# A fast and efficient bioinformatics analysis workflow for processing reads from single-cell multiomics assays captured on a microwell-based platform

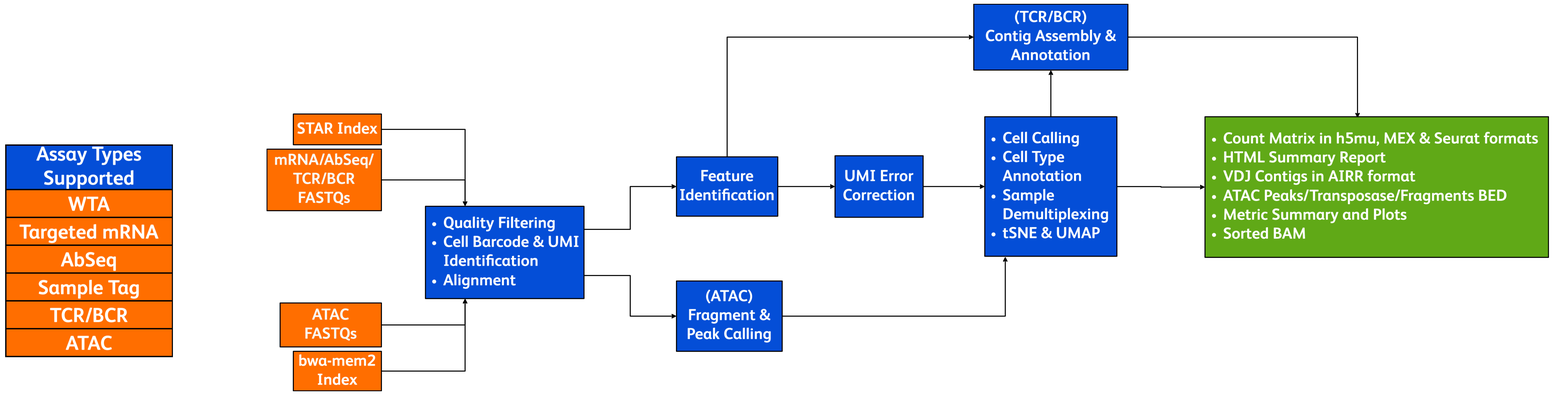
Raghavendra Padmanabhan[1], Brent Weichel[1], Amie Radenbaugh[1], Yuefu Jiang[1], Charles Weeks[1], Thomas McCarthy[1], Youngsook Kim[1], Anthony Berno[1], Devon Jensen[1].
[1]*BD Biosciences, 2350 Qume Drive, San Jose, CA 95131*

**Poster#** 4956

## Abstract

Here, we present a significant update of the BD Rhapsody™ Sequence Analysis Pipeline, version 2—a comprehensive primary analysis workflow that is up to 7X faster and consumes 2X less disk space than the previous release. We used it to process a library with 20,000 human PBMCs sequenced to a depth of 1.2 billion reads in 2.5 hours of wall-clock time. The pipeline can be used with reads from single-cell whole transcriptome, targeted mRNA, surface antigens (AbSeq), TCR/BCR, ATAC-seq and Sample Tag libraries captured on the BD Rhapsody™ System platform. The major processing steps include Quality Filtering, Cell Barcode Identification, Read Alignment, Feature Assignment, Fragment & Peak calling, UMI Error Correction, Identification of Putative Cells, Sample De-Multiplexing, Dimensionality Reduction for Visualization and VDJ Contig Assembly & Annotation. Output files and metrics are available in easy-to-digest formats, including an html report with dynamic visualization. Integrated outputs in Seurat and Scanpy formats combine expression matrices and all cell annotation metadata (e.g., predicted cell type, sample assignment, TCR/BCR sequence and gene segments). These pre-generated files are ready to load into popular single-cell analysis tools. The pipeline uses CWL as the workflow manager and makes use of custom code written in C++ and Python along with various open-source packages for data processing. The pipeline can also process reads from non-human species and includes built-in support for specifying and building custom reference genome indices. In support of high-throughput multiomic discovery studies enabled by the BD Rhapsody™ HT Xpress System, this pipeline has been tested with datasets containing more than 800,000 putative cells and 14.5 billion reads.

