Summary (30 lines max)

Kinases are among the most important drug targets and clinically significant kinase inhibitors have been developed for multiple diseases. A subset of kinases, the understudied dark kinases (DKs), have received little or no attention because foundational data on their biochemical and biological functions is not available. This proposal will collect such data by perturbing DKs genetically and with small molecules and then measuring the cellular consequences using multiplex proteomic, gene expression, metabolomic and imaging assays. A subset of DKs with potential links to human disease will be intensively studied as a means to qualify new therapeutic drug targets. Data collected in this project will be aggregated with existing information from previous NIH-funded large-scale structural and genomic projects to create a *Dark Kinase Knowledgebase* (DKK) that provides gene-by-gene and network-level information on the dark kinome and its interaction with other signal transduction and regulatory networks. Close coordination with the NIH LINCS project will ensure data interoperability and make efficient use of informatics tools. The DKK will be developed in collaboration with the IDG Knowledge Management Center (KMC), adhere to standards for *Findable, Accessible, Interoperable and Reusable* (FAIR) data, and be accessible to human users and machines (via an API). Commercially available DK reagents be validated and extended with new genetic and chemical tools provided to the *Resource Dissemination Center* (RDOC). The overall approach will be iterative, with simpler methods applied first (e.g. simple gene knockout) and more sophisticated methods subsequently (e.g. stable CRIPSRa/i) pursued by an interdisciplinary team of chemists, computational biologists, mass spectroscopists and pharmacologists working on five linked aims. **Aim 1** will develop a computational algorithm for prioritizing DKs, develop and maintain the DKK, and perform network-level analysis on the kinome using supervised and unsupervised machine learning. **Aim 2** will measure kinase abundance in normal and perturbed cells using parallel reaction monitoring with stable isotope dilution (PRM-SID) and RNASeq and data analyzed using network inference tools to provide insight into dark and light kinome in diverse cell types. **Aim 3** will perturb DKs with genetic tools such as CRIPSR/Cas9-mediated gene knockout, CRIPSRa/i to induce more subtle-up and down regulation and inducible gene inaction. The impact on cell fate, morphology and signal transduction will then be determined using PRM-SID, phosphoproteomics, RNASeq, gene reporter assays, metabolomics profiling and highly multiplex single-cell imaging. **Aim 4** will extend DK analysis to small molecule inhibitors by carefully profiling existing drugs against DKs and by designing and synthesizing new chemical ligands. **Aim 5** will involve collaboration with other investigators to assay the expression and function of DKs in primary human cells and tissues relevant to the NIH Precision Medicine Initiative. All aims will be pursued in parallel for a progressively expanding resource of data and tools for continued study of DKs.

Health Relevance

Advancing understanding of understudied kinases, a highly druggable class of proteins, will increase knowledge about signal transduction and control over cellular physiology and is likely to reveal a subset of proteins that should be advanced as targets for new therapeutic drugs.

SPECIFIC AIMS

The overall goals of this proposal are to (i) construct and maintain a machine and human-intelligible Dark Kinase Knowledgebase (DKK) that aggregates available information on DKs with newly collected molecular and functional data on these proteins collected by our *IDG Data and Resource Generation Center*; the DKK will provide hitherto unavailable information on the biochemical and biological functions of DKs (and will be built in collaboration with the *Knowledge Management Center*) (ii) generate, validate, organize and distribute tools and reagents for study of DKs by others (in collaboration with the *Resource Dissemination Center)* (iii) use inference and machine learning tools to generate network-level models of crosstalk among well-studied and dark kinases and ascertain the functions of the dark kinome in regulating different aspects of cellular physiology (iv) identify and intensively studying the subset of DKs likely to play a causal role in the development of human disease; we hope to quality these DKs as targets for novel therapeutic drugs.

Aim 1: Generation, systematization and dissemination of knowledge about dark kinases will aggregate and release all data on DKs to FAIR and emerging NIH LINCS and IDG metadata standards.
Aim 1.1 Prioritizing and systematizing analysis of dark kinases will use an algorithm that weighs (i) evidence of biological function (ii) availability of reagents (iii) input from domain experts and scientific advisors.
Aim 1.2 Creating a dark kinase knowledge base will aggregate and release new and available data on DKs.
Aim 1.3 Network-level understanding of the dark kinome will be generated using diverse computational tools involving supervised and unsupervised machine learning, logic-based models and other methods.
Aim 1.4 Identifying possible therapeutic targets will analyze public data on DKs and results from Aim 5.
Aim 1.5 Reagent validation on hundreds of commercially available DK reagents will accelerate discovery.

Aim 2. Quantitative analysis of DKs using Parallel Reaction Monitoring (PRM) and RNAseq will use precise and innovative methods to quantify kinase abundance.
Aim 2.1 Development of a PRM-SID mass spec assay for the 134 dark kinases will provide a means to precisely measure the absolute abundance of DKs and their interactors.
Aim 2.2 PRM-SID analysis of cell lines, primary cells, tumors and tissues will create a large-scale resource on the distributions of DKs across diverse biological settings.
Aim 2.3 RNAseq will be performed in parallel and will complement PRM and other mass spec assays.

Aim 3: Annotating the dark kinome for cellular phenotypes and functions in signal transduction will determine the consequences of DK perturbation for cellular physiology using profiling and focused assays.
Aim 3.1: Development and testing CRISPR reagents for studying DKs will provide the tools to deplete and mutate kinases individually and in combination using conventional and evolved CRIPSR technologies.
Aim 3.2 Phenotypic analysis of DK perturbation by imaging following genetic or small molecule perturbation will provide a means to ascertain cellular phenotypes at single-cell resolution using novel multiplex methods.
Aim 3.3 Use of an engineered “reporter-world” assay for transcription factors downstream of diverse signal transduction pathways will quantify the functional impact of DK perturbation on cell signaling.
Aim 3.4 Determining the effects of DK perturbation on kinome remodeling will study homeostasis and adaptive remodeling within the kinome by looking for instances of coordinated up and down-regulation of ~330 kinases following selective perturbation of a single kinase gene.
Aim 3.5: Profiling proteomics and phosphoproteomics analysis will extend PRM-SID assays in Aim 2.2 to deep proteomic (~13,000 protein) and phosphoproteomic (~50,000 site) profiling of normal and perturbed cells.
Aim 3.6 Protein-protein interaction studies by mass spectrometry will provide insight into physical interactions among DKs and other cell regulators before and after DK perturbation; these data and results from Aim 3.4 will provide information critical for the construction of new network-level models of the human kinome.
Aim 3.7 Metabolic profiling of cells in which DK activity is perturbed will provide new insight into the role of DKs in regulating fundamental mechanisms of energy production and cellular biosynthesis.

Aim 4: Identifying and characterizing cell active chemical tools for dark kinases will create highly sought research tools for interrogating the biological functions of DKs and for initiating development of therapeutic drugs through careful profiling of existing compounds and synthesis of new molecules.
Aim 4.1 NanoBRET assays for profiling kinase inhibitor activities will enable in-cell measurement of selectivity.
Aim 4.2 Generation and characterization of new chemical tools will allow us to identify or synthesize ligands for up to 70 DKs during the six year IDG program.
Aim 4.3 Screening existing kinase inhibitor collections in cells will allow us to search for hits with (i) high potency for a DK target (ii) selectivity relative to other kinases, all scored using nanoBRET assays.
Aim 4.4 Screening PKIS libraries and synthesizing new compounds will extend Aim 4.3 to additional DKs.

Aim 5: Collaborations to determine the expression and function of DKs in primary human cells and tissues relevant to COPD, cystic fibrosis, arthritis, ALS, diabetes etc. will provide key information for identifying DKs whose dysregulation is associated with human disease and that warrant study as therapeutic drug targets.

RESEARCH STRATEGY

SIGNIFICANCE: The ~500 member human kinome is highly druggable using both competitive small molecule and allosteric inhibitors but the functions of about one-third of the proteins in this family are unknown. We will create and maintain a novel Dark Kinase Knowledgebase (DKK) covering these proteins, generate, organize and distribute reagents for studying dark kinases (DKs; **Table 1**), and identify the biological processes and networks in which DKs function. Study of DKs may also uncover therapeutic drug targets.

Enhancing understanding of essential cellular processes. Preliminary data collected in our labs and available in databases (e.g. GEO) suggest that DKs are widely expressed and change in abundance, localization and phosphorylation status over the course of development, trans-differentiation and disease progression. It is therefore likely that DKs regulate aspects of cellular physiology of wide-spread interest. For example, six members of the human NimA related (NEK) kinases are currently dark, even though their orthologues play critical roles in cell cycle and cilia biology (5). As a class, the DYRK kinases are also understudied, even though select members are being pursued as targets for anti-cancer drugs (6). It is also possible that sets of interacting DKs (based on STRING data) regulate poorly understood aspects of cellular biology that warrant further investigation.

Table 1: Classification of the 134 dark kinases by substrate. Listed alphabetically within each category

Category	Number	Name
Protein kinase	1-98	ADCK2, ADCK4, ADCK5, ALPK2, ALPK3, BCKDK, CAMK1D, CAMK1G, CAMKK1, CDC42BPB, CDC42BPG, CDK10, CDK11B, CDK14, CDK15, CDK17, CDK18, CDKL1, CDKL2, CDKL3, CDKL4, CLK3, CLK4, CSNK1G1, CSNK1G2, CSNK1G3, CSNK2A2, DYRK1B, DYRK2, DYRK3, DYRK4, EEF2K, ERN2, FASTK, HIPK4, LMTK3, LRRK1, LTK, MAP3K10, MAP3K14, MAPK15, MAPK4, MARK4, MAST2, MAST3, MAST4, MKNK2, NEK10, NEK11, NEK4, NEK5, NEK6, NEK7, NIM1K, NRK, PAK3, PAK6, PAK7, PDIK1L, PHKG1, PHKG2, PKMYT1, PKN3, PNCK, PRKACB, PRKACG, PRKAG1, PRKCQ, PRPF4B, PXK, RIOK1, RIOK3, SBK2, SBK3, SGK494, SRPK3, STK17A, STK3, STK32A, STK32B, STK32C, STK33, STK36, STK38L, TESK1, TESK2, TLK1, TLK2, TP53RK, TSSK1B, TSSK3, TSSK4, TSSK6, TTBK1, TTBK2, VRK2, WEE2, WNK2
Pseudo-kinase	99-116	CAMKV, NRBP2, PLK5, POMK, PRKY, PSKH1, PSKH2, RPS6KC1, RPS6KL1, SCYL1, SCYL2, SCYL3, SGK223, STK31, STK40, STKLD1, TBCK, ULK4
Other kinase	117-131	DGKH, GK2, ITPK1, ITPKA, PANK3, PI4KA, PIK3C2B, PIK3C2G, PIP4K2C, PIP5K1A, PIP5K1B, TK2, TPK1, UCK1, UCK2
Regulatory protein	132-134	PHKA1, PRKAB1, PRKRA

Protein kinases: putative serine, threonine, and tyrosine kinases. ***Pseudokinases:*** same fold as the protein kinases, but lack one or more canonical catalytic site residues. ***Other kinases:*** includes inositol, glycerol, nucleotide, and thiamine kinases. ***Regulatory proteins***: phosphotransferase activity not established; modulate other kinases

Identifying new therapeutic targets for human disease. The bulk of the work in this proposal involves systematic DK characterization but the most significant outcome would be identification of one or more DKs with a role in human diseases. We find that some DKs are among the most heavily mutated kinases in The Cancer Genome Atlas (TCGA) For example, ADCK5, MAPK15, NRBP2 are frequently overexpressed in ovarian cancer whereas DGKH and NEK5 are frequently deleted in prostrate adenocarcinoma. PHKG2 is implicated in glycogen storage disease IXc. Given the current state of kinase inhibitor chemistry, it is likely that these DKs can be effectively drugged. The data needed to improve confidence in existing correlations between diseases and DKs will be addressed in this proposal, evaluated by crowd-sourcing and discussed on an ongoing basis with disease experts (Aim 5; see LOS). Our preliminary data also show that DKs are already targets for FDA approved poly-selective kinase inhibitors (Aim 4). The anti-cancer drugs sorafenib (Nexavar®) and sunitinib (Sutent®), for example, inhibit some DKs nearly as efficiently as they do their nominal targets. We will attempt to determine whether these drug activities are biologically significant.

Leveraging little used data and resources. A number of reagents (knock-out cell lines, antibodies, expression vectors etc.) and biological data (crystal structures, protein-protein interactions etc,) on DKs are publically available. The current proposal will organize, validate, and redistribute this data and then fuse it with new IDG data so as to leverage investments that have already been made but yielded insufficient new biology.

INNOVATION: An underlying hypothesis in this proposal is that studying the expression, modification status and localization of DKs in different cellular systems before and after perturbation (activation or inhibition by genetic means or with small molecules) will allow us to elucidate, at least partially, the biological functions of DKs and prioritize the most interesting for in-depth analysis (7). Open-ended functional discovery is not straightforward and we will therefore apply a multi-faceted approach involving several molecular profiling methods (omics), genetic and chemical perturbation and computational analysis. Timely data release will enable crowd-sourced data analysis, increase the likelihood that interesting phenotypes can be discovered and promote follow-on work by others (as envisioned in RFA-RM-16-026).

