MetaMiner™
Cystic Fibrosis Report

Prepared by Yuri Nikolsky

OVERVIEW

MetaMiner™ (CF) is a data analysis platform for a broad range of CF researchers: wet lab biologists, bioinformaticians, clinicians and chemists. The platform was designed for the following purposes: 1. To assemble the most important biological and chemical experimental data available today in cystic fibrosis research in one easy-to-access reference environment, 2. To visualize the key mechanisms leading to and perturbed by the disease in a series of pathway maps and network models, and 3. To provide CF community with a "one stop shop" tool for uploading and analyzing experimental data in disease-centered interface.

MetaMiner™ (CF) represents a version of GeneGo's MetaDiscovery suite enriched with annotated content relevant for cystic fibrosis in a form of gene-disease associations, biologically active compounds relevant for the disease, and disease-specific pathway maps and networks. CF content is supported by a new interface centered on the disease. The platform is "multi-use" and enables different kinds of analyses, "dry lab" research and hypothesis generation workflows. It may be utilized for functional analysis of CF OMICs datasets (gene expression, proteomics, SNPs, metabolomics etc.), browsing of CF maps and networks, combinatorial search for cellular processes, genes, proteins and compounds involved in the disease and could be used as a tool to access relevant literature. The literature references have been curated for the disease relevance by GeneGo. Abstracts of articles are available through the easy "one click" access to PubMed from MetaMiner™ (CF) pages.

The advanced applications of MetaMiner™ (CF) include simulation analysis of CF network models in third party software suites and custom editing of pathways maps and networks. For instance, the users have an option of seamless export of cystic fibrosis networks from MetaMiner™ (CF)to the open source network analysis tool Cytoscape (www.cytoscape.org), followed by applying Cytoscape plug-in modeling packages. The users can apply the integrated MapEditor module for custom editing and drawing of pathway maps and networks. The platform is commercially available for use.

MetaMiner™ (CF) was developed in collaboration with the Cystic Fibrosis Foundation Therapeutics (CFFT) and GeneGo. GeneGo developers assembled and edited CF content and built the disease interface under the guidance of the Committee* [please use a footnote and give page number for committee list] of leading CF experts. This reiterative process was mediated and fully funded by CFFT.

MetaMiner™ (CF) integrates three disease-specific content domains:

1. Gene-disease "causative" associations
2. Bioactive compounds relevant for CF
3. Pathway maps and networks with descriptions

1. Gene-disease Associations

At the moment of the original delivery of MetaMiner™ (CF)**, GeneGo annotated (extracted from primary literature, put in machine readable forms and Oracle tables) information on CF association with 169 genes. These genes are connected with CF by 293 association links, which include:

- Promoter methylation   1
- Rearrangement   18
- Point mutation   8
- Single nucleotide polymorphism (SNP)   17
- RNA splice variant   1
- RNA amount change   130
- Altered interaction   2
- Protein isoform   3
- Protein amount change   113

In total, gene-CF associations were extracted from 220 original articles cited in 393 association's notes. This information can be visualized on networks and also recovered using search functions (Fig. 1).
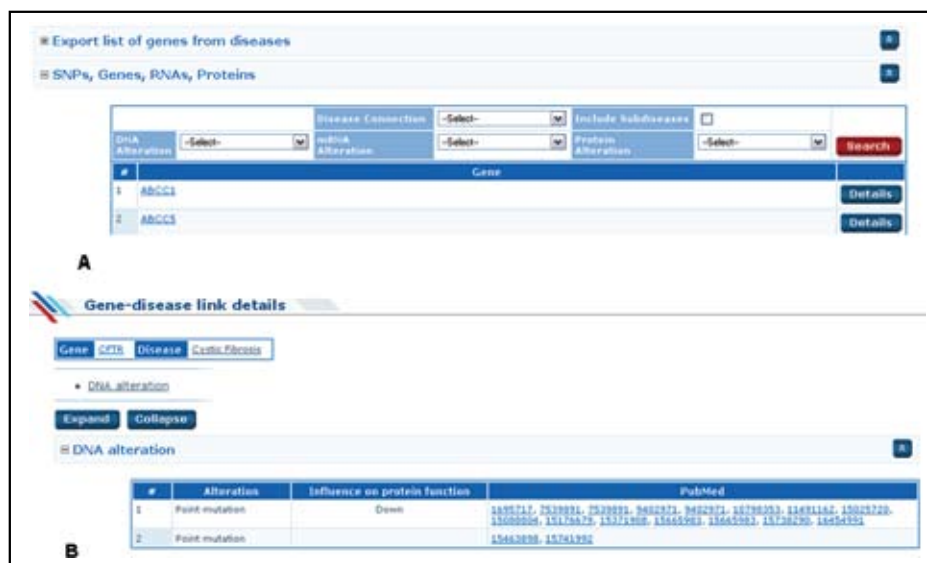

Figure 1. Gene-disease associations in cystic fibrosis. A. A list of genes and proteins relevant to CF at DNA, RNA or protein level. B. Details and references on CFTR mutations in CF.[1]

In MetaMiner™ (CF), the 169 disease-relevant genes feature 26,767 physical and functional interactions with other human proteins, DNA, RNA and compound species. These interactions are annotated from 23,415 original articles.

