Documentation for NormFinder
version 21
January 2024

Documentation authors: Claus Lindbjerg Andersen, Christian Gundesen, Katja Adolf
Excel add-in "NormFinder.xla" authors: Jens Ledet Jensen, Claus Lindbjerg Andersen and Christian Gundesen

Visit moma.dk for additional information on NormFinder.

Content

The "NormFinder" algorithm

"NormFinder" is an algorithm for identifying the optimal normalization gene among a set of candidates. It ranks the set of candidate normalization genes according to their expression stability in a given sample set and given experimental design.

The algorithm is rooted in a mathematical model of gene expression and uses a solid statistical framework to estimate not only the overall expression variation of the candidate normalization genes, but also the variation between sample subgroups of the sample set e.g. normal and cancer samples. Notably, "NormFinder" provides a stability value for each gene, which is a direct measure for the estimated expression variation enabling the user to evaluate the systematic error introduced when using the gene for normalization.

The model and statistical framework underlying "NormFinder" are described in
 *Andersen C.L. et al., "Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets", Cancer Res 2004;64 5245-5250.*The full text article can be downloaded from
http://cancerres.aacrjournals.org/cgi/content/abstract/64/15/5245?etoc

"NormFinder" can analyze expression data obtained through any quantitative method e.g. real time RT-PCR and microarray based expression analysis. The input data is supposed to be on a linear scale. Thus, Ct values from a real time RT-PCR run should not be used directly. Use a standard curve or the delta-Ct method to transform the Ct values to linear scale expression quantities.

License

"NormFinder.xla" is an Add-In for Microsoft Excel which adds the NormFinder functionality directly to the Microsoft Excel software package for easy use.

The software is free to use for both academical and commercial use. Modification or redistribution of the Normfinder.xla is not allowed.

When publishing scientific results, where NormFinder software has been used, please cite the original article:

Andersen C.L. et al., "Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets", Cancer Res 2004;64 5245-5250.

Disclaimer

The Excel Add-in "NormFinder" is provided by the Dept. of Molecular Medicine (MOMA), Aarhus University Hospital, Denmark, former MDL – Molecular Diagnostic Laboratory. "NormFinder" is provided as is, and MOMA does not make any warranty, express or implied, with respect to the use of "NormFinder". By using "Normfinder" you accept that in no event will MOMA be liable for any indirect, punitive, special, incidental or consequential damages however they may arise and even if MOMA has been previously advised of the possibility of such damages. ©2004-2024 MDL/MOMA. All rights reserved.

How to Install the "NormFinder.xla" Add-In

Office 365, Office 2019, 2016, 2010, 2007

Download the file "NormFinder_0953.xla" from moma.dk.
Double-click to add NormFinder to the Add-Ons Tab in Excel. Depending on your security settings, you might need to permit the Add-on in the Security Settings (Start - Excel configuration).
Start NormFinder and agree to the disclaimer in order to use it.
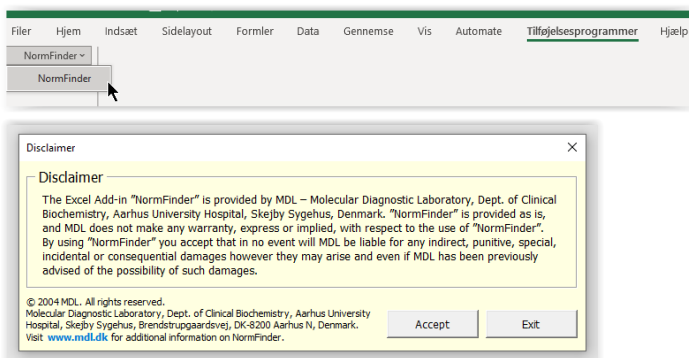
Office 2003

