TIMS Quick Start Tutorial Guide

This quick start quide assumes that TIMS (Translational Informatics Management System) has been installed and the administrator account and password has been setup during the installtion process.

Contents

1. Account Creation

TIMS can be deployed at a multi-institutional level where the specific institutional structure whom an user (e.g. a Principal Investigator, PI) belongs to need to exist first. Otherwise, that institutional structure needs to be first created. Assuming that the TIMS administrator has login id/password (e.g. tims-admin/password)

Step 1: Login as the TIMS administrator

Step 2: Go to the *Admin* drop-down list as shown in Figure 1, then select *Work Unit Management* to get the *Work Unit Management* page (Figure 2)

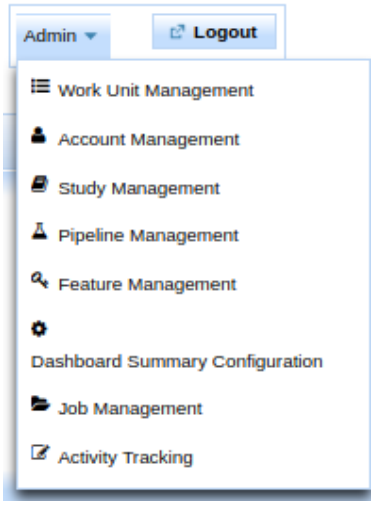

Figure 1.
Admin drop-down menu (left),

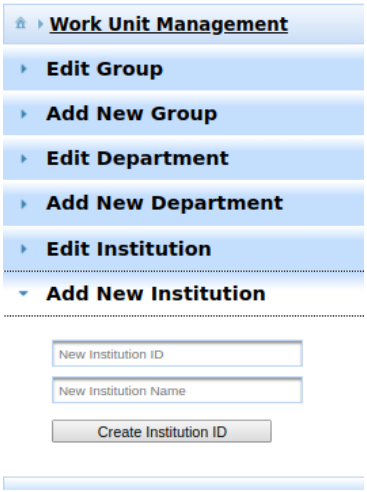

Figure 2.
Group Management Page (right).

Step 3: Select *Add New Institution* and type in the "Institution ID" (eg: *TIMSI*) as well as the "Institution Name" (eg: *TIMS Institute*) accordingly. Then please select *Create Institution ID.*

Step 4: Select *Add New Department* (Figure 3). On the *Select Institution* drop-down list, please select *TIMS Institute*. Subsequently, please type in the "Department ID" (eg: *TIMSD*) and the "Department Name" (eg: *TIMS Division*) accordingly. Then please select *Create Department ID.*

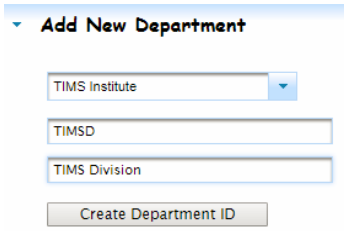

Figure 3: Add New Department

Step 5: Select *Add New Group* (Figure 4). On the *Select Institution* drop-down list, please select *TIMS Institute*. On the *Select Department* drop-down list, please select *TIMS Department*. Subsequently, please type in the "Group ID" (eg: *TIMSG*) and the "Group Name" (eg: *TIMS Group*) accordingly. Then please select *Create Group ID*, [Note: There is no PI account yet, therefore the *Select PI in-charge* drop-down list can be ignored for now. [This can be added in later.]

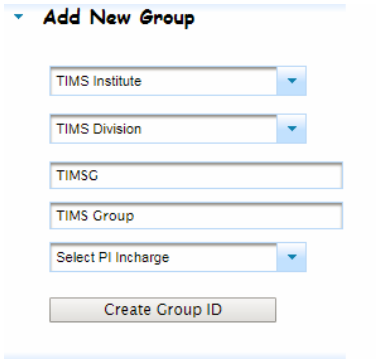

Figure 4: Add New Group

Once the Institute, Department and Group has been successfully created, TIMS administrator can view or edit the work units in the *Edit Institute, Edit Department* and *Edit Group* tabs seen in Figure 2 of the Work Unit Management page.