Integrated multi-omic analysis to FAIR standards. The current proposal is innovative with respect to both computational and experimental approaches. We have consciously chosen scalable proteomic, transcript

profiling, metabolomics and imaging methods that are implemented in our labs (per RFA) and innovation in the current proposal lies in their integration. To achieve this we have allocated resources for sustained and simultaneous progress in five critical computational areas: (i) pipelining data from instruments to permanent data storage without human intervention or data corruption (ii) automated quality control to ensure data precision and accuracy (iii) efficient machine-assisted metadata association and curation (iv) release of data that is *Findable, Accessible, Interoperable and Reusable* (the FAIR standard (7) (v) analysis of IDG data using network-aware inference and simulation methods to promote integrated understanding of the human kinome.
Machine-assisted experimental design. Prioritization of DKs and experimental design will rely on an innovative computational algorithm that continuously collates and evaluates existing and newly acquired data on biological function, reagent availability and disease association. This will guide an iterative profiling approach in which the most feasible experiments are performed first and more sophisticated (and costly) studies reserved for higher-priority targets. We have already built a functioning version of this algorithm (https://github.com/sorgerlab/DarkKinome) that will accept input from the External Scientific Advisory Board and KMC, as envisioned in the RFA. Our approach to prioritizing kinases can be extended to other target classes and represents a natural way to identify areas for intra-IDG collaboration, as envisioned by RFA-RM-16-026 and allocated to the 10% budget set-aside.

APPROACH

Our approach to the dark kinome will start with simpler studies (e.g. CRISPR knockout or RNAi) followed by more complex approaches (e.g. CRISPRa/i or regulated protein degradation) on kinases with a higher priority score. For up to 70 kinases (by Yr 6), we will induce specific dose-dependent inhibition using chemical ligands (Aim 4). Analysis of perturbed cells will involve both systematic profiling in which all 134 proteins (and their interactors) are analyzed in parallel (e.g. by mass-spec Parallel Reaction Monitoring; PRM; Aim 2) as well as more focused kinase-by kinase, approaches. As described below, prioritization will be based on evidence of biological function as well as the availability of experimental tools; the algorithm used for prioritization will be exposed for community-based comment and modification.

Preliminary analysis of DKs by our group reveals extensive information in on-line data repositories (e.g, PhosphoSitePlus, PDB, Bioplex, Hippie, HPRD, I2D, Reactome, The Cancer Genome Atlas-TCGA etc.). In addition, a remarkable number of reagents have been generated by commercial vendors, most of which are untested in publications (Aim 1.5 will validate these), Data include crystal structures for 35 kinases studied by the SGC, recombinant proteins or antibodies for 116 dark kinases and knockout HAP1 cell lines (8) for 101 DKs (**Fig. 1**). We have merged this information with data from the pilot phase of IDG and the NIH LINCS project (http://lincs.hms.harvard.edu) in which Sorger leads a data generation center to create a preliminary information resource (https://github.com/sorgerlab/DarkKinome); reviewers will also find a bioRxiv.org manuscript on DKs at this location. Work on Aims 1.1 to 1.5 will be initiated immediately using this data and continue throughout the performance period as new data and algorithms become available.

Expected Outcomes: We will establish and progressively improve a DKK in close collaboration with the IDG *Knowledge Management Center* (KMC; RFA-RM-16-024; see letter of support from the existing Center), most likely hosted on Pharos (https://pharos.nih.gov). In LINCS we have found that a sustained commitment to data curation and management is required at both data generation centers and the KMC. Primary data will be released through the DKK (and KMC) using established tools or emerging methods and standards being developed by IDG and LINCS. All software and algorithms will be available in the open source, facilitating involvement of the Cutting Edge Informatics Tools (CEITs) planned for the future of IDG; the DKK itself will meet standards for FAIR data (9). The DKK will be initialized with available information and improved progressively using existing and innovative new algorithms. Findings will be incorporated into protein-by-protein datasheets (Aim1.2) as well as gene and protein networks and signatures of kinome perturbation (Aim1.3). Cell-based data will be linked to information on human disease to identify kinases that might warrant qualification as therapeutic targets (Aim1.4). The DKK will also provide a consolidated reference site for information on reagents and assays generated in the current program and for commercial reagents we validate (Aim 1.5). This information will be coordinated with the Resource Dissemination and Outreach Center (RDOC; RFA-RM-16-025). At a minimum, information in the DKK will include items listed in the RFA:

- A catalogue of all available reagents and publically available data.
- A catalogue of cell and tissue-specific protein expression patterns.
- Development and verification of multiplex assays for dark kinases using proteomics, RNA expression and multiplex microscopy in human tissues, hES and iPSC-derived cells, xenografts and single cells.
- Generation and analysis of interactome data covering the dark kinome and its potential regulators or substrates in multiple cell types.
- Rich biochemical and phenotypic data on DKs derived from genetic and small molecule perturbation studies followed by multiplex phenotypic and molecular analysis.

- In-depth analysis of those DKs potentially involved in human disease; this will include analysis of animal models of disease where feasible.

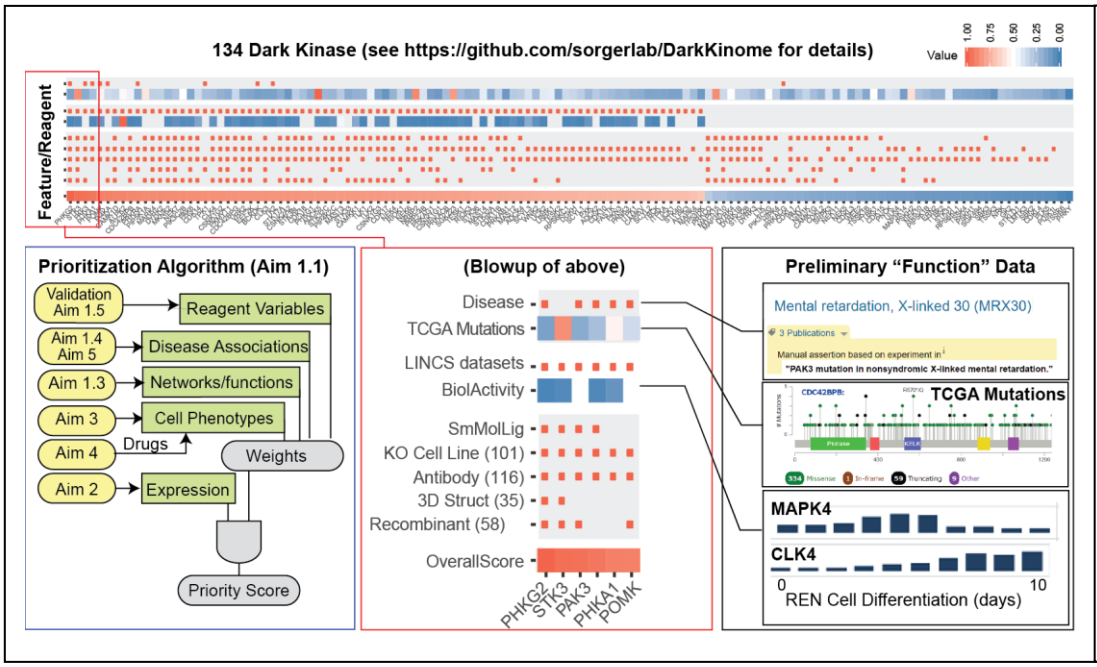


Fig. 1: Prioritizing dark kinases Data and resources from LINCS, IDG Phase I, literature mining and on-line databases consolidated in a preliminary knowledgebase and available at GitHub. This data feeds a prioritization algorithm to guide study of DKs in Aims 2-5. SmMolLig in this chart denotes off-target activities of known drugs; these will be greatly improved and extended in Aim 5. Expression data is by TMT mass spec.

Aim 1: Generation, systematization and dissemination of knowledge about dark kinases, biological networks in which they function and connections to cellular phenotypes and human disease.

***Goal*:** Four data analysis subaims will (i) create and maintain a DKK knowledgebase covering 134 members of the dark kinome (ii) distill primary data into functional biochemical and biological information and promote further research using genetic and chemical tools, which we will also develop and make available (iii) place dark kinases within new and emerging network-level understanding of cell signaling and physiology (iv) identify those dark kinases whose mutation or mis-regulation is associated with human disease, making them possible therapeutic targets. A fifth sub-aim will focus on validation of a large set of existing reagents.

Aim 1.1 Prioritizing and systematizing analysis of dark kinases. Research into DKs will be prioritized using an algorithm (**Fig.1**) that weighs (i) evidence of biological function (ii) availability of internal and external reagents (Aim 1.5) (iii) input from domain experts. The current algorithm considers 2 continuous variables, such as mutational frequency in the TCGA and 7 binary variables such as availability of antibodies (see github.com/sorgerlab/DarkKinome). Evidence of DK expression is particularly important: it is pointless to knock out or inactivate a kinase in a cell line in which it is not expressed. We have preliminary mass spectrometry data for ~90 DKs in breast cancer cell lines, transdifferentiated iPSC cardiomyocytes and neuronal cells. PRM profiling of additional primary and established lines (Aim 2) will improve and expand on this data.

Evidence of biological function is critical to our prioritization scheme and available data provide indirect hints of activity. Several DKs are as heavily mutated in TCGA data as established oncogenes such as ATM. Other DKs are amplified: TLK2 in breast cancer and CDC42BPB in neuroendocrine prostate cancer (TCGA). Unpublished LINCS data show that DKs such as PRPF4B or NRBP2 change in phosphorylation status or abundance during the course of ReNcell Neural Progenitor differentiation (10). Multiple DKs exhibit cancer-specific alternative splicing (75), including 37 DKs that are alternatively spliced only in a narrow range of cancer types. Changes in splicing are likely to alter the connectivity of signaling networks, with potential effects on phenotype. In other cases redundancy is likely: *CDKL1-4* kinases are so homologous that a quadruple knockout would likely be required to probe function; this deprioritizes or complicates genetic approaches. Among reagents, the most valuable are selective small molecule inhibitors; we currently have ~40 such molecules that show varying degrees of selectively and Aim 4 will greatly extend this set. By Yr2 functional data linking DKs to phenotypes will become available based on work in Aims 2 and 3.

Aim 1.2 Creating an information resource on dark kinases. By Yr1, in collaboration with the KMC, we will make the initial DKK available; sufficient useful data is already in hand. The per-kinase view will likely resemble a hybrid of Uniprot and Pharos (or ChEMBL) with links to primary evidence. The DKK will be available for human browsing and programmatic access via an API. Experimental data is useful on a large scale only if it is stored in a standardized manner and tagged with appropriate metadata (11). Curation is relatively straightforward in the case of gene expression data, thanks to years of effort, but data derived from multi-dimensional microscopy or drug dose-response studies have few if any established data standards. The development of tools for pipelining, curating and releasing proteomic, gene expression, imaging and functional

data is a major ongoing effort in LINCS that will be repurposed for IDG. We expect LINCS tools to be useful for all IDG projects not just DKs, and their re-use would bring significant benefits to the program in terms of cost and interoperability.

***Aim 1.3 Network-level understanding of the dark kinome*.** Protein-by protein data summaries (Aim1.2) are familiar and useful in many cases, but future progress requires more sophisticated approaches rooted in understanding cellular networks. We will therefore use inference, regression and other machine learning algorithms (both existing and newly developed) to create network-level representations of DKs and the cellular pathways in which they function. We will link the activities of these networks to biologically consequential cellular phenotypes. DKs will be perturbed in loss of function genetic studies (Aims 3.1 and 3.3) in conjunction with small molecule inhibitors (Aim 4) in diverse cell types. Data on the consequences of DK perturbation will be collected using proteomic, transcriptomic and high dimensional imaging methods (Aim 3) as well as cell growth and transcriptional reporter assays (Aim 3). Data on adaptive reprogramming of DKs (Aim 3.4), which arises from disruption of normal homeostatic mechanisms, and from protein-protein interaction studies (Aims 3.6) will be particularly valuable for inference of network structure.

The Sorger and Gomez labs have published extensively on the use of partial least squares, multi-linear regression and other statistical and machine-learning methods to create and assemble network models from molecular and phenotypic data (12-15), probabilistic and multi-scale statistical mechanics (16-18) and protein structure (18, 19). We have also worked extensively on executable models using Boolean logic (20, 21) and ordinary differential equations (22-26). In addition, Major has developed improved methods for processing 'omic data (e.g., CompPASS, SAINT, HGScore) that extract more accurate information from spectra (66). As demonstrated in a recent DREAM8 challenge, the most accurate network inference typically arises through integration of data-derived methods with a network-level prior (27). We will therefore evaluate all inferred linkages in the context of known interaction databases, such as PathwayCommons or KEGG (28).

