Milestones for IDG2 Mount Sinai KMC (Ma'ayan) - 12/05/2017

Harmonizome Related Milestones

1. Systematically continue the abstraction and integration of data from all possible resources by processing the data into attribute tables, gene-sets, bi-partite graphs, and functional association networks. The data processing scripts will be shared as Jupyter notebooks and on GitHub (S, M, L).
2. Work with the UNM team to make sure selected resources are incorporated within TCRD and Pharos (S, M, L).
3. Make the processing scripts available as live notebooks on the cloud (Google and/or AWS) (M).
4. Release a new version of the Harmonizome with updated datasets and interactive visualizations (M).
5. Incorporate ARCHS4 predictions within the Harmonizome (M).
6. Expand Harmonizome to include associations between: gene/protein, gene-set/pathway, drug/small-molecule, cell/tissue, and patient/disease/side-effects (L).
7. Identify where associations listed in item 6 are signed and directional. Incorporate this information to enable complex queries.
8. Update the Harmonizome mobile application with new features to find information about all human and mouse genes (M, L).

ARCHS4 Related Milestones

9. Periodically update the ARCHS4 resource as more RNA-seq data becomes available (M, L).
10. Create cell type specific co-expression networks from ARCHS4 and GTEx data and predict gene functions from those cell type specific datasets (S).
11. Process all publicly available ChIP-seq data to identify regulatory elements upstream of the dark targets of IDG interest, and incorporate this data within ARCHS4, Harmonizome, TCRD and Pharos (M, L).
12. Create transcript level download files for ARCHS4 users (S).
13. Create ligand and receptor gene set libraries from ARCHS4 and analyze these to obtain a global view (M).

Expression2Kinases Related Milestones

14. Develop the Expression2Kinases (X2K) web application, a tool to link changes in gene expression to upstream protein kinases, including dark kinases (S, M).
15. Validate X2K using data from the Data and Resource Generation Centers (DRGCs) (M, L).
16. Validate X2K using genetic algorithms to learn optimal parameters (S, M).

Mount Sinai EMR Related Milestones

17. Systematically stratify patients based on diagnoses to identify unique cohorts within the Mount Sinai EMR, for example, patients with opioid dependence or a specific form of heart disease, and then, in combination with the Mount Sinai BioMe genomic profiling biobank repository, perform GWAS to identify genes that harbor mutations that present risk (M, L).
18. Harvest the identified mutations in IDG targets and provide the information to be incorporated within Harmonizome, TCRD and Pharos (M, L).

GeneRIF Processing Related Milestones

19. Create gene set libraries by systematically utilizing GeneRIF and sets of biological terms such as cell types, drugs, tissues, diseases, other genes, and more. From these libraries assess the knowledge available about the under-studied IDG targets (S, M).
20. Add this information to be consumed by Harmonizome, TCRD and Pharos resources (M, L).

Processing, Analyzing and Visualizing DRGCs Generated Data

21. Once the new data from the DRGCs is released and become available, we will employ our methods to process, integrate, analyze and visualize it. This effort will be carefully coordinated with the UNM team (M, L).

S- short term; M- midterm; L- long term.