

## Website Glossary

### **Academic:**

A person who works as a researcher/ teacher at a learning institute.

### **ADMET/ ADME-Tox**

ADMET is an abbreviation for **a**bsorption, **d**istribution, **m**etabolism, **e**xcretion and **t**oxicology. This abbreviation is used in compound evaluation in drug development. It is the criteria used to determine optimal drug levels, kinetics of drug exposure and pharmacological activity for clinical use.

### **Algorithm:**

In its simplest form, an algorithm is a process that performs a sequence of operations. The concept of an algorithm is central for explaining how a system can arise, starting from a set of axioms (starting propositions that are not proven but self evident and accepted) and rules.

In GeneGo, algorithms are used to build [pathway networks](#) with a given list. There are 10 network building algorithms and each algorithm is a list of well-defined instructions for completing the network building task. Each algorithm has a defined starting state (axiom) with instructions that compute an end state via a well-defined series of successive steps. All network building algorithms start with a [root node](#) with defined edges within a defined network space. A guide to selecting the best algorithm for you network building needs can be found in our MetaCore help section by searching "algorithm". Details are also summarized in a white paper found [here](#).

[NOTE: check with programmers on type of algorithms / computer language- how it is implemented](#)

### **Amino Acid:**

An amino acid is a molecule that have an amine and carboxyl functional groups. There are 22 natural amino acids and play a crucial role in metabolism, can act as a cofactor in enzymatic reactions and is the basic unit of a [protein](#).

### **Analysis:**

An analysis is the process of dissecting a complex problem into smaller segments to gain a better understanding of the given problem. In GeneGo, to conduct an "analysis" usually applies to dissecting a list of biological or chemical identifiers to determine their collective biological function. GeneGo provides several types to analysis tools to enable different types of analysis outputs including

- 1) Enrichment analysis
- 2) Pathway Network building
- 3) Interactome
- 4) Disease biomarker identification
- 5) Toxicity biomarker identification
- 6) Metabolite prediction
- 7) QSAR modeling
- 8) Chemical target analysis

### **Bioinformatics:**

Bioinformatics can be defined as a field of science in which biology, computer science, and information technology merge into a single discipline to analyze biological information. This includes the use of computers to create life science data bases, implement algorithms to model biological systems, link databases to experimental data, processes experimental data and analyzing experimental data managing genomics data while others insists it is a science developed to solve life science problems with computer language. Often, it involves the creation of extensive electronic databases on genomes, protein sequences etc.

### **Biologics:**

A biologic is a product derived from living organisms, used for therapeutic purposes. Examples of a biologic include solutions of peptides or proteins used in vaccines and monoclonal antibody target therapy and blood test products. Emerging biologics include cell based therapies, such as cell replacement or stem cell treatment. Unlike chemical products, biologics are delicate substances or cells that are sensitive to heat, to light, and to being shaken when in liquid form, and are easily susceptible to contamination.

### **Biomarker:**

In a very broad sense, a biomarker is a substance that can be measured and used as an indicator of a biologic state. Usually biomarkers are used to distinguish between a "normal" and "disease" state. However, this term is also applied to a substance that is used as measure of a pathogenic processes, pharmacologic response, toxicity response etc. Depending on the context of the biomarker, the specific requirements for a protein/gene/miRNA to be called a biomarker varies. For instance, a [disease biomarker](#) in diagnostic clinical sense is usually a protein that is secreted in a measurable fluid (serum, urine) at levels that are significantly higher than "normal". At

GeneGo, if a gene/ protein/miRNA is a biomarker if has been referred to as such in the public domain, with validated experiments to verify its role as an indicator. GeneGo's MetaDiscovery platform contains several filtering tools to allow for further defining "biomarkers" before an analysis is conducted ( ie select for genes/proteins/miRNA that are expressed in a particular fluid or tissue etc.)

### **Biotechnology:**

Biotechnology is the practical application of biological sciences (genetics, microbiology, animal cell culture, molecular biology, biochemistry, embryology, cell biology) to industrial and commercial objectives.

### **Canonical Pathway:**

The term canonical is an adjective to describe an established rule. Traditionally, a canonical pathway is a type of [signaling](#) pathway that is established for a given cellular molecule that participates in a series of linear and stepwise interactions. At GeneGo a canonical pathway is also called a canonical GeneGo Map that has been created internally (not obtained from other sources of canonical pathways). GeneGo's pathways provide an underlying, established story line for a given molecule, or biological process (regulatory or metabolic) or disease mechanisms. Each GeneGo Map is interactive and must have:

- 1) An average of 7 linear, stepwise interactions
- 2) Interactions that have been validated repeatedly and well excepted in the field
- 3) A defined start and end, usually spanning cellular biochemistry ranging from the extracellular space, through ligands and receptors, to intracellular signaling via enzymes and kinases etc) to nuclear functions mediated by transcription factors.

