Glossary

1) Genes/Exons to be sequenced by cancer type 2

1.1) config.json 2

1.2) gene_data.txt 3

1.3) model_data.txt 3

2) TMB Prediction by amount of mutations 5

2.1) config.json 5

2.2) data_by_mutations.txt 6

2.3) gene_data.txt 6

2.4) model_data.txt 6

2.5) predicted_tmb.txt 7

3) TMB Prediction by mutation coordinates 8

4) Graphical Results 8

*4.1) gene-exon_statistics_model_***.pdf* 8

*4.2) TMB-prediction_***.pdf* 8

1) Genes/Exons to be sequenced by cancer type

1.1) config.json

Dictionary format file that includes keys and values related to your choices in the form.

KEYS	KEYS	VALUES
"model"	“cancer”	It contains your selection for cancer type in a three-letter format.
"model"	"quality"	It contains grade model parameters: • “B” (Best): Models with highest accuracy and lowest error. **By default** • “S” (Suggested): Other feasible model options • “X” (All): It includes best models, suggested models and others.
"model"	"databy"	• “G” (Genes). Variable in the models. **By default** • “E” (Exons). Variable in the models. Only available after selecting **All** in the quality model parameter.
"model"	"subset"	• “O” (Our): Our proposed selection of genes to make the models. **By default** • “C” (Cancer Gene Census): Selection of genes/exons to make the models proposed by CGC. Only available after selecting **All** in the quality model parameter. • “F” (Foundation-One panel): Selection of genes/exons to make the models proposed by FO. Only available after selecting **All** in the quality model parameter.
“email”		If you had included an email account, it will be shown here.

1.2) gene_data.txt

Table that includes gene/exon characteristics of the model(s):

COLUMNS	
id_model	Number of identification for the model by cancer type.
gene_exon_id	Gene/Exon ID (Ensembl nomenclature) for each one present in the model.
coefficient	Gene/exon's coefficient from the linear regression algorithm.
hugo_symbol	By Gene Nomenclature Committee (HGNC).
Chromosome	Chromosomal position of the gene/exon.
Start	Start position of the gene/exon.
End	End position of the gene/exon.

1.3) model_data.txt

Table that includes information about the model(s) for the TMB prediction:

COLUMNS	
id_model	Number of identification for the model by cancer type.
cancer_type	Full cancer type name in capital letters.
cancer_id	Cancer type name in a three-letter format.
model_by_gene	(1) for GENES or (0) for EXONS.
Mb	Size of the genomic region expressed in MB.
gene_exon_model	Number of genes/exons required for the TMB prediction for that model.
Constant	Fixed number from the linear regression algorithm for each model.

Strategy	Selection of genes to make the models (Our Selection, Cancer Gene Census, Foundation-One panel).
Classification	Quality parameter for the prediction analysis (Best Model, Suggested Model, Other Model).
R2	Coefficient of Determination between predicted TMB compared to published experimental TMB for each model.
R2_val_interna	Coefficient of Determination for consensus models under 1000 repetitions of Bootstrap internal validation.
R2_val_externa	Coefficient of Determination between consensus models compared to an external validation using samples from COSMIC.
RSE	Relative Standard Error between predicted TMB compared to published experimental TMB.
RSE_val_interna	Relative Standard Error for consensus models under 1000 repetitions of bootstrap internal validation.
RSE_val_externa	Relative Standard Error between consensus models compared to an external validation using samples from COSMIC.

