PC9 WT and AXL KO cells. Note that the signal of CDK1 Y15-p was multiplied by -1 since it is a known inhibitory site of its kinase activity. (F) Phosphorylation signal of phosphosites located in the

activation loop of Erk1/2 in PC9 WT and AXL KO cells treated with the indicated concentrations of dasatinib and EA.

AXL activates upstream kinase YAP regulators which in turn drive AXL expression and kinase activation

To extend our understanding of the mechanism by which the identified phosphoproteomic signaling changes affect AXL-driven phenotypes, we decided to explore the transcriptional changes occurring during switched AXL signaling. We collected bulk RNAseq data of each PC9 AXL Y-to-F cell line treated with E or EA and performed an enrichment. We run PCA analysis of the RNAseq data and found that PC1 represents AXL activation since the scores of all cell lines shift positively along PC 1 in EA-treated cells compared with E. We generated a ranked gene list based on the PC 1 scores to run a ranked GSEA analysis and found a YAP signature in AXL-activated cells (**Figure 6A** and **Supplementary Figure 6A**). This is consistent with the association between AXL and YAP in LUAD patients (**Figure 1B-E**) as well as with our DDMC upstream kinase predictions of C2 and C3, as the role of ABL1 and SFK phosphorylation of YAP is a well-known mechanism of YAP nuclear translocation and activation[59–62]. Moreover, several studies demonstrate the association between YAP activation and the development of drug resistance, including in the context of EGFR-targeted therapy in NSCLC cells[63–67]. For instance, a recent study shows that FAK signaling strongly activates YAP to enable the emergence of drug tolerant persister cells during EGFR-targeted therapies in patient-derived models and in clinical samples[68]. Together, these insights led us to hypothesize that AXL activates key upstream components of the YAP pathway—namely SFK, ABL1, and FAK signaling—to mediate the emergence of drug-tolerant persister cells.

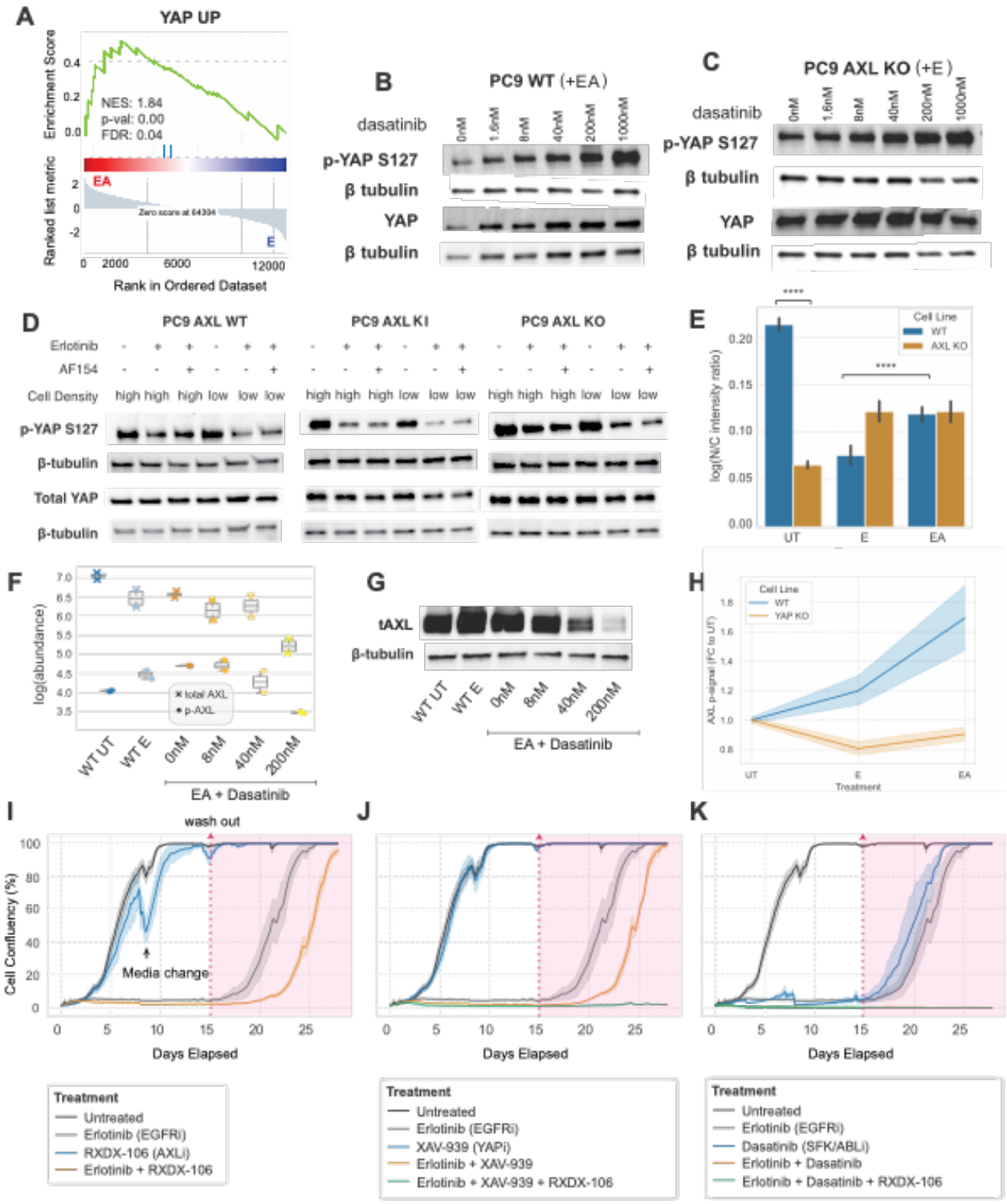


Figure 6. AXL promotes the activation and nuclear translocation of YAP which in turn regulates AXL expression and kinase activity. (A) Ranked GSEA analysis of the RNAseq data of the Y-to-F mutant cell lines ranked by the scores of a PCA analysis (see Supplementary Figure 6A). (B-C) Total and S126-p YAP levels of (B) PC9 WT and (C) AXL KO cells with increasing concentrations of dasatinib in addition to EA. (D) Total and S126-p YAP levels

of PC9 WT, AXL KI, and PC9 AXL KO cells seeded at high or low cell density and treated with E or EA. (E) YAP nuclear translocation ration of immunofluorescence YAP staining after 3 days of treatment with E or EA. Data was quantified using CellProfiler. (F) Total and pan-pY AXL signal in the indicated treatments. (G) Western blot of total AXL levels in the indicated treatments. (H) Total and pan-pY AXL levels in PC9 YAP KO cells. (I-K) Cell viability assay of PC9 WT cells treated with the indicated inhibitors for 15 days. Thereafter, treatment solutions were replaced with media and drug-tolerant persister cells were allowed to regrow for 15 days. Treatment or media were replaced every 3-5 days. *p-value < 0.05, **p-value < 0.001, ***p-value < 0.0001, ****p-value < 0.00001.

To test this hypothesis, we first treated cells with increasing concentrations of dasatinib and blotted for the inhibitory YAP phosphosite S126 which when phosphorylated, is bound by 14-3-3 proteins to sequester YAP in the cytoplasm and prevent its nuclear translocation[59]. Indeed, we found that dasatinib induces an increase in the phosphorylation abundance of the YAP S126-p which is further exacerbated in the absence of AXL**.** Interestingly, AXL KO cells present an upregulation of total YAP protein expression compared with PC9 WT cells which might be indicative of its dysregulation (**Figure 6C/D**).

To further validate the effect of switched AXL signaling on YAP activation, we measured YAP S127-p in PC9 WT, AXL KI and AXL KO cells treated them with E or EA for 24h. We also seeded cells at high or low density as YAP is known to respond to changes in cell confluency, with the transcription factor becoming inactive at higher cell densities[69]. As expected, we found that YAP S127-p most strongly decreases in response to E or EA at low cell density. Importantly, while AXL KO cells still display the same trend, the overall phosphorylation signal of S127-p is substantially higher in AXL KO compared to PC9 WT or AXL KI cells, again showing an AXL-specific YAP activation. Like we observed in PC9 cells treated with dasatinib, AXL KO cells displayed higher levels of total YAP across conditions (**Figure 6D**).

We then directly visualized YAP cellular localization in PC9 WT and AXL KO cells using immunofluorescence. In this experiment we treated cells with E and EA for 3 days to observe YAP localization dynamics as surviving cancer cells persist the inhibitory effects of treatment. PC9 cells treated with EA display significantly higher nuclear YAP than cells treated only with E. Moreover, while most AXL KO cells have cytoplasmic YAP, WT cells YAP displayed a much more nuclear localization. Consistent with the western blots of YAP S127-p, we observed that E administration promoted YAP nuclear translocation in AXL KO cells as well (**Figure 6E** and **Supplementary Figures 6B/C**).

Several studies indicate that AXL can be a target of YAP when in complex with the DNA-binding protein TEAD, thereby proposing that YAP acts upstream of AXL[63,70,71]. Nevertheless, others have reported that it is in fact AXL kinase activity which induces YAP activation[67]. A third hypothesis emerged when another study showed that AXL and the YAP homologue TAZ form a positive feedback loop to promote the development of lung cancer brain metastases[32]. Our results suggest that upon AXL activation, SFK and ABL1 kinases engage a network of cytoskeletal-remodeling, EMT-associated, signaling proteins that affect YAP translocation. To explore whether YAP inhibition influences AXL expression and activation, we quantified total and phosphorylated AXL in PC9 cells treated with dasatinib and found that both decrease with increasing dasatinib (**Figure 4F/G**). Moreover, PC9 YAP KO cells treated with E or EA and fail to activate AXL (**Figure 4H**). Together, this data indicates that AXL and YAP form a positive feedback loop.

We then investigated the influence of AXL signaling in the emergence of drug-tolerant persister cells by inhibiting AXL, SFK and Abl1, YAP, or Erk1/2, through RXDX-106, dasatinib, and

XAV-393 treatment, respectively, in PC9 WT cells for 15 days. Then, we replaced the treatment solutions with complete media and assessed the ability of resistant cells to regrow for 15 more days. We used RXDX-106 instead of R428 given the fact that RXDX-106 is extremely specific to TAM receptors and PC9 cells do not express neither MerTK nor Tyro3 and R428 has been shown to trigger cell death independent of AXL inhibition[72]. We found that whereas AXL or YAP inhibition delayed PC9 regrowth after washout, PC9 cells treated with E were unable to regrow (**Figure 6I-K**). All cell lines treated with E and dasatinib did not regrow. In total, this data supports the hypothesis that AXL and YAP form a positive feedback loop wherein AXL activation leads to the phosphorylation of key upstream components of YAP, which in turn induce its nuclear translocation and effector functions to sustain cancer cell growth.

AXL protects PC9 cells against erlotinib-induced DNA damage in part through CK2- and cell cycle-dependent repair mechanisms

DDMC infers that CK2 regulates C1 since the sequence motif of this cluster is highly acidophilic in its C-terminus which is a known determinant of CK2 kinase specificity. Thus, we sought to elucidate how CK2 influences AXL bypass signaling. Both AXL and CK2 have been previously reported to regulate DDR responses but have never been shown to cooperate in this capacity. AXL inhibition has been shown to downregulate homologous recombination which affects the fidelity of double strand break (DSB) repair[35,36]. On the other hand, CK2 is an established regulator of DSB during mitosis; ATM-dependent phosphorylation of H2AX recruits MDC1 which binds TOPBP1 and CIP2A via its CK2-mediated phosphorylation. Then, TOPBP1 and CIP2A form filamentous structures to tether both DSB ends together[73–75]. We hypothesized that AXL activation promotes to the ability of CK2 to repair DSB induced by E, thereby blocking DNA damage-induced apoptosis and enabling cancer cell survival.

We first asked whether AXL-activated PC9 cells treated with E display decreased DSB lesions. Since reactive oxygen species (ROS) induce DSB events, we monitored the amount of oxidative stress occurring in response to E while activating or inhibiting AXL with A or RXDX-106 (R), respectively. As expected, we observed a decrease in ROS production in EA- versus E-treated cells. Conversely, EGFR and AXL inhibition led to a sustained increase in oxidative stress, displaying a strong generation of ROS after 30 minutes. RXDX-106 alone also triggered an early induction of ROS production which subsequently decreased after 24 hours (**Figure 7A**). Accordingly, we observed many enzymes responsible for ROS detoxification, such as GPX, SOD1, GST, SOD1, or PYCR, upregulated in EA-treated PC9 cells and downregulated in AXL KO cells. HIF-1, also upregulated in AXL-activated cells and downregulated in AXL KO cells, has been shown to sustain cell survival during metastatic colonization by reducing cytotoxic ROS levels[76]. Finally, the lysosomal enzyme PPT1 which is upregulated in response to ROS production, is down- and upregulated in AXL-activated or AXL KO cells, respectively[77] (**Figure 7B**). Thus, this data shows that AXL activation leads to the detoxification of ROS through the activation of antioxidant proteins.

We then used immunofluorescence to stain for p-H2AX and directly quantify the amount of H2AX foci, a proxy for DSB lesions, present in PC9 cells treated under different conditions affecting AXL and CK2 kinase activity in addition to EGFR. As expected, we found a dramatic increase in H2AX foci in E-treated cells which were reduced when AXL is activated. Conversely, combined EGFR and AXL inhibition increased the amount of DDR compared with E. Importantly, we found that AXL activation failed to decrease the amount of H2AX foci induced by the treatment with

CK2 inhibitor CX-4945, indicating that the loss of CK2 kinase activity hinders the ability of AXL to repair E-induced DSB lesions (**Figure 7C**).

Next, we used the RNAseq data of the different AXL Y-to-F mutants to find transcriptional changes indicating an upregulation of DSB repair machinery upon AXL activation. To do so, we performed GSEA analysis using the KEGG 2021 gene set of the transcriptome of the different AXL mutant cell lines treated with E compared with EA. As expected, we found a strong enrichment of a myriad of processes related to DDR (**Figure 7D**). Several of these signatures are associated with the regulation of DDR during cell cycle checkpoints. Since CK2 plays a role in DSB repair during G2 to M checkpoint signaling, we specifically asked whether AXL activation specifically leads to an upregulation of genes involved in this process. We found that EA-treated cells upregulated HR and DSB repair machinery including MDC1, TOPBP1, and MDC1 (**Figure 7D/E**). Interestingly, we also observed an increased expression of CDK1/2, which is consistent with the increased kinase activity of CDK1 during bypass AXL signaling (**Figure 3F**). Consistently, a recent study showed that CDK1 phosphorylates and activates the major HR regulators Rad51 and Rad52 in the G2/M phase. These findings thus suggest that CK2, in cooperation with CDK1/2, promote HR at DSB lesions upon AXL activation (**Figure 7F**).

Given these results, we investigated the ability of PC9 cells to overcome EGFR inhibition in the presence or absence of CK2 activity. We treated PC9 WT or AXL KO cells with E, CX-4945 or a combination of E, CX-4945 or the AXL inhibitor RXDX-106 for 15 days and then replaced the treatment solutions with regular media to observe cancer cell regrowth for 15 days more. We found that CK2 inhibition alone did not affect cancer cell survival and that in combination with E, slightly delayed the emergence of drug-tolerant persister cells **(Figure 7G/H)**.

Conclusively, PC9 cells undergoing AXL bypass signaling display increased protection against DNA damage generated by EGFR targeted therapy. CK2 inhibition blocks this AXL-mediated protection which suggests that AXL requires CK2 to promote DSB repair. However, the addition of CX-4945 barely affected the capacity of PC9 cells to regrow after E treatment.

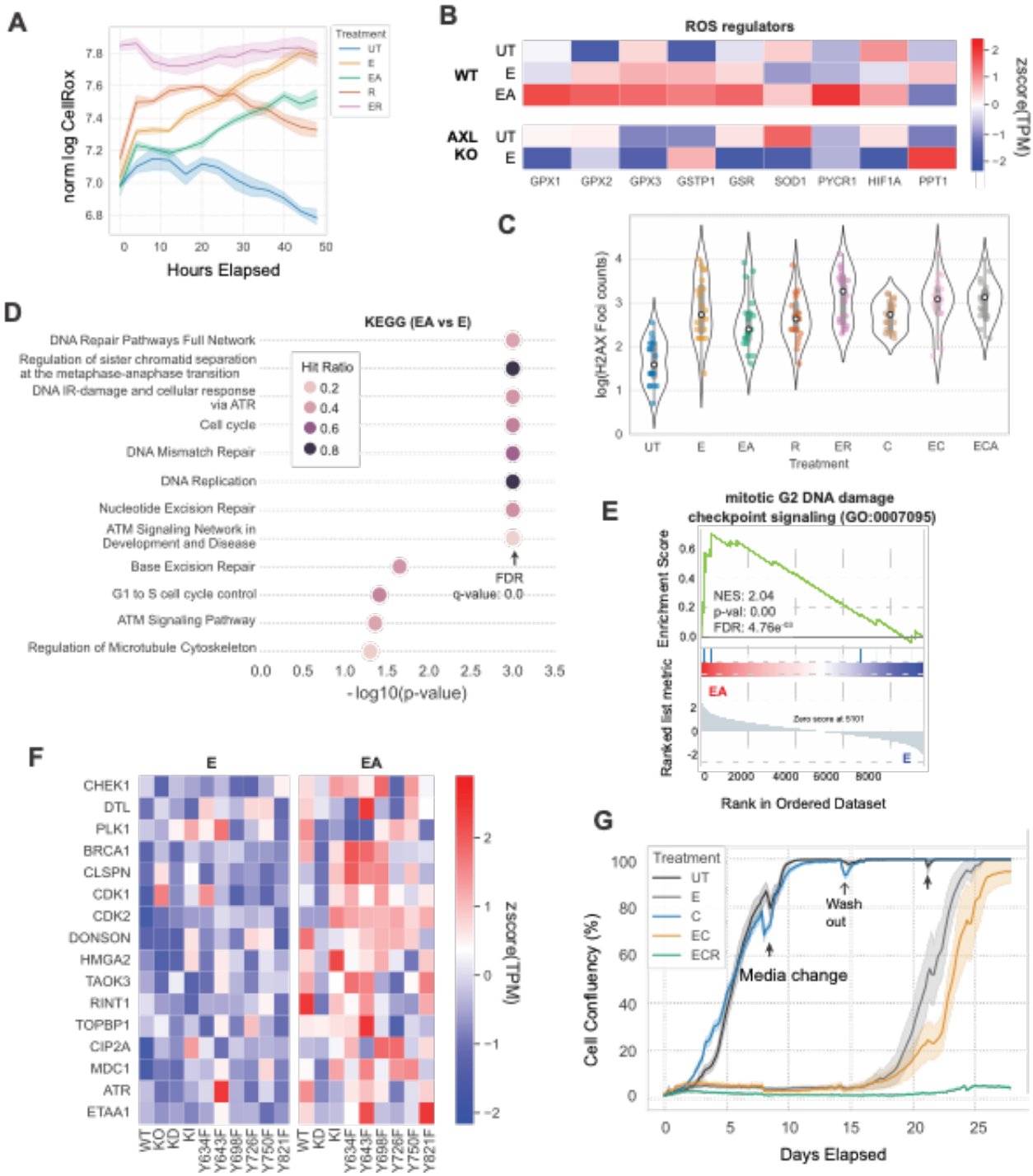


Figure 7. AXL promotes DSB repair through CK2 kinase activity. (A) ROS production normalized to cell number in the indicated treatments. (B) Heatmap of enzymes that regulate ROS levels. (C) Quantification of H2AX foci stained by immunofluorescence. The foci of 30 random cells were counted per condition. (D) GSEA analysis of EA- versus E-treated PC9 using the KEGG 2021 gene library. (E-F) GO analysis showing enrichment of (F) genes involved in mitotic G2 DNA damage checkpoint signaling in cells treated with EA compared with E. (G) Drug-tolerant persister cell assay in PC9 WT cells treated with E and/or CX-4945, or E, CX-4945 and RXDX-106.