After creating the Institution, Department and Group, TIMS administrator can then proceed to create the PI account. In order to create the PI account, please follow the steps below:

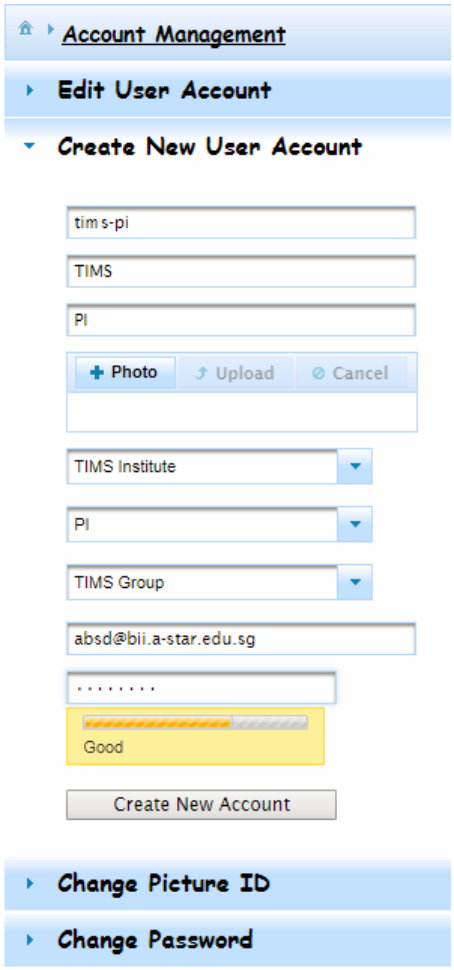

Figure 5: Adding new user account in the account management page

Step 1: Go to the *Admin* drop-down list as shown in Figure 1, then select *Account Management* to get the *Account Management* page (Figure 5).

Step 2: Select *Create New User Account* and type in the "User ID" (eg*: tims-pi*) as well as the "First Name" (eg: *TIMS*) and "Last Name" (eg: *PI*) accordingly. If "Photo" is available, please select *+ Photo* to add photo and subsequently select *Upload*.

Step 3: On the *Select Institution* drop-down list, please select *TIMS Institute*. On the *Select Role* drop-down list, please select "PI". On the *Select Working Unit* drop-down list, please select *TIMS Group*.

[Note: the Working Unit depends on the role of the user. For Director account, the Working Unit will be "TIMS Institute"; for HOD account, the Working Unit will be "TIMS Division"; and for User account the Working Unit will be "TIMS Group" (Refer to TIMS Manual Chapter 2, Section 2,1 for more details.]

Step 4: Please type in the email-address as well as the password accordingly

Step 5: Please select *Create New Account*.

To ensure that the account has been created, please check it from the tab *Edit User Account*. The user-id, last name, first name, role, unit-id, email, status and last login will be shown for the user.

2. Study Creation

Once the institutional structure and the PI user account (i.e. user id: tims-pi as example above) have been created in TIMS. The PI can now request the TIMS administrator to create his/her study in the system.

Firstly, the TIMS Administrator needs to obtain the information related to the study from the PI. This includes all the seven (7) items described in TIMS Manual Chapter 2, Section2.3, i.e. Title, Description, Background, Grant Information, Start and End of the study, Annotation and Disease classification. Once these information are obtained, the TIMS Administrator can follow the steps below to create the Study ID.

