MHH Hannover Medical School

RCU Transcriptomics

Research Core Unit Transcriptomics

Filter tool for single-color and dual-color data

Introduction

By use of our Excel-based filter tool, generated microarray data can be filtered according to specific selection criteria and individually defined thresholds. This allows for an intuitive data assessment with regard to the number of differentially regulated genes. Furthermore, the most promising candidates (those, passing the applied filter criteria) are grouped together, allowing users to generate a gene list for subsequent functional data analyses (e.g. gene ontology analyses, clustering approaches, …) or to generate simple heatmaps as preliminary figures.[*]

The order of microarray data provided within this filter tool resembles data structure as also provided within our standard Excel result files.[†]

The filter tool consists of three worksheets: 1. the 'Data' sheet, containing complete microarray data along with color coded fold change values, 2. the 'Filter' sheet, where individual settings for data filtering are defined and filtering is executed, and 3. the 'Legend' sheet, specifying the color code used to highlight fold change values of differentially regulated candidates.

---

[*] Note that a filtering process performed by use of our Excel tool cannot replace the need for an appropriate statistical test in order to identify the entirety of differentially regulated genes with high accuracy.

[†] We recommend to consult our manuals 'Excel standard result file for single-color studies' and 'Excel standard result file for dual-color studies' regarding the data structure and terms used in this manual.

The 'Data' worksheet[*]

On the worksheet 'Data' you can find the whole data content of your study, similarly arranged as in the Excel standard result file (Figure 1, just scroll to the left to inspect complete data). However, differing from the standard Excel result files, ratio values have been replaced by fold change values within our filter tools (e.g. a ratio value of 0.5 was replaced by a fold change value of -2.0; a ratio of 0.1 was exchanged by -10.0, and so on). Fold change values are color-coded systematically as shown in the 'Legend' worksheet (Figure 6).

Figure 1: Global view on single-color microarray data as contained in the 'Data' worksheet of the single-color filter tool (the dual-color filter tool is not shown).

The 'Filter' worksheet

When you open the filter tool for the first time, the worksheet 'Filter' is active and looks as shown in Figure 2. This worksheet contains 3 panels, the 'Filtering settings panel', the 'Filtering results panel' and the 'Filtering execution panel'.

Functionality as well as the overall layout is largely similar between single-color and dual-color filter tools. However, the two most-striking differences refer to the range of comparisons between samples that can be selected for filtering and one additional filter criterion (p-Value) that is available exclusively for dual-color data (Figure 3).

