

- 1 Introduction
- 2 Installation
- 3 Quick Start
 - 3.1 Get help about the tmbLab R package
 - 3.2 Identify the required data
 - 3.3 Run the function `tmbLab()`
 - 3.4 Run the function `tmbLabWES()`
- 4 Main functions
 - 4.1 `tmbLab()`
 - 4.1.1 Description
 - 4.1.2 Usage
 - 4.1.3 Arguments
 - 4.1.4 Value
 - 4.2 `tmbLabWES()`
 - 4.2.1 Description
 - 4.2.2 Usage
 - 4.2.3 Arguments
 - 4.2.4 Value
- 5 References

Aligning Tumor Mutational Burden (TMB) across diagnostic platforms:

A calibration tool developed in conjunction with Phase 2 of the Friends of Cancer Research TMB harmonization project

BRP Software Development Team

2021-03-29

1 Introduction

This package generates and applies a calibration model as described in Phase 2 of the Friends of Cancer Research (FOCR) tumor mutational burden (TMB) harmonization project [Vega DM 2021]. The package requires the user to supply a tab-delimited text file containing training data consisting of paired TMB values, one from a whole exome sequencing (WES) assay and one from a laboratory-specific panel diagnostic assay, for each of a series of samples. Using these training data, the software estimates a calibration curve with corresponding prediction limits to quantify the average relationship between WES and panel TMB assay values as well as variability around that curve. The user must also input additional TMB values from the same panel diagnostic assay, and the package will provide estimates of TMB that are calibrated to WES TMB using the calibration curve and associated prediction limits calculated from the training data. Intervals of uncertainty are also provided to accompany the corresponding WES-calibrated TMB values. HTML output and plots similar to published results from Phase 2 of the FOCR TMB harmonization project are provided to describe the calibration curve and report WES-calibrated laboratory-specific diagnostic panel assay values.

The calibration model is derived by fitting a weighted least squares linear regression model under assumptions of linear mean structure, Gaussian errors, and power variance structure. The software is designed for this situation based on results from phase 2 of the FOCR TMB Project that supported these assumptions. The source code is made available so that individual laboratories may modify it to explore different statistical approaches to developing calibration curves and prediction limits, which could include generalizations such as nonlinear regression and error structures other than Gaussian and power variance.

2 Installation

After downloading the package to a folder on your computer, you can install the tmbLab R package from the binary or source installer file. In the R console, click “Packages” on the R menu and select “Install package(s) from local files”. In the pop-up window, browse for the file “tmbLab_x.x.x.zip” or “tmbLab_x.x.x.tar.gz” and click “Open”. The installation process will show in the console. If there is no error message, the installation was successful.

3 Quick Start

3.1 Get help about the tmbLab R package

```
library(tmbLab)  
help(package=tmbLab)
```

3.2 Identify the required data

Two data files are required to run the functions `tmbLab()` or `tmbLabWES()`. Additional information about the data files `file.Model.Data.TMB`, `file.Obs.Panel.TMB`, and `file.WES.Panel.TMB` can be found in the function descriptions for `tmbLab()` and `tmbLabWES()` below.

A. `file.Model.Data.TMB` is required input for the `tmbLab()` or the `tmbLabWES()` function. This file describes the laboratory-specific panel TMB training data file name and must be a tab-delimited text file with 3 or more columns:

Column 1: "Sample.ID"; the unique sample identifiers.

Column 2: "Uniform.WES.TMB"; the training set WES values.

Column 3 and more: "Panel.1", "Panel.2," etc. denoting one or more laboratory-specific panel TMB columns.

B. `file.Obs.Panel.TMB` is required input for the `tmbLab()` function. This file describes the name of the data file that contains laboratory-specific panel TMB values at which the function will provide calibrated WES TMB estimates and intervals of uncertainty. The file must be a tab-delimited text file with 2 columns:

Column 1: "Sample.ID"; the unique sample identifiers.

Column 2: " Panel.TMB "; panel TMB values for which calibrated WES TMB estimates and intervals of uncertainty will be calculated. Note that "Panel.TMB" values may be real or hypothetical panel TMB values at which WES-calibrated estimates and intervals of uncertainty are desired.

C. `file.WES.Panel.TMB` is required input for the `tmbLabWES()` function. This file describes the name of the data file that contains WES TMB values at which panel TMB predicted values along with prediction limits will be calculated. The file must be a tab-delimited text file with 2 columns:

Column 1: "Sample.ID"; the unique sample identifiers.

Column 2: " WES.TMB "; WES TMB values for which predicted panel TMB and prediction limits will be calculated. Note that "WES.TMB" values may be real or hypothetical WES TMB values at which predicted panel TMB and prediction limits are desired.

3.3 Run the function `tmbLab()`

```
library(tmbLab)
res <- tmbLab(file.Model.Data.TMB =
              file.path(path.package("tmbLab"), "extdata/Model.WES.Panel.TMB.txt"),
              file.Obs.Panel.TMB =
              file.path(path.package("tmbLab"), "extdata/NewSample.Panel.TMB.txt"))
```

Briefly, running this line of code outputs the following (Note: information about the result files is included in the function descriptions below):

- an HTML result file
(C:/Users/YourUserName/Documents/tmbLab/tmbLab.html) and
- 2 results folders
(C:/Users/YourUserName/Documents/tmbLab/tmbLab_html/ and
C:/Users/YourUserName/Documents/tmbLab/tmbLab_res/) and
- a list `res` including the following elements:

```
names(res)
[1] "Obs.PANEL.TMB.vec" "Obs.PANEL.IDs.vec" "PANEL.vec"          "Calib.Interval"
"trunc.neg.flag"
[6] "dir.resPath"       "dir.output"         "predMethod"        "Lab.All"
```

