

RESEARCH

Open Access



Illuminating the dark kinase: utilizing multiplex peptide activity arrays to functionally annotate understudied kinases

Abdul-rizaq Hamoud¹, Khaled Alganem¹, Sean Hanna¹, Michael Morran¹, Nicholas Henkel¹, Ali S. Imami¹, William Ryan¹, Smita Sahay¹, Priyanka Puvender¹, Austin Kunch¹, Taylen O. Arvay¹, Jarek Meller^{3,4,5,6,7}, Rammohan Shukla⁸, Sinead M. O'Donovan⁹ and Robert McCullumsmith^{1,2*}

Abstract

Protein kinases are critical components of a myriad biological processes and strongly associated with various diseases. While kinase research has been a point of focus in biomedical research for several decades, a large portion of the kinase is still considered understudied or “dark,” because prior research is targeted towards a subset of kinases with well-established roles in cellular processes. We present an empirical and in-silico hybrid workflow to extend the functional knowledge of understudied kinases. Utilizing multiplex peptide activity arrays and robust in-silico analyses, we extended the functional knowledge of five dark tyrosine kinases (AATK, EPHA6, INSRR, LTK, TNK1) and explored their roles in schizophrenia, Alzheimer's dementia (AD), and major depressive disorder (MDD). Using this hybrid approach, we identified 195 novel kinase-substrate interactions with variable degrees of affinity and linked extended functional networks for these kinases to biological processes that are impaired in psychiatric and neurological disorders. Biochemical assays and mass spectrometry were used to confirm a putative substrate of EPHA6, an understudied dark tyrosine kinase. We examined the EPHA6 network and knowledgebase in schizophrenia using reporter peptides identified and validated from the multi-plex array with high affinity for phosphorylation by EPHA6. Identification and confirmation of putative substrates for understudied kinases provides a wealth of actionable information for the development of new drug treatments as well as exploration of the pathophysiology of disease states using signaling network approaches.

*Correspondence:

Robert McCullumsmith
robert.mccullumsmith@utoledo.edu

¹ Department of Neurosciences and Psychiatry, University of Toledo College of Medicine, Toledo, OH, USA

² Neuroscience Institute, ProMedica, Toledo, OH, USA

³ Department of Biomedical Informatics, University of Cincinnati, Cincinnati, OH, USA

⁴ Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

⁵ Division of Biostatistics and Bioinformatics, Department of Environmental and Public Health Sciences, University of Cincinnati, Cincinnati, OH, USA

⁶ Department of Pharmacology and System Biology, College of Medicine, University of Cincinnati, Cincinnati, OH, USA

⁷ Department of Electrical Engineering and Computer Science, College of Engineering and Applied Sciences, University of Cincinnati, Cincinnati, OH, USA

⁸ Department of Zoology and Physiology, University of Wyoming, Laramie, WY, USA

⁹ Department of Biological Sciences, University of Limerick, Castletroy, Limerick, Ireland



Introduction

Protein kinases are integral components of various biological processes and act as master regulators of critical signaling pathways. Disruption of these important enzymes may result in unchecked signaling patterns leading to abnormal cellular behavior. As a result, perturbed kinase signaling, at the genetic or proteomic level, is strongly linked to various diseases. Gain or loss of function mutations in kinase genes are found in several diseases, including developmental and metabolic disorders, as well as some types of cancer [1]. Further, the active kinome, which is the broad-based activity of the complete set of protein kinases, is altered in cancer, psychiatric and neurodegenerative disorders [2–4]. These observations have shifted the focus towards protein kinases as promising druggable targets, resulting in an increase of Food and Drug Administration (FDA) approved small molecules targeting protein kinases over the past ten years [5].

However, while protein kinases have been a central focus of research for several decades, a large portion of the kinome is considered understudied or “dark.” The dark kinome is a consequence of signaling studies being primarily dedicated to a subset of kinases with well-established roles in cellular processes and signaling [6]. “Dark” status is given to a kinase when it is poorly studied compared to the rest of kinome [7, 8]. This disparity of knowledge has left much of the kinome untargeted. For example, less than 10% of kinases have FDA approved small molecule inhibitors or activators [9]. Despite the presence of more than 500 protein kinases in the human kinome, only 76 FDA approved kinase inhibitors have been developed [10]. This is in part due to the complexity of designing selective kinase inhibitors, but also due to poorly studied kinases and kinase networks. Investigating dark kinases will provide insights into this gap of knowledge, highlighted by the Illuminating the Druggable Genome program (IDG) [11]. The purpose of this program is to “illuminate” and target dark kinases and other understudied proteins.

Two important components are needed for dark protein kinases: 1) functional annotation of their downstream substrates, and 2) protein kinase-disease associations. Annotations are key to uncover related cascades and pathways, while disease associations reveal therapeutic opportunities and novel drug targets. We present a hybrid empirical and in-silico workflow to extend the functional knowledge of dark kinases. To empirically identify novel kinase substrates, the PamStation12 (PamGene International B.V.) platform was utilized to profile purified recombinant kinases. The same platform was used to assess the active kinome in disease substrates and to deconvolve their kinomic signatures

utilizing the recombinant kinase screening profiles (Fig. 1).

In the central nervous system (CNS), protein kinases are involved in many critical processes such as long-term potentiation, nutrient sensing, and neuroplasticity [4, 12, 13]. Moreover, perturbed protein kinase networks are associated with several psychiatric and neurological disorders [4]. In this study, we used our hybrid workflow to extend the functional knowledge of five dark tyrosine kinases from different kinase families and their associations with Schizophrenia (SCZ), Alzheimer’s Dementia (AD), and Major Depressive Disorder (MDD).

Methods

Dark kinome

To determine membership in the “dark kinome,” a curated list of all human protein kinases was obtained [14]. This list was queried against the Integrated Network and Dynamical Reasoning Assembler (INDRA) database, a bibliometric knowledge aggregator consisting of unique, de-duplicated statements of causal biological mechanisms drawn from natural language processing of PubMed abstracts, PubMedCentral full texts, and a host of other resources such as Pathway Commons and BioGRID [15]. Consistent with the NIH Illuminating the Druggable Genome (IDG) consortium and prevailing literature, dark kinases were defined as those belonging to the bottom third percentile of the kinome ranked by standardized log-transformed INDRA statements (Supplementary Table 1) [16]. Of these dark kinases, 19 tyrosine kinases were identified from the INDRA database, of which five tyrosine kinases were selected based on recombinant kinase availability, reported tyrosine kinase activity, expression in the brain, and relevance to schizophrenia.

Sample preparation

Tyrosine kinase non-receptor 1 (TNK1, AA range G106-D390) and EPH receptor A6 (EPHA6, AA range R683-V1130) were purchased from ReactionBiology. Leukocyte tyrosine kinase receptor (LTK, AA range S20-P424) and Insulin receptor-related receptor (INSRR, AA range L24-A747) were purchased from R&Dsystems. Apoptosis associated tyrosine kinase (AATK, AA range E1216-A1374) was purchased from LSBio. The recombinant kinases were prepped in fresh 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.25 mM DTT, 0.1 mM EGTA, 0.1 mM EDTA, 0.1 mM PMSF, 25% Glycerol for the PamStation12 as well as the benchtop activity assay validation experiment.

Postmortem dorsolateral prefrontal cortex from non-psychiatric control (17, M:8 and F:9) and schizophrenia (16, M:8 and F:9) subjects (Maryland Brain Collection) were homogenized using Kimble Biomasher II Closed System

Extending the Knowledge of a Dark Kinase Using Functional Kinase Activity Profiling

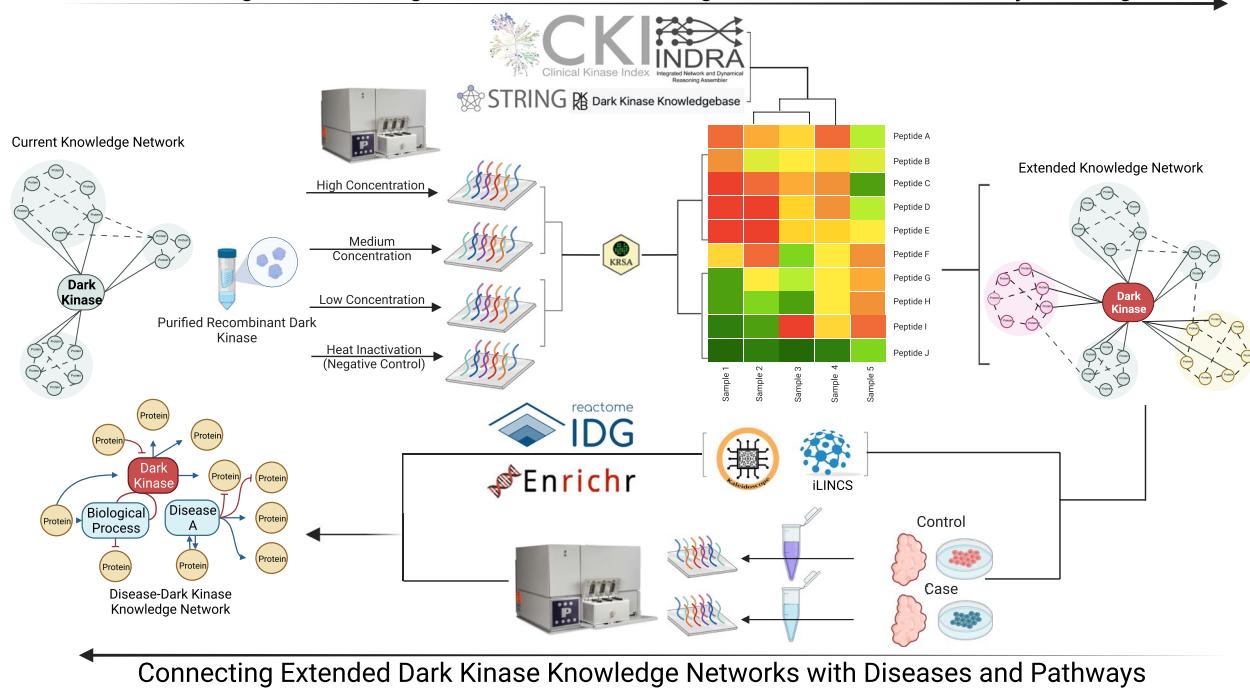


Fig. 1 Extending knowledge of understudied kinase by combining empirical and in-silico approaches. An in-vitro/in-silico hybrid workflow extending functional knowledge of understudied or “dark” kinases was developed. To empirically identify novel kinase substrates, the PamStation12 platform was utilized to profile purified recombinant kinases. In-silico databases and tools were used to analyze and visualize the graph knowledge of dark kinases, as well as to explore connections with diseases and biological pathways. KRSA: Kinome Random Sampling Analyzer, iLINCS: Integrative Library of Integrated Network-Based Cellular Signatures