In addition, 8 small molecules compounds (endogenous metabolites) were associated with the disease, supported by 10 association notes from 8 articles.

7 biomarkers (3 metabolites and 4 proteins) were identified and annotated for CF.

2. Bioactive Compounds

In total, 89 drugs and compounds were linked to CF, including:

- FDA-approved drugs   14
- Drugs in clinical trials   65
- Discontinued drugs   8
- Pre-clinical drug candidates   2

Drugs are covered extensively in MetaMiner™ (CF). Each drug is annotated in a special compound form with information on its chemical, pharmaceutical properties, ADME-Tox, biological activity, including drug-drug and drug-target information. Some of the information on drug page is shown in Fig. 2

3. Pathway Maps and Networks

CF-related genes, proteins and compounds were represented in a form of "canonical" pathway maps and pre-built knowledge networks. In total, 25 maps and 10 networks were generated to visualize different aspects of CF origination and development, and perturbation of cellular processes, signaling, metabolic and transport pathways in the disease. The map and network list:

1. Regulation of CFTR activity (norm and CF)
2. Cytokine production by Th17 cells in cystic fibrosis
3. Bacterial infection in normal airways
4. Bacterial infection in CF airways
5. Mucin expression in CF via TLRs, EGFR signaling pathways
6. Mucin expression in CF via IL-6, IL-17 signaling pathways
7. wtCFTR and delta508-CFTR traffic / Generic schema (norm and CF)
8. Normal wtCFTR traffic / ER-to-Golgi
9. Delta508-CFTR traffic / ER-to-Golgi in CF
10. wtCFTR and deltaF508 traffic / Membrane expression (norm and CF)
11. wtCFTR and delta508 traffic / Clathrin coated vesicles formation (norm and CF)
12. Normal wtCFTR traffic / Sorting endosome formation
13. Delta508-CFTR traffic / Sorting endosome formation in CF
14. wtCFTR and deltaF508 traffic / Late endosome and Lysosome (norm and CF)
15. Regulation of degradation of wt-CFTR (norm)
16. Regulation of degradation of deltaF508 CFTR
17. Intracellular cholesterol transport (norm)
18. Intracellular cholesterol transport in CF
19. CFTR folding and maturation (norm and CF)
20. CFTR-dependent regulation of ion channels in Airway Epithelium (CF and norm)
21. Inhibitory action of Lipoxin A4 on PDGF, EGF and LTD4 signaling
22. Inhibitory action of Lipoxins on Superoxide production in neutrophils
23. Inhibitory actions of Lipoxins on neutrophil migration
24. Inhibitory action of Lipoxins and Resolvin E1 on neutrophil functions
25. Cytokine production by Th17 cells in CF (Mouse Model)
26. Bronchial inflammation in CF
27. Intracellular pattern recognition receptors
28. Regulation of ClC-2
29. TLR signaling in airway epithelium
30. Regulation of CFTR expression
31. Cystic fibrosis and hypoxia
32. Fibrosis signaling: common features
33. Pulmonary fibrosis
34. TGF beta 1 in cystic fibrosis
35. Cystic fibrosis-drugs and drugs targets

Pathway Maps represent interactive interconnected "wire diagrams" of cellular processes and pathways which are considered as "canonical", i.e. generally accepted by experts in the field. The series of CF specific maps is added as the "Cystic fibrosis" folder to the standard set of 650 proprietary mammalian signaling and metabolic maps assembled by GeneGo. The maps are generally used as an important functional ontology for gene list enrichment analysis (see below) and, as templates for mapping experimental data. Within the CF set, the maps for normal and disease state are often paired, in order to represent the specific interactions and pathways perturbed by disease. An example of a pair of such maps (CFTR traffic) is shown on Fig. 3. The maps are interactive. Any object on the map is "clickable" with the link leading to a reference page with protein/gene/compound information or a note for the interaction. Most maps are interconnected (blue and purple rectangles). Experimental data (for instance, microarray expression) and gene lists can be visualized on the map in a form of histograms (Fig. 3).
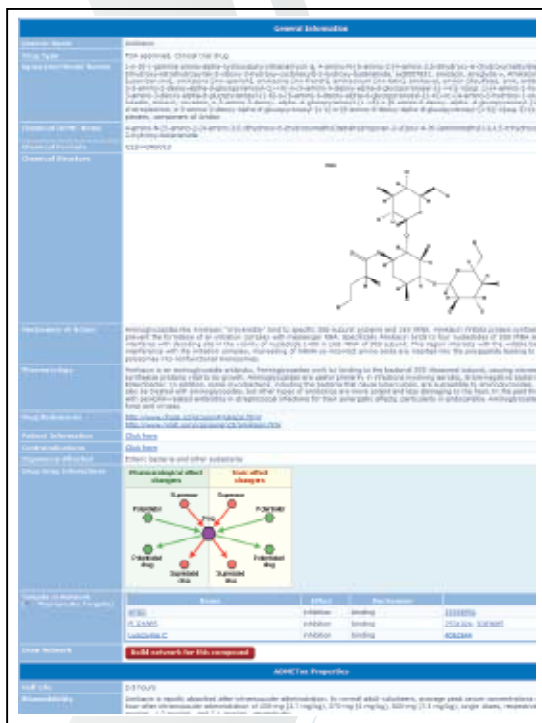

