



Get Started Manual

March 2023

Table of contents

Introduction to GENEVESTIGATOR®.....	2
1.1 What is GENEVESTIGATOR®?.....	2
1.1.1 The concept of meta-profiles.....	2
1.1.2 Software components.....	4
1.1.3 Requirements.....	4
1.2 Types of analysis.....	5
1.3 User interface of the analysis tool.....	6
1.4 Analysis workflow.....	6
1.4.1 Get Started.....	7
1.4.2 Organism and data selection.....	8
Select specific experiments and studies.....	8
1. “Search by keyword, accession number, author/ submitter”.....	9
2. “Search by research area /study design”.....	10
Select data based on annotations.....	11
1. General filters.....	11
2. “Refine selection by conditions”.....	11
3. Filter data by experimental/clinical parameters.....	11
1.4.3 Selecting genes of interest.....	13
Limits on gene selection.....	15
Select genes from predefined gene sets.....	16
1.4.4 Tool selection.....	16
Examples.....	17
1.6 Viewing results.....	19
Saving results and exporting figures & data.....	20
2.1 Saving workspaces.....	20
2.2 Exporting figures.....	20
Working with figures in PDF.....	21
Working with figures in PNG.....	21
2.3 Exporting data from figures.....	21
References.....	22

Introduction to GENEVESTIGATOR®

1.1 What is GENEVESTIGATOR®?

GENEVESTIGATOR® is an innovative search engine to investigate in a single analysis gene transcriptional regulation across thousands of experimental conditions. It summarizes data by condition types such as tissues, cancers, diseases, genetic modifications, or external stimuli based on the concept of meta-profiles as described below. GENEVESTIGATOR® integrates manually curated and quality-controlled gene expression data from public repositories (Figure 1.1) but can also integrate proprietary data (Hruz *et al.*, [1]).

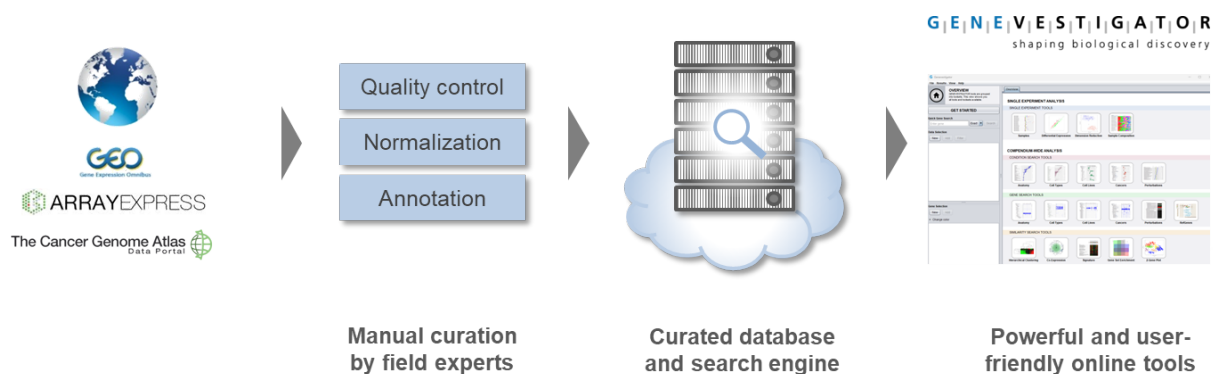


Figure 1.1: GENEVESTIGATOR® is a user-friendly search engine to explore thousands of manually curated and quality-controlled public and private gene expression experiments.

More advanced analyses are also possible. For instance, GENEVESTIGATOR® can search for genes specifically expressed under certain conditions, e.g., in certain tissues, in certain cancer types or in a specific disease. It can also search for genes sharing similar expression regulation with a target gene or group genes with similar expression by means of clustering.

GENEVESTIGATOR® contains data from three species: human, rat, mouse. More information on the database content is available at:

<https://genevestigator.com/gv/doc/content.jsp>

1.1.1 The concept of meta-profiles

Meta-profiles summarize expression levels from many samples according to their biological context. Thus, in a meta-profile, each signal value corresponds to the **average expression level of one gene over a set of samples sharing the same biological context**, e.g., samples from the same tissue type. In contrast, in “standard expression profiles” each signal value corresponds to the expression level of one gene in one sample. GENEVESTIGATOR® uses five types of meta-profiles which group samples according to the following aspects: anatomical parts, cell lines, cell types, cancers types, and perturbations types. The “*Perturbations*” meta-profile comprises responses to various experimental conditions (drugs, chemicals, hormones, etc.), diseases, and genotypes. All samples in the database

are annotated according to these biological dimensions. For each meta-profile, the summarization leads to a matrix of genes versus categories as shown below (Figure 1.2).

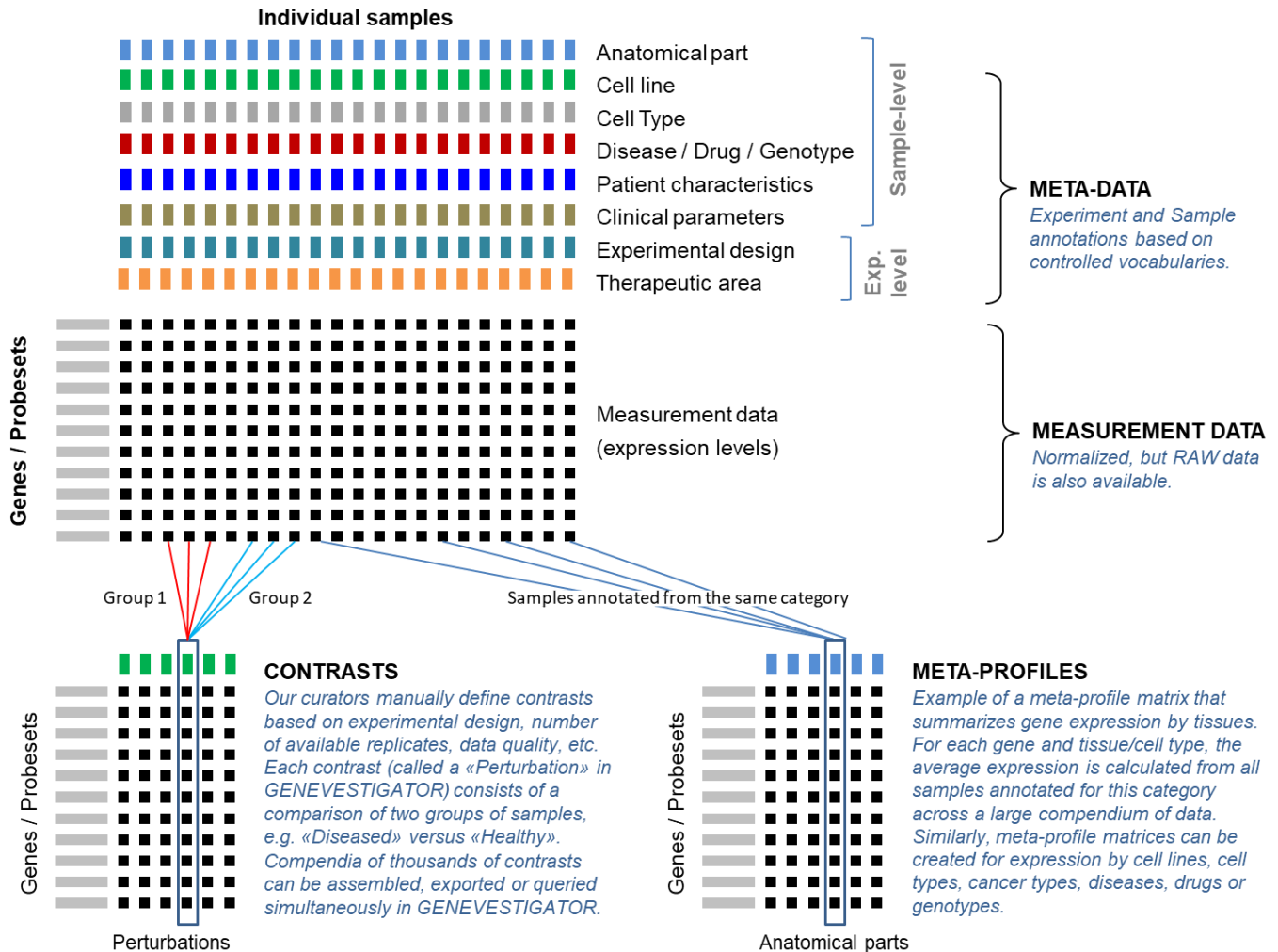


Figure 1.2: Schematic representation of the meta-profile concept. Expression values from many samples (top) are summarized into meta-profiles (bottom) according to their annotations.

Most categories are populated with several samples, thereby providing a robust estimate for the expression of genes within these categories. An analysis based on meta-profiles requires that sample signal values are comparable between the different experiments and, generally, between laboratories. Several lines of evidence support the validity of the summarization into meta-profiles as performed in GENEVESTIGATOR®. For example, Prasad *et al.* [2] demonstrated that transcriptome variations due to the tissue of origin are much larger than the variations due to perturbations or lab effects. Therefore, summarizing tissue-specific expression by grouping data from various experiments provides a relatively good estimate for the expression level of a gene in a given tissue. Likewise, good representative expression meta-profiles can be obtained for cell lines, cell types and cancers. For the «Perturbations» meta-profile, however, results are created by comparing groups of samples from individual experiments. Data from multiple experiments are not mixed to create a single value. As a result, this tool contains large compendia of response types collected from many experiments.