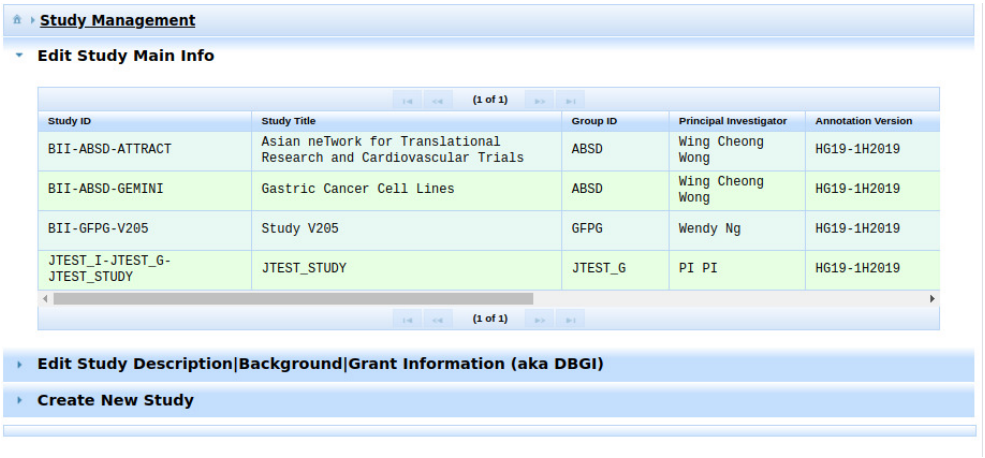

Figure 6. Study Management Page

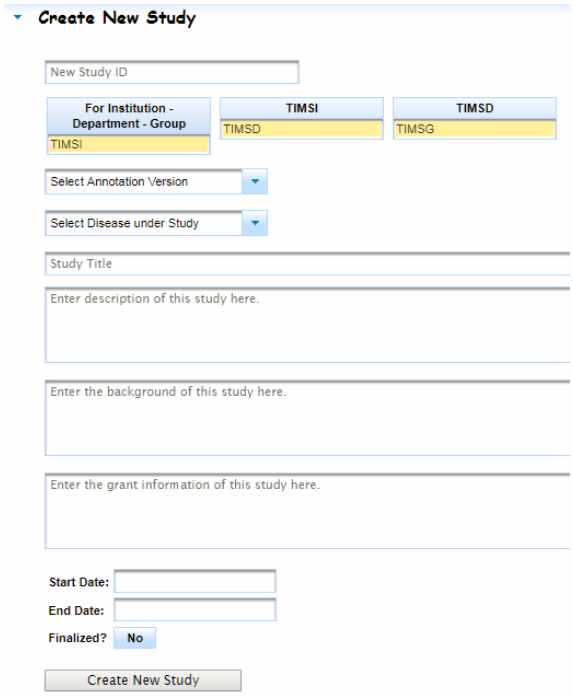

Figure 7: Create New Study

Step 1: Go to the *Admin* drop-down list as shown in Figure 1, then select *Study Management* to get the *Study ID Management page (Figure 6 above).*

Step 2: Please select the tab *Create New Study* to start creating the study (Figure 7).

Step 3: Please fill in all the required fields as shown in Figure . Importantly please select the *Institution – Department – Group* of the PI. This is important to ensure that the Study ID belongs to the PI of the Group.

[Note: For the purpose of this case study, please select the annotation version of "2H2016" and Disease to be "Malignant cancer of stomach". The Institution – Department – Group is

TIMSI – TIMSD – TIMSG]

Step 4: If this is an ad-hoc study, then the toggle the *Finalized* option to be Yes, otherwise, this should always be No. More information on this can be found in TIMS Manual Chapter 3, Section 3.1.5.

Step 5: Please select *Create New Study* to create the study.

3. Pipeline Analysis

With the study created, the PI can now, via his user account (i.e., tims-pi), proceed to analyse the raw data from his project using pipelines available on TIMS.

There are two categories of pipelines available: Array Processing and NGS Processing. The following sections will describe two pipelines available on TIMS, namely: Gene Expression (Affymetrix) and GATK Targeted Sequencing (Germline Mutation), which can be found in the two categories mentioned above respectively.

We have collected 10 gastric cancer cell lines from the public database. The 10 gastric cancer cell lines are available for both microarray gene expression data as well as the whole exome sequencing data.

All data used in this Quick Start Guide can be downloaded from http://mendel.bii.a-star.edu.sg/SEQUENCES/TIMS/Downloads/TestData.tar. Sample annotation files for each dataset used in this tutorial guide are also included in the downloaded tar file.

The downloaded file also contains data for selected other pipelines not described in this tutorial guide. All available test data and corresponding folder names can be seen in the table below for your reference.

Pipeline Category	**Pipeline Name**	**Folder Name containing data**
Array Processing	Gene Expression Pipeline (Affymetrix)	GEx
Array Processing	Methylation Pipeline	Methylation
NGS Processing	GATK Whole-Genome Sequencing (Germline Mutation)	DNAseq-Germline
NGS Processing	GATK Targeted Sequencing (Germline Mutation)	
NGS Processing	GATK Whole-Genome Sequencing (Somatic Mutation)	DNAseq-Somatic
NGS Processing	GATK Targeted Sequencing (Somatic Mutation)	
NGS Processing	RNA Sequencing Pipeline	RNAseq

3.1 Array Processing: Running Gene Expression Pipeline (Affymetrix)

Samples Description

The microarray gene expression data for the ten (10) gastric cancer cell lines are downloaded from the GEO database. Below table shows the ID and location of the CEL files:

Cell Line	GEO ID	URL
AGS	GSM552354	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22183
Hs746T	GSM552359	
MKN1	GSM552363	
MKN45	GSM552364	
SNU1	GSM552372	
SNU16	GSM552373	
SNU719	GSM552375	
YCC11	GSM552380	
YCC2	GSM552385	
YCC3	GSM552387	

All samples are hybridized to Affymetrix HGU133Plus2.0 array.
The downloaded file should include all the 10 Affymetrix CEL files together with sample annotation file named as “samples-annot-gex.txt” in the folder “GEx”.

Uploading Raw Data and Submitting Job

To start the analysis of Affymetrix Gene Expression array, please login and ensure that you are at the *Homepage*, subsequently under the *Array Processing* tab, please select *Gene Expression Pipeline (Affymetrix)*.