---

[*] Note that it is crucial not to alter any kind of data within the worksheet 'Data' (e.g. by altering or adding entries, or, by deleting rows or columns). Otherwise, functionality of the filter tool could be compromised.
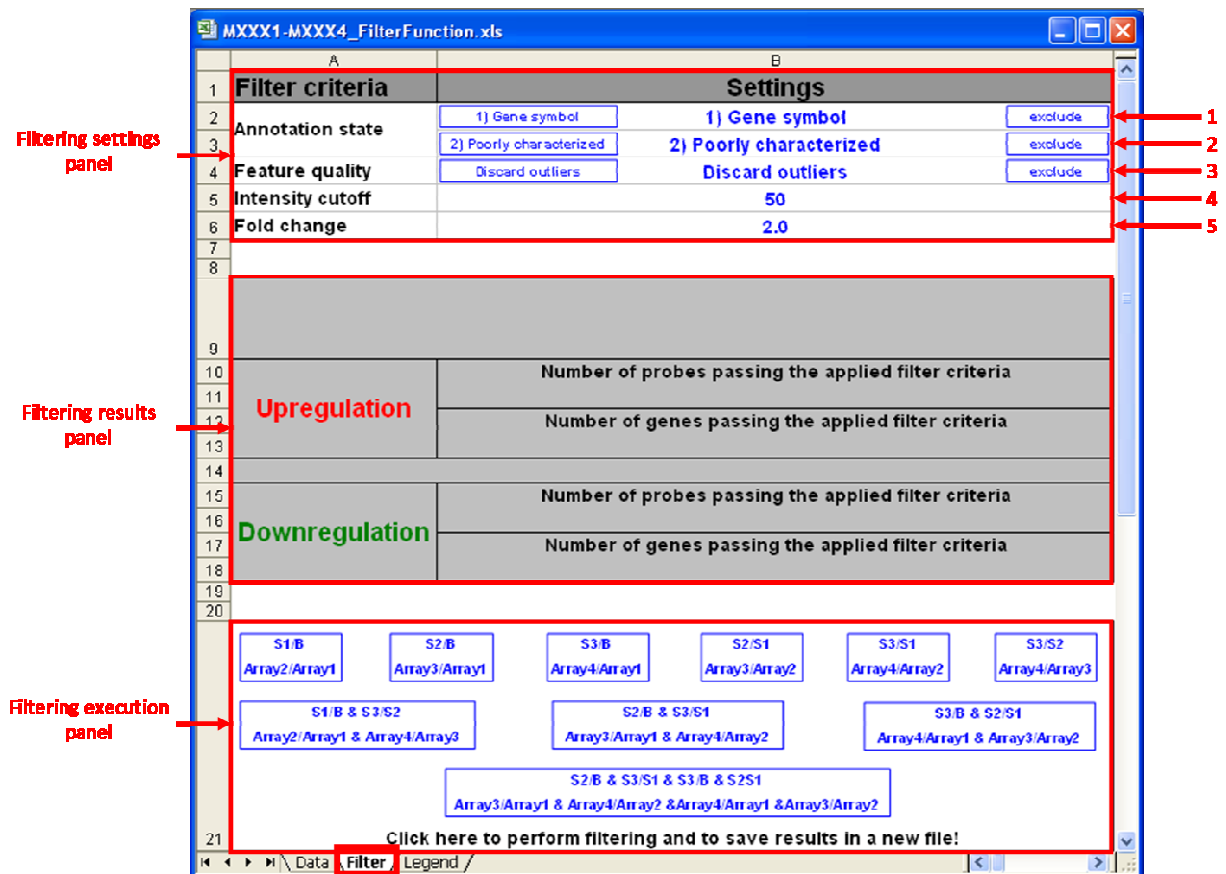
Figure 2: The composition of the 'Filter' worksheet in the filter tool for single-color data.
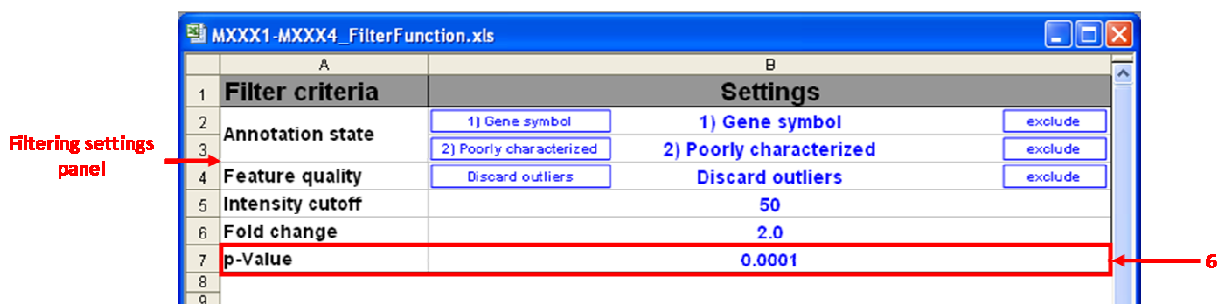

Figure 3: The composition of the 'Filter' worksheet in the filter tool for dual-color data. One additional filter criterion (p-Value) is exclusively available for dual-color data.

The 'Filtering settings panel'

1. You can include / exclude probes (transcripts, genes) that have a high functional characterization or annotation status (official gene symbol have been assigned).

2. You can include / exclude probes that are denoted as 'poorly characterized' in our standard Excel result file (according to our crude probe characterization workflow).

3. You can discard / consider probes that have impairments regarding hybridization performance (such measurements are discarded by default).

