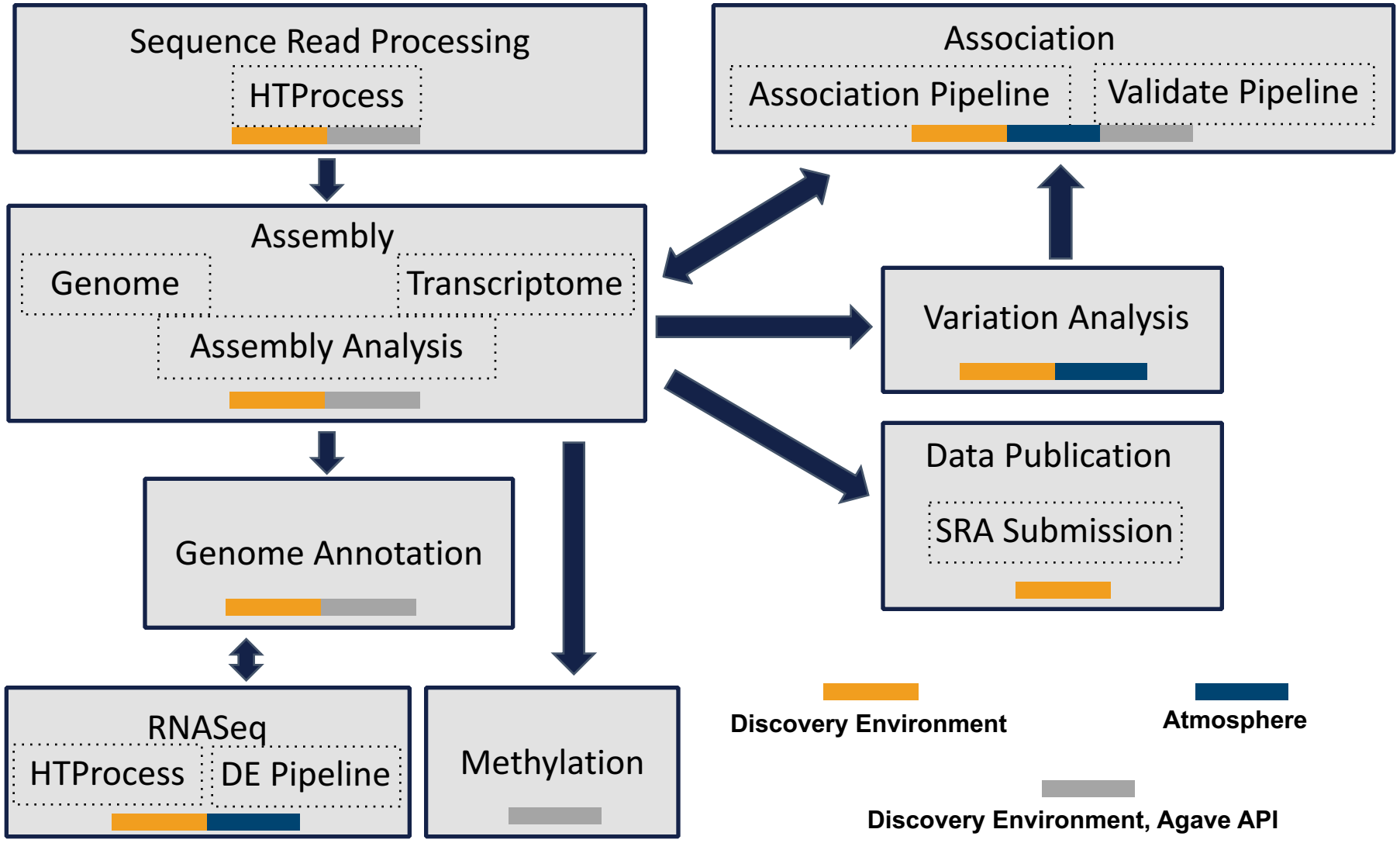
CYVERSE™

Transforming Science Through Data-driven Discovery

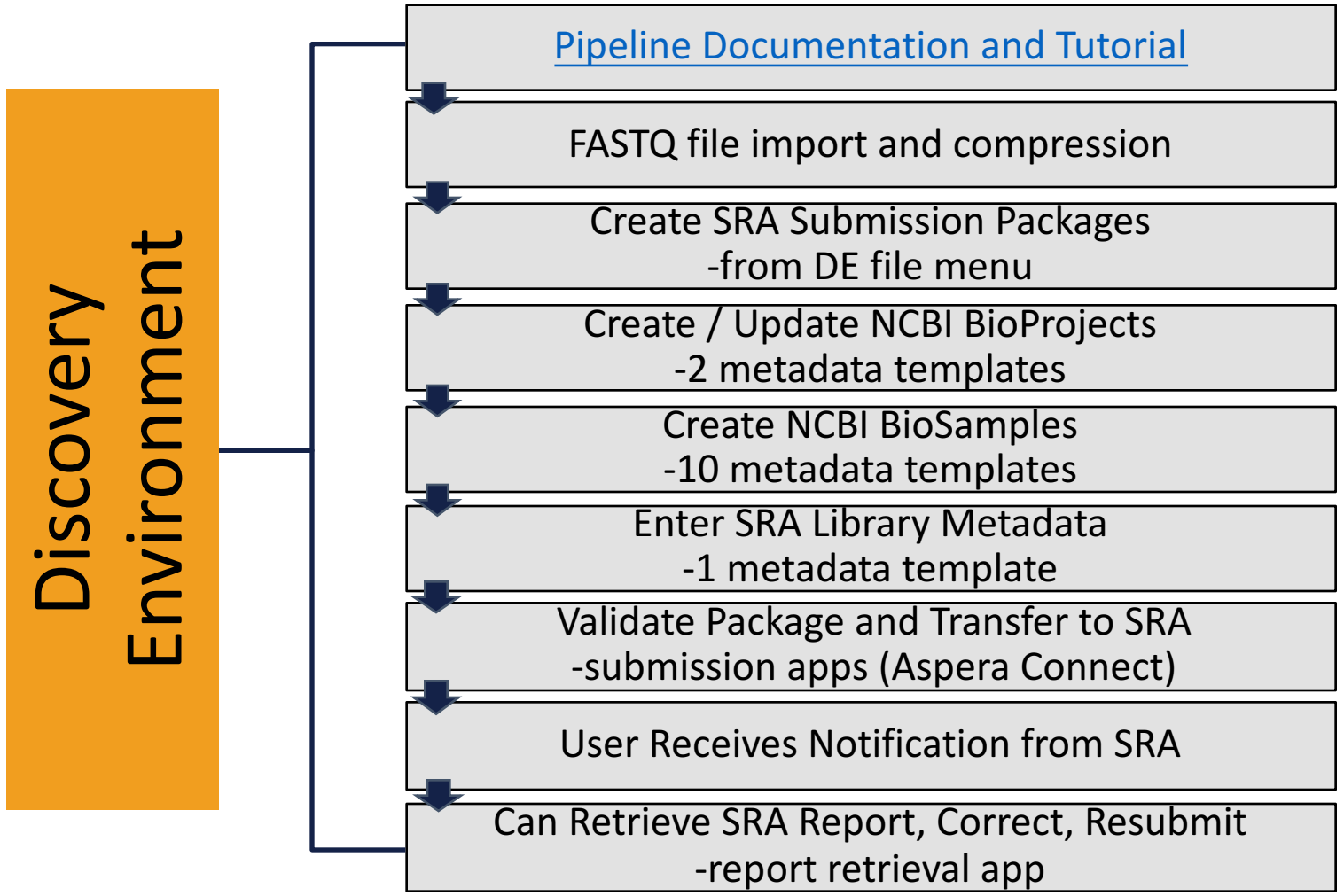
Workflow Overviews

