IDG ILLUMINATING the DRUGGABLE GENOME

Listing of the Digital Tools highlighted during the inaugural IDG Digital Tool Fest 2021 on November 30th, 2021 where IDG members showcased and demonstrated various digital and informatics tools developed within the IDG consortium. Tools are listed in alphabetical order.

Digital Tool Name: Appyters
Link to Digital tool: https://appyters.maayanlab.cloud/
Brief description: Appyters turn Jupyter Notebooks into fully functional standalone web-based bioinformatics applications. Appyters present to users an entry form enabling them to upload their data and set various parameters for a multitude of data analysis workflows. Once the form is filled, the Appyter executes the corresponding notebook in the cloud, producing the output without requiring the user to interact directly with the code. Appyters were used to create many bioinformatics web-based reusable workflows, including applications to build customized machine learning pipelines, analyze omics data, and produce publishable figures. These Appyters are served in the Appyters Catalog.
What demonstration will show/What users will learn: Using the IDG related Appyetrs; Running Appyters locally; Developing Appyters.
IDG Grant: KMC-ISMMS U24CA224260

Digital Tool Name: Appyters to identify perturbations that modulate the expression of single genes
Link to Digital tool:
My Gene's Transcription in Type 2 Diabetes (T2D) Transcriptomics Signatures Appyter
https://appyters.maayanlab.cloud/#/Gene_Expression_T2D_Signatures
Gene Centric GEO Reverse Search Appyter
https://appyters.maayanlab.cloud/#/Gene_Centric_GEO_Reverse_Search
Transformed L1000 to RNA-seq Perturbational Signatures Gene Search Appyter
https://appyters.maayanlab.cloud/#/L1000_RNAseq_Gene_Search
Brief description: The signature data from 3 databases is served as 3 Jupyter Notebook-backed web-based bioinformatics applications (Appyters). These Appyters feature interactive volcano plots that visualize signatures that maximally up- or down-regulate the input gene in the species of interest, as well as full tabulated results for download.
What demonstration will show/What users will learn: Using these 3 Appyters to form hypotheses about ways to modulate the expression of worthy IDG targets in a disease context.
IDG Grant: KMC-ISMMS U24CA224260

Digital Tool Name: ARCHS4
Link to Digital tool: https://maayanlab.cloud/archs4/
Brief description: ARCHS4 provides access to gene counts from HiSeq 2000, HiSeq 2500 and NextSeq 500 platforms for human and mouse experiments from GEO and SRA. The website enables downloading of the data in H5 format for programmatic access as well as a 3-dimensional view of the sample and gene spaces. Search features allow browsing of the data by meta data annotation, ability to submit your own up and down gene sets, and explore matching samples enriched for annotated gene sets. Selected sample sets can be downloaded into a tab separated text file through auto-generated R scripts for further analysis. Reads are

aligned with Kallisto using a custom cloud computing platform. Human samples are aligned against the GRCh38 human reference genome, and mouse samples against the GRCm38 mouse reference genome.
What demonstration will show/What users will learn: Using the ARCHS4 user interface; Accessing ARCHS4 programmatically; Making gene function predictions with data from ARCHS4.
IDG Grant: KMC-ISMMS U24CA224260

Digital Tool Name: Clinical Kinase Index (CKI)
Link to Digital tool: http://cki.ccs.miami.edu/
Brief description: CKI is a simple interactive web application with harmonized datasets extracted from several resources that allows researchers and clinicians to prioritize and evaluate the clinical relevance of kinases as cancer drug targets across solid tumors. It is an open-source platform-independent browser based Shiny application with a menu offering a number of options to query, explore, analyze, and visualize the data.
What demonstration will show/What users will learn: The demonstration will show the different options to rank prospective kinase targets for each cancer and to identify relevant tumors for a given kinase. CKI also provides visualization of the clinical data, various ways to query and filter results, and interactive visualizations of the gene expression and CKI scores across kinases and cancers. Data can also be downloaded.
IDG Grant: RDOC U24TR002278