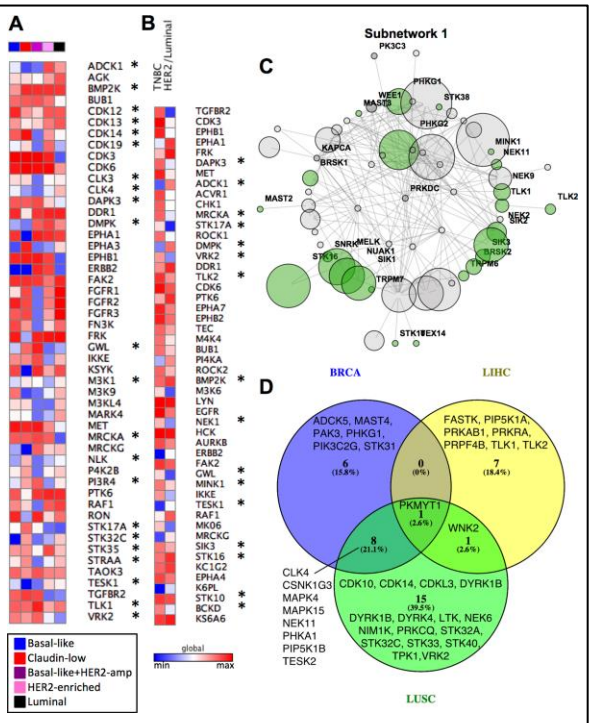


Figure 2. A) Baseline MIB binding signature of the 50 most subtype-predictive kinases across breast cancer cell lines. A "*" denotes understudied kinases. B) Kinase signature of top 50 kinases distinguishing TNBC from HER2/luminal subtypes. C) Kinome subnetwork showing both well-studied (grey) and understudied kinases (green), with node size proportional to degree of on/off-target inhibition by LINCS inhibitors. D) 38 dark kinases showing largely cancer-specific alternative splicing patterns across breast (BRCA), liver hepatocellular carcinoma (LIHC) and lung squamous cell carcinoma (LUSC) within TCGA data.

IDG perturbation data will be combined with public information (e.g GEO, Peptide Atlas) to generate molecular signatures (**Fig 2A, B**) and "archetype maps" that place an individual DK in the broader context of protein-protein interaction and functional networks (e.g., coordinated behaviors in response to perturbation). Directed, undirected and signed relationships will be captured (depending on data type) and linked to phenotypic responses. These studies will yield information on DKs, DK cascades and networks of both DKs and characterized signaling molecules (**Fig. 2C**).

When perturbed, the kinome exhibits extensive, time-dependent remodeling that varies with the type of perturbation, the biological context and disease state. Remodeling is thought to arise from normal homeostatic mechanisms and studying provides data on these mechanisms as well as interactions among kinases (Aim 3.4). We will specifically look for cases in which perturbation of a DK results in remodeling of well-studied kinases, suggesting the existence of previously unknown interactions.

New computational methods will be required for fusing imaging, genomic and proteomic data into better understanding of DK biology, thereby increasing innovation. We will explore probabilistic graphical models to infer linkages between dark kinase activities (as estimated via MIB/MS), interactions with downstream targets (from phosphoproteomics) and activation of transcription factor activity (Aim 3). When preliminary studies suggest that a kinase warrants further study, time series data will be collected and analyzed using executable models such as logical ODEs and Hidden Markov Models.

***Aim 1.4 Identifying possible therapeutic targets among dark kinases*.** Among the most significant outcomes of this research would be identifying one or more new therapeutic targets. There is indirect evidence in TCGA and the literature that this is not an unreasonable goal. We will mine both general-purpose data sets such as GEO, the NHGRI-EBI GWAS Catalog and disease-specific resources such as the AMP-ADKnowledge

Portal (which contains data for multiple NIA-supported programs on Alzheimer's Disease (29); see LOS) for data indicative of DK dysregulation. We will also use our collaborative research groups (see LOS) as a source of domain-specific expertise, leading to in-depth, disease-focused analysis of high priority DKs. Those with the potential to be therapeutic targets will be further prioritized for development of chemical tools and for specialized biological analysis (Aims 3-4). In support of these goals, we will perturb DKs in cells and tissues relevant to diseases such as cancer, COPD, cystic fibrosis, arthritis, ALS, diabetes and dementia (Alzheimer's), all of which are relevant to the Precision Medicine Initiative (Aim 5).

***Aim 1.5 Reagent validation*.** Many commercial reagents are available for DKs, but few have been validated in peer-review papers. We will undertake validation studies in conjunction with the proteomic and functional perturbation experiments in Aims 2-5. Currently, 101 knockout-cell lines for DKs are available in the pseudo-haploid human leukemia cell line, KBM-7, also known as HAP1 cells (8). We will test these lines by PCR and PRM mass spec to confirm loss of gene and protein expression. Antibodies are available for 116 dark kinases and we have partnered with ThermoFisher to validate these reagents and make new ones (ThermoFisher LOS). Antibodies will be tested for their ability to recognize a band of the right molecular weight in a Western blot of a cell line confirmed by PRM to express the kinase and rechecked in a HAP1 knockout. A similar strategy will be used to test antibodies for use in immunofluorescence. All data will be made available on the DKK and provided to reagent vendors.

Potential pitfalls and alternative approaches. An initial prioritization algorithm (Aim1.1) is already available; the primary risk in this subaim is that the algorithm will not prove more useful than human intuition. DKK development (1.2) will be demanding and time-consuming, but is relatively low risk particularly when pursued in collaboration with the KMC. Network inference (Aim1.3) will involve a range of methods, reducing risk, but will nonetheless require some from methodological innovation. The prospects of success in associating a DK with a human disease (1.4) are simply unknown.

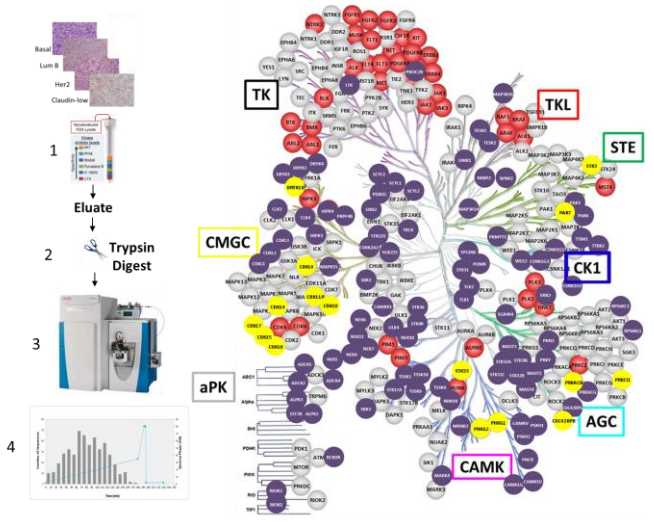


Fig. 3: Configuring a high multiplex PRM assay for kinases targeted drugs and inhibitors. Four breast cancer PDX's were selected that represent the major subtypes of breast cancer (basal, PDX-2; luminal, PDX-20; Her-2+, PDX-8 and claudin low, PDX-12) (77). Tumor lysates were prepared in non-denaturing buffer (1) from sonicated, cryopulverized tissue (60). Multiple lysates (n=3) were prepared from each subtype and passed individually over the multi-inhibitor bead column (Step 1). The bound kinases were eluted and peptides prepared for LC-MS (Step 2). The MS/MS spectra from 24 LC-MS analysis in DDA mode were searched against a RefSeq database (2) Step 3). The spectral library was built in SKYLINE and peptides were selected according to CPTAC criteria and retention time to minimize elution time concurrency (Step 4). The red, grey, and yellow balls represent kinases inhibited by approved drugs in the KIN-200 assay panel. DKs are shown in purple. *We will develop a single panel PRM assay for DKs as described in (4).*

Aim 2. **Quantitative analysis of DKs using Parallel Reaction Monitoring (PRM) and RNAseq**

***Goal*:** The goal is to measure the absolute abundance of DKs relative to better studied kinases before and after perturbation by genetic and chemical means. Parallel RNAseq will provide complementary data.

Approach: Kinome-wide quantification of kinase expression has previously been performed by mass spectrometry without (30) or with chemical (1) or metabolic (31) labeling. In conjunction with Multiplexed Inhibitor Bead) (MIB) enrichment (1) enables profiling of several hundred kinases in the presence and absence of drugs. However, such approaches yield fold-change data rather than absolute protein quantity. Absolute levels are substantially more valuable for cross-tissue comparison and data fusion and can be correlated more reliably with quantitative gene expression data (e.g. RNAseq fkpm units). A high-precision surrogate for protein quantity is available in highly-multiplexed, absolute protein quantification using mass spectrometry of specific peptides (32, 33). The method is based on multiple reaction monitoring with stable isotope dilution (MRM-SID) in which a standard curve is generated using different amounts of synthetic peptides having natural isotopic abundances (L, light) and a constant amount of stable isotope-labeled peptide (H, heavy). This yields the relationship between peak area ratios (PAR's) from LC-MS measurements (L/H) and the quantity of the unlabeled peptide. The same amount of heavy peptide is then spiked into experimental samples, analyzed by LC-MS. Co-eluting internal standard heavy peptides also serves to minimize spectral match false positives. Recent improvements in MRM-SID have increased throughput (cycle time of ~ 3h), degree of multiplexing (~500 H/L pairs), sensitivity (sub-fmolar) and precision (CV < 20%). The use of PRM-MS (34) circumvents

tedious selection of “best” transitions, significantly reducing development time for a multiplex assay. Townsend has already developed a PRM panel for 212 (well-studied) kinases (35) (**Fig. 3**). This KIN-200 assay was configured with 440 H/L synthetic, high purity peptide pairs selected according to Clinical Proteomics Tumor Analysis Consortium (CPTAC) guidelines (36) for uniqueness in a human background proteome; the panel will now be extended to include the DKs (Aim 2.1).

In a PRM-SID assay **(Fig. 4)** kinases affinity enriched from patient-derived xenografts (PDXs) using multi-inhibitor beads (MIBs) (1) were subjected to LC-MS in data-dependent acquisition (DDA) mode on a Q-Exactive mass spectrometer (both Townsend and Johnson have a Q-E MS) Proteotypic peptides (5317) from 336 kinases including 44 DKs were identified. Peptides in the KIN-200 assay are found under “Authentication of Key Reagents” and SOPs are available in the Washington University folder at (https://assays.cancer.gov/. KIN-200 profiling of tumor lysates enables quantification of kinase abundance between ~10 and 10,000 fmol/mg of protein (**Fig 4**) with data for the CDC42BPB DK shown in **Panels A-C**. Boxes denote 11 DKs: PRKACB, STK3, PHKG2, CDK15, CDK14, CDK18, CDK10, PHKG1, PRKCQ, STK33, DYRK1B. In these studies the absolute level of CDC42BPB was found to be ~ 5- 10-fold greater in the Her2^{AMP} positive PDX than the other breast cancer PDXs (**Panel C)**. **Fig. 4, Panel D** shows that the DK CDK18 was only detected in this Her2 positive PDX (LOD for this DK = ~ 150 attomole/mg tissue protein).

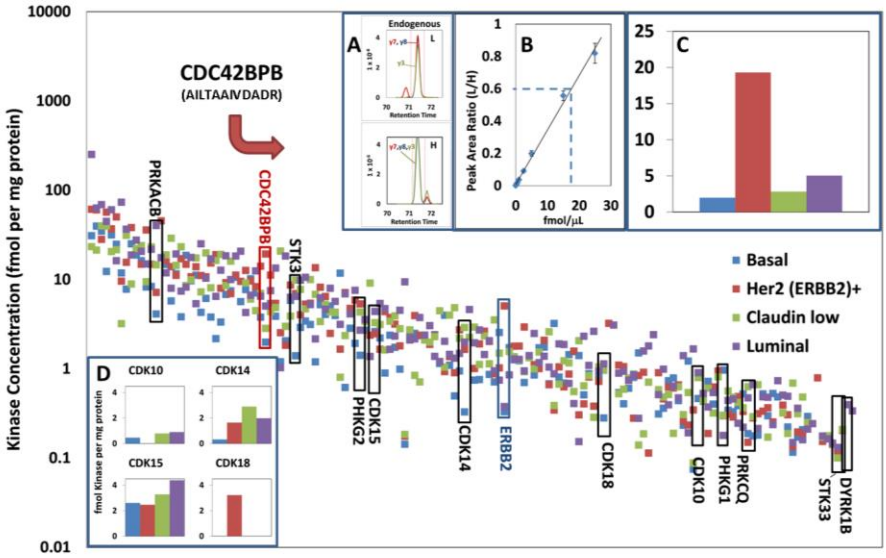


Figure 4: Fig 3: Absolute quantification of kinases from breast cancer PDX’s using PRM-SID. **Panel A**, Ion chromatograms for endogenous and synthetic heavy proteotypic peptide of CDC42BPB. **Panel B**, Response curve for peptide, AILTAAIVDADR. **Panel C**, Absolute levels of dark kinase DCD42BPB in breast cancer PDX’s. **Panel D**, Absolute quantity of dark kinase CDK’s.