Tissue Grinders (DWK Life Sciences, NIPPI Inc., Tokyo, Japan) in M-PER Mammalian Protein Extraction Reagent (ThermoFisher, Rockford, IL) and 100X HALT Protease and Phosphatase Inhibitors. After homogenization, samples were spun down at 13xG for 15 min at 4° C. The supernatant fraction was separated from the pellet fraction, stored and aliquoted in 0.5 mL tubes, and stored at -80° C. Protein concentration was determined using the Pierce BCA Assay Kit (ThermoFisher, Rockford, IL). After the protein concentration was determined for each subject, the supernatant fraction was then pooled by sex and diagnosis with equal protein content for each subject. Each of the four pooled samples was diluted to 2.0 ug/uL. Samples were stored at -80° C until the day of the experiment. On the day of the experiment, 10ug of lysates were loaded onto phosphotyrosine (PTK) chips and run in technical triplicate at. The PTK chips have a total of 196 unique peptides (3 controls, 193 reporter peptides); these peptides are standardized on the chips (ie these are not custom arrays) and are commercially available from the Pamgene Corporation. Prediction algorithms suggests that the PTK chip covers about 96% of protein tyrosine kinases in the human genome [17].

Instruments and experimental conditions

A Pamstation12 (PamGene International B.V., ‘s-Hertogenbosch, North Brabant, The Netherlands) was used to screen recombinant kinases and tissue homogenate on a PTK PamChip (catalog number 86402, 4 arrays/chip). Each array contains 196 immobilized peptides from which phosphorylated tyrosine residues are used as a read-out of kinase activity. The PamChips are blocked with 2% BSA for 30 cycles followed washing of the arrays using PK buffer. The PTK assay master mix consists of PK buffer, 100×BSA, DTT, 10×PTK additive, 4 mM ATP and a phospho-tyrosine antibody. The total protein used for tissue homogenate was 10 ug/array and a concentration-response was used to determine adequate total protein for recombinant kinases. For our human studies, samples were pooled by diagnosis and sex. Pooled samples were run in technical triplicate to account for variability. After the sample is loaded a CCD camera takes a picture every five cycles for 60 cycles with images taken 2 cycles before and after the real-time kinetics determination.

Immunoblotting

30 μ L of EPHA6 (250 ng, 250 ng heat inactivated) and GAB1 (0 ng, 200 ng, 300 ng, 400 ng) were taken from a 96-well plate (Corning) used to co-incubate the proteins at 37 °C and prepared for SDS-PAGE using 6 μ L of Laemmli SDS sample buffer (Boston BioProducts). The protein samples were run on a 1.0 mm thick, 4–12% Bis-Tris gradient gel (Invitrogen) and transferred to a PVDF membrane using a semi-dry transfer cell (Bio-Rad). The membrane was blocked using TBS Odyssey Blocking Buffer (Li-Cor) for 20 min at room temperature, then probed overnight at 4 °C with a primary pan-p-Tyr antibody (1:250, Cell Signaling, catalogue # 9411S). Next, the membrane was incubated for 1 h at room temperature with an anti-mouse secondary antibody (1:5000, Li-Cor, 926–68,070). Following secondary incubation, the membrane was scanned using Odyssey Infrared Imaging System (Li-Cor). Additional samples (1.25ug EPHA6, 2ug GAB1) were sent to the University of Cincinnati Core Research Facility for Mass Spectrometry (Supplementary Material 12).

Data analysis

Kinome array profiling

The raw PamChip images are pre-processed using Bio-Navigator (PamGene International B.V.) to generate numerical values of the signal intensity for each reporter peptide on the chip. To analyze the kinome array profiles, the Kinome Random Sampling Analyzer (KRSA) R package was used to pre-process, apply quality control checks, and select differentially phosphorylated peptides [18]. KRSA was used to analyze the kinome profiles of postmortem dorsolateral prefrontal cortex from control and schizophrenia subjects. The analysis starts with calculating the linear regression slope (signal intensity as a function of camera exposure time) followed by scaling the signal by multiplying by 100 and then log transforming the scaled values. The derived values from the previous steps are used as the final signal (i.e., peptide phosphorylation intensity) in the comparative analyses [19]. Additional quality control filtration steps were carried out to remove peptides with either very low signal intensity (signal < 5) or R^2 of less than 0.9 in the linear model. The signal ratios between pairs of groups (control versus schizophrenia) were used to calculate log₂ fold change (LFC) for each peptide. The LFC was calculated per chip and by sex, and then averaged across chips. To compare the kinase activity, peptides with an average LFC of at least 0.2 were carried forward to the upstream kinase analysis.

To investigate implicated upstream kinases, three different software packages were used: KRSA, upstream kinase analysis (UKA), and Kinase Enrichment Analysis

(KEA3). To look at associated upstream kinase families, KRSA takes the list of differentially phosphorylated peptides and uses a random resampling approach to assign scores for each kinase family [19]. The Upstream Kinase Analysis (UKA) tool from BioNavigator was used to look at individual upstream kinases. The default settings of the standard STK analysis protocol were used with the additional step of upstream kinase analysis. UKA reports the final score as a metric for ranking implicated kinases, which is calculated based on the specificity of the peptides mapped to the kinases and the significance of phosphorylation changes of the peptides between the compared groups. The Kinase Enrichment Analysis Version 3 (KEA3) web tool was also used to perform kinase set enrichment analysis using the corresponding proteins of the top differentially phosphorylated reporter peptides as the input [20].

We used the Creedenzymatic R package to aggregate the results from these three different analytic tools. The Creedenzymatic R package is a pipeline software package that combines, scores, and visualizes the results from multiple upstream kinase analytic tools (<https://github.com/CogDisResLab/creedenzymatic>) [3]. The Creedenzymatic package harmonizes the different metrics used in KRSA, UKA, and KEA3 with percentile rank normalization. This harmonization approach results in a standardized percentile score for each kinase under each tool. Next, the mean and median percentile score for each kinase is calculated by averaging the normalized scores across the three analytic tools. Additionally, kinases are mapped to the official HUGO Gene Nomenclature Committee (HGNC) symbols and subfamilies, ensuring the naming convention is consistent across the three different tools.

For the recombinant kinase screening, semi-supervised clustering was used to categorize peptides into four different groups. An unsupervised hierarchical clustering was first performed followed by extracting the top four clusters. The first cluster, titled “No Affinity,” is a list of peptides that show no phosphorylation activity across all tested protein concentrations. The remaining peptides are clustered under three groups based on their dose-response signal activity: low, medium, and high affinity. The “High Affinity” group is a list of reporter peptides that show phosphorylation activity at the lowest protein concentration. “Medium Affinity” are peptides that only start showing phosphorylation activity at the medium protein concentration. Finally, the “Low Affinity” group is a list of peptides that only show activity at the highest protein concentration.

Principal component analysis (PCA) was also performed to visualize the clusters of peptides and determine the component loading for the first four principal

components (PC1, PC2, PC3, and PC4). To examine the consensus of the amino acid sequence for each cluster of peptides, we first extracted the main tyrosine phosphosite and the ten neighboring amino acids (five amino acids from each side) of all reporter peptides on the PTK PamChip. We then used the ggseqlogo R package to plot the amino acid sequence logos of the three clusters of peptides and for each recombinant kinase screening profile [21].

To investigate the association of the recombinant kinase profiles with the schizophrenia genome dataset, we performed a peptide set enrichment analysis using the tmod R package [22, 23]. Using the average peptide LFC, by comparing control to schizophrenia subjects, we ranked all the peptides that passed quality controls filtration steps based on the absolute values of LFCs. Next, we used the *makeTmod* function from the tmod package to create peptide *modules* based on the five kinase screening profiles and using the three main clusters (low, medium, and high affinity). We also generated combined peptide lists by combining either all peptides from the three clusters or just the medium and high affinity peptides. The combined peptides lists are used to expand the list of peptides for the more selective kinases, mainly INSRR and LTK. To run the enrichment analysis, we utilized the *tmodUtest* function to perform a Mann–Whitney U test (U-test) on the ranks of peptides belonging to the defined clusters and the peptides that do not belong to the peptides sets. We used a modified version of the *evidencePlot* function to visualize the results and highlight the significant enriched peptide sets ($p\text{-value} < 0.05$ and area under the curve (AUC) > 0.5). Additionally, we used a waterfall plot to visualize the peptide LFCs from the schizophrenia genome dataset.

Pathway Analysis and In-silico Data Exploration

The Enrichr web tool was used to perform multiple gene set enrichment analyses across different gene set libraries [24]. The peptides were first mapped to their corresponding proteins, then to the official HGNC gene symbols. The gene set library used in the pathway analysis is the Gene Ontology (GO) Biological Process 2021. The ARCHS4 tool was also used to visualize gene expression levels across various tissues. The Cell Types Database from the Allen Brain Map was used to extract cell-level gene expression in multiple brain regions [25]. The Human Protein Atlas (HPA) web portal was utilized to explore protein expression across tissues, cluster genes based on co-expression, and determine subcellular localization patterns [26, 27]. The web tool Kaleidoscope was used to query multiple curated databases such as STRING, Genotype-Tissue Expression (GTEx), Brain RNASeq, Integrative Library of Integrated Network-Based Cellular

Signatures (iLINCS), and Pharos from the Illuminating the Druggable Genome (IDG) program [28–34].

Results

Identifying dark tyrosine protein kinases

Utilizing the knowledge base INDRA database to assign normalized knowledge scores for all protein kinases, we identified 19 understudied or “dark” protein tyrosine kinases (Supplementary Fig. 1A). Out of these 19 “dark” kinases, we selected five kinases to screen using a multiplex functional kinase activity profiling technology (PamStation12 platform). The kinases we profiled include EPHA6, AATK, INSRR, LTK, and TNK1. The normalized ranks for these kinases range from 0.20 to 0.28, where 0 represents the most understudied kinase and 1 represents the most studied kinase (Supplementary Fig. 1B).

