Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
Cell. Author manuscript; available in PMC 2017 August 11.

Published in final edited form as:
Cell. 2016 August 11; 166(4): 1041–1054. doi:10.1016/j.cell.2016.07.007.

Phosphoproteome Integration Reveals Patient-Specific Networks in Prostate Cancer

Justin M. Drake[#a,n,1], **Evan O. Paull**[#m], **Nicholas A. Graham**[b,c,o], **John K. Lee**[d,e], **Bryan A. Smith**[a], **Bjoern Titz**[b,c], **Tanya Stoyanova**[a,p], **Claire M. Faltermeier**[e], **Vladislav Uzunangelov**[m], **Daniel E. Carlin**[m,q], **Daniel Teo Fleming**[m], **Christopher K. Wong**[m], **Yulia Newton**[m], **Sud Sudha**[l], **Ajay A. Vashisht**[f], **Jiaoti Huang**[g,h,k], **James A. Wohlschlegel**[f], **Thomas G. Graeber**[b,c,g,i], **Owen N. Witte**[a,c,j,k,1], and **Joshua M. Stuart**[m,1]

[a]Department of Microbiology, Immunology, and Molecular Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

[b]Crump Institute for Molecular Imaging, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

[c]Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

[d]Division of Hematology and Oncology, Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

[e]Molecular Biology Institute, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

[f]Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

[g]Jonsson Comprehensive Cancer Center, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

[h]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

[i]California NanoSystems Institute, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

[j]Howard Hughes Medical Institute, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

[1]To whom correspondence should be addressed: Joshua M. Stuart: jstuart@ucsc.edu; Owen N. Witte: owenwitte@mednet.ucla.edu; Justin M. Drake: justin.drake@cinj.rutgers.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

AUTHOR CONTRIBUTIONS
Conceptualization and Methodology, J.M.D., E.O.P., N.A.G., T.G.G, J.M.S., and O.N.W.; Investigation, J.M.D., E.O.P., N.A.G., J.K.L, B.A.S., T.S., and C.M.F.; Resources, A.A.V., J.A.W., S.S., C.K.W., Y.N., E.O.P, and D.T.F; Formal Analysis, J.M.D., E.O.P., N.A.G., J.K.L., B.T., and J.H.; Writing – Original Draft, J.M.D., E.O.P., J.M.S., and O.N.W.; Writing – Review and Editing, All authors.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[k]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, CA 90095

[l]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI 48109.

[m]Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064.

[n]Rutgers Cancer Institute of New Jersey and Department of Medicine, Rutgers-Robert Wood Johnson Medical School, New Brunswick, NJ, 08903.

[o]Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, CA 90089.

[p]Department of Radiology, Canary Center at Stanford for Cancer Early Detection, Stanford University, Palo Alto, CA 94304.

[q]Department of Medicine, University of California, San Diego, CA 92093.

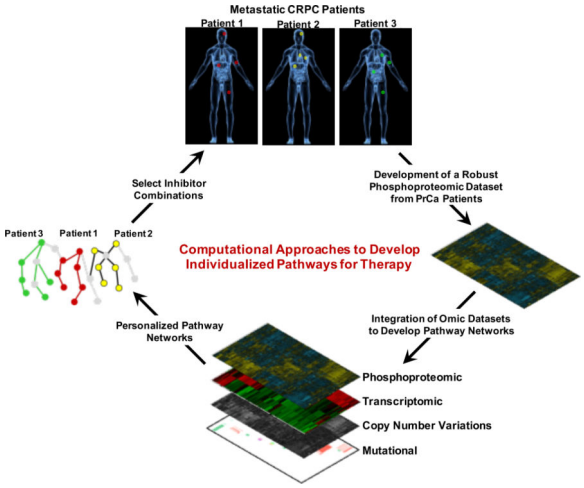
[#] These authors contributed equally to this work.

SUMMARY

We used clinical tissue from lethal metastatic castration resistant prostate cancer (CRPC) patients obtained at rapid autopsy to evaluate diverse genomic, transcriptomic, and phosphoproteomic datasets for pathway analysis. Using Tied Diffusion through Interacting Events (TieDIE), we integrated differentially expressed master transcriptional regulators, functionally mutated genes, and differentially activated kinases in CRPC tissues to synthesize a robust signaling network consisting of druggable kinase pathways. Using MSigDB hallmark gene sets, six major signaling pathways with phosphorylation of several key residues were significantly enriched in CRPC tumors after incorporation of phosphoproteomic data. Individual autopsy profiles developed using these hallmarks revealed clinically relevant pathway information potentially suitable for patient stratification and targeted therapies in late stage prostate cancer. Here we describe phosphorylation-based cancer hallmarks using integrated personalized signatures (pCHIPS) that sheds light on the diversity of activated signaling pathways in metastatic CRPC while providing an integrative, pathway-based reference for drug prioritization in individual patients.

Graphical Abstract

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



INTRODUCTION

DNA and RNA sequencing data have been used to analyze key transcriptional targets, cell surface molecules, or pathways at work in cancer (Aytes et al., 2014; Cancer Genome Atlas Network, 2012a, 2012b; Cancer Genome Atlas Research Network, 2015) (Grasso et al., 2012; Robinson et al., 2015; Taylor et al., 2010; Vaske et al., 2010). One goal from these approaches is to select mutations corresponding to genes or pathways from tumors and then match targeted therapies based on these lesions. However, missing from many genomic or transcriptomic analyses is further measurement and extension of the activated pathways that are found by such approaches using mass spectrometry-based phosphoproteomics.

Protein phosphorylation remains a critical, rate-limiting step for the regulation of signaling pathways over numerous biological events. Determining both the level of phosphorylation and what residues are phosphorylated on a given protein may inform us about the activity of kinases and phosphatases as well as uncover new functional information that was previously underappreciated. Cellular signaling can also be controlled through the recruitment of protein domains (such as SH2 and SH3) to specific phosphorylation sites on kinases (Pawson, 2004). Protein phosphorylation leads to a cascade of downstream signaling events important for cell maintenance and survival and dysregulation of this process has been implicated in many diseases including cancer (Hunter, 2009). It stands to reason that the implementation of phosphoproteomics, coupled with traditional mRNA-based approaches, may provide greater clues to these signaling events than either alone.

Recent computational advances allow for the simultaneous examination of genomic, phosphoproteomic and transcriptional data, in the context of prior pathway knowledge (Cancer Genome Atlas Research Network, 2013, 2014a; Huang et al., 2013). These methods have the advantage of being able to detect events that are below the threshold of statistical significance when examining a single dataset in isolation, as well as finding evidence for functional interactions between proteins. For instance, a multistep systems-level approach was recently used to find genomic events that drive tumorigenesis in glioblastoma by first finding transcriptional ‘master regulators’ that are predicted to control a large number of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

differentially expressed genes, and then reversing pathway database interactions to look 'upstream' for genomic events that may be influencing (and statistically associated with) the activity of regulators active in individual patients (Chen et al., 2014). Similarly, the TieDIE algorithm (Paull et al., 2013) was recently used in a study of thyroid papillary carcinoma to identify signaling pathways linking mutant BRAF and RAS genes to transcription factors and signaling proteins with altered activity in tumor samples. It was found that the small GTPase RHEB, a known regulator of mTOR activity, was a contributing factor to the differences observed between BRAF and RAS mutants (Cancer Genome Atlas Research Network, 2014a). Both of these analyses ranked candidate regulators according to multiple data types and pathway context, though the latter analysis focused on identifying intermediate 'linking' genes that are strongly implicated by the combination of pathway context and the incorporation of multiple data types.

Here, we set out to define the global picture of signaling pathways in lethal prostate cancer through dataset integration. We developed a complete and extensive new dataset of the phosphoproteome in metastatic CRPC by extending our analysis to phosphoserine and phosphothreonine peptides and then combining this information with our previously published phosphotyrosine peptide data (Drake et al., 2013). To develop comprehensive pathway networks that are both enriched and activated in CRPC, we used TieDIE to integrate independent datasets of mutations, transcriptional changes, and phosphoproteome activities in an unbiased manner from a similar set of tumor samples obtained at rapid autopsy (Rubin et al., 2000). The integration of tissue samples from a single autopsy program allowed us to make inferences on the connections between the mRNA and phosphoproteome datasets. In addition, both mRNA and phosphoproteome data were available for several of the patients. Using this information, we introduce a new tool called phosphorylation-based cancer hallmarks using integrated personalized signatures (pCHIPS) to establish patient-specific pathways marking key signaling events for possible targeting.

RESULTS

Development of a robust phosphoproteomic dataset for integration

We analyzed the phosphoproteome of metastatic CRPC tissues and identified 297 phosphotyrosine (pY) peptides, (Drake et al., 2013) and 8,051 phosphoserine/phosphothreonine (pST) peptides from 54 total runs corresponding to 27 samples of interest (11 treatment-naïve, 16 metastatic CRPC; **Data S1a-S1c**) using quantitative label free mass spectrometry (**Fig. 1a**). Hierarchical clustering revealed similarities in the groupings of the samples compared to previously published pY peptide data (Drake et al., 2013). For example, cell lines were distinct from primary tissues and treatment naïve localized prostate cancer clustered independently from metastatic CRPC tissues (**Fig. 1b**). Within this dataset, we were able to directly identify phosphopeptides corresponding to 74 kinases, 18 of which were differentially phosphorylated (FDR<0.05, >1.5 fold) in metastatic CRPC tissues (**Data S1d**). To get an initial sense of the biological processes and pathways enriched in metastatic CRPC, we performed gene set enrichment analysis (GSEA) typically used for RNA-based datasets (Subramanian et al., 2005) as well as kinase-substrate enrichment analysis (KSEA) better tailored for phosphoproteomic-based datasets that we and others have previously

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

established (Drake et al., 2012, 2013) (Casado et al., 2013) (Newman et al., 2013). GSEA of canonical processes and pathways detected over-representation of mRNA splicing and processing, DNA replication, and AR transcription factor pathways as well as loss of integrin signaling, focal adhesion, and axon guidance pathways in metastatic CRPC (**Fig. 1c**). GSEA of transcription factor targets revealed several E2F family members as well as the MYC/Max complex to be over-represented in metastatic CRPC (**Fig. 1d, Data S1e**). The enrichment of E2F target genes is intriguing as we have previously been able to connect a primary basal stem cell signature to small cell neuroendocrine carcinoma with this gene set (Smith et al., 2015). KSEA further implicated enrichment of several kinases in metastatic CRPC including cyclin dependent kinases (CDK2/CDK3), casein kinase 2 (CK2), and β-adrenergic receptor kinases (GRK2/GRK3) (**Fig. 1e, Data S1f**). Many of the genes and kinases identified through GSEA and KSEA have previously been implicated in prostate cancer confirming the validity of our dataset (Gioeli et al., 1999; Li et al., 2014; Lu et al., 1997; Wang et al., 2006). Importantly, the large number of pST identifications enabled an integrated computational approach to identify pathways implicated from this phosphoproteomic dataset.

