Here we lay out the additional annotation datasets due to opt as part of the ENCODE cancer resource.

- ENCODE CANCER [EC]
    - ANNOTATION [AN]
        - [annotation] We describe the principal sublines in ENCODE with mini assays, and also all the other sublines that are cancer-associated that have a normal match associated with them. + [4+1 subset]
            - EC_AN_summary_cancerCellLine.xls :: Key TF summary [WM, XL, LR, DL]: https://docs.google.com/spreadsheets/d/10w0uA5PnHIihlO10hxi1GDz4g--AxvuQBgGbCry_d58/edit?usp=sharing
        - [annotation] key transcription factors and cancer annotation
            - EC_AN_summary_keyTF.xls :: ENCODE cancer cell line summary [DL, Robert Klein, ENCODE consortium]: https://docs.google.com/spreadsheets/d/1gcbSyhrEoZl8RqXfeheacCmD3bq-Zk3eR681Vah_thE/edit?usp=sharing
    - DATA PROVISION [DP]
        - [data provision] additional uniformly processed ENCODE2 resources + raw data
            - EC_DP_uniformE2_ChIP-seq_hm_{sample}_{target}_{treatment}_{lab}_{accession}.bigwig
            - EC_DP_uniformE2_{assay}_{datatype}_{sample}_{target}_{treatment}_{lab}_{accession}.bigwig
            - ...
        - [data provision] processed data summary + deduplication
            - EC_DP_summary_dataAccession.xls
        - [data provision] variant calls for the liver cancer cohort, we provide actual mutation cause [[unclear]] for both germline and normal. This can be used to test many of the calculations. However, because of privacy restrictions, this is not possible with TCGA.
            - EC_DP_variantCall_LIHC.xxx
        - [data provision] We provide a normalized set of gene expression data from TCGA that has been merged to ENCODE specifications.
            - EC_DP_expressionTCGA.zip
    - ANALYSIS [AS]
        - Background Mutation Rate [BMR]
            - [analysis/BMR] We provide a set of normalized functional genomic signal files that are suitable for background mutation rate models. We provide a good match for many of these signals to appropriate sublines. We then provide the code for the model and the results of running the model on many popular sublines in various windows, so one can estimate the

background mutation rate in these regions and then compare to the observed rate in cohorts. We also provide the observed rate of mutations in a number of the TCGA cohorts, and also the corresponding data for the freely available liver cancer cohort.

    - EC_AS_BMR_covMatrix.dat
    - EC_AS_BMR_extGene.tsv
    - EC_AS_BMR_burdenScore_BRCA.tsv
    - EC_AS_BMR_burdenScore_CLL.tsv
    - EC_AS_BMR_burdenScore_LIHC.tsv

- Cis-Regulatory Elements [CRE]
  - [analysis/cis-RE] We make available sets of enhancers and promoters developed in an assay-rich environment for our principal cancer associated cell lines, both germ and normal. These sets of enhancers and promoters are calculated from the signal shape algorithm and also from enhancer-seq data processing and then from ensembling these two to get a best guess. #ESCAPE (EnhancerSeq CAller for PEak) #HISHAPE
    - EC_AS_CRE_HISHAPE_K562.tsv
    - EC_AS_CRE_HISHAPE_GM12878.tsv
    - EC_AS_CRE_ESCAPE_K562.tsv
    - EC_AS_CRE_ESCAPE_GM12878.tsv
    - EC_AS_CRE_ESCAPE_MCF-7.tsv
    - EC_AS_CRE_ESCAPE_HepG2.tsv
  - [analysis/cis-RE] We provide enhancer-to-target gene linkages from the ensemble method. These are from ensembling both Hi-C calculations and also looking at correlations of Histone marks using the method of Cao, et al. #JEME
    - EC_AS_CRE_JEME_K562.tsv
    - EC_AS_CRE_JEME_GM12878.tsv
    - EC_AS_CRE_JEME_HepG2.tsv
    - EC_AS_CRE_JEME_MCF-7.tsv
- Network [NET]
  - [analysis/network] We provide a variety of different transcription factor to promoter and transcription factor - enhancer networks. We provide a full network with all possible linkages which has hundreds of thousands of edges for a number of the main cell lines. Then we also filter this network using the TIP algorithm to give a set of the strongest edges. We then further divide these into distal and proximal edges. #TIP #TOPIC_model/GENE_community
    - EC_AS_NET_proxTSS_K562.tsv
    - EC_AS_NET_proxTIP_K562.tsv
    - EC_AS_NET_proxGTM_K562.tsv
    - EC_AS_NET_distENH_K562.tsv

        - EC_AS_NET_proxTSS_GM12878.tsv
        - EC_AS_NET_proxTIP_GM12878.tsv
        - EC_AS_NET_proxGTM_GM12878.tsv
        - EC_AS_NET_distENH_GM12878.tsv
    - [analysis/network] We provide merge networks. We merge all the cell line specific networks into a single universal network that can be used to across cell line and a rough indication of the regulatory network active in different contexts.
        - EC_AS_NET_mergedTF_K562.tsv
        - EC_AS_NET_mergedTF_GM12878.tsv
        - EC_AS_NET_mergedTF_HepG2.tsv
        - EC_AS_NET_mergedTF_liver.tsv
        - EC_AS_NET_mergedTF_A549.tsv
        - EC_AS_NET_mergedTF_IMR-90.tsv
        - EC_AS_NET_mergedTF_MCF-7.tsv
        - EC_AS_NET_mergedTF_MCF-10A.tsv
        - EC_AS_NET_mergedTF_HeLa-S3.tsv
    - [analysis/network] For all the cell line specific regulatory networks we provide a hierarchical description of each of the networks, with the TFs at the top, middle and bottom networks, from running a variety of different hierarchy algorithms.
        - EC_AS_NET_hierarchyTF_K562_GM12878.tsv
    - [analysis/network] For each of the TFs that are common to the various tumor-normal pair networks, we also provide a rewiring number that describes the degree in which the number of edges, the TF, increase, decrease and remain common between a tumor and normal.
        - EC_AS_NET_rewiringTF_K562_GM12878.tsv
        - EC_AS_NET_rewiringTF_HepG2_LIVER.tsv
        - EC_AS_NET_rewiringTF_A549_IMR90.tsv
        - EC_AS_NET_rewiringTF_MCF7_MCF10A.tsv
    - [analysis/network] We provide a merged regulatory network for K562 and also HepG2 for RNA binding proteins.
        - EC_AS_NET_mergedRBP_K562.tsv
        - EC_AS_NET_mergedRBP_HepG2.tsv

- ### Motif [MTF]
    - Motif disruption score
        - EC_AS_MTF_MotifTools_dscore_allTF.tsv