Fall 2016

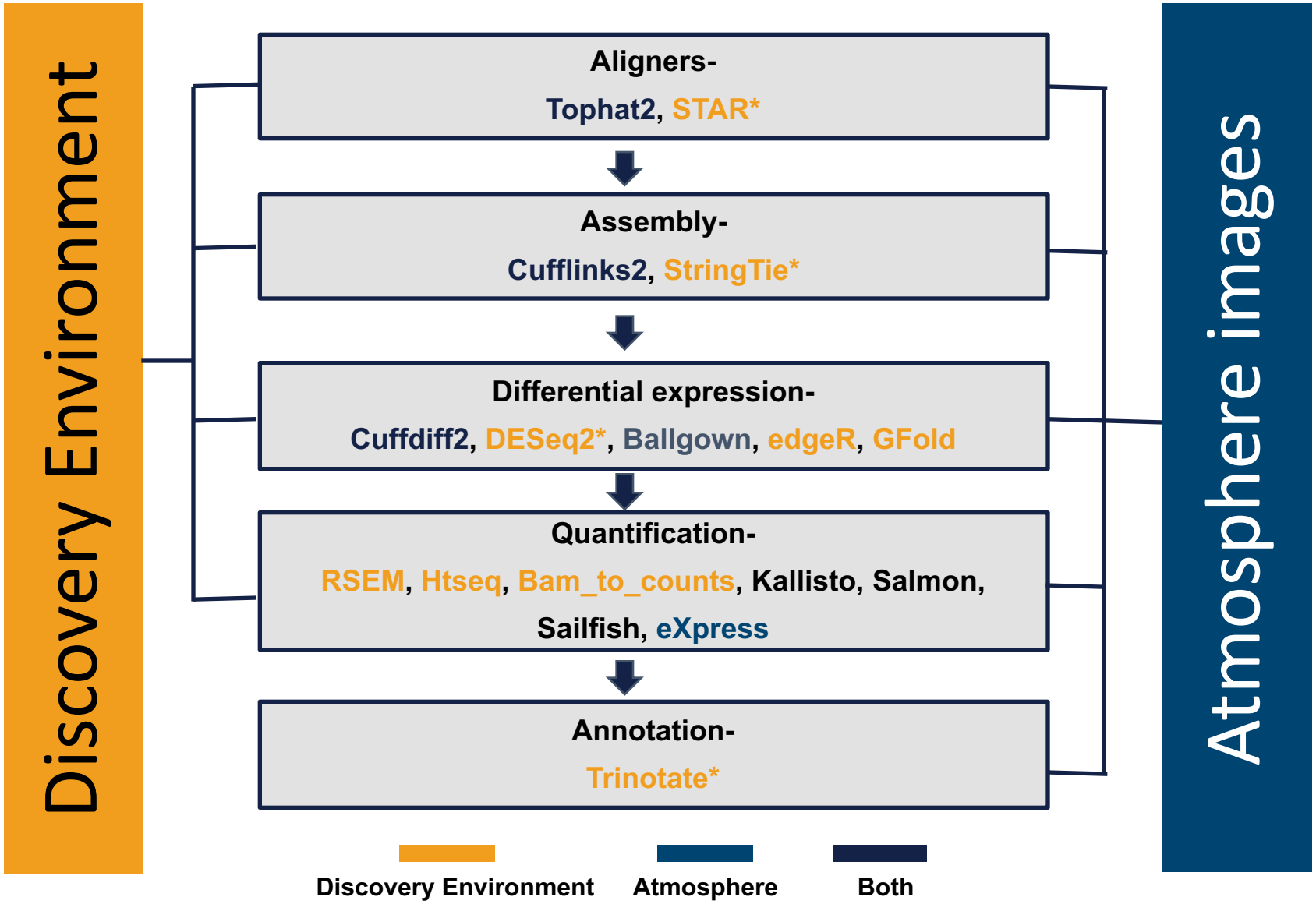
Overview of Genomics Workflows



Data Dissemination: NCBI SRA Submission Pipeline

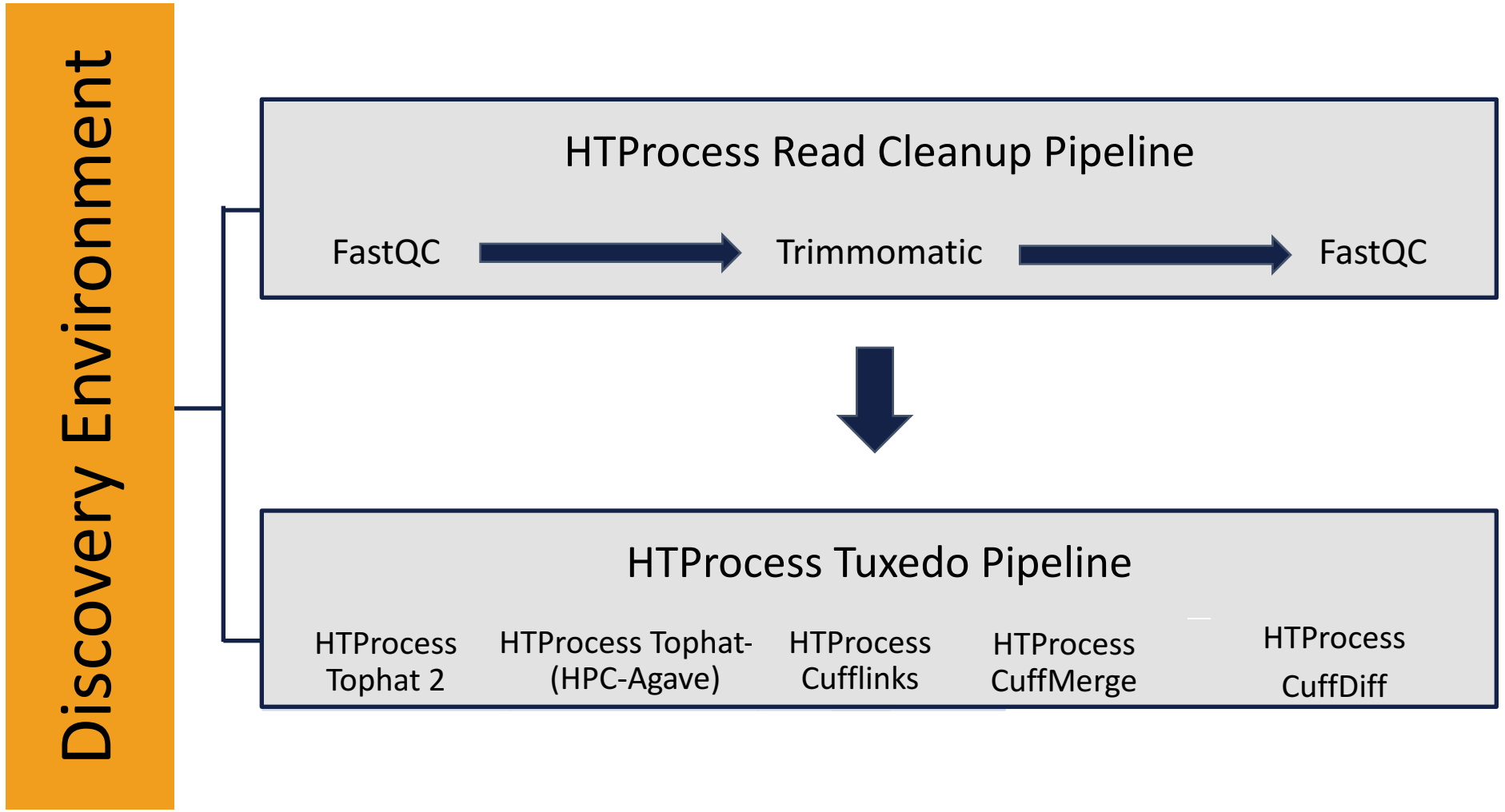