Some brief explanations about the `res` list for the function `tmbLab()` are as follows:

- `res$Obs.PANEL.TMB.vec` observed panel-based TMB values.
- `res$Obs.PANEL.IDs.vec` observed panel-based TMB sample IDs.
- `res$PANEL.vec` the panel names which were provided in the file `file.Model.Data.TMB`.
- `res$Calib.Interval` the calibrated interval level of uncertainty.
- `res$trunc.neg.flag` the option for allowing negative values of panel-based TMB.
- `res$dir.resPath` the file path for the final result files.
- `res$dir.output` the file path for the output files.
- `res$predMethod` the calculation method: 1 for “tmbLab”; 2 for “tmbLabWES”.
- `res$Lab.All` a data frame including the WES-calibrated estimated TMB with intervals of uncertainty:

```
res$Lab.All
  Sample.ID Panel Obs.Panel.TMB CALIB.Est.TMB CALIB.Lower.Lim.TMB CALIB.Upper.Lim.TM
B Range.Indicator
1 NewS0001 Panel.1          2.41          1.4473          0.1519
5.5096          In
2 NewS0002 Panel.1          25.54         22.8244         17.3245          2
9.1249          In
3 NewS0021 Panel.1          52.80         48.0185          0.0000
NA             Out
4 NewS0003 Panel.1          3.06          2.0481          0.2527
6.2459          In
5 NewS0004 Panel.1          26.67         23.8688         18.2891          3
0.2331          In
.....
```

3.4 Run the function `tmbLabWES()`

```
library(tmbLab)
resWES <- tmbLabWES(file.Model.Data.TMB =
                    file.path(path.package("tmbLab"), "extdata/Model.WES.Panel.TMB.
txt"),
                    file.WES.Panel.TMB =
                    file.path(path.package("tmbLab"), "extdata/NewSample.WES.TMB.tx
t"))
```

Briefly, running this line of code outputs the following (Note: information about the result files is included in the function descriptions below):

- an HTML result file
(`C:/Users/YourUserName/Documents/tmbLab/tmbLabWES.html`) and
- 2 results folders
(`C:/Users/YourUserName/Documents/tmbLab/tmbLabWES_html/` and
`C:/Users/YourUserName/Documents/tmbLab/tmbLabWES_res/`) and
- a list `resWES` including the following elements:

```
names(resWES)
[1] "Obs.PANEL.TMB.vec" "Obs.PANEL.IDs.vec" "PANEL.vec"          "Pred.Interval"
"trunc.neg.flag"
[6] "dir.resPath"       "dir.output"         "predMethod"        "Lab.All"
```

Some brief explanations about the `resWES` list for the function `tmbLabWES()` are as follows:

- `resWES$Obs.PANEL.TMB.vec` observed panel-based TMB values.
- `resWES$Obs.PANEL.IDs.vec` observed panel-based TMB sample IDs.
- `resWES$PANEL.vec` the panel names which were provided in the file `file.Model.Data.TMB`.
- `resWES$Pred.Interval` the prediction interval level of uncertainty.
- `resWES$trunc.neg.flag` the option for allowing negative values of panel-based TMB.
- `resWES$dir.resPath` the file path for the final result files.
- `resWES$dir.output` the file path for the output files.
- `resWES$predMethod` the calculation method: 1 for “tmbLab”; 2 for “tmbLabWES”.
- `resWES$Lab.All` a data frame including predicted assay TMB value with prediction intervals:

```
resWES$Lab.All
  Sample.ID Panel WES.TMB WES.Est.TMB WES.Lower.Lim.TMB WES.Upper.Lim.TMB Range.
Indicator
1 NewS0001 Panel.1      5      6.2561           1.9686           10.5435
In
2 NewS0002 Panel.1     10     11.6682           6.5262           16.8103
In
3 NewS0003 Panel.1     15     17.0804          11.3591           22.8017
In
4 NewS0004 Panel.1     17     19.2452          13.3314           25.1591
In
5 NewS0005 Panel.1     20     22.4925          16.3181           28.6670
In
.....
```

4 Main functions

4.1 tmbLab()

4.1.1 Description

This function estimates, from user-supplied training data, a calibration curve with corresponding prediction limits to quantify the average relationship between WES and panel TMB assay results as well as variability around that curve. The user must also input additional TMB values from the same panel diagnostic assay, and the package will provide estimates of TMB that are calibrated to WES TMB using the calibration curve and associated prediction limits calculated from the training data. Intervals of uncertainty are also provided to accompany the corresponding WES-calibrated TMB values.

Maximum likelihood methods implemented in the `gls()` function in the R package “nlme” are used to fit a weighted least squares linear regression model. This model assumes a linear mean structure, Gaussian errors, and power variance structure. Users are advised to check the reasonableness of these assumptions for their own data.