Discussion

The reactivation of oncogenic pathways mediated by RTKs not targeted by therapy, referred as RTK bypass signaling, is a well-known resistance mechanism[5,7–9]. In fact, the FDA recently approved Amivantamab, a bispecific antibody directed against EGFR and MET for patients with advanced or metastatic EGFRm NSCLC (NCT04599712)[78]. Nevertheless, the downstream signaling pathways that bypass RTKs activate to mediate drug resistance remain largely unknown and its investigation could lead to (i) an increased mechanistic understanding of RTK inhibitors reaching the clinic and (ii) the identification of new drug targets. Several challenges intrinsic to the study of RTK downstream signaling such as RTK crosstalk or signaling pathway interdependency hinder the identification of RTK-specific phosphoproteomic and phenotypic changes during bypass resistance[4,5]. To overcome this limitation, we devised a combined computational and experimental approach by generating a panel of PC9 AXL Y-to-F mutant cell lines. Using these cell lines, we measured their phosphoproteomic and phenotypic changes during EGFR inhibition and AXL activation and applied multivariate modeling to identify the most prominent AXL-driven signaling pathways that promote cell fitness. This methodology ensures that any signaling, or phenotypic variation observed across cell lines is specific to AXL.

Our multivariate modeling identified three phosphosite clusters (C1, C2, and C3) defining an "AXL downstream signature", that correlates with cell viability and migration in vitro (**Figure**

3G/**H**). Moreover, the gene expression and phosphorylation signal of this downstream signature were significantly enriched in AXL-high EGFRm LUAD tumors, and C2 and C3, but not C1, correlated with poor treatment response, metastasis, and overall survival in LUAD patients. Interestingly, we observed that unlike its downstream signature, the gene expression of AXL did not correlate with poor clinical outcomes, suggesting that unlike its downstream signature or AXL protein abundance, AXL gene expression does not correlate with its kinase activity. This observation underscores that AXL gene expression might be insufficient to identify AXL-driven LUAD tumors resistant to EGFR-targeted therapies.

The Hippo pathway effector YAP is known to drive resistance to EGFR-targeted therapies in NSCLC[60,61,64,79]. Although YAP has been shown to upregulate AXL gene expression, other studies suggest that it is AXL which acts upstream of the transcription factor[63,67,70]. DDMC identified and we experimentally validated using dasatinib that ABL1 and SFK regulate C2 and C3, respectively, upstream of YAP (**Figures 3J and 5C**/**D**). Moreover, transcriptomic profiling during switched AXL activation revealed a strong enrichment of YAP targets in AXL-activated cells (**Figure 6A**). This is not surprising as YAP phosphorylation by SFK and ABL1 is known to promote its nuclear translocation and effector functions. Importantly, a recent study showed that a previously established FAK signaling signature activates YAP to mediate osimertinib resistance. EA-treated PC9 cells show a dramatic increase in this FAK signature compared with EA-treated AXL KO cells (**Figure 5E**). Genetic or dasatinib inhibition of YAP decreased AXL protein expression and its kinase activity (**Figure 6F-H**). Thus, our data indicates that AXL activates key upstream components of YAP, namely SFK, ABL1, and FAK, and in turn, the transcription factor regulates AXL expression and allows its kinase activation.

Resistant cells to single-agent osimertinib can emerge through both Erk1/2 reactivation or YAP activation, whereas YAP activation becomes the dominant resistance mechanism of cells overcoming combined EGFR and MEK inhibition[65]. Here we show that AXL can mediate both Erk1/2 reactivation and YAP activation during bypass signaling. Intriguingly, here we show that YAP inhibition through dasatinib treatment in EA-treated PC9 cells promote Erk1/2 activation whereas dasatinib inhibited Erk1/2 in AXL KO cells. This suggests that AXL might play a role in enabling cancer cells to rely in either Erk1/2 or YAP to drive drug resistance (**Figure 5F**).

In addition to YAP activation, transcriptomic profiling of the AXL mutant cell lines revealed a strong enrichment in signatures related to DDR pathways. Here we show that AXL protects cells against E-induced DDR by two ways: First, by upregulating antioxidant proteins and enzymes, thereby reducing the amount of ROS production and second, by regulating a mitotic G2 DNA damage checkpoint through the upstream regulator of C1 CK2 as well as with CDK1/2. However, combined EGFR and CK2 inhibition barely affected the emergence of drug-tolerant persister cells (**Figure 7G**), which suggests that this pathway is not required for AXL to provide resistance to anti-EGFR therapy.

In total, this work demonstrates that systematically leveraging Y-to-F mutational studies and combining phosphoproteomic with phenotypic experiments allows dissecting pleiotropic signaling regulators to identify various mechanisms by which AXL drives resistance and collateral phenotypes in lung cancer (**Figure 8**).

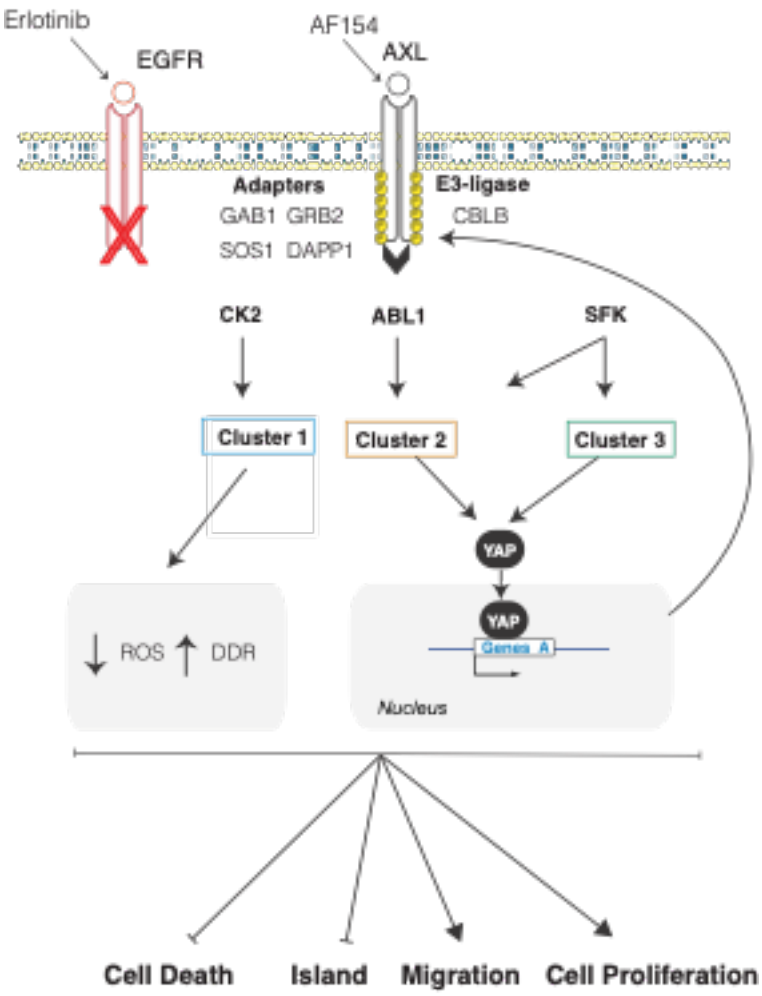


Figure 8. Downstream signaling of AXL bypass resistance to EGFR-targeted therapies in lung cancer. AXL activates ABL1 and SFK that, through phosphorylation of proteins within C2 and C3, promote the nuclear translocation and activation of YAP which in turn upregulated its gene targets, including AXL, and drives cell proliferation and migration. Additionally, AXL activation leads to increased protection against E-induced DNA damage through CK2 kinase activity and the upregulation of antioxidant enzymes..

Methods

All analysis was implemented in Python and can be found at https://github.com/meyer-lab/resistance-MS.

Antibody reagents, inhibitors, and cell culture

Erlotinib (LC Laboratories), RXDX-106 (Selleck Chemicals), R428 (Fisher Scientific), XAV-393 (Selleck Chemicals) were used at 1 μM, dasatinib (LC Laboratories) at 200nM, and CX-4945 and trametinib (both from MedChem Express) were used at 4 μM and 30 nM, respectively. The AXL-activating antibody AF154 (R&D Systems) was used at 300 ng/mL. AXL, YAP, YAP S126-p were purchased from CST and rhodamine-conjugated β-tubulin from Bio-Rad and used for Western blotting. ELISA-based signaling measurements were performed using according to the manufacturer's instructions (Bio-Rad). The Luminex kits EGFR Y1068-p and p-AKT is S473-p were obtained from Bio-Rad. The capture AXL antibody was generated by conjugating a primary AXL antibody (R&D systems MAB154) with magnetic beads (Bio-Rad). For total AXL detection, a biotinylated AXL antibody (R&D systems BAF154) was used and for p-AXL measurements a pan-tyrosine antibody (R&D systems) was used. The primary antibodies YAP (Santa Cruz), p-H2AX (CST), AXL (Abcam), diluted at 1:50, 1:200, and 1:1000, respectively, and the secondary antibodies Alexa Fluor 488 goat anti-rabbit IgG (Invitrogen) and PE goat anti-mouse IgG (Invitrogen), diluted at 1:500 and 1:50, respectively, were used for immunofluorescence as described in[47]. PC9 (Sigma Aldrich) cells and all derivatives were grown in RPMI-1640 media supplemented with 10% fetal bovine serum (FBS) and penicillin/streptomycin. HEK293T cells were grown in DMEM supplemented with 10% FBS and 1% GlutaMAX (Thermo Fisher Scientific). PC9 YAP KO cells were obtained from Passi Jänne's laboratory at Harvard Medical School. The oxidative stress assay was performed using CellROX™ Deep Red Reagent according to the manufacturer's instructions.

Generation of PC9 AXL Y-to-F mutant cell lines

The PC9 AXL KO cell line was generated by transfecting cells with a CRISPR/Cas9 and GFP vector containing a gRNA targeting the AXL kinase domain. The gRNA sequence, cell culturing,

and sorting methods have been previously described[80]. Plasmids containing the AXL phosphosite mutations were generated from an AXL-IRES-Puro vector (Addgene #65627) using site directed mutagenesis. Each mutant was then inserted into a lentiviral vector with a puromycin resistance marker (Addgene #17448).

For viral packaging, HEK 293T cells were seeded at 4.5 x 10^6 per 10 cm dish. After 24 hours, the lentiviral AXL expression vector, VSV-G envelope vector, and packaging vector (Addgene #12259 and #12260 respectively) were combined in a 10:1:10 mass ratio and diluted in Opti-MEM (Thermo Fisher Scientific). TransIT-LT1 (Mirus Bio) was added dropwise, and the solution was mixed gently by swirling and incubated at room temperature for 20 minutes. The solution was then added dropwise to the cells. After 18 hours, transfection media was replaced by media supplemented with 1% BSA fraction V (Thermo Fisher Scientific). Cells were incubated for 24 hours, after which the virus-containing media was removed and stored at 4°C. The media was replaced, and the cells incubated a further 24 hours to generate a second batch of viral media. The harvested batches were then pooled, filtered through a 0.45μm PVDF membrane to remove packaging cells, and flash frozen followed by storage at -80ºC until use.

PC9 AXL KO cells were seeded at 1.5 x 10^5 cells per well with antibiotic-free media in a 6-well plate and incubated for 24 hours. The cells were then infected with viral particles in antibiotic-free media supplemented with polybrene (MilliporeSigma). After 18 hours, the media was replaced with fresh antibiotic-free media. Cells were observed for a GFP positive population and then passaged into a 10 cm plate until confluent. The virally transduced cells were then sorted for based on GFP expression using a BD FACSAria cell sorter. The mutant cell populations were subcultured for later experiments.

Cell Viability and Apoptosis Assays

Cells were seeded in a 96-well plate at a density of 1.05 x 10^3 cells per well. After 24 hours, treatments were added in media containing 300 nm YOYO-3 (Thermo Fisher Scientific). Cells were cultured and imaged every 3 hours using an IncuCyte S3 (Essen Bioscience) at 20x magnification with 9 images per well. The phase, green, and red channels were manually thresholded and then analyzed by IncuCyte S3 software (Essen Bioscience) to determine cell counts and fraction of area covered.

Cell Migration Assay

96-well IncuCyte ImageLock plates (Essen Bioscience) were coated with a Collagen-I solution (Thermo Fisher Scientific), washed twice, and then seeded with 4 x 10^4 cells per well. After a 4-hour incubation, cells were wounded using the IncuCyte WoundMaker, washed twice to remove detached cells, and then treated with respective conditions. Images of the center of the wound were taken every 2 hours at a magnification of 10x, one image per well. The phase and green channels were thresholded and analyzed as above to determine migration measurements.

Cell Island Effect

Phase contrast images used for the cell island measurements were taken from image sets gathered in the cell viability assay. For endpoint readings, images at the 48-hour post-treatment time point were used. Representative images were chosen across experimental replicates. Images were opened in ImageJ and the center of each cell was manually marked. Dead cells, identified using YOYO-3 based fluorescence, were not marked. The 2D coordinates of all cell centers in an image were then exported for analysis. The amount of clustering present in a particular image was then

measured by applying Ripley's K function to the set of coordinates. The implementation of Ripley's K function used was taken from the astropy Python package[49].

Preparation of Cell Lysates for mass spectrometry

Cell lines were grown to confluence in 10 cm dishes over the course of 72-96 hours, washed, and treated by addition of media containing 1 μM erlotinib. Cells were incubated for 4 hours at 37ºC and then additionally treated with media containing 1 μM erlotinib and 300 ng/mL AXL activating antibody for 10 minutes. The cells were then placed immediately on ice, washed with ice-cold phosphate-buffered saline, and lysed with cold 8 M urea containing Phosphatase Inhibitor Cocktail I and Protease Inhibitor Cocktail I (Boston BioProducts). The lysates were then centrifuged at 20,000xg, 4℃ to pellet cell debris, and the supernatants removed and stored at -80℃. A bicinchoninic acid (BCA) protein concentration assay (Pierce) was performed according to the manufacturer's protocol to estimate the protein concentration in each lysate. Cell lysates were reduced with 10 mM DTT for 1 hour at 56℃, alkylated with 55 mM iodoacetamide for 1 hour at RT shielded from light, and diluted 5-fold with 100 mM ammonium acetate, pH 8.9, before trypsin (Promega) was added (20:1 protein:enzyme ratio) for overnight digestion at RT. The resulting solutions were acidified with 1 mL of acetic acid (HOAc) and loaded onto C18 Sep-Pak Plus Cartridges (Waters), rinsed with 10 mL of 0.1% HOAc, and eluted with 10 mL of 40% Acetonitrile (MeCN)/0.1% HOAc. Peptides were divided into 200 aliquots, and sample volume was reduced using a vacuum centrifuge (Thermo) and then lyophilized to dryness for storage at -80℃. TMT labeling for multiplexed analysis was performed according to manufacturer's protocol. Samples, each containing ~200 μg peptides, were resuspended in 35 μL HEPES (pH 8.5), vortexed, and spun down at 13,400 rpm for 1 minute. 400 μg of a given channel of TMT10plex (Thermo) in anhydrous MeCN, was added per sample. Samples were shaken at 400 rpm for 1 hour, after which

the labeling reaction was quenched using 5% hydroxylamine (50%, Thermo). After another 15 minutes on the shaker, all samples were combined using the same pipette tip to reduce sample loss, and sample aliquots were washed twice with 40 µL 25% MeCN/0.1% HOAc which was added to the collection tube to improve yield. Sample volume was reduced using a vacuum centifuge and then lyophilized to dryness for storage at -80ºC.

Phosphopeptide enrichment

Immunoprecipitation (IP) and IMAC were used sequentially to enrich samples for phosphotyrosine containing peptides. TMT-labeled samples were incubated in IP buffer consisting of 1% Nonidet P-40 with protein G agarose beads conjugated to 24 µg of 4G10 V312 IgG and 6 µg of PT-66 (P3300, Sigma) overnight at 4ºC. Peptides were eluted with 25 µL of 0.2% trifluoroacetic acid for 10 minutes at room temperature; this elution was performed twice to improve yield. Eluted peptides were subjected to phosphopeptide enrichment using immobilized metal affinity chromatography (IMAC)-based Fe-NTA spin column to reduce non-specific, non-phosphorylated peptide background. High-Select Fe-NTA enrichment kit (Pierce) was used according to manufacturer's instructions with following modifications. Eluted peptides from IP were incubated with Fe-NTA beads containing 25 µL binding washing buffer for 30 minutes. Peptides were eluted twice with 20 mL of elution buffer into a 1.7 mL microcentrifuge tube. Eluates were concentrated in speed-vac until ~1 µL of sample remained, and then resuspended in 10µL of 5% acetonitrile in 0.1% formic acid. Samples were loaded directly onto an in-house constructed fused silica capillary column [50 µm inner diameter (ID) x 10 cm] packed with 5 µm C18 beads (YMC gel, ODS-AQ, AQ12S05) and with an integrated electrospray ionization tip (~2 µm tip ID).