Functional profiling of dark tyrosine protein kinases

Screening purified recombinant “dark” tyrosine protein kinases under four conditions (low, medium, high protein concentrations, and heat inactivation as negative control) revealed a set of clusters of concentration-dependent reporter peptides. Based on concentration–response patterns and a semi-supervised clustering approach, we grouped peptides into four different categories: zero or no affinity, low, medium, and high affinity (Supplementary Table 2). This analysis identified 18 high affinity peptides, 31 medium affinity peptides, and 4 low affinity peptides for EPHA6 (Fig. 2B). This analysis identified 26, 9, 3 and 3 high affinity peptides, 31, 46, 4 and 1 medium affinity peptides, and 6, 10, 8 and 1 low affinity peptides for TNK1, AATK, INSRR and LTK respectively (Supplementary Fig. 2A, 2G, 2M, 2S). Signal intensity of phosphosites with increasing total protein (2.5 ng – 25 ng) of recombinant protein shows concentration specific increases in reporter peptide phosphorylation (Fig. 2C, Supplementary Fig. 2D, 2J, 2P, 2V).

Chip coverage was calculated by dividing the number of peptide hits by the total number of peptides present on the PTK PamChip (193 peptides). We observed 34%, 27%, 8%, 3%, and 33% chip coverage for AATK, EPHA6, INSRR, LTK, and TNK1 kinases, respectively (Fig. 3A). Next, we examined the intersection of peptide hits across all of the kinase profiles and visualized the overlap as an UpSet plot (Fig. 3B) [35]. The chip coverage and peptide overlap analysis reveals the degree of selectivity of our list of kinases to the reporter peptides printed on the PTK PamChip. For example, the relatively low chip coverage for INSRR and LTK suggests that these kinases are more selective than AATK, EPHA6, and TNK1. Additionally, peptide overlap analysis revealed which reporter peptides are uniquely phosphorylated by the recombinant dark protein kinases. For instance, the peptide overlap

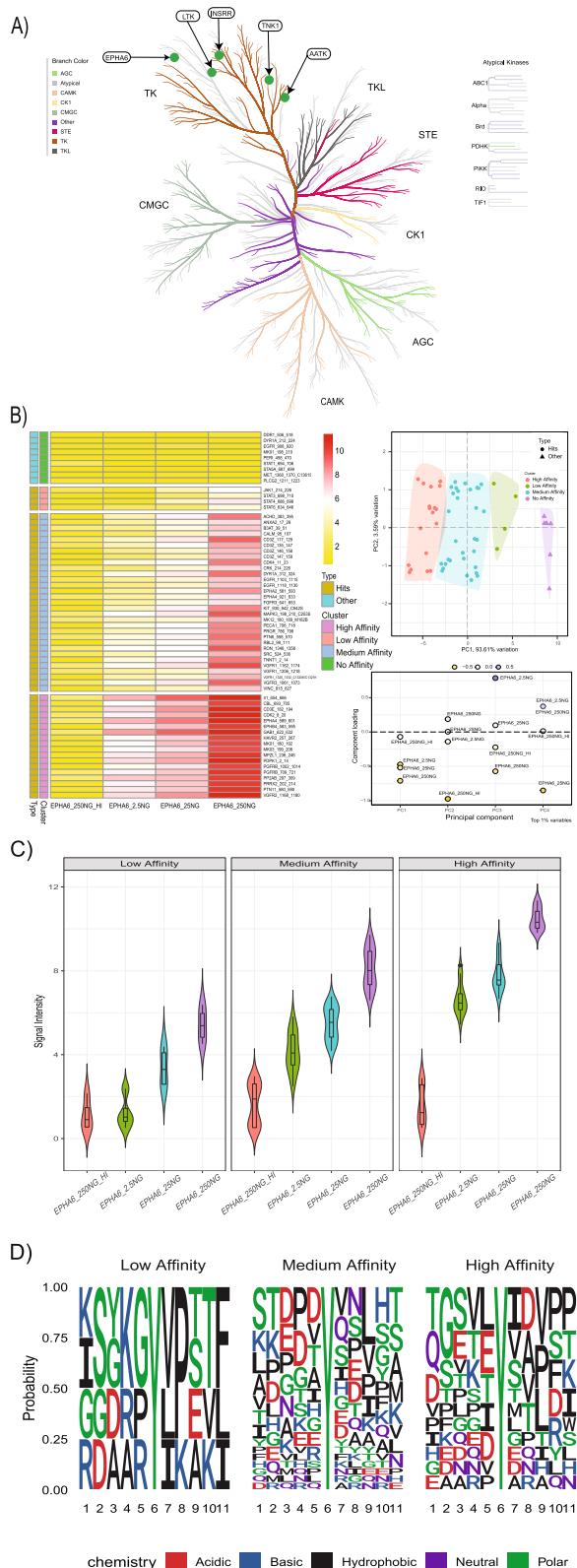


Fig. 2 PamStation12 phospho-tyrosine profiles of purified recombinant dark protein kinases identify novel substrates. **A** Highlighting the five dark kinases (LTK, TNK1, INSRR, AATK, EPHA6) selected for profiling on a kinase phylogenetic tree. **B** 2.5 ng, 25 ng, 250 ng of EPHA6 total protein was screened along with a 250 ng EPHA6 heat inactivated (denatured protein) control. Red indicates increased phosphorylation and yellow indicates less phosphorylation at that peptide. A semi-supervised clustering and principal component analysis (PCA) is visualized for PamChip peptides reporting on EPHA6. **C** Signal intensity plot illustrating increased activity with increased total protein of EPHA6. **D** Peptide sequence logos for EPHA6 peptides identified on the phosphor-tyrosine (PTK) Pamchip. The color of the single letter amino acid code denotes the chemical features of each amino acid map against relative position. High Affinity: list of peptides that show phosphorylation activity at the lowest protein concentration. Medium Affinity: list of peptides that only start showing phosphorylation activity at the medium protein concentration. Low Affinity: a list of peptides that only show phosphorylation activity at the highest protein concentration. No Affinity: a list of peptides that show no phosphorylation activity across all tested protein concentrations. Ephrin Type-A Receptor 6 (EPHA6)

analysis shows that the majority of the peptide hits of AATK are exclusively phosphorylated by AATK and not by the other kinases. On the other hand, the majority of peptides hits of EPHA6 and TNK1 are shared among the two kinases.

Next, we examined the consensus of the amino acid sequence for each cluster of peptides by extracting the main tyrosine phosphosite and the ten neighboring amino acids (five amino acids from each side) of all peptide hits. We visualized the amino acid sequence logos of the three clusters of peptides for each recombinant kinase screening profile (Fig. 2C, Supplementary Fig. 2E, 2K, 2Q, 2W). Pathway analyses were performed using the list of peptide hits for each recombinant kinase screening profile using Enrichr. Using the Gene Ontology (GO) Biological Process 2021 gene set library, we extracted and visualized the top ten pathways (Fig. 2D, Supplementary Fig. 2F, 2L, 2Q, 2X).

In-silico data exploration of dark kinases

Known and predicted functional pathways

We explored several publicly available biological databases to examine the current functional knowledge of our list of kinases in terms of their role as enzymes or substrates, tissue-specific and brain-specific expression at the mRNA and protein levels, known and predicted pathways, subcellular localization, and protein–protein interaction (PPI) networks.

To investigate known downstream substrates for kinases, we queried the iPTMnet database [36]. Unsurprisingly, there are currently no annotated downstream substrates for our list of kinases (Supplementary Table 3).

Next, we examined gene expression of our list of targets at the mRNA and protein level using the GTEx and HPA databases across different tissues and specifically in the brain. The results from these databases indicate that AATK, INSRR, and EPHA6 are highly enriched in the brain. Additionally, the HPA database shows INSRR is enriched in the kidney as well. Interestingly, TNK1 shows low brain expression at the mRNA level but high abundance at the protein level. LTK shows overall low expression across all tissues with the highest being in the lung and intestine (Supplementary Table 4).

In terms of subcellular localization, the HPA database also contains immunofluorescent staining images of human cell lines aiming to fully annotate the subcellular localization of the human proteome. Examining the subcellular localization of our selected dark protein kinases reveals differential localization patterns. AATK is localized to mitochondria, EPHA6 to the nucleoplasm, LTK to vesicles such as endosomes and lysosomes, and TNK1 to cell junctions. No data was available for INSRR.

Next, we explored known and predicted functional pathways of our selected dark protein kinases. For known pathways, we utilized Enrichr and the Gene Ontology (GO) Biological Process 2021 gene set library using the HGNC symbol of the kinase as the single input. The results from this analysis showed several brain-specific pathways for some of our kinases. For example, AATK is involved in brain development (GO:0007420) and central nervous system development (GO:0007417). EPHA6 is involved in axon guidance (GO:0007411) and axonogenesis (GO:0007409). LTK is involved in regulation of neuron differentiation (GO:0045664) and regulation of neuron projection development (GO:0010976 and GO:0010975). INSRR and TNK1 have other pathways that include cellular response to pH (GO:0071467) and actin cytoskeleton reorganization (GO:0031532) for INSRR and innate immune response (GO:0045087) for TNK1 (Supplementary Table 5).

