

Nov 12, 2021

# Running the Titan\_Illumina\_PE Workflow on Terra.bio

Titan Illumina PE SARS-CoV-2 Strain Characterization Workflow for the Terra Platform

Jill V Hagey<sup>1</sup>, Kevin Libuit<sup>2</sup>, Frank J Ambrosio<sup>2</sup>,  
Technical Outreach and Assistance for States Team<sup>1</sup>

<sup>1</sup>Centers for Disease Control and Prevention; <sup>2</sup>Theiagen Genomics

1



[dx.doi.org/10.17504/protocols.io.bzykp7uw](https://dx.doi.org/10.17504/protocols.io.bzykp7uw)

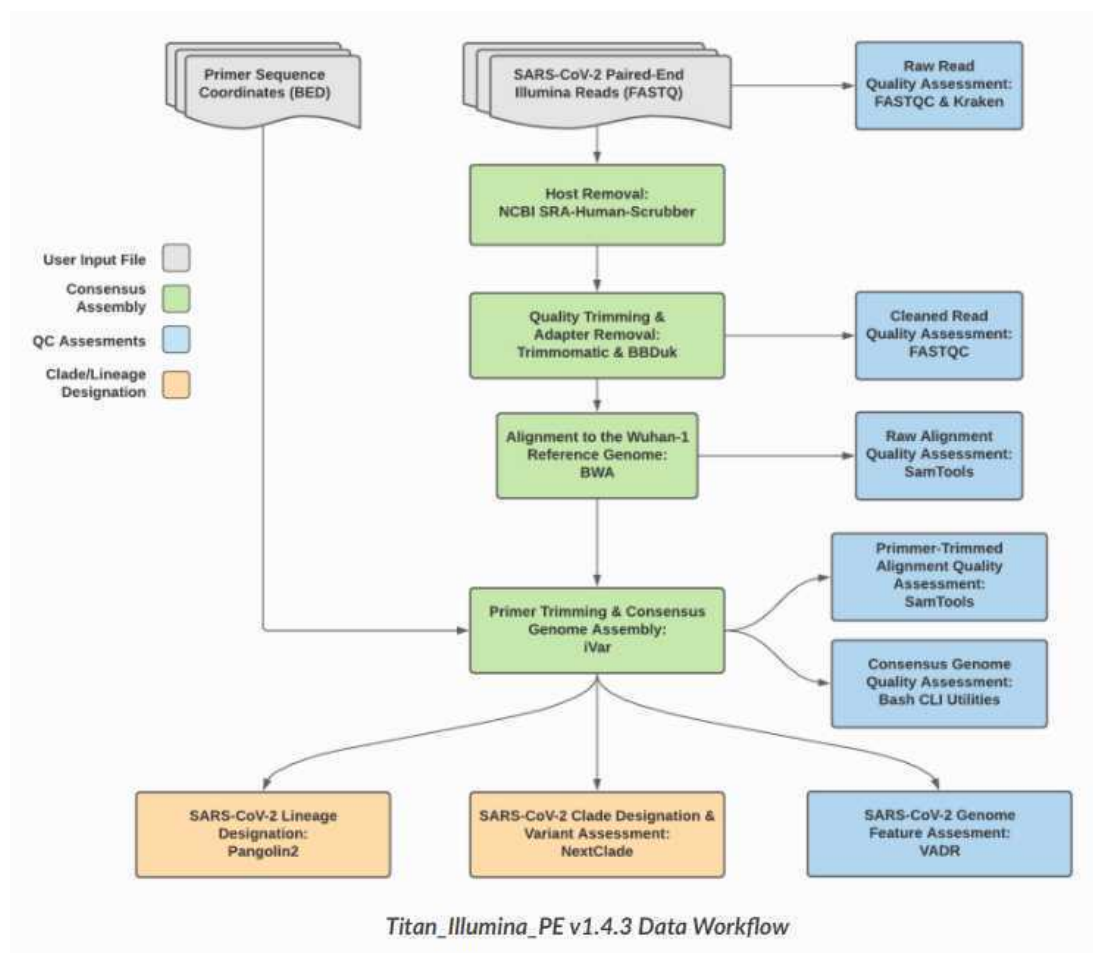
Theiagen

Francis Ambrosio

The opinions expressed here do not necessarily reflect the opinions of the Centers for Disease Control and Prevention or the institutions with which the authors are affiliated. The protocol content here is under development and is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to [protocols.io](https://protocols.io) is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with [protocols.io](https://protocols.io), can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