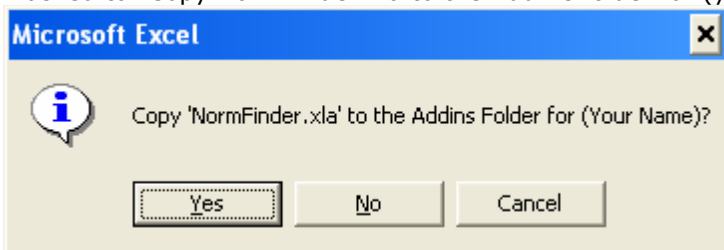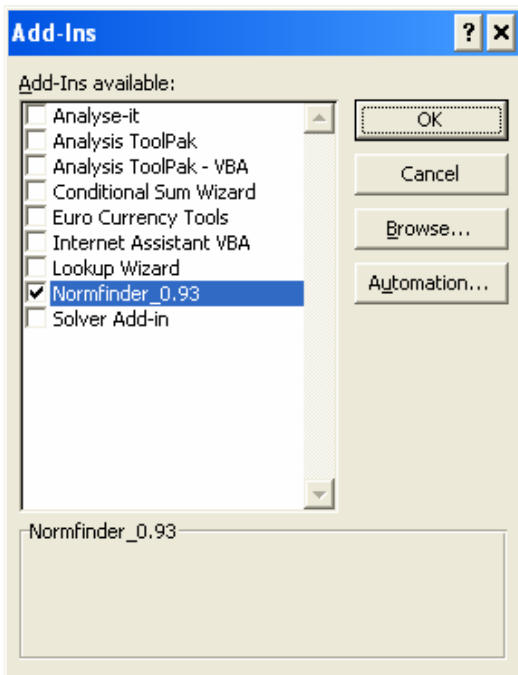Open Excel and choose **Tools** from the menu-bar. Scroll down and select the submenu **Add-Ins...**
From the "Add-Ins" dialog select [Browse…] and locate the "NormFinder.xla" file.
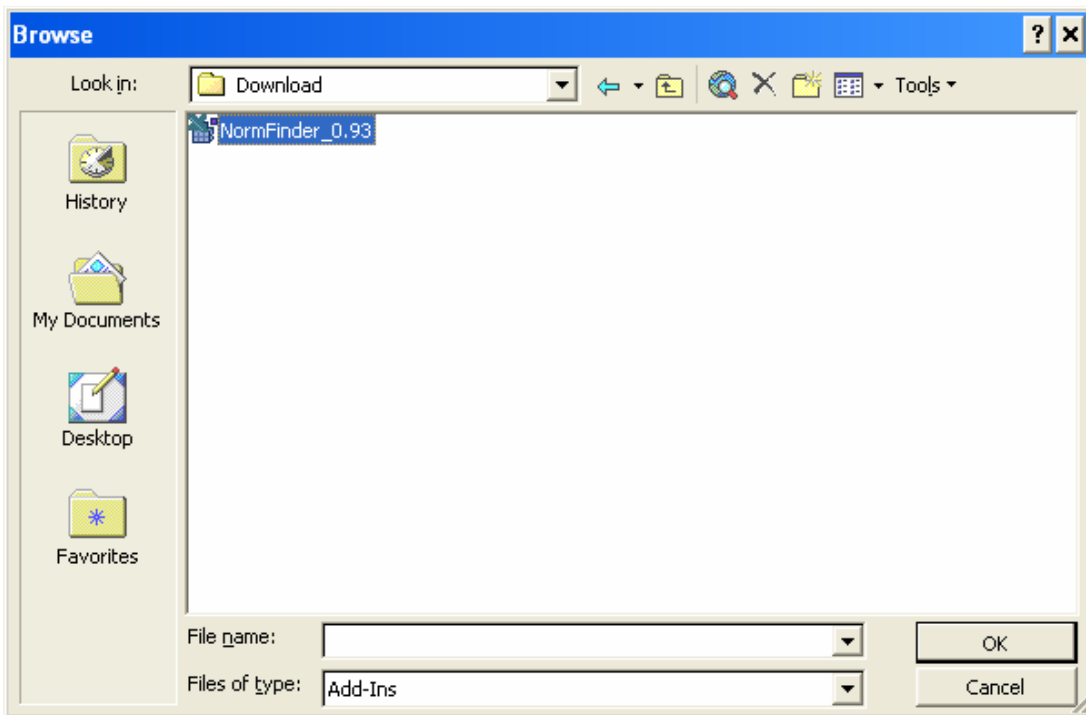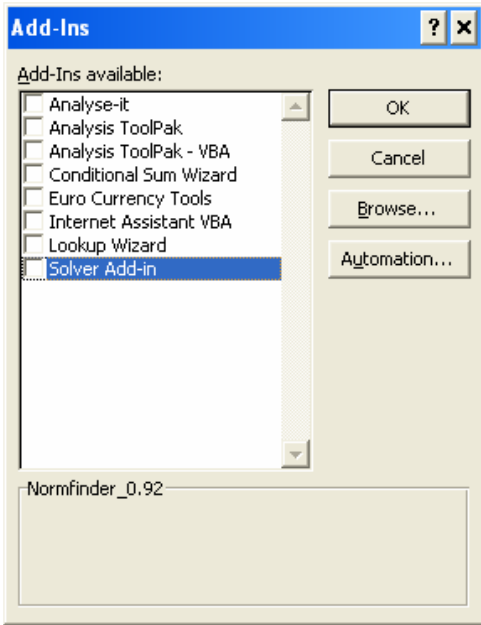
If asked to "Copy NormFinder.xla to the AddIns folder for (you Name)?" choose **[Yes]**.

"NormFinder" will now appear in the list of Add-Ins. Make sure it is checked and then select **[OK]** in the "Add-Ins" Dialog.

If everything went fine – you now have a "NormFinder" menu in your Excel menu-bar.