1.1.2 Software components

GENEVESTIGATOR® comprises three main components (Figure 1.3):

- **The website** to start the analysis tool or to register for a user account. It also contains additional information such as the Get Started manual, publications, video tutorials, training modules, etc.
- **The analysis tool** or the “client application”, with which the analyses are performed. It communicates with the server to get all necessary data for the analyses.
- **The GENEVESTIGATOR® search engine** containing all data content and the software necessary for answering the requests of the analysis tool. Access to the GENEVESTIGATOR® server and to its data is provided via the “client application” and additionally via programmatic access for *Enterprise* customers.

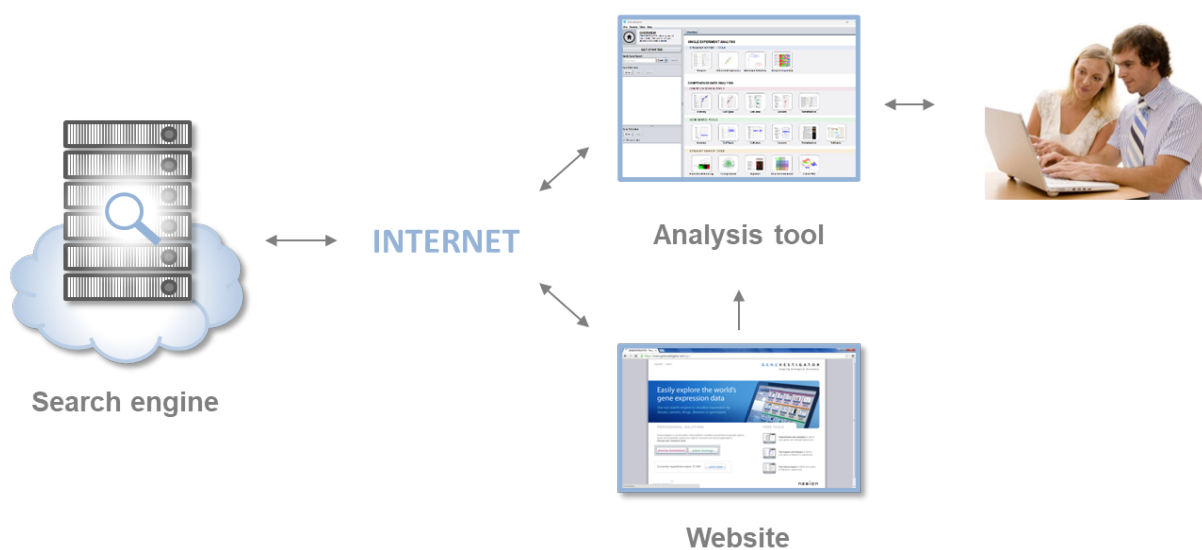


Figure 1.3: Overview of the three main GENEVESTIGATOR® components: website, analysis tool, and search engine (server cluster).

1.1.3 Requirements

The GENEVESTIGATOR® analysis tool is an independent application with an installable file available for Windows, Mac and Linux. On each start-up, it checks for a new version and auto-updates if needed. Working with GENEVESTIGATOR® requires internet access. The Minimal requirements to work with GENEVESTIGATOR®:

- Screen resolution: 1024 x 768
- Memory: 2 GB RAM
- Operating system: 64 bit

1.2 Types of analysis

- **SINGLE EXPERIMENT ANALYSIS:** to visualize the expression of genes across individual experiments (**Samples** tool) or to quickly and easily analyze a particular experiment (**Differential Expression** tool). The **Dimension Reduction** tool allows you to visualize and explore multi-dimensional data in a two-dimensional plot, and thereby to identify clusters of samples that are similar within an experiment. The **Sample Composition** tool allows you to analyze the composition and absolute abundance of distinct cell types in a single sample or groups of samples, and a correlation thereof with different sample traits such as diseases, treatments, etc.
- **COMPENDIUM-WIDE ANALYSIS:** to visualize the expression across thousands of experiments in a single analysis. It contains:
 - The **CONDITION SEARCH TOOLS** allow you to find out which conditions regulate the expression of your genes of interest (up to 400 genes). All the tools allow a rapid identification of the conditions affecting the expression of selected genes, e.g., to identify mutations leading to an up- or down-regulation of these genes. The **CONDITION SEARCH TOOLS** are instrumental in discovery research, validation of existing hypotheses or in generating new hypotheses that can be tested in the laboratory.
 - **GENE SEARCH TOOLS** to search the entire content for genes that are specifically expressed in a chosen set of conditions (a specific tissue type, cell line, cell type, cancer type, or perturbation). It also contains the **RefGenes** tool to find the best reference genes for normalizing RT-qPCR experiments.
 - **SIMILARITY SEARCH TOOLS** to find relationships between genes. This toolset groups various tools for **Hierarchical Clustering**, **Co-Expression**, **Signature** search, **Gene Set Enrichment** analyses and **2-Gene Plot**. It focuses on the identification of groups of genes having similar expression profiles within a dataset, or conditions having similar expression signatures.

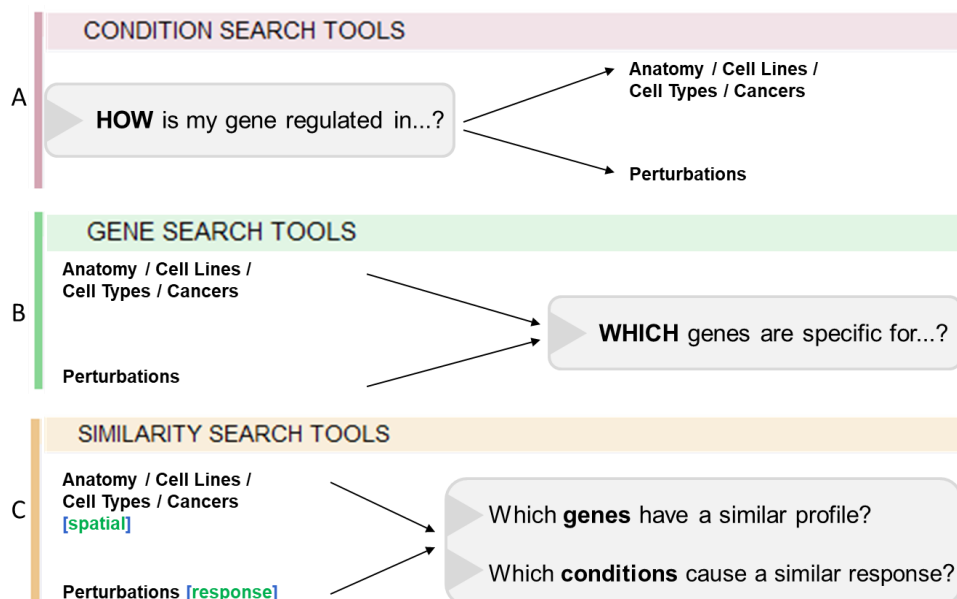


Figure 1.4: The three toolsets of the COMPENDIUM-WIDE ANALYSIS in GENEVESTIGATOR®. **A:** the **CONDITION SEARCH TOOLS** help you find conditions regulating your genes of interest. **B:** the **GENE SEARCH TOOLS** help you find genes having specific expression profiles, such as biomarkers. **C:** the **SIMILARITY SEARCH TOOLS** help you interpret your results and infer regulatory networks.

1.3 User interface of the analysis tool

The interface of the GENEVESTIGATOR® analysis tool has the following main components (Figure 1.5):

1. **“Home” button** and field indicating the toolset currently in use, e.g., **SINGLE EXPERIMENT TOOLS**
2. **“Data Selection” panel** holding folders containing the selected biological samples
3. **“Gene Selection” panel** holding all folders containing the genes of interest that you entered
4. **Tabs** to select analysis tools within a toolset
5. **Tool overview and results panel** where the result of an analysis is displayed

