**Transcription Factors (TFs)**

* ***[ChEA 2016](http://amp.pharm.mssm.edu/Enrichr/)***
  + *Function*: [Part of Enrichr]. ChEA is a tool to identify TFs whose mRNA targets are enriched in a user-supplied gene list (using the same algorithm as ChEA). This allows the identification of up-stream regulators for the given gene lists (e.g. derived from transcriptomic differential expression) based on prior knowledge of these TF-gene relationships..
  + *Data*: ChEA currently contains reference data from 100+ ChIP-chip, ChIP-seq and ChIP-PET experiments in mammalian cell cultures.
* [***ARCHS4 human TF***](http://amp.pharm.mssm.edu/archs4/data.html)
  + *Function:* ARCHS4 provides 1. visualization of aggregated human RNA-seq samples based on metadata, 2. Identification of samples with similar expression profiles to user-supplied lists of up/down-regulated genes. By using the Enrichment feature (Select: Human, Sample, Encode TF ChIP-seq 2015) you can also find which samples have expression profiles that are enriched in genes regulated by a given transcription factor. You can then repeat this for each transcription factor, and download these samples and get the expression profile for them to see if the user-supplied gene list is enriched for those genes.
  + *Data*: Includes 12,254+ human RNA-seq data from HiSeq 2000/2500 platforms as well as ENCODE TF ChIP-seq 2015 data.
    - *GMT in Moshe’s ownCloud (only human, not mouse)****.***
* ***[ARCHS4 mouse TF](http://amp.pharm.mssm.edu/archs4/data.html)***
  + *Function*: Same as ‘ARCHS4 human TF’ but using mouse datasets instead.
  + *Data*: Includes 30,022+ mouse RNA-seq data from HiSeq 2000/2500 platforms.
* ***ENCODE 2016***
  + *Function:* Uses a variety of methods to uncover the detailed structure and function of each part of every gene and its regulatory mechanisms (long-range regulatory elements, promotors, genes, transcripts).
  + Data: Contains data characterizing chromatin (5C ChIA-Pet), long-range regulatory elements (DNase-seq, FAIRE-seq, ATACC-seq, ChIP-seq,), promoters (ChIP-seq, WGBA< RRBS, methyl array), genes (computational predictions), and transcripts (RNA-seq, CLIP-seq, RIP-seq). The ChIP-seq data on regulatory elements is most relevant for the purposes of TF-enrichment analysis.
    - *In Moshe’s ownCloud****.***
* ***Crowd TF***
  + *Same as CREEDS?*
* ***huMAP TF***
  + *Function:*
  + *Data:*
* ***BioGRID low throughput TF***
  + Couldn’t find this exactly, but could perhaps get from PPI file?
* ***JASPAR-TRANSFAC***
* ***CREEDS-TF***

**PPI**

* ***[BioGRID](https://downloads.thebiogrid.org/BioGRID/Release-Archive/BIOGRID-3.4.154/)***
  + *Function*: Data repository for biological interactions in genera (including but not limited to PPIs).
  + *Data*: Currently includes 1,505,899 physical interactions between proteins or genes.
* ***[BioPlex](http://bioplex.hms.harvard.edu/downloadInteractions.php)***
  + *Function*: PPI network database derived from thousands of experiments in cell lines.
  + *Data*: Data for 7,500+ experiments and 73,000+ interaction were collected via immunopurification and detection of associated/tagged proteins by mass spectrometry (AP-MS). BioPlex versions 1.0, 2.0 and unpublished recent data (used unpublished).
    - *Have in 3 different formats*.
* ***[iREF](http://wodaklab.org/iRefWeb/search/index)***
  + *Functions:* Combines 10 other PPI databases that have been curated and allow cross-referencing between these datasets.
  + *Data:* Click ‘Download Intera[c]tome” button to download PPI. Includes…
* BIND
* BioGRID
* CORUM
* DIP
* IntAct
* HPRD
* MINT
* MPact
* MPPI
* OPHID
* ***[hu.MAP](http://proteincomplexes.org/download)***
  + *Function*: Combines two large-scale PPI datasets derived from AP-MS (BioPLEX) and Hein et al. (2015).
  + *Data*:
* ***[MINT](http://mentha.uniroma2.it/download.php)***
  + *Function:* PPI dataset avaiallbe on the Mentha interactome browser.
  + *Data:* Includes interactomes from 8 different species. There is also virusMINT in Enrichr.
    - *Do we want all species? Separate human and mouse data?*
* ***[SNAVI](https://code.google.com/archive/p/snavi/downloads)***
  + *Function*: Desktop application to analyze and visualize protein-protein and ligand-protein interaction networks.
  + *Data:* Includes “Large-Scale Mammalian Cell Signalling Network” (PPIs + ligand-protein interactions?) in sig format.
* ***Ppid***
  + *Can’t find. Found* [*PIPs*](http://www.compbio.dundee.ac.uk/www-pips/dbStats.jsp) *instead though. Or maybe he meant “Pfam\_InterPro\_Domains” in Enrichr?*
* ***KEGG***
  + *Isn’t KEGG an ontological category database?*

