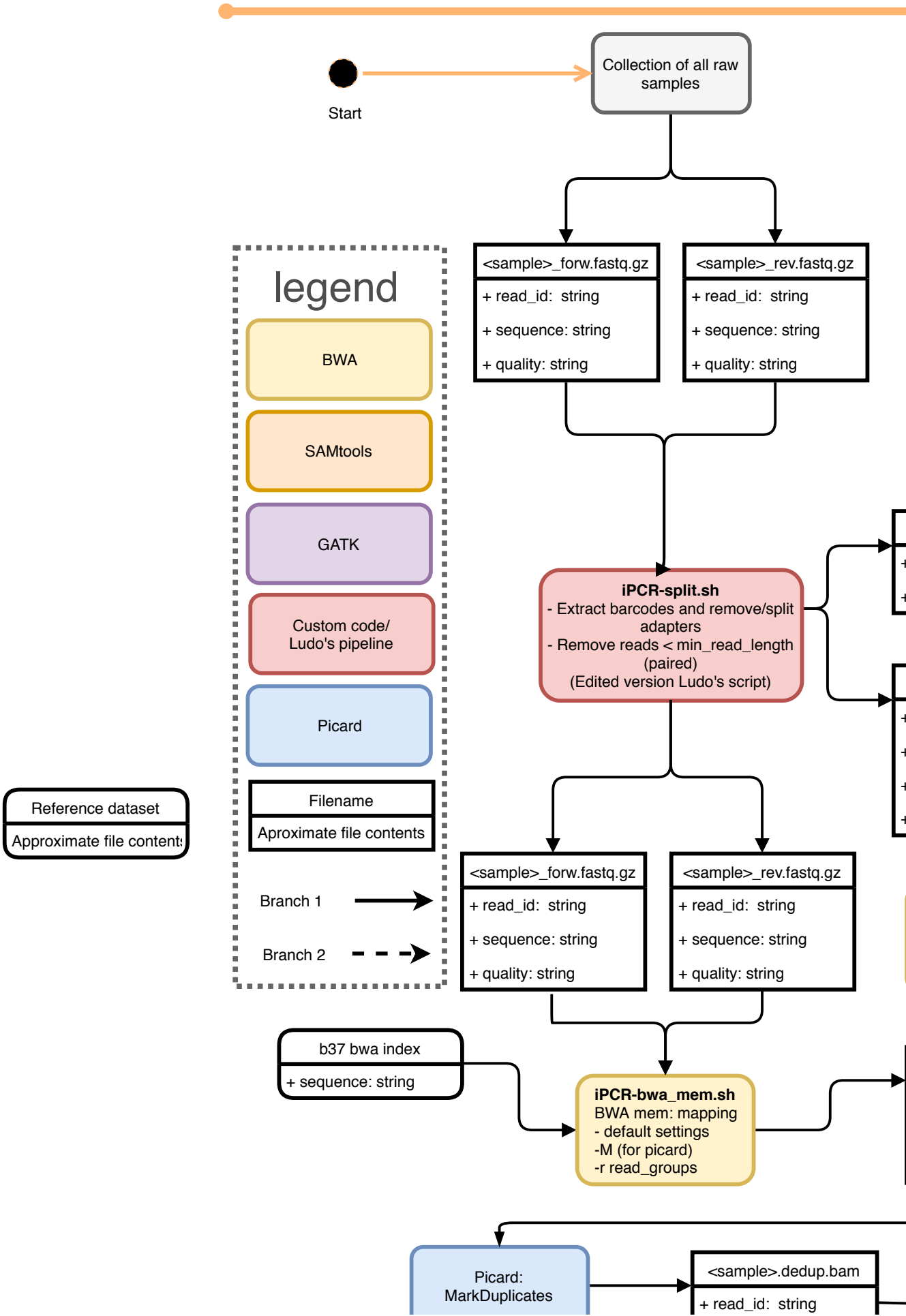