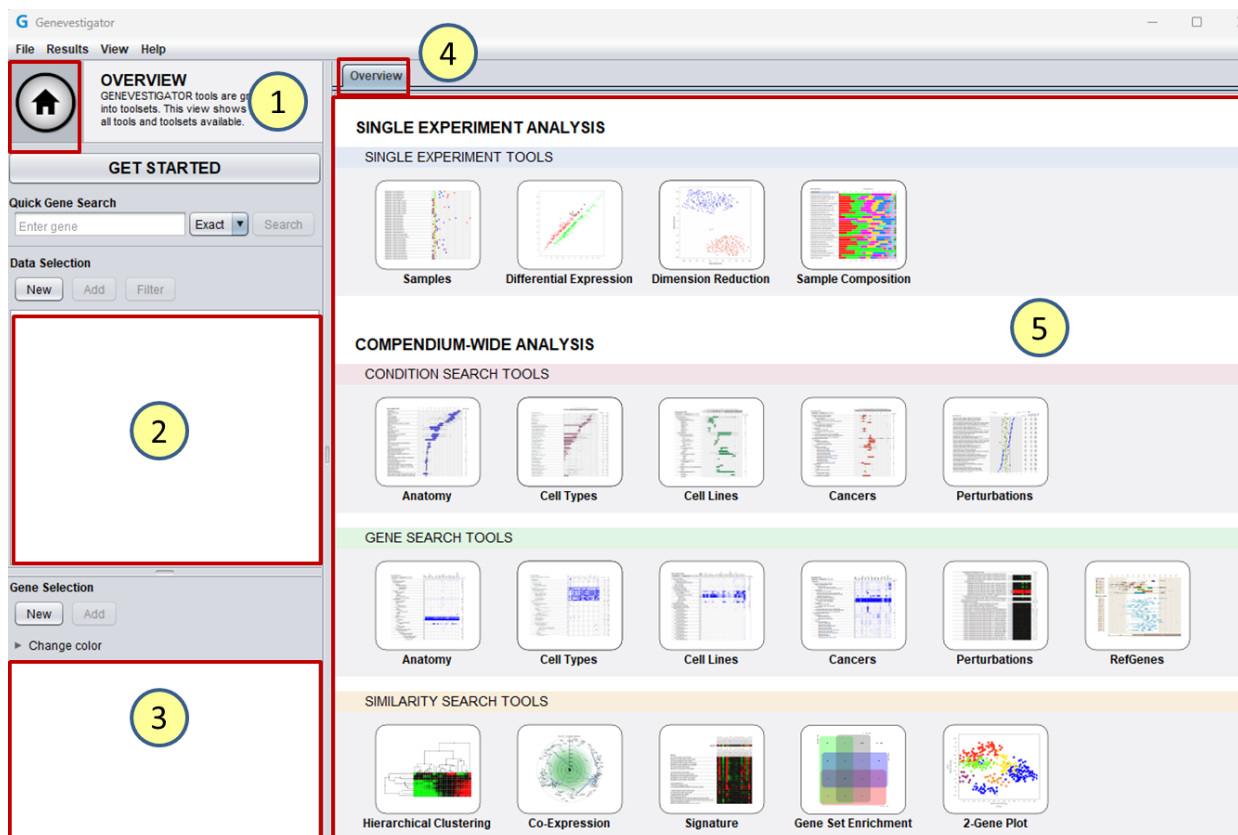


Figure 1.5: Main components of the user interface. 1) “Home” button, 2) “Data Selection” panel, 3) “Gene Selection” panel, 4) Tabs for tools, and 5) Tool overview and results panel.

1.4 Analysis workflow

GENEVESTIGATOR® allows exploring either the entire content or only a subset of it. Typically, an analysis runs on a selection of data and genes consisting of n samples and m genes, creating a $[m \times n]$ matrix. The results are then produced from the combination of data from this matrix and the corresponding sample meta-data. The choice of working with the entire content or only on a subset depends on the type of biological question to be answered (Figure 1.6). For example, to check how five genes respond to the complete compendium of drugs, two selections must be created:

1. All samples (n)
2. The five genes (m) of interest

The following sections provide more details about **each step**.

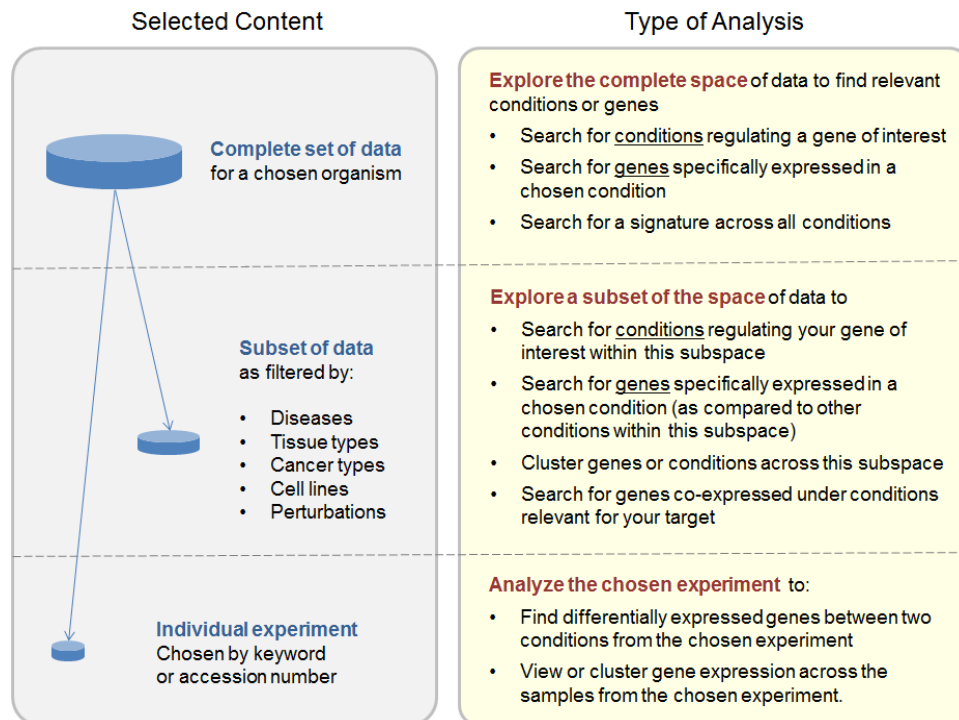


Figure 1.6: Different types of analysis can be done depending on the type and volume of the content selected.

1.4.1 Get Started

By clicking on the “GET STARTED” button a dialogue listing the main uses cases will open (Figure 1.7). On the right panel instructions on how to perform the selected analysis (highlighted in blue in the left panel) can easily be followed.

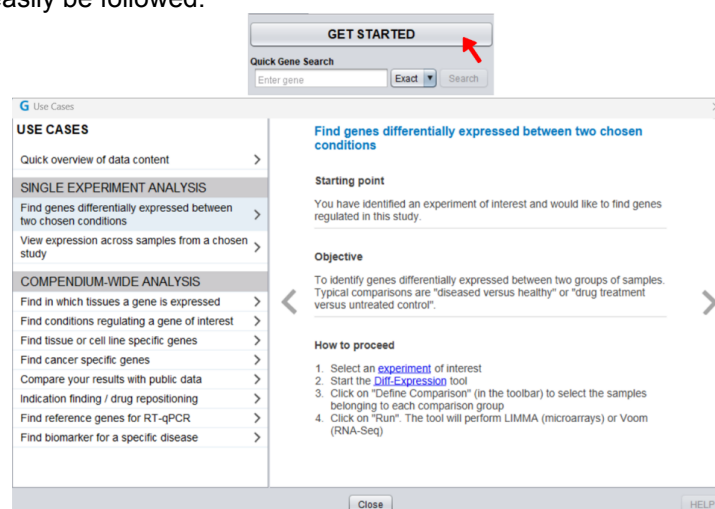


Figure 1.7: “GET STARTED” button and dialogue. The left panel displays a list of the main use cases. The right panel contains some information and instruction on how to perform the selected analysis.

1.4.2 Organism and data selection

To select a set of samples, click on “New” in the “Data Selection” panel (Figure 1.8, *upper image*). A dialogue box will open (Figure 1.8, *lower images*), allowing you to choose an organism and a platform. By default, the platform containing the most samples for the chosen organism is selected.

The “Data Selection” defines which samples will be used for an analysis. For some tools based on meta-profiles, such as the **CONDITION SEARCH TOOLS**, the more samples are selected, the more conditions will be displayed and the more robust the results will be (averages are then calculated from a higher number of single measurements). Therefore, to identify conditions regulating the expression of a gene, it is preferable to start with a maximum number of conditions. For such analyses, just choose the organism you want to work with and click on “OK” without further filtering.

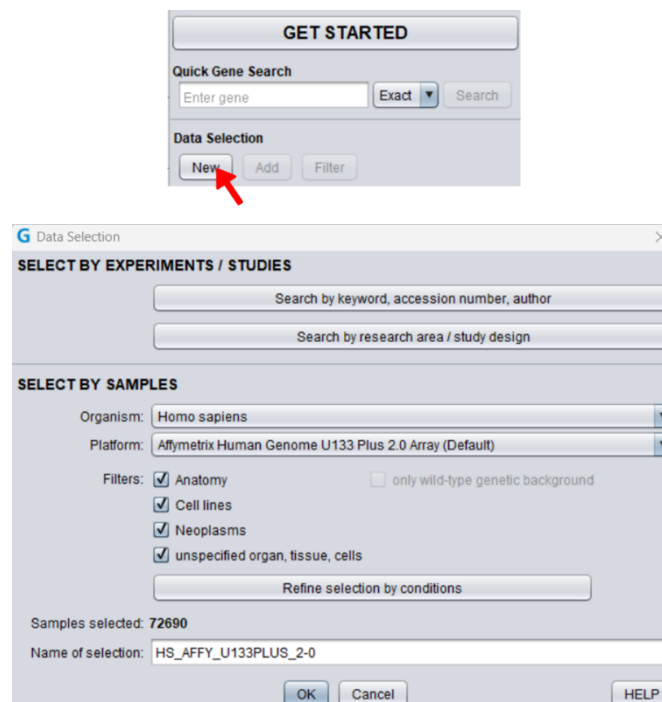


Figure 1.8: Selecting data for analysis. **Upper image:** the “New” button allows you to create a new data selection, while the “Add” button allows you to add samples to an existing selection. **Lower images:** “Data Selection” dialogue boxes where you can choose an organism, the technology platform. The number of samples available for the selection is indicated. Specific filters can be applied.

In other types of analyses, you may want to include only a portion of the database, e.g., a subset of experiments related to a specific disease, or to exclude some experiments or samples that may bias the analysis. To work with only a subset of samples, the general filters or/and the specific filters can be used (Figure 1.8, *lower images*) as described below.