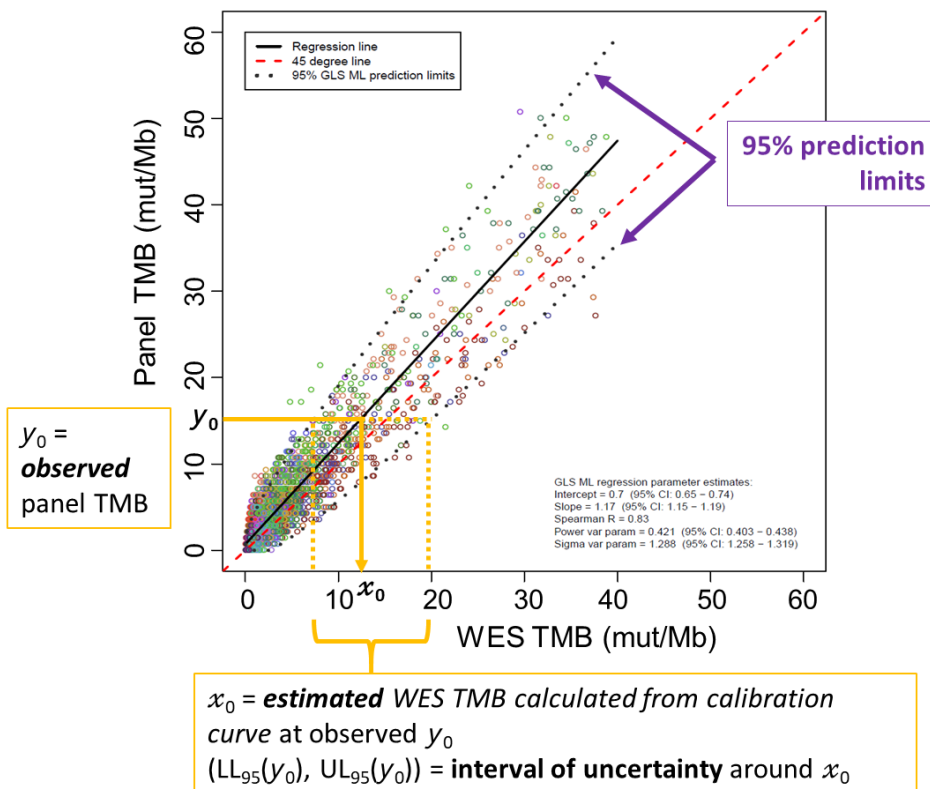
The function requires two input files. The first file, designated below as `file.Model.Data.TMB`, is a tab-delimited text file containing training data consisting of paired TMB values, one from a whole exome sequencing (WES) assay and one from a laboratory-specific panel diagnostic assay, for each of a series of samples. The second is a tab-delimited text file that contains laboratory-specific panel TMB values at which to provide WES-calibrated estimates. These laboratory-specific panel TMB values will be the values at which the function will invert the regression line and prediction limits to obtain WES-calibrated estimates and their corresponding intervals of uncertainty. Further details of the required formats for these files are given below under the “Arguments” section.

To better understand the calibration process, consider Figure 1. The first input file, which contains the training data, will correspond to the points depicted in the scatter plot with x-axis and y-axis corresponding to the WES TMB values `Uniform.WES.TMB` and laboratory-specific panel TMB values `Panel.TMB`, respectively. This first data input file includes all the information necessary to generate the regression line and prediction limits, denoted by the solid black line and the dotted black lines, respectively. The second input file provides a set of y_0 values, which are real or hypothetical panel TMB values at which WES-calibrated estimates and intervals of uncertainty are desired. In Figure 1, a single y_0 input value is depicted on the y-axis with a yellow horizontal line. The

WES-calibrated estimate pertaining to y_0 is derived by using the fitted calibration curve and is depicted on the x-axis as x_0 . In addition, an interval of uncertainty around the WES-calibrated value, $(LL_{95}(y_0), UL_{95}(y_0))$, is provided by projecting the prediction limits onto the WES-axis as shown.

Prior to fitting the regression model, this function drops observations (samples) from the training data supplied via the input `file.Model.Data.TMB` with WES TMB values that are greater than 40. These observations are dropped from the analysis because data collected as part of Phase 2 of the FOCR TMB harmonization project [Vega DM 2021] were insufficient to assess modeling assumptions, particularly linearity, at these high TMB values. The source code is made available so that users may modify this cutoff, if desired.

Figure 1. Calibration for a specific laboratory’s panel-based TMB diagnostic assay. Each plotted point represents WES TMB (x) and Panel TMB (y) for a particular sample in the training data set. The solid black line is the fitted weighted least squares regression line, and the dotted black lines are the corresponding 95% prediction limits. The value y_0 is a hypothetical TMB value measured using the panel assay of interest. The WES-calibrated estimate for y_0 is depicted by x_0 on the x-axis. The interval of uncertainty around the WES-calibrated value, $(LL_{95}(y_0), UL_{95}(y_0))$, is indicated on the x-axis.



4.1.2 Usage

```
tmbLab(  
  file.Model.Data.TMB,  
  file.Obs.Panel.TMB,  
  Calib.Interval = 95,  
  trunc.neg.flag = c("NEG", "TRUNCtoZERO"),  
  dir.output,  
  dir.result = "tmbLab_res",  
  show.HTML = TRUE  
)
```

4.1.3 Arguments

file.Model.Data.TMB Character string

This required file describes the laboratory-specific panel TMB training data file name and must be a tab-delimited text file with 3 or more columns:

Column 1: "Sample.ID"; the unique sample identifiers.

Column 2: "Uniform.WES.TMB"; the training set WES values.

Column 3 and more: "Panel.1", "Panel.2," etc. denoting one or more laboratory-specific panel TMB columns. It is anticipated that most users will provide only one set of laboratory-specific panel TMB data, corresponding to "Panel.1," however, the option is given to include data from multiple laboratory-specific panels by specifying additional columns.

Note also that the model implemented in this package assumes that samples are independent. Inputting multiple replicate TMB measurements from the same sample will not lead to appropriate statistical inference (e.g., prediction limits will be incorrect).

An example of this required file is as follows:

Sample.ID	Uniform.WES.TMB	Panel.1	Panel.2
S0000001	1.3410	1.0132	0.9345
S0000002	1.6414	3.4837	2.0914
S0000003	0.7608	0.3165	0.2429
S0000004	0.8482	1.5107	0.4991
S0000005	0.8301	3.4755	1.8225
.....			

file.Obs.Panel.TMB Character string

This required file describes the name of the data file that contains laboratory-specific panel TMB values at which the function will provide calibrated WES TMB estimates and intervals of uncertainty. The file must be a tab-delimited text file with 2 columns:

Column 1: "Sample.ID"; the unique sample identifiers.