3

4. You can define the lower intensity threshold. Probes for which the arithmetic mean of intensities of the two samples to be compared falls below the specified cutoff, are excluded. (Default: 50)

5. You can specify the absolute fold change value to be regarded as 'true difference' in gene regulation. (Default: 2.0)

6. *(Dual-color only)* You can set the upper limit of the p-value. The lower the p-value, the higher the probability that the measured difference in intensity reflects a true difference in abundance of the two competing labeled RNA species for a given spot (transcript, gene). (Default: 0.0001)

The 'Filtering execution panel'

The data are always filtered based on pairwise comparisons ('ratio comparisons') or combinations thereof. The B, S1, S2, and S3 abbreviations (specific to single-color data) denote array 1, array 2, array 3 and array 4, respectively (Figure 2, Filtering execution panel). For further information regarding microarray order or nomenclature please consult our manual 'Excel standard result file for single-color studies'.

The filtering process is initiated by clicking on the respective blue button.

After having executed the filtering process, filtered data are suggested to be saved by default. At this end, a specific file name is suggested containing all specifically defined filter settings. This might be helpful to administrate files, originated from various filtering processes under different settings.[*]

The results of a filtering procedure are summarized within the 'Filtering results panel', the group of transcripts, passing the applied criteria are specifically ordered and systematically assigned on the 'Data' worksheet (Figure 6).
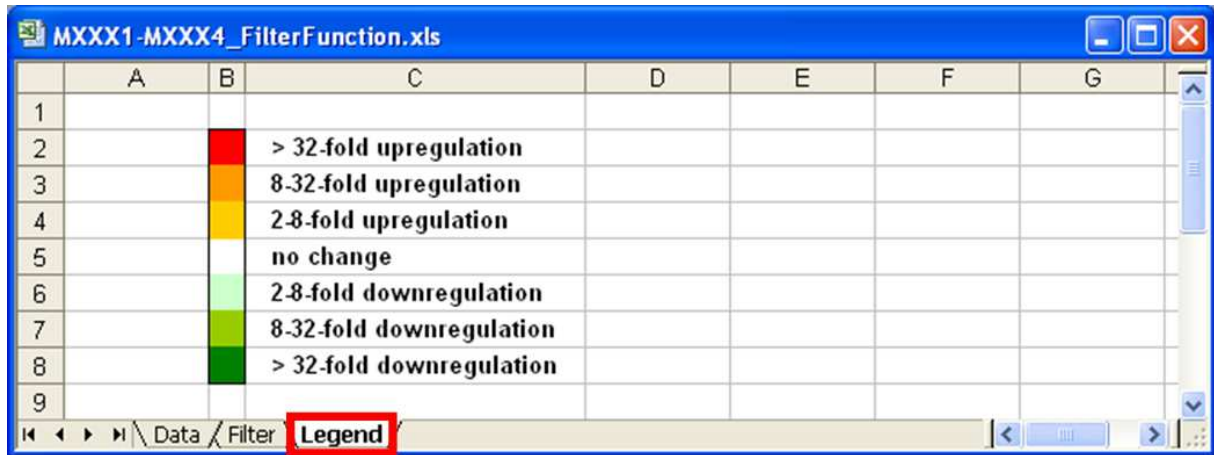
The 'Filtering results panel'

After a filtering process has been performed, numbers of probes and transcripts (genes) indicating differential mRNA expression are counted and depicted within the 'Filtering results panel'. For further information, please go ahead to the paragraph 'Inspecting filtering results'.

---

[*] Note that you can even use an already saved file to perform a new filtering applying different settings, again, the specified settings are integrated into the suggested file name at the end of the new filtering procedure.

The 'Legend' worksheet

Within this worksheet, the color code used to highlight fold change values provided in the worksheet 'Data' is specified.

Figure 4: The meaning of the color code is specified on the worksheet 'Legend'.

