

© 2004 MDL. All rights reserved.

Molecular Diagnostic Laboratory, Dept. of Clinical Biochemistry, Aarhus University Hospital, Skejby Sygehus, Brendstrupgaardsvej, DK-8200 Aarhus N, Denmark.

Visit www.mdl.dk for additional information on NormFinder.

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 paper can be found at <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.

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

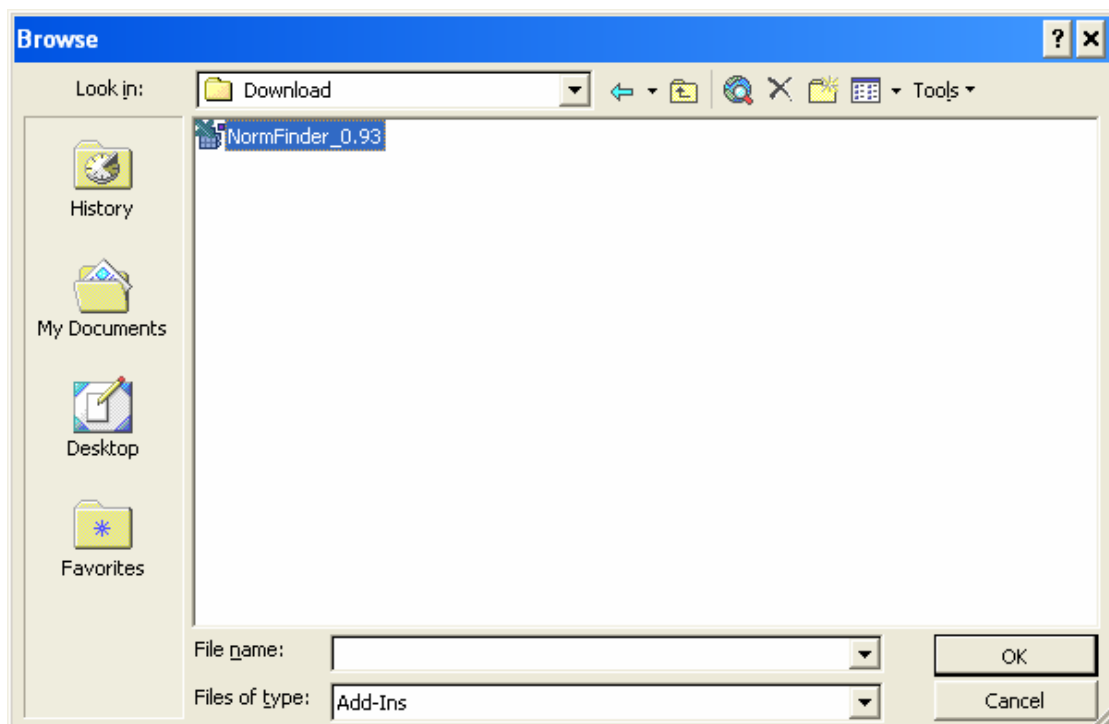
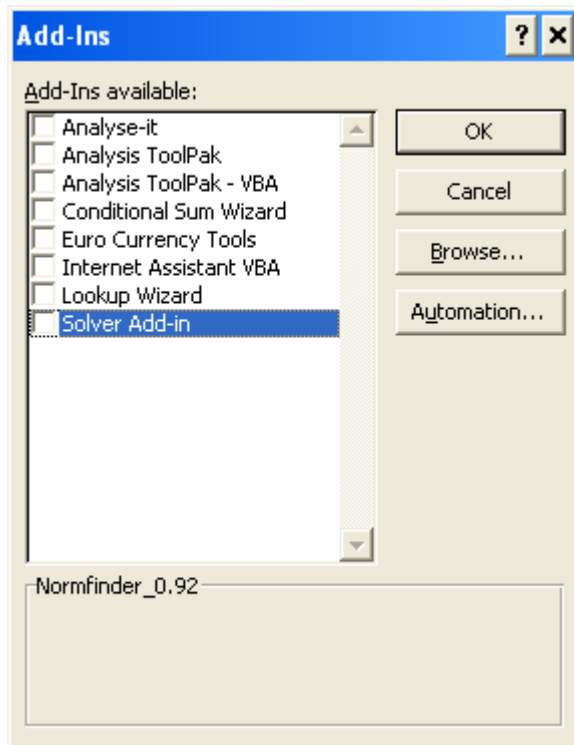
Excel add-in "NormFinder" authors:

Jens Ledet Jensen, Claus Lindbjerg Andersen and Christian Gundesen

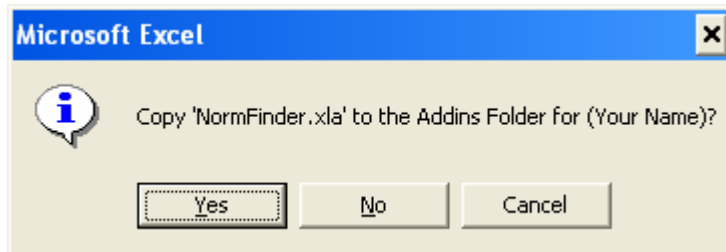
How to Install the “NormFinder.xla” Add-In

Open Excel and choose **Tools** from the menu-bar. Scroll down and select the submenu **Add-Ins...**

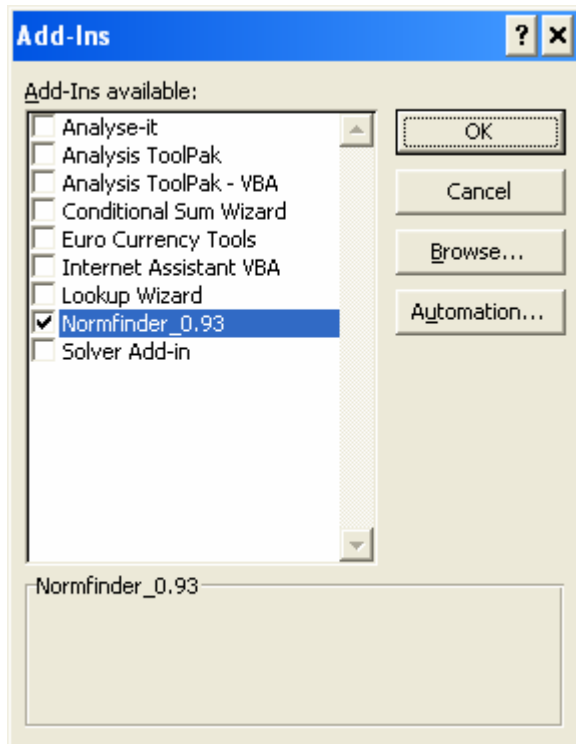
From the “Add-Ins” dialog select [Browse...] and locate the “NormFinder.xla” file.



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 ticked off and then select [OK] in the “Add-Ins” Dialog.



If everything is okay – you now have a “NormFinder” menu in your menu-bar.

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 how to use NormFinder

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.

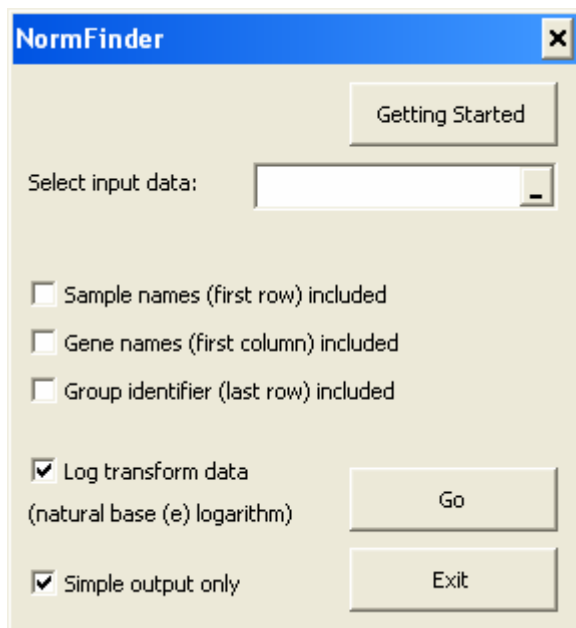
Then select “NormFinder” from the menu-bar.

Use the dialog box to select the input data.

Next tick 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. In case the data are already log transformed (natural base (e) logarithm) this is acceptable too, if the tick in field “log transformed data” is removed.

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



The simple output provides the user with

1. A stability value for each candidate gene and highlights the best gene with the lowest stability value.
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:

In the case group identifiers are included the stability value is the average of ρ_{ig} from Eq. C. For the case with no group identifiers 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 $\hat{\sigma}_{ig}^2$ from Eq. B.

Intergroup variation:

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

Happy NormFinding...