RNA Seq 1 for Differential Expression

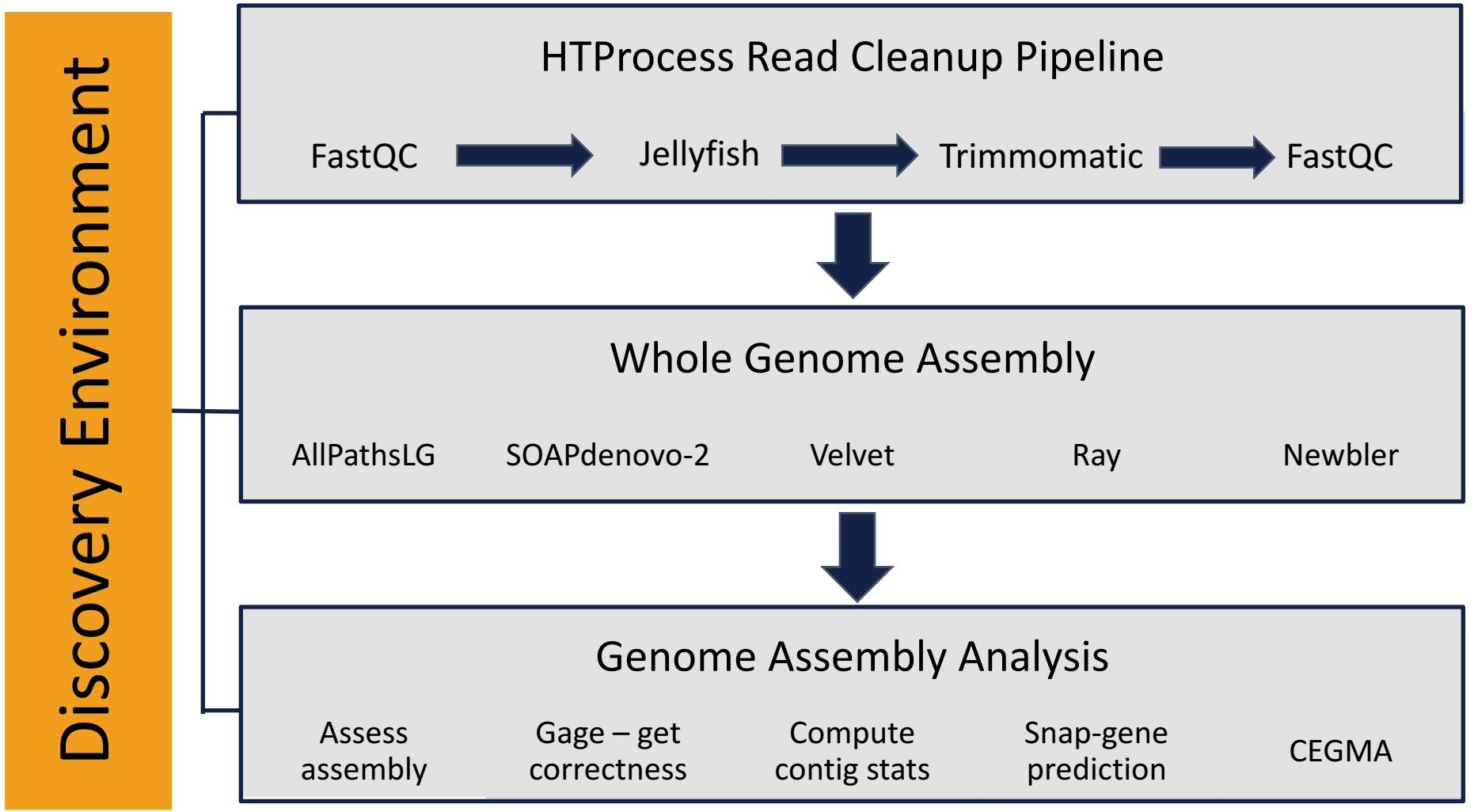


RNA seq 2: High Throughput Process Apps

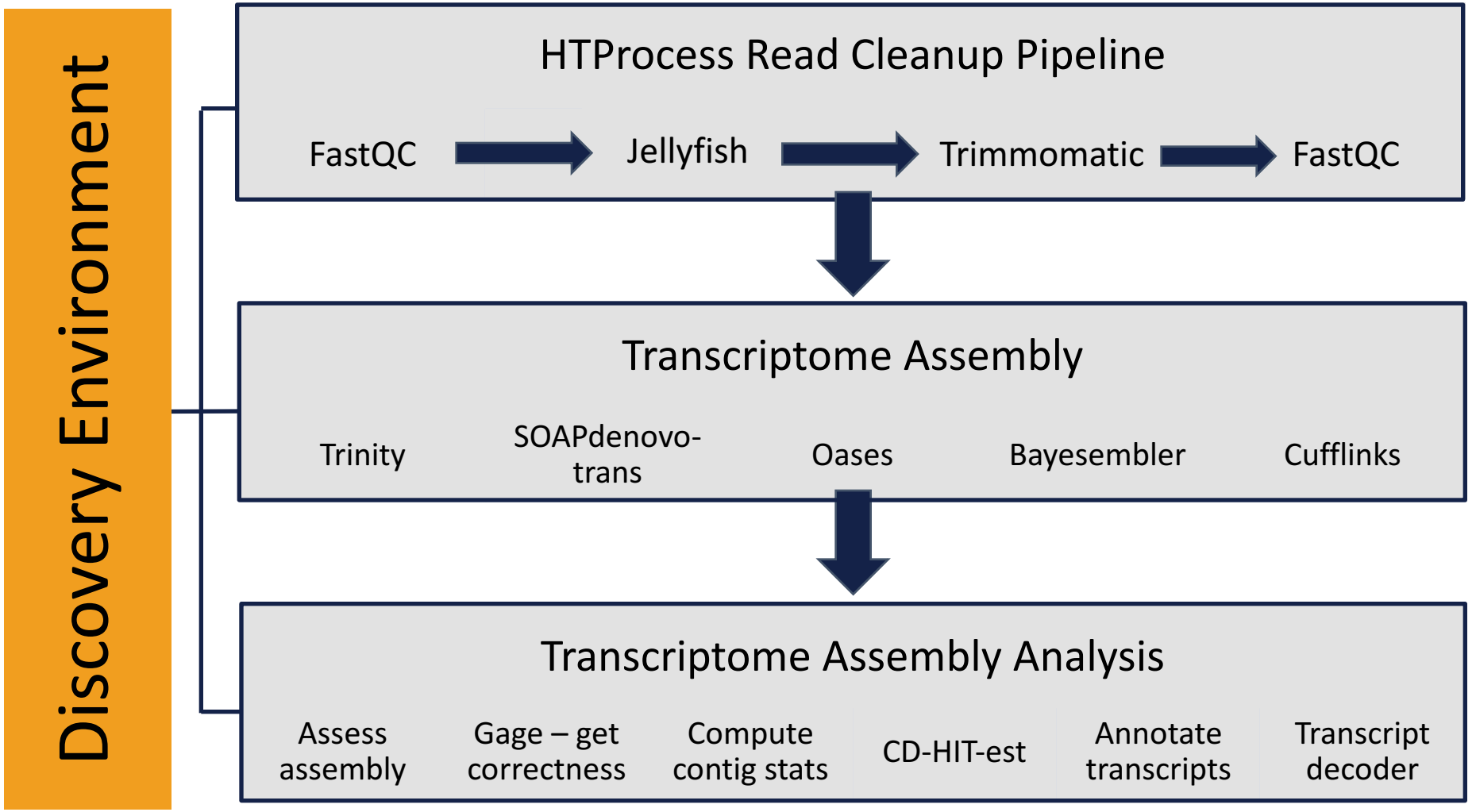
For handling large groups of data and easier workflow management.
Files are managed as a group or library contained in a single directory.



