MetaMiner™
Cystic Fibrosis Report
MetaMiner™ Cystic Fibrosis (CF) represents a version of GeneGo’s MetaDiscovery™ suite that is enriched with content specific for cystic fibrosis. Now in its second phase of development, the content is reflected in several forms of cause-effect relationships including: gene-disease, compounds-disease and compound-protein associations. This information is extracted from primary literature, and manually inputted into an oracle database (1). GeneGo’s MetaDiscovery platform further creates analysis tools to enable a user to search or generate hypothesis-driven outputs against the background of the database. In addition to the underlying content, the MetaMiner CF platform extends the application of the MetaDiscovery platforms to include pre-built disease-specific canonical pathway maps and networks. Therefore, MetaMiner CF is a data analysis platform designed to be used by a broad range of CF researchers including wet lab biologists, bioinformaticians, clinicians and chemists. It is an ongoing project that was initiated in 2007 by GeneGo Inc. in collaboration with Cystic Fibrosis Foundation Therapeutics (CFFT) and a committee of expert scientists that also dictate the content of continuous quarterly updates. This reiterative process is mediated and fully funded by CFFT allowing GeneGo developers to maintain and update CF content under the guidance of the Committee*.

The original goals of this collaboration were to:

1. To assemble and maintain the most relevant cystic fibrosis biological and chemical experimental data available to date in an easy-to-access reference environment
2. To create visualization tools depicting key disease-causative and disease-progressive mechanisms in a series of pathway maps and network models
3. To provide the CF community with a “one-stop-shop” tool for uploading and analyzing experimental data in a disease-centered interface
Update of MetaMiner CF Phase II

We previously reported that the MetaMiner CF content included information on 169 genes connected by 293 association links that were extracted from 220 articles (2). The 169 genes were found to have 26,767 additional physical and functional interactions with other human proteins, DNA, RNA and compound species annotated from 23,415 articles. Also, a total of 89 drugs and compounds were linked to CF at the time. The first MetaMiner CF release also included 25 pre-built canonical maps and 10 networks, that visually described processes of CF origination, development, or perturbation of cellular processes, metabolic and transport pathways.

A summary of the most current release of MetaMiner CF is listed below. This information can be easily accessed via the “one-stop shop” Cystic Fibrosis Disease page (Figure 1) that is complete with interactive links to more detailed descriptions and references.

More Gene-Disease Associations and Biomarkers

The current content count for MetaMiner CF has gone from 169 to 206 disease-relevant genes accompanied by 42,173 physical and functional interactions with other human proteins, DNA, RNA and compounds. These interactions are annotated from 29,807 original articles. A set of publications are particularly noted as key contributions to the current update (3-8). The current platform also includes access to several gene expression profiles [GSE6802, GSE9488] (3,4). A total of 12 biomarkers (7 metabolites and 5 proteins) were identified and annotated for CF, and 457 association links (from 240 original articles), specific to the gene level, were categorized by the following:

- Promoter methylation 1
- Rearrangement 18
- Point mutation 8
- Single nucleotide polymorphism (SNP) 19
- RNA splice variant 1
- RNA amount change 72
- Altered interaction 9
- Protein amount change 107

New CF Specific Maps

The following is a current list of CF-specific GeneGo Canonical Maps, with the new maps listed in bold print. In total, 36 maps and 10 networks have been created since the initiation of the partnership. For a sample map please refer to Figure 4.

10 CFTR traffic and folding related maps:
- Folding and maturation (norm and CF)
- Normal wtCFTR traffic / ER-to-Golgi
- Normal wtCFTR traffic / Sorting endosome formation
- Delta508-CFTR traffic / ER-to-Golgi in CF
- Delta508-CFTR traffic / Sorting endosome formation in CF

Transport_Cathepin-coated vesicle cycle
- wtCFTR and delta508 traffic / Cathepin coated vesicles formation (norm and CF)
- wtCFTR and delta508-CFTR traffic / Generic schema (norm and CF)
- wtCFTR and deltaF508 traffic / Late endosome and Lysosome (norm and CF)
- wtCFTR and deltaF508 traffic / Membrane expression (norm and CF)
5 different Cholesterol and Sphingolipids transport:
- Distribution to the intracellular membrane compartments (normal and CF)
- Generic schema (normal and CF)
- Influx to the early endosome in lung (normal and CF)
- Recycling to plasma membrane in lung (normal and CF)
- Transport from Golgi and ER to the apical membrane (normal and CF)

2 different immune response maps:
- Antigen presentation by MHC class I
- Bacterial infections in normal airways

4 different Inhibitory action of Lipoxin maps:
- Lipoxin A4 on PDGF, EGF and LTD4 signaling
- Lipoxins and Resolvin E1 on neutrophil functions
- On Superoxide production in neutrophils
- On neutrophil migration

2 maps on the influence of Mucin expression in CF:
- via IL-6, IL-17 signaling pathways
- via TLRs, EGFR signaling pathways

3 maps describing CFTR activation:
- Mechanisms of CFTR activation by S-nitrosoglutathione (normal and CF)
- NO-dependent CFTR activation (normal and CF)
- Regulation of CFTR activity (norm and CF)