Select specific experiments and studies

The Data Selection basic dialog (Figure 1.8, lower part) provides direct entry points to two types of searches and selections for entire experiments based on their annotated content and other metadata. Please note that both searches extend to all available studies by default, since there is no organism or platform preselection as in the quick selection case described here above.

1. “Search by keyword, accession number, author/ submitter”

This search option allows you to search experiments by keywords present in any of the sample or experiment descriptions, by accession number either from GENEVESTIGATOR® or from the public repository, or by author/ submitter name. While the default search will be performed across all organisms and platforms, it is still possible to limit the search to specific organisms or platforms using the search options panel (Figure 1.9).

The screenshot shows the 'Data Selection' window with the 'by keyword' tab selected. The search term 'IBD' is entered in the search bar. Below the search bar, there are options to 'Connect terms with' (AND, OR, Search as IDs) and checkboxes for 'Platform' (Homo sapiens, Mus musculus, Rattus norvegicus) and 'Condition' (Research Area, Anatomy, Cell types, Cell Lines, Cancer, Stimulus). A status bar indicates 'Found: 65 experiments from 3 species. This will create 18 selections'. Below this is a table of search results.

Nbr	Matching Fields	Primary L...	Secondar...	Title	Samples
<input type="checkbox"/> HS-00749	Stimulus,Parameter	GSE4183	E-GEOD-...	Inflammation, adenoma and cancer: objective classification of colon bio...	53
<input type="checkbox"/> HS-00897	Publication	GSE11499	E-GEOD-...	Expression data from Crohn's disease (CD) and Ulcerative colitis (UC) ...	11
<input type="checkbox"/> HS-00977	Stimulus	E-MTAB-1...		Whole Blood Gene Expression Analysis of Controls and CRC Patients	199
<input type="checkbox"/> HS-01019	Experiment title,Rese...	GSE33943	E-GEOD-...	Gene expression profiles of pediatric IBD remission patients	57
<input type="checkbox"/> HS-01021	Parameter	GSE1019...	PRJNA10...	Human colon expression in healthy controls, colon-only CD, ileo-colonic...	58
<input type="checkbox"/> HS-01040	Parameter	GSE1183...	E-GEOD-...	Transcription profiling of biopsies from the descending colon of Crohns ...	93
<input type="checkbox"/> HS-01052	Parameter	E-TABM-7...		Transcription profiling by array of primary human macrophages cultured...	66
<input type="checkbox"/> HS-01096	Experiment title	GSE25220	E-GEOD-...	Transcription levels in human colon biopsies in IBS and IBD patients be...	39
<input type="checkbox"/> HS-01256	Stimulus,Parameter	GSE38713	E-GEOD-...	Expression data from intestinal mucosa of patients with UC	41
<input type="checkbox"/> HS-01260	Stimulus	GSE24287	E-GEOD-...	Differential early pathogenic changes in human ileal gene expression in...	100
<input type="checkbox"/> HS-01505	Stimulus,Parameter	GSE66207	GSE6620...	mRNA and small RNA associated with Crohn's disease behavior [RNA...	30
<input type="checkbox"/> HS-01562	Research area,Stimul...	GSE5794...	E-GEOD-...	Core Ileal Transcriptome in Pediatric Crohn Disease	350

Below the table, there is a section for 'Selected: No experiments selected' and a text box for 'Name of selections: DATA'. At the bottom are 'OK', 'Cancel', and 'HELP' buttons.

Figure 1.9: Experiments/ Studies search by keyword, accession number or author. The search is made across all organisms and platforms and allows you to quickly find all the experiments containing a certain keyword. In this example, a search for the keyword IBD resulted in 65 experiments from 3 different species on 18 different platforms.

By default, the terms entered by the user must all appear at the same time for an experiment to be listed in the search results. The “Matching Fields” column shows which annotation fields (e.g., research area, anatomical part, experiment/ sample ID) contain one (or more) of the searched terms and the tooltip shows the matching textual context. This default “AND” operator can be changed to “OR” with the corresponding switch, so that any of the searched terms appearing in an experiment is a sufficient condition for a hit. Note that each term is treated separately, so they can appear in any order and be spread across different annotation fields. Alternatively, when two or more terms are grouped together by enclosing them in double quotes, they must appear next to each other because the quoted part is interpreted as a piece of sentence or multiple-word annotation (for instance, compare searching “CD4 effector memory T cell” with and without quotes).

If the experiments are to be found starting from a given list of identifiers, either internal (e.g., HS-00075, MM-01374...) or external (e.g., GSE12343, SRP148417, SRR1151398 ...) - or a mixture of both types - then the “Search as IDs” option will automatically choose the “OR” connector and interpret a space- separated list of such IDs as intended, additionally providing a feedback on which IDs were not found in the database. Another possibility of specializing the search is the “Condition” filter, which

restricts the search on the chosen annotation fields (e.g., Anatomy, Cell types, Cell Lines, ...). For instance, you might want to search for "T cell" under Cell lines rather than Cell types, or in both annotation fields.

The search by keyword supports a more general query syntax that is provided by the underlying indexing engine Apache Solr. As an example, such syntax allows you to:

- Find experiments where two keywords are mentioned within a number of words of each other by using *Proximity* Searches (with the tilde operator), e.g., "**PBMC kidney**"~3
- Exclude experiments where a certain keyword appears by preceding such keyword with the minus character, e.g., to exclude kidney studies with **PBMC -kidney**
- Logically group terms using parentheses to form sub-queries such as, for instance: **PBMC AND (monocytes OR kidney)**
- Use wildcard characters (* ?) within a single keyword to match prefixes or postfixes (e.g., **cardi*** or ***cytes**), or leave an unknown character in the query (e.g., select Condition = Cell types and search for **CD? T cell** to include both CD4 and CD8 variants)

A more complete description of the query syntax can be found on the Solr website:

https://solr.apache.org/guide/8_11/the-standard-query-parser.html

2. "Search by research area /study design"

This search option (Figure 1.8, lower left image) allows you to search experiments related to a research area and/or based on their study design. The search will be performed across all organisms and platforms by default, but it is possible to limit the search to selected organisms and platforms and/or based on the study type and design (Figure 1.10).

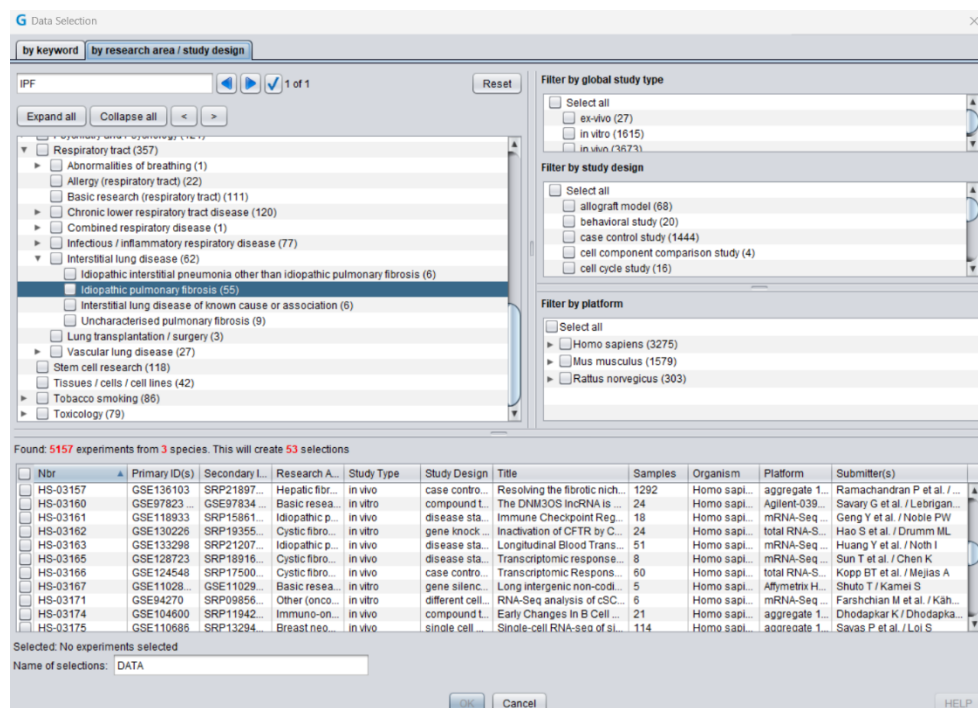


Figure 1.10: Search by research area/ study design. This search option allows you to search for all the experiments related to a specific research area or having a specific experimental design. In this example, the entire content was searched for experiments related to idiopathic pulmonary fibrosis (IPF) without further filtering. 5157 different experiments from 3 different species and on 53 different platforms were found (By resting on a specific experiment, additional information on this experiment will appear in a tooltip).

Select data based on annotations

1. General filters

These filters allow you to quickly select all samples (across multiple studies) having a specific annotation or property (e.g., all Cell Lines samples or only samples with a wild-type genetic background) (Figure 1.8, lower images).

2. “Refine selection by conditions”

The “Refine selection by conditions” button (Figure 1.8, *lower images*) allows you to search subsets of samples associated with specific biological contexts, e.g., all samples from a tissue type.

For each category (Anatomy, Development, Genotype, Perturbations, etc.) you can browse through the corresponding ontology (controlled vocabulary) and select the desired conditions, e.g., liver in the Anatomy ontology (Figure 1.11).