LC-MS/MS analysis

LC-MS/MS of pTyr peptides were carried out on an Agilent 1260 LC coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific). Peptides were separated using a 140-minute gradient with 70% acetonitrile in 0.2 mol/L acetic acid at flow rate of 0.2 mL/minute with approximate split flow of 20 nL/minute. The mass spectrometer was operated in data-dependent acquisition with following settings for MS1 scans: m/z range: 350 to 2,000; resolution: 60,000; AGC target: 3 x 10^6; maximum injection time (maxIT): 50 ms. The top 15 abundant ions were isolated and fragmented by higher energy collision dissociation with following settings: resolution: 60,000; AGC target: 1x10^5; maxIT: 350 ms; isolation width: 0.4 m/z, collisional energy (CE): 33%, dynamic exclusion: 20 seconds. Crude peptide analysis was performed on a Q Exactive Plus mass spectrometer to correct for small variation in peptide loadings for each of the TMT channels. Approximately 30 ng of the supernatant from pTyr IP was loaded onto an in-house packed precolumn (100 μm ID x 10 cm) packed with 10 mm C18 beads (YMC gel, ODS-A, AA12S11) and analyzed with a 70-minute LC gradient. MS1 scans were per-formed at following settings: m/z range: 350 to 2,000; resolution: 70,000; AGC target: 3x10^6; maxIT: 50 ms. The top 10 abundant ions were isolated and fragmented with CE of 33% at a resolution of 35,000.

Peptide identification and quantification

Mass spectra were processed with Proteome Discoverer version 2.5 (Thermo Fisher Scientific) and searched against the human SwissProt database using Mascot version 2.4 (MatrixScience, RRID:SCR_014322). MS/MS spectra were searched with mass tolerance of 10 ppm for precursor ions and 20 mmu for fragment ions. Cysteine carbamidomethylation, TMT-labeled lysine, and TMT-labeled peptide N-termini were set as fixed modifications. Oxidation of methionine and phosphorylation of serine, threonine and tyrosine were searched as dynamic modifications. TMT reporter quantification was extracted, and isotope corrected in Proteome Discoverer. Peptide

spectrum matches (PSM) were filtered according to following para-meters: rank=1, mascot ion score>15, isolation interference<40%, average TMT signal>1,000. Peptides with missing values across any channel were filtered out.

Preprocessing and clustering of phosphoproteomic data

We performed three phosphoproteomic biological replicates of the PC9 AXL Y-to-F mutants treated with EA. The three data sets were concatenated, mean-centered across cell lines, and log2-transformed. To discard phosphosites whose measurements were not reproducible among replicates, all recurrent phosphosites in two or three biological replicates were identified. For those appearing in two biological replicates, any peptides showing a Pearson correlation coefficient smaller than 0.55 were filtered out, whereas for those appearing in all three biological replicates, peptides with standard deviations of 0.5 were discarded. Moreover, to discard any unchanging peptides across cell lines, we filtered any phosphosites out containing a 0.5 cutoff change of the fold-change maximum versus the minimum signal for every cell line. All overlapping peptides were then averaged. The resulting preprocessed phosphoproteomic data set was then fit to Dual Data-Motif Clustering (DDMC) using 5 clusters and a sequence weight of 2.

Prediction of AXL-mediated phenotypes using PLSR

To predict the AXL-mediated phenotypes, a 4-component PLSR model (scikit-learn) was used with the DDMC cluster centers. To assess the predictive performance of PLSR leave-one-out cross-validation was applied wherein the model was trained using the paired cluster centers and phenotypic measurements in all cell lines except one, the cluster centers of that remaining cell line was used to predict its phenotypic measurements and a mean squared error between the predicted and actual value per phenotype was computed. This process was iterated across all cell lines to

obtain a final r score value per phenotype. To benchmark the ability of PLSR to predict the phenotypes using the DDMC clusters, we additionally fit PLSR using either the unclustered phosphoproteomic data set directly, k-means clustering, GMM, DDMC using only the peptide sequence information, or the selected DDMC model combining the phosphorylation and sequence information (referred as DDMC mix). All cluster methods were used with 5 clusters and the number of PLSR components used was optimized for each case. We then applied leave-one-out cross-validation fitting the cluster centers generated with each method to PLSR.

RNAseq sample preparation and sequencing

To generate the RNAseq data, 300.000 cells of each PC9 AXL Y-to-F mutant cell line were seeded in 100 mm dishes. The next day were treated with E or EA and after 24 hours of treatment, cells were lysed with RIPA and RNA was extracted using the RNeasy Mini Kit (Qiagen) and RNA sequencing and read alignment and processing was performed by Novogene. Any genes with less than 10 TPM were filtered out. Ranked or standard GSEA were implemented in python (gseapy).

Clinical data and analysis

Bulk RNAseq, proteomic, and phosphoproteomic data of LUAD patients was obtained from the CTPAC LUAD study[45]. scRNAseq of LUAD patients including cell type and clinical annotations were obtained from Maynard et al[81] which was analyzed using python's package scanpy. The Kaplan-Meier curves displaying the overall survival of LUAD and PAAD patients according to the gene expression of the AXL downstream signature were generated using the web server GEPIA2[82] which uses TCGA/GTEx data.

Drug-tolerant persister cell assay

1000 cells were seeded in a 96 well plate and treated with the indicated treatments and cell lines In four technical replicates per treatment. After 15 days of treatment, the treatment solutions were replaced with complete RPMI-1640 media for additional 15 days to allow drug-tolerant persister cells to regrow. Cells were imaged every 4 hours using an IncuCyte S3 (Essen Bioscience) at 10x magnification with 4 images per well. The phase were manually thresholded and then analyzed by IncuCyte S3 software (Essen Bioscience) to determine cell confluency.

References

1. Atlanta, G. Cancer Facts & Figures 2022. *American Cancer Society* (2022).

2. Hospital, X. *et al.* Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *The Lancet* **12**, 735–777 (2011).

3. Ramalingam, S. S. *et al.* Overall Survival with Osimertinib in Untreated, EGFR -Mutated Advanced NSCLC. *New England Journal of Medicine* **382**, 41–50 (2020).

4. Manole, S., Richards, E. J. & Meyer, A. S. JNK pathway activation modulates acquired resistance to EGFR/HER2-targeted therapies. *Cancer Res* **76**, 5219–5228 (2016).

5. Wilson, T. R. *et al.* Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors. *Nature* **487**, 505–509 (2012).

6. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–34 (2010).

7. Turke, A. B. *et al.* Preexistence and Clonal Selection of MET Amplification in EGFR Mutant NSCLC. *Cancer Cell* **17**, 77–88 (2010).

8. Tao, J. J. *et al.* Antagonism of EGFR and HER3 Enhances the Response to Inhibitors of the PI3K-Akt Pathway in Triple-Negative Breast Cancer. *Sci Signal* (2014).

9. Wilson, T. R., Lee, D. Y., Berry, L., Shames, D. S. & Settleman, J. Neuregulin-1-Mediated Autocrine Signaling Underlies Sensitivity to HER2 Kinase Inhibitors in a Subset of Human Cancers. *Cancer Cell* **20**, 158–172 (2011).

10. Aguilera, T. A. *et al.* Reprogramming the immunological microenvironment through radiation and targeting Axl. *Nat Commun* **7**, 1–14 (2016).

11. Okimoto, R. A. & Bivona, T. G. AXL receptor tyrosine kinase as a therapeutic target in NSCLC. *Lung Cancer: Targets and Therapy* **6**, 27–34 (2015).

12. Chouaib, S. *et al. AXL targeting enhances lymphocyte-mediated cytotoxicity of lung cancer cells*.

13. Aaron S. Meyer, Miller A. Miller, Frank B Gertler & Douglas A Luaffenburger. The Receptor AXL Diversifies EGFR Signaling and Limits the Response to EGFR-Targeted. **6**, (2014).

14. Lotsberg, M. L. *et al.* AXL targeting abrogates autophagic flux and induces immunogenic cell death in drug resistant cancer cells. *Journal of Thoracic Oncology* (2020) doi:10.1016/j.jtho.2020.01.015.

15. Zhang, Z. *et al.* Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. *Nat Genet* **44**, 852–860 (2012).

16. Schoumacher, M. & Burbridge, M. Key Roles of AXL and MER Receptor Tyrosine Kinases in Resistance to Multiple Anticancer Therapies. *Curr Oncol Rep* **19**, (2017).

17. Terry, S. *et al.* AXL targeting overcomes human lung cancer cell resistance to NK- And CTL-mediated cytotoxicity. *Cancer Immunol Res* **7**, 1789–1802 (2019).

18. Elkabets, M. *et al.* AXL mediates resistance to PI3Kα inhibition by activating the EGFR/PKC/mTOR Axis in Head and neck and esophageal squamous cell carcinomas. *Cancer Cell* **27**, 533–546 (2015).

19. Goyette, M. A. *et al.* The Receptor Tyrosine Kinase AXL Is Required at Multiple Steps of the Metastatic Cascade during HER2-Positive Breast Cancer Progression. *Cell Rep* **23**, 1476–1490 (2018).

20. Guo, Z., Li, Y., Zhang, D. & Ma, J. Axl inhibition induces the antitumor immune response which can be further potentiated by PD-1 blockade in the mouse cancer models. *Oncotarget* **8**, 89761–89774 (2017).

21. Dunne, P. D. *et al.* AXL is a key regulator of inherent and chemotherapy-induced invasion and predicts a poor clinical outcome in early-stage colon cancer. *Clinical Cancer Research* **20**, 164–175 (2014).

22. Antony, J. & Huang, R. Y. J. AXL-driven EMT state as a targetable conduit in cancer. *Cancer Res* **77**, 3725–3732 (2017).

23. Gjerdrum, C. *et al.* Axl is an essential epithelial-to-mesenchymal transition-induced regulator of breast cancer metastasis and patient survival. *Proc Natl Acad Sci U S A* **107**, 1124–1129 (2010).

24. Byers, L. A. *et al.* An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. *Clinical Cancer Research* **19**, 279–290 (2013).

25. Kirane, A. *et al.* Warfarin Blocks Gas6-Mediated Axl Activation Required for Pancreatic Cancer Epithelial Plasticity and Metastasis. (2015) doi:10.1158/0008-5472.CAN-14-2887-T.

26. Pénzes, K. *et al.* Combined inhibition of AXL, Lyn and p130Cas kinases block migration of triple negative breast cancer cells. *Cancer Biol Ther* **15**, 1571–1582 (2014).

27. Abu-Thuraia, A. *et al.* AXL confers cell migration and invasion by hijacking a PEAK1-regulated focal adhesion protein network. *Nat Commun* **11**, (2020).

28. Majumder, A. *et al.* Integrated Proteomics-Based Physical and Functional Mapping of AXL Kinase Signaling Pathways and Inhibitors Define Its Role in Cell Migration. *Molecular Cancer Research* 1–14 (2022) doi:10.1158/1541-7786.mcr-21-0275.

29. Taniguchi, H. *et al.* AXL confers intrinsic resistance to osimertinib and advances the emergence of tolerant cells. *Nat Commun* **10**, 2–15 (2019).

30. Noronha, A. *et al.* AXL and Error-Prone DNA Replication Confer Drug Resistance and Offer Strategies to Treat EGFR-Mutant Lung Cancer NOVEMBER 2022 CANCER DISCOVERY | OF2. doi:10.1158/2159-8290.CD-22-0111/3207554/cd-22-0111.pdf.

31. Vasan, N., Baselga, J. & Hyman, D. M. A view on drug resistance in cancer. *Nature* **575**, 299–309 (2019).

32. Hoj, J. P., Mayro, B. & Pendergast, A. M. A TAZ-AXL-ABL2 Feed-Forward Signaling Axis Promotes Lung Adenocarcinoma Brain Metastasis. *Cell Rep* **29**, 3421-3434.e8 (2019).

33. Woo, S. M. *et al.* Axl inhibitor R428 enhances TRAIL-mediated apoptosis through downregulation of c-FLIP and survivin expression in renal carcinoma. *Int J Mol Sci* **20**, (2019).

34. Axelrod, H. & Pienta, K. J. Axl as a mediator of cellular growth and survival. *Oncotarget* **5**, 8818–8852 (2014).

35. Balaji, K. *et al.* AXL inhibition suppresses the DNA damage response and sensitizes cells to PARP inhibition in multiple cancers. *Molecular Cancer Research* **15**, 45–58 (2017).

36. Ramkumar, K. *et al.* AXL inhibition induces DNA damage and replication stress in non-small cell lung cancer cells and promotes sensitivity to ATR inhibitors. *Molecular Cancer Research* molcanres.0414.2020 (2020) doi:10.1158/1541-7786.mcr-20-0414.

37. Zdzalik-Bielecka, D. *et al.* The GAS6-AXL signaling pathway triggers actin remodeling that drives membrane ruffling, macropinocytosis, and cancer-cell invasion. *Proc Natl Acad Sci U S A* **118**, (2021).

38. Yokoyama, Y. *et al.* Tumor Biology and Immunology Immuno-oncological Efficacy of RXDX-106, a Novel TAM (TYRO3, AXL, MER) Family Small-Molecule Kinase Inhibitor. (2019) doi:10.1158/0008-5472.CAN-18-2022.

39. Earp, S., Matsushima, G. K., Seitz, H. M., Camenisch, T. D. & Lemke, G. Clearance of Apoptotic Cells Different Axl/Mertk/Tyro3 Receptors in Macrophages and Dendritic Cells Use. *J Immunol References* **178**, 5635–5642 (2007).

40. Filippakopoulos, P., Müller, S. & Knapp, S. SH2 domains: modulators of nonreceptor tyrosine kinase activity. *Curr Opin Struct Biol* **19**, 643–649 (2009).

41. McDaniel, N. K. *et al.* AXL Mediates Cetuximab and Radiation Resistance Through Tyrosine 821 and the c-ABL Kinase Pathway in Head and Neck Cancer. *Clin Cancer Res* **26**, 4349–4359 (2020).

42. Fridell, Y.-W. C. *et al.* Differential Activation of the Ras/Extracellular-Signal-Regulated Protein Kinase Pathway Is Responsible for the Biological Consequences Induced by the Axl Receptor Tyrosine Kinase. *Mol Cell Biol* **16**, 135–145 (1996).

43. Mark, M. R. *et al.* rse, A novel receptor-type tyrosine kinase with homology to Ax1/Ufo, is expressed at high levels in the brain. *Journal of Biological Chemistry* **269**, 10720–10728 (1994).

44. Braunger, J. *et al.* Intracellular signaling of the Ufo/Axl receptor tyrosine kinase is mediated mainly by a multi-substrate docking-site. *Oncogene* **14**, 2619–2631 (1997).

45. Gillette, M. A. *et al.* Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. *Cell* **182**, 200-225.e35 (2020).

46. Cordenonsi, M. *et al.* The hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. *Cell* **147**, 759–772 (2011).

47. Meyer, A. S., Zweemer, A. J. M. & Lauffenburger, D. A. The AXL Receptor Is a Sensor of Ligand Spatial Heterogeneity. *Cell Syst* **1**, 25–36 (2015).

48. Ripley, B. D. The second-order analysis of stationary point processes. *J Appl Probab* **13**, 255–266 (1976).

49. The Astropy Collaboration *et al.* The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package. (2022) doi:10.3847/1538-4357/ac7c74.

50. Creixell, M. & Meyer, A. S. Dual data and motif clustering improves the modeling and interpretation of phosphoproteomic data. *Cell Reports Methods* 100167 (2022) doi:10.1016/j.crmeth.2022.100167.

51. Adam-Artigues, A. *et al. Targeting HER2-AXL heterodimerization to overcome resistance to HER2 blockade in breast cancer*. *Sci. Adv* vol. 8 https://www.science.org (2022).

52. Kuo, M. T. *et al.* Collaboration Between RSK-EphA2 and Gas6-Axl RTK Signaling in Arginine Starvation Response That Confers Resistance to EGFR Inhibitors. *Transl Oncol* **13**, 355–364 (2020).

53. Meggio, F., Marin, O. & Pinna, L. A. Substrate specificity of protein kinase CK2. *Cell Mol Biol Res* **40**, 401–9 (1994).

54. Wang, C. *et al.* Determination of CK2 Specificity and Substrates by Proteome-Derived Peptide Libraries. *J Proteome Res* **12**, 3813–3821 (2013).

55. Kim, S., Ham, S., Yang, K. & Kim, K. Protein kinase CK2 activation is required for transforming growth factor β-induced epithelial–mesenchymal transition. *Mol Oncol* **12**, 1811–1826 (2018).

56. Hornbeck, P. v. *et al.* 15 years of PhosphoSitePlus ®: Integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res* **47**, D433–D441 (2019).

57. Zhang, Y. *et al.* AXL Inhibitor TP-0903 Reduces Metastasis and Therapy Resistance in Pancreatic Cancer. *Mol Cancer Ther* **21**, 38–47 (2022).

58. Wilson, C. *et al.* Overcoming EMT-associated resistance to anti-cancer drugs via Src/FAK pathway inhibition. *Oncotarget* **5**, 7328–7341 (2014).

59. Sugihara, T. *et al.* YAP tyrosine phosphorylation and nuclear localization in cholangiocarcinoma cells are regulated by LCK and independent of LATS activity. *Molecular Cancer Research* **16**, 1556–1567 (2018).

60. Zanconato, F., Cordenonsi, M. & Piccolo, S. YAP/TAZ at the Roots of Cancer. *Cancer Cell* **29**, 783–803 (2016).

61. Nguyen, C. D. K. & Yi, C. YAP/TAZ Signaling and Resistance to Cancer Therapy. *Trends Cancer* **5**, 283–296 (2019).

62. Levy, D., Adamovich, Y., Reuven, N. & Shaul, Y. Article Yap1 Phosphorylation by c-Abl Is a Critical Step in Selective Activation of Proapoptotic Genes in Response to DNA Damage. doi:10.1016/j.molcel.2007.12.022.