Given that the functional knowledge of understudied kinases is relatively low, it is more suitable to examine predicted functional annotations. To investigate predicted functional pathways, we used three approaches: co-expression clustering, protein–protein interaction networks, and genetic perturbation analysis. The co-expression clustering analysis examines the top pathways

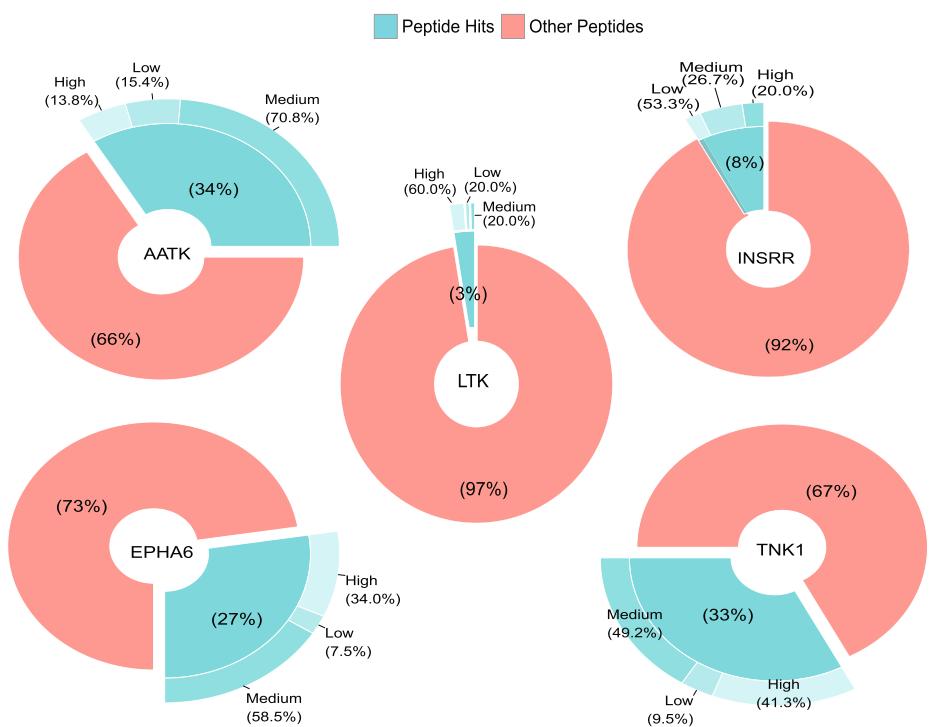
of other well-studied genes that are highly co-expressed with our list of kinases. We utilized the co-expression clustering analyses deployed by the HPA and ARCHS web tools to associate predicted pathways with our targets. The HPA tissue and single-cell co-expression analysis for AATK and INSRR revealed that these protein kinases are co-expressed with genes that are functionally involved in central nervous system myelination, axon guidance, synaptic transmission, and microtubule cytoskeleton organization. Additionally, ARCHS shows central nervous system myelination (GO:0022010) as one of the top predicted pathways for AATK. Other predicted pathways for AATK include regulation of short-term neuronal synaptic plasticity (GO:0048172), synaptic vesicle maturation and exocytosis (GO:0016188, GO:0016079, and GO:2,000,300), and neuronal action potential propagation (GO:0019227 and GO:2,000,463). The predicted pathways for INSRR in ARCHS include Golgi transport vesicle coating (GO:0048200), insulin secretion (GO:0030073), and protein maturation by protein folding (GO:0022417). EPHA6 is part of the “Brain – Ion transport” cluster in the tissue expression clustering analysis in HPA and the “Neurons & Oligodendrocytes – Synaptic function” cluster in the single-cell co-expression analysis. These gene clusters also include other brain-specific pathways such as synaptic transmission, AMPA receptor activity, neuronal action potential, pre-synaptic membrane assembly, glutamate secretion, and regulation of NMDA receptor activity. ARCHS reveals similar pathways such as glutamate receptor signaling pathways (GO:0035235, GO:0007215, and GO:1,900,449) and additional pathways including synaptic plasticity (GO:0048172) and protein localization to synapse (GO:0035418). In HPA, LTK is a member of a cluster of genes that are involved in various immune response pathways such as adaptive and innate immune response, complement activation, and interferon-gamma-mediated signaling pathways and protein assembly and transport (Supplementary Table 6).

The second approach to associate functional pathways to our list of dark kinases is protein–protein interaction networks. STRING is a curated database of known and predicted protein–protein associations based on multiple resources including gene co-expression, experimental data, and text mining. We used STRING to extract the

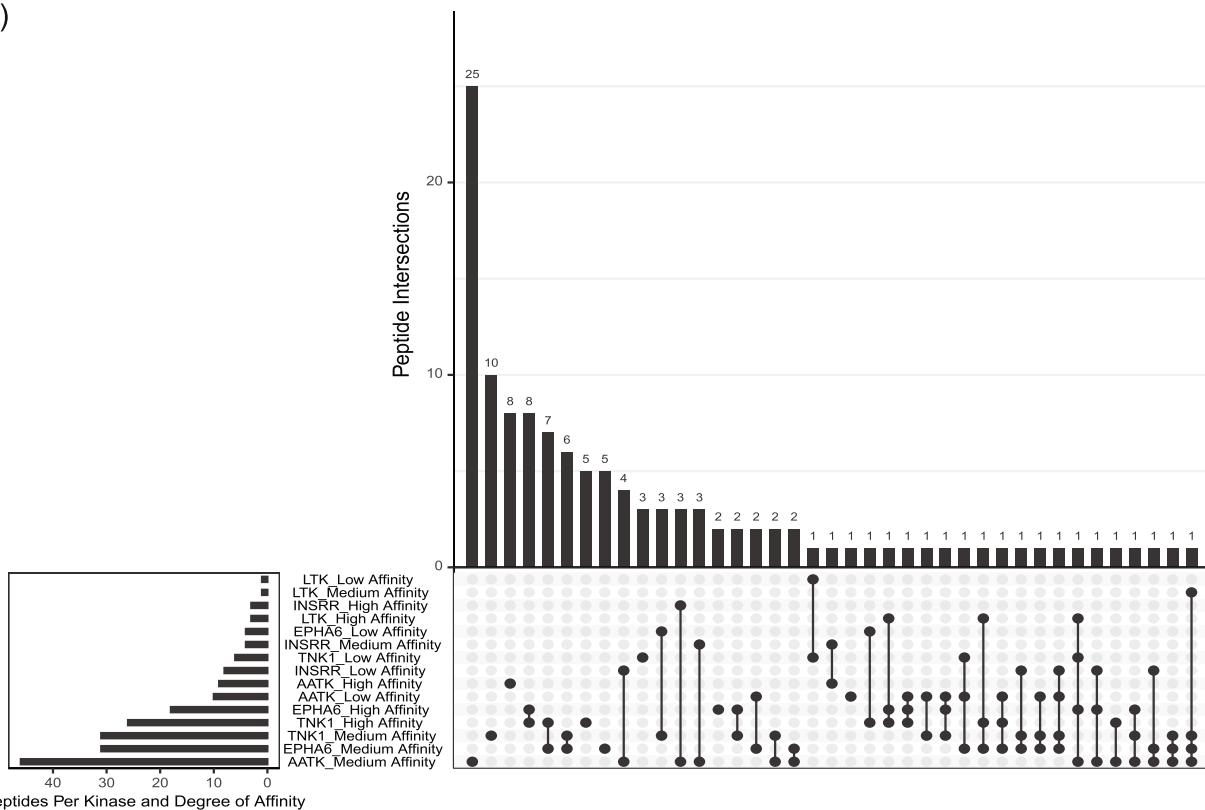
(See figure on next page.)

Fig. 3 Chip coverage and peptide overlap analyses reveal differential kinase-substrate selectivity. **A** Pie charts of chip coverage of AATK, EPHA6, INSRR, LTK, and TNK1. The chip coverage is calculated by dividing the number of peptide hits for each kinase by the total number of peptides present on the protein tyrosine kinase (PTK) PamChip (193 peptides). “Low” denotes low affinity peptides, “Medium” denotes medium affinity peptides, “High” denotes high affinity peptides. **B** UpSet plot showing the overlap of peptide hits across all recombinant kinase profiles and peptide clusters

A)



B)

**Fig. 3** (See legend on previous page.)

top 25 protein interactors with each of our list of kinases, using the default minimum required interaction score (medium confidence=0.400) as the threshold. Next, we used the built-in functional network enrichment analysis in the STRING web tool. Using the previously described parameters, the AATK network had 24 interactors and only three enriched GO (Molecular Function) terms, including Ser-tRNA (Ala) hydrolase activity, alanine-tRNA ligase activity, and Rab guanyl-nucleotide exchange factor activity. EPHA6 network has 24 interactors and 317 enriched GO terms (Biological Process). The top enriched terms for the EPHA6 network include ephrin receptor signaling pathway, axon guidance, and neuron projection development and morphogenesis. The INSRR network contains 24 interactors and 252 enriched GO terms. The top enriched pathways include regulation of protein kinase B signaling, insulin receptor signaling pathway, phosphatidylinositol 3-kinase signaling, and MAPK cascade. There are 12 protein interactors in the LTK network and 85 enriched pathways. The top enriched terms include epidermal growth factor receptor signaling pathway, ERBB2 signaling pathway, generation of neurons, and insulin receptor signaling pathway. Finally, the TNK1 network has 12 interactors and only three enriched GO terms, including regulation of metalloendopeptidase and aspartic-type endopeptidase activity involved in amyloid precursor protein catabolic process (Supplementary Table 7).

To explore the results of our GO analyses, we performed “lookup” validation literature searches for select kinases and associated GO pathways (Supplementary Table 8). Notably, we found that the GO pathways axon guidance (GO:0007411) and innate immune response (GO:0045087) were associated with EPHA6 and TNK1 kinases, respectively. EPHA6 was highly expressed and localized to ganglion cells of the developing human retina. Patterns of EPHA6 expression in the macaque retina related to fovea development and ganglion cell projections suggest a role in the postnatal maintenance of neuronal projections [37]. A group investigating the impact of EPHA6 deletion on neuronal cell morphology in LacZ/LacZ mice discovered extensive impairments in the structure and function of both cells in the brain and spinal cord. These impairments include defects in memory and learning, changes in cellular morphology upon golgi staining at 2 months of age presenting aggregation of cells in the frontal cortical and mid-cortical regions [38]. Similarly, TNK1 was identified as a unique regulator of the ISG (interferon stimulated genes) pathway of the antiviral innate immune response in a high-throughput, genome wide cDNA screening assay used to identify genes regulating the ISG expression. Functionally, activated IFN-receptor complex recruits TNK1 from the cytoplasm,

where upon phosphorylation it potentiates the JAK STAT signaling pathway. The authors noted there was a change in the phosphorylation state of STAT1 at two sites by western blot after 24 h of interferon beta exposure: tyrosine 701, and serine 727, however functional characterization of TNK1 activity was not conducted. Interestingly, the phosphosite containing tyrosine 701 maps to our medium affinity peptide logo (Supplementary Fig. 2E) corroborating the findings in our study [39].