### **Case Study:**

A case study is an exemplary or cautionary model created with intensive detail to stress factors contributing to its success or failure. GeneGo's provides a series of case studies as an instructive and exemplary sample of how to use the MetaDiscovery platform for various analysis needs. These case studies can be found [here](#) and include examples on:

- 1) How to use Metacore for genomic analysis
- 2) How to use Metacore and MetaLink for analysis of novel or predicted interactions
- 3) How to use MetaSearch and MetaCore in combination with other data mining tools for discovering disease biomarkers
- 4) How to use MetaDrug to predict toxic drug outcomes

- 5) How to use MetaDrug to predict drug side effects
- 6) How to use MetaDrug to compare compound metabolites and targets
- 7) How to use MetaDrug to reposition a drug
- 8) How to use MetaCore to analyze metabolite profiles

### **CDX file**

A CDX file is a ChemDraw (CambridgeSoft Corp.) native file format for holding information about the atoms, bonds, connectivity and coordinates of a molecule.

### **Cell:**

A cell is the basic unit of all living organisms. It is a membrane bound structure containing biomolecules (nucleic acids, proteins, and polysaccharides), and chemicals with a specialized function (except in the case of a stem cell where the function is less restricted and can change). A cell is an autonomous self-replicating unit that is organized into cellular compartments or organelles, each of which carries out a particular function important for its maintenance.

### **Central Dogma of biology:**

The principle that genetic information is passed from DNA to RNA to protein.

### **Chemical:**

A chemical substance can either be a pure chemical element or a pure chemical compound (of several elements). Each chemical has a defined atomic or molecular structure that results from, or takes part in, reactions involving changes in its structure, composition, and properties. Each chemical atom is composed of protons, neutrons, and electrons. A chemical in its natural element form, its atom cannot be further broken down by chemical means.

### **Cheminformatics:**

A broad definition of cheminformatics includes the use of computer and informational techniques, to address a range of chemistry oriented problems including high throughput analysis of chemical screens, chemical databases management, Quantitative Structure Activity Relationships (QSAR) prediction and application, 2D and 3D concepts, and drug design.

### **Clinical:**

Any observation or measurement based on patient diagnosis or treatment

**Collaboration:**

A collaboration is the act of several participants working jointly to achieve a common goal.

**Community:**

A social group whose member may reside, work or share common interests.

**Compound:**

A compound is a composition of two or more different elements and therefore can be broken down into simpler type of matter (elements) by chemical but not physical means. Compounds also have properties that are different from its original component elements. In GeneGo the compound category is any endogenous or made compound.

**Conference:**

A conference is a prearranged meeting with the goal of exchanging or discussing information.

**Curation:**

Curation is the process of gathering and organizing information. With respect to biochemical curation, this refers to extraction of public data from the literature. GeneGo's curation process is a manual extraction method to obtain all biological and chemical information from the literature by a PhD and medical experts. The extracted information is from experimental data that validate human, mouse and rat protein-protein, protein-DNA and protein-compound relationships, metabolic and signaling pathways, and the effects of bioactive molecules. For each relationship the effect, mechanism and species must be determined. GeneGo's curation also includes formation of ontologies (GeneGo Maps, GeneGo Processes, and GeneGo Diseases, GeneGo metabolic, toxicity and drug target networks) to store the information in a structured form in an oracle database.

**Dataset:**

A set of variables generated from an experiment usually containing a list items to be measured and the resulting measured values.

**Disease:**

The term disease is used to describe the abnormal function/condition of a part of an organ, or system of an organism that results from various causes including infection,

genetic defect, or environmental stress. A defined set of symptoms may accompany a specific disease to facilitate diagnosis or classification

### **DNA:**

DNA is the abbreviated form of **deoxyribonucleic acid**. It is composed of 4 ribonucleotides called adenine, guanine and cytosine (A, G, C) and thymine (T). Each nucleotide is composed of a phosphate group, a deoxyribose sugar ring, and a nucleobase . DNA is structured as a double helix, found in the nucleus of cells. Each DNA strand contains the basic genetic code of a living organism. DNA is used as a template to produce RNA (ribonucleic acid) which is used as template to produce proteins (the “doers” of the cell).