The Titan\_Illumina\_PE workflow is a part of the Public Health Viral Genomics Titan series for SARS-CoV-2 genomic characterization. Titan\_Illumina\_PE was written specifically to process Illumina paired-end (PE) read data. Input reads are assumed to be the product of sequencing tiled PCR-amplicons designed for the SARS-CoV-2 genome. The most common read data analyzed by the Titan\_Illumina\_PE workflow are generated with the ARTIC V3 protocol. However, alternative primer schemes such as the QIAseq Primer Panel are also suitable for this workflow. The primer sequence coordinates of the PCR scheme utilized must be provided in BED format along with the raw Illumina read data. Upon initiating a Titan\_Illumina\_PE job, the input primer scheme coordinates and raw paired-end Illumina read data provided for each sample will be processed to perform consensus genome assembly, infer the quality of both raw read data and the generated consensus genome, and assign lineage or clade designations as outlined in the Titan\_Illumina\_PE data workflow diagram below.

Additional technical documentation for the Titan\_Illumina\_PE workflow is available at: [https://public-health-viral-genomics-theiagen.readthedocs.io/en/latest/titan\\_workflows.html#titan-workflows-for-genomic-characterization](https://public-health-viral-genomics-theiagen.readthedocs.io/en/latest/titan_workflows.html#titan-workflows-for-genomic-characterization)



### Required input data for Titan Illumina PE:

Illumina paired-end read data (forward and reverse FASTQ files per sample)  
Primer sequence coordinates of the PCR scheme utilized in BED file format

### Video Instruction:

Theiagen Genomics: Titan Genomic Characterization

<https://www.youtube.com/watch?v=zP9I1r6TNrw>

Theiagen Genomics: Titan Outputs QC

<https://www.youtube.com/watch?v=Amb-8M71umw>

For technical assistance please contact us at: **TOAST@cdc.gov**

Titan\_Illumina\_PE.png

DOI

[dx.doi.org/10.17504/protocols.io.bzykp7uw](https://dx.doi.org/10.17504/protocols.io.bzykp7uw)

Jill V Hagey, Kevin Libuit, Frank J Ambrosio, Technical Outreach and Assistance for States Team 2021. Running the Titan\_Illumina\_PE Workflow on Terra.bio.

**protocols.io**

<https://dx.doi.org/10.17504/protocols.io.bzykp7uw>



## Titan Illumina PE SARS-CoV-2 Strain Characterization Workflow for the Terra Platform, Technical Outreach and Assistance for States Team

Illumina, Sequencing, MiSeq, iSeq, MiniSeq, NextSeq, NovaSeq, Paired-End, Next Generation Sequencing, NGS, SARS-CoV-2, Covid, Pangolin, Short-Read, Coronavirus, Genomics, Genetics, Virology, Molecular Biology

protocol ,

Nov 10, 2021

Nov 15, 2021

55020

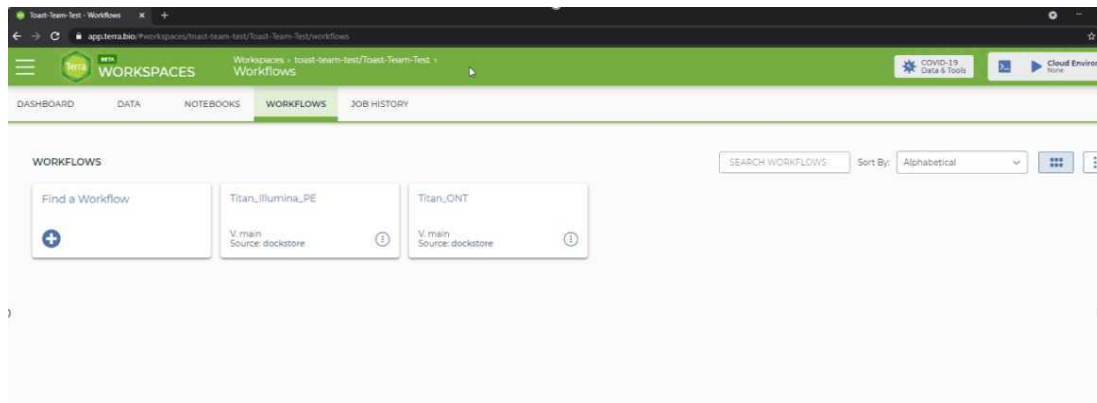
:

The opinions expressed here do not necessarily reflect the opinions of the Centers for Disease Control and Prevention or the institutions with which the authors are affiliated. The protocol content here is under development and is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to [protocols.io](https://dx.doi.org/10.17504/protocols.io.bzykp7uw) is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with [protocols.io](https://dx.doi.org/10.17504/protocols.io.bzykp7uw), can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

### Running the Titan Illumina PE Workflow

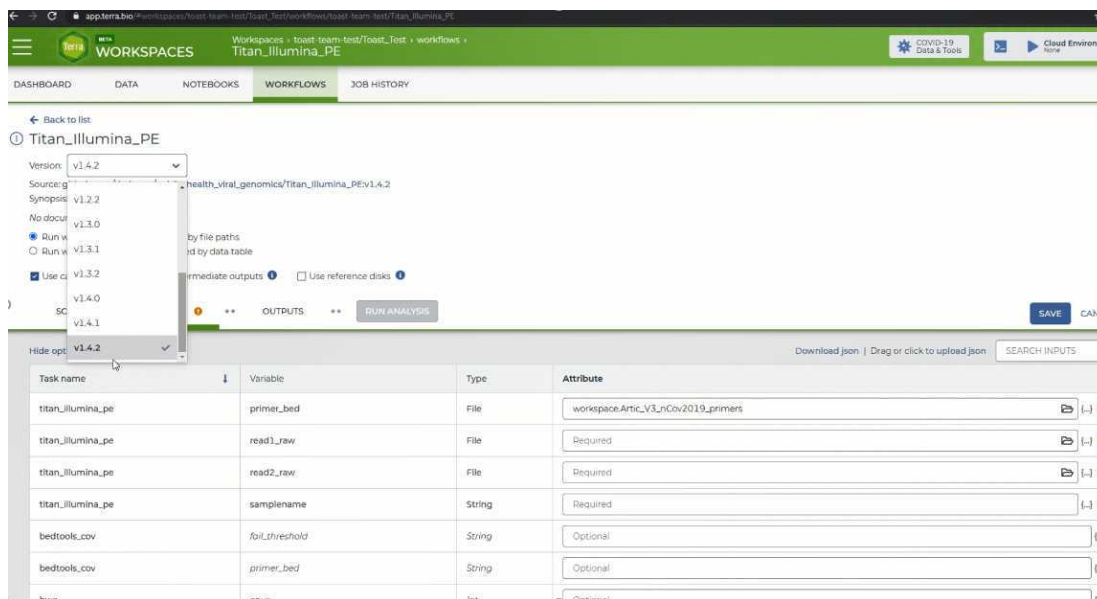
- 1 To run the Titan workflow, click on the 'Workflows' panel in the newly created workspace. It should bring you to the following page. if you do not see the Titan\_Illumina\_PE workflow in your workspace, please see our video on importing a workflow to Terra.bio:

<https://youtu.be/ZRpQyI1DMzo>).



The workflows panel in the newly created Terra workspace

Click on the 'Titan\_Illumina\_PE' tile and it will take you to a new page (if you do not see the Titan\_ClearLabs workflow in your workspace, please see our video on importing a workflow to Terra.bio: <https://youtu.be/ZRpQyI1DMzo>). **Double check that you are using the latest version of the workflow.** Alternately, you may specify another version, **but should only pick a stable version (with numbers) NOT a 'main' or 'dev' version.**



The Titan Illumina PE workflow page. Check version you are using.

Click the 'Run workflow(s) with inputs defined by data table' option button and then in the 'Select root entity type' pull-down menu, select 'sample' or whichever entity type you specified in the sample table. The first four rows in the 'Inputs' tab should each have a yellow caution icon. **These four rows represent variables that have to be provided by the user.** This is the information that we populated the sample data table with in the previous steps.

Make sure "Use call caching" is checked and then select the root entity type for the data you wish to analyze (here it's sample).

**NOTE:** Call caching allows Terra to identify and skip jobs that have been run previously; this option is by default enabled to avoid unnecessary compute costs. More information on Terra call caching, including examples of when you may want to disable this feature, is available through the [Terra Support Documentation](#).

The screenshot shows the Terra web interface for the 'Titan\_Illumina\_PE' workflow. The 'INPUTS' tab is selected, displaying a table of inputs. The 'Attribute' column has dropdown menus for each input. A 'SCRIPT' tab is also visible on the left.