Subsequently please follow the steps below:
Step 1: Please select the study to work on, subsequently please select *New data to upload? Yes* because we are uploading new data [Note: if you are rerunning the pipeline with different parameters, you can select *New data to upload? No*] as shown inFigure 8. Then please select Proceed.

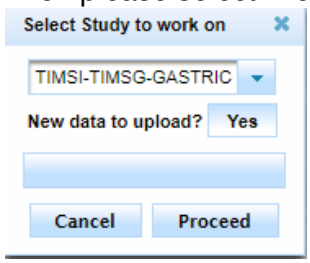

Figure 8: Upload new data

Step 2: On the *Gene Expression Pipeline* page, under the *Upload Sample Probe File(s)* section, please select *+ Select Sample File* and then *navigate* to the folder that contains the CEL files. Please select all the CEL files and add them to the import list.

Step 3: Please type in the Description field as “10 Gastric Cancer Cell Lines”. Then please select *Upload* to upload the CEL files (Figure 9)

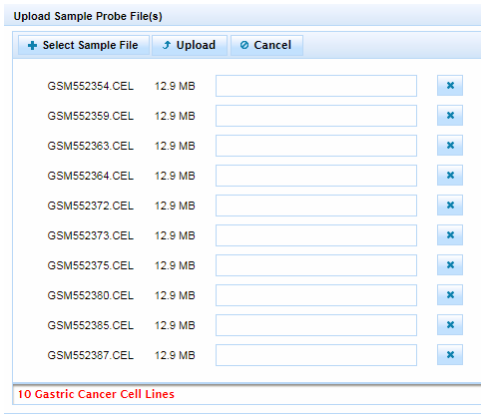

Figure 9: Uploading 10 CEL files

Step 4: Under the *Upload Samples Annotation file* section, please select *+ Select Samples Annotation File* then navigate to the folder that contains the samples annotation file (i.e. samples-annot-GEx.txt) and add the samples annotation file to the import list. Then please select *Upload* to upload the samples annotation file (Figure 10).

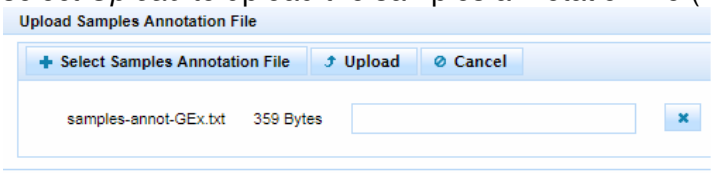

Figure 10: Upload samples annotation file

Step 5: On the Parameters section, please select *Type: HG-U133-Plus2.0* for the type of array. Then please select *Select Normalization Method: RMA* for the normalization method (Figure 11).

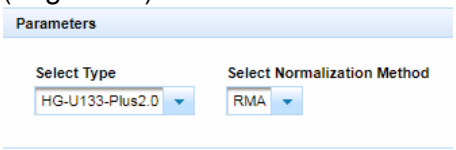

Figure 11: Parameters for Gene Expression Pipeline (Affymetrix)

Step 6: Please select *Submit* and check if all samples have been uploaded (Figure 12) in the *Pipeline Configuration Review* dialog box. If there are no errors, select *Confirm* to start the analysis.

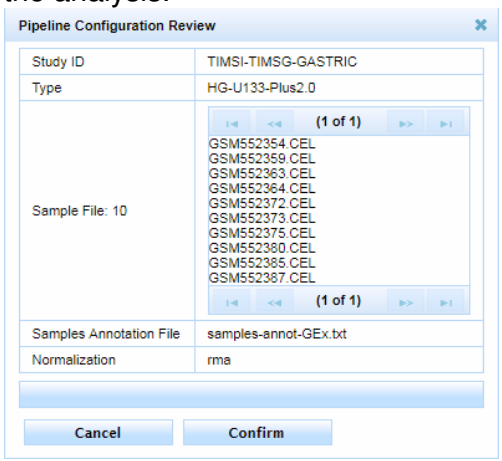

Figure12: Pipeline configuration review dialog box

Checking Pipeline Job Status

Once the job has been submitted, you can go to the *Homepage* and select the *My Work Area* tab, then please select *Pipeline Job Status* to check the status of your job. It should be shown as "In-progress" to ensure that the job is running (Figure 13).