CF ANNOTATED CONTENT

Figure 2. Compound table for CF-relevant drugs[2]

Interaction annotation   mapped data

comment for a normal process

link to another "normal" map

link to another CF-perturbed map

A

B

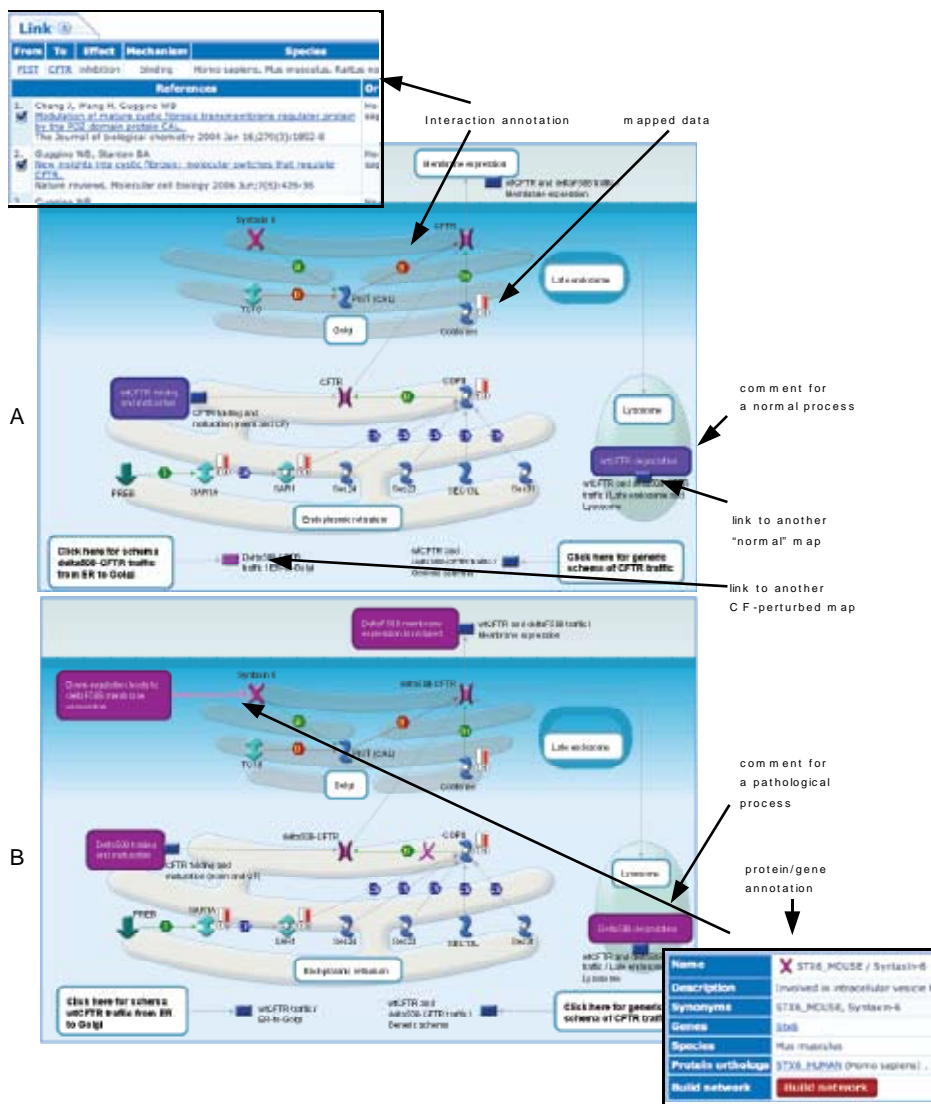comment for a pathological process

protein/gene annotation

Figure 3. Pathway map "CFTR traffic: ER to Golgi". A. Traffic of normal CFTR; B. Traffic of Delta 508 mutant CFTR, Map "Normal wtCFTR traffic/ER-to-Golgi"[3]

White Paper - MetaMiner Cystic Fibrosis (CF)

The pathway maps are manually drawn by GeneGo annotators using the MapEditor™ tool. Maps are easily editable and customizable (see below).