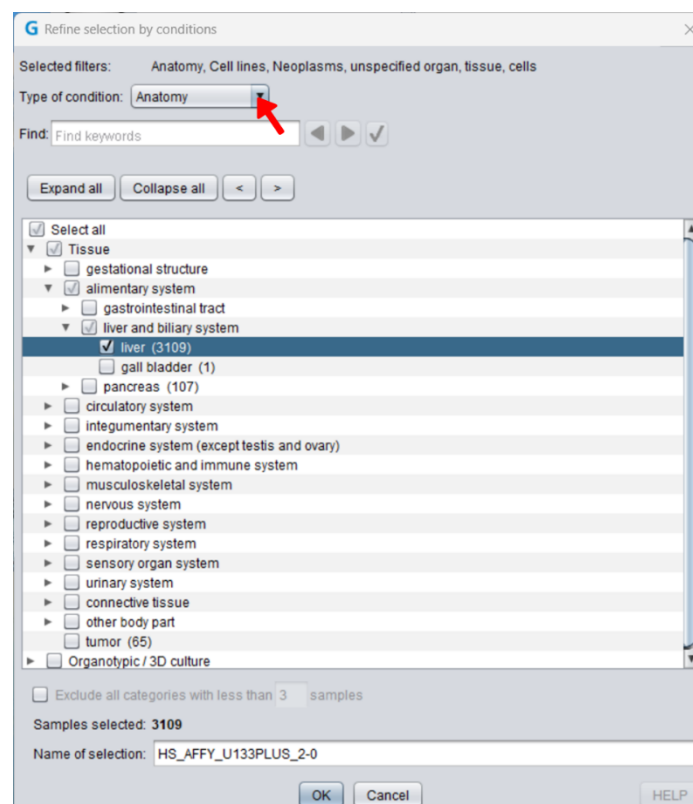


Figure 1.11: The “Refine selection by conditions” dialogue box. With this filter, samples can be selected based on their annotation, for instance all samples from liver can be quickly selected.

3. Filter data by experimental/clinical parameters

Once a data selection has been created, the samples from this selection can be filtered based on experimental/clinical parameters by clicking on “Filter” in the “Data Selection” panel (Figure 1.12). The samples (from the selection in focus) containing the deselected parameters will be removed from the analysis, e.g., if you deselect “diseased”, all samples from diseased patients will be deselected.

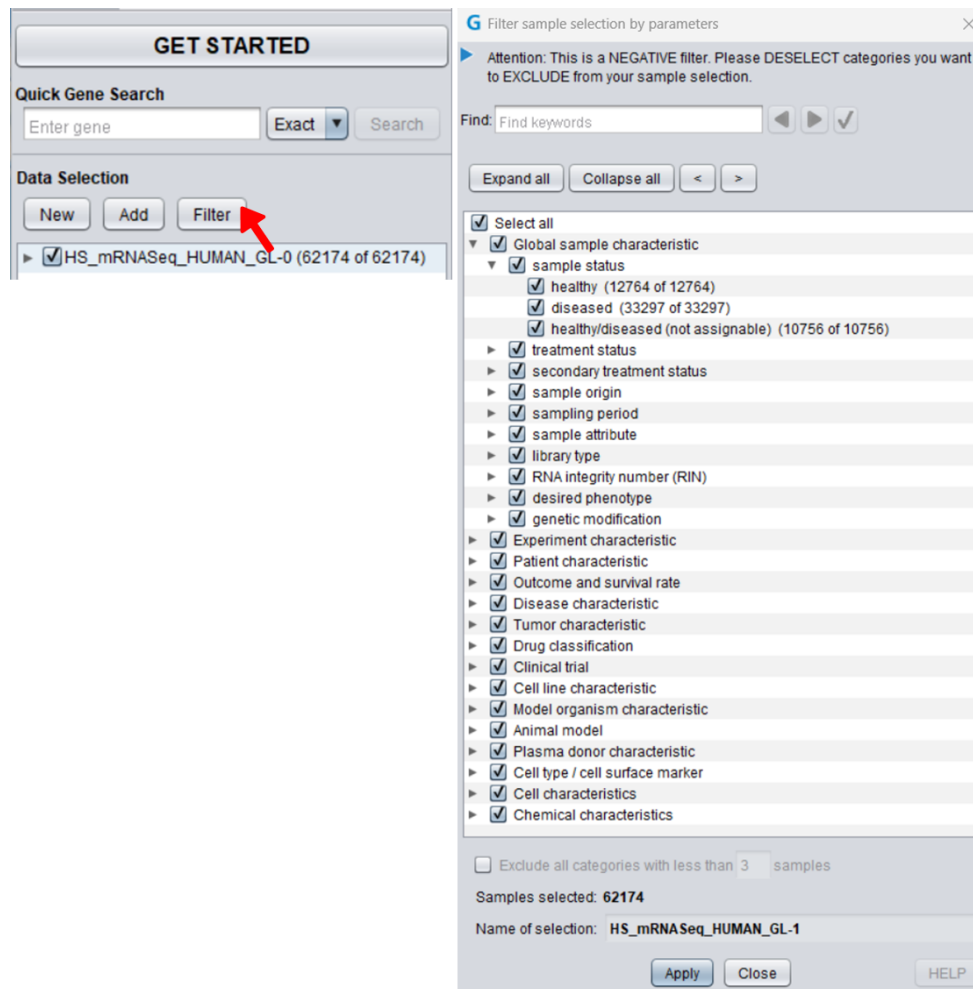


Figure 1.12: Experimental/ clinical parameters filter. This filter allows you to quickly select subsets of samples based on the experimental or clinical parameter across the whole “Data Selection”, e.g., only samples from healthy patients.

A note about the platforms:

For some organisms, GENEVESTIGATOR® contains expression data from multiple types of platforms, e.g., different generations of Affymetrix GeneChip® arrays. On these arrays, individual genes are frequently represented by a different set of probes. Expression measurements for a given gene may be targeting different transcript regions or splice variants and are not necessarily comparable between platforms. Therefore, to keep the analysis results easily interpretable, data from different array types are not mixed. Thus, a sample selection always contains only datasets from the same platform.

A note about the replicates:

Experiments are often repeated to increase the robustness of the results (biological replicates) or the same samples from a given experiment are measured multiple times (technical replicates). In GENEVESTIGATOR®, almost all replicates are biological replicates. Replicates have the same sample name up to the suffix, e.g. “_1-1”, “_1-2”, “_1-3” for replicates 1 to 3 of sample 1.

1.4.3 Selecting genes of interest

The *Samples* tool, the **CONDITION SEARCH TOOLS** and the **SIMILARITY SEARCH TOOLS** help you prioritize or interpret a list of genes by visualizing these genes against various biological contexts or grouping them according to their similarity of expression. Therefore, a list of genes must be entered prior to the analysis. In contrast, no list of genes is required for the **GENE SEARCH TOOLS** as these tools aim at identifying novel genes with particular properties.

The “Gene Selection” panel (Figure 1.5, *number 3*) holds all created gene selections from the current analysis session. To create a new selection of genes, click “New” in the “Gene Selection” panel. The below “Gene Selection” dialogue will open (Figure 1.13).

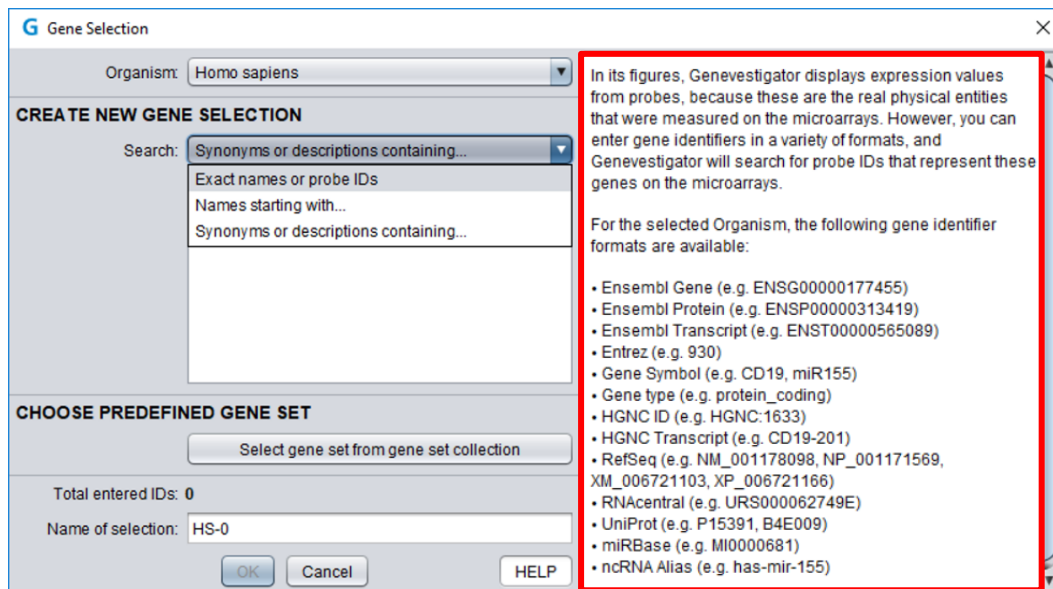


Figure 1.13: The “Gene Selection” dialogue. Genes can be entered in different formats, e.g., probe IDs, Entrez gene, UniGene, gene symbol, or identifiers from other gene models. All recognized models are listed on the right side of the window (marked in red) upon clicking “HELP”. Additional search options such as “Exact names or probe IDs” or “Names starting with...” are available.

