How to Operate the BD Rhapsody™ Targeted Analysis VDJ CDR3 Pipeline

Introduction

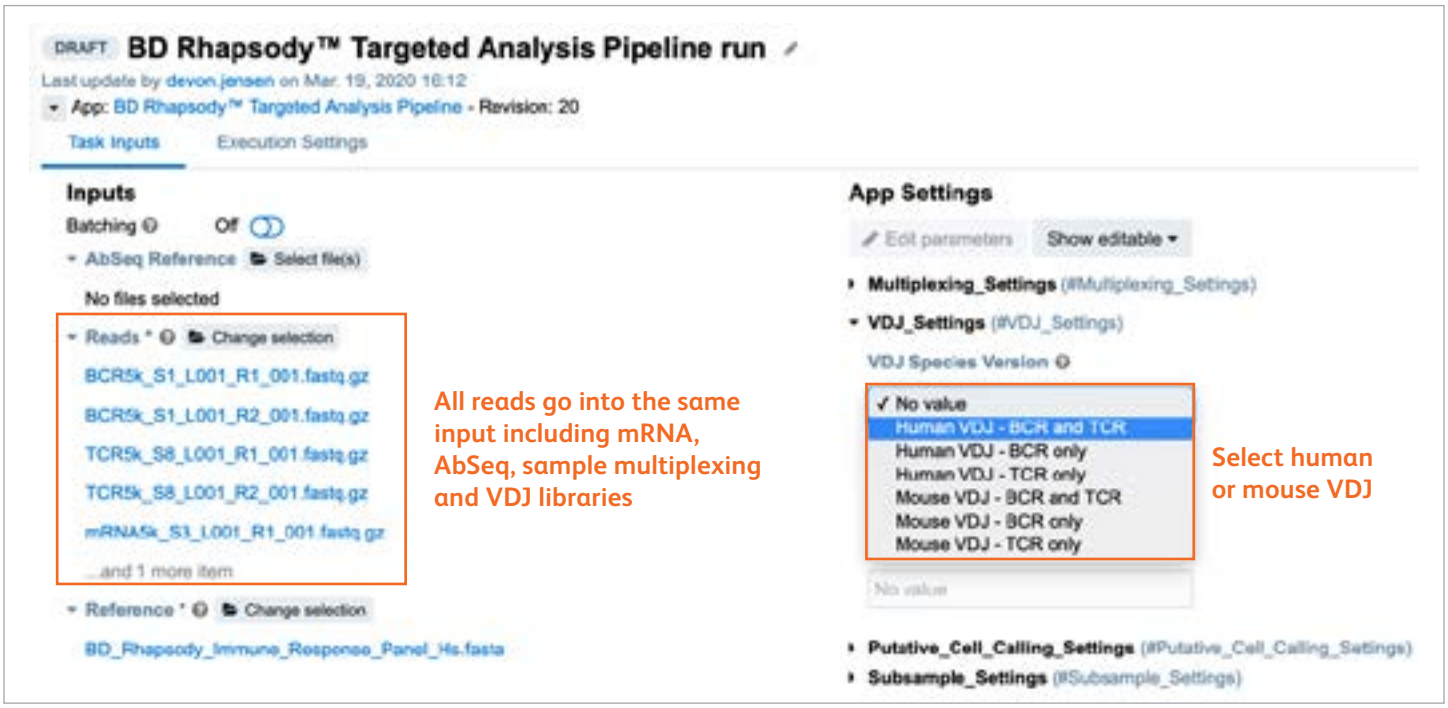
Starting the pipeline for a VDJ protocol experiment is very similar to a standard BD Rhapsody™ Targeted Pipeline run. This technical bulletin is a supplement to the **BD® Single-Cell Multiomics Bioinformatics Handbook** and provides a step-by-step instructions on how to operate the BD Rhapsody Targeted Analysis VDJ CDR3 Pipeline. The steps detailed below are designed to work with BD Rhapsody Targeted Analysis Pipeline, version 1.9 beta and for use with BD Rhapsody VDJ CDR3 Protocols.

Please note that the VDJ analysis is not included in the BD Rhapsody Targeted Pipeline version 1.8 and previous. For access to version 1.9 beta of the BD Rhapsody Targeted Pipeline with VDJ, contact BD single-cell multiomics support at **scomix@bdscomix.bd.com**.