***Aim 2.1 Development of a PRM-SID assay for the 134 dark kinases*.** We will develop a PRM panel for as many of the 134 DKs as technically possible (**Table 1**). Using MIB enrichment of endogenous kinases in cell lines and PDXs, we currently detect expression of ~90 DKs. Improving on this poses a chicken and egg problem: to develop a PRM assay we need to detect suitable peptides in the mass spec by shotgun methods. For kinases we cannot currently detect, we will express DKs using pooled lentiviruses (~30 per pool) in HEK293 cells (or other cell types if necessary), perform LC/MS and select peptides. Our lentiviral library currently contains genes for 59 DKs and will be expanded to the full set of 134 In Yr1. Kinases will be enriched using MIBs or Flag-tags (present in the lentivirus vector) followed by tryptic digestion and LC-MS analysis in DDA mode. The highest intensity proteotypic peptides with physical and chemical characteristics already established for KIN-200 will be used to build a high-sensitivity, multiplex PRM assay.

Aim 2.2 *PRM-SID assay of cell lines, primary cells, tumors and tissues*. With completion of the DK PRM Assay we will screen 36-48 human cell lines representing different tissues of origin as well as selected tissues available by autopsy. We will assay primary human lung alveolar epithelium, liver, fibroblast, hES cells, hES cells as well as iPS cell-derived neurons and cardiomyocytes, all of which are currently grown in our labs or those of our collaborators (see LOS) (37, 38). These data will provide absolute quantitation of DK levels and, in addition to addressing a requirement of the RFA, they are a necessary prelude to CRISPR/Cas9 and pharmacologic perturbation.

Aim 2.3 RNASeq to determine patterns of DK gene expression. RNASeq will be performed in parallel with PRM and other mass spec assays to establish patterns of DK expression in normally growing and perturbed cell lines. Extensive evidence confirms that protein and RNA data are highly complementary. Conventional deep RNASeq is similar in economics and a counterpart to deep proteomics (Aim 3.5). As a lower cost, high-throughput counterpart to PRM-SID we will used DGE-UMI RNASeq profiling. In 3’ Digital Gene Expression (3’-DGE), cDNA is generated by oligo dT priming, unique molecular identifiers (UMIs) are incorporated and strand-specificity is preserved. With the help of the Broad Institute that developed the method, the Sorger lab has implemented this technique on his robots at a reagent cost of ~$5 per sample. Extensive analysis has demonstrated that DGE-UMI preserves a large fraction of the information content of deep sequencing. When

there evidence of bifurcation in cell state or fate, single-cell RNA sequencing will be used. Sorger has applied this method previously (76) and it is available on a fee-for-service basis from a HMS core.

Potential pitfalls and alternative approaches. We do not anticipate major challenges in developing PRM assays for DKs since we have extensive experience with the method and suitable peptides for 44 DKs have already been developed. Should we nonetheless encounter difficulties, we will use the recently described TOMAHAQ TMT method (39) developed in the Gygi lab and recently implemented in the Sorger group.

Aim 3: Annotating the dark kinome for cellular phenotypes and functions in signal transduction.

Goal: The goal of sub aims 3.1 to 3.6 is to determine the consequences of DK perturbation for cellular physiology using profiling and focused assays. Initial studies will be performed on a panel of six immortalized cell lines we know can be transfected or infected efficiently and reliably grown in culture. When chemical tools are available (Aim 4) we will perform assays on a wider range of cell types. As necessary, selected DKs will be studied in lines in which they appear to have a biological function (e.g ADCK5 in ovarian cancer cells). These data will be analyzed using many of the same tools as PRM studies in Aim 2. Kinome reprogramming (Aim 3.4) and PPI studies (Aim 3.6) are expected to be particularly valuable for network inference.

Aim 3.1: Development and testing CRISPR reagents for studying DKs. To study the biological functions of DKs we will deplete and mutate kinases individually and in some cases in combination using a range of CRIPSR technologies. The complex physiology of kinases (and the language in RFA-RM-16-026) suggests that fairly subtle modulation of kinase activity is required. Nonetheless, it makes sense to proceed from simple to complex approaches, making maximal use of available reagents. As shown in the flow chart (**Fig 5**) we will first study validated (Aim 1.5) HAP1 knockout cells, 101 of which are potentially available; for other kinases we will use Dharmacon synthetic guide RNAs (four per target gene) to generate knockouts (40, 41). If we see a phenotype (see below) we will then proceed to less dramatic up and down-regulation using CRISPRi/a (42). Selective activation or inactivation will be achieved using doxycycline-induced Cas9 expression (43). In the case of DKs with particularly interesting functions, we will introduce site-specific mutations in kinase and other domains and express the modified DKs in cells in which the endogenous gene is knocked out. We expect to use lentivirus vectors for these experiments. We note that CRISPR technology is developing extremely rapidly and that it is likely that the precise approach we use will vary over the course of the grant; fortunately both HMS and UNC have active and expert core facilities and Johnson, Sorger and Major actively use CRIPSR to study kinase biology. Nonetheless, it is likely to remain true that regulated and selective modulation of specific kinases will remain relatively time-consuming and very difficult for at least some genes (for largely unknown reasons). Thus, we will reserve the creation of stable lines involving subtle up and down regulation, and gene mutation, for those DKs that appear to have interesting cellular or disease phenotypes.

As an alternative approach, we will use classical RNAi techniques for transient and stable gene depletion (with all of the attendant caveats). Small molecules will also be used to induce expression or degradation of DKs, for example, by fusing the DK to a mutant form of DHFR that requires trimethoprim for stability (44). Inducible gene deletion is possible using hydroxytamoxifen and a CRISPR-Cas9 fusion to the estrogen receptor (ERT2) (45) and using light (46). DKs can also be fused to a degron using CRIPSR /Cas and then degraded in a controlled manner (47).

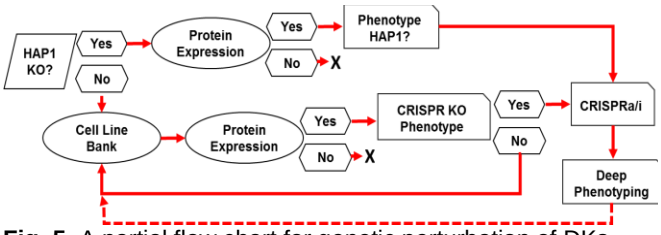


Fig. 5. A partial flow chart for genetic perturbation of DKs involving the use of selectively more powerful CRISPR methods for gene inactivation and modulation (see text)

Since the IDG RFA envisions high-throughput protein analysis it might seem reasonable to choose one method and then implement it *en masse*. This appears to work for simple deletion but it is our overwhelming experience that the utility of more nuanced approaches varies with the protein under study. We therefore believe that it will be necessary to use a gene modulation tool kit. We expect to have simple deletions for all DKs by Yr2 and to subject ~10-20 DKs to more nuanced and intensive genetic analysis each year thereafter. In parallel, we can also look for synthetic lethal phenotypes by performing CRIPSR-based screening experiments using libraries and cell lines in which one or more DKs has been previously knocked out.

Aim 3.2 Phenotypic analysis of DK perturbation by imaging. Following genetic or small molecule perturbation, cells will be subjected to rapid, and inexpensive analysis using high content imaging using the "Cell Painting" method (48). This involves staining cells with multiple dyes to score shape and morphology in multiple channels (degree of spreading, vesicle content, nuclear morphology shape of the cytoskeleton, etc.) Automated image analysis routines developed at UNC and HMS will be used to detect differences in cell state over time and, in the case of small molecules, across a dose range. Cells will be plated on solid substrates varying in composition and stiffness to elicit different phenotypes. Live-cell assays using white-light cameras or

dye-labelled cells and fluorescence (depending on cell type) will be used to score motility and adhesion. Cell viability assays will also be performed in the presence and absence of drugs to look for kinase interaction (49).

To enable deep phenotyping at a single cell level we will perform cyclic immunofluorescence (CycIF) in which iterative four to six-color imaging creates 20-30 channel, high-resolution images of cells grown in 96 and 384 well plates (50). To date, nearly 200 antibodies have been validated for proteins involved in the cell cycle, cytoskeleton, receptor-mediated signal transduction, transcription factors etc. including those in the MEK-ERK,

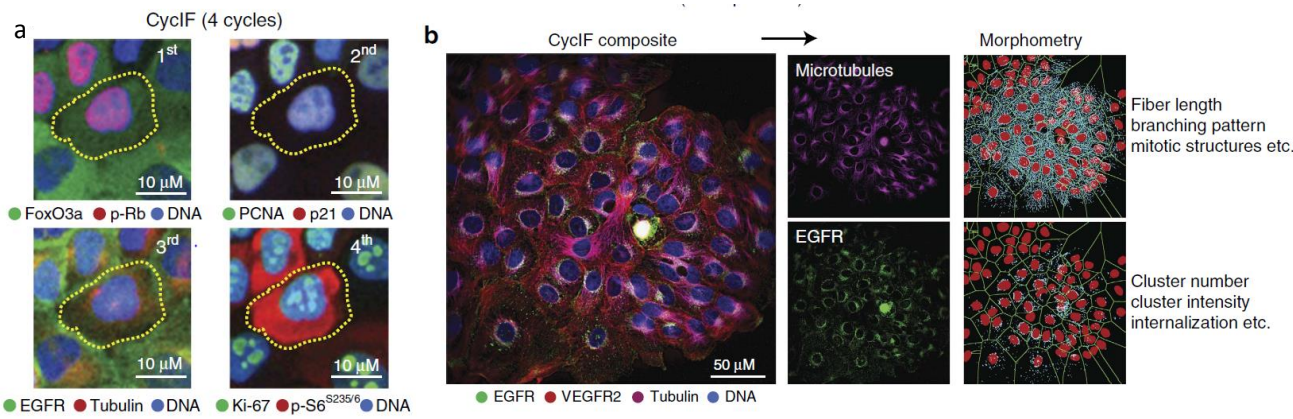


Fig.6:**A**. Highly multiplexed (20-30 channel) imaging of cells before and after small molecule or genetic perturbation using the method of cyclic immunofluorescence developed in the Sorger Lab (https://youtu.be/H6tMEsdwt1Y). In this example, MCF10A cells were fixed and subjected to four-cycle CycIF using fluorophore-conjugated antibodies p-Rb$^{S807/S811}$ (cycle 1), p21 and PCNA (cycle 2), EGFR and b-tubulin (cycle 3) and Ki-67 and p-S6$^{S235/6}$ (cycle 4). **B**. Morphometric analysis of CycIF-generated images of single cells. Three rounds of CycIF staining were were binarized and passed through different filters (sharpen, skeletonized and maximum, and so on) for extracting texture features (length, branches, enrichment and clusters). CycIF, combined with algorithms developed for flow cytometry and CyTOF, can exploit natural cell-to-cell fluctuations in protein levels and activities to uncover aspects of drug and ligand response that are obscured by population average measurement. ~200 antibodies have been validated for CyCIF of cell cycle regulators, signal transduction kineses etc. (3).

PI3K-AKT, WNT, NFκB, JAK/STAT, TGFβ and other pathways (http://lincs.hms.harvard.edu/db/antibodies/?page=1)(**Fig. 6**). CycIF will be powerful tool for DK analysis because it represents a means to look broadly for single cell phenotypes in a wide range of cell types and conditions at low cost. Patterns of fluctuation and co-variation among proteins at a single-cell level represent a powerful means for constructing protein-interaction maps. CycIF experiments will be integrated and guided by our results from genetic, protein (PRM), other assays. CRISPRi/a experiments will make extensive use of imaging. Data dissemination will be made possible by tools the Sorger lab is currently developing (in collaboration with the BD2K consortium) to manage and analyze large sets of image data on the Cloud (currently Amazon EC2).

Aim 3.3 Use of an engineered "reporter-world" assay to monitor signal transduction. These studies will use engineered reporters for specific signal transduction pathways to quantify the functional impact of DK perturbation on signal transduction. The assays will involve engineered lentiviral transcriptional reporters (**Table 2**) each containing 12 concatemerized binding sites for pathway-specific transcription factors upstream of a 'swappable' reporter module (**Fig. 7A**). The reporters are introduced into target cells via lentivirus and if needed, stable lines selected by puromycin resistance (**Fig. 7A**) (51-53). In a recent proof-of-principle siRNA screen of 3,700 disease-associated genes (including 306 kinases) in HEK293T cells harboring the NRF2 reporter (**Fig. 7B**) numerous hits were uncovered and validated in follow-up siRNAs and gain-of-function studies. For example, the *DPP3* protein was shown to bind KEAP1 to sterically displace NRF2 from KEAP1,

Table 2: Engineered reporters of signal transduction pathway activity

[1]Pathway	Reporter Name	12X DNA sequence	Agonist	[2]Cell Line	Fold Δ	Target genes
Wnt/β-catenin	pBAR	AGATCAAAGG	Wnt	A375, HEK293T	~800	AXIN2, LGR5, SP5
Retinoic Acid	pRARE	GGTTCACCGAAAGTTCA	Retinoic Acid	DLD1, SW480	~700	HOXA1, CDX1
Keap1/Nrf2	pARE	TCACAGTGACTCAGCA	Oxidative stress	KBM7, H2228	~11	HMOX1, NQO1
TGFβ	pSBE	AGCCAGACA	TGFβ	HaCaT, MCF10A	~80	PAI1, GADD45β
Notch	pCSL	CGTGGGAA	NICD transfection	HEK293T	~30	HES1, HEY2
TNFα/various	pNFκB	GGGAATTTCC	TNFα	A375, HEK293T	~300	IκBα, PTGS2,

[1] Not shown: STAT reporter, AP1 reporter, serum response factor (ELK1/SRF).
[2]These cell lines harbor intact signaling pathway and robust response following pathway agonist.

thus activating NRF2 signaling (54). Several kinases were identified and validated, including SLK, MAPK11 and the stress activated protein kinase MAPK12, all of which repress NRF2 signaling. The method can also detect other forms of perturbation. For example, over-expression studies using the same reporters identified 14

kinases whose expression impacted NRF2-dependent transcription in HEK293T cells (**Fig.7C**). A screen of the PKIS chemical library (Aim 4) revealed 5 compounds that inhibited and 9 that potentiated NRF2 transcriptional activity >2-fold. These approaches will be applied directly to analysis of DKs using reagents generated in Aim 3.1 (CRISPR) and Aim 4 (small molecule inhibitors).