Integration of transcriptomic and phosphoproteomic datasets using TieDIE

To prioritize kinases that are likely to regulate the observed gene expression profile of metastatic samples, and be related to genomic aberrations observed in prostate cancer, we developed an original computational pipeline using the TieDIE algorithm, a pathway-based method developed to find protein and gene interactions related to disease (Cancer Genome Atlas Research Network, 2014b) (Paull et al., 2013). The approach integrates complementary transcriptomic and genomic datasets collected from different metastatic CRPC tissues or patients, as well as prior knowledge in the form of pathway databases, to find subnetworks of related proteins implicated by multiple forms of biological evidence (**Fig. 2, Fig. S1a**). We first applied the Master Regulator Inference algorithm (MARINa) algorithm (Alvarez et al., 2015), a method to infer the activity of a given protein based on the differential expression/phosphorylation of the targets it regulates. This allowed us to identify transcription factors with differential activity (repression/activation) as well as differentially activated kinase regulators (based on the predicted upstream kinases for each phosphopeptide) in metastatic CRPC samples as compared with treatment naïve prostate cancers (**Data S1g, S1h).** In addition, kinases directly identified by the mass spectrometer in our phosphoproteomic dataset (phosphorylated kinases) were merged with the kinase regulators before input to TieDIE.

From this differential analysis, we were able to incorporate 74 transcription factor (TF) regulators, 14 inferred kinases regulators, and 24 differentially phosphorylated kinases (**Fig. 2**). The dataset used for analysis included a matrix of inferred and measured kinases for the same 16 metastatic CRPC samples, and a matrix of inferred transcription factors for 27 metastatic samples. Several patients with marked differences in response to therapy (e.g. anti-androgens or chemotherapy, **Data S1a**) had highly similar transcriptomes as evidenced by the transcription factors identified by MARINa for the 16 metastatic CRPC samples. Thus, the differences in protein level signaling could help explain these differences as well as offer new treatment options that could abrogate the signaling upstream of these TF-driven

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

circuits (**Fig. S1b**). These phosphoproteomic and transcriptomic matrices only overlapped for seven patient samples (from six unique patients) and are used for our patient-specific networks. As a third input to TieDIE, a background of somatic mutations and copy-number aberrations was collected from a large number of prostate cancer samples from multiple datasets. The strength of our approach is that it unites these diverse data, collected on different patient samples, to identify pathways implicated by several viewpoints.

We asked if the kinases inferred by master regulator analysis or identified by phosphorylation status were significantly interrelated to the set of genes involved in somatic mutations or to those genes implicated as transcription factors by master regulator analysis of the expression data. A conservative test that permuted the input gene sets over 1,000 replications demonstrated that the kinases are indeed “nearby” in pathway space to genes with genomic or transcriptomic alterations (**Fig. S2a**). Thus, despite the fact that the inferred TF regulators are not directly targeted for phosphorylation by the kinase regulators more than we expect by chance, the TFs are “close” in network space suggesting longer paths are needed to encompass the signaling transduced from the phosphoproteome to the transcriptome. The TieDIE solutions were robust to changes in the method's single parameter (alpha) that controls the size of the network solutions (**Fig. S2b**). Using varying settings for alpha, we selected a compact network with a high level of specificity (**Fig. S2b**), which consisted of 338 nodes -- 40 kinases, 53 transcription factors, 86 amplified/deleted/mutated genes, and 163 linking proteins – connected by 1,889 edges. To simplify this network, interactions that were supported by the phosphoproteomic data were retained. This resulted in a network we refer to as the “scaffold network” for metastatic CPRC that contained 122 nodes and 256 edges (**Fig. S2c, Data S1i**). TieDIE used 61 genes that were not included in the input set, termed “linker” proteins, to produce the scaffold network. Consistent with their predicted embedding in metastatic signaling, these 61 linkers were found to have phospho-residues with significantly higher phosphorylation abundance in metastatic CRPC compared to treatment naïve prostate cancer (**Fig. S2d**; $p<4.5 \times 10^{-6}$).

The diffusion process employed by TieDIE controls for the spurious inclusion of “hub” genes -- those genes with many connections in the generic background network potentially as a result of study-bias. However, it was possible other factors could influence a linker's inclusion that would undermine the network's relevance to the given input set. Thus, we explicitly tested for inclusion bias in the linker genes by quantifying the frequency with which they were included in random TieDIE solutions constructed using simulated arbitrary input gene sets of the same sizes as the provided inputs. 1,000 simulations demonstrated that the linker genes were included at frequencies no higher than other background genes (**Fig. S2e,f**). Furthermore, no inclusion bias was observed for linkers with higher connectivity or centrality.

The scaffold network was found to have sub-networks significantly represented by cancer-related MSigDB cancer hallmarks gene sets including AKT/mTOR/MAPK signaling, nuclear receptor signaling (which includes the androgen receptor (AR) pathway), the cell cycle, DNA repair, stemness, and migration (**Fig. S2g-l, Data S1j**) as well as established prostate cancer-specific pathways recorded in the KEGG pathway database (8.8 fold enrichment, p<4.8e-15 or based on DAVID overlap analysis, david.ncifcrf.gov). Sub-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

network views in Figures 3 and S2 show only genes that fall within both the curated hallmark gene sets and the previously generated scaffold network, with grey nodes representing genes that are in the scaffold network but not in the respective hallmark.

To determine the distinct biology revealed by the phosphoproteomic data in metastatic CRPC, we re-ran the same TieDIE analysis to obtain a comparably sized scaffold network without the phosphoproteomics information and compared its cancer hallmark enrichment against the one found when all the data was included (**Fig. S3a, b**). We found significant enrichment of AKT/mTOR/MAPK signaling pathways when the phosphoproteomic data was included whereas enrichment was only marginal without inclusion of this data (**Fig. 3a**; 4.6 vs. 1.6 −log10 hypergeometric p-value). In addition, we found higher relative enrichment of proteins involved in cell cycle, DNA repair, and nuclear receptor pathways when the phosphoproteomic data was included (**Fig. 3a**; cell cycle: 20.5 vs. 14.7, −log10 hypergeometric p-value; DNA repair: 6.6 vs. 5.1; nuclear receptor signaling: 8.1 vs 5.8). Inspecting each sub-network through our phosphoproteomic data revealed several newly discovered enzymatically active phospho-residues enriched in metastatic CRPC. This included MAPK signaling targets (RPS6KA4 S^{343}/S^{347}, S^{682}/T^{687}), cell cycle targets (MCM2 $S^{40/41}$, S^{27}), and the DNA repair kinase PRKDC T^{2609}, S^{2612} (**Fig. 3b-i**). Several other kinases within these sub-networks were hyperphosphorylated at residues with unknown function in metastatic CRPC including PRKAA2 S^{337}, MAPK14 S^{2}, STK39 S^{385}, NIPBL S^{318}, and SNW1 S^{14} implicating several more new targets for investigation. Lower relative enrichment in metastatic CRPC was observed for TGF-β or WNT/β-catenin signaling pathways when the phosphoproteomic data was included. This can be partially explained by the lack of overlap between kinases identified directly by the phosphoproteomic data, lowering the relative importance of these gene sets after its inclusion. However, our differential analysis of metastatic CRPC to treatment naïve prostate cancer did observe strong enrichment of both TGF-β and WNT/β-catenin signaling pathways after the integration of the phosphoproteomic data in metastatic CRPC (**Fig. S3b**). We also observed that the fraction of proteins overlapping with any of the “hallmark” gene sets to be higher when including the phosphoproteomic data, accounting for any potential study bias. These results provide evidence of actionable phosphorylation events in metastatic CRPC, several of which have previously been implicated in this disease including PRKDC, PRKAA2, and AKT (Goodwin et al., 2015; Yu et al., 2015) (Park et al., 2009) while others such as RPS6KA4 and MCM2 represent new drug targets.

Identification of patient-specific integrated networks

To identify patient-specific signaling routes we used the integrated phosphoproteome-transcriptome network to analyze the 6 metastatic CRPC patients that had both transcriptomic and phosphoproteomic data available. We ran the VIPER algorithm, a sample-specific version of MARINa that infers the activity of proteins based on measurements of the targets they regulate (Alvarez et al., 2015), to summarize the transcriptomic and phosphoproteomic data vectors of each patient into protein activity inferences of a relatively small number of transcriptional and kinase ‘master regulators’, respectively (**Fig. 4a**). Similar to the full dataset, the transcriptional master regulators were highly similar across this patient cohort but the inferred and phosphorylated kinases were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

somewhat different between each of the individual patients (**Fig. 4b**). Importantly, we found that phosphoproteomic-driven VIPER inferences of protein activity for patient RA55 were highly correlated and consistent across two metastatic sites (**Fig. S4a, b**). Similarly, a second patient, RA43, was found to have higher pairwise correlations (on average) between samples when comparing inferred protein activities from VIPER, than when comparing the relative phosphorylation of peptides (**Fig. S4c-h**). We asked if the high differential kinase activities in CRPC compared to primary prostate cancer inferred by VIPER were concordant with the measured phosphorylation levels. We measured the correlation between VIPER-inferred and measured activity for 26 phosphoresidues for which functional annotations could be found recorded in the phosphosite.org database (**Fig. S4i**). Of these, 10 had significant positive correlations with VIPER activity (BH FDR < 0.1); none had significant anti-correlations. Of the 10 phosphoresidues with positive correlations, 8 were annotated on enzymatically active sites, consistent with the higher activity predicted by VIPER for the metastatic samples.