Column 2: " Panel.TMB "; panel TMB values for which calibrated WES TMB estimates and intervals of uncertainty will be calculated. Note that "Panel.TMB" values may be real or hypothetical panel TMB values at which WES-calibrated estimates and intervals of uncertainty are desired. Typical hypothetical TMB values of interest would be 5, 10, 15, and 20. The file must contain at least one Panel.TMB value.

An example of this required file is as follows:

Sample.ID	Panel.TMB
NewS0001	2.41
NewS0002	25.54
NewS0003	3.06
NewS0004	26.67
NewS0005	37.38
.....	

Calib.Interval Numeric value

Specifies the level of uncertainty of the calibrated interval. For example, Calib.Interval = 95 means that 95% prediction limits will be projected onto the WES TMB axis to derive 95% calibration intervals of uncertainty. The inputted value must be a number between 1 and 100. Common values are 90, 95, and 99. Default is 95.

trunc.neg.flag Character string

If value is “NEG”, the code allows negative values of panel-based TMB to be used in the regression modeling. Although TMB values cannot actually be negative, some laboratories apply correction factors to panel TMB values that could cause small values to become negative. Negative values are not a problem for fitting the regression line, but they must be set to zero for purposes of variance and prediction limit calculation. Nonetheless, WES-calibrated estimates of TMB will always be truncated at zero so that they are not reported as negative values.

If “TRUNCtoZERO”, the code truncates negative panel-based TMB values to zero prior to their use in regression modeling.

`dir.output` Character string

This argument can be used to specify the directory output file path. Default is <user’s home directory>/tmb.

`dir.result` Character string

This argument can be used to specify the folder name in which the results will be placed. This folder will be found within the folder “dir.output”. Default is “tmbLab_res”.

`show.HTML` Logical

If TRUE, the HTML summary document will automatically pop up in the system default browser. Default is TRUE. A copy of this output is also saved for later viewing (see tmbLab.html output file).

4.1.4 Value

The output list includes the following elements:

- `Obs.PANEL.TMB.vec` : observed panel-based TMB values.
- `Obs.PANEL.IDs.vec` : observed panel-based TMB sample IDs.
- `PANEL.vec` : the panel names which were provided in the file `file.Model.Data.TMB`.
- `Calib.Interval` : the calibrated interval level of uncertainty.
- `trunc.neg.flag` : the option for allowing negative values of panel-based TMB.
- `dir.resPath` : the file path for the final result files.
- `dir.output` : the file path for the output files.
- `predMethod` : the calculation method: 1 for "tmbLab"; 2 for "tmbLabWES".
- `Lab.All` : a data frame including the WES-calibrated estimated TMB with intervals of uncertainty.

The following output files are generated:

`tmbLab.html`

The HTML output popup, which appears automatically if `show.HTML=TRUE`, includes a summary of relevant information. This same information is saved in a file that will be written to the specified `dir.output`. The preprocessing section of the output file describes the total number of samples, the number of samples that were dropped because they had WES TMB values >40 , and the range of data values observed. The table provides the second data file input values ("`Sample.ID`" and "`Obs.Panel.TMB`") along with the WES-calibrated estimates ("`Calib.Est.TMB`") and the lower and upper intervals of uncertainty corresponding to these values ("`CALIB.Lower.Lim.TMB`" and "`CALIB.Upper.Lim.TMB`"). The next column, "`Range.Indicator`," indicates whether the specified "`Obs.Panel.TMB`" value falls within the range of laboratory-specific TMB values (e.g., "`Panel.1`") described in `file.Model.Data.TMB`. "`Range.Indicator`" will be designated "In" if the specified "`Obs.Panel.TMB`" value falls within the range of laboratory-specific TMB values in the training data (e.g., the range of "`Panel.1`" described in `file.Model.Data.TMB`) and will be designated "Out" if the specified "`Obs.Panel.TMB`" value falls outside this range. For reference, this range is also described in the preprocessing section of the HTML output. The upper and lower limits of uncertainty are set to "NA" or 0, respectively, if the upper limit cannot be interpolated from within the limits of the plot or if the lower limit is below 0. In the last column, a link is provided to a .pdf plot of the scatterplot along with the calibration fitted regression line (solid black line), calibration regression line parameters (text on bottom right), and the prediction limits (dotted black line). Portions of the information that are summarized in this HTML file are also included in various output files in the `dir.output` folder, "`tmbLab_html`" and "`tmbLab_res`," as described below.

`Panel.n.GLS.ML.scatterplot.pdf`

Scatterplots depicting laboratory-specific TMB values (y-axis) versus WES TMB values (x-axis), a 45 degree line (dashed red line), the superimposed regression line (solid black line), and prediction intervals (dotted black line) are in the tmbLab_html folder and are linked to in the final column of the tmbLab.html file. The level of the prediction intervals is specified through the Calib.Interval (Default = 95)% input variable because the prediction interval level controls the calibration interval level. GLS ML regression parameter estimates are also provided in the bottom righthand corner of the plots, including the intercept, slope, Spearman R, Power variance parameter, and sigma variance parameter. The x and y axes of the plots range from 0 to 55 as these lower values are of greatest relevance for clinical decision making. Users are advised to check for possible outliers that lie beyond the upper range of the y-axis as these outliers have the potential to be influential. The naming convention of the scatterplots is such that n in the name "Panel.n.GLS.ML.scatterplot.pdf" refers to the corresponding panel number specified in the columns of file.Model.Data.TMB. If only "Panel.1" is specified, only one plot will be generated with the name "Panel.1.GLS.ML.scatterplot.pdf."

