KG Lab T2DREAM Submission Protocol

Version 2

1. Submit **EXPERIMENT form** containing pre-database-defined information such as “Award” (grant or funding information), “lab” (lab information), , “assay\_term” (ATAC-seq/ChIP-seq/DNase-seq/WGBS/RNS-seq/need to be put into the DB ahead), “biosample\_term\_name” (adipocyte/liver/etc.), “biosample\_term\_id” (CL:#/UBERON:#/etc.), “biosample\_type” (tissue/primary cell/etc.), “target” (need to be put into the DB ahead; can leave blank), “possible\_controls” (possible control used; can leave blank), and etc. After filling in all the information, submit through “T2DREAM🡪ALL/Selected🡪Post to T2DREAM”; for modification, submit through “T2DREAM🡪ALL/Selected🡪Patch to T2DREAM”. Column B needs to leave blank when doing submission. If information has been submitted to T2DREAM without any problems, column B will show as “success”. **”EXPERIMENT”** key can be gotten from T2DREAM through “T2DREAM🡪ALL/Selected🡪Get from T2DREAM” for later use. Each **EXPERIMENT** represents a group of samples either are biological replicates or technical replicates and the same target. For examples, ChIP-seq data “e10.5\_heart\_H3K4me3” and “e10.5\_limb\_H3K4me3” will be two experiments. Also Column C “experiment aliases” can be used exchangeable with “EXPERIMENT” key, it highly recommended to use this make-more-sense key to track the data.
2. Submit **LIBRARY form** containing pre-database-defined information and get a **”LIBRARY”** key from the database if needed. Column C “library aliases” can be used exchangeable with “LIBRARY” key. Each **LIBRARY** represents an individual/sample from one **EXPERIMENT**. For examples, ChIP-seq data “e10.5\_heart\_H3K4me3” might have 2 samples and will generate 2 library records such as “e10.5\_heart\_H3K4me3\_YS001” and “e10.5\_heart\_H3K4me3\_YS002”. ATTENTION: **EXPERIMENT form** and **LIBRARY form** can be submitted depending on user preference, but it might be easier to submit the **EXPERIMENT form** first and add some postfix to “experiment aliases” as “library aliases”
3. Submit **REPLICATE form** containing the user-generated unique **“ALIASES: ARRAY”** key which connect column D the **”EXPERIMENT”** key from step 1 and column G the **”LIBRARY”** key from step 2 together. Each **REPLICATE** represents a single sample, and each biological or technical replicates will have it’s own replicate recode.
4. **FILES form** contains an unique 1-to-1 **“FILES ALIASES: ARRAY”** key for each file and after filling in all the information it will be used as input for gz2T2D.pl on Linux to submit files (fastq/bam/bigBed/bigWig/tagAlign/etc.) to T2DREAM.
5. Pre-defined information in ENCODE:
   1. <https://www.encodeproject.org/help/getting-started/#organization>
   2. <https://www.encodeproject.org/awards/>
   3. <https://www.encodeproject.org/labs/>
   4. <https://www.encodeproject.org/datasets/>
   5. <https://www.encodeproject.org/biosamples/>
   6. <https://www.encodeproject.org/targets/>
   7. <https://www.encodeproject.org/documents/>
   8. <https://www.encodeproject.org/projects/>
6. Pipelines
   1. <https://www.encodeproject.org/pipelines/>
   2. <https://www.encodeproject.org/atac-seq/#overview>
   3. <https://github.com/kundajelab/atac_dnase_pipelines>