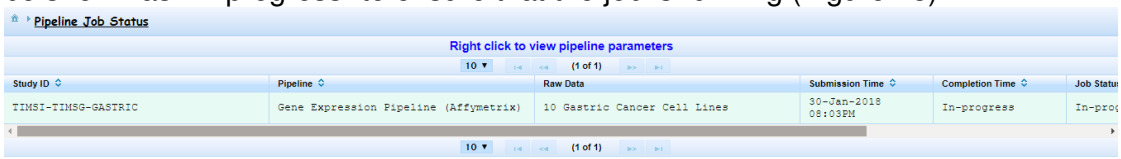

Figure 13: Pipeline job status shows the gene expression pipeline is in progress

Obtaining Pipeline Analysis Outputs

Once the job has finished, you will receive the email specifying that the job has been completed. You can then login to TIMS and go to *Homepage* > *My Work Area* > *Pipeline Job Status* to check the status. The job status should be shown as "Completed" and you can download the results from the *Output File*, *Detail Output* and *Report* field. The *Output File* is in the gene-level output such that genes are in rows and subjects are in columns. The*Detail Output* shows the detailed output for each gene and each subject. Each row represents a gene and all the probesets belong to that gene and the expression of all samples belong to that subject will be shown as colon-delimited value. The (QC) *Report*specific for the Affymetrix gene expression array can be downloaded to evaluate the quality of each sample in this study.

3.2 NGS Processing: GATK Targeted Sequencing (Germline Mutation) Pipeline

Samples Description

The whole-exome sequencing data for the ten (10) gastric cancer cell lines are downloaded from the SRA database. Below table shows the ID and location of the SRA files:

Cell Line	SRR ID	URL
AGS	SRR3929829	https://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR392/SRR3929829/SRR3929829.sra
Hs746T	SRR3929830	https://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR392/SRR3929830/SRR3929830.sra
MKN1	SRR3929852	https://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR392/SRR3929852/SRR3929852.sra
MKN45	SRR3929874	https://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR392/SRR3929874/SRR3929874.sra
SNU1	SRR3929835	https://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR392/SRR3929835/SRR3929835.sra
SNU16	SRR3929836	https://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR392/SRR3929836/SRR3929836.sra
SNU719	SRR3929844	https://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR392/SRR3929844/SRR3929844.sra
YCC11	SRR3929845	https://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR392/SRR3929845/SRR3929845.sra
YCC2	SRR3929847	https://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR392/SRR3929847/SRR3929847.sra
YCC3	SRR3929848	https://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR392/SRR3929848/SRR3929848.sra

The FASTQ files were extracted from the SRA files, subsequently trimmed for bad quality reads. The trimmed-fastq files were aligned to the UCSC hg19 genome using BWA-MEM following GATK recommended parameters.

For the proof-of-concept purpose, we have selected only 92 genes from the whole-exome sequencing data. These 92 genes are selected based on the highest gene expression variation across the 10 gastric cancer cell lines. The genomic coordinates of the 92 genes are shown in the interval.bed file accompanied this tutorial.
The resultant BAM files above will be pre-processed and subsequently the reads aligned to the 92 genes (interval.bed) will be extracted.

The data in this section can be found in the downloaded data file in the folder named "DNAseq-Germline". This should include all the 10 BAM files together with sample annotation file named as "samples-annot-exome.txt". [Note: Many of the steps are the same as the Gene Expression pipeline, therefore account user is expected to be familiar with the dialog box and therefore the screenshots won't be included unless necessary]

Uploading Raw Data and Submitting Job
To start the analysis of GATK Targeted Sequencing (Germline Mutation), please login and ensure that you are at the *Homepage*, subsequently under the *NGS Processing* tab, please select *GATK Targeted Sequencing (Germline Mutation)*.

Subsequently please follow the steps below:
Step 1: Please select the study to work on, subsequently please select *New data to upload? Yes* because we are uploading new data [Note: if you are rerunning the pipeline with different parameters, you can select *New data to upload? No*] as shown in 8. Then please select *Proceed*.

Step 2: On the *GATK Targeted Sequencing (Germline Mutation)* page, under the *Upload Sample Probe File(s)* section, please select *+ Select Sample File* and then navigate to the folder that contains the BAM files. Please select all the BAM files and add the BAM files to the import list.

Step 3: Please type in the Description field as "10 Gastric Cancer Cell Lines (BAM files)". Then please select *Upload* to upload the BAM files.

Step 4: Subsequently, please select *+ Select Interval File* and then navigate to the folder that contains the interval.bed file. Please select the interval.bed file and add it to the import list..

Step 5: Under the *Upload Samples Annotation file* section, please select *+ Select Samples Annotation File* and navigate to the folder that contains the samples annotation file (i.e. samples-annot-Exome.txt). Please select OK to add the samples annotation file. Then please select Upload to upload the samples annotation file.