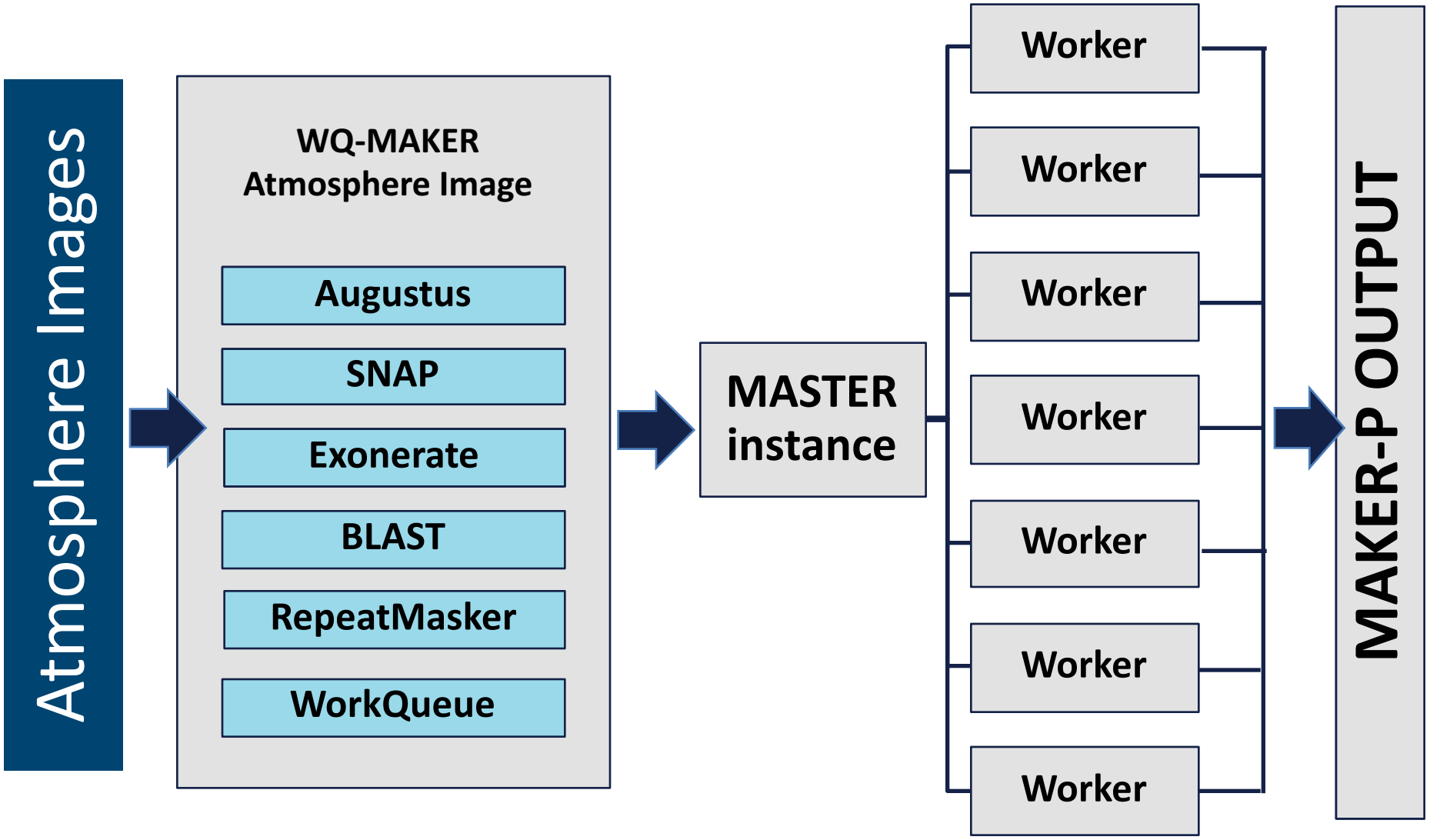
Genome Assembly



Transcriptome Assembly

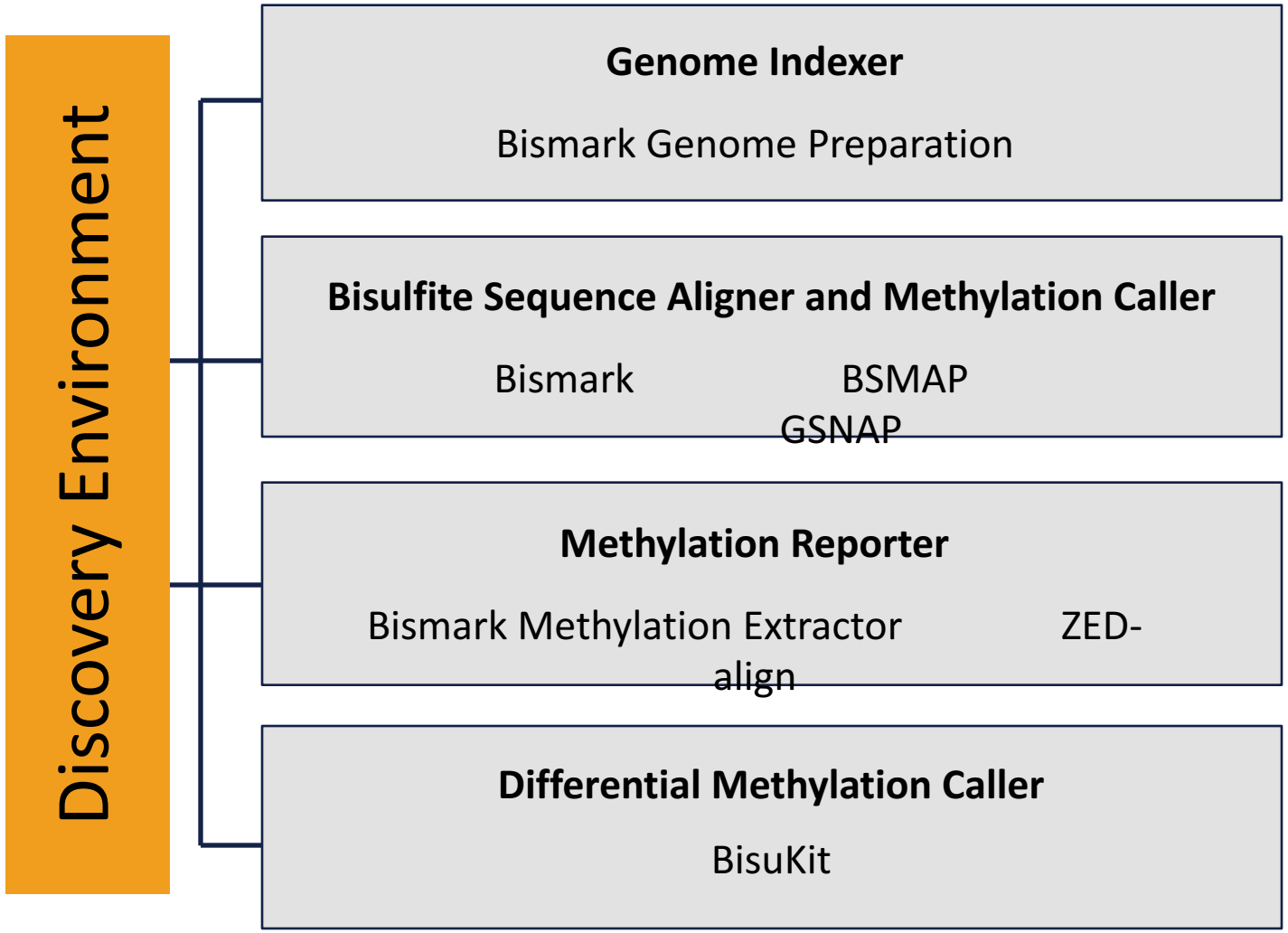


