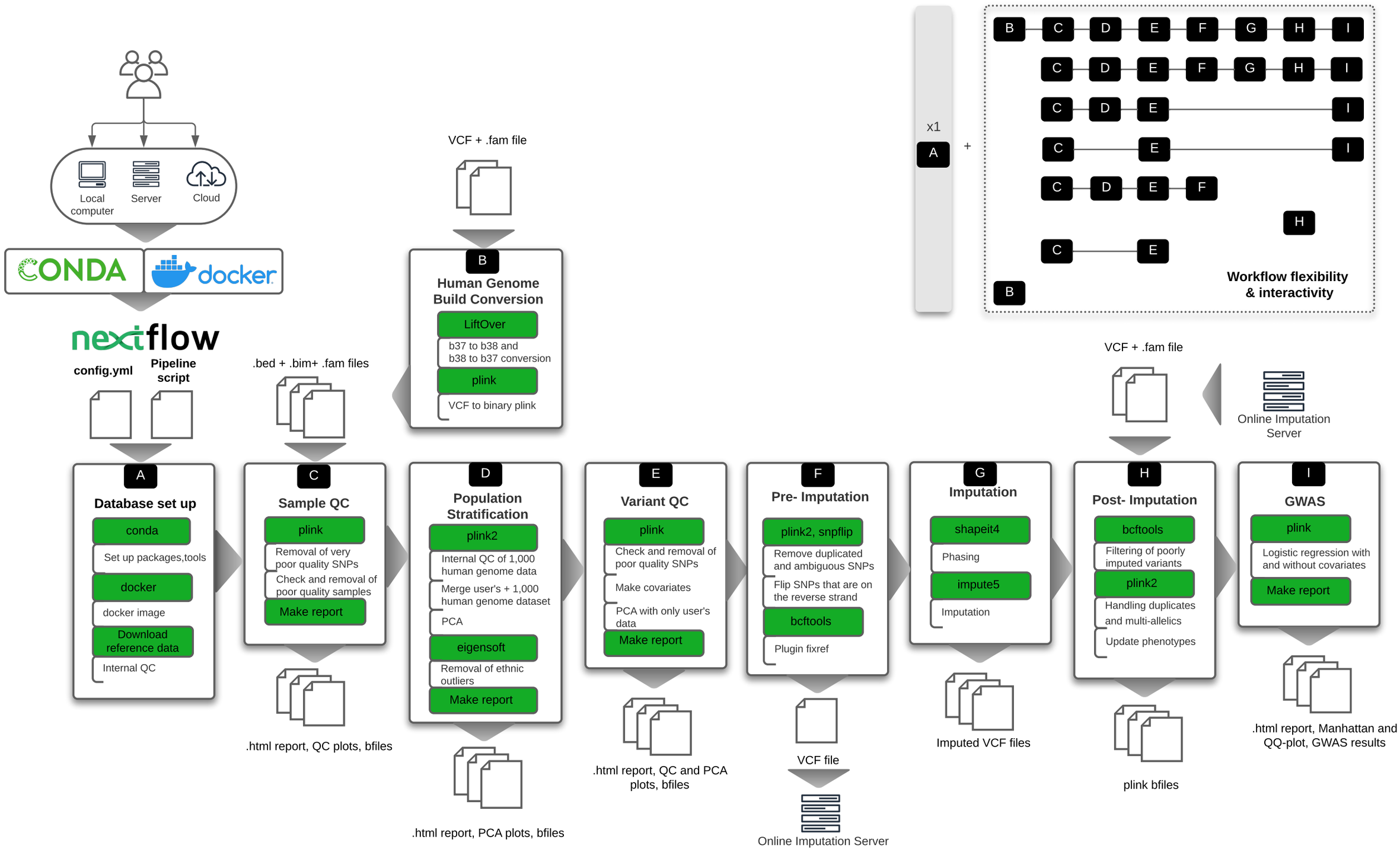
Local computer
Server
Cloud
CONDA
docker
nextflow
config.yml
Pipeline script
A
Database set up
conda
Set up packages,tools
docker
docker image
Download reference data
Internal QC
.bed + .bim+ .fam files
C
Sample QC
plink
Removal of very poor quality SNPs
Check and removal of poor quality samples
Make report
.html report, QC plots, bfiles
VCF + .fam file
B
Human Genome Build Conversion
LiftOver
b37 to b38 and b38 to b37 conversion
plink
VCF to binary plink
D
Population Stratification
plink2
Internal QC of 1,000 human genome data
Merge user's + 1,000 human genome dataset
PCA
eigensoft
Removal of ethnic outliers
Make report
.html report, PCA plots, bfiles
E
Variant QC
plink
Check and removal of poor quality SNPs
Make covariates
PCA with only user's data
Make report
.html report, QC and PCA plots, bfiles
F
Pre- Imputation
plink2, snpflip
Remove duplicated and ambiguous SNPs
Flip SNPs that are on the reverse strand
bcftools
Plugin fixref
VCF file
Online Imputation Server
G
Imputation
shapeit4
Phasing
impute5
Imputation
Imputed VCF files
VCF + .fam file
Online Imputation Server
H
Post- Imputation
bcftools
Filtering of poorly imputed variants
plink2
Handling duplicates and multi-allelics
Update phenotypes
plink bfiles
I
GWAS
plink
Logistic regression with and without covariates
Make report
.html report, Manhattan and QQ-plot, GWAS results
x1
A
+
B C D E F G H I
C D E F G H I
C D E I
C E I
C D E F
H
C E
B
Workflow flexibility & interactivity