Bio-IT hackathon

Method Optimization for NEAT WDL

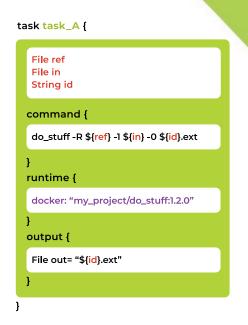Diversifying BAMSurgeon WDL to add functionality

Method Optimization

- Generating synthetic reads with 1000 genomes and additional parameters to provide even coverage, etc.
- This tool is meant to support FAIR principles
- Decreasing the cost will aid in adoption

```
workflow myWorkflowName {

    File my_ref
    File my_input
    String name

    call task_A {
        input: ref= my_ref, in= my_input, id= name
    }
    call task_B {
        input: ref= my_ref, in= task_A.out
    }
}

task task_A {              ...        }
task task_B {              ...        }
```

```
task task_A {

    File ref
    File in
    String id

    command {
        do_stuff -R ${ref} -1 ${in} -0 ${id}.ext
    }
    runtime {
        docker: "my_project/do_stuff:1.2.0"
    }
    output {
        File out= "${id}.ext"
    }
}
```

Runtime and cost

Workflow Name	Cost for 1 sample (range)	Cost for 100 samples (total)	Time to run 1 sample	Time to run 100 samples in parallel (wallclock)
Collect1000GParticipant	$1.64 to $2.90	$193.75	4.5 hours	12 hours
GenerateSyntheticReads	$2.40 to $3.44	$405.67	4.5 hours	12 hours
MutateReadsWithBAMSurgeon	$0.02 to $0.15	$5.72	.5 hours	2.5 hours

Diversifying BamSurgeon options

- A WDL uses Bamsurgeon for spiking variants
- Currently the WDL only supports SNVs
- Adding INDELs & SVs