With microarrays, expression levels are measured by probes targeting particular transcript regions. For this reason, the results displayed in GENEVESTIGATOR® correspond to the expression values of these probes (see below for the definition of a probe). The GENEVESTIGATOR® database contains a mapping of these probes to gene identifiers (gene model) such as Entrez, ENSEMBL, UniProt, or gene symbol, so you can enter gene identifiers in any of these supported formats. Lists of identifier formats available for the selected organism are available in the “Gene Selection” dialogue under “HELP”.

A note about the probe definition:

In some microarray technologies, the expression level of a gene is measured by a single probe. In others, multiple probes are measured and summarized into a single value per gene. Furthermore, different terms are used for groups of multiple probes, e.g., “probe set” for Affymetrix. As GENEVESTIGATOR® contains data from several types of microarray platforms, we are using the term “probe” to define a probe or a probe set representing a given gene. Whenever a piece of information is specific to Affymetrix, we use the Affymetrix “probe set” terminology. By extension, the term “probe” is used in GENEVESTIGATOR® in the sense of measurement unit for other technologies like RNA-Seq and proteomics.

Additional options to enter gene IDs are provided:

1. Exact names or probe IDs
2. Names starting with...: to search all genes of a family, e.g., COX
3. Synonyms or descriptions containing...: to create a list of genes with similar functions, e.g., "receptor kinase"

Enter your list of genes. By default, the best probe for the entered genes will be chosen automatically (see below for the definition of best probe). By automatically selecting the best probes for a gene, the selected genes are linked to an organism but not to a specific platform. Consequently, the same gene selection can be used with data selections from different platforms. Nevertheless, as some platforms do not measure all genes and as all gene models are not mapped on each platform, it can happen that some genes of the selection will not be found on all platforms (Figure 1.14). In this case, a warning will be displayed on top of the results panel.

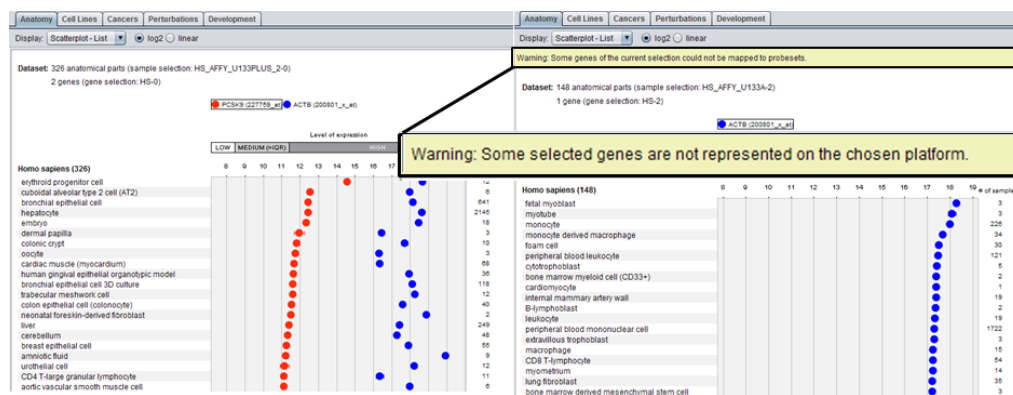


Figure 1.14: Gene Selection used for analyses on different platforms. As some genes are not measured on all platforms, it may happen that genes from a "Gene Selection" are not measured on certain platforms.

A note about the best probe definition:

Every platform supported in GENEVESTIGATOR® makes use of specific measurement probes to target genes, transcripts, or proteins in a sample. GENEVESTIGATOR® allows users to input any desired gene/protein/transcript identifier (ID) of interest for which we have an entry. This ID may or may not directly map to a measurement probe of a supported platform. To facilitate data exploration, GENEVESTIGATOR® will try to automatically determine the most specific measurement probe that can best represent the relative expression rate of the entered ID. The probe which is returned as the "best" probe for a given ID may only approximate the actual expression rate as each technology is limited to a subset of all measurable content in a sample.

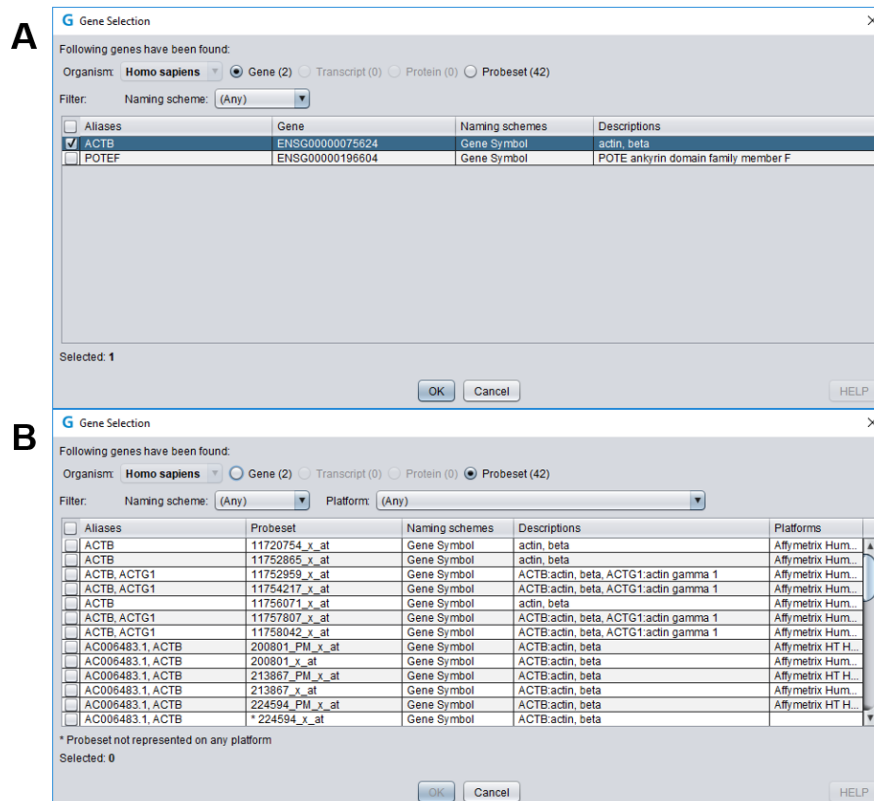


Figure 1.15: “Gene Selection” dialogue. A: Some genes are represented by multiple probes on the microarray platforms as indicated by the higher number of “Probeset” (29). **B:** By switching to the “Probeset” information level, it is possible to see which probes are available for a given gene on the different platforms. In this dialogue box, one or several probes per gene can be selected for the analyses.

For example, a user may enter a protein ID using the UniProt gene model and search for relevant probes on the Affymetrix Human Genome U133 Plus 2.0 Array. The Affymetrix array does not directly measure the quantity of protein in the sample but rather the relative mRNA content. A pre-calculated mapping of UniProt IDs to Affymetrix probe sets is used to determine the best measurement probe for representing the relative expression rate of the protein. There is no guarantee that this mapping is always accurate or direct. The mapping can only be as accurate as the underlying genome and protein information on which it is based. In this regard, GENEVESTIGATOR® attempts to provide the best effort mapping of a given ID to a measurement probe; the user is free to choose (a) different probe(s) by switching to the “Probeset” information level (Figure 1.15).

Limits on gene selection

Depending on your license type, the number of genes (or of probes if a gene is mapped by several probes) per “Gene Selection” is limited to 1 or 400. Please visit the GENEVESTIGATOR® website (www.genevestigator.com) to find out how many genes can be analyzed in parallel for each type of license.

Creation of a list of orthologs from an existing gene selection: It is possible to translate a gene selection from one species to another, using a mapping by OMA. To create a new selection with orthologs from another species, simply **right click** on the gene selection of interest, go on “Create Orthologs” and select the desired organism. A new gene selection containing the corresponding gene orthologs will appear as new folder in the “Gene Selection” panel.

Select genes from predefined gene sets

In addition to the above mentioned methods, genes can also be selected from predefined gene sets such as Reactome pathways or Gene Ontology categories. To do this, click “Select gene set from gene set collection” in the “Gene Selection” dialogue (Figure 1.13) and search for gene sets of interest, e.g. by keyword (Figure 1.16).