To generate patient-specific network models, we intersected sample-specific VIPER inferences, the phosphorylation abundance of select phosphoresidues, mutations, copy number gains, and copy number deletions with the integrated TieDIE "scaffold network" solution. Proteins could then be prioritized by their activities and by their ability to regulate (or be regulated by) other genes implicated in a patient's network. The use of the scaffold allows the cohort-level data to inform the analysis of a single patient's data, which improves the accuracy and robustness of the resulting networks (**Fig. S4j,k**). The scaffold network was also found to generalize to unseen patient data based on a leave-one-out test in which the scaffold network was rebuilt after removing the data for each patient in turn as assessed by multiple different sub-sampling tests (**Fig. S4l, m**).

Assessment of actionable pathways for personalized medicine predictions

We created a visualization scheme we refer to as phosphorylation-based cancer hallmarks using integrated personalized signatures, or pCHIPS (**Fig. 5a, Fig. S5, Data S1k**). pCHIPS enables visual inspection and prioritization of the signaling pathways specific to each individual patient and is useful for suggesting personalized treatment options. Dissecting the pCHIPS of patient RA40, we observed 4 significantly enriched subnetworks including a large active network related to cell cycle processes (**Fig. 5b-f**). Interestingly, this was the only patient that we analyzed with a missense mutation and deletion in the tumor-suppressor gene APC. While frequently observed in colorectal cancers, APC mutations can occur in other cancers (Kandoth et al., 2013) where its inactivation leads to increased β-catenin activity (Morin et al., 1997). Indeed, we observed strong phosphorylation of the enzymatic active site of β-catenin (S^{675}). The putative activation of EZH2 is also linked to β-catenin activation in several cancers including hepatocellular carcinoma and breast cancer (Chang et al., 2011; Cheng et al., 2011) . EZH2 activation in this patient is supported by both low level amplification (Mermel et al., 2011) and hypophosphorylation of residue T^{487} (a marker for ubiquitination of EZH2) as well as amplification of DNA methyltransferase 3 (DNMT3) and predicted transcriptional activity of DNTM1 (Ning et al., 2015) (**Fig. 5c**). Further, the amplification and predicted transcriptional activity of SUV39H1 correlates with EZH2 expression in tumor development (Pandey et al., 2014), consistent with our observations. Mechanistically, EZH2 activity is sufficient for activation of AKT1 (Gonzalez et al., 2011),

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

which we observed through both hyperphosphorylation of the enzymatic active site T^{308} as well as indirectly through the prediction of high AKT activity by VIPER analysis (**Data S1h**). Together, this information implicates the involvement of β-catenin, AKT1, and EZH2 in contributing to altered cell cycle regulation and growth and suggests that targeted inhibition within this network could have been useful in this patient. Similar mechanisms related to other signaling pathways for other patients can also be described (**Data S2a-f**) as well as inter-patient pathway differences within the same hallmark (**Fig. 6**).

Given a complex patient specific network, how do we use it to select an optimal treatment strategy? Under the assumption that we seek to reverse as many altered gene activities found in a patient, we consider here the idea of using a minimum combination of targets that influence the largest area in a patient's network. Understanding the nesting of gene regulatory signals provides information about how to select genes for this purpose. Therefore, we developed a hierarchy of therapeutic kinase targets based on KSEA-derived relationships between kinases and the sets of peptides each kinase regulates, as well as evaluation of the cancer hallmarks and pathways of each individual patient (**Fig. 7a**). The hierarchy reveals the top kinase targets for every individual patient that we analyzed and the corresponding therapeutic intervention. Given this structure, targeting of a single kinase such as PRKDC may be sufficient to blunt the activity of other kinases that phosphorylate many of the same targets (NEK3, PRKCZ) and those that are, additionally, predicted to be phosphorylated by PRKDC (ATM, IKBKB).

To assess the validity of our kinase predictions, we developed kinase hierarchies for prostate cancer cell lines, LNCaP, 22Rv1, and DU-145 for which external data were available and for which we had transcriptomic and phosphoproteomic data. Using existing *in vitro* drug response data from the Genomics of Drug Sensitivity in Cancer (Yang et al., 2013) (GDSC, http://www.cancerrxgene.org/, **Data S1l**), we compared the relative sensitivity, measured in -ln(IC_{50}) values, for all of the inhibitors that target the predicted kinases. The relative rank of the personalized network prediction score was significantly correlated to kinase inhibitor sensitivity in an aggressive DU-145 cell line ($p<0.024$; kendall-tau rank correlation) but not for a second aggressive cell line 22Rv1 (**Fig. S7a-e**). In the case of DU-145 cells, the highest activity corresponded well with the strongest response (MAPK14, EGFR, and PTK2) (**Fig. S7a**). Interestingly, for 22Rv1, PRKDC had the highest inferred activity of all kinases and was found to be essential for 22Rv1 survival in a genome-wide RNA silencing screen performed by the Achilles project (**Fig. S7f**). In addition, the predictions for 22Rv1 were also found to be weakly positively correlated overall with the gene essentiality data ($p<0.07$; **Fig. S7f**). Indeed, a recent publication evaluated PRKDC function in a panel of prostate cancer cell lines, including 22Rv1, and observed that inhibiting PRKDC activity was effective at delaying metastasis formation after tail vein injection (Goodwin et al., 2015). This result provides evidence that PRKDC activity in the 22Rv1 cell line, as predicted in our models, is essential for development of metastases *in vivo* and targeting this kinase with a PRKDC selective inhibitor was effective at blocking this process. Taken together, both aggressive cell line predictions could be corroborated with either the external drug sensitivity or gene essentiality data despite the known sources of inherent noise in both

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

profiling studies. Future in vitro and in vivo experiments will be necessary to further confirm the results of these data-induced networks.

DISCUSSION

Targeting the synthesis of androgens or AR directly is the current standard of care in advanced prostate cancer and most tumors are responsive to these therapies. Our network models identified and implicated AR signaling as active in this cohort of patient samples. However, current clinical inhibitors targeting AR alone in late stage prostate cancer patients provide survival benefits of only 3-4 months (de Bono et al., 2011; Scher et al., 2012). Previously, we analyzed the abundance of phosphotyrosine peptides using unbiased quantitative mass spectrometry to identify tyrosine kinase signaling pathways in metastatic CRPC (Drake et al., 2013). Together with this work, we have provided clues into the signaling pathways that are activated in metastatic prostate cancer patients who had received, and became resistant to, anti-androgen therapy and that individual patients with multiple metastatic lesions displayed similar kinase signaling profiles (Drake et al., 2013). If kinase activity is one mechanism by which prostate tumors bypass anti-androgen therapy, then an interesting concept would be the implementation of kinase inhibitor therapies in combination with AR targeted agents. One exception would be patients who develop a lethal variant of CRPC termed small cell neuroendocrine carcinoma (SCNC) as these tumors, on average, lack AR signaling and have been shown to be driven by oncogenes such as MYCN or aurora kinase A (AURKA) (Beltran et al., 2011; Lee et al., 2016). For patients with intact AR signaling, several clinical trials are underway to address combinatorial therapy in metastatic prostate cancer including inhibition of AKT, MET/VEGFR2, or SRC in combination with AR blockade (ClinicalTrials.gov Identifiers: NCT01485861, NCT01995058, NCT01685125). While the results of these trials are still pending, the need for models to predict combinatorial therapies through joint analysis of high-throughput datasets that interrogate multiple aspects of the cell in clinical tissues are essential to identify the key biomarkers for patient stratification and therapy.

We presented pCHIPS as a method to capture multiple perspectives of cellular biology from phosphoproteomic and transcriptomic data integration and present this data at the individual level. Our analysis implicated several signaling proteins such as PRKDC, PRKAA2, PTK2, RPS6KA4, and CDK family members within these pathways as possible new therapeutic targets and/or biomarkers in prostate cancer. In nearly every case, we note a different implicated therapy suggested by the phosphoproteomic data. Interestingly, the transcriptional regulators were found to be more consistent across the metastatic samples while the kinase activities were found to vary. This suggests that the dominant signaling networks driving the biology of each patient may converge on the downstream transcriptional programs identified by the gene expression data. Several patients with marked differences in response to therapy (e.g. anti-androgens or chemotherapy, **Data S1a**) have highly similar transcriptomes as evidenced by the transcription factors identified by VIPER for the 16 metastatic CRPC samples. The differences in protein level signaling could help explain the variable responses and offer new treatment options to abrogate the signaling upstream of these TF-driven circuits.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

An intriguing question is whether network-based approaches like the one presented here yield similar or complementary information about treatment strategy compared to those based on so-called actionable mutations. First, for the five patient samples for which we had genomics data, we found cases in which different hallmarks were implicated with the patient-specific networks compared to using only the mutational information. Seven hallmarks were concordant across the patients, seven were discordant, and five agreed in a subset of patients (see **Fig. S6b**). Second, we used the models derived from cell lines to investigate whether the presence of mutations or inferred activated kinases were more informative about drug sensitivity. We tabulated the data and found that the inferred phospho-based activities were as indicative of drug response as the presence of somatic mutations in those pathways and, when averaged across pathways and cell lines, these data suggest one type of data is sufficient to implicate pathway targets (see **Fig. S7g, Data S1l, S1m**). Importantly, for an individual patient afflicted with a tumor that lacks mutations in known actionable pathways, phosphoproteomic data could be informative to prioritize treatment.

Continued development of these computational strategies will enable better determination of the specific vulnerabilities in individual tumors as our work sheds light on the diversity of the activated signaling pathways in metastatic CRPC tumors. These data and resulting pathway-based inferences establish a window into the regulation of protein signaling of aggressive tumors and a valuable reference for further investigation. Specifically, cell-line specific networks and kinase targets could be selected to inhibit cell growth or to test whether inhibition of kinases at higher levels can abrogate those at lower levels of the signaling hierarchy. Ultimately, further interrogation of these networks in appropriate pre-clinical models to assess co-targeting or combination therapies are necessary and warrant future investigation into patient stratification prior to clinical intervention. To facilitate such follow-up investigations, we have made available several modalities of the data and results. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD002286 (Vizcaíno et al., 2014). In addition to these data, the results are available through the UCSC TumorMap portal (http://tumormap.ucsc.edu/), providing public access to the assayed and predicted phosphorylation levels for primary and metastatic prostate cancer datasets from several public sources. Finally, we provide an online tool (https://sysbiowiki.soe.ucsc.edu/pchips) for users to input gene expression data to develop their own phosphoproteome-guided networks without the need for their own phosphoproteome data.