The third approach to assess functional annotation of understudied kinases is with controlled genetic manipulation. The LINCS database hosts over a million transcriptional signatures of various cell lines with either genetic or chemical perturbations. We queried the iLINCs web portal to find and analyze gene knockdown or overexpression signatures of our list of dark kinases. Four of our five dark kinases have at least one gene knockdown signature, while TNK1 only has a gene overexpression signature. To keep the comparisons consistent, we selected the VCAP cell line as the primary cell line for the connectivity analysis because it was the only cell line that has gene knockdown signatures for all four kinases (AATK, EPHA6, INSRR, and LTK). To examine and functionally annotate the transcriptional “echo” of knocking down these kinases, we extracted the top differentially expressed genes ($p\text{-value} < 0.05$ and \log_2 fold change > 0.5 or < -0.5) from the LINCS knockdown signatures and performed gene set enrichment analysis using Enrichr. Using methods described in Sect. “[Data analysis](#)” of the *dark kinome supplement*, we selected top differentially expressed genes and extracted 99 genes from the AATK knockdown signature (LINC-SKD_29946). Sorting by Enrichr’s Combined Score, the top associated pathways of this list of genes include regulation of astrocyte differentiation (GO:0048711, GO:0048710), regulation of insulin receptor signaling pathway (GO:0046627, GO:1,900,077), and regulation of extrinsic apoptotic signaling pathway via death domain receptors (GO:1,902,041). We extracted 113 differentially expressed genes from the EPHA6 knockdown signature (LINC-SKD_30785). The top implicated pathways for these genes include protein insertion into mitochondrial membrane (GO:0001844, GO:0097345, and GO:0051204), regulation of inositol trisphosphate biosynthetic process (GO:0032960), and glutamate receptor signaling pathway (GO:0035235). For the INSRR knockdown signature (LINC-SKD_31354), we extracted 98 differentially expressed genes and revealed pathways involved in regulation of hormone metabolic process (GO:0032352), sequestering of NF-kappaB (GO:0007253), and regulation of glial cell proliferation (GO:0060251). We extracted 100 differentially expressed genes from the LTK knockdown signature

(LINCSKD_31563). The top implicated pathways for this list of genes include DNA ligation (GO:0006266), central nervous system projection neuron axonogenesis (GO:0021952), and acetyl-CoA metabolic process (GO:0006084). For the TNK1 overexpression signature (LINCSOE_9396), we extracted 278 differentially expressed genes and revealed pathways involved in regulation of protein import (GO:1,904,589), axo-dendritic transport (GO:0008088), and glycolytic process (GO:0006096) (Supplementary Table 8).

Connectivity analysis with psychiatric and neurogenerative diseases

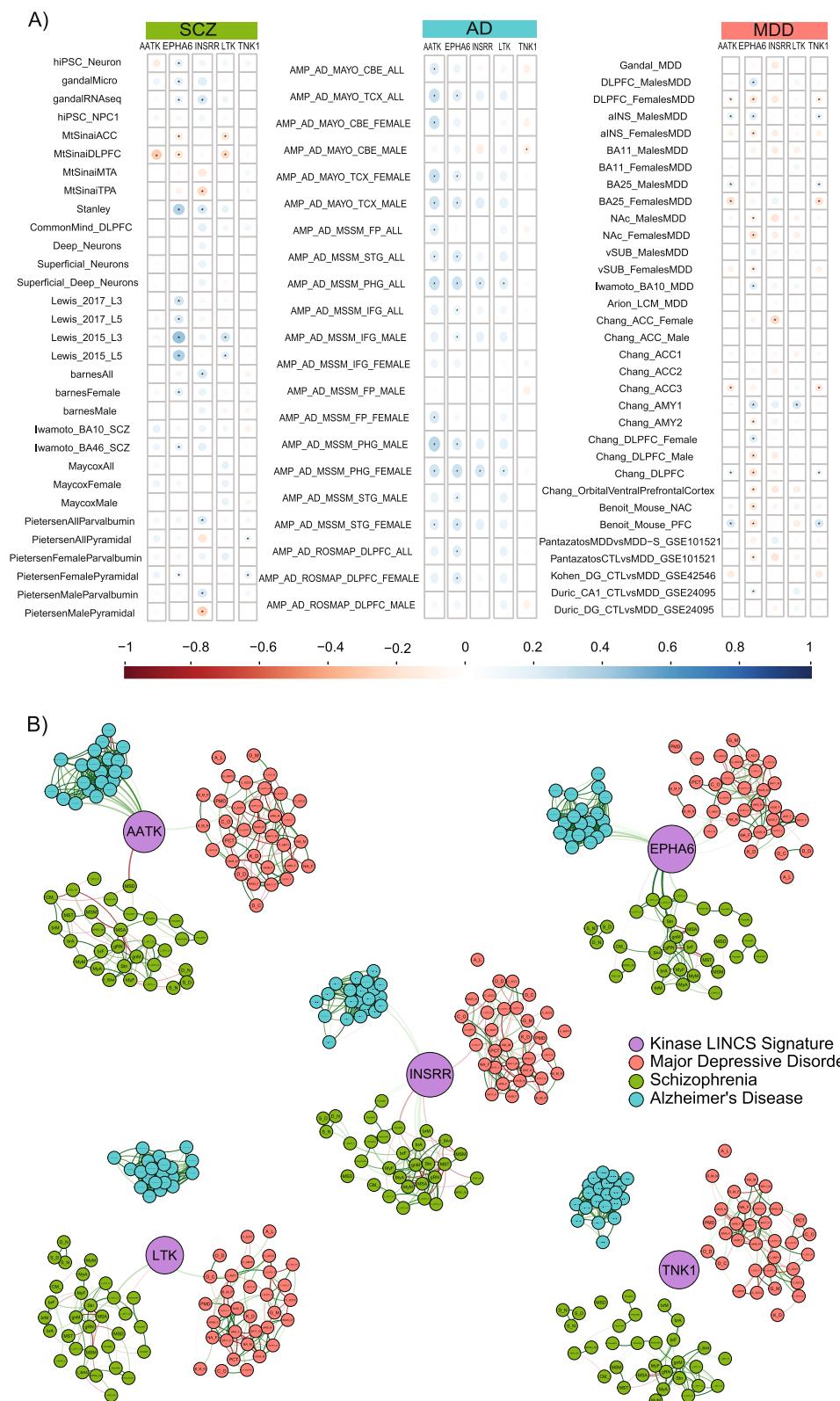
To elucidate the association of our list of dark kinases with diseases, more specifically with psychiatric and neurogenerative disorders, we performed a transcriptional connectivity analysis between gene perturbation signatures (from LINCS) and previously published disease signatures from schizophrenia (SCZ), major depressive disorder (MDD), and Alzheimer's Disease (AD) datasets. Utilizing the disease signatures available in Kaleidoscope, we performed a Pearson correlation analysis to calculate the concordance scores between the LINCS gene perturbation and disease signatures using the top differentially expressed genes in each kinase knockdown or overexpression signature (Fig. 4A). Each node represents a transcriptional signature, and the edges are the Pearson correlation coefficients (Fig. 4B). Overall, the EPHA6 gene signature showed the most concordance with the disease groups, most notably with the SCZ and AD datasets. Some kinases showed relatively higher correlation with one disease group compared to the other diseases. For example, the AATK knockdown signature showed similarity with only the AD datasets. Additionally, we observed that the INSRR gene signature showed exclusively higher concordance with the SCZ datasets. The LTK knockdown signature showed the least amount of connectivity with the disease groups. The TNK1 overexpression signature also had low connectivity, especially with the SCZ and AD datasets.

Profiling the activity of the tyrosine subkinome in schizophrenia

Utilizing the KRSA R package [19], the kinome array profiles of postmortem dorsolateral prefrontal cortex from pooled samples of control and SCZ subjects were fully analyzed starting from preprocessing and quality control filtering to clustering and extracting differentially phosphorylated peptides. To account for the technical variability between chips and biological variation between male and female subjects, the analysis was performed per chip and separated by sex. To determine the global phosphorylation signals in the control and SCZ subjects, we used KRSA to generate global heatmaps with unsupervised hierarchical clustering of pooled samples run in technical triplicate (Fig. 5A, D). To better highlight the differences between the two groups, the values are scaled by row/peptide (Z score transformation). The clustering analysis shows a clear difference between the control and SCZ subjects. Overall, we observed a strong reduction of global phosphorylation in both male and female subjects. To investigate implicated upstream kinases, we used KRSA, UKA, and KEA3 to perform upstream kinase enrichment analysis and the Creedenzymatic R package to aggregate and harmonize the results from each analytic tool. We used the quartile plot function from the Creedenzymatic R package to plot the top associated kinases that explain the differences between the control and SCZ subjects (Fig. 5C, F). The upstream kinases analysis revealed a list of differentially active kinases in SCZ. Moreover, we observed a list of kinases that were differentially active in only the male subjects, only female subjects, and in both male and female subjects. For example, ABL1, DDR2, FYN, LCK, SYK, and NTRK1 kinases were differentially active in male and female SCZ subjects. TYRO3, IGF1R, INSR, MET, PDGFRA, HCK, and SRC kinases were only differentially active in the female subjects, whereas ALK, MERTK, PTK2, PTK2B, JAK2, and FLT3 kinases were only differentially active in the male subjects. Next, we used KinMap to plot the top implicated upstream kinases on the kinome phylogenetic tree (Fig. 5B, E) [40].

(See figure on next page.)

Fig. 4 Transcriptional connectivity analysis explores concordance between AATK, EPHA6, TNK1, LTK, INSRR gene knockdown signatures and schizophrenia and Alzheimer's Disease. **A** Pearson correlation analysis of LINCS knockdown signatures of AATK, EPHA6, INSRR, LTK, and TNK1 against 31 schizophrenia (SCZ), 21 Alzheimer's Disease (AD), and 33 major depressive disorder (MDD) datasets extracted from the Kaleidoscope web application. The values represent correlation coefficients and color represent the sign of the coefficients (blue: positive correlation, red: negative correlation). The “**” within each node denotes the significance of the correlation analysis where the corrected p-value was under 0.05. **B** The connectivity analysis visualized as networks where each node represents a transcriptional signature and edges represent the correlation coefficients between signatures. The central node in each network represents the gene knockdown or overexpression signature of the corresponding kinase (AATK, EPHA6, INSRR, LTK, and TNK1) retrieved from iLINCS

**Fig. 4** (See legend on previous page.)

Connecting the recombinant kinase profiles with the schizophrenia active kinase

To connect the recombinant kinase profiles of our dark kinases with the SCZ kinase dataset, we performed a peptide set enrichment analysis using the tmod R package (Supplementary Fig. 3). Similar to the upstream kinase analysis, the recombinant kinase profile enrichment analysis also revealed sex differences in peptide set enrichment scores and directionality of activity. The AATK, EPHA6, INSRR, LTK and TNK1 peptide sets were enriched in the female subjects and mostly show decreased activity, whereas the same peptides sets were not enriched in the male kinase profiles (Supplementary Fig. 3A-B).