63. Ghiso, E. *et al.* YAP-Dependent AXL Overexpression Mediates Resistance to EGFR Inhibitors in NSCLC. *Neoplasia* **19**, 1012–1021 (2017).

64. Park, H. S. *et al.* Targeting YAP-p62 signaling axis suppresses the EGFR-TKI-resistant lung adenocarcinoma. *Cancer Med* **10**, 1405–1417 (2021).

65. Kurppa, K. J. *et al.* Treatment-Induced Tumor Dormancy through YAP-Mediated Transcriptional Reprogramming of the Apoptotic Pathway. *Cancer Cell* **37**, 104-122.e12 (2020).

66. Miao, J. *et al.* YAP regulates PD-L1 expression in human NSCLC cells. *Oncotarget* **8**, 114576–114587 (2017).

67. Saab, S., Chang, O. S.-S., Nagaoka, K., Hung, M.-C. & Yamaguchi, H. *The potential role of YAP in Axl-mediated resistance to EGFR tyrosine kinase inhibitors*. *Am J Cancer Res* vol. 9 www.ajcr.us/ (2019).

68. Haderk, F. *et al.* A focal adhesion kinase-YAP signaling axis drives drug tolerant persister cells and residual disease in lung cancer 2 3. doi:10.1101/2021.10.23.465573.

69. Zhao, B. *et al.* Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes Dev* **21**, 2747–2761 (2007).

70. Guo, Y. *et al.* CK2-induced cooperation of HHEX with the YAP-TEAD4 complex promotes colorectal tumorigenesis. *Nat Commun* **13**, 4995 (2022).

71. Chang, L. *et al.* The SWI/SNF complex is a mechanoregulated inhibitor of YAP and TAZ. *Nature* **563**, 265–269 (2018).

72. Chen, F., Song, Q. & Yu, Q. Axl inhibitor R428 induces apoptosis of cancer cells by blocking lysosomal acidification and recycling independent of Axl inhibition. *Am J Cancer Res* **8**, 1466–1482 (2018).

73. Groelly, F. J., Fawkes, M., Dagg, R. A., Blackford, A. N. & Tarsounas, M. Targeting DNA damage response pathways in cancer. *Nat Rev Cancer* (2022) doi:10.1038/s41568-022-00535-5.

74. Siddiqui-Jain, A. *et al.* CK2 inhibitor CX-4945 suppresses DNA repair response triggered by DNA-targeted anticancer drugs and augments efficacy: Mechanistic rationale for drug combination therapy. *Mol Cancer Ther* **11**, 994–1005 (2012).

75. Rabalski, A. J., Gyenis, L. & Litchfield, D. W. Molecular pathways: Emergence of protein kinase CK2 (CSNK2) as a potential target to inhibit survival and DNA damage response and repair pathways in cancer cells. *Clinical Cancer Research* **22**, 2840–2847 (2016).

76. Zhao, T. *et al.* HIF-1-mediated metabolic reprogramming reduces ROS levels and facilitates the metastatic colonization of cancers in lungs. *Sci Rep* **4**, (2014).

77. Yun, H. R. *et al.* Palmitoyl Protein Thioesterase 1 Is Essential for Myogenic Autophagy of C2C12 Skeletal Myoblast. *Front Physiol* **11**, (2020).

78. Neijssen, J. *et al.* Discovery of amivantamab (JNJ-61186372), a bispecific antibody targeting EGFR and MET. *Journal of Biological Chemistry* **296**, (2021).

79. Hsu, P.-C. *et al. YAP promotes erlotinib resistance in human non-small cell lung cancer cells*. www.impactjournals.com/oncotarget.

80. Zweemer, A. J. M. *et al.* Apoptotic Bodies Elicit Gas6-Mediated Migration of AXL-Expressing Tumor Cells. *Molecular Cancer Research* (2017) doi:10.1158/1541-7786.MCR-17-0012.

81. Maynard, A. *et al.* Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. *Cell* **182**, 1232-1251.e22 (2020).

82. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* **47**, W556–W560 (2019).

Supplementary Figures

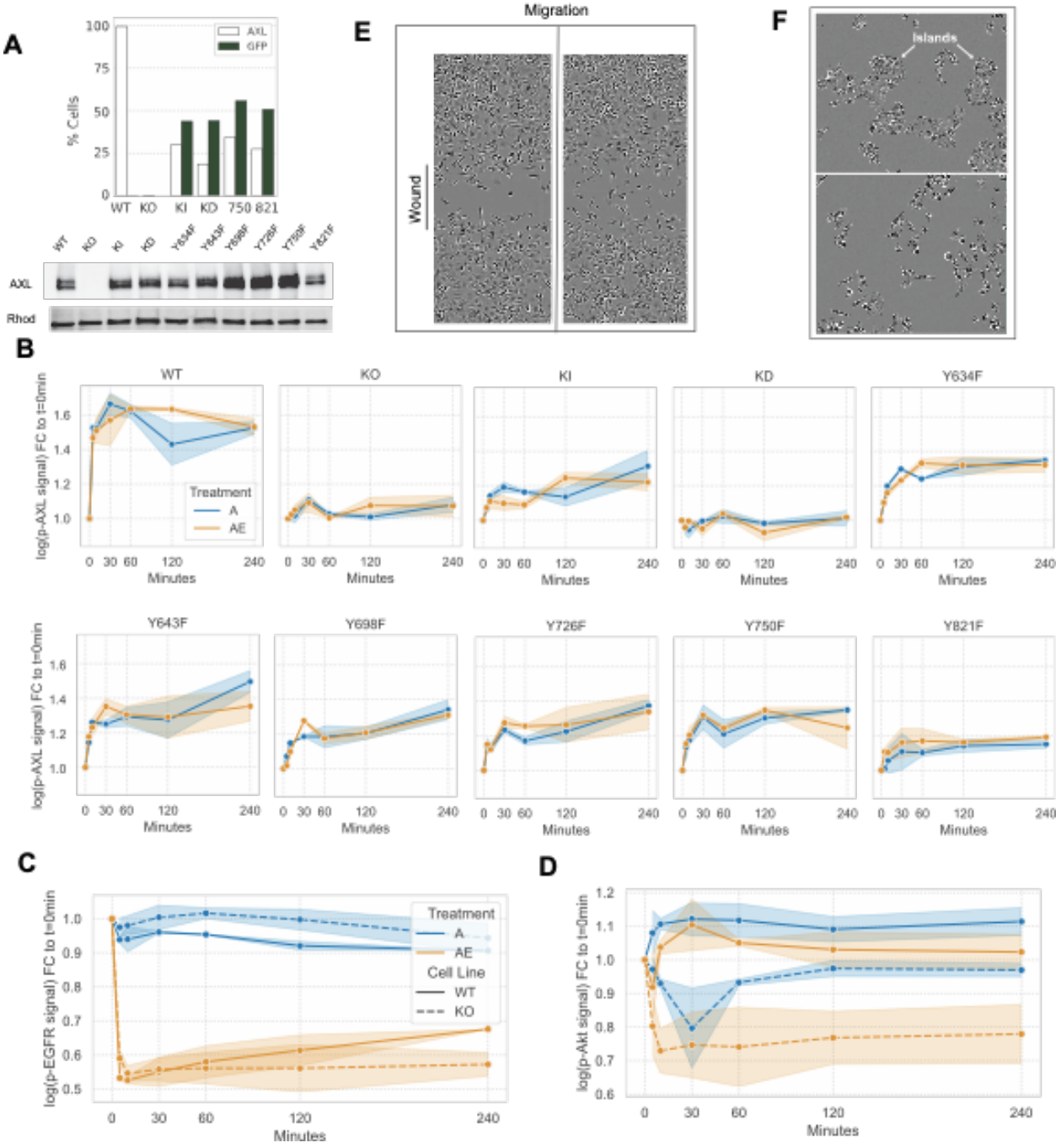


Figure S1. (A) Cell surface and total protein AXL expression in the indicated PC9 cell lines. (B) AXL phosphorylation signal in PC9 AXL mutant cell lines in response to A alone or in combination with E. (C) EGFR inhibition in response to E. (D) AXL-specific activation of Akt. (E-F) Images of a scratch wound assay after 12h (E) as well as the E-induced cell island effect (F). Related to **Figure 2**.

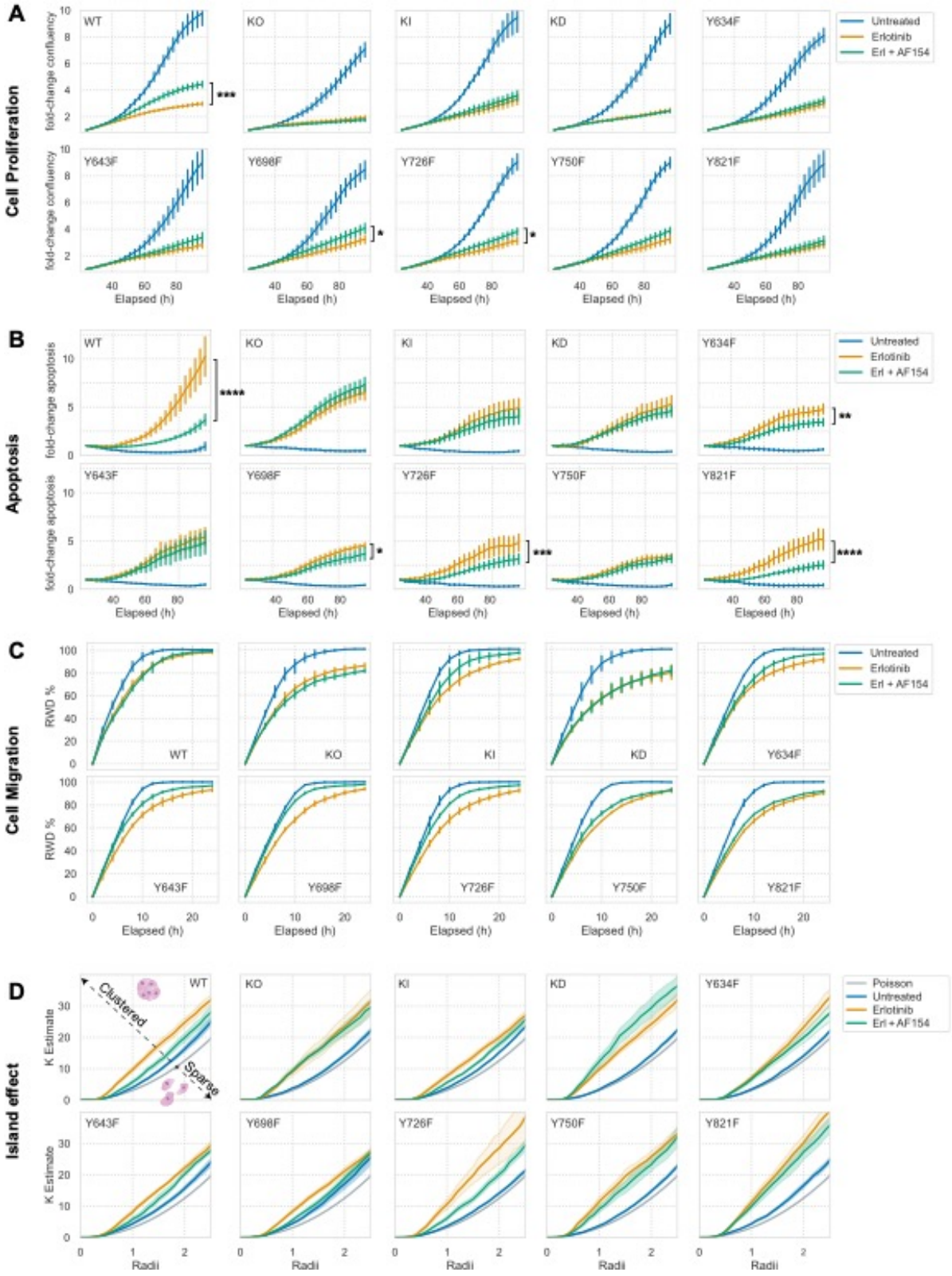


Figure S2. (A-B) Cell proliferation and cell death quantified for 96 hours using live cell imaging in response to E or EA. (C) Relative wound density (RWD) measured by a scratch wound assay across all PC9 cell lines treated with E

or EA. (D) Extent of a E-induced cell island effect upon AXL activation measured by Ripley's K function. Related to **Figure 2**.

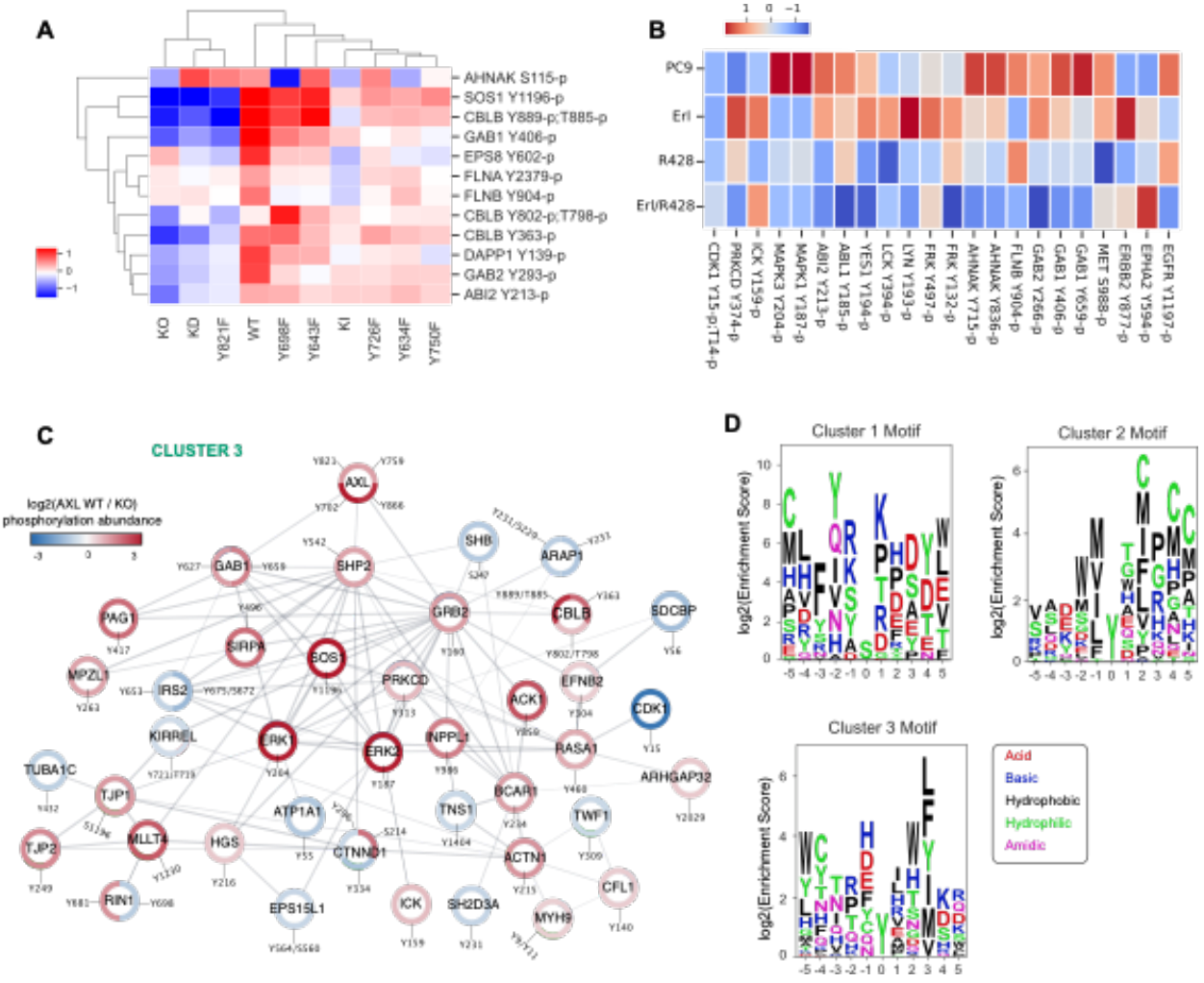


Figure S3. A) Hierarchical clustering of the phosphorylation signal of RTK adapters across AXL Y-to-F mutants. B) Heatmap of the indicated phosphosites in PC9 cells treated with E, R428, or the combination of both C) STRING network of DDMC's C3 D) Motif logo plots using the PSSM of C1, C2, and C3. Related to **Figure 3**.