All.GLS.ML.ReggressionParameters.txt

Detailed regression parameters, found in the tmbLab_res and tmbLab_html folder. The model is fit as described in the background section. Each Panel for which data are supplied in file.Model.Data.TMB will correspond to a row in this file. The information and parameters described in this file include N, the number of paired laboratory-specific TMB and WES TMB points included in the training data used for fitting the calibration model; N.NEG.TMB, number of negative panel TMB values provided (for most laboratories this number will be zero); MODEL.AIC, the Akaike information criterion for the fitted regression model; MODEL.BIC, the Bayesian information criterion for the fitted regression model; BETA0, BETA0.SE, BETA0.95CI.LL, BETA0.95CI.UL, which describe the estimated intercept parameter, the corresponding standard error, lower 95% confidence limit, and upper 95% confidence limit, respectively; BETA1, BETA1.SE, BETA1.95CI.LL, BETA1.95CI.UL, which describe the estimated slope parameter, the corresponding standard error, lower 95% confidence limit, and upper 95% confidence limit, respectively; SPEARMAN.R, which describes the Spearman correlation coefficient estimating the association between panel TMB and WES TMB in the training data; POW.PARAM, POW.SE, POW.95CI.LL, POW.95CI.UL, which describe the estimated power parameter, the corresponding standard error, lower 95% confidence limit, and upper 95% confidence limit, respectively; SIGMA.PARAM, SIGMA.95CI.LL, SIGMA.95CI.UL, which describe the estimated sigma parameter, the corresponding lower 95% confidence limit, and upper 95% confidence limit, respectively; LSIGMA.PARAM, LSIGMA.SE, which describe the estimated sigma parameter on the natural log (base e) scale and the corresponding standard error, respectively; RESID.VAR, which describe the standard error of the residuals corresponding to the fitted model. Refer to the package nlme() for more information.

tmbLab.ALL.GLS.ML.FIT.TMB.txt

File containing points lying on estimated regression lines and prediction limits calculated at the originally supplied WES TMB values. Columns are "Sample.ID", the panel identifier "Panel.n", regression estimates "FIT.Est.TMB", lower prediction limits "FIT.Lower.Lim.TMB", and upper prediction limits "FIT.Upper.Lim.TMB" calculated at the WES values "Uniform.WES.TMB" provided in training data file.Model.Data.TMB, found in the tmbLab_res folder. If training data from multiple laboratory-specific panels are provided, this file will include regression results for all laboratory-specific panels for which training data were supplied ("Panel.1", "Panel.2", etc.) in a long data format so that each panel's information is appended with a variable specifying the panel number. Note that this output file provides results similar to those provided by the function tmbLabWES, except that tmbLabWES allows users to specify WES values for which this information is output, whereas, this file provides results using only the WES values that were input as part of the training data in file.Model.Data.TMB.

tmbLab.All.GLS.ML.CALIB.txt

File containing WES-calibrated TMB estimates and intervals of uncertainty calculated at user-supplied laboratory-specific panel TMB values that were input via file.Obs.Panel.TMB. Columns are "Sample.ID", the panel identifier "Panel.n", panel TMB values at which WES-calibrated values will be calculated ("Obs.Panel.TMB" from file.Obs.Panel.TMB), WES-calibrated values "CALIB.Est.TMB", lower limit of uncertainty "CALIB.Lower.Lim.TMB," upper limit of uncertainty "CALIB.Upper.Lim.TMB," and the indicator ("Range.Indicator") designating whether the specified "Obs.Panel.TMB" value falls within the range of laboratory-specific TMB values supplied in the training data (i.e., in file.Model.Data.TMB). If multiple laboratory-specific panels are specified, this file will include regression results for all laboratory-specific panels that were entered ("Panel.1", "Panel.2", etc.) in a long data format so that each panel's information is appended with a variable specifying the panel number.

All.GLS.ML.RegistrationCurveFit.txt ,
All.GLS.ML.RegistrationCurveFit.Grid.txt

Background files describing the predicted regression curve estimates calculated at “Uniform.WES.TMB” values supplied for the training data in the file.Model.Data.TMB and on a function-generated fine grid (denoted by “.Grid” in the file name) to enable easy plotting of smooth curves. These files are located in the tmbLab_html folder. Note that the information in the file “All.GLS.ML.RegressionCurveFit.txt” is also summarized for use in the file “tmbLab.All.GLS.ML.FIT.TMB.txt” in the tmbLab_res folder.

All.GLS.ML.LowerLimitPredictionCurve.txt ,
All.GLS.ML.LowerLimitPredictionCurve.Grid.txt ,
All.GLS.ML.UpperLimitPredictionCurve.txt ,
All.GLS.ML.UpperLimitPredictionCurve.Grid.txt

Background files describing the lower and upper prediction limits for “Uniform.WES.TMB” values supplied for the training data in the file.Model.Data.TMB and on a function-generated fine grid (denoted by “.Grid” in the file name) to enable easy plotting of smooth curves. The fine grid will be the same as that used to generate the output file All.GLS.ML.RegressionCurveFit.Grid.txt. These files are located in the tmbLab_html folder. Note that the information in the files “All.GLS.ML.LowerLimitPredictionCurve.txt” and “All.GLS.ML.UpperLimitPredictionCurve.txt” is also summarized for use in the file “tmbLab.Lab.All.GLS.ML.FIT.TMB.txt” in the tmbLab_res folder.