Instructions how to use NormFinder

Basic requirements for NormFinder analysis

The mathematical method applied in NormFinder requires the analysis of a minimum of 3 genes and a minimum of 2 samples per group. Generally we recommend analysis of 5-10 candidate genes and at least 8 samples per group.

Instructions

To evaluate the expression stability of your candidate normalization genes open the file with your input data. The input data should be organized with the first column containing the gene names, and the first row containing the sample names. If appropriate the last row should contain the sample-group identifiers, which are supposed to be integers.

Then select "NormFinder" from the menu-bar.

Use the dialog box to select the input data.

Next check the fields "Sample names", "Gene names", "Group identifier", "log transform data" and "simple output only" as appropriate.
The data are supposed to be on a linear scale. If your data are already log transformed (natural base (e) logarithm) this is acceptable too, if you uncheck the field "log transformed data".

Finally, indicate whether a simple or extended output is wanted.

NormFinder

Getting Started

Select input data: [   ] [_]

☐ Sample names (first row) included

☐ Gene names (first column) included

☐ Group identifier (last row) included

☑ Log transform data
(natural base (e) logarithm)

Go

☑ Simple output only

Exit
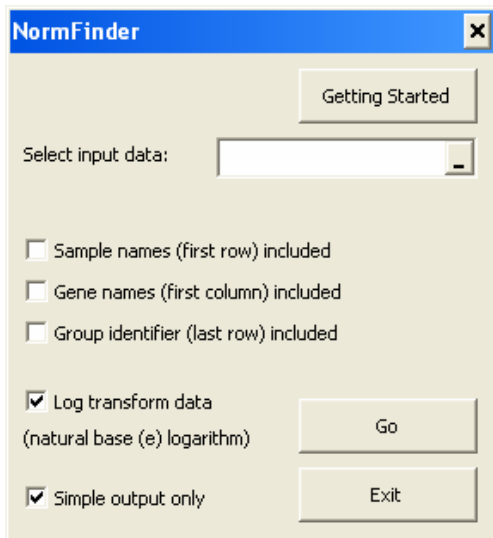
The simple output provides the user with:

1. A stability value for each candidate gene. The best gene with the lowest stability value is highlighted.
2. The best combination of two genes for a two gene normalization factor (only if group identifiers were included).
3. A stability value for the best combination of two genes (only if group identifiers were included).

The extended output provides in addition:

4. The standard error of the stability value (only if group identifiers were not included)
5. The estimated intragroup variations for each gene and for each group (only if group identifiers were included).
6. The estimated intergroup variations for each gene (only if group identifiers were included).

Description of the output

Stability values:
If group identifiers are included, the stability value is the average of ρ_{ig} from Eq. C.

If no group identifiers are used, the stability value is $\hat{\sigma}_i$ from Eq. B. In the latter case the extended output provides in addition the standard error for $\hat{\sigma}_i$.

The stability value for the best combination of two genes is ρ_A from equation (1.10) in supplementary information.

Intragroup variation:
The intragroup variation is given as from Eq. B. $\hat{\sigma}_{ig}^2$

Intergroup variation:
The intergroup variation is given as d_{ig} which enters Eq. C.

Example

The top ranked gene (which has the smallest stability value) is the candidate gene most stably expressed in the sample set investigated.

Whether the top ranked gene is sufficiently stably expressed for the purpose intended, cannot be determined from the stability value. The users of NormFinder will have to evaluate this themselves.

The following example illustrates how such an evaluation can be performed.

Example data: *See the file "Example Bladder data variance estimation.xls" (available from* [*https://www.moma.dk/software/normfinder*](https://www.moma.dk/software/normfinder)*).*

In the example the experimental design consists of two sample groups – the simplest situation.

In this case you can evaluate the effect of using a given candidate for normalization by plotting the data produced by NormFinder.

Run Normfinder and uncheck "simple output only". Then NormFinder will output the estimated inter- and intra-group variances.

Plotting the inter-group variances will show the effect of using any of the candidates for normalization. We look for a candidate gene with an inter-group variance as close to zero as possible. To obtain confidence intervals on the inter-group variances, you will have to plot the average of the intra-group variances as error bars to inter-group variances.

Still what we are aiming at is a candidate with an inter-group variation as close to zero as possible, and at the same time having as small errors bars as possible. The candidate found this way, will be the top-ranked candidate picked by NormFinder.

Generally, the larger expression differences you are searching for, the larger inter-group variation you can accept! However, it is clear that you want to minimize the inter-group variation.

If data from a gene or two, known to be deregulated, is included in the plot described above, then the effect of using the different normalization candidates can directly be evaluated by the ability of NormFinder to identify these genes as deregulated. Such a plot will in most instances enable the user to evaluate whether one or more of the tested candidates behave sufficiently well for the intended purpose.

Happy NormFinding…