EXPERIMENTAL PROCEDURES

Quantitative analysis of phospho-serine and -threonine peptides by quantitative mass spectrometry

Phosphopeptide enrichment was performed as previously described (Zimman et al., 2010) with minor modifications. The desalted peptide mixture was fractionated online using EASY-spray columns (25cm × 75 μm ID, PepMap RSLC C18 2μm). The gradient was delivered by an easy-nLC 1000 ultra high-pressure liquid chromatography (UHPLC) system (Thermo Scientific). MS/MS spectra were collected on a Q-Exactive mass spectrometer

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Thermo Scientific) (Kelstrup et al., 2012; Michalski et al., 2011). Samples were run in technical duplicates, and raw MS files were analyzed using MaxQuant version 1.4.1.2 (Cox and Mann, 2008). MS/MS fragmentation spectra were searched using ANDROMEDA against the Uniprot human reference proteome database with canonical and isoform sequences (downloaded January 2012 from uniprot.org). N-terminal acetylation, oxidized methionine, and phosphorylated serine, threonine, or tyrosine were set as variable modifications, and carbamidomethyl cysteine (*C) was set as a fixed modification. The false discovery rate was set to 1% using a composite target-reversed decoy database search strategy. Group-specific parameters included max missed cleavages of 2 and label free quantitation (LFQ) with an LFQ minimum ratio count of 1. Global parameters included match between runs with a match time window and alignment time window of 5 and 20 minutes, respectively, and match unidentified features selected.

MS data analysis—Quantitative, label-free phosphopeptide data from MaxQuant were $\log_{10}$ transformed and missing data were imputed using random values generated from a normal distribution centered on the 1% quantile and the median standard deviation of all phosphopeptides (Deeb et al., 2012). After missing value imputation, phosphopeptides were quantile normalized. For clustering, phosphopeptide data was filtered using an FDR-corrected ANOVA p-value of 0.05. Hierarchical clustering was performed using the Cluster 3.0 program with the Pearson correlation and pairwise complete linkage analysis (Eisen et al., 1998). Java TreeView was used to visualize clustering results (Saldanha, 2004). Quantitative data for each phosphopeptide can be found in **Data S1b-S1d**.

TieDIE pathway analysis of clinical prostate cancer samples

We used the TieDIE algorithm (Paull et al., 2013) to connect 35 kinases and ‘kinase regulators’, 108 putative cancer driver genes with genomic perturbations in CRPC and 74 transcription factors, using the ‘Multinet’ (Khurana et al., 2013) pathway database consisting of a diverse set of literature-based gene-gene interactions (43,722 protein-protein interactions; 27,900 direct phosphorylation; 27,914 transcriptional/regulatory; 9,714 metabolic; genetic interactions excluded). Each of these three inputs were treated as a separate, equally weighted, input set for the algorithm, while the gene members of each input set were weighted by the total evidence for each protein: kinases by combined SAM d-statistic and MARINa inferred activity level, transcription factors by MARINa inferred activity level, and genomic events by the number of mutations and copy-number alterations observed in the 49 metastatic prostate cancer samples. The kinase, genomic event and TF gene sets were found to be significantly close in pathway space ($p<0.012$; **Fig. S2a**), according to a conservative background model run with 1,000 permutations of the input data.

The resulting network consisted of 338 nodes and 1,889 edges (597 direct phosphorylation; 1,184 protein-protein interaction; 102 transcriptional/regulatory; 6 metabolic). This network was filtered further by restricting to protein-protein edges with at least one pair of constituent phosphopeptides with at least modest correlation (Spearman rank correlation, Rho ≥ 0.3), resulting in a final ‘scaffold’ network of 122 nodes and 256 edges (190 protein-protein interaction; 131 phosphorylation).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cancer hallmark enrichment analysis

Cancer hallmark definitions were downloaded from the GSEA/MSigDB (http://www.broadinstitute.org/gsea/msigdb) database, and reduced to hallmarks highly linked to cancer (**Data S1j**). Enrichment analysis was performed by calculating the probability of overlap between the test set (defined by the set of genes in a network model) and the hallmark sets, using the hypergeometric distribution. Hallmark 'wheels' were colored proportionally to the negative log p-value returned by the hypergeometric test.

Patient-specific network generation and kinase target prediction

To generate sample-specific networks, we used the VIPER package (Alvarez et al., 2015) to infer sample-specific activity and applied thresholds derived from the MARINa analysis to each sample's data, generating binary calls for each of the 35 kinase regulators and 74 TFs, respectively. Scores for the 24 peptides with significant differential phosphorylation activity were z-normalized by gene and thresholded at a z-score of 1.0 or above, while VIPER pseudo z-scores were thresholded at the level corresponding to a 0.1 FDR cutoff in each corresponding Network Enrichment Score for MARINa analysis (supplemental methods). Functional "high-level" copy number gain and loss was assessed with the GISTIC algorithm (Mermel et al., 2011). For each sample, we searched all paths connecting any active kinase, mutation or high-level copy number gain or deletion to any active TF over edges contained in the scaffold network, using the NetworkX python package (Schult and Swart, 2008).

For all proteins in each patient-specific network, we performed three independent rankings based on the phosphorylation activity of functionally annotated peptides, VIPER inferred activity scores, and the network connectivity as measured by the shortest-path betweenness centrality for all genes. These three independent rankings were averaged for each protein providing a patient-specific network (PNET) score, from which a final combined ranking of all proteins for each patient was derived (**Fig. 7**;).

Statistical analysis

All statistical data were presented after either t-tests or one-way ANOVA as described in the figure legends. Correlation analysis was performed for each pair of proteins with an edge in the TieDIE network, by calculating the pairwise Spearman correlation between all corresponding peptides; only protein-protein edges with at least moderate positive or anti-correlation ($Rho \geq 0.3$; $Rho \leq 0.3$) between one pair of respective peptides were retained.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank members of the O.N.W. and J.M.S. laboratories for helpful comments and discussion on the manuscript and the University of Michigan for supplying material from their rapid autopsy program. J.M.D. is supported by the Department of Defense Prostate Cancer Research Program W81XWH-14-1-0148 and Prostate Cancer Foundation Young Investigator Award; N.A.G. is supported by UCLA Scholars in Oncologic Molecular Imaging (SOMI) program, NIH grant R25T CA098010; J.K.L. is supported by Specialty Training and Advanced Research (STAR) Program at UCLA, Prostate Cancer Foundation Young Investigator Award, and Tower Cancer Research Foundation Career Development Award; B.A.S. is supported by UCLA Tumor Immunology Training Grant #T32 CA009120;

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

T.S. is supported by Prostate Cancer Foundation Young Investigator Award and National Institute of Health/National Cancer Institute K99 Pathway to Independence Award 4R00CA184397; C.M.F. is supported by the UCLA Medical Scientist Training Program; J.H. is supported by NIH grants 5R01CA172603-02, 2P30CA016042-39, 1R01CA181242-01A1, 1R01CA195505, the Department of Defense Prostate Cancer Research Program W81XWH-12-1-0206, UCLA SPORE in prostate cancer, Prostate Cancer Foundation Honorable A. David Mazzone Special Challenge Award, and UCLA Jonsson Comprehensive Cancer Center Impact Grant; J.A.W. is supported by NIH GM089778; T.G.G. is supported by the NCI/NIH P01 CA168585, an American Cancer Society Research Scholar Award RSG-12-257-01-TBE, the CalTech-UCLA Joint Center for Translational Medicine, the UCLA Jonsson Cancer Center Foundation, the National Center for Advancing Translational Sciences UCLA CTSI Grant UL1TR000124, and a Concern Foundation CONquer CanCER Now Award; J.M.S. is supported by NCI/NIH U24-CA143858, NCI/NIH 1R01CA180778, NHGRI/NIH 5U54HG006097, and NIGMS/NIH 5R01GM109031 grants; O.N.W. is an Investigator of the Howard Hughes Medical Institute and is supported by the Zimmerman Family, the Concern Foundation and by a Prostate Cancer Foundation Challenge Award. J.H., T.G.G., J.M.S., and O.N.W. are supported by the West Coast Prostate Cancer Dream Team supported by Stand Up to Cancer/AACR/Prostate Cancer Foundation SU2C-AACR-DT0812 (O.N.W. co-PI). This research Grant is made possible by the generous support of the Movember Foundation. Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research.

REFERENCES

Alvarez, MJ.; Giorgi, F.; Califano, A. Using viper, a package for Virtual Inference of Protein-activity by Enriched Regulon analysis. 2015.

Aytes A, Alvaro A, Antonina M, Celine L, Alvarez MJ, Mireia C-M, Tian Z, Eastham JA, Anuradha G, Pienta KJ, et al. Cross-Species Regulatory Network Analysis Identifies a Synergistic Interaction between FOXM1 and CENPF that Drives Prostate Cancer Malignancy. Cancer Cell. 2014; 25:638–651. [PubMed: 24823640]

Beltran H, Rickman DS, Park K, Chae SS, Sboner A, MacDonald TY, Wang Y, Sheikh KL, Terry S, Tagawa ST, et al. Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. Cancer Discov. 2011; 1:487–495. [PubMed: 22389870]

de Bono JS, Logothetis CJ, Molina A, Fizazi K, North S, Chu L, Chi KN, Jones RJ, Goodman OB Jr, Saad F, et al. Abiraterone and increased survival in metastatic prostate cancer. N. Engl. J. Med. 2011; 364:1995–2005. [PubMed: 21612468]

Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012a; 487:330–337. [PubMed: 22810696]

Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012b; 490:61–70. [PubMed: 23000897]

Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013; 499:43–49. [PubMed: 23792563]

Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. Cell. 2014a; 159:676–690. [PubMed: 25417114]

Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. Nature. 2014b; 507:315–322. [PubMed: 24476821]

Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015; 163:1011–1025. [PubMed: 26544944]

Casado P, Rodriguez-Prados J-C, Cosulich SC, Guichard S, Vanhaesebroeck B, Joel S, Cutillas PR. Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. Sci. Signal. 2013; 6:rs6. [PubMed: 23532336]

Chang C-J, Yang J-Y, Xia W, Chen C-T, Xie X, Chao C-H, Woodward WA, Hsu J-M, Hortobagyi GN, Hung M-C. EZH2 promotes expansion of breast tumor initiating cells through activation of RAF1-β-catenin signaling. Cancer Cell. 2011; 19:86–100. [PubMed: 21215703]