Table of contents

VDJ analysis using the Seven Bridges Genomics platform

BD Rhapsody Analysis Pipelines are available on the **Seven Bridges Genomics platform** or on a local installation. All libraries from a single cartridge should go into the same pipeline run, including mRNA targeted, AbSeq, sample multiplexing, and BCR and TCR VDJ libraries. To turn on VDJ analysis in the pipeline, select the new app setting called "VDJ Species Version".



Local pipeline with CWL-runner

Follow the instructions in the **BD® Single-Cell Genomics Analysis Setup User Guide**, using the following docker command:

docker pull bdgenomics/rhapsody:1.9-beta

- Retrieve the 1.9 beta CWL document at **https://bitbucket.org/CRSwDev/cwl/src/master/v1.9-beta/**
- Create a .yml file to specify pipeline inputs, including the new "VDJ_Version" option.

Example: template_targeted_1.9-beta.yml

Reads (required) - Path to your read files in the FASTQ.GZ format. You may specify as many R1/R2 read pairs as you want.

```
Reads:
- class: File
location: "test/mySample_R1_.fastq.gz"
- class: File
location: "test/mySample_R2_.fastq.gz"
```

All reads go into the same input including mRNA, AbSeq, sample multiplexing and VDJ libraries

Reference (required) - Path to mRNA reference file for pre-designed, supplemental or custom panel, in FASTA format.
Reference:
- class: File
location: "test/BD_Rhapsody_Immune_Response_Panel_Hs.fasta"
AbSeq_Reference (optional) - Path to the AbSeq reference file in FASTA format. Only needed if BD AbSeq Ab-Oligos are used.

AbSeq_Reference:
- class: File
location: "test/AbSeq_reference.fasta"

VDJ Version (optional) - If VDJ is run, specify species: human, mouse, humanBCR, humanTCR, mouseBCR or mouseTCR
VDJ_Version: human **New option**

Overview of VDJ analysis

1. Identify and separate VDJ reads (Bowtie2 alignment)

A reference for aligning all reads is created by combining the VDJ gene segment sequences with mRNA panel, sample tag and AbSeq reference targets. Human and mouse BCR and TCR gene segments are built into the pipeline. Only necessary gene segments are added (BCR only, TCR only or both as appropriate). Gene segments are from the international ImMunoGeneTics information system® (**IMGT.org**). Reads that align to BCR or TCR gene segments are separated from other alignment types for further processing.

2. Identify constant region (Bowtie2 alignment)

A second bowtie2 alignment will identify the constant region used. This is purposefully a separate step, in preparation for future pipeline improvements. This is especially important for BCR-heavy chains, which switch to different isotypes.

3. Utilize IGBlast for VDJ segment and CDR3 determination

Each read sequence is analyzed by IGBlast to:

- Determine V, D, J genes that are used, along with alignment quality scores (e-value)
- Identify CDR3 nucleotide and amino acid sequence
- Check if sequence is productive, i.e., it is in-frame and contains no stop codon

4. Filter out low-quality reads

Reads are removed from further processing if they have a quality e-value of >1.0e-3 and a CDR3 was not identified.

5. Correct UMI and CDR3 nucleotide errors

UMI errors that are single base substitution errors are identified and adjusted to the parent UMI barcode using recursive substitution error correction (RSEC). This is identical to the RSEC algorithm described in the **BD® Single-Cell Multiomics Bioinformatics Handbook**. In addition, the CDR3 nucleotide sequences from the same cell index and UMI undergo a RSEC correction to remove sequencing errors.

6. Determine dominant CDR3 clone per cell-chain

For each cell index and chain type, such as TCRA and IGK, a dominant CDR3 clone is selected based on its respective UMI count and read count. Non-dominant clones are still output in the unfiltered data.

7. Create a per cell table and add experimental cell type

All dominant chain information for each cell index is compiled into a single row to facilitate downstream analysis. Then, an experimental cell type is inferred in one of two ways. First, if the experiment contained a targeted mRNA library using the Human Immune Response Panel, cells are labeled by a machine learning cell classifier trained on PBMC samples. Otherwise, the cell type is inferred from the relative molecule count of BCR chains vs TCR chains in each cell.

8. Run an additional distribution-based error correction per chain

Read counts from each chain type go through an additional round of error correction, like distribution-based error correction (DBEC) as described in the bioinformatics handbook. A histogram of chain read counts per cell is generated and a multimodal curve fit will identify a threshold of separation between signal and noise.

9. Generate metrics and file outputs

VDJ specific metrics are generated at several levels—overall VDJ, chain level metrics and cell type metrics. CSV data tables are output, representing filtered and unfiltered data per cell. Putative cells are determined by the mRNA targeted library and not by VDJ information.