CF network models are a different type of functional visualization. They are built manually by MetaMiner™ (CF) network tools, using genes, proteins and compounds as seed objects (represented as nodes) and protein interactions as edges. The networks are a more flexible tool than maps, and usually they are used for visualizing compilations of pathways or more specialized processes (for instance, drug action). Typically, networks are more complex than maps (featuring more objects and connectors). An example of a pre-built network is shown on Fig. 4. Within MetaMiner™ (CF), networks are used for data mapping and enrichment analysis. Importantly, experimental data can be uploaded for network analysis by entering individual genes of interest or uploading data in batches such as an excel spreadsheet or zip files with multiple experiments. This utility may be beneficial for the researchers who have large lists of genes for analysis (with associated fold expression) such as microarray or proteomic data. Networks can be also transferred for simulation analysis and visualization into third party software packages such as Cytoscape and its plug-ins.
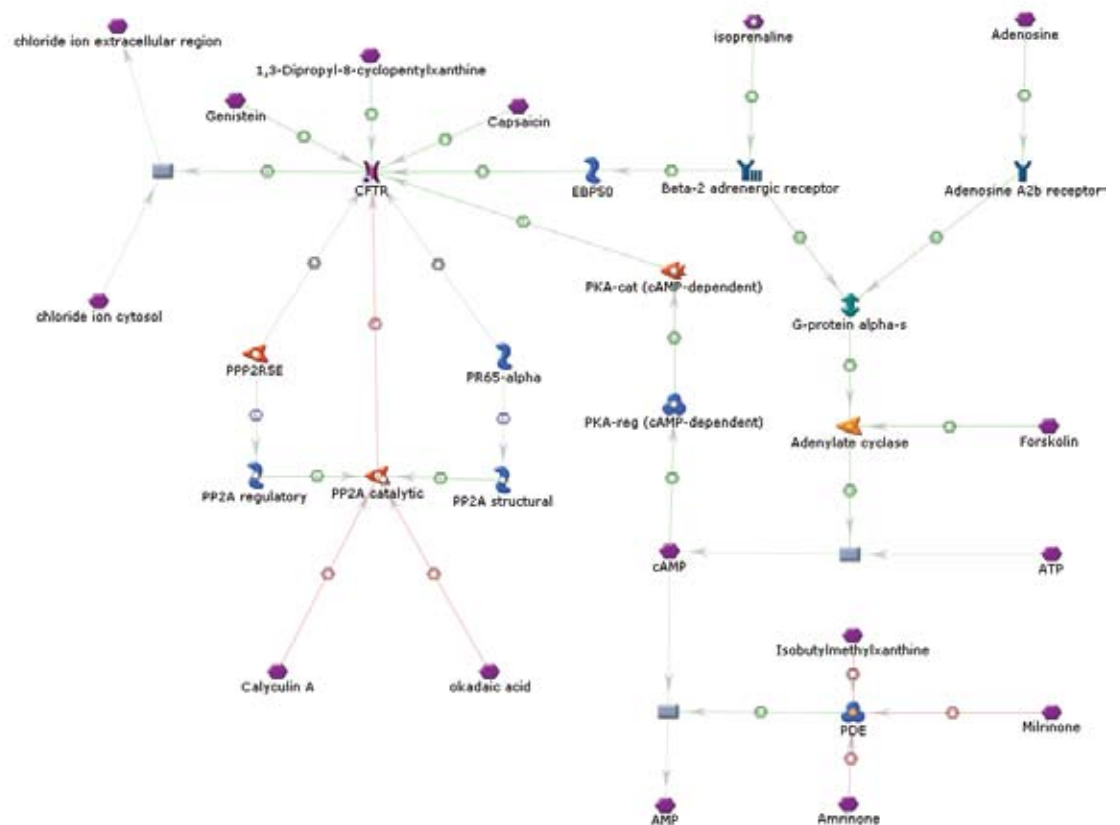

Figure 4. The pre-built network "cystic fibrosis drugs and drug targets". The nodes on the network represent proteins (colored symbols), compounds (purple hexagons) and metabolic reactions (grey boxes). The edges correspond to one-step protein-protein and protein-compound interactions. Green color is for activation interactions; red color is for inhibition.[4]

MetaMiner™ (CF) features complete functionality of the MetaCore™ platform including mouse and rat specific database, MetaRodent™ and MapEditor™. A comprehensive tutorial and learning videos for these tools are available from GeneGo. In this report, we will focus on tools for search, browsing and analysis of CF specific content and mapping experimental data.