Chen JC, Alvarez MJ, Talos F, Dhruv H, Rieckhof GE, Iyer A, Diefes KL, Aldape K, Berens M, Shen MM, et al. Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. Cell. 2014; 159:402–414. [PubMed: 25303533]

Cheng ASL, Lau SS, Chen Y, Kondo Y, Li MS, Feng H, Ching AK, Cheung KF, Wong HK, Tong JH, et al. EZH2-mediated concordant repression of Wnt antagonists promotes β-catenin-dependent hepatocarcinogenesis. Cancer Res. 2011; 71:4028–4039. [PubMed: 21512140]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 2008; 26:1367–1372. [PubMed: 19029910]

Deeb SJ, D'Souza RCJ, Cox J, Schmidt-Supprian M, Mann M. Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles. Mol. Cell. Proteomics. 2012; 11:77–89. [PubMed: 22442255]

Drake JM, Graham NA, Stoyanova T, Sedghi A, Goldstein AS, Cai H, Smith DA, Zhang H, Komisopoulou E, Huang J, et al. Oncogene-specific activation of tyrosine kinase networks during prostate cancer progression. Proc. Natl. Acad. Sci. U. S. A. 2012; 109:1643–1648. [PubMed: 22307624]

Drake JM, Graham NA, Lee JK, Stoyanova T, Faltermeier CM, Sud S, Titz B, Huang J, Pienta KJ, Graeber TG, et al. Metastatic castration-resistant prostate cancer reveals intrapatient similarity and interpatient heterogeneity of therapeutic kinase targets. Proc. Natl. Acad. Sci. U. S. A. 2013; 110:E4762–E4769. [PubMed: 24248375]

Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U. S. A. 1998; 95:14863–14868. [PubMed: 9843981]

Gioeli D, Mandell JW, Petroni GR, Frierson HF Jr, Weber MJ. Activation of mitogen-activated protein kinase associated with prostate cancer progression. Cancer Res. 1999; 59:279–284. [PubMed: 9927031]

Gonzalez ME, DuPrie ML, Krueger H, Merajver SD, Ventura AC, Toy KA, Kleer CG. Histone methyltransferase EZH2 induces Akt-dependent genomic instability and BRCA1 inhibition in breast cancer. Cancer Res. 2011; 71:2360–2370. [PubMed: 21406404]

Goodwin JF, Kothari V, Drake JM, Zhao S, Dylgjeri E, Dean JL, Schiewer MJ, McNair C, Jones JK, Aytes A, et al. DNA-PKcs-Mediated Transcriptional Regulation Drives Prostate Cancer Progression and Metastasis. Cancer Cell. 2015; 28:97–113. [PubMed: 26175416]

Grasso CS, Wu Y-M, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, Quist MJ, Jing X, Lonigro RJ, Brenner JC, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature. 2012; 487:239–243. [PubMed: 22722839]

Huang S-SC, Clarke DC, Gosline SJC, Labadorf A, Chouinard CR, Gordon W, Lauffenburger DA, Fraenkel E. Linking proteomic and transcriptional data through the interactome and epigenome reveals a map of oncogene-induced signaling. PLoS Comput. Biol. 2013; 9:e1002887. [PubMed: 23408876]

Hunter T. Tyrosine phosphorylation: thirty years and counting. Curr. Opin. Cell Biol. 2009; 21:140–146. [PubMed: 19269802]

Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013; 502:333–339. [PubMed: 24132290]

Kelstrup CD, Young C, Lavallee R, Nielsen ML, Olsen JV. Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. J. Proteome Res. 2012; 11:3487–3497. [PubMed: 22537090]

Khurana E, Fu Y, Chen J, Gerstein M. Interpretation of genomic variants using a unified biological network approach. PLoS Comput. Biol. 2013; 9:e1002886. [PubMed: 23505346]

Lee JK, Phillips JW, Smith BA, Park JW, Stoyanova T, McCaffrey EF, Baertsch R, Sokolov A, Meyerowitz JG, Mathis C, et al. N-Myc Drives Neuroendocrine Prostate Cancer Initiated from Human Prostate Epithelial Cells. Cancer Cell. 2016; 29:536–547. [PubMed: 27050099]

Li W, Ai N, Wang S, Bhattacharya N, Vrbanac V, Collins M, Signoretti S, Hu Y, Boyce FM, Gravdal K, et al. GRK3 is essential for metastatic cells and promotes prostate tumor progression. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:1521–1526. [PubMed: 24434559]

Lu S, Tsai SY, Tsai MJ. Regulation of androgen-dependent prostatic cancer cell growth: androgen regulation of CDK2, CDK4, and CKI p16 genes. Cancer Res. 1997; 57:4511–4516. [PubMed: 9377562]

Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2. 0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011; 12:R41. Others. [PubMed: 21527027]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Michalski A, Damoc E, Hauschild J-P, Lange O, Wieghaus A, Makarov A, Nagaraj N, Cox J, Mann M, Horning S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol. Cell. Proteomics. 2011; 10:M111.011015. [PubMed: 21642640]

Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, Vogelstein B, Kinzler KW. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science. 1997; 275:1787–1790. [PubMed: 9065402]

Newman RH, Hu J, Rho H-S, Xie Z, Woodard C, Neiswinger J, Cooper C, Shirley M, Clark HM, Hu S, et al. Construction of human activity-based phosphorylation networks. Mol. Syst. Biol. 2013; 9:655. [PubMed: 23549483]

Ning X, Shi Z, Liu X, Zhang A, Han L, Jiang K, Kang C, Zhang Q. DNMT1 and EZH2 mediated methylation silences the microRNA-200b/a/429 gene and promotes tumor progression. Cancer Lett. 2015; 359:198–205. [PubMed: 25595591]

Pandey M, Sahay S, Tiwari P, Upadhyay DS, Sultana S, Gupta KP. Involvement of EZH2, SUV39H1, G9a and associated molecules in pathogenesis of urethane induced mouse lung tumors: potential targets for cancer control. Toxicol. Appl. Pharmacol. 2014; 280:296–304. [PubMed: 25168426]

Park HU, Suy S, Danner M, Dailey V, Zhang Y, Li H, Hyduke DR, Collins BT, Gagnon G, Kallakury B, et al. AMP-activated protein kinase promotes human prostate cancer cell growth and survival. Mol. Cancer Ther. 2009; 8:733–741. [PubMed: 19372545]

Paull EO, Carlin DE, Niepel M, Sorger PK, Haussler D, Stuart JM. Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion Through Interacting Events (TieDIE). Bioinformatics. 2013; 29:2757–2764. [PubMed: 23986566]

Pawson T. Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. Cell. 2004; 116:191–203. [PubMed: 14744431]

Robinson D, Van Allen EM, Wu Y-M, Schultz N, Lonigro RJ, Mosquera J-M, Montgomery B, Taplin M-E, Pritchard CC, Attard G, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015; 161:1215–1228. [PubMed: 26000489]

Rubin MA, Putzi M, Mucci N, Smith DC, Wojno K, Korenchuk S, Pienta KJ. Rapid ("warm") autopsy study for procurement of metastatic prostate cancer. Clin. Cancer Res. 2000; 6:1038–1045. [PubMed: 10741732]

Saldanha AJ. Java Treeview--extensible visualization of microarray data. Bioinformatics. 2004; 20:3246–3248. [PubMed: 15180930]

Scher HI, Fizazi K, Saad F, Taplin M-E, Sternberg CN, Miller K, de Wit R, Mulders P, Chi KN, Shore ND, et al. Increased survival with enzalutamide in prostate cancer after chemotherapy. N. Engl. J. Med. 2012; 367:1187–1197. [PubMed: 22894553]

Schult, DA.; Swart, P. Exploring network structure, dynamics, and function using NetworkX.. Proceedings of the 7th Python in Science Conferences; SciPy. 2008; 2008. p. 11-16.

Smith BA, Sokolov A, Uzunangelov V, Baertsch R, Newton Y, Graim K, Mathis C, Cheng D, Stuart JM, Witte ON. A basal stem cell signature identifies aggressive prostate cancer phenotypes. Proc. Natl. Acad. Sci. U. S. A. 2015; 112:E6544–E6552. [PubMed: 26460041]

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 2005; 102:15545–15550. [PubMed: 16199517]

Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010; 18:11–22. [PubMed: 20579941]

Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, Haussler D, Stuart JM. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics. 2010; 26:i237–i245. [PubMed: 20529912]

Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, Dianes JA, Sun Z, Farrah T, Bandeira N, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat. Biotechnol. 2014; 32:223–226. [PubMed: 24727771]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Wang G, Ahmad KA, Unger G, Slaton JW, Ahmed K. CK2 signaling in androgen-dependent and -independent prostate cancer. J. Cell. Biochem. 2006; 99:382–391. [PubMed: 16598768]

Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013; 41:D955–D961. [PubMed: 23180760]

Yu G, Lee Y-C, Cheng C-J, Wu C-F, Song JH, Gallick GE, Yu-Lee L-Y, Kuang J, Lin S-H. RSK promotes prostate cancer progression in bone through ING3, CKAP2, and PTK6-mediated cell survival. Mol. Cancer Res. 2015; 13:348–357. [PubMed: 25189355]

Zimman A, Chen SS, Komisopoulou E, Titz B, Martínez-Pinna R, Kafi A, Berliner JA, Graeber TG. Activation of aortic endothelial cells by oxidized phospholipids: a phosphoproteomic analysis. J. Proteome Res. 2010; 9:2812–2824. [PubMed: 20307106]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

HIGHLIGHTS

- Provides a new phosphoproteomic encyclopedia of prostate cancer cells & tissues
- Integration of omics datasets reveal a map of activated signaling pathways in CRPC
- Development of personalized cancer hallmarks termed pCHIPS
- Patient-specific hierarchy of clinically actionable pathways for therapy

