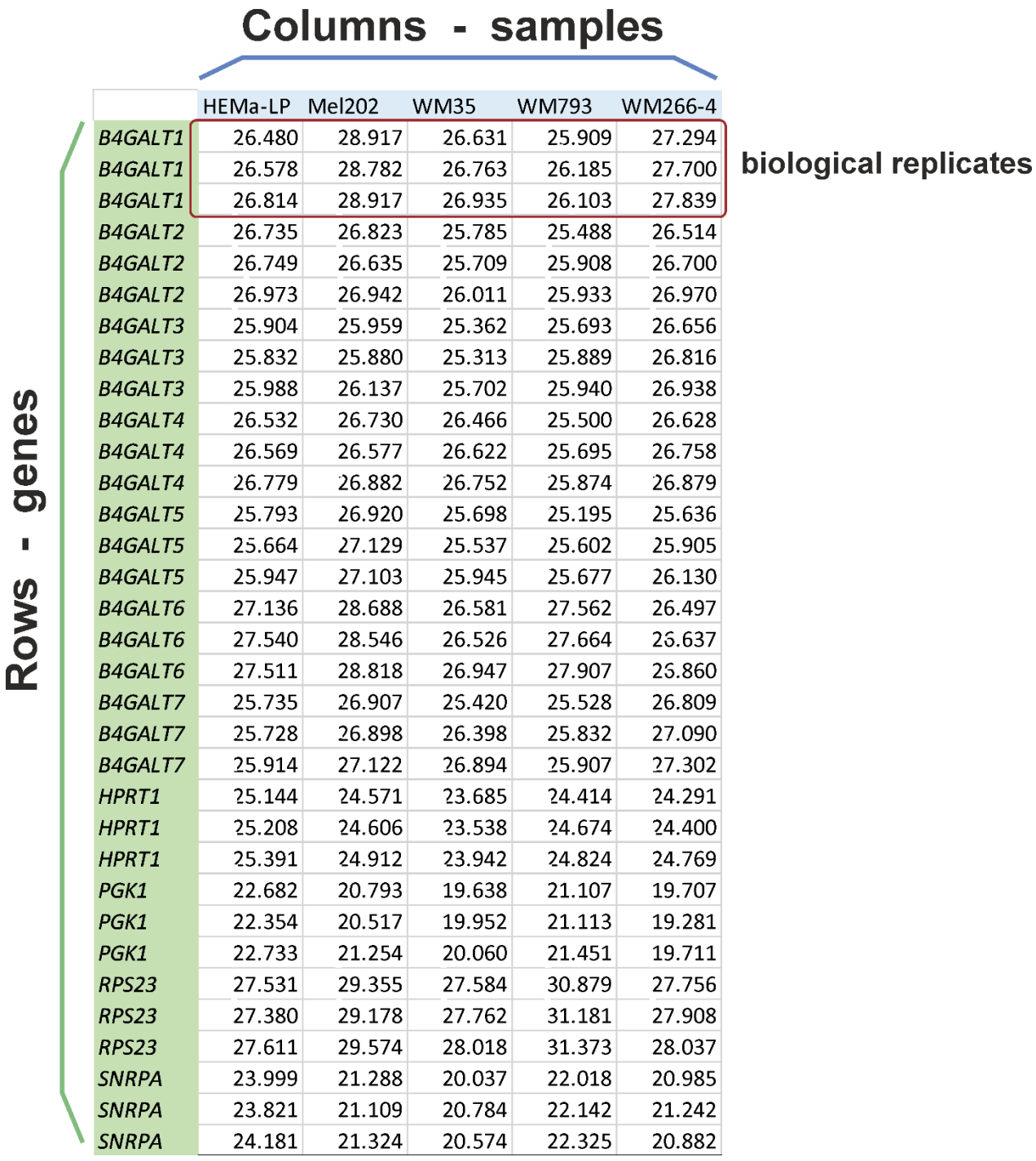
Data format

Input data format

The program accepts data in the form of .xls or .xlsx (excel spreeadsheet) file with the following format:



	HEMa-LP	Mel202	WM35	WM793	WM266-4
B4GALT1	26.480	28.917	26.631	25.909	27.294
B4GALT1	26.578	28.782	26.763	26.185	27.700
B4GALT1	26.814	28.917	26.935	26.103	27.839
B4GALT2	26.735	26.823	25.785	25.488	26.514
B4GALT2	26.749	26.635	25.709	25.908	26.700
B4GALT2	26.973	26.942	26.011	25.933	26.970
B4GALT3	25.904	25.959	25.362	25.693	26.656
B4GALT3	25.832	25.880	25.313	25.889	26.816
B4GALT3	25.988	26.137	25.702	25.940	26.938
B4GALT4	26.532	26.730	26.466	25.500	26.628
B4GALT4	26.569	26.577	26.622	25.695	26.758
B4GALT4	26.779	26.882	26.752	25.874	26.879
B4GALT5	25.793	26.920	25.698	25.195	25.636
B4GALT5	25.664	27.129	25.537	25.602	25.905
B4GALT5	25.947	27.103	25.945	25.677	26.130
B4GALT6	27.136	28.688	26.581	27.562	26.497
B4GALT6	27.540	28.546	26.526	27.664	26.637
B4GALT6	27.511	28.818	26.947	27.907	26.860
B4GALT7	25.735	26.907	26.420	25.528	26.809
B4GALT7	25.728	26.898	26.398	25.832	27.090
B4GALT7	25.914	27.122	26.894	25.907	27.302
HPRT1	25.144	24.571	23.685	24.414	24.291
HPRT1	25.208	24.606	23.538	24.674	24.400
HPRT1	25.391	24.912	23.942	24.824	24.769
PGK1	22.682	20.793	19.638	21.107	19.707
PGK1	22.354	20.517	19.952	21.113	19.281
PGK1	22.733	21.254	20.060	21.451	19.711
RPS23	27.531	29.355	27.584	30.879	27.756
RPS23	27.380	29.178	27.762	31.181	27.908
RPS23	27.611	29.574	28.018	31.373	28.037
SNRPA	23.999	21.288	20.037	22.018	20.985
SNRPA	23.821	21.109	20.784	22.142	21.242
SNRPA	24.181	21.324	20.574	22.325	20.882

Illustration 1: Input data format

- **COLUMNS** - samples of the experimental model to be analysed. First-row should contain a name of each sample for one column.
- **ROWS** - reference and target genes for analysis (selection is done inside the program). The first column should contain a name for the gene in a given row. Genes should be organized in continuous blocks and the biological replicates should be one below the other.
- **VALUES** - For 'combined input' data values are expected to be Ct values.

