Data processing and analysis workflow

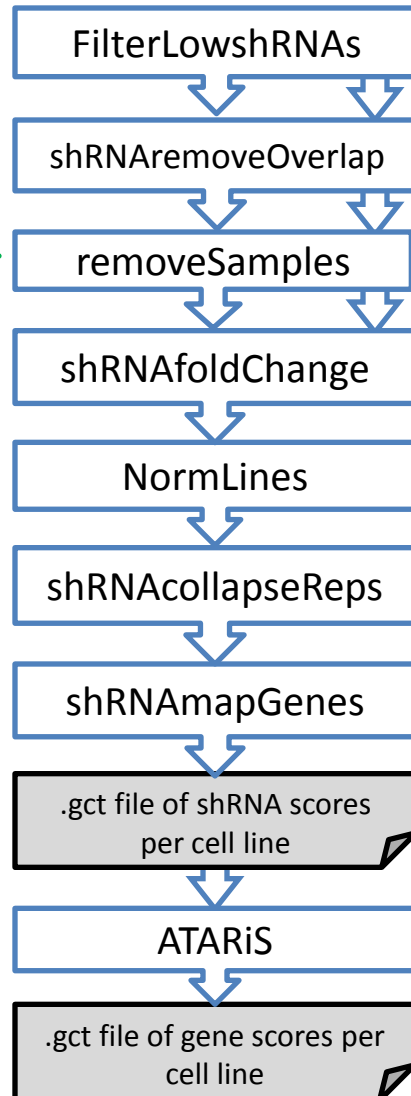**Achilles data processing GenePattern pipeline**

The *FilterLowshRNAs* module produces two matched output files: 1) initial DNA pool values and 2) cell line values, per shRNA. All modules until the *shRNAfoldChange* module are run on both.

Quality control GenePattern modules

FPmatching

ReplicatesQC

Supplemental Figure 1 outlines the GenePattern modules available for quality control, data processing, normalization, and downstream data analysis of Achilles data. All are available as individual GenePattern modules (http://genepattern.org)

FilterLowshRNAs

shRNAremoveOverlap

removeSamples

shRNAfoldChange

NormLines

shRNAcollapseReps

shRNAmapGenes

.gct file of shRNA scores per cell line

ATARiS

.gct file of gene scores per cell line

PARIS

Removes undesirable shRNAs and cell line replicates:
• Those shRNAs low in initial DNA pool or that overlap in sequence.
• Replicate samples that fail QC

• Calculates fold change values between initial DNA pool values and final cell line values.
• Normalizes cell lines to the same scale (quantile, ZMAD, PMAD)
• Collapses replicate cell line samples to a single values per shRNA
• Maps shRNAs to gene symbols based on a mapping file