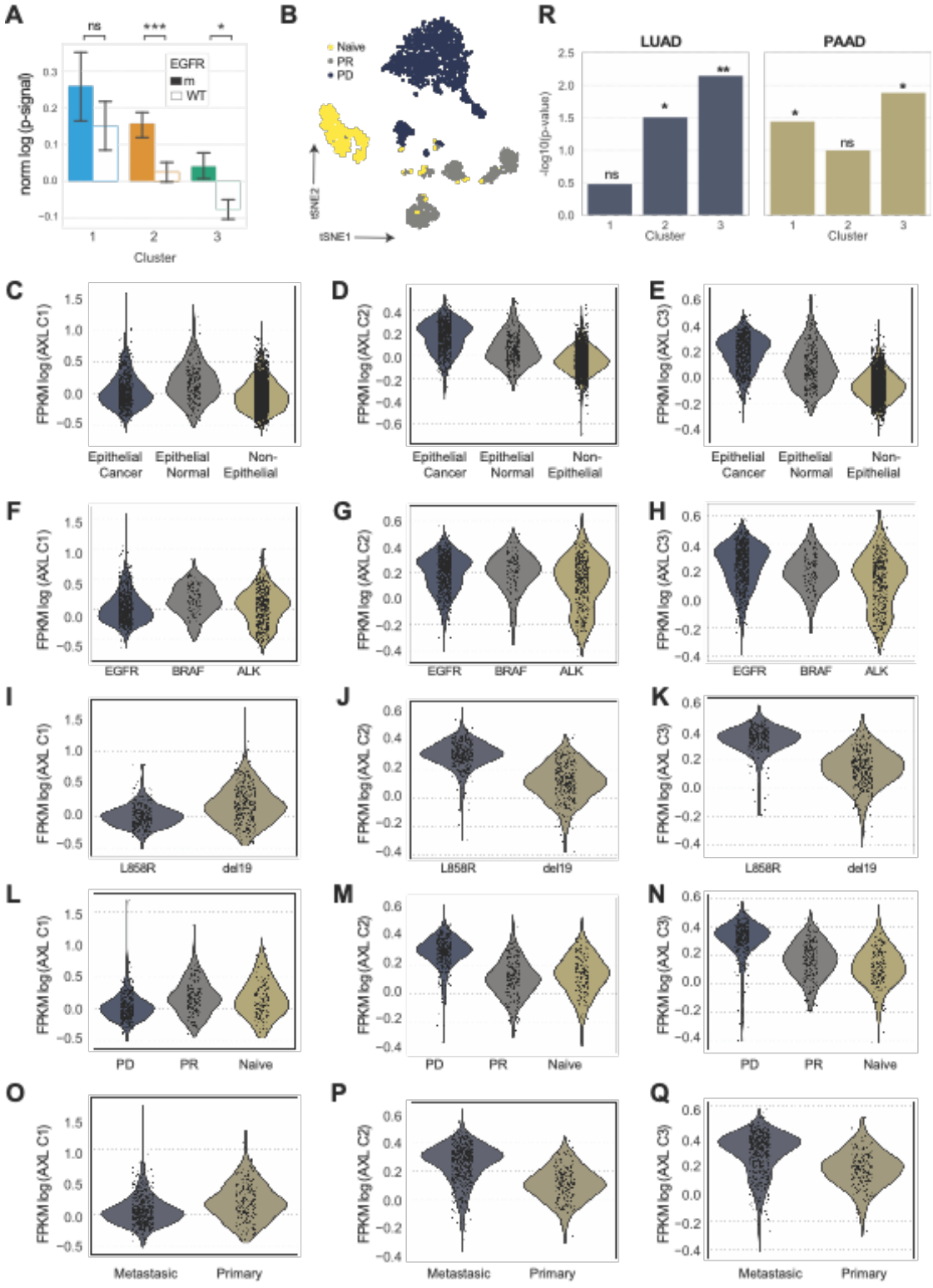


Figure S4 (A) Phosphorylation signal of C1, C2, or C3 in EGFRm or EGFR WT tumors. (B) tSNE plot of cancer cells grouped by treatment response. (C-Q) Gene expression of C1, C2, C3 in (C-E) epithelial cancer cells, epithelial normal cells, or non-epithelial, (F-H) Gene driver (I-K) EGFR mutation, (L-N) Tumor response, and (O-Q) Metastasis

status. R) p-value of the overall survival differences of LUAD or PAAD patients with high or low AXL signature scores. Related to **figure 4**.

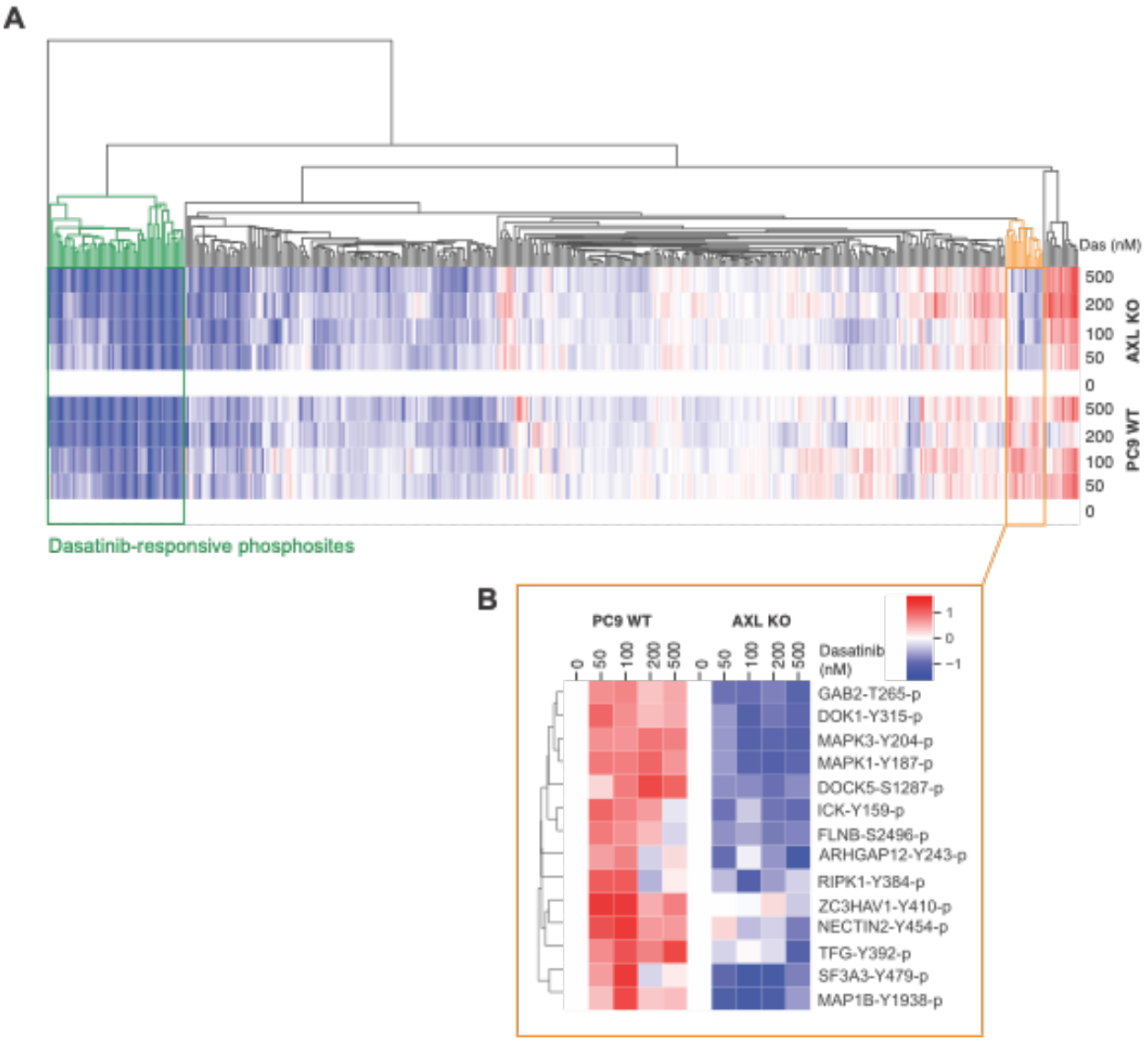


Figure S5 (A) Heatmap of full phosphoproteomic data of PC9 WT and AXL KO cells treated with 0, 50, 100, 200, or 500nM or dasatinib and EA. Two clusters are highlighted: Green indicates the dasatinib-responsive phosphosites shown in **Figure 4C.** B) Cluster shown in orange includes phosphosites increased in PC9 WT but decreased AXL KO cells. Related to **Figure 5**.

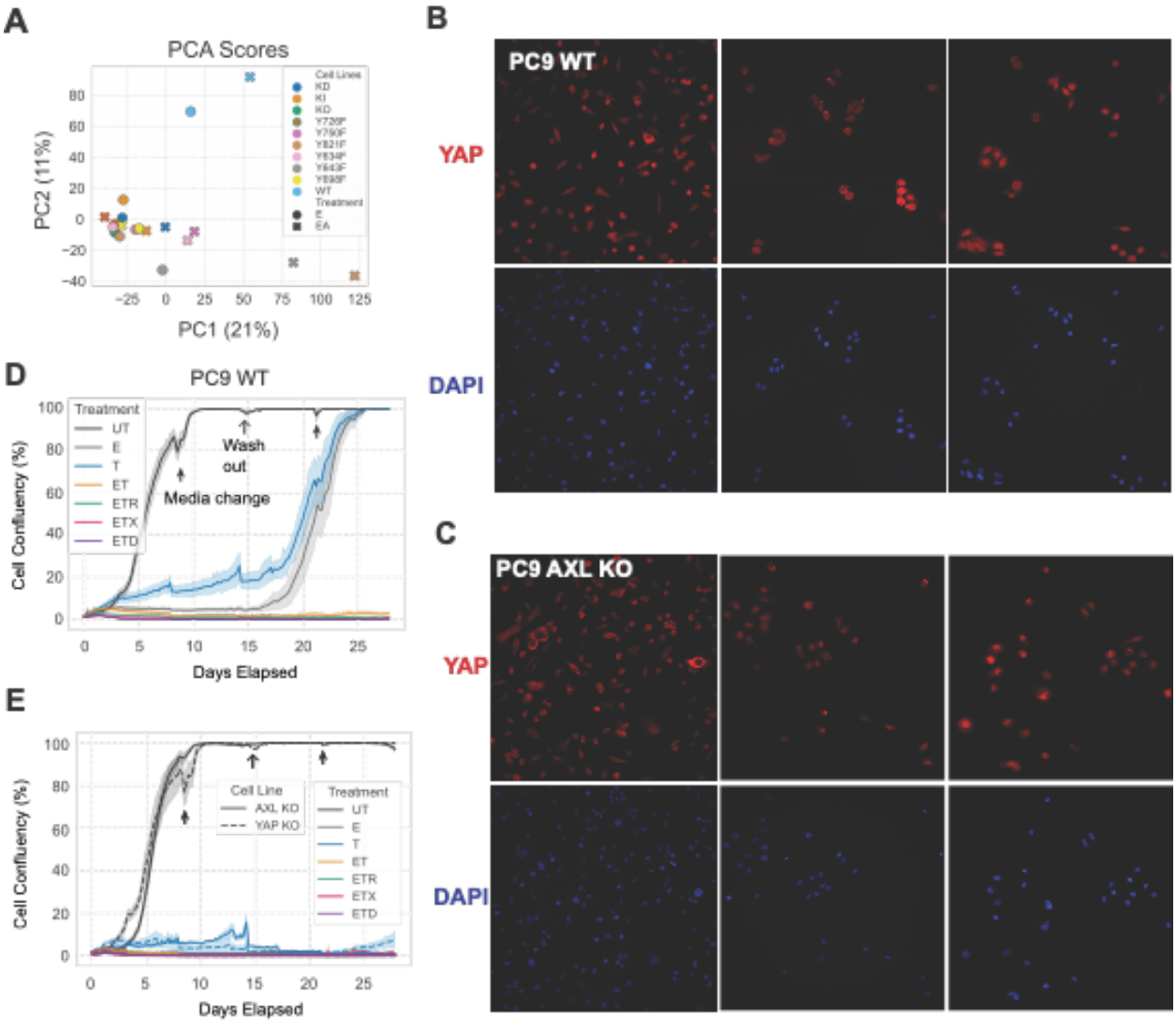


Figure S6 (A) Scores of a PCA analysis of the RNAseq data of the PC9 Y-to-F mutant cell lines treated with E or EA. (B-C) Immunofluorescence staining of YAP and DAPI in (B) PC9 WT and (C) AXL KO cells. (D-E) Drug-tolerant persister cells assay with the indicated treatments in (D) PC9 WT, as well as in (E) AXL KO and YAP KO cells. Related to **Figure 6**.

Chapter 4:

Systems approaches to uncovering the contribution of environment-mediated drug resistance

Creixell M.[1], Kim H[2,3], Mohammadi F[1], Peyton SR[2,3], Meyer A[1]

[1]Department of Bioengineering, University of California Los Angeles

[2]Molecular and Cellular Biology Graduate Program, University of Massachusetts Amherst, United States

[3]Department of Chemical Engineering, University of Amherst, United States

Current Opinion in Solid State and Materials Science 26 (2022) 101005

Contents lists available at ScienceDirect

Current Opinion in Solid State & Materials Science

journal homepage: www.elsevier.com/locate/cossms

Systems approaches to uncovering the contribution of environment–mediated drug resistance

Marc Creixell[a], Hyuna Kim[b], Farnaz Mohammadi[a], Shelly R. Peyton[b,c], Aaron S. Meyer[a,*]

[a] Department of Bioengineering, University of California Los Angeles, United States
[b] Molecular and Cellular Biology Graduate Program, University of Massachusetts Amherst, United States
[c] Department of Chemical Engineering, University of Massachusetts Amherst, United States

ARTICLE INFO

Keywords:
ECM
Cell signaling
Drug resistance
Tensor decomposition

ABSTRACT

Cancer drug response is heavily influenced by the extracellular matrix (ECM) environment. Despite a clear appreciation that the ECM influences cancer drug response and progression, a unified view of how, where, and when environment–mediated drug resistance contributes to cancer progression has not coalesced. Here, we survey some specific ways in which the ECM contributes to cancer resistance with a focus on how materials development can coincide with systems biology approaches to better understand and perturb this contribution. We argue that part of the reason that environment–mediated resistance remains a perplexing problem is our lack of a wholistic view of the entire range of environments and their impacts on cell behavior. We cover a series of recent experimental and computational tools that will aid exploration of ECM reactions space, and how they might be synergistically integrated.

1. Tumor cell–environment interactions are multi–factorial

Environmental factors play a crucial role in dictating sensitivity or resistance of cancer cells to drug treatment (Fig. 1). Advances in biomaterials and bioprinting technologies have made it possible to mimic the 3D tumor microenvironment and evaluate how specific extracellular matrix (ECM) compositions, ligands, geometries, and biophysical properties influence drug response more accurately. However, a consensus understanding of how environment–mediated drug resistance (EMDR) limits the efficacy of targeted therapies remains elusive. The presence of cell–to–cell intrinsic differences, multi–factorial environmental inputs, and combinations of resulting phenotypes in and around tumors present an overwhelmingly large parameter space affecting drug response. Given these challenges, we propose that a combination of modeling techniques tailored for high–dimensional analysis, alongside high–throughput experimental approaches to interrogate engineered tumor microenvironment (TME) variations, is critical to integrate and summarize this vast space, unify findings across studies, and generate data–driven hypotheses and biological insights related to EMDR.

1.1. Cell intrinsic differences lead to clonal diversification

Before directly addressing EMDR, we must acknowledge how clonal heterogeneity drives drug response characteristics. Intratumoral heterogeneity can be subdivided in two broad categories—environmental and cellular—resulting in complexity that must be considered when studying ECM–cell interactions. Isolation and systematic characterization of cancer cells has significantly advanced our understanding of the cell–autonomous heterogeneity present within the tumor [1,2]. Genome sequencing studies have uncovered that tumors are both spatially and temporally heterogeneous, leading to major challenges in finding therapies effective for all cells in a tumor. As just one of numerous examples, Yates et al. analyzed the spatial distribution of subclones for twelve cancers. Eight showed significant spatial heterogeneity in cells with point mutations, and two showed heterogeneity in copy–number changes. Interestingly, they noted the degree of heterogeneity in triple–negative breast cancer correlated to tumor size, although the causality is not clear [3].

Environmental heterogeneity arises when regions within the TME contain distinct environments that modulate disease progression. For instance, the pancreatic TME consists of locally confined sub–tumor microenvironments with distinct fibroblast plasticity, tumor immunity, and treatment response states revealed by regional transcriptomic and proteomic measurements [4]. Additionally, tumors are formed by heterogenous subpopulations of intrinsically different cells that evolve in response to their environment; consequently, each subpopulation has

* Corresponding author.

https://doi.org/10.1016/j.cossms.2022.101005
Received 10 December 2021; Received in revised form 26 April 2022; Accepted 27 April 2022
Available online 13 May 2022

distinct input–output responses to the TME. A substantial amount of work has been dedicated to investigating the emergence of subclonal populations defined by their genetic diversification in response to the selective pressure imposed by drugs [5–8]. In breast cancer, sequencing 303 tumor samples from 50 patients enabled investigators to map the spatial and temporal clonal evolution of tumors. They were able to identify causal genetic alterations present in specific subclones with increased resistance to chemotherapy and metastatic capacity [3]. However, while clonal subpopulations within tumors have now been extensively profiled for their selection and relative fitness within the tumor ecosystem,this has yet to be evaluated within the context of their local TME (Fig. 1) [9].

1.2. Cell–ECM interactions

The ECM is a collection of sugars, proteoglycans, and proteins that collectively provide attachment, scaffolding, and, via integrin binding, initiate receptor–mediated signaling in resident cells, which plays a critical role in regulating tumor growth and metastasis[10]. Tissuesthat nurture the formation of successful metastases have a distinct ECM profile [11], and thorough analysis of these sites has revealed that each tissue has a distinct profile of adhesive matrix proteins and physiological stiffness [12–20]. Inspired by this, Barney et al. designed bone–,brain–, and lung–inspiredbiomaterial platforms by varying the composition and density of ECM proteins [21]. When they cultured bone–,brain–,or lung–tropic cell lines on these ECMs, they found coordination of ECM with EGF regulation of spreading rate, area, displacement, and chemotactic index. They also demonstrated that these ECM–specific responses could predict breast cancer metastasis. The clear next step is to take these ECM–specific biomaterials to drug response studies.