2) TMB Prediction by amount of mutations

2.1) config.json

Dictionary format file that includes keys and values related to your choices in the form. There could be more than one dictionary in the "config.json", these are shown in brackets:

KEYS	KEYS	VALUES
"efile"	“filename”	It is your input file named as "data_by_mutations.txt".
"efile"	“delimiter”	The type of delimitation you selected on the website, according to the type of your input file.
"efile"	"header"	Presence of heading, according to your input file.
“column”	"gene_exon_id"	Positional number of the column containing information about “gene/exon ID” of your input file.
“column”	"mutations"	Positional number of the column containing information about mutations in your input file.
"model"	“cancer”	It contains your selection for cancer type in a three-letter format.
"model"	"quality"	It contains grade model parameters: • “B” (Best): Models with highest accuracy and lowest error. **By default** • “S” (Suggested): Other feasible model’s options • “X” (All): It includes best models, suggested models and others.
"model"	"databy"	• “G” (Genes). Variable in the models. **By default** • “E” (Exons). Variable in the models. Only available after

		selecting **All** in the quality model parameter.
"model"	"subset"	• “O” (Our): Our proposed selection of genes to make the models. **By default** • “C” (Cancer Gene Census): Selection of genes/exons to make the models proposed by CGC. Only available after selecting **All** in the quality model parameter. • “F” (Foundation-One panel): Selection of genes/exons to make the models proposed by FO. Only available after selecting **All** in the quality model parameter.
“email”		If you had included an email account it will appear on here
"warning_gene_exon_**"		There could be several alerts depending on the characteristics of your input file. They will be detailed in the "text" section inside de warning box. These alerts mean that the TMB prediction results could be altered if the input file did not satisfy the required conditions for a specific cancer type.

2.2) data_by_mutations.txt

Your input file.

2.3) gene_data.txt

Table that includes gene/exon characteristics of the model(s). Idem item 1.2

2.4) model_data.txt

Table that includes information about the model(s) for the TMB prediction: Idem item 1.3

2.5) predicted_tmb.txt

Prediction results for the analysis of your data input:

COLUMNS	
id_model	Number of Identification for the model by cancer type.
sample_id	Input file named as "data_by_mutations.txt".
predicted_TMB	*Tumor Mutation Burden prediction **
cancer_type	Full cancer type name in capital letters.
cancer_id	Cancer type name in three-letter format.
model_by_gene	yes/no
Strategy:	Selection of genes to make the models (Our Selection, Cancer Gene Census, Foundation-One panel).
Classification	Grade parameter for the prediction analysis (Best Model, Suggested Model, Other Model).
gene_exon_model	Number of genes/exons required for the TMB prediction for that model.
gene_exon_submited	Number of genes/exons uploaded in the input file.
gene_exon_not_used	Number of genes/exons not necessary for the prediction for that model.
gene_exon_as_cero_mutations	Number of genes/exons required in the model to calculate the TMB prediction, but not present in the input file. These will be considered with zero mutations.
gene_exon_used	Number of genes/exons used from your input file to calculate the TMB prediction for that model.

Threshold value suggested for medical use: 150 mutations.

3) TMB Prediction by mutation coordinates

In this case,"data_by_coordenates.txt" is your input file which will be converted into "data_by_mutations.txt" to calculate the TMB prediction. Consequently, the resulting files are the same as in “TMB Prediction by amount of mutations”

4) Graphical Results

*4.1) gene-exon_statistics_model_***.pdf*

It sums up graphically the statistics related to the TMB calculus included in “predicted_tmb.txt”. It is relevant to point out that for a reliable TMB prediction it is advisable to upload the mutations in the genes/exons that the models require, even if the number of mutations in those genes is zero. On the other hand, in case you don’t include genes/exons needed in the model, the reliability of the prediction can not be assured.

Those genes/exons which were uploaded but whose mutations are not relevant for the model for the specified cancer type will not be considered to execute the calculus.

We highly recommend to check in the website the genes/exons to be sequenced by cancer type to have previous information on what genes/exons to upload to have a solid TMB prediction for your sample.

*4.2) TMB-prediction_***.pdf*

It returns the TMB prediction for the model(s) as dots. The threshold value suggested for medical use (150 mutations) is plotted as a dotted red line. This would collaborate with the discrimination between patients responding (dots in red) and not-responding (dots in black) to ICB therapy .