4.2 tmbLabWES()

4.2.1 Description

This function estimates, from user-supplied data, a calibration curve with corresponding prediction limits to quantify the average relationship between WES and panel TMB assay results as well as variability around that curve. For a set of input WES values, this function also provides estimates of laboratory-specific panel diagnostic TMB and prediction limits for the laboratory-specific panel diagnostic assay. This function differs from the all-encompassing function `tmbLab()` in that it estimates the regression line and prediction limits, but does not further provide estimates of laboratory-specific TMB that are calibrated to WES TMB.

Maximum likelihood methods implemented in the `gls()` function in the R package “nlme” are used to fit a weighted least squares linear regression model. This model assumes a linear mean structure, Gaussian errors, and power variance structure. Users are advised to check the reasonableness of these assumptions for their own data.

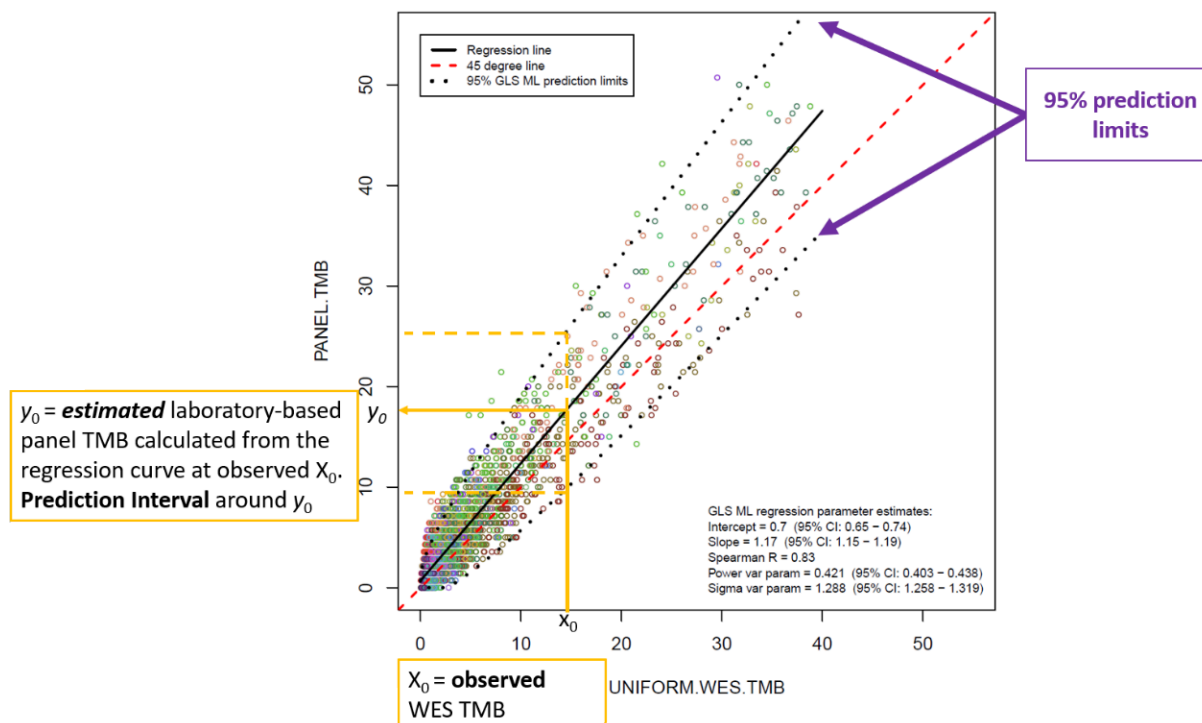
This function requires two input files. The first file, designated below as “file.Model.Data.TMB,” is a tab-delimited text file containing training data consisting of paired TMB values, one from a whole exome sequencing (WES) assay and one from a laboratory-specific panel diagnostic assay, for each of a series of samples. The second is a tab-delimited text file that contains WES values at which to output predicted laboratory-specific panel TMB values along with their corresponding prediction limits. Further details of the required formats for these files are given below under the “Arguments” section.

To better understand the model-fitting process, consider Figure 2. The first input file, which contains the training data, will correspond to the points depicted in the scatter plot with x-axis and y-axis corresponding to the WES TMB values (“Uniform.WES.TMB”) and laboratory-specific panel TMB values (“Panel.TMB”), respectively. This first data input file will include the information necessary to generate the regression line and prediction limits, denoted by the solid black line and the dotted black lines, respectively. The second input file provides a set of x_0 values, which are real or hypothetical WES values at which estimated laboratory-specific panel TMB values and prediction intervals are desired. In Figure 2, a single x_0 input value is depicted on the x-axis with a yellow vertical line. The laboratory-based panel TMB estimate pertaining to x_0 is derived as the predicted mean value on the

regression line and is depicted on the y-axis as y_0 . In addition, the prediction interval is provided for x_0 as depicted by the horizontal dashed lines.

Prior to regression fitting, this function drops observations (samples) from the training data supplied via the input file file.Model.Data.TMB with WES TMB values that are greater than 40. These observations are dropped from the analysis because data collected as part of Phase 2 of the FOCR TMB harmonization project [Vega DM 2021] we insufficient to assess modeling assumptions, particularly linearity, at these high TMB values. The source code is made available so that individual laboratories may modify this cutoff, if desired.

Figure 2. Predicted laboratory panel TMB for specific WES TMB values. Each plotted point represents WES TMB (x) and Panel TMB (y) for a particular sample in the training data set. The solid black line is the fitted weighted least squares regression line, and the dotted black lines are the corresponding 95% prediction limits. The value x_0 is a hypothetical WES TMB value. The predicted laboratory panel TMB value for x_0 is depicted by y_0 on the y-axis and is derived as the predicted mean value on the regression. The prediction limits around the laboratory panel TMB value are indicated on the y-axis with the dashed horizontal lines.



