Data Preprocessing Workflow
W = White
B/AA = Black or African American
N = Normal tissue
T = Primary tissue
GDC Data Portal - NCI
TCGA-LUAD cohort
Input
W: 387 (T) and 40 (N) samples
B/AA: 52 (T) and 6 (N) samples
BAM to FASTQ conversion
FASTQ Quality Control
FASTQ mapping with miRge2.0
Data Filtering and Mapping
Single Nucleotide Variant (SNV) Data Annotation
114,312 dbSNP (v154)
10,131 COSMIC (v92)
43,726 TCGA Somatic mutations (v20)
348 TARGET Somatic mutations (v20)
2,885 MiREDiBase known A-to-I miRNA Editing sites (v1)
Mapped miRNA isoforms
Data Filtering
Data Annotation and Filtering
Downstream Analyses