Digital Tool Name: Drug Central
Link to Digital tool: https://drugcentral.org
Brief description: DrugCentral is an on-line compendium that brings together active pharmaceutical ingredients and drug formulations. For active pharmaceutical ingredients, this includes major properties of drugs, such as chemical information; patent status and FDA Orange Book information; pharmacokinetic properties; sex-separated side effects from FAERS; bioactivites for mode-of-action and other human and non-human targets. For formulations, it includes links to DailyMed and approval information. Multiple identifiers such as ATC codes, INNs, RxNorm, ChEBI, ChEMBL, DrugBank and MeSH are also provided.
What demonstration will show/What users will learn: Following a brief introduction to major DrugCentral fields (such as side effects, drug target information, etc.), the demonstration will show users how to use REDIAL-2020 (machine-learning based prediction of anti-SARS-CoV-2 activities) and L1000 gene signatures. The users will also learn about summary statistics and data downloads.
IDG Grant: KMC-UNM U24CA224370

Digital Tool Name: Drugshot
Link to Digital tool: https://maayanlab.cloud/drugshot/
Brief description: DrugShot is a web-based server application and an Appyter that enables users to enter any biomedical search term into a simple input form to receive ranked lists of drugs and other small molecules based on their relevance to the search term. To produce ranked lists of small molecules, DrugShot cross-references returned PubMed identifiers (PMIDs) with DrugRIF or AutoRIF, curated resources of drug-PMID associations, to produce an associated small molecule list where each small molecule is ranked according to total co-mentions with the search term from shared PubMed IDs. Additionally, using two types of drug-drug similarity matrices, lists of small molecules are predicted to be associated with the search term. Such predictions are based on literature co-mentions and signature similarity from LINCS L1000 drug-induced gene expression profiles.

What demonstration will show/What users will learn: Using the Drugshot user interface; Accessing Drugshot programmatically; Making predictions about the properties of drugs with Drugshot.
IDG Grant: KMC-ISMMS U24CA224260

Digital Tool Name: Geneshot
Link to Digital tool: https://maayanlab.cloud/geneshot/
Brief description: Geneshot is a search engine developed to illuminate this gap and to promote attention to the under-studied genome. Through a simple web interface, Geneshot enables researchers to enter arbitrary search terms, to receive ranked lists of genes relevant to the search terms. Returned ranked gene lists contain genes that were previously published in association with the search terms, as well as genes predicted to be associated with the terms based on data integration from multiple sources. The search results are presented with interactive visualizations. To predict gene function, Geneshot utilizes gene–gene similarity matrices from processed RNA-seq data, or from gene–gene co-occurrence data obtained from multiple sources. In addition, Geneshot can be used to analyze the novelty of gene sets and augment gene sets with additional relevant genes.
What demonstration will show/What users will learn: Using the Geneshot user interface; Accessing Geneshot programmatically; Making gene function predictions with data from Geneshot.
IDG Grant: KMC-ISMMS U24CA224260

Digital Tool Name: Gene and Drug Landing Page Aggregator
Link to Digital tool: https://cfde-gene-pages.cloud/
Brief description: The goal of the Gene and Drug Landing Page Aggregator (GDLPA) is to gather all primary resources and databases that provide gene and drug landing pages into one website. The GDLPA simple search engine only requires the user to submit a gene or a drug name. The website returns a collection of resources presented as cards with links to the gene and drug landing pages from each resource.
What demonstration will show/What users will learn: Case studies of using the resource for finding knowledge about IDG targets.
IDG Grant: KMC-ISMMS U24CA224260

Digital Tool Name: Pharos
Link to Digital tool:https://pharos.nih.gov/
Brief description: Pharos is the user interface for the IDG generated TCRD (Target Central Research Database). Pharos displays collated information on targets, diseases and ligands, and allows users to filter and browse by a multitude of filters to create and explore lists of interest. Further details are viewable on the specific target, disease and ligand level, and new updates allow for deeper list analysis, focusing on filter value enrichment, predicted activities, disease associations and target interactions.
What demonstration will show/What users will learn: Participants will learn how to browse through Pharos, upload and save lists, analyze lists and get predicted ligand interactions.
IDG Grant: KMC-UNM U24CA224370

Digital Tool Name: Pharos GraphQL API
Link to Digital tool:https://pharos.nih.gov/api
Brief description: Pharos is the user interface for the IDG generated TCRD (Target Central Research Database). In a recent update, Pharos moved from a traditional REST API to the Facebook created GraphQL

API methodology. This allows more fine-tuned queries, so users are able to fetch only the data they need, but also fetch deeply nested data with ease.
What demonstration will show/What users will learn: Participants will learn how to query the Pharos API using GraphQL.
IDG Grant: KMC-UNM U24CA224370