10 additional maps:
- Bacterial infections in CF airways
- Beta-2 adrenergic-dependent CFTR expression
- CFTR translational fidelity (class I mutations)
- CFTR-dependent regulation of ion channels in Airway Epithelium (norm and CF)
- Cytokine production by Th17 cells in CF
- Cytokine production by Th17 cells in CF (mouse model version)
- IL-1 beta-dependent CFTR expression
- Regulation of CFTR gating (normal and CF)
- Regulation of degradation of deltaF508 CFTR in CF
- Regulation of degradation of wt-CFTR

CF specific networks include:
- CFTR expression regulation
- Cystic fibrosis - drugs and drugs targets
- Cystic fibrosis and hypoxia
- Fibrosis signaling: common features
- Intracellular pattern recognition receptors
- Pulmonary Fibrosis
- Regulation of CIC-2
- TGF beta-1 in cystic fibrosis
- TLR signaling in airway epithelium

CF Drugs and Compounds
In addition, 11 endogenous metabolites were annotated in the disease association context, supported by 11 articles. These small molecule compounds are accompanied by 473 physical and functional interactions (and reactions) with human proteins, annotated from 506 original articles. In total, 90 drugs and compounds were linked to CF, including:

- FDA-approved drugs 15
- Drugs in clinical trials 63
- Discontinued drugs 9
- Pre-clinical drug candidates 2
Guide to Analysis Approach using MetaMiner CF Content and the GeneGo MetaDiscovery Platform

The CF content is supported by an interface centered on the disease (the CF Disease Detail page: see Figure 1). However, the GeneGo MetaDiscovery platform also enables different types of analyses ranging from a general search to workflows that assist with “dry lab” research and hypothesis generation. Furthermore, the platform can be used for complete functional analysis of CF OMICs datasets (gene expression, proteomics, SNPs, metabolomics etc.), browsing of CF maps and networks, or combinatorial search for cellular processes, genes, proteins and compounds, as well as for accessing the relevant literature. With a simple “click” abstracts of articles are available through the interface that is linked to PubMed® from MetaMiner CF pages or from any map, network or search result. Here we provide an abbreviated guide showing how to use the MetaDiscovery platform in the context of MetaMiner CF.

Note: data used for enrichment, network and interactome analysis taken from a proteomics analysis by Balch et al (6).

Figure 1. Accessing the CF Disease Detail page: To access the CF Disease Detail page, from the EZ Start window select “Browse Content” (1) then the “GeneGo Disease-Specific Content (Cystic Fibrosis)” option (2). As shown above, the Disease Detail page lists categories for further searching: SNPs, genes, proteins, compounds, biomarkers, drugs, disease maps and disease pathway network. To access more information select the heading of choice for more options. Shown above, is the option to search more SNPs, genes and proteins (3).
1) Search and Data Mining Approaches

The MetaDiscovery interface allows a user to start their analysis or hypothesis generation with a simple question. By searching the database, and in this case CF content in particular, a series of observations can be drawn and lists can be generated for further analysis. There are several starting points to browse and search:

Using the CF Disease Detail Page

MetaMiner CF content can be accessed directly from the EZ Start Browse Content section of the interface (Figure 1).

Using EZ Search

Another way to access CF information (or information on your favorite gene, protein or another disease) is the EZ Search tool. This can be accessed from the Search tab of the EZ Start interface. By selecting EZ Search and searching “cystic fibrosis”, a list of results is tabulated according to type of annotated information that exists in the underlying database. For instance in Figure 2, 23 genes match the search terms. 4 proteins and 90 drugs (etc.) are tabulated. When selecting a category in the left hand panel, the right hand panel of the results page provides a list of specific results, each in a “clickable” form to link to more details.

Figure 2: Using EZ Search to Search for CF Content: MetaCore’s EZ Search tool enables the user to find all annotated or map or network information about “your favorite gene/protein/disease/miRNA/compound. Shown above is how to access the EZ Search tool from EZ Start (1) and an example of searching for cystic fibrosis (2).
Using MetaSearch

MetaMiner (CF) in conjunction with MetaCore applications also features a powerful Boolean search engine, which can be launched from the Search tab of the EZ Start interface under the search tab (Figure 2-1). MetaSearch enables specific queries to call upon the entire database of interactions and functional ontologies, including CF-specific information. Query results can be easily exported in several formats including MetaCore-compatible list formats, or to Excel. The power of MetaSearch includes: 1) accessing information from the underlying database and 2) enables the generation of new lists of data for hypothesis generation from several perspectives, including drug discovery. For instance, a query for “find all compounds for any category of cystic fibrosis” can be designed as follows (Figure 3) to generate a list of 11 compounds.