4.2.2 Usage

```
tmbLabWES(  
  file.Model.Data.TMB,  
  file.WES.Panel.TMB,  
  Pred.Interval = 95,  
  trunc.neg.flag = c("NEG", "TRUNCtoZERO"),  
  dir.output,  
  dir.result = "tmbLabWES_res",  
  show.HTML = TRUE  
)
```

4.2.3 Arguments

file.Model.Data.TMB Character string

This required file describes the laboratory-specific panel TMB training data file name and must be a tab-delimited text file with 3 or more columns:

Column 1: "Sample.ID"; the unique sample identifiers.

Column 2: "Uniform.WES.TMB"; the training set WES values.

Column 3 and more: "Panel.1", "Panel.2," etc. denoting one or more laboratory-specific panel TMB columns. It is anticipated that most users will provide only one set of laboratory-specific panel TMB data, corresponding to "Panel.1," however, the option is given to include data from multiple laboratory-specific panels by specifying additional columns.

Note also that the model implemented in this package assumes that samples are independent. Inputting multiple replicate TMB measurements from the same sample will not lead to appropriate statistical inference (e.g., prediction limits will be incorrect).

An example of this required file is as follows:

Sample.ID	Uniform.WES.TMB	Panel.1	Panel.2
S0000001	1.3410	1.0132	0.9345
S0000002	1.6414	3.4837	2.0914
S0000003	0.7608	0.3165	0.2429
S0000004	0.8482	1.5107	0.4991
S0000005	0.8301	3.4755	1.8225
.....			

file.WES.Panel.TMB Character string

This required file describes the name of the data file that contains WES TMB values at which panel TMB predicted values along with prediction limits will be calculated. The file must be a tab-delimited text file with 2 columns:

Column 1: "Sample.ID"; the unique sample identifiers.

Column 2: " WES.TMB "; WES TMB values for which predicted panel TMB and prediction limits will be calculated. Note that "WES.TMB" values may be real or hypothetical WES TMB values at which predicted panel TMB and prediction limits are desired. Typical hypothetical WES values of interest would be 5, 10, 15, and 20. The file must contain at least one WES.TMB value.

An example of this required file is as follows:

```
Sample.ID      WES.TMB
NewS0001      5
NewS0002      10
NewS0003      15
NewS0004      17
NewS0005      20
.....
```

Pred.Interval Numeric value

Specifies the prediction interval level of uncertainty. For example, Pred.Interval = 95 means that 95% of the observed values of panel TMB for a sample with the designated WES.TMB are predicted to fall within the interval. The inputted value must be a number between 1 and 100. Common values are 90, 95, and 99. Default is 95.

trunc.neg.flag Character string

If value is “NEG”, the code allows negative values of panel-based TMB to be used in the regression modeling. Although TMB values cannot actually be negative, some laboratories apply correction factors to panel TMB values that could cause small values to become negative. Negative values are not a problem for fitting the regression line, but they must be set to zero for purposes of variance and prediction limit calculation. Nonetheless, WES-calibrated estimates of TMB will always be truncated at zero so that they are not reported as negative values.

If “TRUNCtoZERO”, the code truncates negative panel-based TMB values to zero prior to their use in regression modeling.

`dir.output` Character string

This argument can be used to specify the directory output file path. Default is <user’s home directory>/tmb.

`dir.result` Character string

This argument can be used to specify the folder name in which the results will be placed. This folder will be found within the folder “dir.output”. Default is “tmbLab_res”.

`show.HTML` Logical

If TRUE, the HTML summary document will automatically pop up in the system default browser. Default is TRUE. A copy of this output is also saved for later viewing (see tmbLabWES.html output file).

4.2.4 Value

The output list includes the following elements:

- `Obs.PANEL.TMB.vec` : observed panel-based TMB values.
- `Obs.PANEL.IDs.vec` : observed panel-based TMB sample IDs.
- `PANEL.vec` : the panel names which were provided in the file `file.Model.Data.TMB`.
- `Pred.Interval` : the prediction interval level of uncertainty.
- `trunc.neg.flag` : the option for allowing negative values of panel-based TMB.
- `dir.resPath` : the path for the final result files.
- `dir.output` : the path for the output files.
- `predMethod` : the calculation method: 1 for "tmbLab"; 2 for "tmbLabWES".
- `Lab.All` : a data frame including predicted assay TMB value with prediction intervals.

The following output files are generated:

`tmbLabWES.html`

The HTML output popup, which appears automatically if show.HTML=TRUE, includes a summary of relevant information. This same information is saved in a file that will be written to the specified dir.output. The preprocessing section describes the total number of samples, the number of samples that were dropped because they had WES TMB values >40, and the range of data values observed. The table provides the second data file input values ("Sample.ID" and "WES.TMB") along with the regression estimates ("Est.TMB") and the lower and upper prediction intervals corresponding to these values ("Lower.Lim.TMB" and "Upper.Lim.TMB"). The next column, "Range.Indicator", indicates whether the specified "WES.TMB" value falls within the range of WES TMB values (e.g., "Panel.1") described in file.Model.Data.TMB. "Range.Indicator" will be designated "In" if the specified "WES.TMB" value falls within the range of WES TMB values in the training data (e.g., the range of "Panel.1" described in file.Model.Data.TMB) and will be designated "Out" if the specified "WES.TMB" value falls outside this range. For reference, this range is also described in the preprocessing section of the HTML output. The lower prediction interval limits are set to 0 if the lower limit is below 0. In the last column, a link is provided to a .pdf plot of the scatterplot along with the calibration fitted regression line (solid black line), calibration regression line parameters (text in bottom right of plot), and the prediction limits (dotted black line). Portions of the information that are summarized in this HTML file are also included in various output files in the dir.output folder, "tmbLabWES_html" and "tmbLabWES_res," as described below.