Several studies have indeed demonstrated that integrin–binding to the ECM drive therapy resistance and blocking these interactions has emerged as a potential strategy to overcome integrin–mediated drug resistance [22–24]. For instance, administration of a β_1 integrin inhibitory antibody (AIIB2) reduced tumor volume and increased the efficacy of ionizing radiation in human breast cancer xenografts [25]. A 3D spheroid system demonstrated that matrix–attached ovarian cancer cells developed resistance to dual inhibition of PI3K/mTOR, while inner matrix–deprived cells underwent apoptosis. The ECM–driven adaptive mechanism included upregulation of nutrient deprivation and cellular stress programs that sensitized cells when targeted [26]. Lee et al. developed a miniaturized 3D cell–culture array (DataChip), which spotted up to 1080 cell–seededcollagen or alginate gels onto glass slides [27]. This assay, although miniaturized, yielded accurate cytotoxicity information when comparing its IC_{50} values to those in conventional well–plate assays. These high–throughput 3D systems will enable toxicity analysis of drug candidates and their metabolites at the early–stage of drug development process(Fig. 2) [28]. Examples of combinatorial ECM screening approaches include that from Beachley et al., who spotted ECM from 11 different tissues and quantified cell growth over a large panel of cell types, including three different cancer cell lines [29]. Although not explored yet, this would be an excellent system to screen for EMDR across tissue specific ECMs. Mabry et al. designed a synthetic hydrogel in a high–throughput multi–well plate, where they can vary

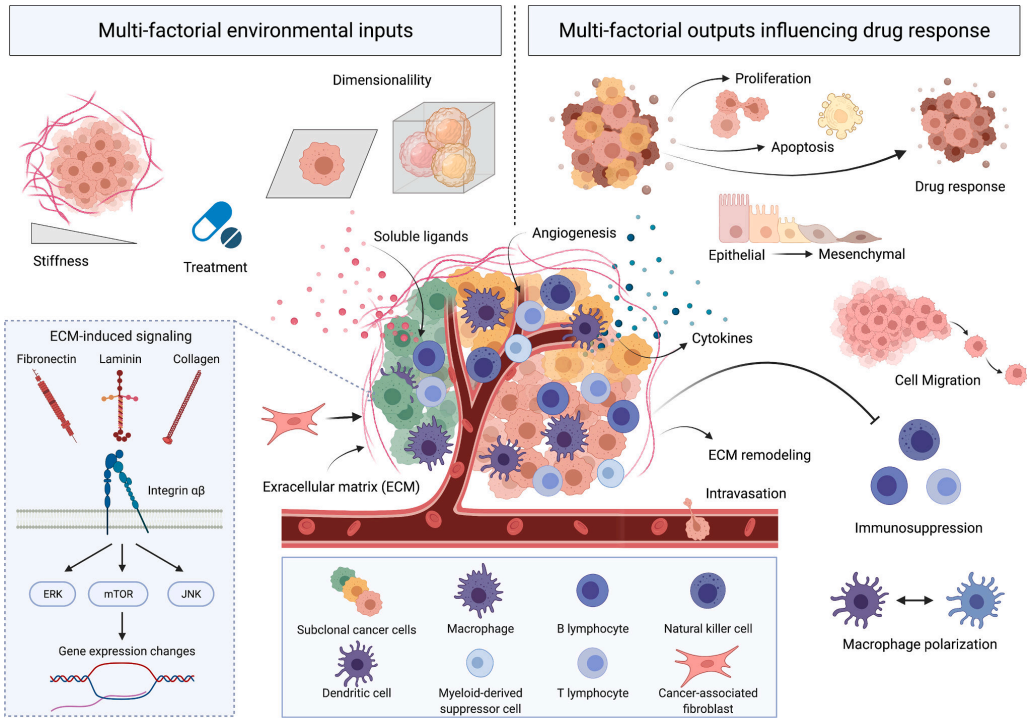


Fig. 1. Tumor cell responsesto their environment are multi–factorial. There are multiple factors that can affect tumor cell behaviors. Dimensionality (2D or 3D), stiffness of the culturing materials, topography of the surface, addition of growth factors, shear forces or interstitial fluid pressure can change how tumor cells behave. Outputs are also multi–factorial. Within the same tumor, cells can show different proliferation profiles, drug responses,immune responses,and ECM remodeling due to their intra–and inter–tumor heterogeneity.

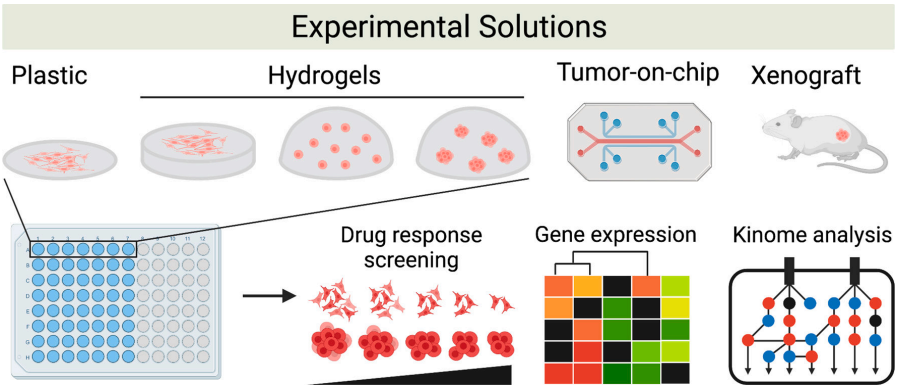


Fig. 2. Experimental solutions to reflect multiple environmental inputs and outputs. To consider physical and biological factors that affect cancer cell behavior, different culturing platforms such as 2D plastic, 2D hydrogels, 3D hydrogels with single cells, 3D hydrogels with spheroids can be used. These can be made on high-throughput system, like 96-well plates. Tumor-on-chip is another platform that reflects tumor microenvironment with microfluid system. Tumor xenografts, although not high throughput, are physiologically closer to the actual tumor microenvironment. With these platforms, drug response, gene expression, and kinome analysis can be done to study multi-factorial outputs.

several parameters including the types of peptides, peptide concentration, and matrix modulus dynamically while cells are embedded [30]. Thus far, this lab has used this and similar approaches to study vascular interstitial cells during fibrosis, and it would be an excellent tumor ECM progression model for cancer.

These studies and others also emphasized the importance of dimensionality, or the geometry in which cells grow, when attempting to recreate the TME [31,32]. In fact, directly comparing CRISPR screens in 2D monolayers and 3D lung-cancer spheroids revealed that many 3D cancer growth-specific vulnerabilities are not recapitulated in 2D [33]. The role of focal adhesion proteins in driving cancer cell dissemination is similarly distinct between 2D and 3D environments [34]. However, systematic comparisons of different dimensionalities can reveal how individual cellular components retain their functional role, but differing importance comes from a shift in the overall rate limiting processes. For example, Meyer et al. identified that, while migration growth factor responses differ between 2D and 3D, initial membrane protrusion responses were identical [35]. A mechanistic understanding of which specific components within the ECM and molecular interactors in cancer cells modulate these phenotypic responses will improve our ability to manipulate these responses.

1.3. The physical traits of the TME

The biophysical properties of the TME (e.g., ECM modulus) generate forces that are sensed by cells via integrins, the cytoskeleton, and mechanosensitive ion channels, and strongly influence cell behavior [36]. Modulus is an intrinsic property that describes the extent to which the ECM resists deformation in response to an applied force. Increased modulus is commonly observed in tumors relative to healthy tissue across many cancer types [37–40], typically through ECM deposition by stromal cells, especially cancer associated fibroblasts (CAFs) [41], or via mechanical stresses from cell contraction and tumor growth [36]. The density and identity of ECM proteins deposited by tumor stroma dynamically stiffens the ECM: e.g., type I collagen imparts tensile strength, and elastin gives tissue its elasticity. Several studies support a prevailing hypothesis that ECM stiffness impacts tumor metastases [42,43]. Solid stresses can also be generated by increased tissue volume and increased interstitial fluid pressure. The impact of these mechanical factors in treatment response has gained continued attention. However, their role in EMDR and cancer recurrence has been underappreciated and a deeper molecular understanding of the pathways engaged might lead to further therapeutic opportunities. We will highlight some of the emerging studies and tools developed for this exact topic here.

Schwartz et al. developed an approach to study drug responses in cells cultured on a 2D environments or in 3D hydrogels [44]. This study revealed that cells were more resistant to receptor tyrosine kinase (RTK)-targeting drugs, such as sorafenib and lapatinib, when cells were cultured on high modulus 2D hydrogels or when spheroids were cultured in high modulus 3D hydrogels. This agrees with a prior study from Nguyen et al. who showed that several cancer cell lines were more resistant to sorafenib when they were cultured on a stiffer 2D biomaterials [45]. They went further to show that gene expression differences were not sufficient to explain the observed drug resistance. They combined multiple linear regression analysis with quantification of the activation of a small panel of receptor tyrosine kinases in and on those environments to reveal that blocking MEK phosphorylation alongside these targeted therapies could overcome EMDR and reduce tumor burden in vivo.

Both prior studies show the value in incorporating tunable biomaterials into high-throughput formats, but their limitation is that although they tuned modulus and geometry, they were limited to single ECM proteins. Brooks et al. overcame this limitation when they studied ovarian cancer drug resistance in an omentum-inspired 3D hydrogel [46]. They generated 3D tumor spheroids in microwells and encapsulated them in the 3D hydrogels that had the stiffness and protein composition of omentum tissue. The tumor spheroids showed increased sensitivity to Carboplatin, Doxorubicin, LY2606368, and Mafosfamide compared to those cultured on 2D tissue culture polystyrene, which they were able to correlate to observed clinical response with paired primary patient derived samples [46].

1.4. Soluble ligands are released by cancer and stromal cells

Paracrine signaling between cancer and stroma also initiates receptor-mediated signaling to tumor cells, further driving disease progression. For example, TGFβ, IL-6, IL-8, EGF, and hepatocyte growth factor (HGF) produced by CAFs have emerged as potential drug resistance targets [41,47]. IL-6 and IL-8 exert pro-migratory effects on cancer cells [48], and IL-6 further affects apoptosis, cell growth, angiogenesis, and antioxidant metabolic programs that ultimately protect cells from therapy [49]. Similarly, autocrine and paracrine release of TGFβ promotes epithelial-to-mesenchymal transition (EMT), cancer cell migration, and invasion [50–52]. HGF is the cognate ligand of the RTK c-Met,

which regulates tumor growth, migration, and resistance to tyrosine kinase inhibitors (TKIs) [50,51,53,54]. Importantly, crosstalk between response to combinations of these soluble factors can additionally serve as a drug resistance mechanism. For instance, CAFs respond to TGFβ inhibition by increasing release of HGF, which in turn enhances cell invasion [55–58]. Furthermore, cancer cells reprogram their energy metabolism from mitochondrial oxidative phosphorylation to glycolysis–mediated ATP production to fuel uncontrolled cell proliferation in tumors. This metabolic switch occurs in response to hypoxia common to tumors, activating a response system that upregulates glucose transporters and glycolytic enzymes, and increases the release of reactive oxygen species (ROS) [59,60]. Recent work by Oren et al. observed a strong relationship between the proliferative capacity of persister cell subpopulations and antioxidant expression signatures. They subsequently validated that the EGFR inhibitor osimertinib induces an increase in ROS and showed that the fraction of proliferative persister cells increased significantly upon ROS neutralization using the ROS scavenger N–acetylcysteine [61].

Clearly, soluble factors in the TME are crucial mediators of EMDR. Whether and how these signals should be targeted, however, will depend on their exact role in the TME. For instance, these soluble factors might alter tumor cell responses to their environment, promote survival and other phenotypes directly, or promote tumorigenic responses indirectly through changes to the TME. Systematic studies of soluble and ECM signals will help to identify the unique or overlapping role of each. A sophisticated study to do exactly this, Lin et al. generated 210 unique microenvironments to explore in parallel how they shape the inhibitory effects of lapatinib [62]. They tested four ECM compositions and seven ligands, on either 2,500 Pa or 40 kPa elastic modulus substrate polyacrylamide gels and measured cell proliferation, HER2 expression and phosphorylation, and cell morphology across four different cancer cell lines. Interactions between soluble factor responses can be observed in vivo as well. For instance, trastuzumab or lapatinib elicited responses in HER2–amplified breast cancer metastases when they were situated in mammary fat pad but not in the brain, where the combination of a HER2 inhibitor with an anti–VEGFR2 antibody was required to slow tumor growth [63]. In a second study, HER3 blockade was needed to overcome resistance conferred by the brain microenvironment to PI3K inhibition [64].

1.5. Multiple cell phenotypes, in combination, define the resistance capacity of cancer cells

Drug resistance is typically quantified via a direct or surrogate measure of cell number. However, the ability of tumors to sustain growth in the presence of treatment is defined by cooperation between phenotypic responses. Drug treatment can lead to potent differences in cell response, hidden by simply quantifying cell number [65]. Resistant tumor cells that undergo EMT result in both drug resistance and increased metastatic capacity [66,67]. Resistance to cancer immunotherapies is driven, at least in part, by mechanisms such as T–cell exhaustion, immune suppressive cell populations, and inhibitory cytokines and metabolites that immunologically suppress the tumor microenvironment [68]. Furthermore, cancer and other resident cells (e.g., CAFs) reciprocally respond to tumor environmental signals by remodeling their ECM. This response can be subdivided in three broad categories: (1) ECM deposition affecting the biochemical and mechanical properties of the environment, (2) the proteolytic degradation of the ECM weakening migratory barriers and promoting the invasive ability of cancer cells, and (3) integrin–mediated cell–ECM binding causing non–proteolytic physical breaching of the base membrane facilitating cancer cell invasion [69]. For these reasons, drug resistance must be viewed as the coordination of multiple phenotypes to overcome the inhibitory effects of therapy.

The drug response and resistance mechanisms of RTK inhibitors effectively illustrates the importance of determining the role of individual factors within an overall environment. Our labs and others have shown that a common form of RTK inhibitor resistance is derived from alternative "bypass" RTKs that can reactivate essential survival signaling [70–73]. While each RTK has some propensity for this signaling redundancy, different growth factors vary in their resistance–conferring capacities. For instance, lung adenocarcinoma cells treated with the EGFR inhibitor erlotinib can be made resistant by HGF or FGF much more so than EGF, IGF, HRG, or PDGF, and each growth factors' resistance conferring capacity could be explained by its ability to activate a common downstream pathway [70,74]. Beyond merely inducing cell proliferation, RTK–mediated bypass resistance can also activate collateral malignant programs that, in combination, limit therapy efficacy and direct disease progression. For instance, resistance through AXL activation associates with more advanced diseases stages as the receptor sustains cells that have undergone EMT resulting in increased metastatic capacity [66,73,75,76]. The TME can also modulate drug response through these same signals. While these bypass mechanisms mean that there are many possible environmental factors that can contribute to resistance, systematically identifying which factors may lead to similar downstream consequences would help to build a catalog of which are capable of conferring bypass resistance.

Together, these observations indicate that no one environment will faithfully represent the complex cell–ECM interactions that exist in a living tumor. High–throughput screening methods that systematically evaluate the impact of hundreds of tumor–inspired microenvironments can circumvent these limitations to help uncover resistance or sensitivity–driving environmental cues. These data can then be complemented with transcriptomics, proteomics, and/or phosphoproteomics information to uncover the signaling programs that transduce such cues and give rise to the resulting phenotypic responses (Fig. 2). Provided this huge space of inputs and outputs, a set of experimental and computational methods to systematically disentangle what specific microenvironmental factors, signaling changes, and phenotypic expressions drive drug resistance will be necessary.

2. Unique challenge of multi–factorial problems like environment–mediated resistance: Exploring the space of possibilities

A central challenge in understanding how environment contributes to tumor cell resistance is that the "space" of environmental factors is so large. Different environmental factors interact, creating context–specific responses unless one considers the entire range of possible environments. On the other hand, it becomes impossible to measure response to all possible environments. Even in large–scale profiling studies, it can be difficult to "map" between studies to determine which responses occur through common mechanisms, and which represent distinct pathways. A similar problem in mapping arises when considering the relevance of in vitro systems to in vivo responses: first, there are many factors that can be mimicked in vitro, but it is not clear which are essential to creating an accurate (or minimally necessary) representation of the in vivo environment; second, the in vivo environment itself is not monolithic, and contains microenvironments that vary in their composition. The answer cannot simply be larger studies, as the scale difference is modest compared to the overwhelming range of environmental factors. For instance, testing 10 parameters defined by the categories above, in combination and in triplicate, would require 300,000 measurements (Fig. 3A).

A similar challenge exists in the use of many–parameter computational models. Optimization requires exploring which among many possible inputs leads to optimal outputs. However, the curse of dimensionality dictates that problems quickly become intractable, even for computational exploration, as the number of parameters increases (Fig. 3B). Solutions to this problem from computational applications offer lessons for studying the ECM environment.

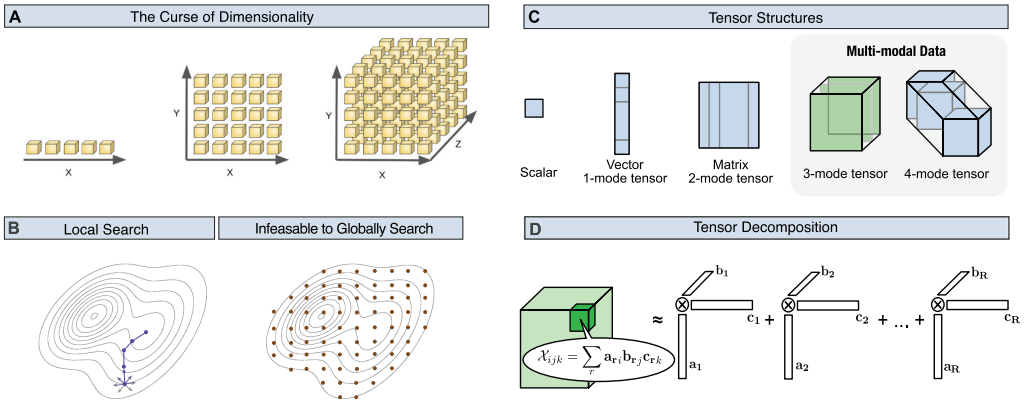


Fig. 3. Computational solutions to mapping EMDR. A) The vast "space" of environmental factors that affect cell response prevents an accurate experimental and computational representation of EMDR. B) Optimization is achieved through a local exploration within the neighborhood of an existing optimal solution. However, because of the curse of dimensionality, this can require an impossibly large number of local searches to fully explore the multi-factorial space of EMDR. C) Multi-modal data includes different variables that combinatorically influence the ECM-mediated cell responses. D) Dimensionality reduction using tensor decomposition can vastly simplify multi-modal patterns.

3. Computational solutions to reduce the space of ECM responses

Computational approaches can provide a solution to parsing the overwhelming space of ECM environment possibilities and their effects on drug response. These methods can help visualize multi-factorial space, optimize experimental design for desired outcomes, and integrate observations with prior knowledge. To do so, computational methods work off two general principles—recognition of certain general patterns, and the use of pre-defined patterns based on prior knowledge. Both approaches have shown success in making sense of cells' responses to their environment and have future opportunities for improvement.