Digital Tool Name: Protein Kinase Ontology (ProKinO)
Link to Digital tool: https://prokino.uga.edu/nb/
Brief description: The protein kinase ontology (ProKinO) is an integrated knowledge graph conceptualizing the wealth of information connecting protein kinase sequence, structure, function, evolution and disease (cancer) in human and machine readable format.
What demonstration will show/What users will learn: Demonstrations will focus on ProKinO knowledge graph mining using SPARQL and visualization tools for mining kinase sequence and regulatory features in light of structural models (AlphaFold) and crystal structures. Attendees will learn to formulate new testable hypotheses on understudied dark kinases using evolutionary and functional context of the entire kinome in one place.
IDG Grant: CEIT-UGA U01CA239106
Digital Tool Name: Reactome-IDG Portal
Link to Digital tool: https://idg.reactome.org
Brief description: Placing understudied proteins in the context of biological pathways facilitates the generation of experimentally testable hypotheses to infer potential functions of these proteins. The Reactome Pathway Diagram View is a web-based tool, providing a biologist-friendly way to visualize proteins, complexes, and reactions in high-quality Reactome pathways. In order to put understudied proteins in the context of Reactome pathways, we have extended the Pathway Diagram View to overlay tissue-specific expression data, protein pairwise relationships, and drug/target interactions. We implemented new interfaces for users to select tissue-specific mRNA and protein expression data from 19 data sources collected in the Target Central Resource Database (TCRD). The pairwise relationship overlay allows users to display positive and negative relationships from multiple sources in the same view. We have also implemented a new visualizer via the use of Cytoscape.js, allowing a pathway to be displayed as a set of functional interactions. Drugs can be overlayed in our pairwise view and in the new functional interaction visualizer. The new features we have introduced in the Reactome Pathway Diagram View pave the way for us to predict and visualize functions of understudied proteins based on Reactome pathways.
What demonstration will show/What users will learn:
Search the idg.reactome.org web portal with understudied protein identifiers to identify primary and interacting pathways reachable via one-hop pairwise relationships.
Visualize understudied proteins location in Reactome's annotated pathways and interacting pathways and network visualisations..
Overlay multiple tissue specific gene or protein expression values from 19 data sources collected in the Target Central Resource Database onto Reactome pathway and network visualisations.
Overlay protein/protein pairwise relationships or drug/target interactions onto Reactome pathways and network visualisations.
IDG Grant: CEIT-OHSU U01CA239069

Digital Tool Name: TIGA
Link to Digital tool: https://unmtid-shinyapps.net/shiny/tiga/
Brief description: TIGA (target illumination GWAS analytics) is an algorithm, automated workflow, and web application providing rational ranking, filtering and interpretation of inferred gene–trait associations and aggregated GWAS evidence by leveraging existing curation and harmonization efforts from the NHGRI-EBI GWAS Catalog. Each gene–trait association is evaluated for confidence, with scores derived solely from aggregated statistics, linking a protein-coding gene and phenotype. We propose a method for assessing confidence in gene–trait associations from evidence aggregated across studies, including a bibliometric assessment of scientific consensus based on the iCite relative citation ratio, and meanRank scores, to aggregate multivariate evidence.
What demonstration will show/What users will learn: This demo will feature the TIGA web app as a tool for conveniently finding, prioritizing, and exploring gene-trait associations from aggregated GWAS evidence, in the context of drug target illumination scientific use cases.
IDG Grant: KMC-UNM U24CA224370

Digital Tool Name: TIN-X
Link to Digital tool: https://newdrugtargets.org
Brief description: TIN-X (Target Importance and Novelty eXplorer) is an interactive visualization tool for illuminating associations between diseases and potential drug targets. TIN-X uses natural language processing to identify disease and protein mentions within PubMed content using previously published tools for Named Entity Recognition (NER) of gene/protein and disease names. Target data is obtained from Target Central Resource Database (TCRD). Two important metrics, novelty and importance, are computed from this data, and when plotted as log(importance) vs log(novelty), aid the User in visually exploring the novelty of drug targets and their associated importance to diseases. The original TIN-X [Cannon et al 2017] has been significantly improved with an expanded dataset, modernized architecture, a REST API, and UI enhancements.
What demonstration will show/What users will learn: Using TIN-X to visually explore understudied drug targets; Demonstration of interactive features of the TIN-X User Interface.
IDG Grant: KMC-UNM U24CA224370