**Kinase**

* ***[Phosphosite](https://www.phosphosite.org/staticDownloads.action)***
  + *Function*: Data repository to gather information on posttranslational modifications (PTMs), including phosphorylation, ubiquitination, acetylation and methylation.
  + *Data*: All datasets are downloadable. The Kinase\_Substrate\_Dataset is a list of kinases and the protein they phosphorylate/interact with. Species in this data include,
    - *Do we just want humans and/or mice*? Separate?
* ***NetworkKlN***
  + ***Can’t find?....***
* ***[KEA 2015](http://www.maayanlab.net/KEA2/)***
  + ***Function:*** KEA is a tool to identify kinases that phosphorylate proteins/genes in a user-supplied protein/gene list (using the same algorithm as ChEA).
  + ***Data:*** Currently contains 2 phosphite level datasets (1. Literature-based kinase-substrate library with phosphites, and 2. Biological terms associated with phosphites from literature mining) and 7 protein-level datasets (1. Literature-based kinase-substrate library, 2. Biological terms associated with phosphorylated protein from literature mining, 3. SILAC experiment gene set library, 4/5. Up/down-regulated genes following kinase perturbation from GEO, 6/7. Up/down-regulated genes following kinase perturbation from the L1000 connectivity map
    - ***Lots of different files, not sure which one to download***
    - ***NOTE: Can’t use some of the buttons (e.g. Downloads) in Safari. Had to use Chrome instead.***
* ***Kathy Wu***
  + *Additional dataset that Avi may get from Kathy Wu (check in).*
* ***L1000 Kinases***

**Tests**

* ***Fisher:*** Both ChEA and KEA use Fisher’s exact test.
* ***Alex odds ratio adjustment:*** This is in additional method used to sort TFs hits as well as Kinase hits (according to Alex).

**Validation Datasets**

***Drug-Target Data***

* ***[DrugBank](https://www.drugbank.ca/)***
  + *Data*: Drug perturbation data. The target of each drug from DrugBank combined with drug perturbation expression results from 407 different experiments (done by Moshe).
    - *Have Moshe’s data as well as the edge lists Alex gave me.*
* [***Target Central***](http://juniper.health.unm.edu/tcrd/)
  + *Function*: Database that contains the molecular target(s) for a huge list of drugs with a focus on several families of targets: GPCRs, kinases, ion channels and nuclear receptors (and olfactory receptors).
  + *Data*: Currently includes 20,031 targets of 1,795 unique drugs. Drugs are put into four categories to describe what kind of knowledge we currently have about them: Tclin (targets in DrugCentral), Tchem (targets in CheMBL or DrugCentral), Tbio (little is known about these targets), Tdark (virtually nothing is known about these targets).
* ***[Drug Repurposing Hub](https://clue.io/repurposing-app)***
  + *Data*: Lots of metadata on each drug including SMILES.

***Drug Perturbation Transcriptomic Datasets***

* ***[L1000](http://amp.pharm.mssm.edu/Enrichr/" \l "stats)***
  + *Data:* Kinase Perturbations datasets for both gene knowndown and gene knockups in separate files (is are they up- and down-regulation?). Located in Enrichr downloads.
* ***[CREEDS](http://amp.pharm.mssm.edu/CREEDS/)***
  + *Data*: Contains Up and downregulation of Located in CREEDS downloads.
* ***[RNA-seq sig](http://amp.pharm.mssm.edu/Enrichr/" \l "stats)***
  + *Data*: Disease and drug perturbation experiments separated into up- and down- regulated genes. Located in Enrichr downloads as ‘RNA-Seq\_Disease\_Gene\_and\_Drug\_Signatures\_from\_GEO’.