Cystic Fibrosis Disease Page

CF disease page is the main hub for CF-related information in MetaMiner™ (CF). The disease page is accessible from many pages within the platform, depending on the analysis workflow (for instance, from maps, networks objects, proteins, drugs etc.). One standard path to open CF disease pages is through "search" in the main menu (Fig. 5).
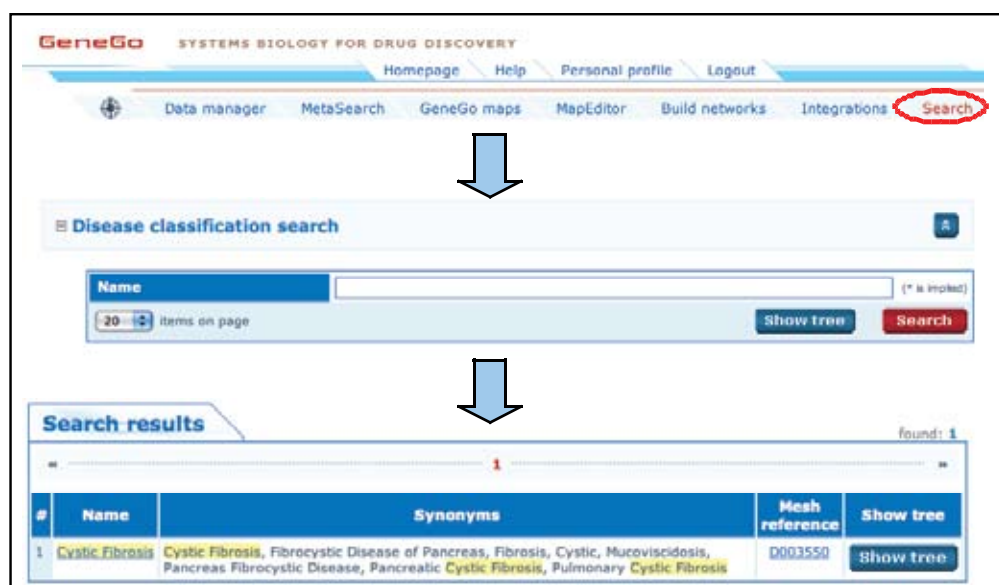
Figure 5. Opening CF disease page from the main menu of MetaMiner (CF). Click the Search link. Then search for 'Cystic Fibrosis' in the "Disease classification search" section.

The disease page is shown at Fig. 6, B. It contains a brief description, links to relevant disease classifications and databases such as WTO's ICD-9, 10, Mesh, OMIM and Disease DB. The link to the Disease tree opens GeneGo disease classification with CF "leafs" and corresponding branches (Fig. 6,A). The links below lead to internal MetaMiner™ (CF) pages with CF-associated gene lists, gene-disease associations, biomarkers, drugs, CF maps and networks and, finally to all standard GeneGo maps which contain CF-related genes.

Browse CF Maps from Main Menu

CF maps assembled into a folder "Cystic fibrosis" can be also viewed by clicking on "GeneGo maps" tab in MetaMiner™ (CF) main page (Fig. 7).

White Paper - MetaMiner Cystic Fibrosis (CF)

A

B

Figure 6.
A. GeneGo disease classification tree with CF.
B. Cystic fibrosis disease page.[5]

A

B

Figure 7. Browsing CF maps in MetaMiner™ (CF).
A. Maps category is selected from the main menu;
B. Position and content of CF maps in the Disease folder
CF networks can be opened from CF disease page.

Combinatorial Search for CF-Related Information

MetaMiner™ (CF) features a powerful Boolean search engine, which can be launched from the main menu called MetaSearch (MS). MS can be launched from the main menu and enables specific queries within the whole database of interactions and functional ontology's, including CF-specific information. MS is supported by an intuitive GUI for query design, export of the results, creating lists of found genes, proteins and compounds etc. For instance, a query "Find all drugs for ion channel receptors involved in cystic fibrosis" can be designed as follows (Fig. 8,B) and finds 2 compounds.

White Paper - MetaMiner Cystic Fibrosis (CF)

A
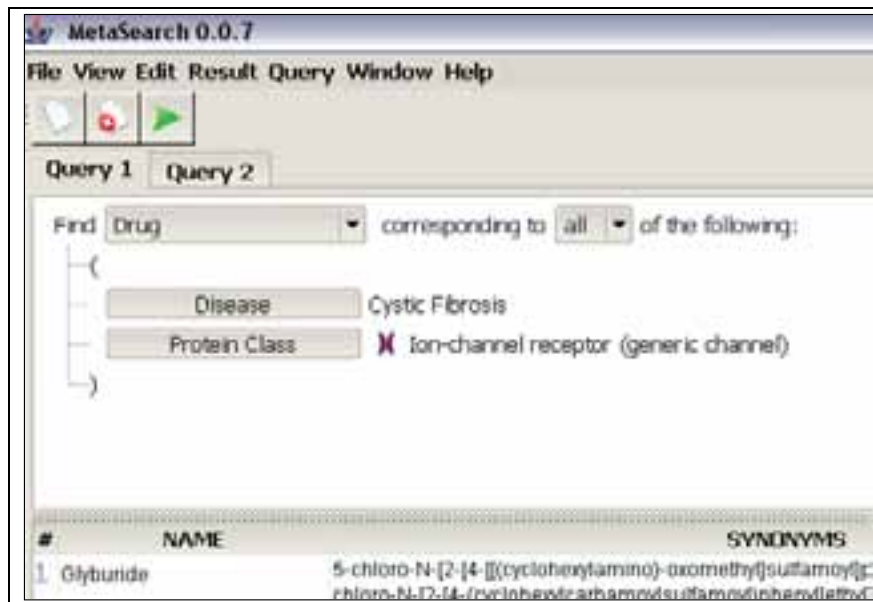
B

Figure 8. **A**. Launching MetaSearch from main menu.
B. Design and execution of a query in MetaSearch.[6]