Aim 3.4 Determining the effects of DK perturbation on kinome remodeling. These studies will use the DK and KIN-200 PRM-SID assays to measure the effects of perturbing well studied and dark kinases on the kinome as a whole (~330 kinases). The logic behind this approach is that the kinome is highly inter-connected, so even a project focused on DKs will benefit greatly from parallel measurements of as many kinases as possible. In addition, the kinome exhibits a high level of adaptive reprogramming in which inhibition of one kinase results in network rewiring and upregulation of other kinases, a phenomenon that has been most extensively studied in the case of resistance to targeted anti-cancer drugs (55, 56). Adaptive kinome reprogramming (already studied by Johnson and Sorger) is thought to reflect the disruption of feedback and feedforward regulatory loops that normally maintain cellular homeostasis. Its analysis in cancer has led to substantial advances in our understanding of immediate-early signal transduction and cellular homeostasis (1,57,58). The DK and KIN-200 PRM assays provide an unparalleled opportunity to measure reprogramming over time in multiple cell types. We will perturb well-studied kinases, particularly those that are clinically important, and measure cross-talk between the dark and light kinomes. We will also perturb DKs using genetic and small molecule reagents (Aims 2 and 4) and then study the impact on all kinases. Depending on reagent availability and suitability for CRISPR-based gene modification, such studies will be conducted in breast, lung, melanoma and ovarian cell lines (59). Because reprogramming is a dynamic process, data will be collected over time (typically 6-72 hr) exploiting the high through-put of PRM methods. Data will be analyzed using inference algorithms as described in Aim 1.3

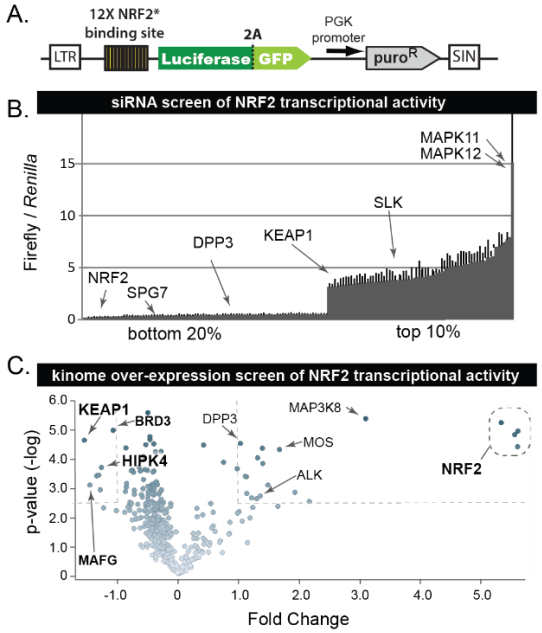


Fig. 7: Functional genomic screens of NRF2-dependent transcription as an example of reporter world assay. **A**. Novel NRF2- dependent transcriptional reporter. **B** HEK293T cells carrying the NRF2 reporter were transfected with siRNAs in 384 plates 72 hours before lysis and quantification. Three non-overlapping siRNAs were pooled at 20 nM. Data were normalized to CMV-driven Renilla luciferase. Error bars represent SD across biological triplicate experiments. **C.** Volcano plot representation of an over-expression screen of 384 kinases in HEK293T cells carrying the NRF2 reporter. P-values calculated from five biological repeats.

Aim 3.5: Profiling proteomics and phosphoproteomics analysis. These studies extend the PRM assays in Aim 2,2 by performing deep proteomics (~10-13,000 proteins per sample) and phosphoproteomic (~50,000 sites) on the most informative subset of DKs using a validated metal ion affinity enrichment protocol in combination with off-line fractionation and ten-channel isobaric peptide labeling (TMT-10) (59, 60) that increases LC-MS throughput 3-fold. In addition to the gains in throughput, combining the nine samples and the comparator reference pool (prepared from digests of all relevant cell lines) yields highly reliable relative fold-change data for the entire proteome. In addition, higher throughput survey phosphoproteomics (~10,000 phosphosites) will be performed in a 96 well format (61). Survey phosphoproteomics will be used to study dose-time relationships and to optimize conditions for more costly and time-consuming deep proteomics. In cases in which activation loop phosphopeptides can be detected, synthetic phosphopeptides will be prepared, stable isotope dilution response curves generated and phosphosite occupancy determined using PRM methods, as recently demonstrated using MRM-MS (62, 63). This allows us to infer kinase activity directly. For example, in the MIB experiment on four breast cancer PDX models (**Fig. 3**), allowed us to monitor activation loop phosphorylation on several DKs: CDK10-T196; DYRK1B-Y273; and PRKACB-T196/T198. As an alternative method, CST PTMScan motif-specific antiphosphopeptide antibodies will be used for enrichment of active kinases followed by phospho-PRM. Differences detected following perturbation of DKs will be integrated in network-level models as described in Aim 1.

Aim 3.6 Protein-protein interaction studies by mass spectrometry. Dynamic regulation of protein-protein interactions (PPI) governs the phenotypic response to extracellular information. Similarly, genomic alterations reorganize PPI networks to promote disease phenotypes. To identify and quantify protein connectivity (and post-translational modifications), we will employ immunoprecipitation (IP/MS) and biotin-ligase proximity labeling (BioID/MS). IP/MS and BioID are highly complementary. IP/MS captures relatively stable PPIs and produces datasets that are easily scored for true and false-positive discoveries. The method, it is susceptible to post-lysis artifacts and does not capture weak or transient PPIs. In contrast, BioID/MS leverages the

promiscuous BirA* ligase to biotinylate lysine residues within proximal proteins in live cells, improving detection of low affinity and transient interactions. The Major lab has significant expertise in IP/MS and growing experience with BioID/MS (54, 64-66). To physically position the dark kinases within the global protein interactome, we will integrate PPI data from the literature with experimentally-derived interaction maps obtained using both IP/MS and BioID.

This will be accomplished by using a BirA*-FLAG-phage lentiviral clone library to create at least two independent stable cell lines for each DK. Cell line identity will be bait-specific, based on the following considerations: 1) endogenous expression of the kinase (PRM assay/RNAseq), 2) varied tissue origin, 3) ease of culture, and 4) functional response to perturbation of the kinase by CRISPR/Cas9 KO and/or small molecule perturbation (Aims 3.1 to 3.5) . FLAG-based IP/MS and BioID/MS will be performed in parallel in biological duplicate. Tryptic peptides will be multiplexed using tandem mass tags (TMTs) to allow relative quantitation while minimizing instrument time on the Thermo Fusion Lumos mass spec in the Major lab. Raw MS data and resulting protein co-complex probabilistic scoring will be processed through our informatics pipeline and results integrated with data from the BioPlex project as well as other PPI studies (Aim 1).

Aim 3.7 Mass spectrometry-based metabolite profiling of DK perturbed cells is motived by data showing that therapeutic kinase inhibitors cause dramatic changes in cellular metabolism. Sophisticated metabolic profiling (e.g. with labels) is beyond the scope of this study but we will perform basic profiling for DKs, prioritized by our deep proteomics/phospho-proteomics data (Aim 3.5), using LC-MS/MS assays at our core facilities (http://www.bidmcmassspec.org (78) This will provide data on 200-300 identified polar and non-polar analytes that will be mapped to standard genome-wide metabolic models (79). Additional analysis will be performed in collaboration with our disease experts (see LOS).

Potential pitfalls and alternative approaches. This Aim involves a sufficient number of alternative approaches that technical difficulties with one method or DK will not hold up progress on others. The process of determining a function for a little studied protein is nonetheless fraught with uncertainties. For some kinases we may not be able to detect a phenotype in a convenient cell line using even the most sensitive assays. This will argue for testing other cell types, but even that may be insufficient. Navigating the cost-benefit analysis of pursuing a single kinase in greater depth, or moving onto another kinases is the primary function of the algorithm described in Aim1.1

Aim 4: Identifying and characterizing cell active chemical tools for dark kinases. Small molecule inhibitors are among the most valuable and highly sought research tools for interrogating the biological function of novel proteins and for initiating development of therapeutic drugs. Chemical tools allow temporal and dose-dependent inhibition of kinase activity in cells without long-term changes to the genome and associated compensatory mechanisms. Small molecules are active in a wide range of cells, including primary cells and tissues in which genetic tools can be hard to deliver. Importantly, use of chemical tools generates a view of kinase biology that is complementary to the result of genetic manipulation of cells (67). We will develop cellular assays for all expressed DKs and identify potent, selective cell-active inhibitors (chemical tools) for a set of prioritized DKs based in part from work in Aims 1-3, the literature and collaboration with the KMC. Our approach capitalizes on the recent development of nanoBRET assays to profile kinase inhibitors in living cells (see LOS from Promega), and the medicinal chemistry expertise of the kinase inhibitor team led by Tim Willson as part of the Structural Genomics Consortium (SGC). DKs prioritized by RFA-RM-16-026 can be grouped into four categories based on predicted substrates; three proteins are non-catalytic

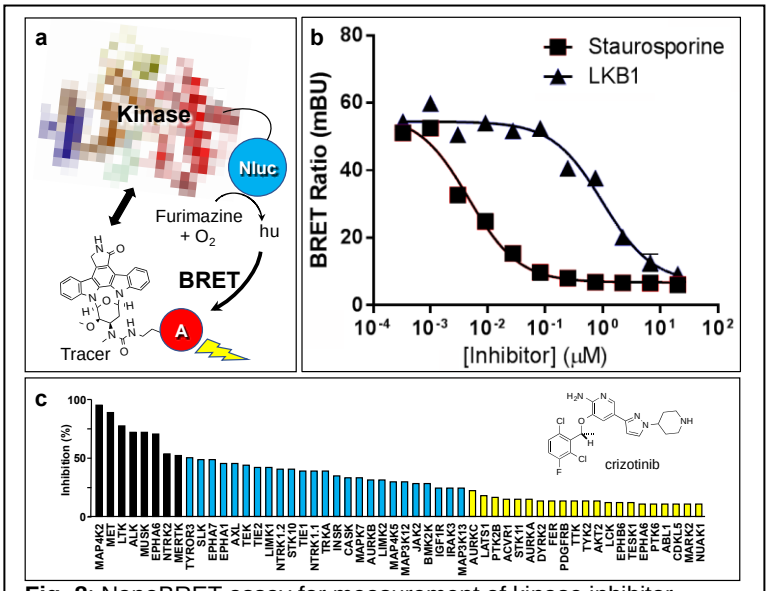


Fig. 8: NanoBRET assay for measurement of kinase inhibitor activity in cells. **a**. Illustration of the nanoBRET assay. A red-shifted fluorescent tracer binds to the kinase-Nluc fusion, resulting in BRET when the Nluc substrate generates light. **b**. Addition of a competitor ligand results in inhibition of the BRET signal. AAK1-Nluc was expressed in Hek293 cells. Fluorescent tracer and test inhibitor (staurosprine or LKB1) were added, followed by furimazine substrate. BRET signal was dose-dependently inhibited. **c**. Profile of the multikinase inhibitor crizotinib across 180 kinases. Crizotinib (1 μM) showed >50% displacement of tracer on 8 kinases (black bars), 50-25% on 25 kinases (blue bars), <25% on the remaining kinases (yellow bars and not shown).

regulatory subunits (**Table 1**). We expect to have developed cell-active chemical tools for up to 70 of the DKs during the six years of this grant, with approximately 40 selective inhibitors being characterized currently.