Task name	Variable	Type	Attribute
titan_illumina_pe	primer_bed	File	Required
titan_illumina_pe	read1_raw	File	Required
titan_illumina_pe	read2_raw	File	Required
titan_illumina_pe	samplename	String	Required
bedtools_cov	failThreshold	String	Optional

The Titan Illumina PE workflow input panel after defining inputs by the sample data table

For the first row, click on the 'Attribute' text box and then click 'workspace.Artic\_V3\_primers' (or whatever you named the primer bed file). In the second row we input our primer\_bed file of choice, which should be saved as a workspace element in Terra (if you do not have this saved as a workspace element please see our video on uploading a workspace element: <https://youtu.be/Orbz7uRtwmQ>). For the next row, the 'read1\_raw' variable, click on the 'Attribute' text box and then click 'this.read1'. In the third row, the 'read2\_raw' variable, click on the 'Attribute' text box and then click 'this.read2'. Finally, in the third row, the 'samplename' variable, click on the 'Attribute' text box and then click option that displays your root entity in the notation: 'this.root\_entity'. Click the 'Save' button on the top right-hand side of the page. The yellow caution icons should disappear.

The screenshot shows the Terra web interface for the 'Titan\_Illumina\_PE' workflow. The 'INPUTS' tab is selected, displaying a table of inputs. The 'Attribute' column now shows the selected workspace elements: 'workspace.Artic\_V3\_primers', 'this.read1', 'this.read2', and 'this.Titan\_Illumina\_PE\_validation\_id'.

Task name	Variable	Type	Attribute
titan_illumina_pe	primer_bed	File	workspace.Artic_V3_primers
titan_illumina_pe	read1_raw	File	this.read1
titan_illumina_pe	read2_raw	File	this.read2
titan_illumina_pe	samplename	String	this.Titan_Illumina_PE_validation_id

## VERSION UPDATE:

Beginning with v1.4.4, you have the option to specify either [USHER](#) or pangoLEARN

methods for PANGO lineage assignment. The default is UShER unless "pangolearn" is entered into the "inference\_engine" row.

Visit the [CDC COVID-19 Genomic Epidemiology Toolkit](#) for an introduction to UShER:

### Module 3.3: Real-time phylogenetics with UShER

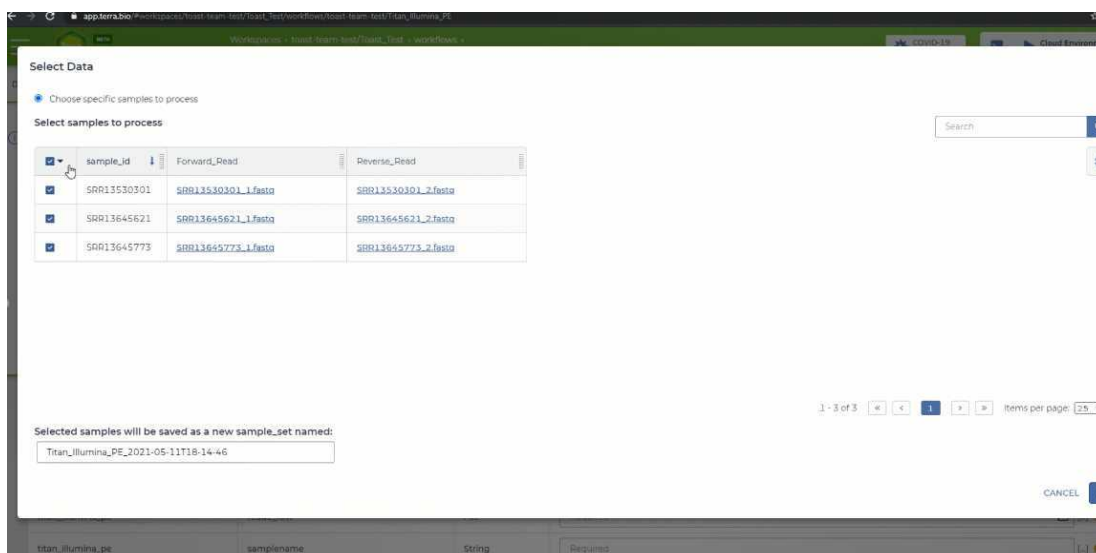
Next, click the "output" tab and select "use defaults" to autofill in the names of the output files according to the sample names. If you forget to select this option, you will have trouble accessing the results! **Save changes by clicking the 'Save' button.**

The screenshot shows the Terra.bio workspace for the 'Titan\_Illumina\_PE' workflow. The 'OUTPUTS' tab is active, displaying a table of output variables. The table has columns: Task name, Variable, Type, and Attribute. Two rows are shown, both with 'aligned\_bai' and 'aligned\_bam' variables. The 'Attribute' column has a dropdown menu with 'Use defaults' selected. A 'SAVE' button is at the bottom right.