TOOLS

Editing and Customization

MetaMiner™ (CF) is an interactive environment. A user can edit and customize any map, network, protein/gene list, add or remove new experiments, and exchange data and results with colleagues etc. Customization and exchange functionality is described in the MetaCore™ manual.

MapEditor™

MapEditor™ is a stand alone Java application available within MetaMiner™ (CF) license. MapEditor™ combines a functionality of an interactive graphic editor which draws high resolution pathway maps with a search/retrieval tool which fetches proteins, genes, compounds and interactions from the underlying MetaMiner™ (CF) knowledge base. A user can draw maps from scratch, convert networks into a map view or edit any annotated map included in MetaMiner™ (CF). Every user is encouraged to build their own versions of CF maps and draw additional maps using MapEditor™. Importantly, the researchers will have complete control over access to the custom maps they build. The maps will be securely stored on individual accounts and shared only at Investigator's will. All 30 original CF maps assembled in the "Cystic fibrosis" folder are available for editing by a straightforward upload procedure from MapEditor™ (Fig. 9).

When the map is uploaded, a user can add and remove objects and interactions, change sub-cellular localizations for objects and create objects and interactions if they are not found in the knowledge base etc. On Fig. 10, a novel "custom" interaction is added for CFTR to the map "CFTR folding and maturation."

As soon as the custom map is created, it can be "published", i.e. added to the list of all maps in MetaMiner™ (CF) in the "Custom maps" folder. All visualization and mapping tool in MetaMiner™ (CF) will work on the custom map. It will be included with the rest of maps in enrichment analysis of uploaded experimental data and gene lists (Fig. 11).

White Paper - MetaMiner Cystic Fibrosis (CF)

8

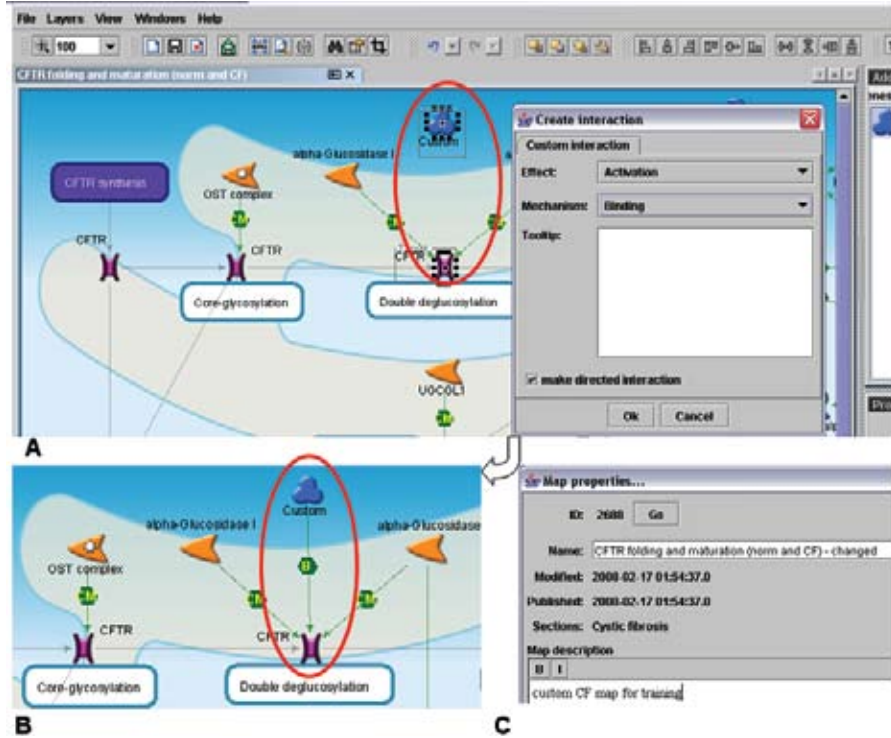**Figure 9.** Uploading CF maps for editing in MapEditor™