## Workflow Overview




## Performance Benchmarks

### Comparison to previous pipeline release

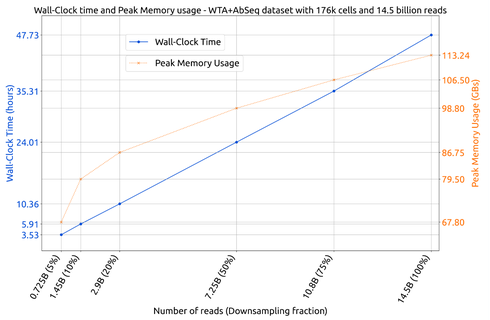






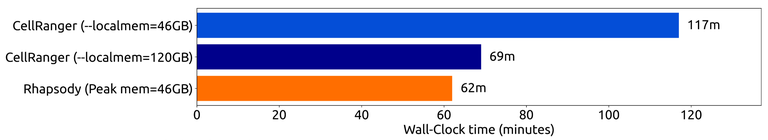
### Runtime and Memory Usage

| Dataset | Wall-Clock Time (32 threads) | Peak Memory Usage |
|---|---|---|
| WTA<br>20k cells. 1.2B reads | 2h19m | 43GB |
| WTA+TCR/BCR+AbSeq<br>10k cells, 422M reads | 7h18m | 60GB |
| WTA+AbSeq<br>176k cells, 14.5B reads | 47h43m | 113GB |
| Targeted<br>800k cells, 1B reads | 3h44m | 39GB |
| ATAC<br>15k cells, 590M reads | 5h50m | 70GB |




Note: All datasets were run on a m6i.8xlarge AWS instance with 32 CPUs, 128GB RAM and a 4 TB SSD EBS disk

### Comparison b/w Rhapsody 2.2 and CellRanger 7.2.0
### Dataset: 10k PBMCs, 496M WTA reads




Both programs were run with 32 threads

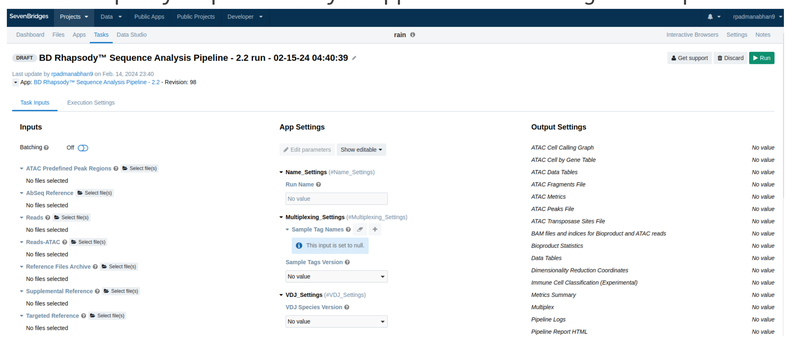
10X dataset: https://www.10xgenomics.com/resources/datasets/10k-human-pbmcs-3-v3-1-chromium-x-with-intronic-reads-3-1-high

Rhapsody dataset: https://bd-rhapsody-public.s3.amazonaws.com/Rhapsody-Demo-Data-Inputs/09WTA-ABC-EB-10kPBMC.tar (09-WTA library downsampled to 496M reads)

## Convenient deployment options

1. SevenBridges cloud platform
2. Local deployment using Docker container
3. Local deployment using Dockerfree bundle on Ubuntu 16.04+ / CentOS 8+ / RHEL 7+

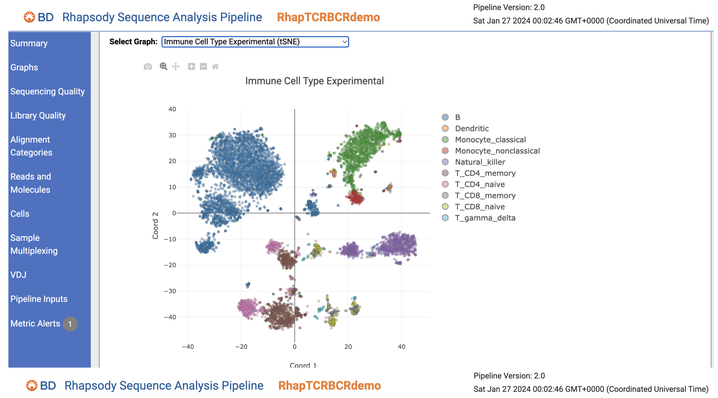
### BD Rhapsody Sequence Analysis App on the SevenBridges cloud platform

