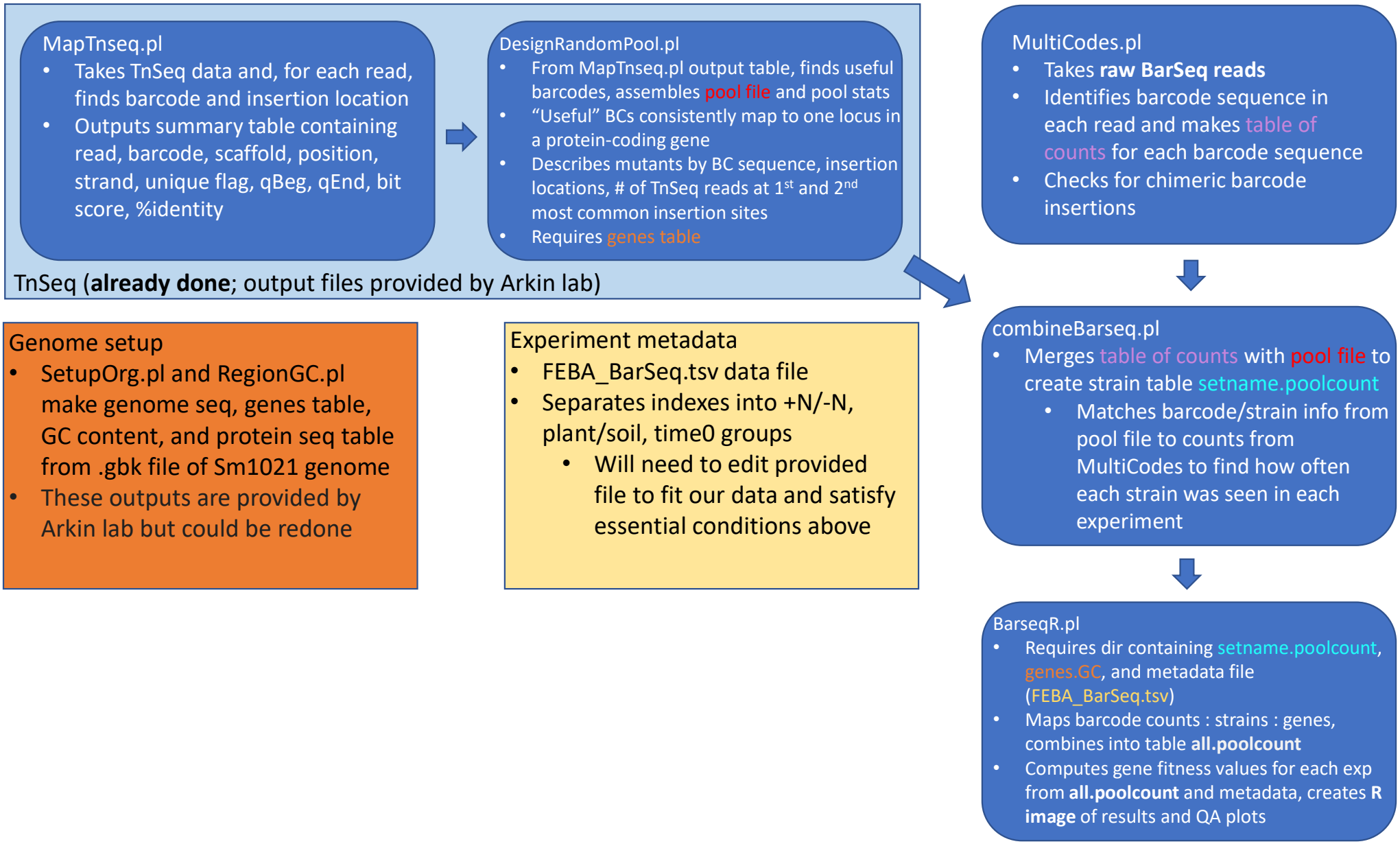
MapTnseq.pl
• Takes TnSeq data and, for each read, finds barcode and insertion location
• Outputs summary table containing read, barcode, scaffold, position, strand, unique flag, qBeg, qEnd, bit score, %identity
DesignRandomPool.pl
• From MapTnseq.pl output table, finds useful barcodes, assembles pool file and pool stats
• "Useful" BCs consistently map to one locus in a protein-coding gene
• Describes mutants by BC sequence, insertion locations, # of TnSeq reads at 1st and 2nd most common insertion sites
• Requires genes table
TnSeq (already done; output files provided by Arkin lab)
MultiCodes.pl
• Takes raw BarSeq reads
• Identifies barcode sequence in each read and makes table of counts for each barcode sequence
• Checks for chimeric barcode insertions
Genome setup
• SetupOrg.pl and RegionGC.pl make genome seq, genes table, GC content, and protein seq table from .gbk file of Sm1021 genome
• These outputs are provided by Arkin lab but could be redone
Experiment metadata
• FEBA_BarSeq.tsv data file
• Separates indexes into +N/-N, plant/soil, time0 groups
• Will need to edit provided file to fit our data and satisfy essential conditions above
combineBarseq.pl
• Merges table of counts with pool file to create strain table setname.poolcount
• Matches barcode/strain info from pool file to counts from MultiCodes to find how often each strain was seen in each experiment
BarseqR.pl
• Requires dir containing setname.poolcount, genes.GC, and metadata file (FEBA_BarSeq.tsv)
• Maps barcode counts : strains : genes, combines into table all.poolcount
• Computes gene fitness values for each exp from all.poolcount and metadata, creates R image of results and QA plots

Glossary

- Genes table
  - From SetupOrg.pl with Sm1021 .gbk file
  - Fields: locusId, sysName, scaffoldId, begin, end, strand
  - Whole genome
  - Provided already
- Genes.GC
  - From RegionGC.pl with genes table
  - Adds GC and nTA (# TA dinucleotides) fields to genes table
  - Provided already
- Pool file
  - From DesignRandomPool.pl with genes table, processed TnSeq data
  - Fields: barcode, rcbarcode, nTot (total reads), n1, scaffold/strand/pos, n2, scaffold2/strand2/pos2, pastEnd
  - Describes deletion mutants with consistent BC insertion into a protein-coding gene; accounts for 1st and 2nd most common insertion locations; checks for quality and designs a randomized starting pool from these data
  - Provided already
- Table of counts
  - From MultiCodes.pl with **raw BarSeq data**
  - Finds and counts the barcode sequence in each read, totals how many times each seq was seen
- Setname.poolcount
  - From combineBarseq.pl given table of counts, pool file
  - Combines pool file with counts table to show how often each strain from the pool was seen in each treatment; also records total count
- Experiment metadata file FEBA_BarSeq.tsv
  - Fields: SetName, Index, Description, and Date_pool_expt_started
    - SetName describes the whole Smeli experiment; the same as in setname.poolcount
  - Other fields optional; script only requires these fields to run
  - Designates each index by experimental treatment/time0; description should be eg Lib +N +p, Lib –N –p, time0
- All.poolcount
  - From BarSeqR.pl with setname.poolcount, metadata file, genes.GC
  - Combines genes table with setname.poolcount to describe the inactivated genes harboring each barcode
  - BarSeqR.pl calls the R script RunFEBA.R to calculate gene-level fitness, experiment quality plots, etc. from all.poolcount
  - RunFEBA.R generates **R image** containing all results, biggest cofitness hits, and tools for further analyses