Figure 10. Customization of CF pathway maps in MapEditor. **A, B.** A new interaction between a "custom object" and CFTR is added to the standard map; **C.** The edited map is annotated with new information
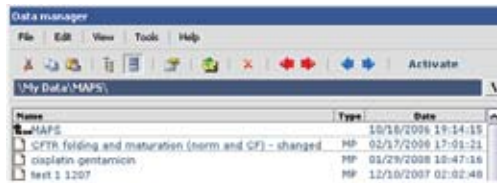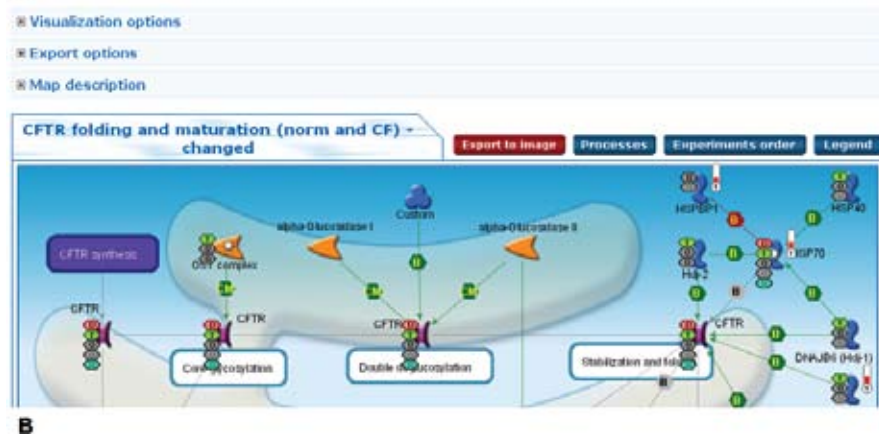

Figure 11. Custom edited CF map in MetaMiner™ (CF).

A. *The edited map is added to the "Maps" folder in Data Manager.* **B.** The edited map with visualization options (disease, orthologs, tissue, drugs) and the gene expression data superimposed (red histograms)[8]

White Paper - MetaMiner Cystic Fibrosis (CF)

Figure 12. General schema of functional analysis of experimental data and gene/protein lists in MetaMiner™ (CF).
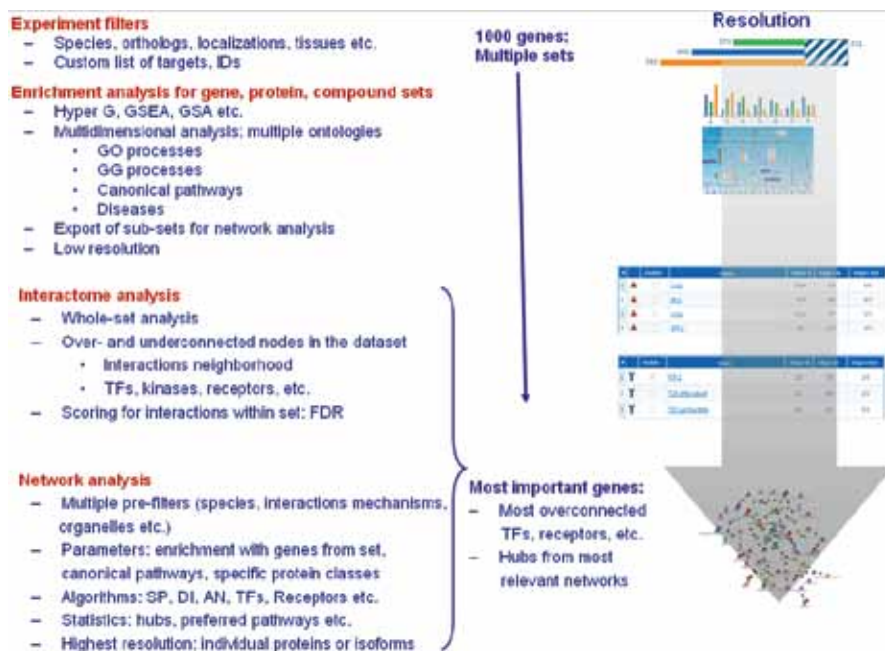
Analysis of experimental data in MetaMiner™ (CF)

MetaMiner™ (CF) is designed primarily for functional analysis of OMICs experimental data (expression arrays, proteomics, metabolomics), gene and protein lists. There are many ways to analyze the data; some of which are automated in workflows. All necessary information is available in the MetaCore™ user manual and learning materials. Here, we will provide a brief overview of MetaMiner™ (CF)'s capabilities.

The general schema of functional analysis is presented in Fig. 12. The uploaded experiments or gene/protein/compound lists are subjected to four levels of analysis:

- The experimental set(s) is custom "filtered" according to the user's needs. Filters include gene expression in human tissues and cellular organelles, matching with orthologs in 10 organisms, specific cellular processes etc. In addition, the uploaded gene lists can be normalized against the microarray content or a custom dataset.

- Enrichment analysis (EA) in multiple functional ontologies. EA is a "classical" tool which shows relative prevalence of genes from certain cellular processes, pathways, diseases etc. in the uploaded dataset(s). In MetaMiner™ (CF), 8 different functional ontologies are used for analysis. Relative impact of different entities within an ontology calculated as a p-value of hyper-geometric distribution.

