**February 24, 2020 DRAFT**

**Analysis of Cancer Cell Line Omics Data**

Four papers with omics data

1. CRISPR screen: 20 ESCC cell lines, 18009 genes
2. RNAseq, WES, WGS, RRPA, RRBS, microRNA, histone modification profiling, SNP array data, GDCS data, structural variants, gene-fusion events, shRNA and sgRNA dependence (Project Achilles and Project DRIVE): 22 ESCC cell lines
3. Metabolomics data: 24 ESCC cell lines, 225 metabolites
4. Drug sensitive data: 25 ESCC cell lines, 265 drugs

Other databases:

1. Project Achilles
2. Project DRIVE
3. GDSC
4. PharmacoDB

Research questions:

1. What omics data can we download? Quality? Usability?
2. NRF2 in ESCC cells
   1. General characterization of omics data
   2. Determine NRF2 pathway activity status (RNAseq data analysis using clustering assay and hsNRF2hiESCC\_234 gene set/mmNRF2ChIPseqES\_1923 gene set)
   3. Compare NRF2high vs NRF2low ESCC cells
      1. What causes NRF2high?
         1. Due to mutations of NRF2/KEAP1/CUL3 (correlate WES/WGS data)
         2. Due to other molecular events (How to analyze?)
      2. Potential drug targets in NRF2high ESCC
         1. Correlate NRF2 status with CRISPR data
         2. Validate with Project Achilles and Project DRIVE data
      3. Metabolism
         1. Correlate NRF2 status with 225 metabolites data
         2. Compare with data in two published papers (ESCC, NSLC)
      4. Drug sensitivity (265 drugs, GDSC, PharmacoDB):
         1. Correlate NRF2 status with drug sensitivity data
         2. Find potential drug combinations for NRF2high ESCC
      5. Correlate NRF2 status with protein expression (RRPA data)
      6. Correlate NRF2 status with gene methylation (RRBS data)
      7. Correlate NRF2 status with microRNA expression (microRNA profiling data)
      8. Correlate NRF2 status with histone modification (histone profiling data)
   4. Validate above discoveries in NRF2high HNSCC/OSCC vs NRF2low HNSCC/OSCC cells
3. p53 in ESCC cells
4. Other pathways in ESCC cells
5. Prediction of drug sensitivity of ESCC cells with AI
   1. Training set and test set?
   2. Develop a searchable database for commercial use