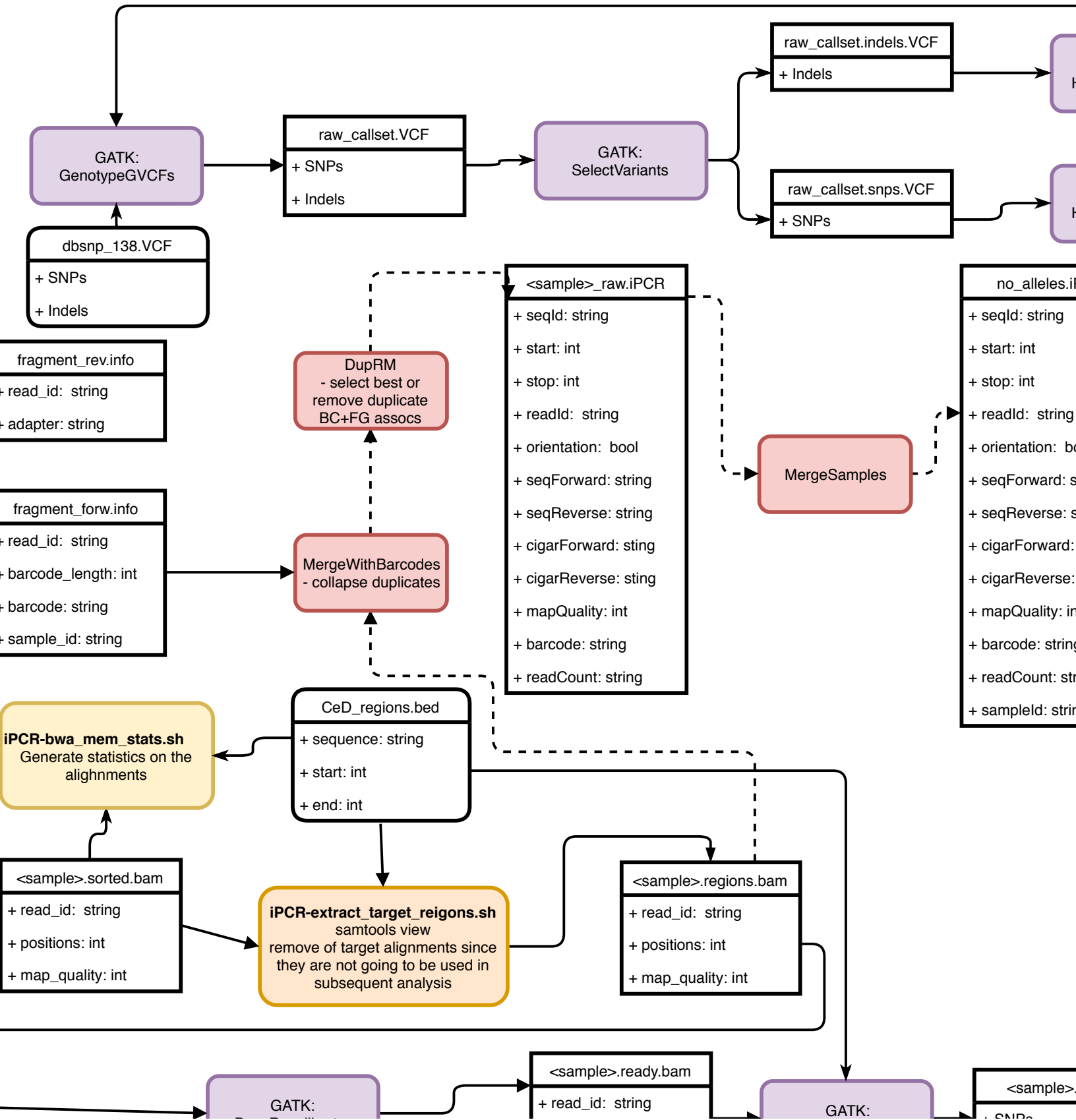
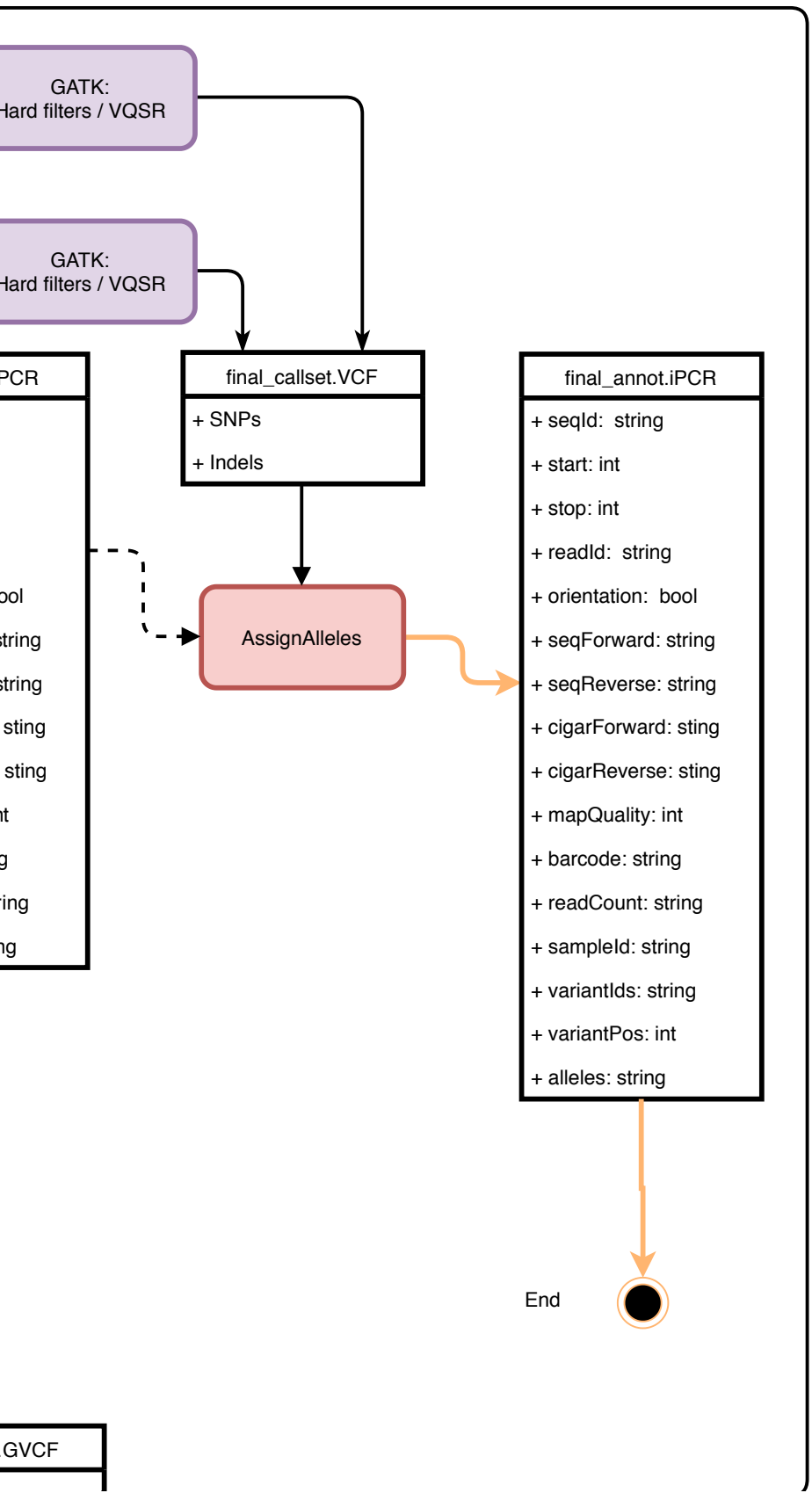