Output	File	Content
VDJ metrics	<sample_name>_VDJ_metrics.csv	Report containing metrics associated with VDJ sequencing
VDJ per cell	<sample_name>_VDJ_perCell.csv	Data table containing dominant and error-corrected VDJ chain information for putative cells
VDJ per cell unfiltered	<sample_name>_VDJ_perCell_unfiltered.csv.gz	Data table containing dominant VDJ chain information for each cell index, putative and non-putative
VDJ per cell-chain unfiltered	<sample_name>_VDJ_perCellChain_unfiltered.csv.gz	Data table containing all clones for each cell index, dominant and non-dominant

VDJ metrics output

File: <sample_name>_VDJ_metrics.csv

The VDJ metrics file contains metrics at several levels including overall VDJ, chain level metrics and cell type metrics.

Section/metric	Definition	Major contributing factors
Overall VDJ metrics		
Reads_Cellular_Aligned_to_VDJ	Number of reads from all libraries with a valid cell label and UMI that aligned to a VDJ gene segment	• Sequencing quality • Library quality
Reads_CDR3_Valid_Unfiltered	Number of cellular VDJ aligned reads that had a valid CDR3 sequence	• Cell viability • Library quality
Reads_CDR3_Valid_Putative	Number of cellular VDJ aligned reads that had a valid CDR3 sequence and came from a putative cell	• Cartridge workflow performance
Pct_Reads_CDR3_Valid_from_Putative_Cells	Percent reads of the above metric relative to Reads_ Cellular_Aligned_to_VDJ	• Cartridge workflow performance
Reads_CDR3_Valid_Putative_Corrected	Number of cellular VDJ aligned reads that had a valid CDR3 sequence, came from a putative cell, belonged to the dominant clone for that cell-chain and passed DBEC	• Cell viability • Library quality
Pct_Reads_CDR3_Valid_Corrected_from_ Putative_Cells	Percent reads of the above metric relative to Reads_ Cellular_Aligned_to_VDJ	• Cell viability • Library quality
Mean_Reads_CDR3_Valid_Corrected_per_ Putative_Cell	Mean of reads_CDR3_Valid_Putative_Corrected per putative cell	• Cell viability • Library quality
Molecules_Unfiltered	Number of molecules represented by cellular VDJ aligned reads that had a valid CDR3 sequence after RSEC	• Cell viability • Library quality
Molecules_Corrected_Putative	Number of molecules from putative cells that also passed DBEC	• Cartridge workflow performance
Mean_Molecules_Corrected_per_Putative_Cell	Mean molecules per putative cell after DBEC	• Cell viability • Cartridge workflow performance
Chain type metrics		
Chain type metrics are identical to overall metrics except that they are split by VDJ chain type, such as TCR Alpha and BCR Kappa.		
Cell type metrics		
Cell_Type_Experimental	Cell types that were identified among all putative cells Cell type is inferred, either from the mRNA targeted panel expression data or from relative counts of BCR vs TCR	• Sample type • mRNA panel
Number_cells	Number of cells for each cell type	• Sample type
BCR_Paired_Chains_Percent	Percent of cells of each type that had both a BCR heavy chain and BCR light chain (Kappa or Lambda)	• Cell viability • Library quality
TCR_Paired_Chains_Percent	Percent of cells of each type that had either TCR Alpha and TCR Beta, or TCR Gamma and TCR Delta	• Cell viability • Library quality
<chain_type>_Percent_Cells_Positive	Percent of cells of each cell type that had at least one valid corrected molecule of the listed chain type	• Cell viability • Library quality
<chain_type>_ Mean_Molecules_per_Cell	Mean number of corrected molecules in each cell type of the listed chain type	• Cell viability • Library quality

VDJ per cell output

File: <sample_name>_VDJ_perCell.csv

The VDJ per cell output data table contains dominant and error-corrected (RSEC and DBEC) VDJ gene segment and CDR3 information for putative cells. Each row represents one cell. The cell indexes and order below are identical to the order of the cell index in the targeted gene and AbSeq expression data tables (e.g., <sample_name>_DBEC_MolsPerCell.csv). That makes this file an easy way to combine gene/AbSeq expression with VDJ information for downstream tools. To aid in downstream analysis, data from BCR Kappa and Lambda, TCR Alpha and Gamma, and TCR Beta and Delta are consolidated to one set of columns for each pair. Fully separate information for each chain is available in an output file, which is often labeled as _VDJ_perCell_unfiltered.csv file.