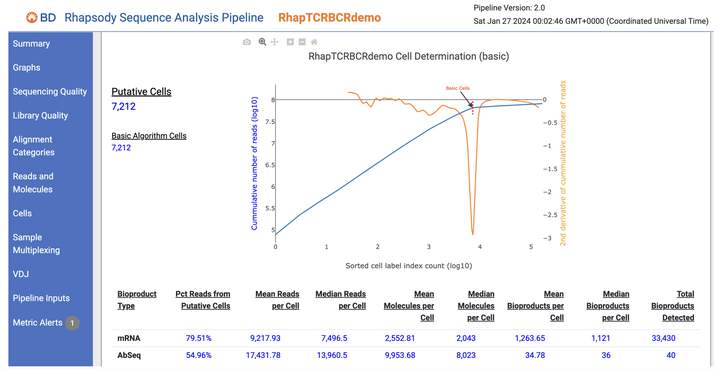



Visit the link below to access the setup guide and documentation

https://bd-rhapsody-bioinfo-docs.genomics.bd.com/

## Visualize results with the interactive HTML report




| Bioproduct Type | Pct Reads from Putative Cells | Mean Reads per Cell | Median Reads per Cell | Mean Molecules per Cell | Median Molecules per Cell | Mean Bioproducts per Cell | Median Bioproducts per Cell | Total Bioproducts Detected |
|---|---|---|---|---|---|---|---|---|
| mRNA | 79.51% | 9,217.93 | 7,496.5 | 2,552.81 | 2,043 | 1,263.65 | 1,121 | 33,430 |
| AbSeq | 54.96% | 17,431.78 | 13,960.5 | 9,953.68 | 8,023 | 34.78 | 36 | 40 |

## Output files are ready for downstream analysis

```
> library("Seurat")
> rhapsody_pbmc <- readRDS("RhapTCRBCRdemo-dev_Seurat.rds")
> rhapsody_pbmc
An object of class Seurat
33471 features across 7212 samples within 2 assays
Active assay: RNA (33431 features, 0 variable features)
 2 layers present: counts, data
 1 other assay present: ADT
 2 dimensional reductions calculated: tsne, umap
> head(rhapsody_pbmc@meta.data[,c("orig.ident", "Cell_Type_Experimental", "Sample_Tag", "nCount_RNA", "nCount_ADT", "T
otal_VDJ_Molecule_Count")], 5)
             orig.ident Cell_Type_Experimental     Sample_Tag nCount_RNA nCount_ADT Total_VDJ_Molecule_Count
3091 RhapTCRBCRdemo-dev          T_CD4_memory SampleTag06_hs       2191       9310                        3
3517 RhapTCRBCRdemo-dev          T_CD4_memory SampleTag06_hs       3051       8559                       20
3921 RhapTCRBCRdemo-dev           T_CD4_naive SampleTag05_hs       1671       9321                       11
4283 RhapTCRBCRdemo-dev           T_CD4_naive SampleTag05_hs       2091      11314                        2
4318 RhapTCRBCRdemo-dev                     B SampleTag06_hs       2331      12214                       43
```

```
import muon
rhapsody_pbmc = muon.read_h5mu("Rhapsody-TCRBCR-Demo.h5mu")
rhapsody_pbmc
```

0.6s Python

```
MuData object with n_obs × n_vars = 7212 × 33471
  var:  'Raw_Reads', 'Raw_Molecules', 'Raw_Seq_Depth', 'RSEC_Adjusted_Molecules', 'RSEC_Adjusted_Reads_non-singleton', 'RSEC_
  2 modalities
    rna:      7212 x 33431
      obs:    'Cell_Type_Experimental', 'Sample_Tag', 'Sample_Name', 'SampleTag01_hs_Read_Count', 'SampleTag02_hs_Read_Coun
      var:    'Raw_Reads', 'Raw_Molecules', 'Raw_Seq_Depth', 'RSEC_Adjusted_Molecules', 'RSEC_Adjusted_Reads_non-singleton'
      uns:    'Pipeline_Inputs', 'Pipeline_Metrics'
      obsm:   'X_tsne', 'X_umap'
    prot:     7212 x 40
      obs:    'Cell_Type_Experimental', 'Sample_Tag', 'Sample_Name', 'SampleTag01_hs_Read_Count', 'SampleTag02_hs_Read_Coun
      var:    'Raw_Reads', 'Raw_Molecules', 'Raw_Seq_Depth', 'RSEC_Adjusted_Molecules', 'RSEC_Adjusted_Reads_non-singleton'
      uns:    'Pipeline_Inputs', 'Pipeline_Metrics'
      obsm:   'X_tsne', 'X_umap'
```