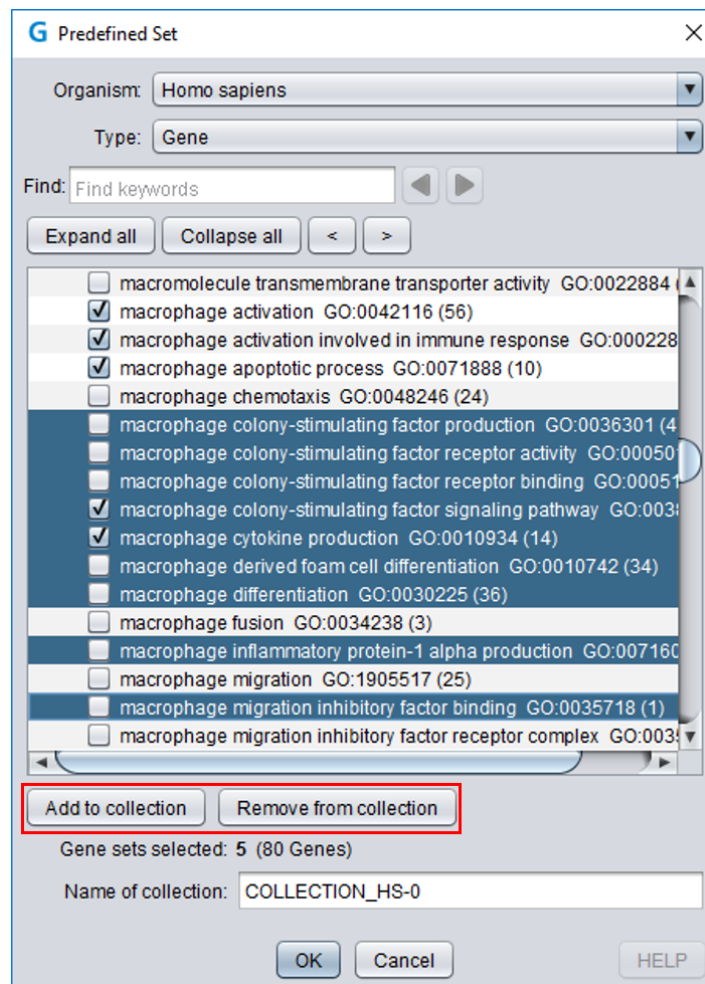


Figure 1.16: Selecting genes from predefined gene sets. In the “Gene Selection” dialogue (Figure 1.13), click “Select gene set from gene set collection” to select gene sets such as Reactome pathways or Gene Ontology categories. Selection can be done by checking individual boxes as well as by multiple selections using SHIFT- and CTRL-clicking in combination with the “Add to...” and “Remove from...” buttons (red rectangle).

1.4.4 Tool selection

Once a “Data Selection” and eventually a “Gene Selection” (mandatory for the **Samples** tools, the **CONDITION SEARCH TOOLS** and the **SIMILARITY SEARCH TOOLS**) have been defined, several types of analyses can be performed. As mentioned above, four main toolsets exist and each toolset comprises several individual tools (Figure 1.17). To start working with a tool, click on the corresponding icon. You can switch to another tool by going on the “Home” button (Figure 1.5, *number 1*) and selecting a new tool. Alternatively, you can change the tool within a toolset by selecting a different Tab (Figure 1.5, *number 4*). The different types of queries are described for each tool on the top right “HELP” button.

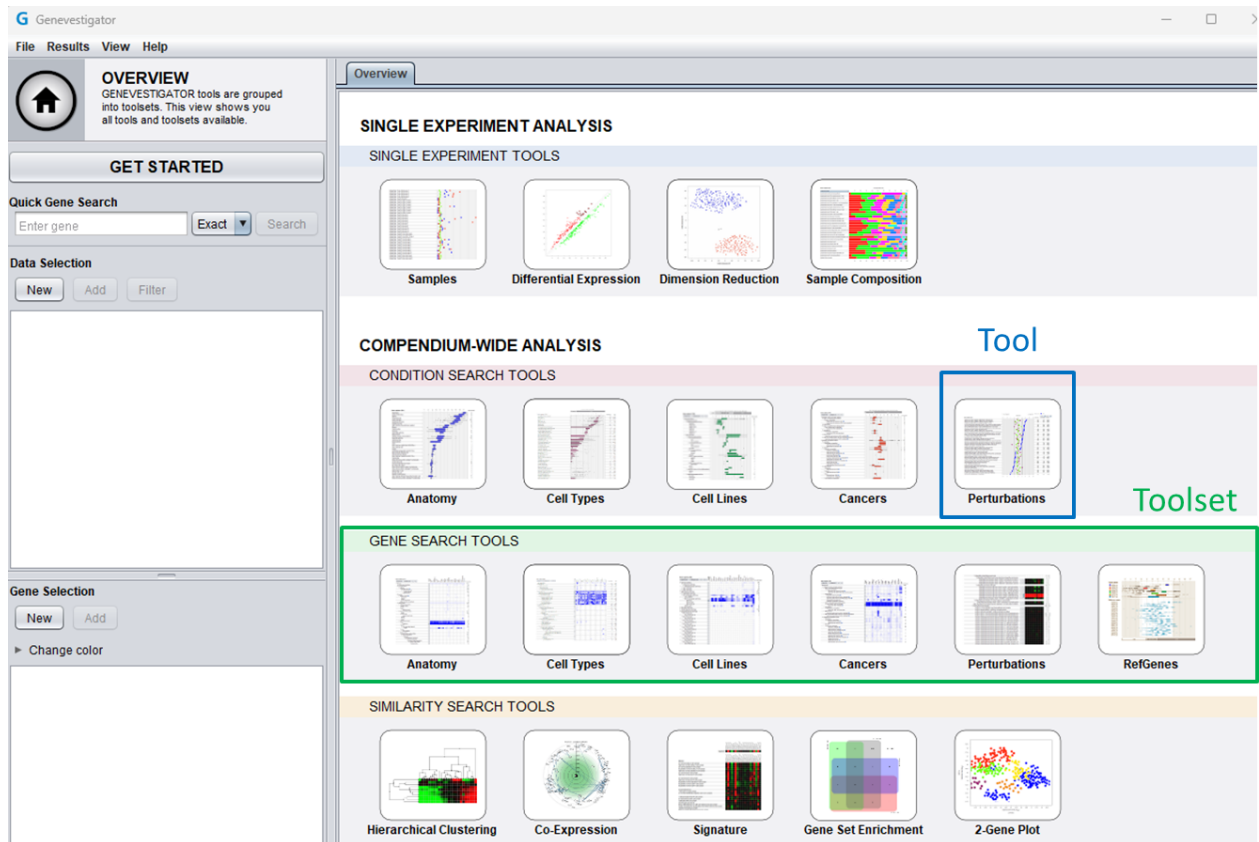


Figure 1.17: User interface of GENEVESTIGATOR®, with its four toolsets and their individual tools.

Examples

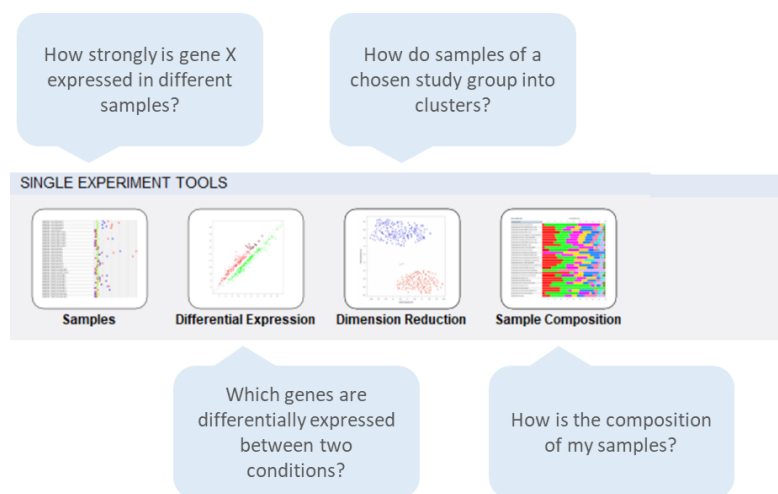


Figure 1.18: Examples of biological questions for specific tools in the **SINGLE EXPERIMENT TOOLSET**.

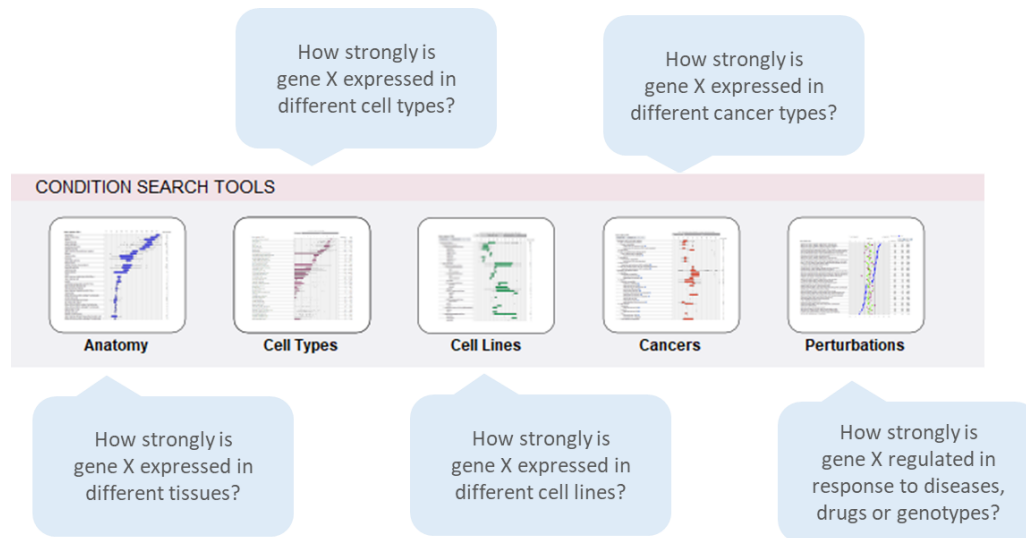


Figure 1.19: Examples of biological questions for specific tools in the **CONDITION SEARCH TOOLSET**.