Section/metric	Definition	Major contributing factors
Cell_Index	Unique cell ID for the cell represented by this row Cell index will match between VDJ data and gene/AbSeq expression data tables	• Sequencing quality • Library quality
Total_VDJ_Read_Count	Total number of error-corrected VDJ reads for all chains in the cell	• Cell viability • Library quality
Total_VDJ_Molecule_Count	Total number of error-corrected VDJ molecules for all chains in the cell	• Cell viability • Library quality
<chain_type>_V_gene_Dominant	Dominant V gene segment identified for this chain type in the cell	• VDJ recombination
<chain_type>_D_gene_Dominant	Dominant D gene segment identified for this chain type in the cell	• VDJ recombination
<chain_type>_J_gene_Dominant	Dominant J gene segment identified for this chain type in the cell	• VDJ recombination
<chain_type>_C_gene_Dominant	Dominant C gene segment identified for this chain type in the cell	• VDJ recombination
<chain_type>_CDR3_Nucleotide_Dominant	Nucleotide sequence of the dominant clone for this chain type in the cell	• VDJ recombination
<chain_type>_CDR3_Translation_Dominant	Amino acid sequence of the dominant clone for this chain type in the cell	• VDJ recombination
<chain_type>_Read_Count	Number of error-corrected reads for this chain type in the cell	• Cell viability • Library quality
<chain_type>_Molecule_Count	Number of unique error-corrected molecules (UMI) for this chain type in the cell	• Cell viability • Library quality
BCR_Paired_Chains	True/False—this cell contains at least one error-corrected molecule of each BCR heavy and light (Kappa or Lambda)	• Cell viability • Library quality
TCR_Paired_Chains	True/False—this cell contains at least one error-corrected molecule of each TCR Alpha and TCR Beta, or TCR Gamma and TCR Delta	• Cell viability • Library quality
Cell_Type_Experimental	Inferred cell type of this cell index Cell type is inferred, either from the mRNA targeted panel expression data or from relative counts of BCR vs TCR	• Sample type • mRNA panel

VDJ per cell unfiltered output

File: <sample_name>_VDJ_perCell_unfiltered.csv.gz

The VDJ per cell unfiltered output data table contains the dominant VDJ gene segment and CDR3 information for each cell index, putative and non-putative. Each row represents one cell. UMI barcode and CDR3 correction are applied using RSEC but not DBEC.

Section/metric	Definition	Major contributing factors
Shared column definitions are identical to the VDJ per cell file		
Putative	True/False—this cell index was selected as a putative cell based on the mRNA panel	• Cell viability • mRNA panel

VDJ per cell-chain unfiltered output

File: <sample_name>_VDJ_perCellChain_unfiltered.csv.gz

The VDJ per cell-chain unfiltered output data table contains the VDJ gene segment and CDR3 information for all unique clones (dominant and non-dominant) for each cell index (putative and non-putative). UMI barcode and CDR3 correction are applied using RSEC but not DBEC.

Section/metric	Definition	Major contributing factors
Cell_Index	Unique cell ID for the cell represented by this row Cell index will match between VDJ data and the gene/AbSeq expression data tables	• Sequencing quality • Library quality
Chain_Type	Type of VDJ sequence: one of TCR_Alpha, TCR_Beta, TCR_Gamma, TCR_Delta, BCR_Heavy, BCR_Kappa and BCR_Lambda	• Cell viability • Library quality
V_gene	V gene segment identified for this cell-chain combination	• VDJ recombination
D_gene	D gene segment identified for this cell-chain combination	• VDJ recombination
J_gene	J gene segment identified for this cell-chain combination	• VDJ recombination
C_gene	C gene segment identified for this cell-chain combination	• VDJ recombination
CDR3_Nucleotide	Nucleotide sequence of the CDR3 for this cell-chain combination	• VDJ recombination
CDR3_Translation	Amino acid sequence of the CDR3 for this cell-chain combination	• VDJ recombination
Read_Count	Number of reads for this cell-chain combination	• Cell viability • Library quality
Molecule_Count	Number of unique molecules (UMI) for this cell-chain combination	• Cell viability • Library quality
Putative	True/False—this cell index was selected as a putative cell based on the mRNA panel	• Cell viability • mRNA panel

Additional resources

For any additional questions, contact **scomix@bdscomix.bd.com** or access BD VDJ CDR3 Pipeline protocols online at **scomix.bd.com.**

23-22817-00

BD Life Sciences, San Jose, CA, 95131, USA

bdbiosciences.com