Aim 4.1 NanoBRET assays for profiling kinase inhibitor activities. The scientific impact of chemical tools is directly proportional to their degree of characterization (68). Conventional mechanisms for characterizing kinase inhibitors, which rely on biochemical assays, have two shortcomings. First, there are over 50 DKs for which no biochemical assay is available (69). Many of the missing assays are for pseudokinases, which are difficult to express or lack catalytic activity (70). Second, biochemical assays cannot determine whether an inhibitor has potent activity in cells, an essential requirement for elucidating biologic function. To overcome these limitations, we will use an innovative approach developed by Promega - 'nanoBRET assays' (**Fig. 8a**). The nanoBRET assay is a robust and scalable measure of kinase activity in living cells (71, 72) that does not require catalytic activity of the kinase and is agnostic to mode of inhibition. It monitors kinase inhibition in the setting of high cellular ATP concentrations, and can detect the influence of potentially unknown partner proteins. It depends on NanoLuc (Nluc), a small and extremely bright luciferase (73) that is fused to a kinase of interest. In NanoBRET use of a fluorescent tracer to cells expressing a Nluc fusion generates a signal that is competed off in a dose-dependent manner by small molecules that bind to the same kinase (**Fig. 8b**). We have synthesized 4 cell permeable fluorescent tracers that generate strong BRET signals for 180 kinase-Nluc fusions; this set includes 3 pseudokinases and 27 DKs. The nanoBRET assay has also been formatted for determining the selectivity profile of kinases by use of panels of cells, each expressing different kinase-Nluc protein and grown in 384-well assay plates. Using this protocol, the multikinase inhibitor crizotinib was screened against 180 protein kinases to give a direct measurement of selectivity (**Fig. 8c**).

We will develop nanoBRET assays for all of the DKs in the first year of the grant. Expression vectors for the kinases will be generated with N-terminal or C-terminal fusion to NLuc. Many DK clones are already in hand and commercial gene synthesis will be used to generate the others. Fluorescent tracers will be synthesized from 20 promiscuous kinase inhibitors selected from PKIS/PKIS2, each of which bind to >100 protein kinases, or from candidate molecules that bind to non-protein kinases (**Fig. 9a**). NLuc-kinase fusions will be transiently expressed in HEK293 cells and tested against the library of tracers to select those that yield an optimal BRET signal. The specificity of the signal will be demonstrated by addition of an excess of unmodified promiscuous kinase inhibitor.

Aim 4.2 Generation and characterization of new chemical tools. We will identify high quality chemical tools for up to 70 of the dark kinases during the six year program, and already have identified candidate tools for 40 of the dark kinases. High quality chemical tools require a) high potency for their target, b) selectivity over related proteins, and c) activity in a wide range of cells. We define a chemical tool as a molecule that demonstrates an IC_{50} of <1 μM against its primary target and cross-activity on <10 kinases using the in-cell nanoBRET assay. Where it is not possible to remove important off-target kinase activity, a negative control compound (a close analog that lacks activity on the target kinase) will be generated. All chemical tools will be produced in mg quantities and made available to investigators as DMSO solutions for use in cell assays.

Traditionally, chemical tools that meet our quality criteria are expensive and difficult to produce. However for protein kinases the generation of these tools is greatly facilitated by exploiting decades of biochemical and structural studies in academia and industry. For example, there are now many approved kinase inhibitors (e.g. crizotinib, **Fig. 8c**), some with exquisite selectivity (e.g. lapatinib for EGFR/HER2 or SGX523 for MET (68)). Chemogenomic sets, such as PKIS/PKIS2 (69, 74), provide a plentiful source of starting points for the development of chemical tools for the dark kinases.

By examining PKIS, PKIS2, and the literature we have already identified candidate tools for 40 of the dark kinases (**Fig. 9b**). We will test these candidates in the live cell nanoBRET assays and make those that pass available as tools, starting in Yr 2. Compounds that do not pass our criteria for a tool, will be subject to medicinal chemistry optimization.

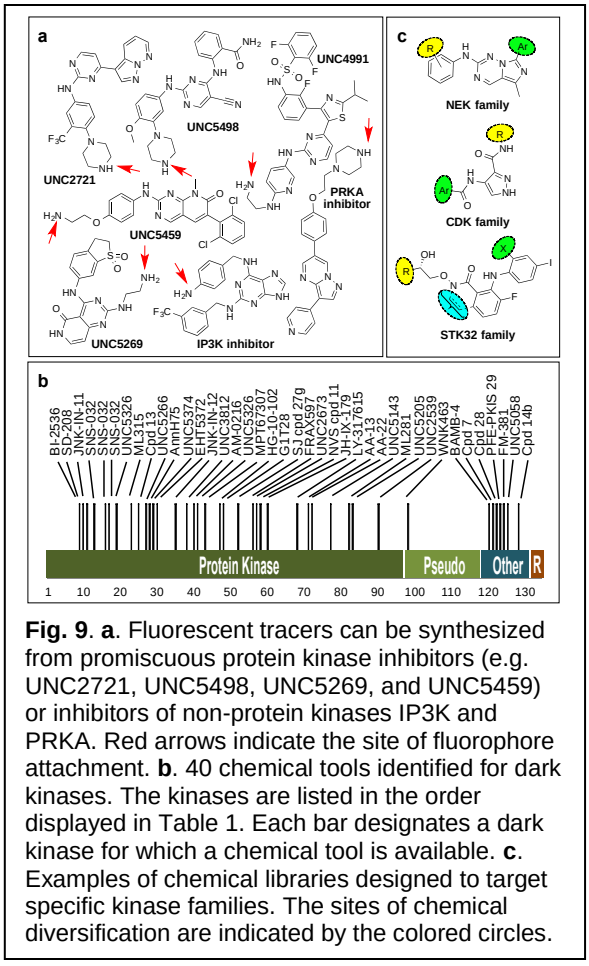


Fig. 9. **a**. Fluorescent tracers can be synthesized from promiscuous protein kinase inhibitors (e.g. UNC2721, UNC5498, UNC5269, and UNC5459) or inhibitors of non-protein kinases IP3K and PRKA. Red arrows indicate the site of fluorophore attachment. **b**. 40 chemical tools identified for dark kinases. The kinases are listed in the order displayed in Table 1. Each bar designates a dark kinase for which a chemical tool is available. **c**. Examples of chemical libraries designed to target specific kinase families. The sites of chemical diversification are indicated by the colored circles.

During the Yr 3-6 of the grant, we will identify chemical tools for an additional 30 of the remaining DKs, yielding a total of 70 tool compounds.

Aim 4.3 Screening existing kinase inhibitor collections in cells. PKIS/PKIS2 or non-protein kinase targeted compound sets will be screened at 1 µM to identify hits showing >90% displacement of the tracer, which will then be confirmed in dose-response assays. Compounds displaying IC_{50} <1 µM will be counter screened on >200 nanoBRET assays to determine selectivity. Inhibitors that fail to meet potency and selectivity criteria will be subject to medicinal chemistry optimization.

Aim 4.4 Screening new kinase inhibitor libraries*.* Many of the DKs belong to families that are known to be chemically tractable, and can be tackled by targeted library synthesis and screening. We have selected chemical templates with activity on these families (see examples in **Fig. 9c**), and will synthesize libraries of 100-200 analogs for screening in cells and subsequent chemical optimization.

Potential pitfalls and alternative approaches. Given the productive collaboration with Promega (see LOS) we do not anticipate significant difficulties in developing nanoBRET assays for the DKs. It is possible several kinases and pseudokinases will prove difficult for developing selective inhibitors. We will prioritize our chemistry efforts based on DK biology defined in Aims 1-3 to synthesize inhibitors for the most kinases relevant to the cellular function and potential involvement in disease.

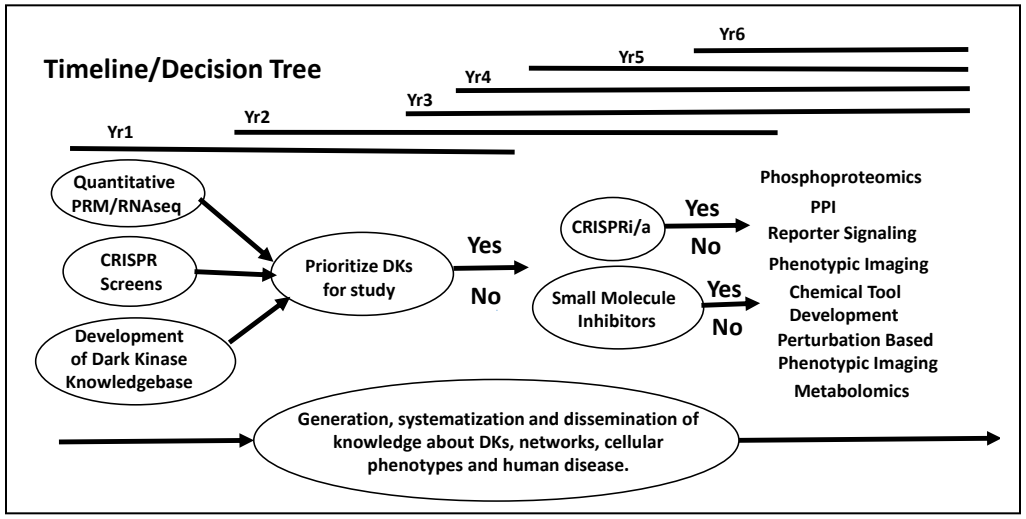
Aim 5: Collaborations to determine the expression and function of DKs in primary human cells and tissues. We will leverage collaborations with disease experts (see LOS) to study the functions of DKs in primary cells relevant to COPD, cystic fibrosis, arthritis, ALS, diabetes and Alzheimer's disease (the list will expand each year as we solicit new collaborations). We will mine data being collected by the Precision Medicine Initiative and large-scale cancer genetic studies to identify the subset of DKs whose dysregulation is associated with disease; studies in cells will then attempt to validate this association

Potential pitfalls and alternative approaches. Assays in Aims 2-3 are readily amenable to study of primary cells and tissues. We can also use the warm cadaver program at HMS and UNC to obtain different tissues and organ samples for analysis; patient-derived xenografts (PDXs) are readily available for many different tumor types.

MILESTONES AND DELIVERABLES

First year milestones: Aim 1: create a first generation open-source dark kinome knowledgebase (DKK). Aim 2: develop a PRM assay panel for all 134 proteins. Aim 3: begin loss-of-function CRISPR screens. Aim 4: establish nanoBRET assays for all 134 proteins and define selectivity profiles of 10 DK inhibitors. Aim 5: Begin collaborations for DK analysis in primary cells and tissues.

Milestones Years 2 and 3: Aim 1: Load new IDG data into DKK; begin network inference; develop data API with KMC; investigate disease associations; validate commercial reagents. Aim 2: PRM and RNAseq to define quantitative expression profiles of the 134 DKs. Aim 3: Complete survey of HAP1 and CRISPR knockout cells for all kinases. Initiate generation of CRSPRi/a cell lines or methods for inducible DK inactivation (~20 lines per year); analyze cells by CyCIF imaging, RNAseq, PRM and metabolomics profiling. Aim 4: complete selectivity profiles for 40 existing small molecule inhibitors in nanoBRET assays; perform PRM, phosphoproteomic, imaging and RNAseq measurements for these small molecule drugs; synthesize 10 new selective inhibitors for prioritized DKs.



Milestones Years 4-6: Aim 1: Build integrated, network-level models of the dark kinome and its interaction with known signaling networks; initiate validation studies on disease associations Aim 2: develop and apply phosphor-PRM assays to improve throughput of phosphoproteomics. Aim 3: Large-scale use of PRM and RNAseq for measuring dynamic changes in protein expression in response to CRISPRi/a or small molecule perturbation of DKs. Extend and deepen analysis of DKs cell and animals disease models Aim 4: Complete synthesis and selectivity profiles for 30 DK inhibitors. Aim 5: Continue disease model collaborations.

References

1. Duncan JS, Whittle MC, Nakamura K, Abell AN, Midland AA, Zawistowski JS, Johnson NL, Granger DA, Jordan NV, Darr DB, Usary J, Kuan PF, Smalley DM, Major B, He X, Hoadley KA, Zhou B, Sharpless NE, Perou CM, Kim WY, Gomez SM, Chen X, Jin J, Frye SV, Earp HS, Graves LM, Johnson GL. Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer. Cell. 2012;149(2):307-21. PMID: 22500798; PMC3328787.

2. Huang K-L, Li S, Mertins P, Cao S, Gunawardena HP, Ruggles KV, Mani DR, Clauser KR, Tanioka M, Usuary J, Kavuri SM, Xie L, Yoon C, Qiao JW, Wrobel J, Wyczalkowski M, Erdmann-Gilmore P, Snider JE, Hoog J, Singh P, Niu B, Guo Z, Sun SQ, Sanati S, Kawaler E, Wang X, Scott A, Ye K, McLellan MD, Wendl MC, Malovannaya A, Held JM, Gillette MA, Fenyö D, Kinsinger CR, Mesri M, Rodriguez H, Davies SR, Perou CM, Ma C, Townsend RR, Chen X, Carr SA, Ellis MJ, Ding L. Proteogenomic integration illuminates therapeutic targets in breast cancer xenografts. Nature Communications (in press).