A particularly promising solution to exploring many-parameter responses is finding low-rank structure within a space of possible responses. For instance, when exploring how cells respond to a panel of soluble factors, the responses can be visualized with a technique like principal component analysis (PCA), which identifies a low-rank structure within matrix data (conditions ×measurements). However, cell responses are not only multi-dimensional but multi-modal, meaning that types of variables, such as soluble ligands, treatment, and ECM biophysical or biochemical properties, can be present in all possible combinations, and induce cell phenotype changes through their combination (Fig. 3C and 4). For instance, Lin et al examined cell responses to lapatinib with different soluble ligands, ECMs, and matrix stiffnesses, with or without lapatinib treatment. These different groups of factors were tested in all possible combinations. Such structured data can be organized as a 4-mode tensor where axes represent ECMs, ligands, matrix stiffness, and treatments. Moreover, this data is still multi-dimensional given that many measurements were made. For these cases, higher-mode generalizations of PCA exist. These methods, broadly referred to as tensor decomposition techniques, can be remarkably effective at dimensionality reduction even beyond PCA (Fig. 3D) [77]. For instance, Farhat et al. recently profiled the response of several immune cell populations to cytokines at several concentrations over time, resulting in a four-mode tensor (cell types ×time × ligand ×concentration) [78]. Tensor-based decomposition reduced the data to ~ 2% of its original size, while preserving 90% of the variance in the original 2,880 measurements, by recognizing that just two overall patterns existed across the different parameters [78]. Importantly, this reduction is far beyond what would have been achieved by PCA, and additionally helped visualize the effect of each mode by separating their effects (e.g., the effect of time). Tan et al. similarly explored how tensor decompositions can reduce antibody serology data in which the antigen targeting and immune interactions of antibodies are profiled across subjects and time [79]. There too, they observed that dimensionality reduction in tensor form could further reduce data into component patterns, and that its visualization was improved by separation of each mode's effect. Dimensionality reduction in this form can also help to harmonize data across studies. For instance, it has been used to identify patterns across microbiome studies, overcoming inter-study batch effects and sparse, irregular sampling [80]. Common patterns exist among all the ECM factors we have described above, as well, where substantial dimensionality reduction can reduce the search space of experiments.

As an illustrative example of the biological conclusions one can derive from multi-modal dimensionality reduction, we performed PARAFAC decomposition on a microenvironment microarray (MEMA) dataset of MCF10A breast epithelial cells. 182 cellular properties were measured across 57 ligand treatments and 48 ECM environments, forming a 3-mode tensor [81] (Fig. 4A). First, PARAFAC can reduce the data more effectively than PCA—by 98% while still explaining 60% of the variance (Fig. 4B). This makes it easier to identify patterns in how ECM composition, soluble factors, and interactions between both affect the measured cell response. The influence of each mode within a component is multiplied together to reconstruct the data, making interpretation straightforward. For instance, component 2 illustrates the effect of elastin and nidogen, across a range of soluble ligands, affecting cytoplasmic shape properties. These results additionally show how one can draw overall conclusions about the data. For example, component 2 varies more extremely with ECM environment, while component 1 is much more responsive to soluble ligands' presence. Finally, tensor decomposition can facilitate the integration of various data sets, provided they have one mode in common. Here, one could expand the shown tensor (Fig. 4A) into a 4-mode tensor incorporating measurements from different cell lines across the same ligands and ECMs [82,83]. This can help ensure that the cell-ECM interactions discovered are conserved across relevant cell line models.

Alternative solutions to the curse of dimensionality are methods that still enable optimization for some desired output. Broadly, optimization is often performed computationally using either gradients, which indicate the direction of iterative improvement, or surrogate models that act as a stand-in to represent the search space. Both techniques have been applied, for example, in protein design. For instance, directed evolution

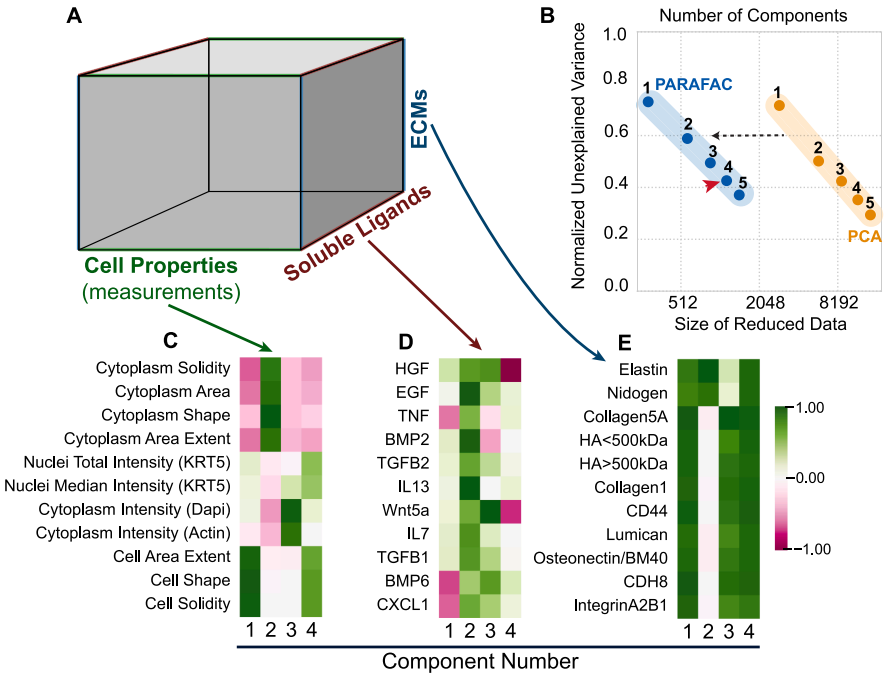


Fig. 4. Illustrative example of multi-modal microenvironment microarray (MEMA) dimensionality reduction. (A) MEMA data in a 3-mode tensor format containing the measured cell properties across various ECM environments and soluble ligands [79]. (B) PARAFAC reduces the MEMA data set by 98% compared to its original size, clearly outperforming PCA across different component numbers. (C–E) Upon decomposition using 4 components (red arrowhead in B), the resulting weights explain the combinatorial effects of the different factors within each mode, namely soluble ligands (D) and ECMs (E) affecting cell properties (C). Note that only a part of the data was shown for simplicity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

is essentially a series of local searches within the neighborhood of the existing optimal solution (Fig. 3B) [84]. Surrogate modeling using generative models like Gaussian processes has been widely applied for optimizing protein design when combined with evaluating libraries of variants [85–87]. These techniques might similarly be applied to ECM environments to optimize some desired outcome such as in the design of regenerative tissue scaffolds.

4. Outlook

Though the TME is seemingly vast in the possible combinations of environments and responses, solutions exist for systematically characterizing such problems. Overcoming drug resistance induced by the tumor environment will continue to benefit from tighter integration of experimental solutions to systematically measure environmental responses, and computational solutions to make sense of these measurements. As outlined above, multi-modal dimensionality reduction solutions will be especially critical to this effort as they provide a solution for globally characterizing the space of possible ECM environments. There is good reason to expect that, despite the large number of environmental factors, common patterns exist in the types of responses they elicit, and therefore significant gains are possible through dimensionality reduction. We expect that, as progress is made in characterizing these responses, the ECM environment will one day be considered tractable to profile in a comprehensive way. Comprehensively capturing the influence of the ECM environment will in turn greatly improve our ability to elicit durable treatment responses in cancer.

CRediT authorship contribution statement

Marc Creixell: Writing – original draft, Visualization, Writing – review & editing. **Hyuna Kim:** Writing – original draft, Visualization, Writing – review & editing. **Farnaz Mohammadi:** Visualization, Writing – review & editing. **Shelly R. Peyton:** Writing – original draft, Visualization, Supervision, Writing – review & editing. **Aaron S. Meyer:** Conceptualization, Writing – original draft, Visualization, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by NIH U01-CA215709 to A.S.M., a grant from the Jayne Koskinas Ted Giovanis Foundation for Health and Policy to S.R.P. and A.S.M., and in part by the UCLA Jonsson Comprehensive Cancer Center (JCCC) grant NIH P30-CA016042.

References

[1] R.M. Neve, K. Chin, J. Fridlyand, J. Yeh, F.L. Baehner, T. Fevr, L. Clark, N. Bayani, J.-P. Coppe, F. Tong, T. Speed, P.T. Spellman, S. DeVries, A. Lapuk, N.J. Wang, W.-L. Kuo, J.L. Stilwell, D. Pinkel, D.G. Albertson, F.M. Waldman, F. McCormick, R.

B. Dickson, M.D. Johnson, M. Lippman, S. Ethier, A. Gazdar, J.W. Gray, A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes, Cancer Cell 10 (6) (2006) 515–527.
[2] A. Lapuk, H. Marr, L. Jakkula, H. Pedro, S. Bhattacharya, E. Purdom, Z. Hu, K. Simpson, L. Pachter, S. Durinck, N. Wang, B. Parvin, G. Fontenay, T. Speed, J. Garbe, M. Stampfer, H. Bayandorian, S. Dorton, T.A. Clark, A. Schweitzer, A. Wyrobek, H. Feiler, P. Spellman, J. Conboy, J.W. Gray, Exon-Level Microarray Analyses Identify Alternative Splicing Programs in Breast Cancer, Mol Cancer Res 8 (7) (2010) 961–974.
[3] L.R. Yates, M. Gerstung, S. Knappskog, C. Desmedt, G. Gundem, P. Van Loo, T. Aas, L.B. Alexandrov, D. Larsimont, H. Davies, Y. Li, Y.S. Ju, M. Ramakrishna, H. K. Haugland, P.K. Lilleng, S. Nik-Zainal, S. McLaren, A. Butler, S. Martin, D. Glodzik, A. Menzies, K. Raine, J. Hinton, D. Jones, L.J. Mudie, B. Jiang, D. Vincent, A. Greene-Colozzi, P.-Y. Adnet, A. Fatima, M. Maetens, M. Ignatiadis, M.R. Stratton, C. Sotiriou, A.L. Richardson, P.E. Lønning, D.C. Wedge, P. J. Campbell, Subclonal diversification of primary breast cancer revealed by multiregion sequencing, Nat. Med. 21 (7) (2015) 751–759.
[4] B.T. Grünwald, A. Devisme, G. Andrieux, F. Vyas, K. Aliar, C.W. McCloskey, A. Macklin, G.H. Jang, R. Denroche, J.M. Romero, P. Bavi, P. Bronsert, F. Notta, G. O'Kane, J. Wilson, J. Knox, L. Tamblyn, M. Udaskin, N. Radulovich, S.E. Fischer, M. Boerries, S. Gallinger, T. Kislinger, R. Khokha, Spatially confined sub-tumor microenvironments in pancreatic cancer, Cell 184 (22) (2021) 5577–5592.e18, https://doi.org/10.1016/j.cell.2021.09.022.
[5] D.A. Landau, S.L. Carter, G. Getz, C.J. Wu, Clonal evolution in hematological malignancies and therapeutic implications, Leukemia 28 (1) (2014) 34–43.
[6] M. Greaves, C.C. Maley, Clonal evolution in cancer, Nature 481 (7381) (2012) 306–313.
[7] B. Zhao, J.C. Sedlak, R. Srinivas, B. Tidor, D.A. Lauffenburger, M.T. Hemann, B. Zhao, J.C. Sedlak, R. Srinivas, P. Creixell, et al., Exploiting Temporal Collateral Sensitivity in Tumor Clonal Evolution Exploiting Temporal Collateral Sensitivity in Tumor Clonal Evolution, Cell 165 (2016) 234–246.
[8] N. Bolli, H. Avet-Loiseau, D.C. Wedge, P. Van Loo, L.B. Alexandrov, I. Martincorena, K.J. Dawson, F. Iorio, S. Nik-Zainal, G.R. Bignell, J.W. Hinton, Y. Li, J.M.C. Tubio, S. McLaren, S. O' Meara, A.P. Butler, J.W. Teague, L. Mudie, E. Anderson, N. Rashid, Y.-T. Tai, M.A. Shammas, A.S. Sperling, M. Fulciniti, P. G. Richardson, G. Parmigiani, F. Magrangeas, S. Minvielle, P. Moreau, M. Attal, T. Facon, P.A. Futreal, K.C. Anderson, P.J. Campbell, N.C. Munshi, Heterogeneity of genomic evolution and mutational profiles in multiple myeloma, Nature Communications 5 (1) (2014), https://doi.org/10.1038/ncomms3997.
[9] D.L. Schaff, S. Singh, K.-B. Kim, M.D. Sutcliffe, K.-S. Park, K.A. Janes, Fragmentation of Small-Cell Lung Cancer Regulatory States in Heterotypic Microenvironments, Cancer Res 81 (2021) 1853–1867.
[10] D.G. Duda, A.M.M.J. Duyverman, M. Kohno, M. Snuderl, E.J.A. Steller, D. Fukumura, R.K. Jain, Malignant cells facilitate lung metastasisby bringing their own soil, Proc Natl Acad Sci U S A 107 (50) (2010) 21677–21682.
[11] F.I. Staquicini, M. Cardó-Vila, M.G. Kolonin, M. Trepel, J.K. Edwards, D.N. Nunes, A. Sergeeva, E. Efstathiou, J. Sun, N.F. Almeida, S.-M. Tu, G.H. Botz, M.J. Wallace, D.J. O'Connell, S. Krajewski, J.E. Gershenwald, J.J. Molldrem, A.L. Flamm, E. Koivunen, R.D. Pentz, E. Dias-Neto, J.C. Setubal, D.J. Cahill, P. Troncoso, K.-A. Do, C.J. Logothetis, R.L. Sidman, R. Pasqualini, W. Arap, Vascular ligand-receptor mapping by direct combinatorial selection in cancer patients, Proc Natl Acad Sci U S A 108 (46) (2011) 18637–18642.
[12] K. Miller, K. Chinzei, G. Orssengo, P. Bednarz, Mechanical properties of brain tissue in-vivo: experiment and computer simulation, J Biomech 33 (11) (2000) 1369–1376.
[13] B. Clarke, Normal bone anatomy and physiology, Clin J Am Soc Nephrol 3 (Supplement 3) (2008) S131–S139.
[14] U.A. Gurkan, O. Akkus, The mechanical environment of bone marrow: a review, Ann Biomed Eng 36 (12) (2008) 1978–1991.
[15] J. Zamecnik, The extracellular space and matrix of gliomas, Acta Neuropathol 110 (5) (2005) 435–442.
[16] K. Saha, A.J. Keung, E.F. Irwin, Y. Li, L. Little, D.V. Schaffer, K.E. Healy, Substrate Modulus Directs Neural Stem Cell Behavior, Biophys. J . 95 (9) (2008) 4426–4438.
[17] Y.T. Wei, W.M. Tian, X. Yu, F.Z. Cui, S.P. Hou, Q.Y. Xu, I.-S. Lee, Hyaluronic acid hydrogels with IKVAV peptides for tissue repair and axonal regeneration in an injured rat brain, Biomed Mater 2 (3) (2007) S142–S146.
[18] J.B. Lian, G.S. Stein, Development of the osteoblast phenotype: molecular mechanisms mediating osteoblast growth and differentiation, Iowa Orthop J 15 (1995) 118–140.
[19] R.P. Dickenson, W.C. Hutton, J.R. Stott, The mechanical properties of bone in osteoporosis, J Bone Joint Surg Br 63-B(2) (1981) 233–238.
[20] J.P. Winer, P.A. Janmey, M.E. McCormick, M. Funaki, Bone marrow-derived human mesenchymal stem cells become quiescent on soft substrates but remain responsive to chemical or mechanical stimuli, Tissue Eng Part A 15 (1) (2009) 147–154.
[21] L.E. Barney, E.C. Dandley, L.E. Jansen, N.G. Reich, A.M. Mercurio, S.R. Peyton, A cell-ECM screening method to predict breast cancer metastasis, Integr Biol (Camb) 7 (2) (2015) 198–212.
[22] R. Fridman, G. Giaccone, T. Kanemoto, G.R. Martin, A.F. Gazdar, J.L. Mulshine, Reconstituted basement membrane (matrigel) and laminin can enhance the tumorigenicity and the drug resistance of small cell lung cancer cell lines, Proc Natl Acad Sci U S A 87 (17) (1990) 6698–6702.
[23] Y. Mori, N. Shimizu, M. Dallas, M. Niewolna, B. Story, P.J. Williams, G.R. Mundy, T. Yoneda, Anti-α4 integrin antibody suppresses the development of multiple myeloma and associated osteoclastic osteolysis, Blood 104 (2004) 2149–2154.
[24] C.C. Park, H. Zhang, M. Pallavicini, J.W. Gray, F. Baehner, C.J. Park, M.J. Bissell, B1 Integrin Inhibitory Antibody Induces Apoptosis of Breast Cancer Cells, Inhibits Growth, and Distinguishes Malignant from Normal Phenotype in Three Dimensional Cultures and in Vivo, Cancer Res. 66 (3) (2006) 1526–1535.
[25] C.C. Park, H.J. Zhang, E.S. Yao, C.J. Park, M.J. Bissell, B1 Integrin Inhibition Dramatically Enhances Radiotherapy Efficacy in Human Breast Cancer Xenografts, Cancer Res. 68 (11) (2008) 4398–4405.
[26] T. Muranen, L. Selfors, D. Worster, M. Iwanicki, L. Song, F. Morales, S. Gao, G. Mills, J. Brugge, Inhibition of PI3K/mTOR Leads to Adaptive Resistance in Matrix-Attached Cancer Cells, Cancer Cell 21 (2) (2012) 227–239.
[27] M.-Y. Lee, R.A. Kumar, S.M. Sukumaran, M.G. Hogg, D.S. Clark, J.S. Dordick, Three-dimensional cellular microarray for high-throughput toxicology assays, Proc. Natl. Acad. Sci. U.S.A. 105 (1) (2008) 59–63.
[28] T.J. Puls, X. Tan, M. Husain, C.F. Whittington, M.L. Fishel, S.L. Voytik-Harbin, Development of a Novel 3D Tumor-tissue Invasion Model for High-throughput, **High-content Phenotypic Drug Screening**. Scientific Reports 8 (2018) 13039.
[29] V.Z. Beachley, M.T. Wolf, K. Sadtler, S.S. Manda, H. Jacobs, M.R. Blatchley, J. S. Bader, A. Pandey, D. Pardoll, J.H. Elisseeff, Tissue matrix arrays for high-throughput screening and systems analysis of cell function, Nat. Methods 12 (12) (2015) 1197–1204.
[30] K.M. Mabry, M.E. Schroeder, S.Z. Payne, K.S. Anseth, Three-DimensionalHigh-Throughput Cell Encapsulation Platform to Study Changes in Cell-Matrix Interactions, ACS Appl. Mater. Interfaces 8 (34) (2016) 21914–21922.
[31] J. Rodrigues, M.A. Heinrich, L.M. Teixeira, J. Prakash, 3D In Vitro Model (R) evolution: Unveiling Tumor-Stroma Interactions, Trends in Cancer 7 (3) (2021) 249–264.
[32] G. Rijal, W. Li, A versatile 3D tissue matrix scaffold system for tumor modeling and drug screening, Sci. Adv. 3 (2017) 1–16.
[33] K. Han, S.E. Pierce, A. Li, K. Spees, G.R. Anderson, J.A. Seoane, Y.-H. Lo, M. Dubreuil, M. Olivas, R.A. Kamber, M. Wainberg, K. Kostyrko, M.R. Kelly, M. Yousefi, S.W. Simpkins, D. Yao, K. Lee, C.J. Kuo, P.K. Jackson, A. Sweet-Cordero, A. Kundaje, A.J. Gentles, C. Curtis, M.M. Winslow, M.C. Bassik, CRISPR screens in cancer spheroids identify 3D growth-specific vulnerabilities, Nature 580 (7801) (2020) 136–141.
[34] S.I. Fraley, Y. Feng, R. Krishnamurthy, D.-H. Kim, A. Celedon, G.D. Longmore, D. Wirtz, A distinctive role for focal adhesion proteins in three-dimensional cell motility, Nat. Cell Biol. 12 (6) (2010) 598–604.
[35] A.S. Meyer, S.K. Hughes-Alford, J.E. Kay, A. Castillo, A. Wells, F.B. Gertler, D. A. Lauffenburger, 2D protrusion but not motility predicts growth factor-induced cancer cell migration in 3D collagen, J. Cell Biol. 197 (2012) 721–729.
[36] H.T. Nia, L.L. Munn, R.K. Jain, Physical traits of cancer, Science 2020(1979) 370.
[37] L. Huwart, N. Salameh, L. ter Beek, E. Vicaut, F. Peeters, R. Sinkus, B.E. Van Beers, MR elastography of liver fibrosis: preliminary results comparing spin-echoand echo-planar imaging, Eur Radiol 18 (11) (2008) 2535–2541.
[38] M.J. Paszek, N. Zahir, K.R. Johnson, J.N. Lakins, G.I. Rozenberg, A. Gefen, C. A. Reinhart-King, S.S. Margulies, M. Dembo, D. Boettiger, D.A. Hammer, V. M. Weaver, Tensional homeostasisand the malignant phenotype, Cancer Cell 8 (3) (2005) 241–254.
[39] M.J. Paszek, V.M. Weaver, The tension mounts: mechanics meets morphogenesis and malignancy, J Mammary Gland Biol Neoplasia 9 (4) (2004) 325–342.
[40] K.R. Levental, H. Yu, L. Kass, J.N. Lakins, M. Egeblad, J.T. Erler, S.F.T. Fong, K. Csiszar, A. Giaccia, W. Weninger, M. Yamauchi, D.L. Gasser, V.M. Weaver, Matrix crosslinking forces tumor progression by enhancing integrin signaling, Cell 139 (5) (2009) 891–906.
[41] R. Kalluri, The biology and function of fibroblasts in cancer, Nat. Rev. Cancer 16 (9) (2016) 582–598.
[42] A. Kostic, C.D. Lynch, M.P. Sheetz, S.A. Aziz, Differential matrix rigidity response in breast cancer cell lines correlates with the tissue tropism, PLoS ONE 4 (7) (2009) e6361, https://doi.org/10.1371/journal.pone.0006361.
[43] I. Indra, V. Undyala, C. Kandow, U. Thirumurthi, M. Dembo, K.A. Beningo, An in vitro correlation of mechanical forces and metastatic capacity, Phys Biol 8 (1) (2011) 015015, https://doi.org/10.1088/1478-3975/8/1/015015.
[44] A.D. Schwartz, L.E. Barney, L.E. Jansen, T.V. Nguyen, C.L. Hall, A.S. Meyer, S. R. Peyton, A biomaterial screening approach reveals microenvironmental mechanisms of drug resistance, Integr. Biol. 9 (12) (2017) 912–924.
[45] T.V. Nguyen, M. Sleiman, T. Moriarty, W.G. Herrick, S.R. Peyton, Sorafenib resistance and JNK signaling in carcinoma during extracellular matrix stiffening, Biomaterials 35 (22) (2014) 5749–5759.
[46] E.A. Brooks, M.F. Gencoglu, D.C. Corbett, K.R. Stevens, S.R. Peyton, An omentum-inspired 3D PEG hydrogel for identifying ECM-drivers of drug resistant ovarian cancer, APL Bioengineering 3 (2019) 26106.
[47] C.M. Garvey, R. Lau, A. Sanchez, R.X. Sun, E.J. Fong, M.E. Doche, O. Chen, A. Jusuf, H.-J. Lenz, B. Larson, S.M. Mumenthaler, Anti-EGFR Therapy Induces EGF Secretion by Cancer-Associated Fibroblasts to Confer Colorectal Cancer Chemoresistance, Cancers (Basel) 12 (6) (2020) 1393, https://doi.org/10.3390/cancers12061393.
[48] N.P. Jobe, D. Rösel, B. Dvořánková, O. Kodet, L. Lacina, R. Mateu, K. Smetana, J. Brábek, Simultaneous blocking of IL-6 and IL-8 is sufficient to fully inhibit CAF-induced human melanoma cell invasiveness, Histochem. Cell Biol. 146 (2) (2016) 205–217.
[49] N. Kumari, B.S. Dwarakanath, A. Das, A.N. Bhatt, Role of interleukin-6 in cancer progression and therapeutic resistance, Tumor Biology 37 (9) (2016) 11553–11572.
[50] B. Erdogan, D.J. Webb, Cancer-associated fibroblasts modulate growth factor signaling and extracellular matrix remodeling to regulate tumor metastasis, Biochem. Soc. Trans. 45 (2017) 229–236.

