Files for kit comparison

1. Accuracy – synth\_data\_generation.rmd, synth\_plots\_SStype3.rmd, synth\_stats\_seq\_features.rmd
   1. normalized\_synth.rmd – created by[synth\_fold\_output.txt(need methods for this), synth\_seqs\_named.fa, Synthdata\_five\_percent\_double\_mapped/miR.Counts.csv]
   2. Var\_exp\_Table (was created from same code but with diff factors in the lm so that all could be included)
2. Composition – May22\_Percentage.Rmd
   1. Percentages\_other\_RNAs
3. Detection – Detection\_Plot.Rmd
   1. Pheno<- read.table(here("Brain\_samples\_missingClontech\_batch2/Pheno\_2.csv"), header = T, sep = ",")
   2. miR\_counts <- read.table(here("Brain\_samples\_missingClontech\_batch2/miR.Counts.csv"),header = T, sep = ",")
   3. KitVenndiagram\_may22\_2batches – seq logo doesn’t work for unique detected seqs because so few across batches that are 22 nuc long with DEseq instead of TMM –
   4. miRNA\_detection\_June.Rmd
      1. – seq logo for unique sequences for each kit
      2. across starting amount This has the inconsistency plot!!!
   5. IsomiR\_analysis\_new\_dataJune.Rmd – has the isomiR detection plot

Starting amount script… listed just below has code for venn diagrams of each starting amount

1. Starting Amount – Starting\_amt\_June.Rmd
   1. Pheno<- read.table(here("Brain\_samples\_missingClontech\_batch2/Pheno\_2.csv"), header = T, sep = ",")
   2. miR\_counts <- read.table(here("Brain\_samples\_missingClontech\_batch2/miR.Counts.csv"),header = T, sep = ",")

Does this have within batch consistency ???

1. Consistency across batch and within batch– StandardDeviation1000kg\_type3.Rmd and Starting\_amt\_June.Rmd
   1. Pheno<- read.table(here("Brain\_samples\_missingClontech\_batch2/Pheno\_2.csv"), header = T, sep = ",")
   2. miR\_counts<-read.table(here("Brain\_samples\_missingClontech\_batch2/miR.Counts.csv"), header = TRUE, sep = ",")
   3. load(here("miRNA\_hsa\_lengths.rda"))
   4. load(here("UnionInfo.rda"))####NEED to redo this… using: miRNA\_info\_May22.Rmd[mature.fa, NEED folding results on computer at work (might be here /Users/carriewright/Documents/miRNA seq Projects/fastas\_and\_foldinginfo/SOMETHING\_fold\_output.txt), and need to use synth\_data\_generation.rmd as a guide to generate – will just generate for all miRNAs- not just union at any given point in time)
   5. maybe use miRNA\_hsa\_lengths\_and mirNA\_hsa\_GC
2. Isomir Analysis
   1. isomiRs\_new\_Data incomplete
3. Overall analysis:
   1. Kit\_variance\_explained\_and\_Dendrogram.Rmd

* Variance plot and Dendrogram options
  1. Correlation\_plot.Rmd with GGally

1. UMI – synth\_preprocessing\_and\_UMI\_length.Rmd