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

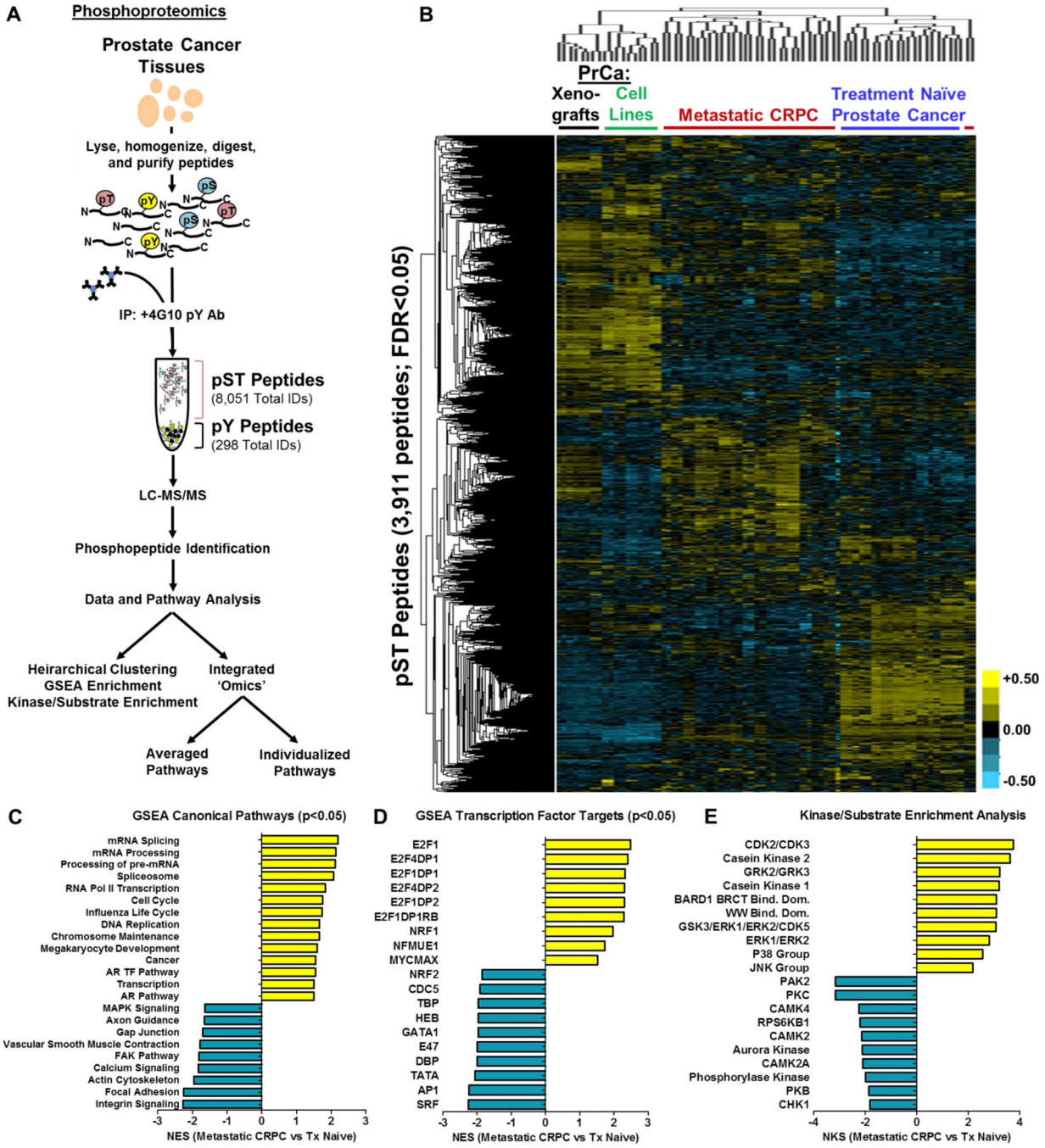


Figure 1. Characterization of the phosphoproteome in metastatic CRPC tissues

(A) General workflow of the phosphopeptide enrichment and quantitative mass spectrometry protocol followed by data and pathway analyses. Analyses described in the text. (B) Unsupervised hierarchical clustering heatmap of phosphoserine and phosphothreonine peptides identified from prostate cancer cell lines and tissues. 3,911 unique phosphopeptides (rows) were significantly identified from over 36 samples (columns). Unsupervised hierarchical clustering was performed using the Cluster program with the Pearson correlation and pairwise complete linkage analysis. (C-E) Gene set enrichment analysis

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(GSEA) was performed to identify canonical pathways (C) and transcription factor targets (D) with activity either higher (right yellow bars) or lower (left blue bars) in metastatic CRPC compared to primary tissue. (E) Kinase/substrate enrichment analysis (KSEA) identified several unique kinases that were not directly sequenced by the mass spectrometer in the phosphoproteomic data; NES = normalized enrichment score, Yellow = hyperphosphorylation, Blue = hypophosphorylation in the heatmap (B). See also Data S1a-S1f.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

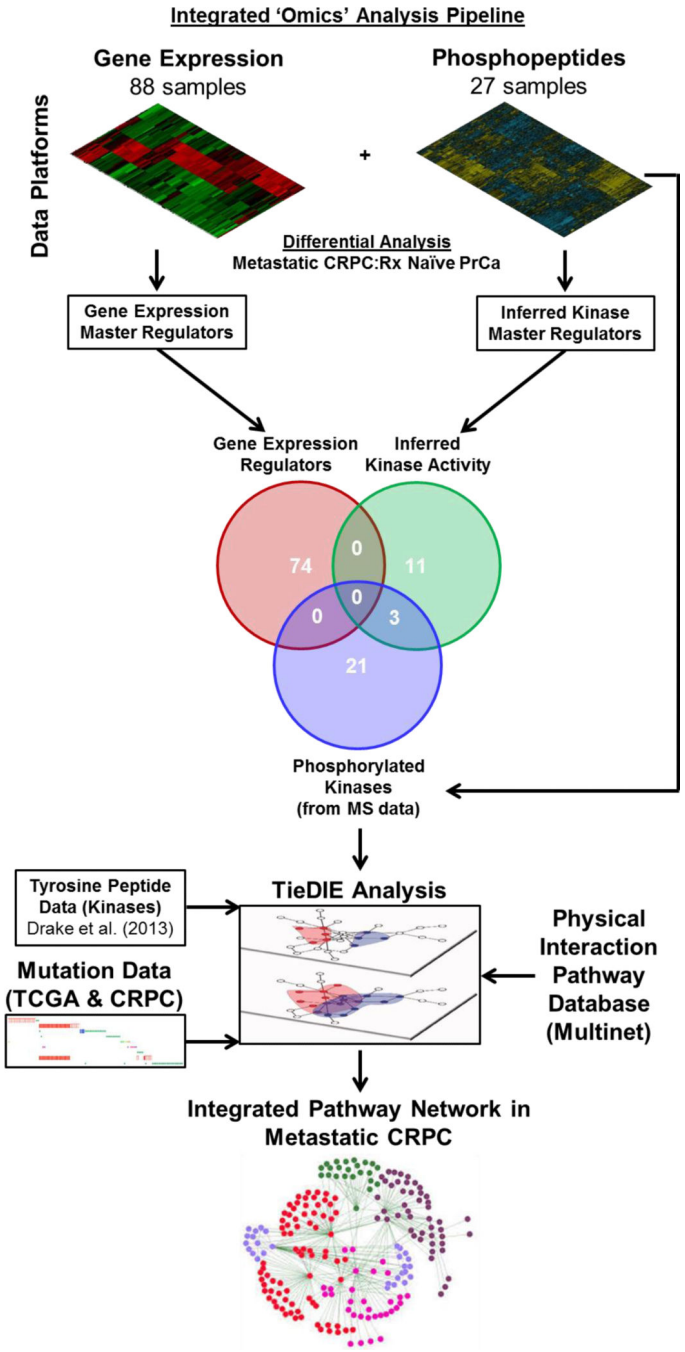


Figure 2. Pipeline for omic dataset integration

Flow diagram depicting the integration pipeline. 27 gene expression and 16 phosphoproteomic CRPC patient datasets were integrated with mutational data and combined using TieDIE to generate the resulting integrated network. The overlay of input gene expression, kinase master regulators, and phosphorylated kinases are shown as a Venn diagram. See also Figures S1-S2, Data S1g-S1j.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