Task name	Variable	Type	Attribute
titan_illumina_pe	aligned_bai	File	Required
titan_illumina_pe	aligned_bam	File	Required

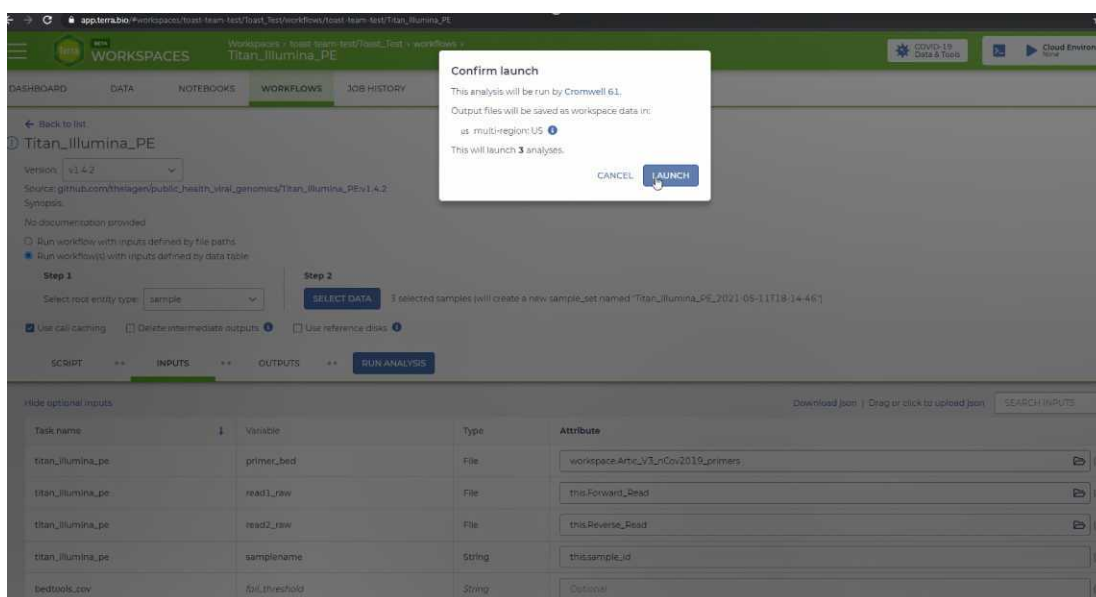
Setting output names by clicking "use defaults"

You are now ready to run the Titan\_Illumina\_PE workflow! Click the blue 'select data' button and select the samples you want to run or click the check box at the very top left to select all of them to be run.



Check the box to have all samples run through the Titan\_Illumina\_PE workflow.

Click on the 'Run Analysis' button to the right of the 'Outputs' tab. A popup window should appear titled 'Confirm launch'. **If the 'Run Analysis' button is greyed out, you need to save your recent changes by clicking the 'Save' button.**



The 'Confirm launch' popup window

Click the 'Launch' button to access the 'Job History' panel where each sample will be queued for the Titan\_Illumina\_PE analysis.



The screenshot shows the Terra.bio 'Job History' page for a submission with ID a90d1b08-26a3-4fb7-b06d-86c7c4999b62. The page displays various workflow statistics and a table of data entities.

Data Entity	Last Changed	Status	Run Cost	Messages	Workflow ID	Links
SR013530301 (sample)	May 11, 2021, 2:15 PM	Queued	N/A			
SR013645621 (sample)	May 11, 2021, 2:15 PM	Queued	N/A			
SR013645773 (sample)	May 11, 2021, 2:15 PM	Queued	N/A			

The 'Job History' page showing each sample queued for the Titan\_Illumina\_PE run

The status will change from 'Queued' to 'Submitted' to 'Running.' After the workflow finishes, the status column will indicate either 'successful' or 'failed.'

If you run into any trouble at all please feel free to check out our video on Running the Titan\_Illumina\_PE Workflow, or reach out to our support email: [support@terrapublichealth.zendesk.com](mailto:support@terrapublichealth.zendesk.com)