Figure 3. Using Boolean-based queries to search for CF content: Shown above is a snapshot of the MetaSearch interface where the question “find all compounds for any category of cystic fibrosis” was proposed. The results are listed in the bottom panel (red box).
2) Enrichment Analysis

If your starting point is a list of genes, proteins, SNPs or miRNAs, one common analysis approach is to ask “What is the most representative “X” of my list?”, where “X” can be a function from one of the GeneGo ontologies: GeneGo canonical Maps, CF disease-specific maps (or networks), GeneGo processes, diseases, metabolic pathways, toxic or drug target pathways or ontologies from the public domain such as Gene Ontology. The MetaDiscovery platform addresses this approach by providing enrichment analysis tools where a list is parsed across an ontology to generate histograms with a quantitative assessment of relevance (p-value of significance). The histograms are interactive and further enable a user to visualize pre-built maps or networks representing a biological process with the input data overlaid or marked. Therefore a user can input a list from CF targeting experiments and determine key perturbed functions across one ontology at a time or across all simultaneously using a workflow. The addition of CF-specific maps creates a more specific CF-rich environment in order to determine enriched CF-specific processes. Enrichment analysis can also be used for several different data types or similar data types for comparison. The steps required to conduct an enrichment analysis workflow are summarized in Figure 4.

Figure 4A.
3) Addressing Data Connectivity and Drawing Pathways

**Network Building Algorithms and Interactome Analysis**

Another common approach in determining the collective function of a list of genes, proteins, SNPs, miRNA (etc) is to ask “How are these items connected?”, “What is their biochemical relationship?”, “Is there a common denominator?”, and “Is their connectivity relevant?” GeneGo’s network-building algorithms (Figure 5a) provide the flexibility to “connect-the-dots” between genes, etc. on a list with a level of control that cannot be compared to alternative pathway-building tools. For instance, it is possible that a set of genes from a wild-type CFTR over-expressing cell line differs from the Delta508-expressing cell line (which can be determined with enrichment analysis too). As a result, the interactions that co-exist also differ at various levels (cell or tissue or process-dependent). With the direct interaction algorithm, differences in the closest-knit associations between the items on the list can be determined. However, changes may occur several steps away from a direct interaction of a gene set and a different algorithm such as auto-expand is required. Additional algorithms give the ability to expand the biological space associated with the input list, determine most connected hubs of signaling and identify regulatory factors that may impact the list as a whole (i.e. common transcription factor - Figure 5c). For a white paper on network algorithms, email training@genego.com.
Figure 5. Building a network using the Analyze Network-Receptor Algorithms: To assess connectivity between protein expression from HEK293 cells expressing wild-type or DF508 CFTR, with the specific goal of identifying common hubs of regulation by transcription factors, the “analyze network” algorithm was used (a). In this algorithm, sub-networks generated based on highly-saturated items from the input list, are ranked by a p-value and G-Score and interpreted in terms of Gene Ontology. The user has a choice of sub-network to visualize from a network list (b). In this example, a sub-network generated around CFTR is shown (c). The networks are similar to the maps where the colored arrows represent directionality, effect and mechanism of interaction. Any interactions that are part of a canonical map are highlighted in teal blue.
**Interactome Analysis**

Moreover, the interactome tool assesses connectivity amongst the data versus the database in a tabular form within a single data set, or between data sets (Figure 6). The MetaDiscovery interactome options include tabulating according to protein class or amount of over versus under connectivity. Below is a sample of the interactome analysis by proteins function.

Figure 6. Interactome Analysis of CF data: Protein expression from HEK293 cells expressing wild-type or DF508 CFTR were subjected to the interactome by protein class function in MetaCore. Shown is a sample the resulting output including a legend (a) and the list of connected objects (b). A sample observation is marked in red: CFTR is connected to 22 out of 23 items from the data itself and 303 out of 19836 interactions in the database. Similarly, GLUT4 (a glucose transporter) is connected to 7/23 data objects and 666/19836 within the database indicating that CFTR is more connected to the data set than GLUT4, despite GLUT4 having more interactions with the database.
**Additional Translational Approaches**

More translational applications of MetaMiner CF include:

1) Access to drug details from a search or from a gene/protein/disease of interest
2) Biomarker targeting workflows
3) Toxicity analysis workflows
4) Customization of visual outputs using MetaLink to map novel interactions (unknown to the public domain).
5) Customization of visual outputs using MapEditor to create maps with custom objects, text, links and cellular localizations – ideal for presentations

For a training session on how to execute these approaches please contact training@genego.com. The advanced applications of MetaMiner CF also include integration of GeneGo CF analyses with third-party software suites to enable seamless export of cystic fibrosis networks or analysis lists from MetaMiner CF to Cytoscape, GeneSpring, DecisionSite, and Resolver, following the addition of the appropriate plug-ins.

**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. Go to http://www.genego.com/trials, and complete the online form.

*Cystic Fibrosis Committee Members:*

Jerry Wright, Ph.D. Tzyh-Chang Hwang, M.D., Ph.D.
Christopher Karp, M.D. Benjamin Gaston, M.D.
Pamela Zeitlin, MD, Ph.D. Thomas Kelley, Ph.D.
Ann Harris, Ph.D., M.A.
Diana Wetmore, Ph.D. and Elizabeth Joseloff, Ph.D. (CFFT Sponsors)

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