### **Drug:**

A drug is a substance used in the diagnosis, treatment, or prevention of a disease or as a component of a medication. In GeneGo the “drug” category is specific to compounds that are approved by a governing agency, in clinical trials or discontinued. This category is further divided into biologics (composed of biological material-usually nucleic acid composition) and small molecules (chemical composition)

### **Drug Target:**

A drug target is usually a protein that may be implicated in a disease whose activity may be altered by artificial means (drug administration).

### **Edge:**

The term “edge” in GeneGo refers to the definition used in graphing mechanics in generating a biological network and is used in combination with the term “node”. An edge denotes the association of any biological/chemical component that is annotated in GeneGo's data base to have an effect on another object ( a node), and is marked by an arrow.

### **Enrichment:**

There are several definitions for enrichment all based on the act of making something more rich or full. In GeneGo, enrichment or the [enrichment analysis tool](#) is used to determine the most populated pre-built map or network created by GeneGo, with a given/ user dataset. Each prebuilt map or network exists in an organized ontology of which there are nine GeneGo ontologies and three Gene Ontology options. Therefore in enrichment analysis, a given user data list is first parsed against the data base to match gene/ proteins identifiers then through the selected ontology and the result is list of most data populated maps or networks.

**Experiment:**

A scientific experiment is the process in which causal relationships among variables is investigated. The term experiment is also used in GeneGo to refer to an uploaded dataset containing miRNA, RNA, gene or protein identifier ( in any publically available form including Entrez, Swiss prot, and specific chip probe ids such as affymetrix, agilent, illumina).

**FDA:**

FDA is an abbreviation for **F**ood and **D**rug **A**dministration which is a United States federal agency in the Department of Health and Human Services that is responsible for regulating the release of new foods, drugs and other health-related products.

**Gene:**

A gene is the basic hereditary unit in a living organism that encodes for a given protein product. It is composed of DNA and each gene stretch contains a coding (exon ) and non-coding (intron) region. Genes are compacted into chromosomal structures. Some regulatory units that each gene has include a promoter region and an enhancer region to facilitate initiate [transcription](#).

**Genetics:**

Genetics is the study of heredity or the transmission and variation of inherited characteristics, in particular chromosomes and DNA.

**Genome:**

A genome is the complete set of genes of an organism, usually described in the context of number of chromosomes or base pairs.

**Genomics:**

Genomics is the branch of genetics that studies full genomes of an organism.

**Grant:**

A grant is financial assistance which does not generally have to be repaid. Grants are usually obtain after an application and competition process.

**Hidden nodes:**

A hidden node is an object in a biological system that maybe overlooked due to the lack of changes in expression. In biology, many proteins play integral roles in signaling and are regulated not by expression but other modification, such as phosphorylation, or change in cellular localization. More importantly, in systems biology modeling,

characterization of a set of proteins requires a method to account for a better correlation between levels of protein and gene expression. In other words, cellular processes such as mRNA degradation, alternative splicing, translational regulation, post-translational modifications and protein degradation also need to be accounted for. GeneGo Inc has developed an algorithm to identify "hidden nodes" affected by such process in a given data set. To use this algorithm click [here](#).

### **Hypergeometric distribution statistic:**

Hypergeometric distribution is a type of statistic determined for population based analyses. It describes the probability of the number of successful events would occur when selections are made from two groups without replacing members of the groups from a finite population. It is used in random sampling for statistical quality control. A simple everyday example would be the random selection of members for a team from a population of girls and boys.

### **Integration:**

At GeneGo this word is used in the context of computing integration and refers to the creation of links between previously separate computer systems, applications, services or processes. GeneGo's MetaDiscovery platform integrated with several software applications for pre and post analysis or data capturing of the results obtained from MetaCore or MetaDrug. Integration enables easy import from or export to other programs. The MetaDiscovery platform is integrated with:

- 1) Endnote- for exporting and managing references obtained from MetaDiscovery analysis
- 2) Resolver from Rosetta Biosoftware- importing and exporting lists. It requires that you have a license to the Rosetta Resolver Software Developers Kit (SDK) . It is a useful tool for statistically processing, managing and sharing large data sets. For details on how to use Resolver please refer to <http://www.rosettabio.com/products/resolver> for their technical support.
- 3) Expressionist from Genedata- for importing and exporting lists. It is a useful tool that supports high throughput, standardized and workflow-based data processing and statistical analysis of large data sets. For details on how to use Expressionist please refer to <http://www.genedata.com/products/expressionist/> for their technical support.
- 4) Genespring from Agilent- GeneSpring GX is another statistical tool for fast visualization and analysis of expression data. More information on it's deliverables can be found here <http://www.chem.agilent.com/en-US/products/software/lifesciencesinformatics/genespringgx/Pages/default.aspx>.

