Sequencing DRT

2019-09

Who we are

HIVE - Matt Ruffalo
TMC-Florida - Maigan Brusko
TMC-UCSD - Blue Lake
TMC-UW - Hannah Pliner
TMC-Stanford - Aaron Horning
TMC-Stanford - Amir Bahmani

What is our goal?

- Establish technology-specific analytic pipelines
  - Custom demux for many assays
- Work with HIVE group to run analysis steps on IEC resources
  - Provide tools/analysis scripts for HIVE to containerize and wrap in workflow description language (for tools/pipelines not already containerized)
- Standardize analysis steps between assays when appropriate
  - Possibilities: short read alignment, quantification, normalization
- Define QC metrics
  - Assay-specific and more general (e.g. FastQC for all sequencing reads)
- Define minimum standards of data release

What we have done so far

Goal	% Done
Identify Assays & Centers & Reps	100%
Specific workflow for assay at each Center	~80%
Definition of data levels	80-90%, assay-specific
File formats defined	40%: defined for raw data (FASTQ), not for secondary analysis
Assay metadata & file format defined	10%
Processing pipeline defined	~70%
Identify potentially common processing steps & who/where it will run.	~50%; some common steps identified (e.g. alignment, quantification, normalization), but no consensus yet on specifics of standardization
Assay & data QA/QC criteria	40%: FastQC for reads uploaded to HIVE, unsure about assay-specific QA/QC
Understand how to upload data/metadata to HIVE	For pipeline development/testing: 100% (Globus), final production upload: 5%?
Understand how to validate and transfer processing pipeline to HIVE.	90%

What we hope to achieve today

- Points of contact between TMCs and HIVE members
- Extend metadata standards to include assay-specific information (sequencing hardware, reagents, anything appropriate)
- Consensus about standardized analysis steps (e.g. alignment tool, reference genome, quantification)