Interface

Main Window

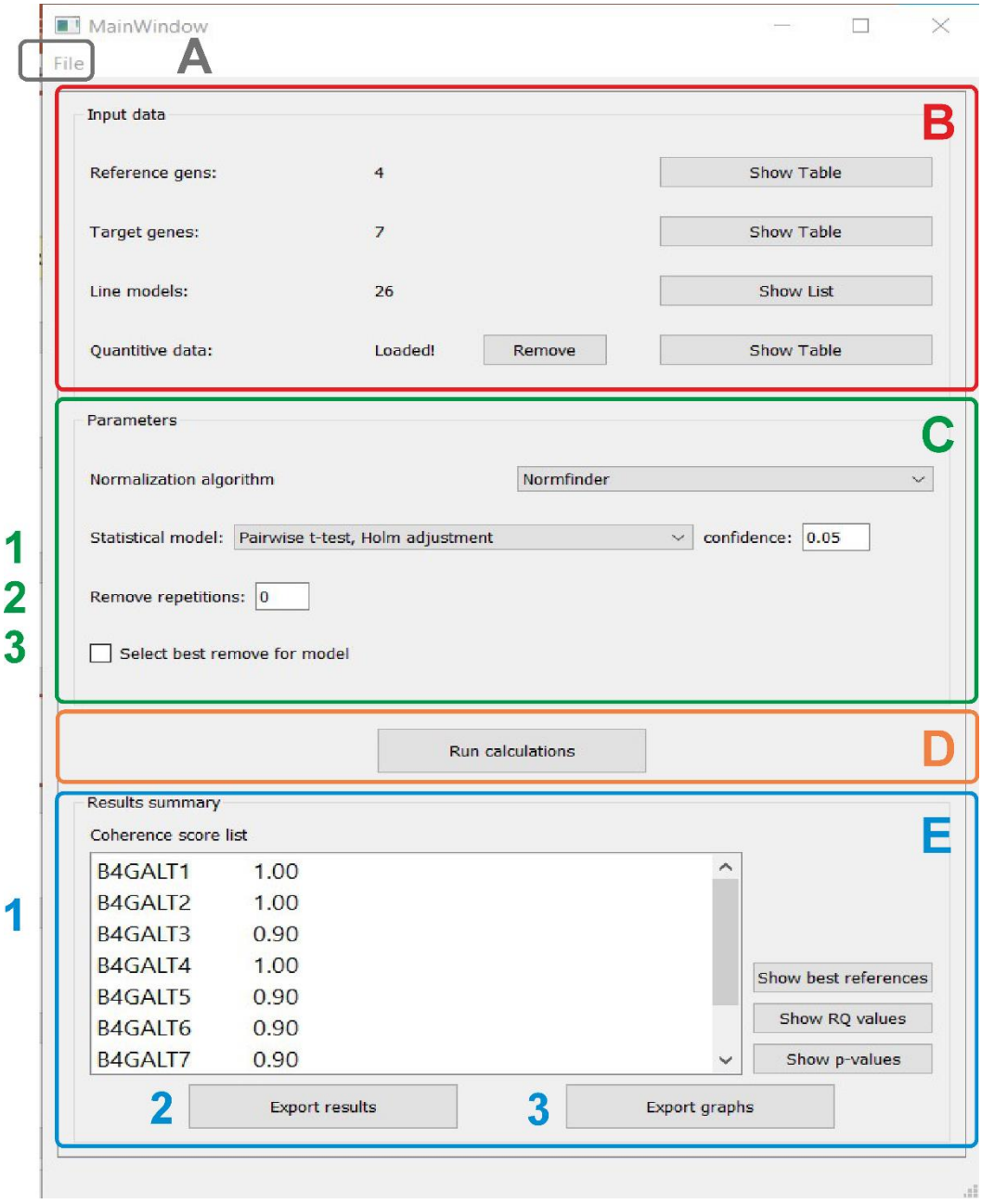


Illustration 2: Main window

- **A** – *Menu* from which we can perform 2 actions:
  - Read from combined input – a method for loading data into the program
  - Load Quantified – a method for loading quantified data, this action can be performed only after normal data is loaded.
- **B** – *Input data* section for displaying information about loaded data. "Show ***" buttons can be used to look up loaded information in the form of tables.
- **C** - *Parameters* section for selecting parameters for the algorithm. This section is active only after the input data is properly loaded.
  - 1) Selection of statistical models with the confidence value for evaluating if the difference between samples for a given target gene is statistically significant.
  - 2) Selecting the repetition number of removing the worst reference gene from the list passed to the normalization algorithm.
  - 3) Checkbox for selecting algorithm mode:
    - checked – for each remove repetition series only remove with best stability score will be selected for further analysis,
    - unchecked – all remove repetition levels will be used.
- **D** - Run calculations button - this action will be available only if input data is loaded and a proper parameters are selected
- **E** - Results summary section
  - 1) List displaying calculated coherence score for each target gene and all target genes average.
  - 2) Export results in form of single excel file
  - 3) Export results in form of boxplots graphs in the form of separate .png files for each target gene-remove repetition.