Downstream target validation

To validate EPHA6 phosphorylation of novel substrates identified on the array, kinase array experimental conditions were replicated using a benchtop assay with recombinant EPHA6 and one of the “hits” for this dark protein kinase from the PTK array, Growth Factor Receptor Bound Protein 2-Associated Protein 1 (GAB1). Following incubation with an in vitro reaction assay, samples were run on an SDS-PAGE gel and sent for mass spectrometry (Fig. 6A). A pan-phospho-tyrosine antibody was used to identify autophosphorylation sites for EPHA6 as well as novel and predicted phosphorylation sites on GAB1. Interestingly, a p-Tyr band was observed at the expected GAB1 molecular weight only with active EPHA6 and not in the heat inactivated EPHA6 control. Additionally, a p-Tyr band was observed with active EPHA6 alone and wasn’t observed in the heat inactivated control suggesting EPHA6 autophosphorylation at a Tyrosine residue (Fig. 6B).

Mass spectrometry of co-incubated recombinant proteins revealed phosphorylation of GAB1 and EPHA6 at several tyrosine residues. Tyrosine residue pY627 was initially identified as a PTK array reporter peptide (from GAB1) for recombinant EPHA6 on the array (Fig. 2B). This site and two others were identified on recombinant GAB1 via mass spectrometry. Additionally, two putative autophosphorylation sites on EPHA6 were identified at pY259 and pY484 (Fig. 6C, D). Autophosphorylation of

EPHA6 occurred at the ATP binding N-lobe between the EPH ligand binding domain and the fibronectin type III conserved regions. GAB1 was phosphorylated mostly at the C-terminus, along several intrinsically disordered regions suggesting changes in localization and protein–protein interactions (Fig. 6E). Changes in GAB1 function and localization are yet to be determined.

Discussion

In this study, we used a hybrid *in silico* and *in vitro* approach to extend the functional knowledge of five understudied (i.e., dark) tyrosine kinases (AATK, EPHA6, INSRR, LTK, and TNK1) and their associations with schizophrenia (SCZ), Alzheimer’s dementia (AD), and major depressive disorder (MDD). Our hybrid approach to identifying and characterizing novel substrates for dark kinases is unique compared to traditional methods. In contrast to traditional approaches, we used a multiplex platform that allows for a large number of substrates to be examined at the same time and we identified peptide substrates based on kinase activity profiles. Most approaches to kinase substrate characterization require prior selection of a kinase of interest. This is followed by knockdown, overexpression, and/or drug inhibition studies of the targeted kinase with biochemical assays and/or mass spectrometry. While these approaches are tried and true they may be difficult to scale up to the subkinome or whole kinase level. Further, with mass spectrometry-based phospho-peptide detection, low abundance may mask subtle changes in peptide abundance, making inference of kinase activity based on this technique challenging. Our kinase array profiling platform avoids these potential limitations, directly measuring kinase activity, rather than peptide abundance.

In silico gene expression enrichment and pathway prediction analyses revealed that AATK, INSRR, and EPHA6 have high expression levels in the brain and are associated with important CNS biological pathways such as neurodevelopment and axon guidance. Transcriptional connectivity analyses show that the EPHA6 gene signature has the most concordance with the disease groups (SCZ, AD, and MDD). EPHA6 is a member of the ephrin receptor family which is the largest family

(See figure on next page.)

Fig. 5 Sex specific kinase perturbation in the DLPFC of schizophrenia subjects. **A-C** Pooled technical triplicate females, **D-F** Pooled technical triplicate males: Unsupervised clustering of the active kinase profiles of postmortem dorsolateral prefrontal cortex (DLPFC) samples from control and SCZ subjects. Color represents the relative phosphorylation activity at each reporter peptide on the PTK PamChip. Yellow represents relatively lower phosphorylation activity and red indicates relatively higher phosphorylation activity. The phosphorylation signals are scaled by row/peptide (Z score transformation) to better highlight the differences between the two groups. This approach shows the upstream kinase analysis on the kinase phylogenetic tree with an emphasis on the tyrosine kinase subgroup. Creedenzymatic plots featuring the quartile scores of upstream kinases using three analytic tools: Kinome Random Sampling Analyzer (KRSA), upstream kinase analysis (UKA), and Kinase Enrichment Analysis (KEA3)

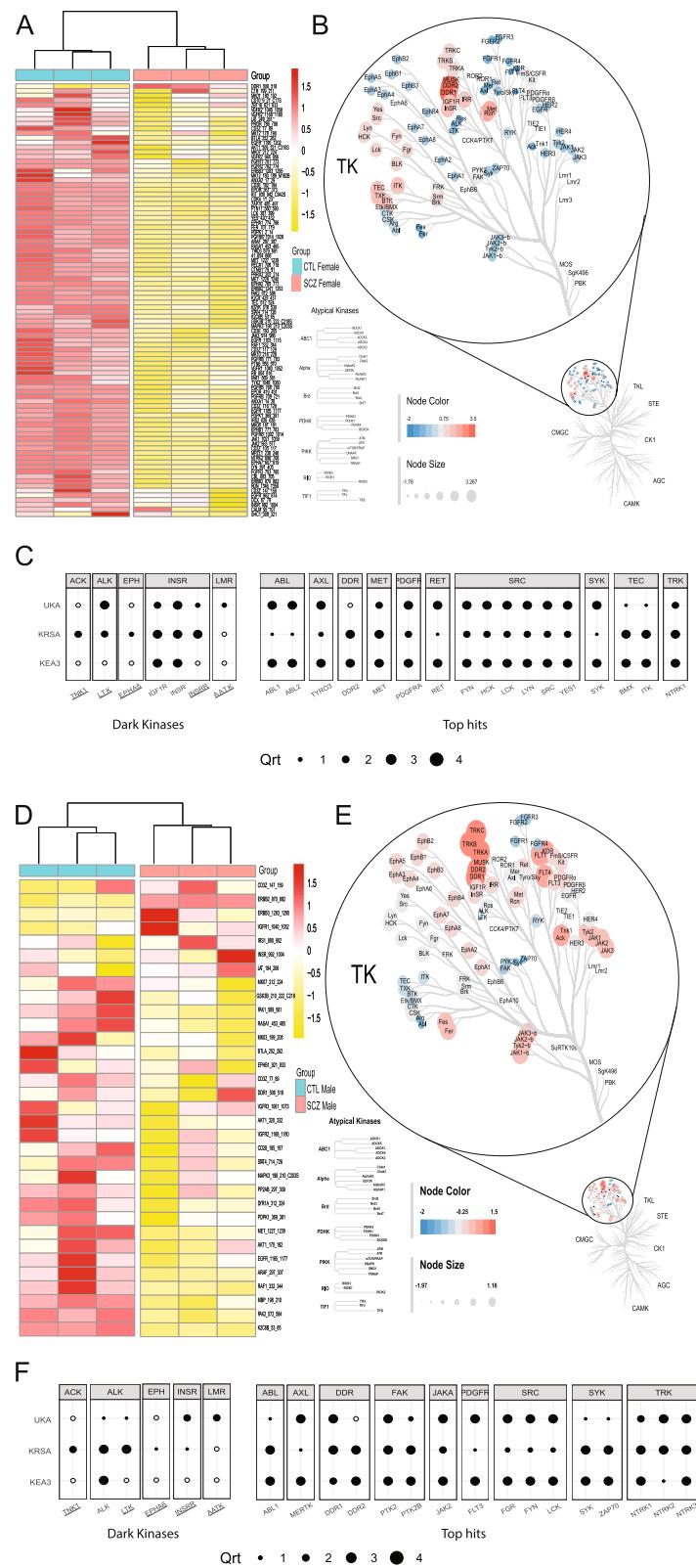


Fig. 5 (See legend on previous page.)