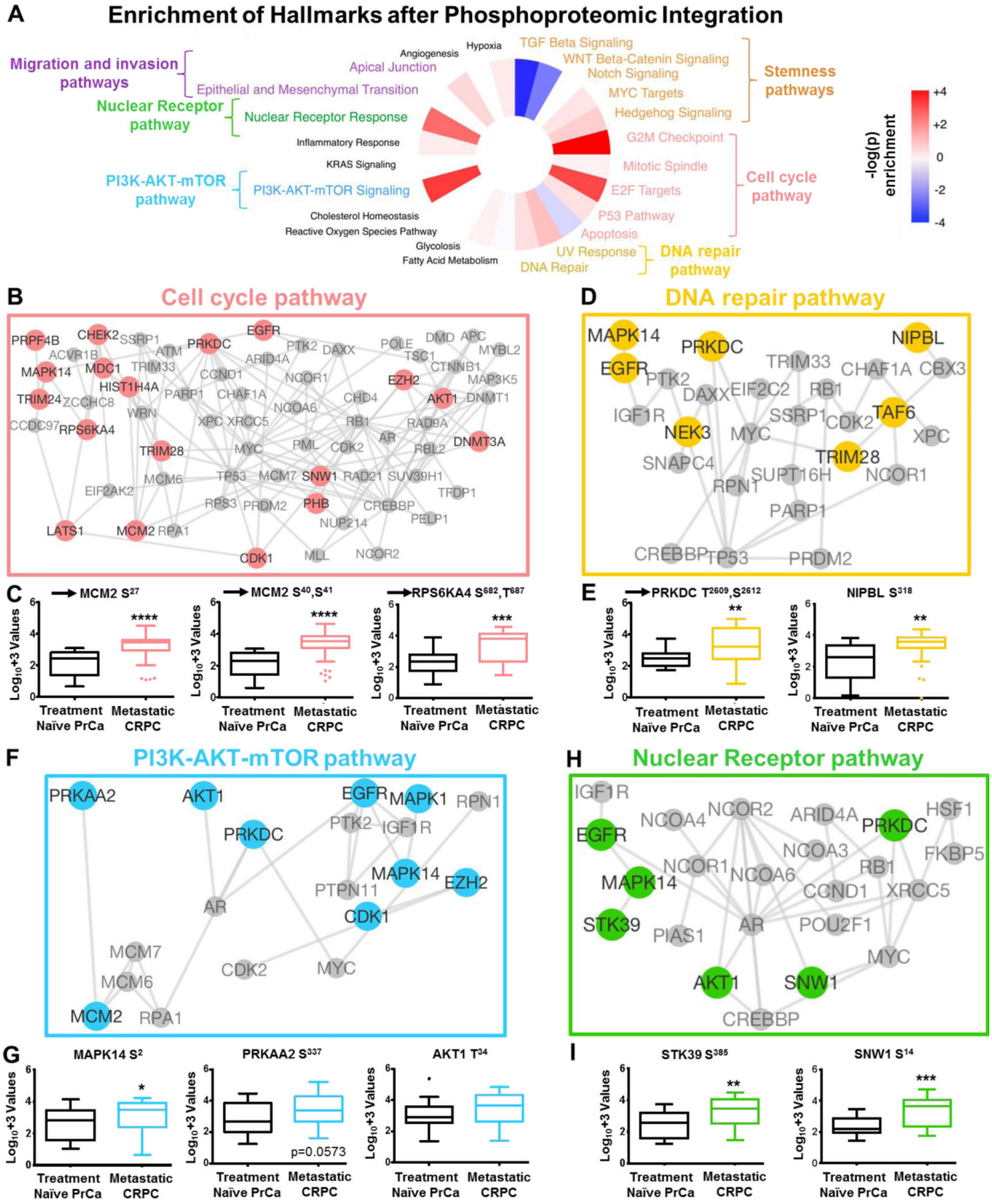


Figure 3. Pathway analysis of metastatic CRPC

Enriched cancer hallmarks generated by dataset integration using TieDIE after inclusion of the phosphoproteomic and gene expression data relative to gene expression data alone (A). Several cancer hallmarks were enriched after inclusion of the phosphoproteomic data including the cell cycle pathway (B, red nodes), DNA repair pathway (D, yellow nodes), AKT/mTOR/MAPK pathway (F, blue nodes), and the nuclear receptor pathway (H, green nodes). Detailed analysis of each of these pathways revealed several common and unique players with high connectivity. Assessment of a select number of kinases and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

phosphoproteins from each network confirmed their elevated phosphorylation state (C, E, G, I) including some with direct phosphorylation on their enzymatic active residue (C, E). This supports the activation state of the networks observed. Black arrow represents phosphoresidues that result in enzymatic activity of the given protein. These defined subnetworks only contain genes that fall within both the curated hallmark gene sets and the previously generated scaffold network, with colored nodes corresponding to genes that are members of a hallmark and exclusive to the integrated network solution containing the phosphoproteomic data; grey genes are other scaffold members in the surrounding region. T-test was performed to calculate significance. *=$p<0.05$, **=$p<0.01$, ***=$p<0.001$, ****=$p<0.0001$. See also Figure S3.

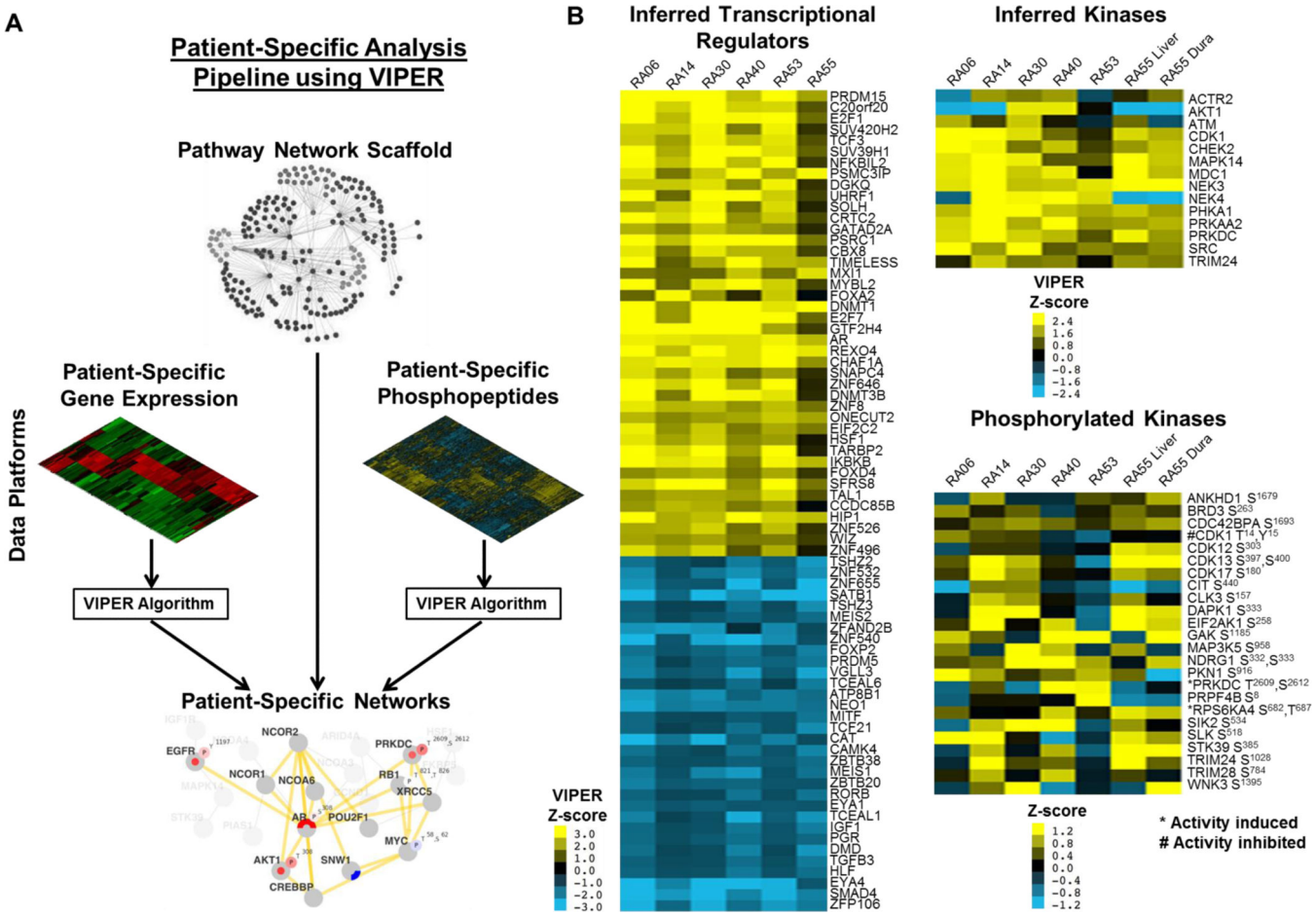


Figure 4. Development of a patient-specific network using VIPER

(A) Flow diagram depicting the integration of gene expression and phosphoproteomic datasets for VIPER analysis. (B) Heatmap of the gene expression and kinase master regulators and phosphorylated kinases for all 6 patients. This data was used as the input for patient-specific network analysis. See also Figure S4.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