Step 6: Please type in *Read Depth 100, Variant Depth 10,* and *Exclude DB: Yes* to filter the detected variants. These parameters filter to include only variants with read depth at least 100, the alternative variant depth is at least 10. Subsequently exclude the detected variant if it is inside dbSNP. (Figure 14)

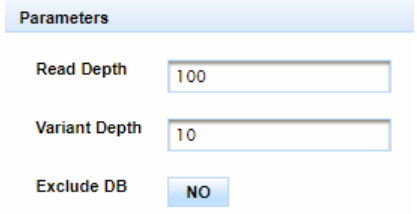

Figure 14: Parameters for GATK Targeted Sequencing (Germline Mutation) pipeline

Step 7: Please select *Submit* and check if all samples have been uploaded in the Pipeline Configuration Review dialog box. If there are no errors, select *Confirm* to start the analysis.

Checking Pipeline Job Status
Once the job has been submitted, you can go to the *Homepage* and select the *My Work Area* tab, then please select *Pipeline Job Status* to check the status of your job. It should be shown as "In-progress" to ensure that the job is running.

Obtaining Pipeline Analysis Outputs
Once the job has finished, you will receive the email specifying that the job has been completed. You can then login to TIMS and go to *Home* > *My Work Area* > *Pipeline Job Status* to check the status. The job status should be shown as "Completed" and you can download the results from the *Output File*, *Detail Output* and *Report* field. The Output File is in the simplified MAF (Mutation Annotation Format) file to import to cBioPortal; whereas the Detail Output contains the full MAF file. The (QC) Report specific for the Germline mutation can be downloaded to evaluate the quality of each sample in this study.
Other pipelines not described in this tutorial guide follow a similar process to the two pipelines mentioned in Section 3.1 and 3.2 above.

4. Setting up Visualization

Once analysis is done, the account user is able to visualize the data in cBioPortal. TIMS provides the function to export the results from the analysis to cBioPortal interface. In order to visualize the results in cBioPortal, the user first needs to setup the visualization first by selecting the output results to be exported. cBioPortal can then be launched following a successful export for visualization of the exported data.

Please follow the steps below to setup visualization and subsequently visualize it in cBioPortal.

Step 1: From the *Homepage*, under *Visualization* tab, please select *Setup Study for Visualization*. Subsequently, the dialog box *Select visualizer to work with* will be shown. Please select *cBioPortal* and then select *Proceed* (Figure 15).

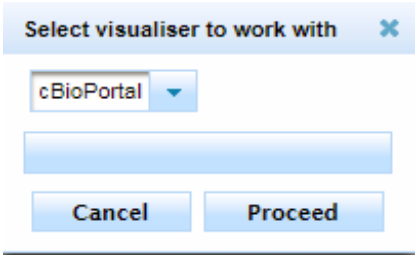

Figure 15: Selecting visualizer dialog box

Step 2: On the *Job(s) Selection for Visualization* page, first select for study to be visualized under the *Select Study to setup for visualization* section. The sections below will then be automatically populated with all completed jobs from pipelines. The user should then select the job to visualize for each profile. The job will be categorized based on the specific technology pipelines. For each technological pipeline, only one job can be selected. Multiple technology pipelines can be selected to be exported and only the selected results are exported for visualization (Figure 16). Once jobs of interes have been selected, click on *Export Data to cBioPortal.*

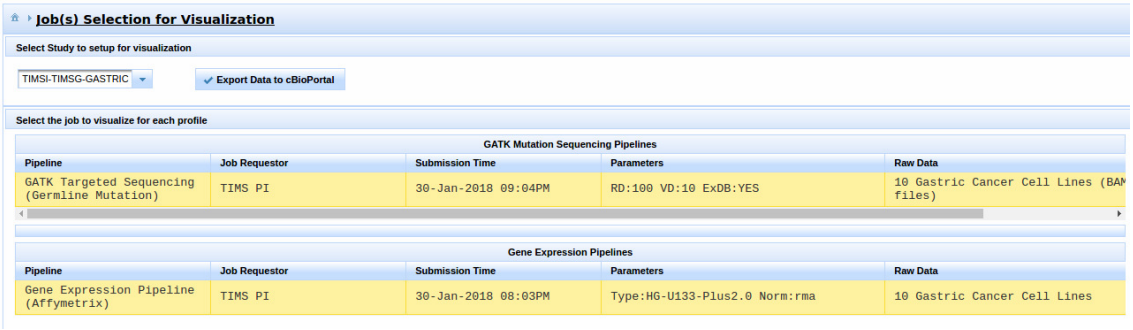

Figure16: Job(s) selection for visualization page

Step 3: After the results have been exported to cBioPortal, you will receive an email indicating that the visualization is ready. In order to visualize the exported data, please go to *Homepage* > *Visualization* > *Visualize my Study Data* to get the *Visualization* Page.