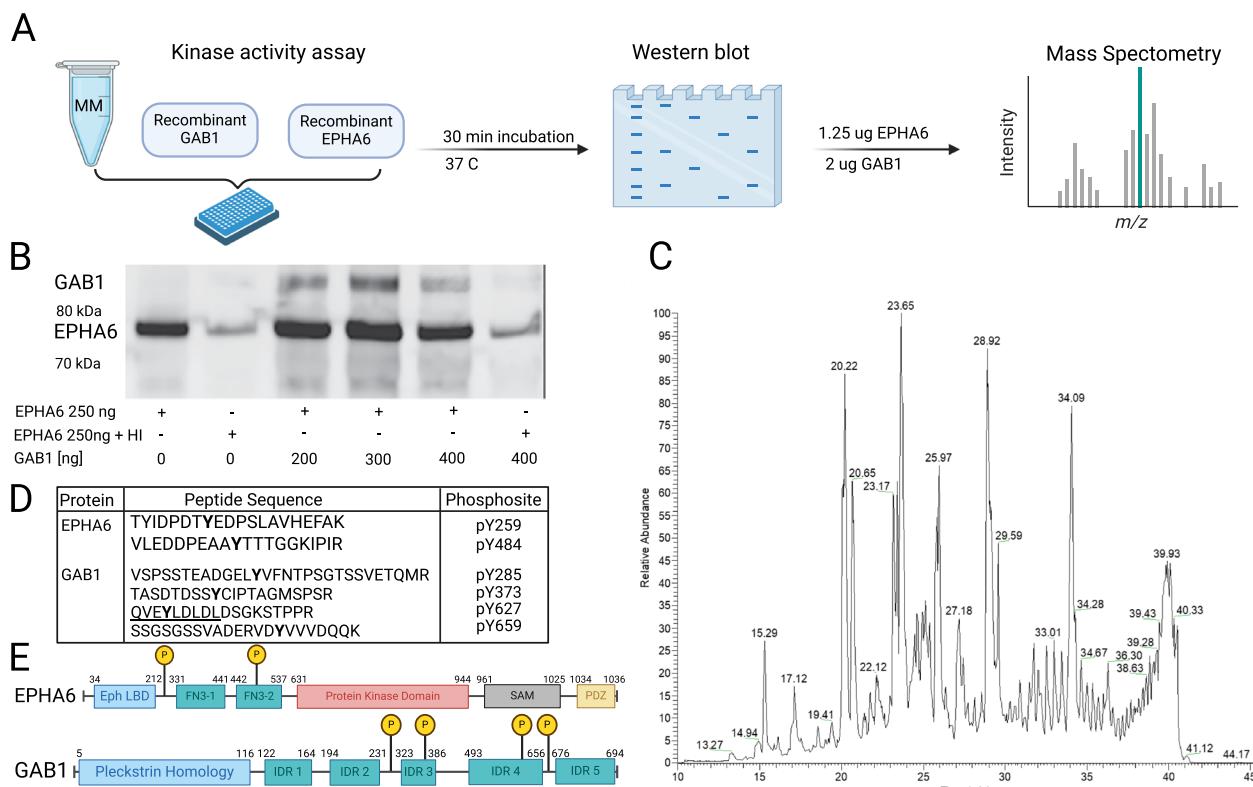


Fig. 6 Downstream Target Validation of EPHA6 predicted phosphosites. **A** Experimental workflow for the benchtop kinase activity assay, western blot validation experiment, and mass spectrometry phospho-site identification. **B** Representative western blot for two recombinant proteins, GAB1 (substrate) and EPHA6 (kinase). All western blot samples were collected from the benchtop activity assay. 250 ng of active or heat inactivated EPHA6 was run on a gel with and without increasing total protein of GAB1 (0 ng, 200 ng, 300 ng, 400 ng). A pan-phospho-tyrosine antibody (1:250, Cell Signaling, catalogue # 9411S) was used to identify any tyrosine site phosphorylated by EPHA6. **C** Relative abundance of peptides following protease digestion of the kinase substrate reaction was analyzed by LC-MSMS over a 60-min gradient. **D** Peptide sequences shown were identified from mass spectrometry and mapped to EPHA6 and GAB1. The underlined sequence corresponding to phosphosite pY627 is present as a substrate on the PTK Pamchip. **E** Placement of GAB1 and EPHA6 phosphosites on full protein sequence domain maps

of receptor tyrosine kinases. The ephrin receptor family is involved in various cellular processes during development and is a key regulator of adult tissue homeostasis [41]. More specifically, pathway analysis of our recombinant kinase screening of EPHA6 is consistent with what is known about its role in axon guidance and development. Moreover, the high concordance of an EPHA6 knockdown signature with the AD datasets complements a known association of EPHA6 with learning and memory [42]. Further, a genome-wide analysis of an early-onset familial Alzheimer's disease (EO-FAD) case identified a pathogenic copy number variation (CNV) located in close proximity to EPHA6 [43]. Finally, axon guidance pathway genes, including EPHA6, are associated with schizophrenia based on genome-wide association studies (GWAS) [44].

Interestingly, some kinases had more connections with one disease group than others. For instance, the AATK knockdown signature showed similarity with only AD

datasets, whereas the INSRR gene signature showed exclusively higher concordance with the SCZ datasets. AATK has recently been identified as part of a novel mechanism controlling dendritic spine formation via endosome trafficking [12]. Two novel potential loci for frontotemporal dementia (FTD), which is the second most prevalent form of early onset dementia after Alzheimer's dementia (AD), exert pathological effects by decreasing expression of AATK [45].

INSRR belongs to the insulin receptor family alongside the insulin and insulin-like growth factor receptors [46]. Impaired glucose metabolism and insulin resistance are two well-known hallmarks of schizophrenia [47]. INSRR functions as an extracellular pH sensor and plays an important role in pH homeostasis [48]. It is well documented that the pH in brains of schizophrenia patients are lower than normal controls and possibly altering dopamine release and uptake at synapses, NMDA receptor activity, and glutamate release [49]. However, the

exact cause for lower pH in schizophrenia still remains unclear [49]. Dysregulation of INSRR in schizophrenia, as we observed in this study, may offer a potential explanation of the impaired pH sensing in this illness.

The co-expression clustering analysis we performed for our list of dark kinases grouped AATK and INSRR in clusters of genes that are functionally involved in central nervous system myelination, axon guidance, and synaptic transmission. All of these biological processes have been associated with schizophrenia [50]. Abnormalities of oligodendrocytes and myelin sheath surrounding axons have been observed in postmortem studies in schizophrenia [51, 52]. While there is evidence for the role of insulin related proteins such as insulin-like growth factor-I in oligodendrocyte development and altered myelin content, the role of INSRR in myelination is unclear [53]. Interestingly, given the fact that INSRR might act as pH sensor, research suggests that extracellular acidic pH decreases survival of oligodendrocyte precursor cells (OPCs) and reduces their differentiation into oligodendrocytes adding to impaired myelination in schizophrenia [54]. The co-expression clustering analysis of AATK also validates the proposed role of AATK as regulator of oligodendrocyte differentiation and myelination [13].

The chip coverage and peptide overlap analysis of the recombinant kinase screening profiles revealed a variable degree of substrate selectivity. This was expected as some kinases target a wide range of downstream targets, whereas other kinases are very selective [55]. Additionally, peptide overlap analysis revealed limited overlap between the five kinases, consistent with what we expected given that these kinases are members of five different kinase families based on phylogeny. One interesting observation from this study is that AATK, despite being labeled as a serine/threonine kinase, had strong phosphorylation activity on the tyrosine PamChip suggesting that this kinase is a dual-specificity kinase. This observation suggests that the other kinases in the subfamily, including LMTK2 and LMTK3, may be dual-specificity as well.

Notably, our five dark kinases were not screened for serine/threonine substrate activity. Provided the limited knowledgebase of substrate selectivity for dark kinases, it is possible for these kinases to have dual specificity for substrates. Future studies should include functional assessment of axon generation in response to perturbation of EPHA6 expression. Such studies will leverage the workflow established in this study to determine novel substrates for dark kinases. Of note, the observed sex-effect of dark kinases showing decreased activity in female, but not male, schizophrenia subjects will require further examination. Broadly decreased kinase activity in female, but not male, schizophrenia subjects suggest

a sex-specific signaling alteration for this often-devastating illness. Mechanistic studies examining EPHA6's role in axon guidance and development along with projects focused on sex-specific kinase signaling networks would provide important insight to the functional roles of this kinase.

Our study is not without limitations. Reporter peptides on the PTK chip were preselected and limited to 193 unique sequences, making conclusions about consensus peptide target sequences limited, as this is only a small fraction of the entire proteome. Further, the predicted coverage for the PTK chip is 96% for the tyrosine kinase subkinome, which may leave some kinases untargeted. Interactions between recombinant proteins may differ compared to cell-based systems; our promising findings, particularly for EPHA6, need to be confirmed using cell culture or animal models.

In conclusion, we identified 195 putative novel kinase-substrate interactions with variable degrees of affinity for five understudied protein tyrosine kinases. The hybrid workflow we deployed corroborated existing knowledge for these kinases and revealed potentially novel functional associations. We posit that this approach may be deployed to provide strong functional annotation leads for understudied serine/threonine kinases as well. Extending the substrate annotations of dark kinases will close the gap of knowledge and disparity in cell signaling and kinase networks, offsetting the bias of existing biological databases towards well-studied “light” protein kinases.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12964-024-01868-4>.

Supplementary Material 1. Supplementary Figure 1: Identification of dark kinases. A) Utilizing the knowledge-based INDRA database to assign normalized knowledge scores for all protein kinases. 19 understudied or “dark” protein tyrosine kinases were identified. Of these 19 dark protein tyrosine kinases, the five shown in blue were the ones of interest in our study. B) Normalized INDRA statements of our five dark kinases amongst light and dark kinases. LTK (Leukocyte Receptor Tyrosine Kinase), AATK (Apoptosis Associated Tyrosine Kinase), INSRR (Insulin Receptor Related Receptor), TNK1 (Tyrosine Kinase Non-Receptor 1), and EPHA6 (EPH Receptor A6).

Supplementary Material 2. Supplementary Figure 2: PamStation12 kinase profiles of purified recombinant kinases identify novel substrates. Four of the remaining understudied or “dark” kinases were screened using the PamStation12 (PamGene International B.V.) platform: A, G, M, and S. 2.5-ng, 25-ng, 250-ng of recombinant total protein was screened along with a 250-ng heat inactivated (denatured protein) control. Red indicates increased phosphorylation and yellow indicates less phosphorylation at that peptide. A semi-supervised clustering and principal component analysis (PCA) is visualized for PamChip peptides reporting on kinase activity. D, J, P and V) Signal intensity plot illustrating increased activity with increased total protein of recombinant kinase. E, K, Q and W) Peptide sequence logos for each recombinant protein peptide sequence identified on the PTK Pamchip. F, L, R and X) The color of the single letter amino acid code denotes the chemical features of each amino acid map against relative position. High Affinity: list of peptides that show phosphorylation activity at the lowest protein concentration. Medium Affinity: list of

peptides that only start showing phosphorylation activity at the medium protein concentration. Low Affinity: a list of peptides that only show phosphorylation activity at the highest protein concentration. No Affinity: a list of peptides that show no phosphorylation activity across all tested protein concentrations. Apoptosis-associated tyrosine kinase (AATK), Insulin Receptor-related Protein (INSRR), Leukocyte Tyrosine Kinase Receptor (LTK), and Non-receptor Tyrosine Protein Kinase (TNK1).

Supplementary Material 3. Supplementary Figure 3: A) Females B) Males. Peptides reporting on dark kinase activity in schizophrenia versus controls postmortem samples. The peptides are ranked by log₂ fold change (LFC) calculated by comparing control to schizophrenia subjects. SCZ – Schizophrenia, CTRL – Control, AATK - Apoptosis-associated tyrosine kinase, EPHA6 - EPH Receptor A6, INSRR - Insulin Receptor-related Protein, LTK - Leukocyte Tyrosine Kinase Receptor, and TNK1 - Non-receptor Tyrosine Protein Kinase'.