- Interactome analysis is a unique feature of MetaMiner™ (CF) not found in any other data analysis platform. Interactome algorithms calculate relative connectivity (number of interactions) of individual proteins/genes within the set compare to the whole database. Proteins are divided onto protein classes such as transcription factors, receptors, ligands (secreted proteins), kinases, phosphatases, proteases and endogenous metabolic enzymes. Connectivity can be calculated for individual datasets and between the datasets.

- Network analysis. Genes/proteins in the dataset(s) can be connected to each other via protein interactions, forming signaling and metabolic networks. The network topology and composition vary depending on the chosen algorithms, filters and purpose of analysis. Networks provide the highest resolution among functional analysis tools.

- Eventually, functional analysis results in compilation of genes/proteins "most relevant" for the experimental condition. "Relevance" may be defined at all 4 levels of analysis. For instance, interactome analysis finds most over-connected TFs, receptors and ligands; network analysis results in conversion and diversion hubs, longest pathways on the networks etc., and enrichment analysis can identify genes in a particular process of interest. The gene/protein lists of interest can be extracted at any level and saved within MetaMiner™ (CF) or exported as an Excel file.

How to access the MetaMiner™ Cystic Fibrosis Platform

Access to the CF maps, networks and disease pages are available via MetaCore™ that can be licensed from GeneGo Inc by contacting sales@genego.com, Tel: 858 756 7996. All CF researchers receive discounts so please state that you are a CF researcher.

We also have 2 week free trials available by downloading trial paperwork, http://www.genego.com/productTrials.php signing and faxing to 760 479 2059.

*Cystic Fibrosis Committee Members:

Jerry Wright, Ph.D., Chairman	William B. Guggino, Ph.D.
Christopher Karp, M.D.	Jerry Nick, M.D.
Eric Sorscher, M.D.	Bill Balch, Ph.D.
Jay Kolls, M.D.	Kim Folger-Bruce, Ph.D.
Pamela Zeitlin, MD, Ph.D.	David Fischer, Ph.D.
Paul McCray, M.D.	Diana Wetmore, Ph.D., CFFT, Sponsor
Ray Frizzell, Ph.D.	

** The cystic fibrosis content is updated with the regular quarterly updates of MetaCore™ database. Therefore, the actual numbers may be higher that stated in the report.

1. The "SNPs, Genes, RNAs, Proteins" section (Figure 1A) is in the Disease Info page for Cystic Fibrosis. To get to this page, from the Data Manager, click the 'Search' link in the upper left. Then search for 'cystic fibrosis' in the "Disease classification search" section. Then click the "Cystic Fibrosis" link on the search results page. The page that opens is the Disease Info page for Cystic Fibrosis. To get to the "Gene-disease link details: DNA alteration" section (Figure 1B), click the 'Details' button next to the desired gene in the "SNPs, Genes, RNAs, Proteins" section.

2. To access the drug details page (Figure 2), from the Data Manager, click the 'Search' link in the upper left. Then search for 'Amikacin' in the "Drug search" section. Then click on the word "Amikacin" in the compound name list.

3. To access these maps (Figure 3 A and B), from the Data Manager, click the "GeneGo maps" link at the top of the page. Then click "Cystic fibrosis" in the "Disease maps" folder. Then click on the map name link as listed in the Figure 3 caption.

4. The "Cystic fibrosis – drugs and drug targets" pre-built network can be found in the "Disease Pathway Networks" section of the Disease Info page for Cystic Fibrosis. See Note 1 for a description on how to get to this page.

5. To get to these pages, from the Data Manager click the "Search" link in the upper left. Then search for 'cystic fibrosis' in the "Disease classification search" section. In the window that opens, click the "Show Tree" button to access the table in Figure 6A. Click the "Cystic Fibrosis" name link to access the page shown in Figure 6B.

6. To set up a query as displayed in Figure 8B from the initial MetaSearch window, first select 'Drug' in the drop-down menu next to "Find". Then click the "Select" button. In the window that appears choose "Diseases". Then search for 'cystic fibrosis' at the top. Then double-click the "Cystic Fibrosis" search result, single-click "Cystic Fibrosis" in the lower section, and click "Ok". Next, click the "Species" button. Then choose "Protein Class" in the window that appears. Then double-click the "Ion-channel receptor (generic channel)" option.

7. Custom objects can be added to a map in MapEditor by dragging the "Custom" object from the "Custom objects" tab of the "Add objects" panel to the currently open map.

8. Further instructions for uploading data and visualizing data on maps can be found in the "How to Upload Experiments" and "How to use Maps with Uploaded Data" tutorials accessible in the "Help" section of MetaCore.

GeneGo, Inc.

500 Renaissance Drive, Suite 106
St. Joseph, Michigan 49085 USA
(888) 592-3124 , (858)756-7996
44-778-615-0699 United Kingdom
www.genego.com
sales@genego.com