Genome Annotation: WQ-MAKER in Atmosphere

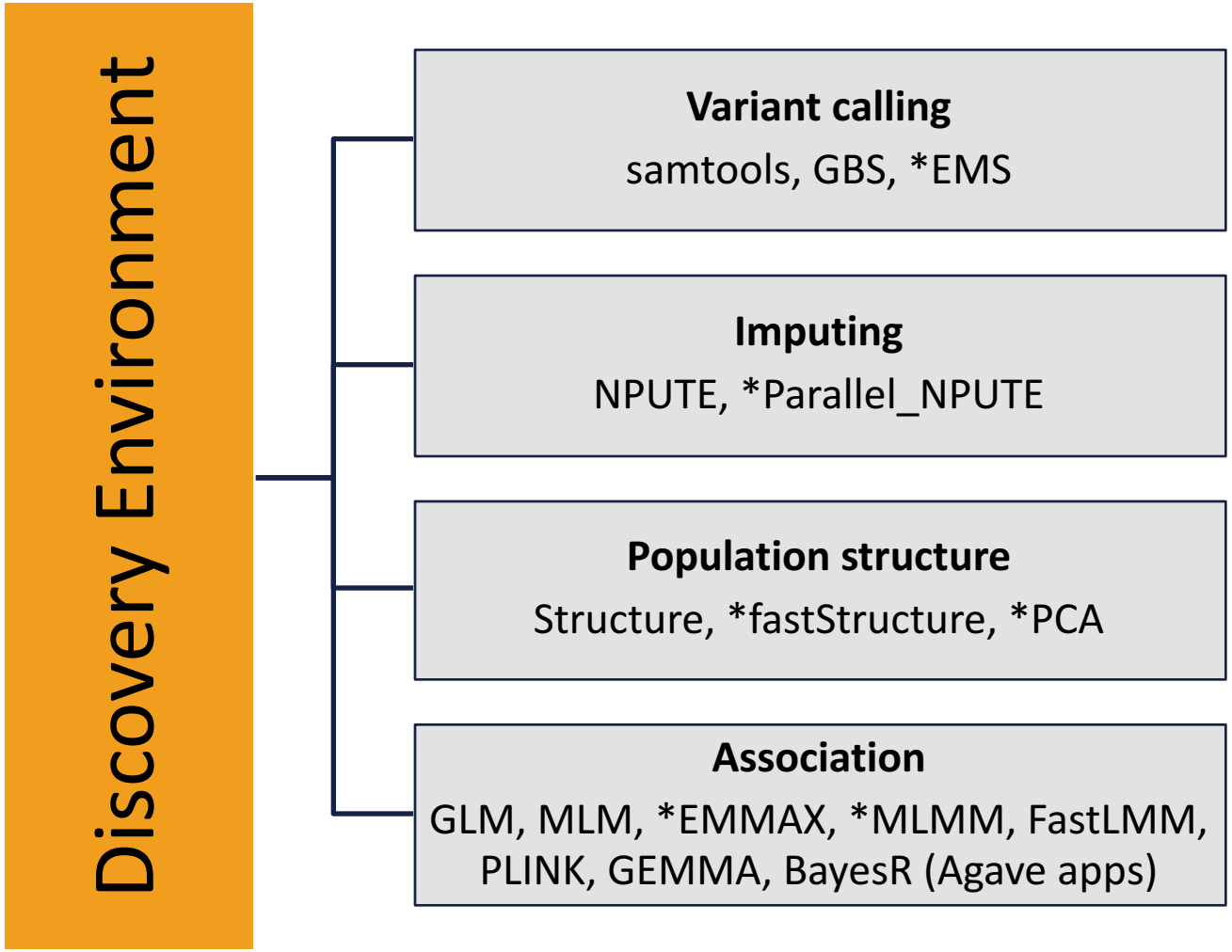


In collaboration with the Douglas Thain lab (http://www3.nd.edu/~dthain/)