Supplementary Material 4.

Supplementary Material 5.

Supplementary Material 6.

Supplementary Material 7.

Supplementary Material 8.

Supplementary Material 9.

Supplementary Material 10.

Supplementary Material 11.

Supplementary Material 12.

Supplementary Material 13.

Supplementary Material 14.

Supplementary Material 15.

Supplementary Material 16.

Received: 19 May 2024 Accepted: 2 October 2024

Published online: 17 October 2024

References

- Lahiry P, et al. Kinase mutations in human disease: interpreting genotype-phenotype relationships. *Nat Rev Genet.* 2010;11(1):60–74.
- Bhullar KS, et al. Kinase-targeted cancer therapies: progress, challenges and future directions. *Mol Cancer.* 2018;17(1):48.
- Creeden JF, et al. Kinome array profiling of patient-derived pancreatic ductal adenocarcinoma identifies differentially active protein tyrosine kinases. *Int J Mol Sci.* 2020;21(22):8679.
- Ping L, et al. Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. *Sci Data.* 2020;7(1):315.
- Santos R, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov.* 2017;16(1):19–34.
- Fedorov O, Muller S, Knapp S. The (un)targeted cancer kinome. *Nat Chem Biol.* 2010;6(3):166–9.
- Vella V, Giamas G, Ditsiou A. Diving into the dark kinase: lessons learned from LMTK3. *Cancer Gene Ther.* 2022;29(8–9):1077–9.
- Axtman AD. Characterizing the role of the dark kinase in neurodegenerative disease - A mini review. *Biochim Biophys Acta Gen Subj.* 2021;1865(12):130014.
- Essegian D, et al. The clinical kinase index: a method to prioritize understudied kinases as drug targets for the treatment of cancer. *Cell Rep Med.* 2020;1(7):100128.
- Roskoski RJ Jr. Properties of FDA-approved small molecule protein kinase inhibitors: a 2023 update. *Pharmacol Res.* 2023;187:106552.
- Rodgers G, et al. Glimmers in illuminating the druggable genome. *Nat Rev Drug Discov.* 2018;17(5):301–2.
- Nishino H, et al. The LMTK1-TBC1D9B-Rab11A cascade regulates dendritic spine formation via endosome trafficking. *J Neurosci.* 2019;39(48):9491–502.
- Jiang C, et al. AATYK is a novel regulator of oligodendrocyte differentiation and myelination. *Neurosci Bull.* 2018;34(3):527–33.
- Moret N, et al. A resource for exploring the understudied human kinome for research and therapeutic opportunities. *bioRxiv.* 2021;2020.04.02.022277. <https://doi.org/10.1101/2020.04.02.022277>.
- Gyori BM, et al. From word models to executable models of signaling networks using automated assembly. *Mol Syst Biol.* 2017;13(11):954.
- Oprea TI, et al. Unexplored therapeutic opportunities in the human genome. *Nat Rev Drug Discov.* 2018;17(5):377.
- Imami A.S. *CDRL Kinome Analysis Omnibus*. Kinome omnibus 2023; Available from: https://cogdisreslab.github.io/kinome_omnibus/data_generation.html.
- DePasquale EAK, et al. KRSA: An R package and R Shiny web application for an end-to-end upstream kinase analysis of kinome array data. *PLoS One.* 2021;16(12):e0260440.
- DePasquale EAK, Alganem K, Bentea E, Nawreen N, McGuire JL, Tomar T, Naji F, Hilhorst R, Meller J, McCullumsmith RE. KRSA: An R package and R Shiny web application for an end-to-end upstream kinase analysis of kinome array data. *PLoS One.* 2021;16(12):e0260440. <https://doi.org/10.1371/journal.pone.0260440>.
- Kuleshov MV, et al. KEA3: improved kinase enrichment analysis via data integration. *Nucleic Acids Res.* 2021;49(W1):W304–16.
- Wagih O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics.* 2017;33(22):3645–7.
- Zyla J, et al. Gene set enrichment for reproducible science: comparison of CERNO and eight other algorithms. *Bioinformatics.* 2019;35(24):5146–54.
- Weiner 3rd J, Domaszewska T. tmod: an R package for general and multivariate enrichment analysis. *PeerJ Preprints.* 2016;4:e2420v1. <https://doi.org/10.7287/peerj.preprints.2420v1>.
- Xie Z, et al. Gene set knowledge discovery with enrichr. *Curr Protoc.* 2021;1(3):e90.
- Tasic B, et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature.* 2018;563(7729):72–8.
- Uhlen M, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.
- Thul PJ, et al. A subcellular map of the human proteome. *Science.* 2017;356(6340):eaal3321.

Acknowledgements

Thank you to the Advanced Microscopy & Imaging Center (AMIC) at the University of Toledo Integrated Core Facility Health Science Campus, led by Dr. Andrea Kalinoski, for their assistance with all multi-plex array experiments. We also thank the University of Cincinnati Proteomics Laboratory (UCL), led by Dr. Ken Greis, for his assistance on the mass spectrometry studies.

Authors' contribution

A.H., K.A., S.H., M.M., N.H., and R.M. all contributed to the conception of the project. A.H., K.A., S.H., M.M., N.H., and R.M. all contributed to the design of the work. A.H., A.I., W.R., K.A., A.K., R.S. and R.M. helped with the acquisition, analysis of bioinformatics data. J.M., A.I., S.O., W.R., K.A., and R.M. helped to interpret data. A.H., K.A., S.H., A.I., S.S., P.P. all contributed to the drafted of the manuscript.

Funding

NIH: AG057598, NIMH: MH107487, NIMH: MH121102.

Data Availability

The datasets generated and/or analyzed during the current study are available in the CogDisResLab/SCZ_Dark_Kinome repository, hosted on GitHub. The data can be freely accessed at https://github.com/CogDisResLab/SCZ_Dark_Kinome. This repository includes all relevant raw data, processed data, and code necessary to reproduce the analyses and results presented in this manuscript. For further inquiries or additional information, please contact the corresponding author.

Declarations

Competing interests

The authors declare no competing interests.

28. Alganem K, et al. Kaleidoscope: a new bioinformatics pipeline web application for in silico hypothesis exploration of omics signatures. *bioRxiv*. 2020;2020.05.01.070805. <https://doi.org/10.1101/2020.05.01.070805>.
29. Szklarczyk D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607–13.
30. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580–5.
31. Zhang Y, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci.* 2014;34(36):11929–47.
32. Zhang Y, et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron.* 2016;89(1):37–53.
33. Pilarczyk M, et al. Connecting omics signatures of diseases, drugs, and mechanisms of actions with iLINCs. *bioRxiv*. 2019;13:826271.
34. Nguyen DT, et al. Pharos: collating protein information to shed light on the druggable genome. *Nucleic Acids Res.* 2017;45(D1):D995–1002.
35. Lex A, et al. UpSet: visualization of intersecting sets. *IEEE Trans Vis Comput Graph.* 2014;20(12):1983–92.
36. Huang H, et al. iPTMnet: an integrated resource for protein post-translational modification network discovery. *Nucleic Acids Res.* 2018;46(D1):D542–50.
37. Kozulin P, et al. Gradients of Eph-A6 expression in primate retina suggest roles in both vascular and axon guidance. *Mol Vis.* 2009;15:2649–62.
38. Das G, et al. EphA5 and EphA6: regulation of neuronal and spine morphology. *Cell Biosci.* 2016;6:48.
39. Ooi EL, et al. Novel antiviral host factor, TNK1, regulates IFN signaling through serine phosphorylation of STAT1. *Proc Natl Acad Sci U S A.* 2014;111(5):1909–14.
40. Eid S, et al. KinMap: a web-based tool for interactive navigation through human kinome data. *BMC Bioinformatics.* 2017;18(1):16.
41. Darling TK, Lamb TJ. Emerging Roles for Eph Receptors and Ephrin Ligands in Immunity. *Front Immunol.* 2019;10:1473.
42. Savelieva KV, et al. Learning and memory impairment in Eph receptor A6 knockout mice. *Neurosci Lett.* 2008;438(2):205–9.
43. Hooli BV, et al. Rare autosomal copy number variations in early-onset familial Alzheimer's disease. *Mol Psychiatry.* 2014;19(6):676–81.
44. Wang Z, et al. Axon guidance pathway genes are associated with schizophrenia risk. *Exp Ther Med.* 2018;16(6):4519–26.
45. Ferrari R, et al. A genome-wide screening and SNPs-to-genes approach to identify novel genetic risk factors associated with frontotemporal dementia. *Neurobiol Aging.* 2015;36(10):2904 e13–26.
46. Deyev IE, et al. Deficient response to experimentally induced alkalosis in mice with the inactivated insrr gene. *Acta Naturae.* 2011;3(4):114–7.
47. Henkel ND, et al. Schizophrenia: a disorder of broken brain bioenergetics. *Mol Psychiatry.* 2022;27:2393.
48. Tilak M, et al. Receptor tyrosine kinase signaling and targeting in glioblastoma multiforme. *Int J Mol Sci.* 2021;22(4):1831.
49. Park HJ, Choi I, Leem KH. Decreased brain pH and pathophysiology in schizophrenia. *Int J Mol Sci.* 2021;22(16):8358.
50. Landek-Salgado MA, Faust TE, Sawa A. Molecular substrates of schizophrenia: homeostatic signaling to connectivity. *Mol Psychiatry.* 2016;21(1):10–28.
51. Flynn SW, et al. Abnormalities of myelination in schizophrenia detected in vivo with MRI, and post-mortem with analysis of oligodendrocyte proteins. *Mol Psychiatry.* 2003;8(9):811–20.
52. Takahashi N, et al. Linking oligodendrocyte and myelin dysfunction to neurocircuitry abnormalities in schizophrenia. *Prog Neurobiol.* 2011;93(1):13–24.
53. Ye P, et al. Myelination is altered in insulin-like growth factor-I null mutant mice. *J Neurosci.* 2002;22(14):6041–51.
54. Jagielska A, Wilhite KD, Van Vliet KJ. Extracellular acidic pH inhibits oligodendrocyte precursor viability, migration, and differentiation. *PLoS One.* 2013;8(9):e76048.
55. Bradley D, Beltrao P. Evolution of protein kinase substrate recognition at the active site. *PLoS Biol.* 2019;17(6):e3000341.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.