Load input data window

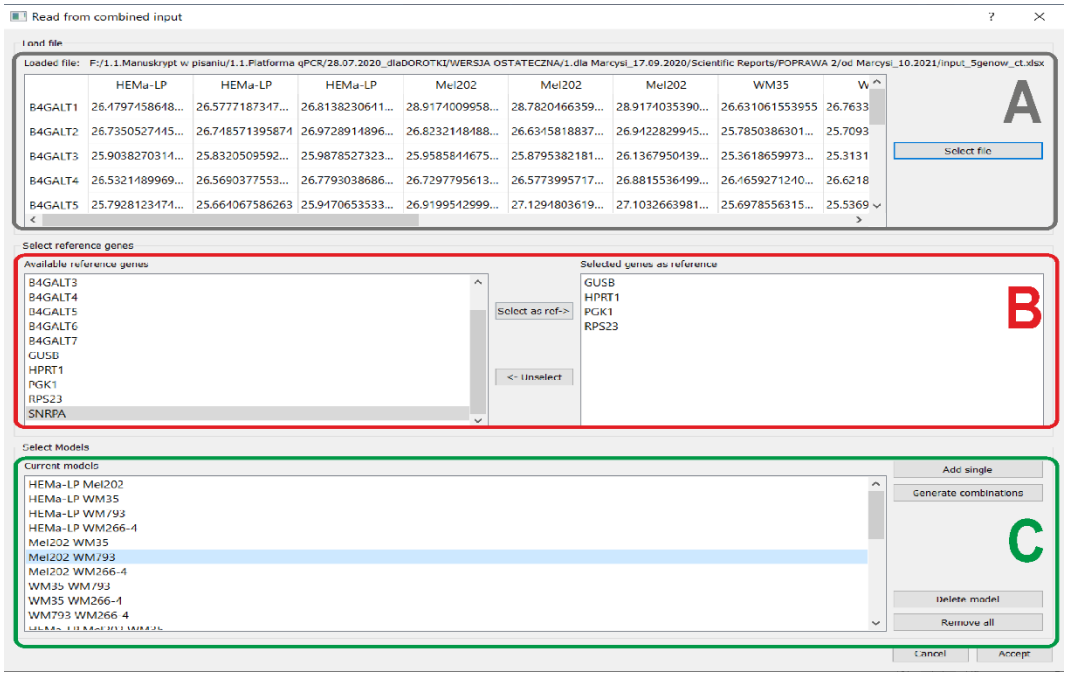


Illustration 3: Load input data window

- **A** - Section for selecting the file from the drive to load (Select file)
- **B** - Section for selecting which genes should be used as reference genes. After selecting the gene from the available gene list (left window) press the "*Select as ref->*" button to add this gene to the reference gene list. Unselected genes are assumed to be target genes.
- **C** - Selection of samples models combinations. For automatic generation press "Generate combinations" which will generate all possible combinations of samples. For manual addition of model form list of possible samples use the "*Add single*" button.

After a successful file load, selecting a target and reference genes and models use "Accept" to load this data to the main application.

Standard use case

1. From "*File*" menu select "*Read from combined input*" to open the window for loading data.
2. In "*Read from combined input*" use "*Select file*" button to load data from the spreadsheet.
3. In "*Select reference genes*" select reference genes to be used in analyze, unselected genes are assumed to be target genes.
4. In "*Select Models*" section create models of samples.
5. Use "*Accept*" button to load data into program memory and return to main window.
6. * If quantified data is available it can be loaded using "*File → Load Quantified*"
7. "*Parameters*" section set parameters for analysis.
8. Press "*Run calculations*" to perform analysis.
9. Export results in form of excel spreadsheet using "*Export results*" or as boxplot representation using "*Export graphs*".

If you use this software, please cite it as below:

1) Janik ME, Szwed S, Grzmil P, Kaczmarek R, Czerwiński M, Hoja-Łukowicza D. RT-qPCR analysis of human melanoma progression-related genes – A novel workflow for selection and validation of candidate reference genes. Int. J. Biochem. Cell Biol. 101 (2018) 12–18; doi: https://doi.org/10.1016/j.biocel.2018.05.007.
2) Hoja-Łukowicz, D., Maciążek, D., Kościelniak, P. et al. Innovative GenExpA software for selecting suitable reference genes for reliable normalization of gene expression in melanoma. Sci Rep 12, 3331 (2022). https://doi.org/10.1038/s41598-022-07257-6.