Inspecting filtering results

After a specific filtering has been performed, the 'Filter' worksheet looks as exemplified in Figure 5 for single-color data (in the dual-color filter tool results are arranged identically). Within the 'Filtering results panel' you can find information about

1. the respective pairwise comparison(s) 'ratio comparisons' initially selected
2. the number of 'upregulated probes' that passed the filter criteria
3. the number of upregulated transcripts (genes) that passed the filter criteria
4. the number of 'downregulated probes' that passed the filter criteria
5. the number of downregulated transcripts (genes) that passed the filter criteria.
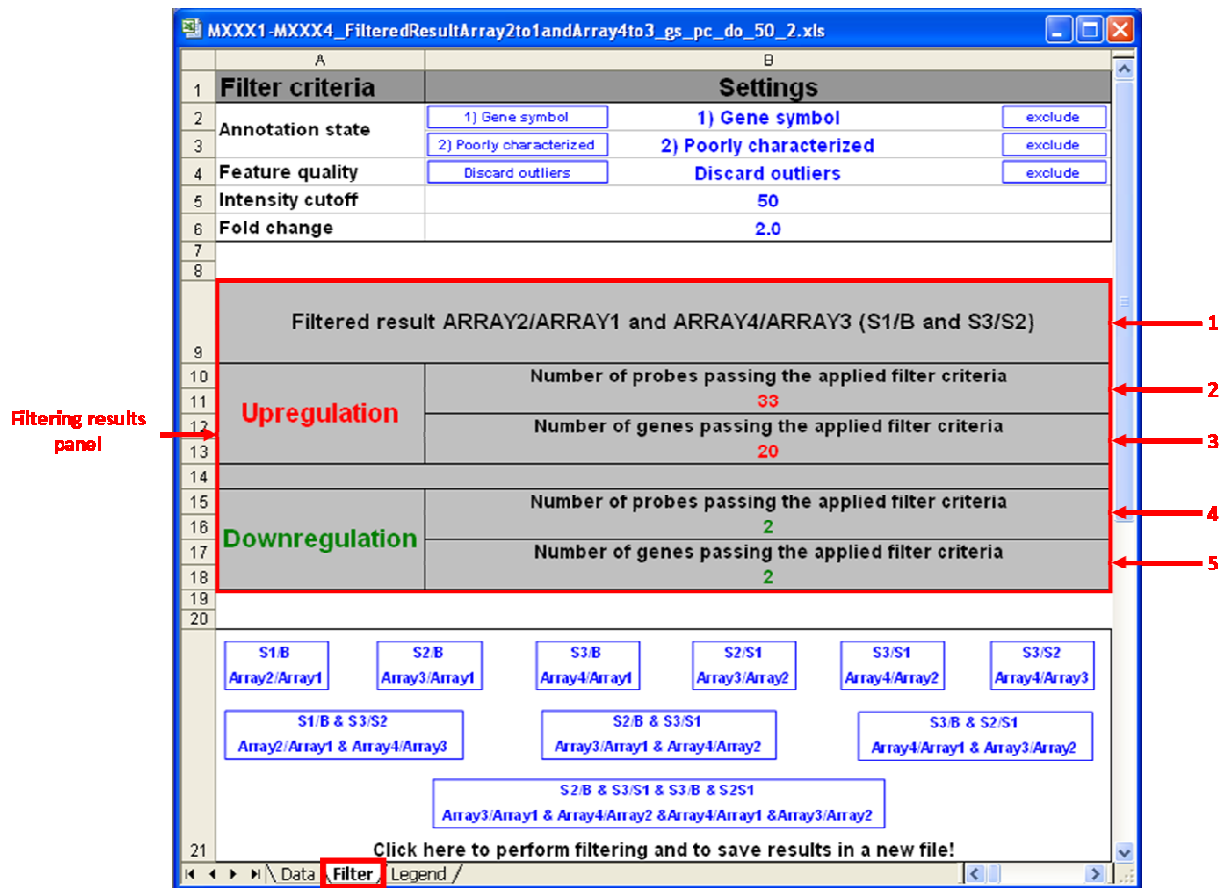
Figure 5: Results of a specific filtering process are depicted within the 'Filtering results panel'.

Additionally, microarray data (contained in the worksheet 'Data') are ordered alphabetically according to the GeneName column. Probe measurements that passed the applied filter criteria are marked by the entries 'up' or 'down' within the column: 'Filtering probes …' (Figure 6). If more than one probe directed against the same transcript fulfilled the applied filter criteria, one probe, systematically selected out of all 'positive' probes (the probe with the lowest FeatureNum entry), was used for counting the number of regulated transcripts/genes (rather than the number of probes). This probe measurement receives the additional assignment 'up' or 'down' within column: 'Filtering genes …. The resulting group of (non-redundant) probe measurements is arranged at the very top of the table and is designated as 'Non-redundant probe group' (Figure 6). This group represents the entirety of regulated transcripts/genes in a non-redundant way. All probes that additionally fulfill the applied criteria but were not selected for the 'Non-redundant probe group' constitute a second block of measurements, designated as 'Supplemental evidence probe group' (Figure 6).