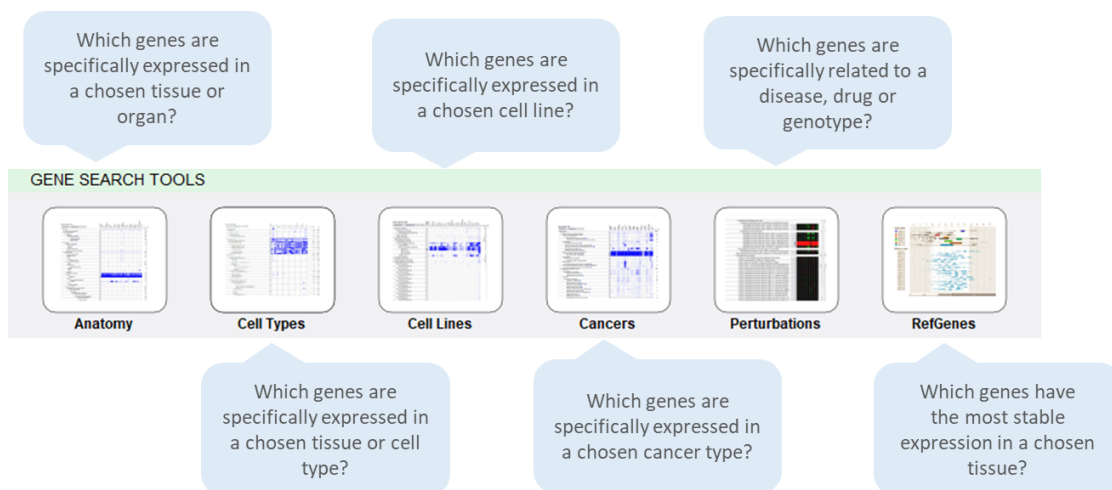


Figure 1.20: Examples of biological questions for specific tools in the **GENE SEARCH TOOLSET**.

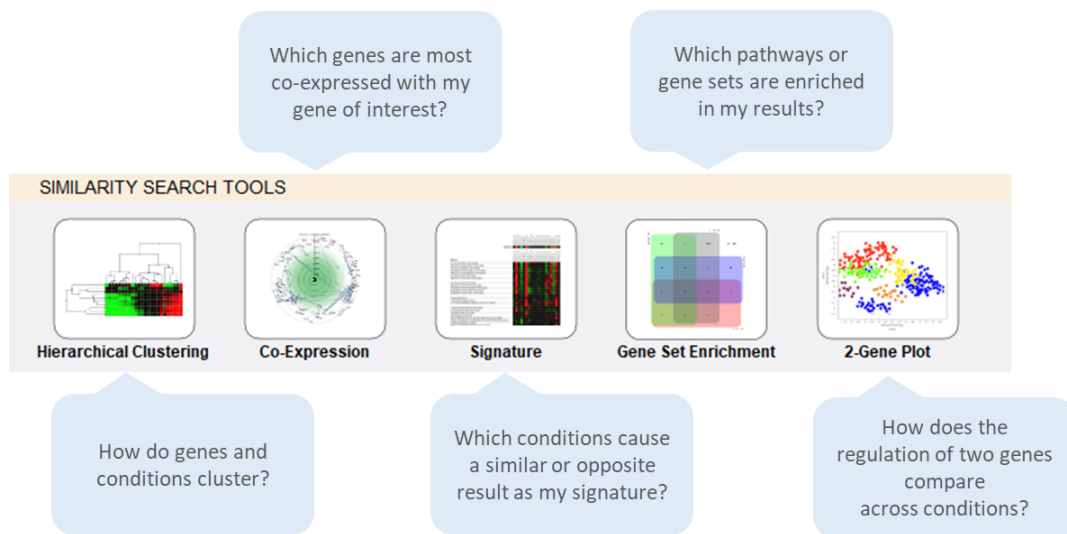


Figure 1.21: Examples of biological questions for specific tools in the **SIMILARITY SEARCH TOOLSET**.

1.6 Viewing results

For the **Samples** tool and the **CONDITION SEARCH TOOLS**, the results are immediately displayed based on the data and gene selections. The **Differential Expression** tool, the **GENE SEARCH TOOLS** and the **SIMILARITY SEARCH TOOLS** require additional choice and therefore have to be triggered using the "Run" button.

The results can be displayed in various formats, the choice of which mainly depends on the number of genes you would like to visualize.

Saving results and exporting figures & data

GENEVESTIGATOR® has several functions to save your analysis results and to allow you to continue working with it in another session or to export results for use in other applications. The current state of your analysis can be saved as a workspace and figures or data can be exported.

2.1 Saving workspaces

The “Save workspace” function in the “File” menu saves all sample selections and all gene selections (on the level of probes) of the current analysis to a local file on your computer. Workspaces are not stored on the GENEVESTIGATOR® servers. Via the “Load workspace” menu function such a file can be opened again and the analysis re-established. Workspaces are not associated with a user account, i.e., you can send such a file to a colleague to share your results or to collaborate on an analysis.

2.2 Exporting figures

In the “Results” menu under “Export Image”, GENEVESTIGATOR® provides extended possibilities to export the figures from your analysis results for further use (Figure 2.1). After previewing and scaling the figure, you have the choice between four file formats: PDF, PNG, GIF or JPG. The first is ideally suited for print publications, since it allows highest possible resolution and resizing without loss of information (vector graphics). The latter three are designed for publishing figures on the Web. Their resolution is lower, especially if figures are subsequently increased in size using image processing software (see Figure 2.2 for a comparison of the formats).

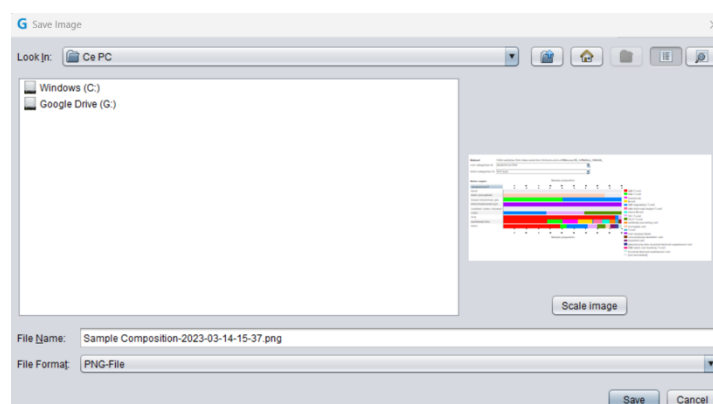


Figure 2.1: The figure export dialogue. Figures can be resized to the desired size and shape and exported either in PNG, GIF, JPG or PDF format.

Working with figures in PDF

PDF (Portable Document Format) is a graphics file format, making it easier to embed within another document. A PDF file can contain any combination of graphics, images and text, being therefore the most versatile format currently available. The main advantage of PDF for exporting figures from an application like GENEVESTIGATOR® is that it is vector-based and not pixel-based. For example, a line is stored as a stretch with start and end points, color, and thickness. In an image format such as JPEG, GIF or PNG, a line does not explicitly exist, it simply appears as a line because the pixels along this line possess the same color, but basically, they are “independent” (Figure 2.2).

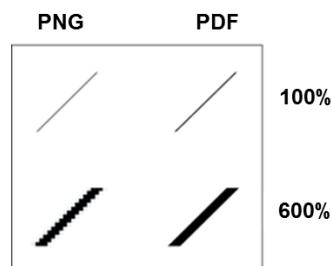


Figure 2.2: Comparison of the PNG and PDF formats for exporting figures. Since browsers cannot interpret PDF format directly, and PNG uses high compression, PNG is best suited for websites or in a presentation. PDF is the preferred choice for print publications.

Working with figures in PNG

PNG (Portable Networks Graphics) is a bitmap image format that uses lossless data compression. PNG performs better and now often replaces the older but more familiar GIF format. PNG supports palette-based, grayscale and RGB images, but not CMYK. Its lossless data compression allows you to open, edit, and re-save the image without loss of quality (in contrast, for example, to JPEG, which deteriorates the image after re-saving). Most browsers support the format and correctly render PNG image files.

2.3 Exporting data from figures

This feature is only available if you have a **Professional** or **Enterprise** license of GENEVESTIGATOR®. It is not available in trial and **Educational** licenses.

In every tool, you can export the expression data underlying a graph using the “Export data” function in the “Results” menu. The results can be exported to three different file types: Microsoft Excel (*.xlsx), comma-separated values (*.csv) or tab-delimited (*.txt). The latter two formats can be directly read in spread sheet applications like Microsoft Excel, Open Office Calc and in most statistical analysis software. This feature is ideal if you want to use separate software for creating the figures based on the same data. In the export dialogue, you can choose which gene aliases or descriptions should be added to the signal data in the export.

Note: Several spreadsheet software products, such as Microsoft Excel, allow a smaller number of columns (e.g., max. 256 columns in some older Excel versions) than lines (e.g., max. 65,000 lines in older Excel versions). Depending on the data size you want to export, please check how many genes and conditions are being exported and choose the layout accordingly so that you can open it in your preferred spreadsheet application.

References

- [1] Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Grissem W and Zimmermann P. Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. **Adv Bioinformatics** 2008, 420747 [\[Full Text\]](#)
- [2] Prasad A, Kumar SS, Dessimoz C, Bleuler S, Laule O, Hruz T, Grissem W, and Zimmermann P. Global regulatory architecture of human, mouse and rat tissue transcriptomes. **BMC Genomics** 2013, 14:716 [\[Full Text\]](#)