Step 4: Each study can only have one exported profile at any one time. Please select Launch to open cBioPortal to visualize the data for selected study. The cBioPortal home page should look like Figure 17.

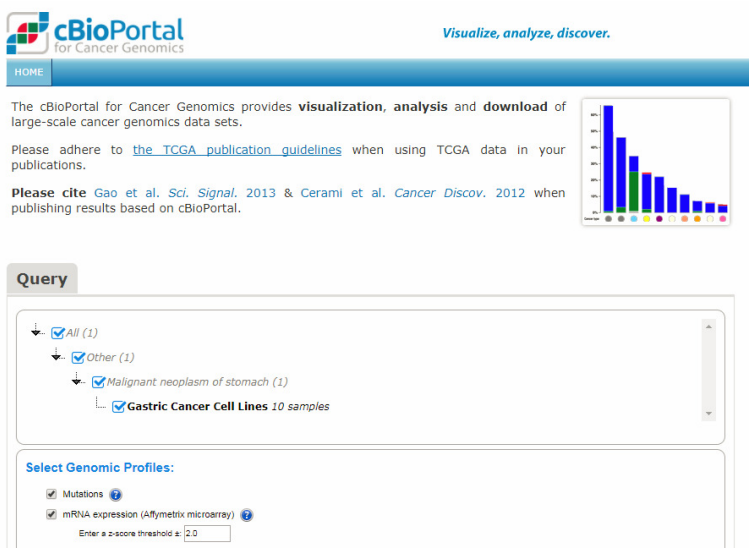

Figure 17: cBioPortal Home Page for the exported data

5. Uploading Meta Data

To associate the raw data with individual subjects, subject metadata can be uploaded onto TIMS as well. Three files are required for the upload of subject metadata on TIMS: 1. Core Data Tag File, 2. Meta Data File and 3. Study Specific Fields File. Details on all three files and their required formats (if any) can be found in TIMS manual Chapter 4, Section 4.1.1.

When uploading all three files mentioned above, select the *Subject Data* tab > *Meta Data Management* from the *Homepage*. A dialog box will pop up and the PI should select the relevant study and click on *Proceed* to open up the *Meta Data Management* page (Figure 18).

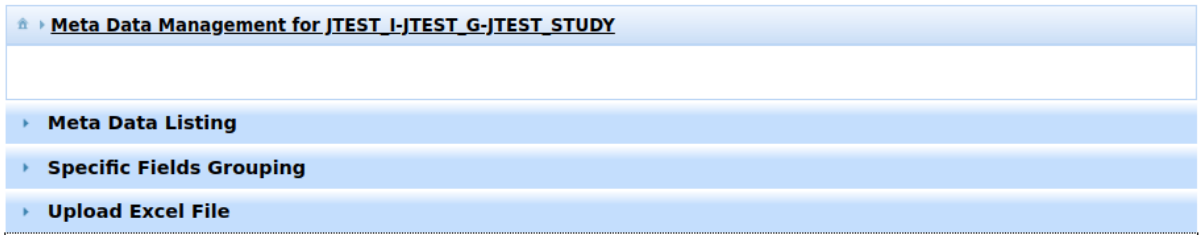

Figure 18. Meta Data Management Page.

Under the *Upload Excel File* tab, user should then proceed to upload all three required files by selecting *+ Core Data Tag File, + Meta Data File and + Study Specific Fields File* respectively (Figure 19). Notifications on the top right will be shown on the uploading of Core Data and Specific Fields files to indicate the status. A dialog box will pop up on the uploading of the Meta Data with a preliminary overview of the data, then select *Proceed*. Detailed interpretation of this dialog box can be found in TIMS Manual Chapter 4, Section 4.1.2.

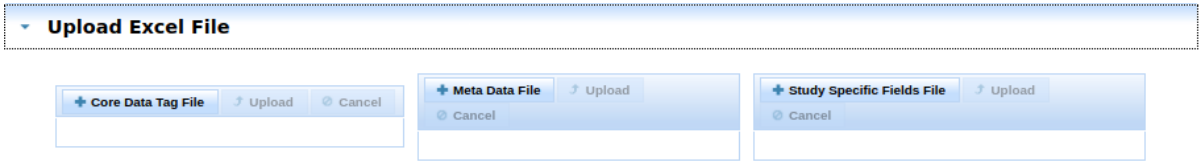

Figure 19. Upload Excel File

Once Meta Data File has beensuccessfully uploaded, an email will be sent to the account user. This can be confirmed by checking the uploaded data through the *Meta Data Listing* tab on the *Meta Data Management* page shown in Figure 1, whereAccount user should be able to view all successfully uploaded data.