Probe measurements indicating an upregulation are always followed by measurements showing a downregulation.
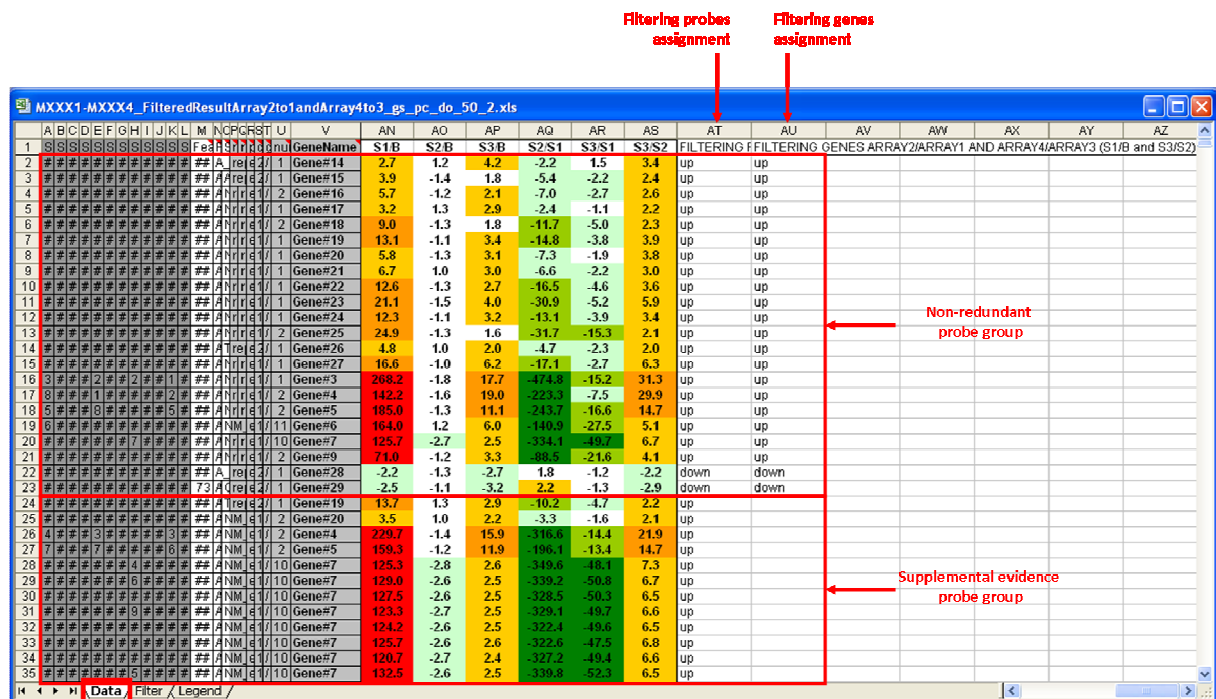

Figure 6: Single-color microarray data arrangement after a filtering process has been executed (dual color data not shown). Two blocks of probe measurements can be distinguished, the 'Non-redundant probe group' at the top and the 'Supplemental evidence probe group' below.

Kontakt:

Heike Schneider
Technische Assistentin
Gebäude I3, Ebene 1, Raum 1080
Tel: +49 (0)511-532 2830
E-Mail: Schneider.Heike@mh-hannover.de

Torsten Glomb
Diplom-Bioinformatiker
Gebäude I3, Ebene 1, Raum 1300
Tel: +49 (0)511-532 2869
E-Mail: Glomb.Torsten@mh-hannover.de

Dr. rer. nat Oliver Dittrich-Breiholz
Leiter der Zentralen Forschungseinrichtung Transcriptomics
Gebäude I3, Ebene 1, Raum 1270
Tel: +49 (0)511-532 5814
E-Mail: Dittrich.Oliver@mh-hannover.de