Panel.n.GLS.ML.scatterplot.pdf

Scatterplots depicting laboratory-specific TMB values (y-axis) versus WES TMB values (x-axis), a 45 degree line (dashed red line), the superimposed regression line (solid black line), and Pred.Interval (Default = 95)% prediction intervals (dotted black line) are in the tmbLabWES_html folder and are linked to in the final column of the tmbLabWES.html file. GLS ML regression parameter estimates are also provided in the bottom righthand corner of the plots, including the intercept, slope, Spearman R, Power variance parameter, and sigma variance parameter. The x and y axes of the plots range from 0 to 55 as these lower values are of greatest relevance for clinical decision making. Users are advised to check for possible outliers that lie beyond the upper range of the y-axis as these outliers have the potential to be influential. The naming convention of the scatterplots is such that n in the name "Panel.n.GLS.ML.scatterplot.pdf" refers to the corresponding panel number specified in the columns of file.Model.Data.TMB. If only "Panel.1" is specified, only one plot will be generated with the name "Panel.1.GLS.ML.scatterplot.pdf."

All.GLS.ML.RegressionParameters.txt

Detailed regression parameters, found in the `tmbLabWES_res` and `tmbLabWES_html` folder. The model is fit as described in the background section. Each Panel for which data are supplied in `file.Model.Data.TMB` will correspond to a row in this file. The information and parameters described in this file include `N`, the number of paired laboratory-specific TMB and WES TMB points included in the training data used for fitting the calibration model; `N.NEG.TMB`, number of negative panel TMB values provided (for most laboratories this number will be zero); `MODEL.AIC`, the Akaike information criterion for the fitted regression model; `MODEL.BIC`, the Bayesian information criterion for the fitted regression model; `BETA0`, `BETA0.SE`, `BETA0.95CI.LL`, `BETA0.95CI.UL`, which describe the estimated intercept parameter, the corresponding standard error, lower 95% confidence limit, and upper 95% confidence limit, respectively; `BETA1`, `BETA1.SE`, `BETA1.95CI.LL`, `BETA1.95CI.UL`, which describe the estimated slope parameter, the corresponding standard error, lower 95% confidence limit, and upper 95% confidence limit, respectively; `SPEARMAN.R`, which describes the Spearman correlation coefficient reflecting the association between panel TMB and WES TMB in the training; `POW.PARAM`, `POW.SE`, `POW.95CI.LL`, `POW.95CI.UL`, which describe the estimated power parameter, the corresponding standard error, lower 95% confidence limit, and upper 95% confidence limit, respectively; `SIGMA.PARAM`, `SIGMA.95CI.LL`, `SIGMA.95CI.UL`, which describe the estimated sigma parameter, the corresponding lower 95% confidence interval, and the corresponding upper 95% confidence interval, respectively; `LSIGMA.PARAM`, `LSIGMA.SE`, which describe the estimated sigma parameter on the natural log (base e) scale and the corresponding standard error, respectively; `RESID.VAR`, which describe the standard error of the residuals corresponding to the fitted model. Refer to the package `nlme()` for more information.

`tmbLab.All.GLS.ML.PRED.txt`

Results file containing columns pertaining to “Sample.ID”, the panel identifier “Panel.n”, WES values at which predicted panel TMB values will be estimated (“WES.TMB” from file.WES.Panel.TMB), predicted panel TMB values “Est.TMB”, lower limit of the prediction interval “Lower.Lim.TMB,” upper limit of the prediction interval “Upper.Lim.TMB”, and the indicator (“Range.Indicator”) designating whether the specified “Obs.WES.TMB” value falls within the range of WES TMB values supplied in the training data (i.e., in file.Model.Data.TMB). If multiple laboratory-specific panels are specified, this file will include regression results for all laboratory-specific panels for which data were entered (“Panel.1”, “Panel.2”, etc.) in a long data format so that each panel’s information is appended with a variable specifying the panel number.

All.GLS.ML.RegistrationCurveFit.txt ,
All.GLS.ML.RegistrationCurveFit.Grid.txt

Background files describing the predicted regression curve estimates for “Uniform.WES.TMB” values supplied for the training data in the file.Model.Data.TMB and on a function-generated fine grid (denoted by “.Grid” in the file name) to enable easy plotting of smooth curves. These files are located in the tmbLabWES_html folder. Note that the information in the file “All.GLS.ML.RegistrationCurveFit.txt” is also summarized for use in the file “tmbLab.All.GLS.ML.PRED.txt” in the tmbLabWES_res folder.

All.GLS.ML.LowerLimitPredictionCurve.txt ,
All.GLS.ML.LowerLimitPredictionCurve.Grid.txt ,
All.GLS.ML.UpperLimitPredictionCurve.txt ,
All.GLS.ML.UpperLimitPredictionCurve.Grid.txt

Background files describing the lower and upper prediction limits for “Uniform.WES.TMB” values supplied for the training data in the file.Model.Data.TMB and on a function-generated fine grid (denoted by “.Grid” in the file name) to enable easy plotting of smooth curves. The fine grid will be the same as that used to generate the output file All.GLS.ML.RegistrationCurveFit.Grid.txt. These files are located in the tmbLabWES_html folder. Note that the information in the files “All.GLS.ML.LowerLimitPredictionCurve.txt” and “All.GLS.ML.UpperLimitPredictionCurve.txt” is also summarized for use in the file “tmbLab.Lab.All.GLS.ML.PRED.txt” in the tmbLabWES_res folder.

5 References

Vega DM, Yee LM, McShane LM, et al. Aligning Tumor Mutational Burden (TMB) quantification across diagnostic platforms: Phase 2 of the Friends of Cancer Research TMB Harmonization Project. Submitted to *Annals of Oncology* 2021.