3. Lin JR, Fallahi-Sichani M, Chen JY, Sorger PK. Cyclic Immunofluorescence (CycIF), A Highly Multiplexed Method for Single-cell Imaging. Curr Protoc Chem Biol. 2016;8(4):251-64. PMID: 27925668; PMC5233430.

4. Whiteaker JR, Halusa GN, Hoofnagle AN, Sharma V, MacLean B, Yan P, Wrobel JA, Kennedy J, Mani DR, Zimmerman LJ, Meyer MR, Mesri M, Rodriguez H, Clinical Proteomic Tumor Analysis C, Paulovich AG. CPTAC Assay Portal: a repository of targeted proteomic assays. Nat Methods. 2014;11(7):703-4. PMID: 24972168; PMC4113142.

5. Moniz L, Dutt P, Haider N, Stambolic V. Nek family of kinases in cell cycle, checkpoint control and cancer. Cell Div. 2011;6:18. PMID: 22040655; PMC3222597.

6. Zhou Q, Phoa AF, Abbassi RH, Hoque M, Reekie TA, Font JS, Ryan RM, Stringer BW, Day BW, Johns TG, Munoz L, Kassiou M. Structural Optimization and Pharmacological Evaluation of Inhibitors Targeting Dual-Specificity Tyrosine Phosphorylation-Regulated Kinases (DYRK) and CDC-like kinases (CLK) in Glioblastoma. J Med Chem. 2017. PMID: 28206758.

7. Fedorov O, Muller S, Knapp S. The (un)targeted cancer kinome. Nat Chem Biol. 2010;6(3):166-9. PMID: 20154661.

8. Kotecki M, Reddy PS, Cochran BH. Isolation and characterization of a near-haploid human cell line. Exp Cell Res. 1999;252(2):273-80. PMID: 10527618.

9. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, t Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018. PMID: 26978244; PMC4792175.

10. Choi SH, Kim YH, Hebisch M, Sliwinski C, Lee S, D'Avanzo C, Chen H, Hooli B, Asselin C, Muffat J, Klee JB, Zhang C, Wainger BJ, Peitz M, Kovacs DM, Woolf CJ, Wagner SL, Tanzi RE, Kim DY. A three-dimensional human neural cell culture model of Alzheimer's disease. Nature. 2014;515(7526):274-8. PMID: 25307057; PMC4366007.

11. AlQuraishi M, Sorger PK. Reproducibility will only come with data liberation. Sci Transl Med. 2016;8(339):339ed7. PMID: 27194726; PMC5084089.

12. Fallahi-Sichani M, Becker V, Izar B, Baker GJ, Lin JR, Boswell SA, Shah P, Rotem A, Garraway LA, Sorger PK. Adaptive resistance of melanoma cells to RAF inhibition via reversible induction of a slowly dividing de-differentiated state. Mol Syst Biol. 2017;13(1):905. PMID: 28069687; PMC5248573.

13. Fallahi-Sichani M, Moerke NJ, Niepel M, Zhang T, Gray NS, Sorger PK. Systematic analysis of BRAF(V600E) melanomas reveals a role for JNK/c-Jun pathway in adaptive resistance to drug-induced apoptosis. Mol Syst Biol. 2015;11(3):797. PMID: 25814555; PMC4380931.

14. Niepel M, Hafner M, Pace EA, Chung M, Chai DH, Zhou L, Schoeberl B, Sorger PK. Profiles of Basal and stimulated receptor signaling networks predict drug response in breast cancer lines. Sci Signal. 2013;6(294):ra84. PMID: 24065145; PMC3845839.

15. Wu Y, Johnson GL, Gomez SM. Data-driven modeling of cellular stimulation, signaling and output response in RAW 264.7 cells. J Mol Signal. 2008;3:11. PMID: 18498628; PMC2441624.

16. Carlson DE, DeMaria EJ, Campbell RW, Gann DS. Behavioral and hormonal influence on blood volume restitution after hemorrhage in swine. Am J Physiol. 1989;256(1 Pt 2):R207-16. PMID: 2536248.

17. Gomez SM, Lo SH, Rzhetsky A. Probabilistic prediction of unknown metabolic and signal-transduction networks. Genetics. 2001;159(3):1291-8. PMID: 11729170; PMC1461852.

18. Gomez SM, Noble WS, Rzhetsky A. Learning to predict protein-protein interactions from protein sequences. Bioinformatics. 2003;19(15):1875-81. PMID: 14555619.

19. Doolittle JM, Gomez SM. Mapping protein interactions between Dengue virus and its human and insect hosts. PLoS Negl Trop Dis. 2011;5(2):e954. PMID: 21358811; PMC3039688.

20. Saez-Rodriguez J, Alexopoulos LG, Zhang M, Morris MK, Lauffenburger DA, Sorger PK. Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. Cancer Res. 2011;71(16):5400-11. PMID: 21742771; PMC3207250.

21. Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, Sorger PK. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. Mol Syst Biol. 2009;5:331. PMID: 19953085; PMC2824489.

22. Lopez CF, Muhlich JL, Bachman JA, Sorger PK. Programming biological models in Python using PySB. Mol Syst Biol. 2013;9:646. PMID: 23423320; PMC3588907.

23. Eydgahi H, Chen WW, Muhlich JL, Vitkup D, Tsitsiklis JN, Sorger PK. Properties of cell death models calibrated and compared using Bayesian approaches. Mol Syst Biol. 2013;9:644. PMID: 23385484; PMC3588908.

24. Albeck JG, Burke JM, Spencer SL, Lauffenburger DA, Sorger PK. Modeling a snap-action, variable-delay switch controlling extrinsic cell death. PLoS Biol. 2008;6(12):2831-52. PMID: 19053173; PMC2592357.

25. Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, Lauffenburger DA, Sorger PK. Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Mol Syst Biol. 2009;5:239. PMID: 19156131; PMC2644173.

26. Xu K, Morgan KT, Todd Gehris A, Elston TC, Gomez SM. A whole-body model for glycogen regulation reveals a critical role for substrate cycling in maintaining blood glucose homeostasis. PLoS Comput Biol. 2011;7(12):e1002272. PMID: 22163177; PMC3233304.

27. Hill SM, Heiser LM, Cokelaer T, Unger M, Nesser NK, Carlin DE, Zhang Y, Sokolov A, Paull EO, Wong CK, Graim K, Bivol A, Wang H, Zhu F, Afsari B, Danilova LV, Favorov AV, Lee WS, Taylor D, Hu CW, Long BL, Noren DP, Bisberg AJ, Consortium H-D, Mills GB, Gray JW, Kellen M, Norman T, Friend S, Qutub AA, Fertig EJ, Guan Y, Song M, Stuart JM, Spellman PT, Koeppl H, Stolovitzky G, Saez-Rodriguez J, Mukherjee S. Inferring causal molecular networks: empirical assessment through a community-based effort. Nat Methods. 2016;13(4):310-8. PMID: 26901648; PMC4854847.

28. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, Schultz N, Bader GD, Sander C. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res. 2011;39(Database issue):D685-90. PMID: 21071392; PMC3013659.

29. Hodes RJ, Buckholtz N. Accelerating Medicines Partnership: Alzheimer's Disease (AMP-AD) Knowledge Portal Aids Alzheimer's Drug Discovery through Open Data Sharing. Expert Opin Ther Targets. 2016;20(4):389-91. PMID: 26853544.

30. Urisman A, Levin RS, Gordan JD, Webber JT, Hernandez H, Ishihama Y, Shokat KM, Burlingame AL. An Optimized Chromatographic Strategy for Multiplexing In Parallel Reaction Monitoring Mass Spectrometry: Insights from Quantitation of Activated Kinases. Mol Cell Proteomics. 2017;16(2):265-77. PMID: 27940637; PMC5294213.

31. Oppermann FS, Gnad F, Olsen JV, Hornberger R, Greff Z, Keri G, Mann M, Daub H. Large-scale proteomics analysis of the human kinome. Mol Cell Proteomics. 2009;8(7):1751-64. PMID: 19369195; PMC2709199.

32. Kennedy JJ, Abbatiello SE, Kim K, Yan P, Whiteaker JR, Lin C, Kim JS, Zhang Y, Wang X, Ivey RG, Zhao L, Min H, Lee Y, Yu MH, Yang EG, Lee C, Wang P, Rodriguez H, Kim Y, Carr SA, Paulovich AG. Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. Nat Methods. 2014;11(2):149-55. PMID: 24317253; PMC3922286.

33. Abbatiello SE, Schilling B, Mani DR, Zimmerman LJ, Hall SC, MacLean B, Albertolle M, Allen S, Burgess M, Cusack MP, Gosh M, Hedrick V, Held JM, Inerowicz HD, Jackson A, Keshishian H, Kinsinger CR, Lyssand J, Makowski L, Mesri M, Rodriguez H, Rudnick P, Sadowski P, Sedransk N, Shaddox K, Skates SJ, Kuhn E, Smith D, Whiteaker JR, Whitwell C, Zhang S, Borchers CH, Fisher SJ, Gibson BW, Liebler DC, MacCoss MJ, Neubert TA, Paulovich AG, Regnier FE, Tempst P, Carr SA. Large-Scale Interlaboratory Study to Develop, Analytically Validate and Apply Highly Multiplexed, Quantitative Peptide Assays to Measure Cancer-Relevant Proteins in Plasma. Mol Cell Proteomics. 2015;14(9):2357-74. PMID: 25693799; PMC4563721.

34. Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. Mol Cell Proteomics. 2012;11(11):1475-88. PMID: 22865924; PMC3494192.

35. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP. Comprehensive analysis of kinase inhibitor selectivity. Nat Biotechnol. 2011;29(11):1046-51. PMID: 22037378.

36. Hoofnagle AN, Whiteaker JR, Carr SA, Kuhn E, Liu T, Massoni SA, Thomas SN, Townsend RR, Zimmerman LJ, Boja E, Chen J, Crimmins DL, Davies SR, Gao Y, Hiltke TR, Ketchum KA, Kinsinger CR, Mesri M, Meyer MR, Qian WJ, Schoenherr RM, Scott MG, Shi T, Whiteley GR, Wrobel JA, Wu C, Ackermann BL, Aebersold R, Barnidge DR, Bunk DM, Clarke N, Fishman JB, Grant RP, Kusebauch U, Kushnir MM, Lowenthal MS, Moritz RL, Neubert H, Patterson SD, Rockwood AL, Rogers J, Singh RJ, Van Eyk JE, Wong SH, Zhang S, Chan DW, Chen X, Ellis MJ, Liebler DC, Rodland KD, Rodriguez H, Smith RD, Zhang Z, Zhang H, Paulovich AG. Recommendations for the Generation, Quantification, Storage, and Handling of Peptides Used for Mass Spectrometry-Based Assays. Clin Chem. 2016;62(1):48-69. PMID: 26719571; PMC4830481.

37. Bonora M, Wieckowsk MR, Chinopoulos C, Kepp O, Kroemer G, Galluzzi L, Pinton P. Molecular mechanisms of cell death: central implication of ATP synthase in mitochondrial permeability transition. Oncogene. 2015;34(12):1608. PMID: 25790189.

38. Sendoel A, Dunn JG, Rodriguez EH, Naik S, Gomez NC, Hurwitz B, Levorse J, Dill BD, Schramek D, Molina H, Weissman JS, Fuchs E. Translation from unconventional 5' start sites drives tumour initiation. Nature. 2017;541(7638):494-9. PMID: 28077873; PMC5287289.

39. Erickson BK, Rose CM, Braun CR, Erickson AR, Knott J, McAlister GC, Wuhr M, Paulo JA, Everley RA, Gygi SP. A Strategy to Combine Sample Multiplexing with Targeted Proteomics Assays for High-Throughput Protein Signature Characterization. Mol Cell. 2017;65(2):361-70. PMID: 28065596; PMC5250569.

40. Tan J, Martin SE. Validation of Synthetic CRISPR Reagents as a Tool for Arrayed Functional Genomic Screening. PLoS One. 2016;11(12):e0168968. PMID: 28030641; PMC5193459

41. Evers B, Jastrzebski K, Heijmans JP, Grernrum W, Beijersbergen RL, Bernards R. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. Nat Biotechnol. 2016;34(6):631-3. PMID: 27111720.

42. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, Qi LS, Kampmann M, Weissman JS. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell. 2014;159(3):647-61. PMID: 25307932; PMC4253859.

43. Horlbeck MA, Gilbert LA, Villalta JE, Adamson B, Pak RA, Chen Y, Fields AP, Park CY, Corn JE, Kampmann M, Weissman JS. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. Elife. 2016;5. PMID: 27661255; PMC5094855.

44. Iwamoto M, Bjorklund T, Lundberg C, Kirik D, Wandless TJ. A general chemical method to regulate protein stability in the mammalian central nervous system. Chem Biol. 2010;17(9):981-8. PMID: 20851347; PMC2943492.

45. Liu KI, Ramli MN, Woo CW, Wang Y, Zhao T, Zhang X, Yim GR, Chong BY, Gowher A, Chua MZ, Jung J, Lee JH, Tan MH. A chemical-inducible CRISPR-Cas9 system for rapid control of genome editing. Nat Chem Biol. 2016;12(11):980-7. PMID: 27618190.