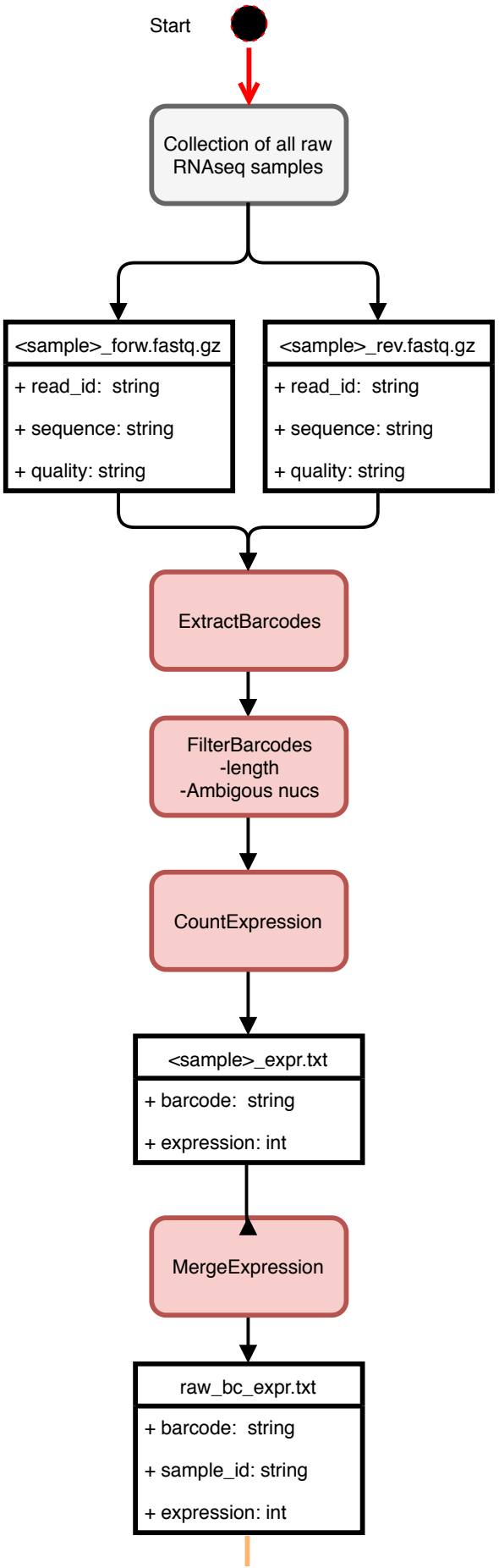