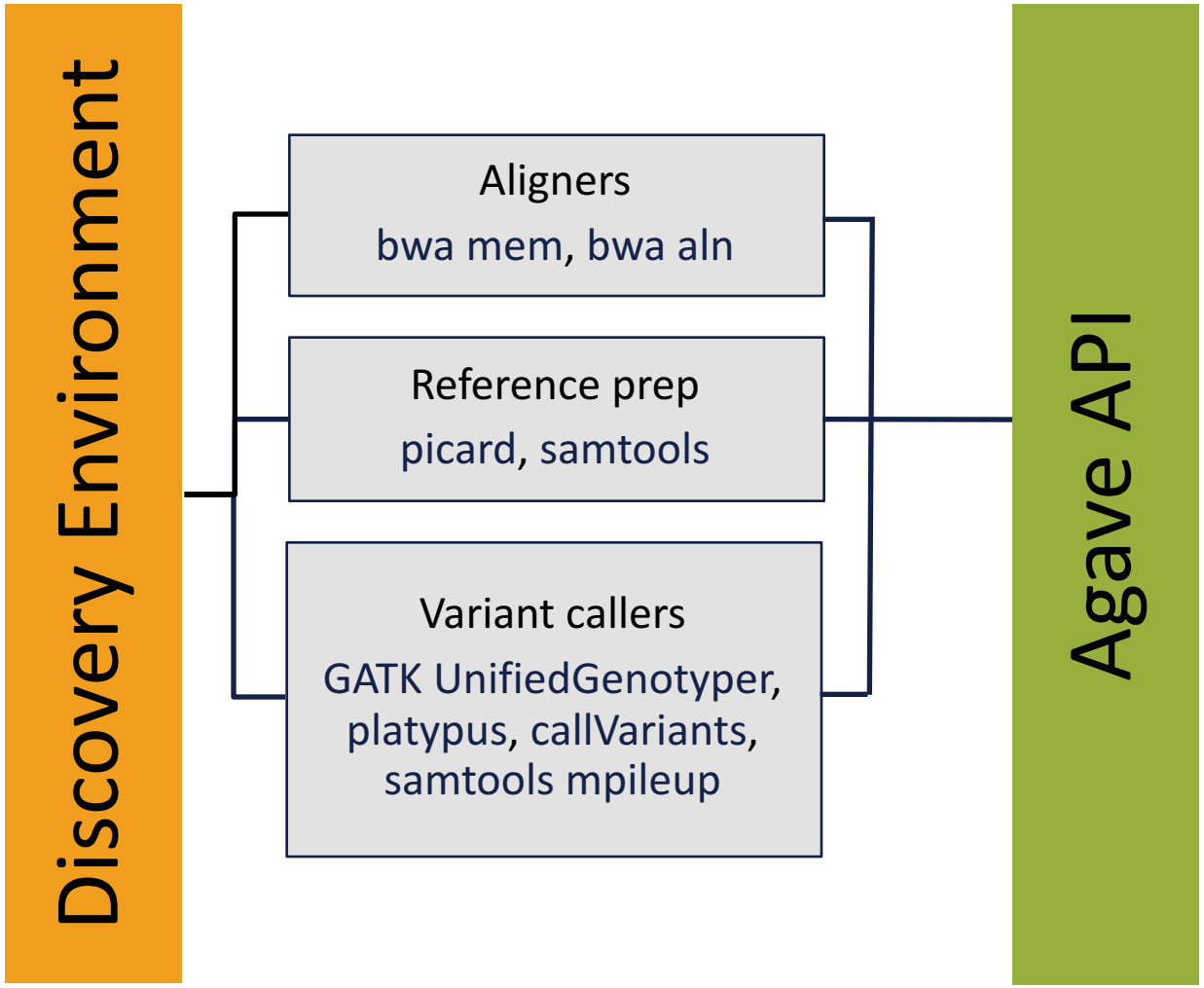
DNA Methylation Analysis



Association Analysis

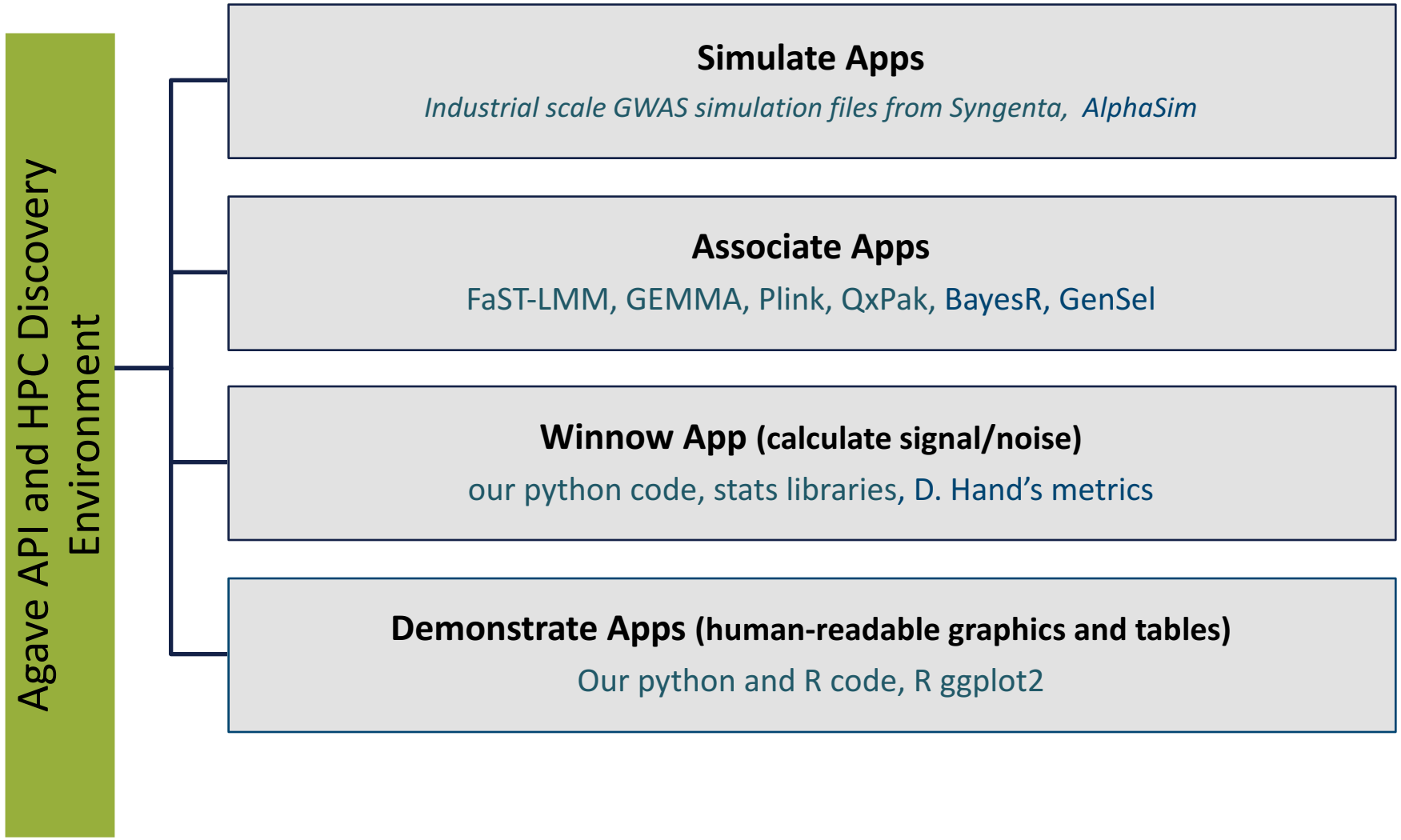


Variant Caller Pipeline



Validate Pipeline

-extensible, scalable testing of tool accuracy and precision



Genomics Workflows

Quick Reference Guide

Workflow	Platform	Limits
Genome Assembly	DE, Agave	--
Transcriptome Assembly	DE, Agave	48 hrs run time max
Genome Annotation WQ-MAKER	Atmo	--
RNA Seq	DE, Atmo	--
HT Process	DE, Agave	150 GB input data max
Methylation Analysis	DE, Agave	--
Association Analysis	DE, Agave	--
Variant Caller	DE, Agave	48 hrs run time max
Validate	Atmo, Agave	--
SRA Submission	DE	--

DE - Discovery Environment
Atmo - Atmosphere
Agave - Agave API