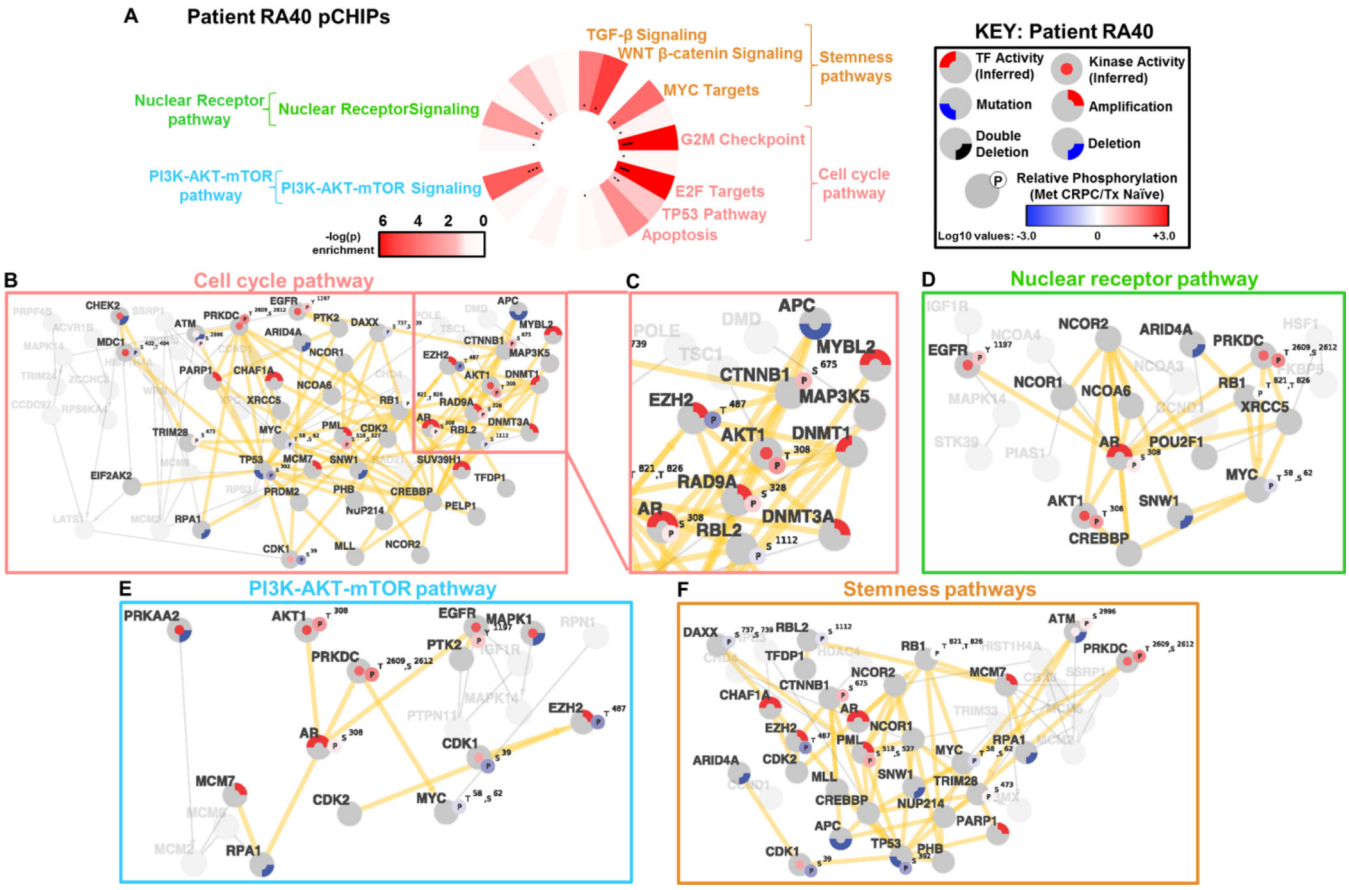


Figure 5. Integrated pathway network of patient RA40

(A) Phosphorylation-based cancer hallmarks using integrated personalized signatures (pCHIPS) analysis for patient RA40 revealed strong enrichment of cell cycle and PI3K-AKT-mTOR pathway networks. The pCHIPS wheel summarizes enrichment between genes in each patient-specific network and the corresponding pCHIPS category: labels indicate categories with significant enrichment after multi-hypothesis correction (FDR < 0.1). Black dots indicate SNV and copy-number genomic events in this patient. Patient-specific network nodes and edges related to cell cycle pathway (B, C), nuclear receptor pathway (D), PI3K-AKT-mTOR pathway (E), and stemness pathways (F). Edges belonging to both the patient-specific network model and the cell cycle related scaffold network are shown as thick yellow edges, while corresponding genes are shaded in dark grey. Yellow arrows indicate that the upstream kinase directly phosphorylates the downstream substrate. “Circleplot” quadrants for each gene summarize genomic, transcriptomic and phosphoproteomic activity relevant to metastatic CRPC phenotype (upper right: amplification; lower right: deletion; lower left: mutation; upper left: transcriptional regulatory activity; center: kinase regulatory activity). Node “ears” peripherally attached to circleplots represent relative phosphorylation of specific, functionally annotated peptides sites on each protein. Genes and edges that are not represented in the patient-specific network but are in the scaffold network are shown in light grey. See also Figure S5, Data S1k, and Data S2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

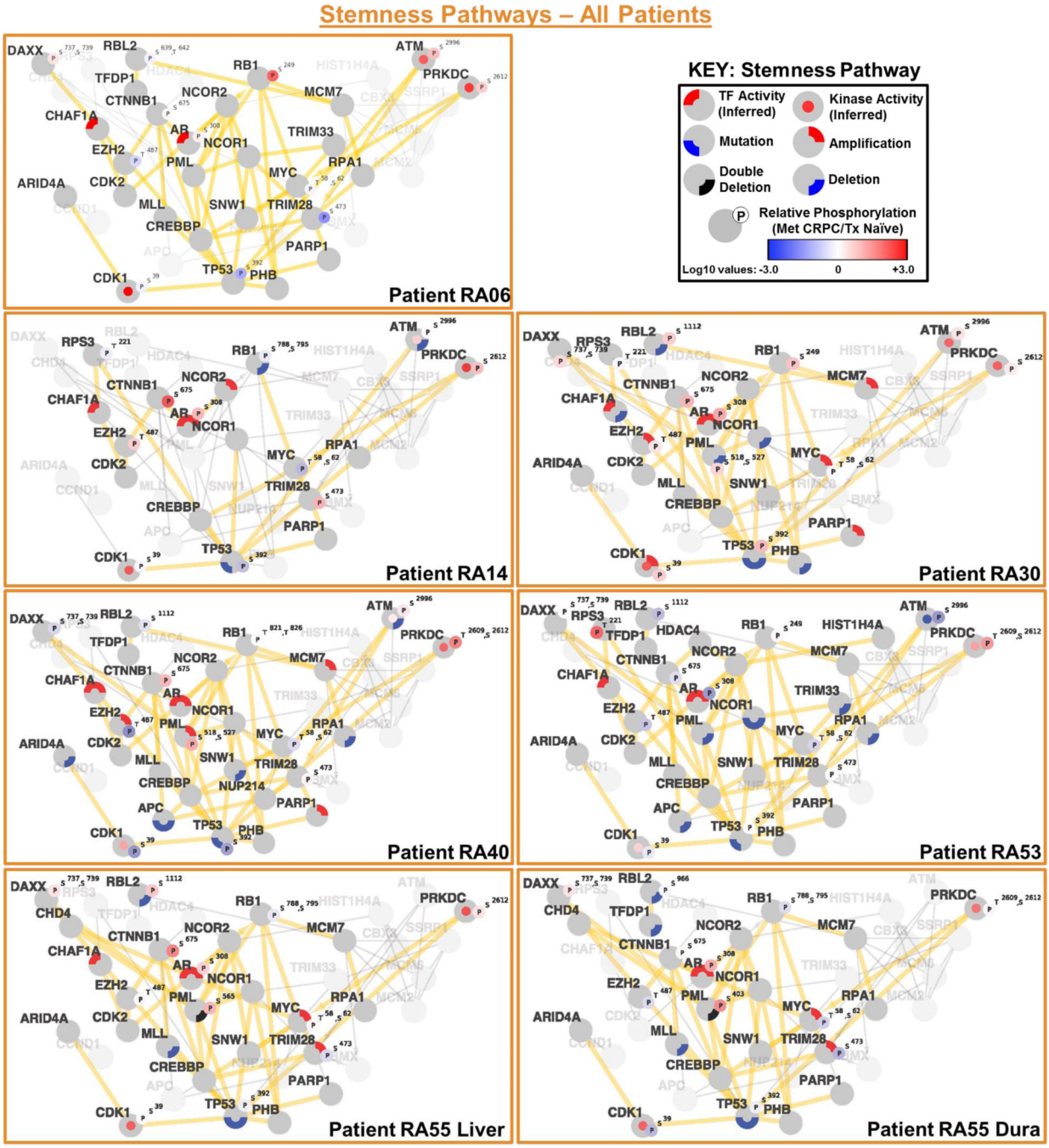


Figure 6. Comparison of the stemness pathway hallmarks across all 7 patient samples

Patient-specific networks were developed from the stemness pathway hallmarks and revealed distinct regions of the network were differentially activated across the CRPC patient samples. This suggests that while the stemness pathway hallmarks were enriched in all the patients evaluated, a patient-specific evaluation is needed to determine the precise targets for therapy. Genes and edges that are not represented in the patient-specific network but are in the scaffold network are shown in light grey.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

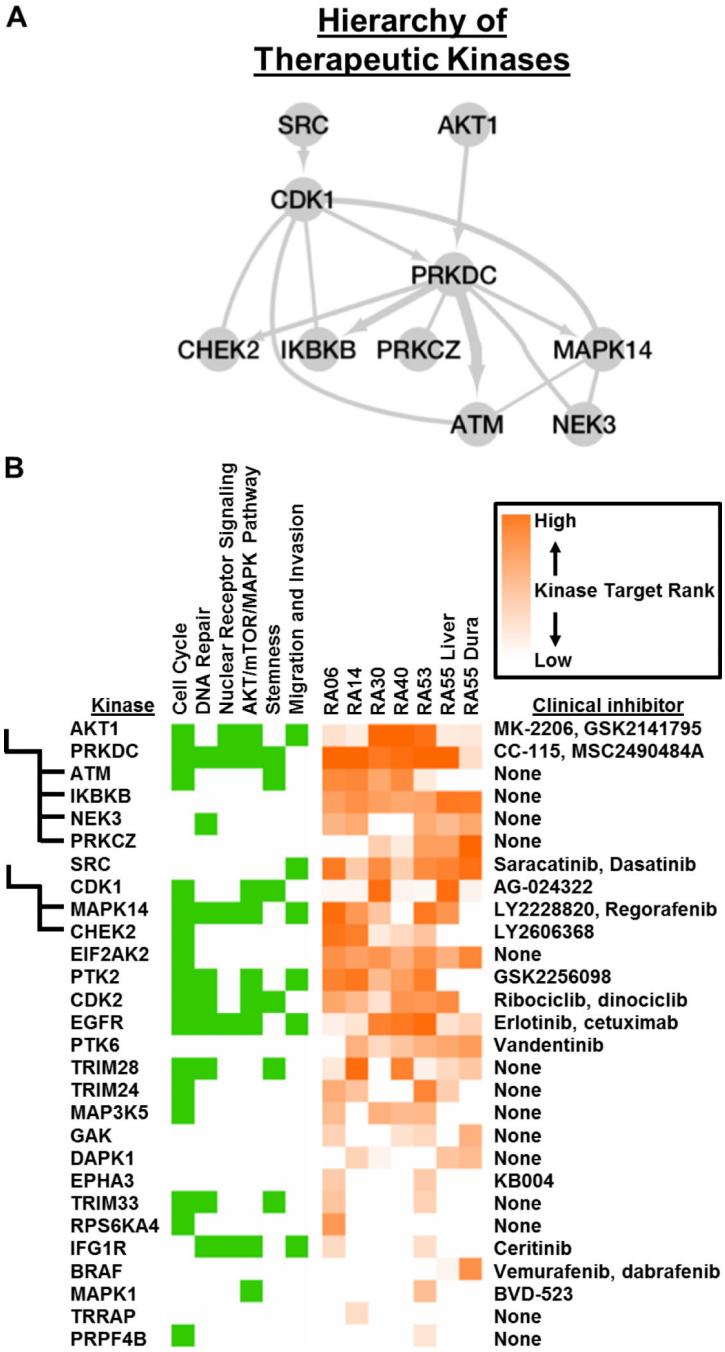


Figure 7. Summary of kinase target potential in patient-specific networks

(A) Network diagram of hierarchy between kinase targets derived from KSEA interactions and potential "coverage" of phosphopeptides activated in CRPC. The thickness of each edge represents the degree of overlap in the set of protein targets that each kinase is predicted to phosphorylate. Directed arrows indicate predicted phosphorylation from a (source) kinase, at a residue on the corresponding target kinase. (B) Therapeutic potential and summary of kinase targets. Far left: the hierarchy of therapeutic kinase targets shown in part (A) is briefly summarized. Left: Green boxes indicate kinases (rows) that are members of each of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the six major hallmark subnetworks (columns) shown in Figure 3. Right: Orange boxes indicate the predicted importance of kinase targets based on the combined evidence from VIPER-inferred kinase activity, phosphorylation status of functionally annotated peptides, and connectivity, for each patient specific network (columns). Currently available clinical inhibitors for each are listed on the right. See also Figures S6-S7, Data S1l-S1m.