[51] Y. Yi, S. Zeng, Z. Wang, M. Wu, Y. Ma, X. Ye, B. Zhang, H. Liu, Cancer-associated fibroblasts promote epithelial-mesenchymal transition and EGFR-TKI resistance of non-smallcell lung cancers via HGF/IGF-1/ANXA2 signaling, Biochimica et Biophysica Acta -Molecular Basis of Disease 1864 (3) (2018) 793-803.
[52] Y. Yu, C.-H. Xiao, L.-D. Tan, Q.-S. Wang, X.-Q. Li, Y.-M. Feng, Cancer-associated fibroblasts induce epithelial-mesenchymal transition of breast cancer cells through paracrine TGF-β signalling, Br. J. Cancer 110 (3) (2014) 724-732.
[53] A.B. Turke, K. Zejnullahu, Y.-L. Wu, Y. Song, D. Dias-Santagata,E. Lifshits, L. Toschi, A. Rogers, T. Mok, L. Sequist, N.I. Lindeman, C. Murphy, S. Akhavanfard, B.Y. Yeap, Y. Xiao, M. Capelletti, A.J. Iafrate, C. Lee, J.G. Christensen,J. A. Engelman, P.A. J änne, Preexistence and Clonal Selection of MET Amplification in EGFR Mutant NSCLC, Cancer Cell 17 (1) (2010) 77-88.
[54] D. Seneviratne, J. Ma, X. Tan, Y.-K. Kwon, E. Muhammad, M. Melhem, M. C. DeFrances, R. Zarnegar, Genomic Instability Causes HGF Gene Activation in Colon Cancer Cells, Promoting Their Resistance to Necroptosis, Gastroenterology 148 (1) (2015) 181-191.e17, https://doi.org/10.1053/j.gastro.2014.09.019.
[55] K. Matsumoto, H. Tajima, H. Okazaki, T. Nakamura, Negative regulation of hepatocyte growth factor gene expression in human lung fibroblasts and leukemic cells by transforming growth factor-β1 and glucocorticoids, J. Biol. Chem. 267 (1992) 24917-24920.
[56] O. Mungunsukh, R.M. Day, W.P. Tansey, Transforming growth factor-β1selectively inhibits hepatocyte growth factor expression via a micro-RNA-199-dependent posttranscriptional mechanism, Mol. Biol. Cell 24 (13) (2013) 2088-2097.
[57] N. Cheng, A. Chytil, Y.u. Shyr, A. Joly, H.L. Moses, Transforming growth factor-β signaling-deficient fibroblasts enhance hepatocyte growth factor signaling in mammary carcinoma cells to promote scattering and invasion, Mol. Cancer Res. 6 (10) (2008) 1521-1533.
[58] J. Oyanagi, N. Kojima, H. Sato, S. Higashi, K. Kikuchi, K. Sakai, K. Matsumoto, K. Miyazaki, Inhibition of transforming growth factor-βsignaling potentiates tumor cell invasion into collagen matrix induced by fibroblast-derived hepatocyte growth factor, Exp. Cell Res. 326 (2) (2014) 267-279.
[59] R.J. DeBerardinis, J.J. Lum, G. Hatzivassiliou, C.B. Thompson,The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation, Cell Metab. 7 (1) (2008) 11-20.
[60] R.G. Jones, C.B. Thompson, Tumor suppressorsand cell metabolism: A recipe for cancer growth, Genes Dev. 23 (5) (2009) 537-548.
[61] Y. Oren, M. Tsabar, H.F. Cabanos, M.S. Cuoco, E. Zaganjor, P.I. Thakore, M. Tabaka, C.P. Fulco, S.A. Hurvitz, D.J. Slamon, et al., Cycling cancer persister cells arise from lineages with distinct transcriptional and metabolic programs, bioRxiv (2020), https://doi.org/10.1101/2020.06.05.136358.
[62] C.-H. Lin, T. Jokela, J. Gray, M.A. LaBarge, Combinatorial Microenvironments Impose a Continuum of Cellular Responses to a Single Pathway-Targeted Anti-cancer Compound, Cell Rep 21 (2) (2017) 533-545.
[63] D.P. Kodack, E. Chung, H. Yamashita, J. Incio, A.M.M.J. Duyverman, Y. Song, C. T. Farrar, Y. Huang, E. Ager, W. Kamoun, S. Goel, M. Snuderl, A. Lussiez, L. Hiddingh, S. Mahmood, B.A. Tannous, A.F. Eichler, D. Fukumura, J. A. Engelman, R.K. Jain, Combined targeting of HER2 and VEGFR2 for effective treatmentof HER2-amplified breast cancer brain metastases,Proc Natl Acad Sci U S A 109 (45) (2012), https://doi.org/10.1073/pnas.1216078109.
[64] D.P. Kodack, V. Askoxylakis, G.B. Ferraro, Q. Sheng, M. Badeaux, S. Goel, X. Qi, R. Shankaraiah, Z.A. Cao, R.R. Ramjiawan, D. Bezwada, B. Patel, Y. Song, C. Costa, K. Naxerova, C.S.F. Wong, J. Kloepper, R. Das, A. Tam, J. Tanboon, D.G. Duda, C. R. Miller, M.B. Siegel, C.K. Anders, M. Sanders, M.V. Estrada, R. Schlegel, C. L. Arteaga, E. Brachtel, A. Huang, D. Fukumura, J.A. Engelman, R.K. Jain, The brain microenvironment mediates resistance in luminal breast cancer to PI3K inhibition through HER3 activation, Sci. Transl. Med. 9 (391) (2017), https://doi.org/10.1126/scitranslmed.aal4682.
[65] S.Y. Bae, N. Guan, R. Yan, K. Warner, S.D. Taylor, A.S. Meyer, Measurement and models accounting for cell death capture hidden variation in compound response, Cell Death Dis. 11 (4) (2020), https://doi.org/10.1038/s41419-020-2462-8.
[66] L.A. Byers, L. Diao, J. Wang, P. Saintigny, L. Girard, M. Peyton, L.i. Shen, Y. Fan, U. Giri, P.K. Tumula, M.B. Nilsson, J. Gudikote, H. Tran, R.J.G. Cardnell, D. J. Bearss, S.L. Warner, J.M. Foulks, S.B. Kanner, V. Gandhi, N. Krett, S.T. Rosen, E. S. Kim, R.S. Herbst, G.R. Blumenschein, J.J. Lee, S.M. Lippman, K.K. Ang, G. B. Mills, W.K. Hong, J.N. Weinstein, I.I. Wistuba, K.R. Coombes, J.D. Minna, J. V. Heymach, An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance, Clin. Cancer Res. 19 (1) (2013) 279-290.
[67] D.F. Quail, J.A. Joyce, Microenvironmental regulation of tumor progressionand metastasis, Nat. Med. 19 (11) (2013) 1423-1437.
[68] P. Sharma, S. Hu-Lieskovan, J.A. Wargo, A. Ribas, Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy, Cell 168 (4) (2017) 707-723.
[69] J. Winkler, A. Abisoye-Ogunniyan, K.J. Metcalf, Z. Werb, Conceptsof extracellular matrix remodelling in tumour progression and metastasis, Nat. Commun. 11 (2020) 1-19.
[70] S. Manole, E.J. Richards, A.S. Meyer, JNK pathway activation modulatesacquired resistance to EGFR/HER2-targeted therapies, Cancer Res. 76 (18) (2016) 5219-5228.
[71] M. Miyawaki, H. Yasuda, T. Tani, J. Hamamoto, D. Arai, K. Ishioka, K. Ohgino, S. Nukaga, T. Hirano, I. Kawada, K. Naoki, Y. Hayashi, T. Betsuyaku, K. Soejima, Overcoming EGFR bypass signal-induced acquired resistance to ALK tyrosine kinase inhibitors in ALK-translocated lung cancer, Mol. Cancer Res. 15 (1) (2017) 106-114.
[72] K.R. Amato, S. Wang, L.i. Tan, A.K. Hastings, W. Song, C.M. Lovly, C.B. Meador, F. Ye, P. Lu, J.M. Balko, D.C. Colvin, J.M. Cates, W. Pao, N.S. Gray, J. Chen, EPHA2 blockade overcomes acquired resistance to EGFR kinase inhibitors in lung cancer, Cancer Res. 76 (2) (2016) 305-318.
[73] Z. Zhang, J.C. Lee, L. Lin, V. Olivas, V. Au, T. LaFramboise, M. Abdel-Rahman, X. Wang, A.D. Levine, J.K. Rho, Y.J. Choi, C.-M.Choi, S.-W. Kim, S.J. Jang, Y. S. Park, W.S. Kim, D.H. Lee, J.-S. Lee, V.A. Miller, M. Arcila, M. Ladanyi, P. Moonsamy, C. Sawyers, T.J. Boggon, P.C. Ma, C. Costa, M. Taron, R. Rosell, B. Halmos, T.G. Bivona, Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer, Nat. Genet. 44 (8) (2012) 852-860.
[74] T.R. Wilson, J. Fridlyand, Y. Yan, E. Penuel, L. Burton, E. Chan, J. Peng, E. Lin, Y. Wang, J. Sosman, A. Ribas, J. Li, J. Moffat, D.P. Sutherlin, H. Koeppen, M. Merchant, R. Neve, J. Settleman, Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors, Nature 487 (7408) (2012) 505-509.
[75] P.D. Dunne, D.G. McArt, J.K. Blayney, M. Kalimutho, S. Greer, T. Wang, S. Srivastava, C.W. Ong, K. Arthur, M. Loughrey, K. Redmond, D.B. Longley, M. Salto-Tellez, P.G. Johnston, S. Van Schaeybroeck, AXL is a key regulator of inherent and chemotherapy-induced invasion and predicts a poor clinical outcome in early-stage colon cancer, Clin. Cancer Res. 20 (1) (2014) 164-175.
[76] O.-Y. Revach, O. Sandler, Y. Samuels, B. Geiger, Cross-talk between receptor tyrosine kinases AXL and ERBB3 regulates invadopodia formation in melanoma cells, Cancer Res. 79 (10) (2019) 2634-2648.
[77] T.G. Kolda, B.W. Bader, Tensor Decompositionsand Applications, SIAM Rev 51 (3) (2009) 455-500.
[78] A.M. Farhat, A.C. Weiner, C. Posner, Z.S. Kim, B. Orcutt-Jahns, S.M. Carlson, A. S. Meyer, Modeling cell-specific dynamics and regulation of the common gamma chain cytokines, Cell Rep 35 (4) (2021) 109044, https://doi.org/10.1016/j.celrep.2021.109044.
[79] Z.C. Tan, M.C. Murphy, H.S. Alpay, S.D. Taylor, A.S. Meyer, Tensor-structured decomposition improves systems serology analysis, Mol Syst Biol 17 (9) (2021), https://doi.org/10.15252/msb.202110243.
[80] C. Martino, L. Shenhav, C.A. Marotz, G. Armstrong, D. McDonald, Y. Vázquez-Baeza, J.T. Morton, L. Jiang, M.G. Dominguez-Bello, A.D. Swafford, E. Halperin, R. Knight, Context-aware dimensionality reduction deconvolutes gut microbial community dynamics, Nat Biotechnol 39 (2) (2021) 165-168.
[81] A.B. Keenan, S.L. Jenkins, K.M. Jagodnik, S. Koplev, E. He, D. Torre, Z. Wang, A. B. Dohlman, M.C. Silverstein, A. Lachmann, et al., The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations, Cell Systems 6 (2018) 13-24.
[82] P.-P. Kuang, X.-H. Zhang, C.B. Rich, J.A. Foster, M. Subramanian, R.H. Goldstein, Activation of elastin transcription by transforming growth factor-in human lung fibroblasts, Am J Physiol Lung Cell Mol Physiol 292 (2007) 944-952.
[83] M. J önsson, J. Dejmek, P.-O. Bendahl, T. Andersson, Loss of Wnt-5a Protein Is Associated with Early Relapse in Invasive Ductal Breast Carcinomas, Cancer Res 62 (2) (2002) 409-416.
[84] K.K. Yang, Z. Wu, F.H. Arnold, Machine-learning-guideddirected evolution for protein engineering, Nat Methods 16 (8) (2019) 687-694.
[85] J.-E. Shin, A.J. Riesselman, A.W. Kollasch, C. McMahon, E. Simon, C. Sander, A. Manglik, A.C. Kruse, D.S. Marks, Protein design and variant prediction using autoregressive generative models, Nat Commun 12 (2021) 2403.
[86] B.L. Hie, K.K. Yang, Adaptive machine learning for protein engineering, Curr. Opin. Struct. Biol. 72 (2022) 145-152.
[87] P.A. Romero, A. Krause, F.H. Arnold, Navigating the protein fitness landscape with Gaussian processes, Proc Natl Acad Sci U S A 110 (2013) E193-201.