46. Polstein LR, Gersbach CA. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. Nat Chem Biol. 2015;11(3):198-200. PMID: 25664691; PMC4412021.

47. Sheridan RM, Bentley DL. Selectable one-step PCR-mediated integration of a degron for rapid depletion of endogenous human proteins. Biotechniques. 2016;60(2):69-74. PMID: 26842351; PMC4893940.

48. Bray MA, Singh S, Han H, Davis CT, Borgeson B, Hartland C, Kost-Alimova M, Gustafsdottir SM, Gibson CC, Carpenter AE. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc. 2016;11(9):1757-74. PMID: 27560178; PMC5223290.

49. Hafner M, Niepel M, Chung M, Sorger PK. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat Methods. 2016;13(6):521-7. PMID: 27135972; PMC4887336.

50. Lin JR, Fallahi-Sichani M, Sorger PK. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. Nat Commun. 2015;6:8390. PMID: 26399630; PMC4587398.

51. Walker MP, Stopford CM, Cederlund M, Fang F, Jahn C, Rabinowitz AD, Goldfarb D, Graham DM, Yan F, Deal AM, Fedoriw Y, Richards KL, Davis IJ, Weidinger G, Damania B, Major MB. FOXP1 potentiates Wnt/beta-catenin signaling in diffuse large B cell lymphoma. Sci Signal. 2015;8(362):ra12. PMID: 25650440; PMC4356208.

52. Madan B, Walker MP, Young R, Quick L, Orgel KA, Ryan M, Gupta P, Henrich IC, Ferrer M, Marine S, Roberts BS, Arthur WT, Berndt JD, Oliveira AM, Moon RT, Virshup DM, Chou MM, Major MB. USP6 oncogene promotes Wnt signaling by deubiquitylating Frizzleds. Proc Natl Acad Sci U S A. 2016;113(21):E2945-54. PMID: 27162353; PMC4889410.

53. Major MB, Roberts BS, Berndt JD, Marine S, Anastas J, Chung N, Ferrer M, Yi X, Stoick-Cooper CL, von Haller PD, Kategaya L, Chien A, Angers S, MacCoss M, Cleary MA, Arthur WT, Moon RT. New regulators of Wnt/beta-catenin signaling revealed by integrative molecular screening. Sci Signal. 2008;1(45):ra12. PMID: 19001663.

54. Hast BE, Goldfarb D, Mulvaney KM, Hast MA, Siesser PF, Yan F, Hayes DN, Major MB. Proteomic analysis of ubiquitin ligase KEAP1 reveals associated proteins that inhibit NRF2 ubiquitination. Cancer Res. 2013;73(7):2199-210. PMID: 23382044; PMC3618590.

55. Stuhlmiller TJ, Earp HS, Johnson GL. Adaptive reprogramming of the breast cancer kinome. Clin Pharmacol Ther. 2014;95(4):413-5. PMID: 24413269; PMC4091669.

56. Johnson GL, Stuhlmiller TJ, Angus SP, Zawistowski JS, Graves LM. Molecular pathways: adaptive kinome reprogramming in response to targeted inhibition of the BRAF-MEK-ERK pathway in cancer. Clin Cancer Res. 2014;20(10):2516-22. PMID: 24664307; PMC4024346.

57. Stuhlmiller TJ, Miller SM, Zawistowski JS, Nakamura K, Beltran AS, Duncan JS, Angus SP, Collins KA, Granger DA, Reuther RA, Graves LM, Gomez SM, Kuan PF, Parker JS, Chen X, Sciaky N, Carey LA, Earp HS, Jin J, Johnson GL. Inhibition of Lapatinib-Induced Kinome Reprogramming in ERBB2-Positive Breast Cancer by Targeting BET Family Bromodomains. Cell Rep. 2015;11(3):390-404. PMID: 25865888; PMC4408261.

58. Zawistowski JS, Bevill SM, Goulet DR, Stuhlmiller TJ, Beltran AS, Olivares-Quintero JF, Singh D, Sciaky N, Parker JS, Rashid NU, Chen X, Duncan JS, Whittle MC, Angus SP, Velarde SH, Golitz BT, He X, Santos C, Darr DB, Gallagher K, Graves LM, Perou CM, Carey LA, Earp HS, Johnson GL. Enhancer Remodeling during Adaptive Bypass to MEK Inhibition Is Attenuated by Pharmacologic Targeting of the P-TEFb Complex. Cancer Discov. 2017;7(3):302-21. PMID: 28108460; PMC5340640.

59. McAlister GC, Huttlin EL, Haas W, Ting L, Jedrychowski MP, Rogers JC, Kuhn K, Pike I, Grothe RA, Blethrow JD, Gygi SP. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. Anal Chem. 2012;84(17):7469-78. PMID: 22880955; PMC3715028.

60. Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, Wang X, Qiao JW, Cao S, Petralia F, Kawaler E, Mundt F, Krug K, Tu Z, Lei JT, Gatza ML, Wilkerson M, Perou CM, Yellapantula V, Huang KL, Lin C, McLellan MD, Yan P, Davies SR, Townsend RR, Skates SJ, Wang J, Zhang B, Kinsinger CR, Mesri M, Rodriguez H, Ding L, Paulovich AG, Fenyo D, Ellis MJ, Carr SA, Nci C. Proteogenomics connects somatic mutations to signalling in breast cancer. Nature. 2016;534(7605):55-62. PMID: 27251275; PMC5102256.

61. Humphrey SJ, Azimifar SB, Mann M. High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. Nat Biotechnol. 2015;33(9):990-5. PMID: 26280412.

62. Kennedy JJ, Yan P, Zhao L, Ivey RG, Voytovich UJ, Moore HD, Lin C, Pogosova-Agadjanyan EL, Stirewalt DL, Reding KW, Whiteaker JR, Paulovich AG. Immobilized Metal Affinity Chromatography Coupled to Multiple Reaction Monitoring Enables Reproducible Quantification of Phospho-signaling. Mol Cell Proteomics. 2016;15(2):726-39. PMID: 26621847; PMC4739685.

63. Whiteaker JR, Zhao L, Yan P, Ivey RG, Voytovich UJ, Moore HD, Lin C, Paulovich AG. Peptide Immunoaffinity Enrichment and Targeted Mass Spectrometry Enables Multiplex, Quantitative Pharmacodynamic Studies of Phospho-Signaling. Mol Cell Proteomics. 2015;14(8):2261-73. PMID: 25987412; PMC4528251.

64. Mulvaney KM, Matson JP, Siesser PF, Tamir TY, Goldfarb D, Jacobs TM, Cloer EW, Harrison JS, Vaziri C, Cook JG, Major MB. Identification and Characterization of MCM3 as a Kelch-like ECH-associated Protein 1 (KEAP1) Substrate. J Biol Chem. 2016;291(45):23719-33. PMID: 27621311; PMC5095425.

65. Kim TY, Siesser PF, Rossman KL, Goldfarb D, Mackinnon K, Yan F, Yi X, MacCoss MJ, Moon RT, Der CJ, Major MB. Substrate trapping proteomics reveals targets of the betaTrCP2/FBXW11 ubiquitin ligase. Mol Cell Biol. 2015;35(1):167-81. PMID: 25332235; PMC4295375.

66. Goldfarb D, Hast BE, Wang W, Major MB. Spotlite: web application and augmented algorithms for predicting co-complexed proteins from affinity purification--mass spectrometry data. J Proteome Res. 2014;13(12):5944-55. PMID: 25300367; PMC4360886.

67. Weiss WA, Taylor SS, Shokat KM. Recognizing and exploiting differences between RNAi and small-molecule inhibitors. Nat Chem Biol. 2007;3(12):739-44. PMID: 18007642; PMC2924165.

68. Uitdehaag JC, Verkaar F, Alwan H, de Man J, Buijsman RC, Zaman GJ. A guide to picking the most selective kinase inhibitor tool compounds for pharmacological validation of drug targets. Br J Pharmacol. 2012;166(3):858-76. PMID: 22250956; PMC3417414.

69. Drewry DH, Wells CI, Andrews D, Richard Angell, Al-Ali H, Axtman AD, Capuzzi SJ, Elkins J, Ettmayer P, Frederiksen M, Gileadi O, Grey N, Hooper A, Knapp S, Laufer S, Luecking U, Muller-Knapp S, Muratov E,

Rajiah D, Saikatendu K, Trieber D, Zuercher WJ, Willson TM. Progress Towards a Public Chemogenomic Set for Protein Kinases: A Call for Contributions. bioaRxiv. 2017.

70. Hammaren HM, Virtanen AT, Silvennoinen O. Nucleotide-binding mechanisms in pseudokinases. Biosci Rep. 2016;36(1):e00282.

71. Machleidt T, Woodroofe CC, Schwinn MK, Mendez J, Robers MB, Zimmerman K, Otto P, Daniels DL, Kirkland TA, Wood KV. NanoBRET - a novel BRET platform for the analysis of protein-protein interactions. ACS Chem Biol. 2015;10(8):1797-804. PMID: 26006698.

72. Robers MB, Dart ML, Woodroofe CC, Zimprich CA, Kirkland TA, Machleidt T, Kupcho KR, Levin S, Hartnett JR, Zimmerman K, Niles AL, Ohana RF, Daniels DL, Slater M, Wood MG, Cong M, Cheng YQ, Wood KV. Target engagement and drug residence time can be observed in living cells with BRET. Nat Commun. 2015;6:10091. PMID: 26631872; PMC4686764.

73. Hall MP, Unch J, Binkowski BF, Valley MP, Butler BL, Wood MG, Otto P, Zimmerman K, Vidugiris G, Machleidt T, Robers MB, Benink HA, Eggers CT, Slater MR, Meisenheimer PL, Klaubert DH, Fan F, Encell LP, Wood KV. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. ACS Chem Biol. 2012;7(11):1848-57. PMID: 22894855; PMC3501149.

74. Elkins JM, Fedele V, Szklarz M, Abdul Azeez KR, Salah E, Mikolajczyk J, Romanov S, Sepetov N, Huang XP, Roth BL, Al Haj Zen A, Fourches D, Muratov E, Tropsha A, Morris J, Teicher BA, Kunkel M, Polley E, Lackey KE, Atkinson FL, Overington JP, Bamborough P, Muller S, Price DJ, Willson TM, Drewry DH, Knapp S, Zuercher WJ. Comprehensive characterization of the Published Kinase Inhibitor Set. Nat Biotechnol. 2016;34(1):95-103. PMID: 26501955.

75. Tsai Y, Dominguez D, Gomez S, and Wang Z. Transcriptome-wide identification and study of cancer-specific splicing events across multiple tumors. *Oncotarget*. 2015. Feb 5 PMID: 25749525 PMC4466652

76. Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, Fallahi-Sichani M, Dutton-Regester K, Lin JR, Cohen O, Shah P, Lu D, Genshaft AS, Hughes TK[1], Ziegler CG[1], Kazer SW, Gaillard A, Kolb KE, Villani AC, Johannessen CM, Andreev AY, Van Allen EM, Bertagnolli M, Sorger PK, Sullivan RJ, Flaherty KT, Frederick DT, Jané-Valbuena J, Yoon CH, Rozenblatt-Rosen O, Shalek AK, Regev A, Garraway LA Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016 Apr 8;352(6282):189-96. PMID: 27124452 PMC4944528

77. Li S, Shen D, Shao J, Crowder R, Liu W, Prat A, He X, Liu S, Hoog J, Lu C, Ding L, Griffith OL, Miller C, Larson D, Fulton RS, Harrison M, Mooney T, McMichael JF, Luo J, Tao Y, Goncalves R, Schlosberg C, Hiken JF, Saied L, Sanchez C, Giuntoli T, Bumb C, Cooper C, Kitchens RT, Lin A, Phommaly C, Davies SR, Zhang J, Kavuri MS, McEachern D, Dong YY, Ma C, Pluard T, Naughton M, Bose R, Suresh R, McDowell R, Michel L, Aft R, Gillanders W, DeSchryver K, Wilson RK, Wang S, Mills GB, Gonzalez-Angulo A, Edwards JR, Maher C, Perou CM, Mardis ER, Ellis MJ Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. Cell Rep. 2013 Sep 26;4(6):1116-30. PMID: 24055055 PMC3881975

78. Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat Protoc. 2006;1(1):387-96. PMID: 17406261

79. Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, Thorleifsson SG, Agren R, Bölling C, Bordel S, Chavali AK, Dobson P, Dunn WB, Endler L, Hala D, Hucka M, Hull D, Jameson D, Jamshidi N, Jonsson JJ, Juty N, Keating S, Nookaew I, Le Novère N, Malys N, Mazein A, Papin JA, Price ND, Selkov E Sr, Sigurdsson MI, Simeonidis E, Sonnenschein N, Smallbone K, Sorokin A, van Beek JH, Weichart D, Goryanin I, Nielsen J, Westerhoff HV, Kell DB, Mendes P, Palsson BØ. A community-driven global reconstruction of human metabolism. Nat Biotechnol. 2013 May; 31(5):419-25. PMID: 23455439 PMC3856361