Start
Collection of all raw samples
<sample>_forw.fastq.gz
+ read_id: string
+ sequence: string
+ quality: string
<sample>_rev.fastq.gz
+ read_id: string
+ sequence: string
+ quality: string
legend
BWA
SAMtools
GATK
Custom code/ Ludo's pipeline
Picard
Filename
Aproximate file contents
Branch 1
Branch 2
Reference dataset
Approximate file content
iPCR-split.sh
- Extract barcodes and remove/split adapters
- Remove reads < min_read_length (paired)
(Edited version Ludo's script)
<sample>_forw.fastq.gz
+ read_id: string
+ sequence: string
+ quality: string
<sample>_rev.fastq.gz
+ read_id: string
+ sequence: string
+ quality: string
b37 bwa index
+ sequence: string
iPCR-bwa_mem.sh
BWA mem: mapping
- default settings
-M (for picard)
-r read_groups
Picard: MarkDuplicates
<sample>.dedup.bam
+ read_id: string

iPCR pipeline



cDNA


GATK:
Hard filters / VQSR
GATK:
Hard filters / VQSR
PCR
ool
tring
string
sting
sting
nt
g
ring
ng
final_callset.VCF
+ SNPs
+ Indels
AssignAlleles
final_annot.iPCR
+ seqId: string
+ start: int
+ stop: int
+ readId: string
+ orientation: bool
+ seqForward: string
+ seqReverse: string
+ cigarForward: sting
+ cigarReverse: sting
+ mapQuality: int
+ barcode: string
+ readCount: string
+ sampleId: string
+ variantIds: string
+ variantPos: int
+ alleles: string
End
GVCF


Start
Collection of all raw RNAseq samples
<sample>_forw.fastq.gz
+ read_id: string
+ sequence: string
+ quality: string
<sample>_rev.fastq.gz
+ read_id: string
+ sequence: string
+ quality: string
ExtractBarcodes
FilterBarcodes
-length
-Ambigous nucs
CountExpression
<sample>_expr.txt
+ barcode: string
+ expression: int
MergeExpression
raw_bc_expr.txt
+ barcode: string
+ sample_id: string
+ expression: int

+ positions: int

+ map_quality: int

Marking of duplicates may not be ideal since the complexity of the library is limeted due to the enrichment of target fragments. For more detaills: https://gatkforums.broadinstitute.org/gatk/discussion/6747/how-to-mark-duplicates-with-markduplicates-or-markduplicateswithmatecigar


BaseRecalibrator
+ positions: int
+ map_quality: int
HaplotypeCaller
+ SNPs
+ Indels
dbsnp_138.VCF
+ SNPs
+ Indels
b37.fa
+ sequence: string


End