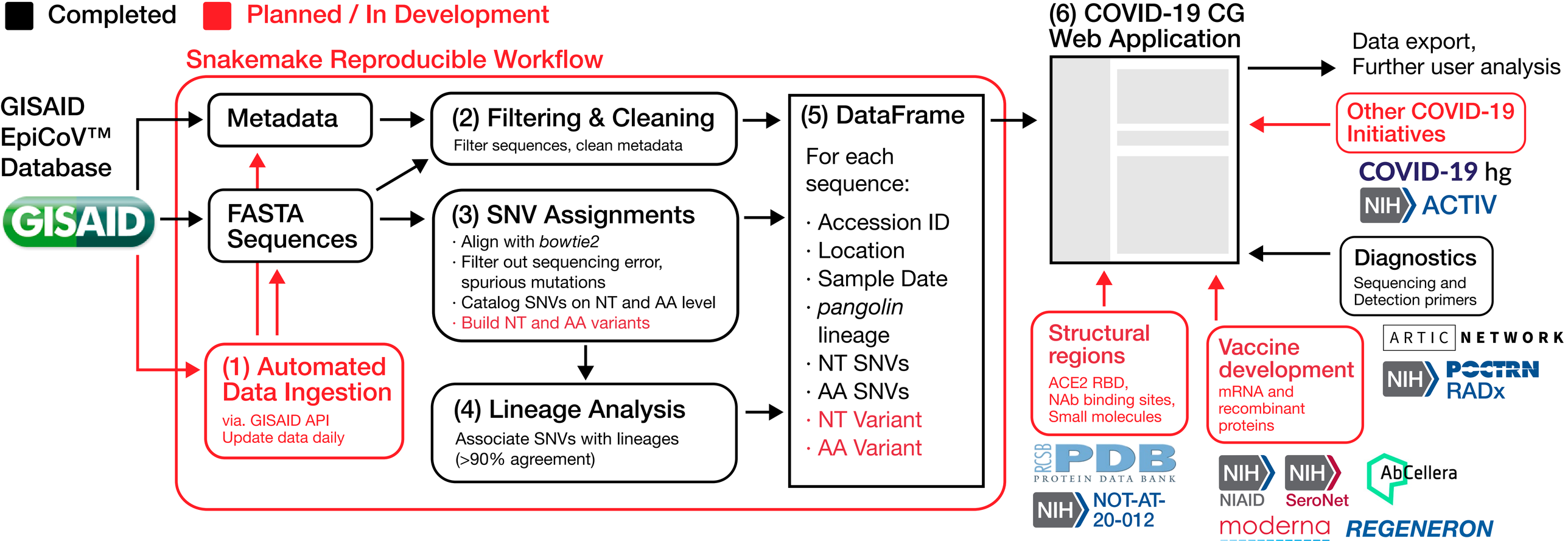
Completed
Planned / In Development
Snakemake Reproducible Workflow
GISAID EpiCoV™ Database
GISAID
Metadata
FASTA Sequences
(1) Automated Data Ingestion
via. GISAID API
Update data daily
(2) Filtering & Cleaning
Filter sequences, clean metadata
(3) SNV Assignments
· Align with bowtie2
· Filter out sequencing error, spurious mutations
· Catalog SNVs on NT and AA level
· Build NT and AA variants
(4) Lineage Analysis
Associate SNVs with lineages (>90% agreement)
(5) DataFrame
For each sequence:
· Accession ID
· Location
· Sample Date
· pangolin lineage
· NT SNVs
· AA SNVs
· NT Variant
· AA Variant
(6) COVID-19 CG Web Application
Data export, Further user analysis
Other COVID-19 Initiatives
COVID-19 hg
NIH ACTIV
Diagnostics
Sequencing and Detection primers
ARTIC NETWORK
NIH POETRN RADx
Structural regions
ACE2 RBD, NAb binding sites, Small molecules
RCSB PDB PROTEIN DATA BANK
NIH NOT-AT-20-012
Vaccine development
mRNA and recombinant proteins
NIH NIAID
NIH SeroNet
AbCellera
moderna
REGENERON