Account user can download previously uploaded data using *Download Meta Data*. A report on the quality of previously uploaded data can also be obtained using *Quality Report of Last Data Upload* (Figure 20). Detailed interpretation of the Quality Report (Figure 21) can be found in TIMS Manual Chapter 4, Section 4.1.3.

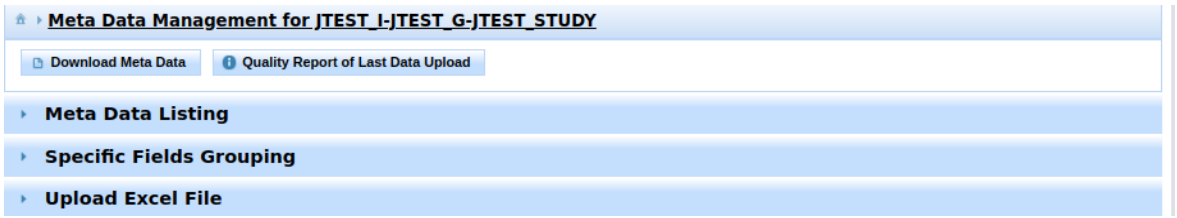

Figure 20: Edit Meta Data after upload

```
1 Overview of the quality of data (Uploaded by JTEST_ADMIN@01-Jun-18 12:00PM)
2
3 Records with invalid date: 3/11 (27.3%)
4 2, 3, 4,
5
6 Records with missing data: 5/11 (45.5%)
7 8, 9, 10, 11, 12,
8
9 Missing visits detected:
10 MKN45=2011-02-14,
11 Affected records: 1/11 (9.1%)
12 6,
13
14 Records with invalid data: 2/11 (18.2%)
15 5, 7,
16
```
Figure 21: Example of Quality Report of Last Data Uploaded

6. Finalization of Study

Having completed analysis on raw data as well as uploaded corresponding meta-data, account user is now able to select for relevant pipeline outputs to finalize data for storage into database.

This is accessible from *Homepage* > *My Study* > *Finalize Study* > *Select Study*.

User is then able to select for specific outputs of interest from various pipelines to be included in the database before proceeding to finalize (Figure 22.).

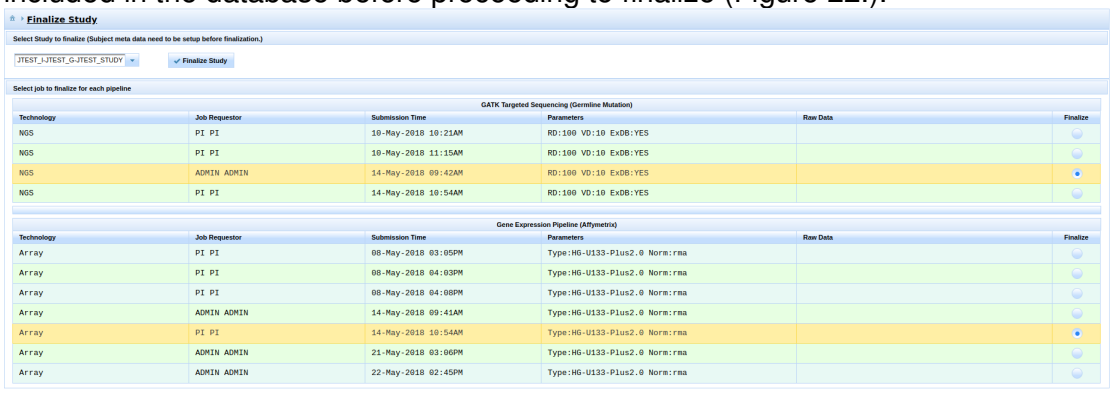

Figure 22: Selecting specific outputs of interest from various pipelines for finalization.

Once finalized, user can then access collated outputs from *Homepage* > *My Work Area* > *Completed Study Output* Figure 23).

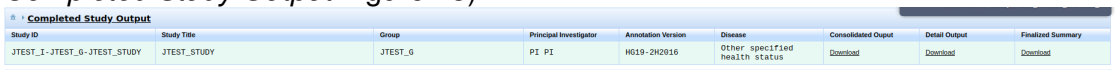

Figure 23: Completed Study Output

7. Unfinalize Study & Close Study

These options are only available Administrator accounts as mentioned in TIMS Manual Chapter 2. . This is accessible from *Homepage* > *My Study* > *Unfinalize Study/Close Study.*