Custom Perl data analysis workflow

Python

Galaxy

Perl

| Remove seqs. that do not begin with CTTCA | → | Map reads with Bowtie | → | Filter seq errors / Calculate read count and breakpoints |

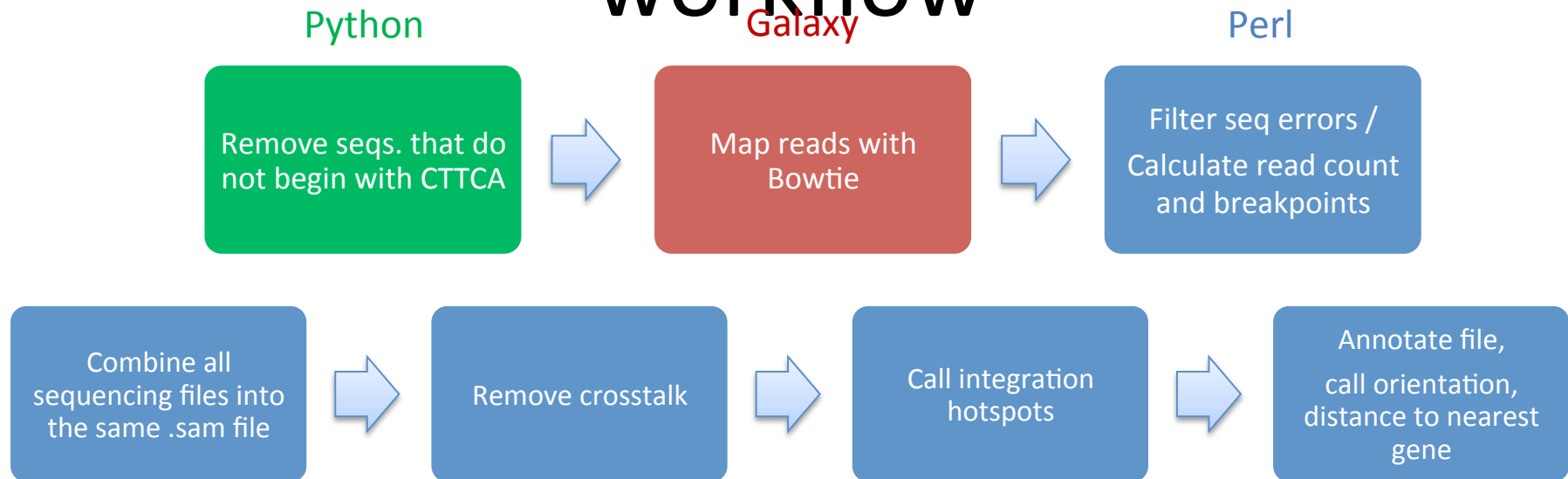| Combine all sequencing files into the same .sam file | → | Remove crosstalk | → | Call integration hotspots | → | Annotate file, call orientation, distance to nearest gene |

Output for each integration:
- Location of integrations
- Distance to nearest gene
- Orientation of virus with respect to nearest gene
- Number of breakpoints for each integration
- Number of reads for each integration
- Call integration hotspots and common integration sites