- 5) Spotfire DecisionSite- This tool allows for the integration of different data types (includes chemical) and has query capabilities. For more information visit: <http://stn.spotfire.com/stn/Platform/DecisionSite.aspx>
- 6) Discovery Gate – ideal to search for chemical based information. Refer to <https://www.discoverygate.com/iss/servlet/App?Action=ShowApp> for more details on how to access.
- 7) Inforsense: tool that allows database integration, statistically gene expression and literature search. A good reference in its uses can be found here [http://www.oracle.com/technology/industries/life\\_sciences/presentations/olsug\\_june04/olsug\\_june04\\_inforsense.pdf](http://www.oracle.com/technology/industries/life_sciences/presentations/olsug_june04/olsug_june04_inforsense.pdf)
- 8) SciTegic Pipeline Pilot- for linking databases. This tool is ideal if you have access to MetaBase and would like to search information and combine the information with other database functions.
- 9) Illumina GenomeStudio- enables efficient data analysis for whole-genome sequencing and resequencing experiments performed using the Genome Analyzer. It displays genome-wide consensus alignments using single-read and paired-end sequencing data as well as SNP discovery and validation experiments using output files generated in the CASAVA module of the Genome Analyzer Pipeline analysis software.
- 10) Cytoscape- This is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data. <http://www.cytoscape.org/>

Instructions on how to use these integrations are listed at [http://portal.genego.com/help/Help\\_Pages/mc\\_integrations.html](http://portal.genego.com/help/Help_Pages/mc_integrations.html).

### **Interaction:**

An interaction is kind of action that occurs between two or more objects that have an effect upon one another. In GeneGo an interaction can also be thought revered to the interconnectivity or relationship between two objects. We define an interaction to be directional, with a defined effect (positive or negative) and a specific mechanism. The types of annotated mechanisms include: binding, catalysis, cleavage, competition, covalent modifications, phosphorylation, dephosphorylation, microRNA binding, transcription regulation, transformation, transport for physical interactions and influence on expression, competition, pharmacological effect, toxic effect for functional interactions

Therefore, on any GeneGo pathway, an Interaction can be physical or functional and are represented on pathway maps and networks by arrows between network objects (nodes). The mechanism is marked by a symbolic marking (e.g., B, CM, +P, -P) in a

hexagon on the arrow. All symbols and markings are described in legends for maps and networks and can be obtained [here](#).

### **Interactome:**

An interactome is the defined set of all molecular interactions in cells. GeneGo's MetaCore platform enables interactome analysis with the interactome tool that allows a user to:

- 1) Estimating interconnectedness of an experimental dataset (density of interactions),
- 2) Finding statistically significant interactions in the set
- 3) Enrichment of the dataset with protein classes
- 4) Estimating interconnectedness between several datasets,
- 5) Identification of statistically significant data objects that interact with and regulate other proteins,
- 6) Analysis of different types of data (SNP, genomic, proteomic, etc.) simultaneously

### **Journal:**

A journal is a set of articles published periodically and is dedicated to a particular subject. A list of recent GeneGo journal publications can be found [here](#).

### **Knowledge Mining:**

Knowledge mining is the process of extracting information from an existing data source (either experimental data or established databases). Knowledge mining is applied to several types of "searching" including finding patterns from several datasets ("signature set"), or finding lists of information for further analysis finding known genes/ proteins for a specific disease or targets for a specific drug.

### **Manual:**

A manual is written book of instructions on how to use a particular program, machine etc.

### **Map (see Canonical Pathway)**

#### **MapEditor:**

MapEditor™ is a unique bioinformatics module which enables custom editing of existing canonical pathway maps, conversion of networks into map views for pathway analysis, and drawing regulatory and metabolic maps from scratch. A user may choose the objects for their maps from the manually curated MetaCore™ database of genes, proteins, compounds and interactions, or introduce new objects. Custom maps then

can be securely “published” (added to MetaCore’s standard maps collection on the user’s server), used as a template for mapping experimental data, saved and shared with colleagues. Users can focus and organize their wet lab or in silico research around a set of interactive maps created in MapEditor™ and linked to MetaCore™.

MapEditor™ for pathway analysis is an “add-on” tool seamlessly integrated with both MetaCore™ and MetaDrug™. More product highlights can be found [here](#).

### **Medicine:**

There are many definitions to the term “medicine”. For instance a drug that is used to treat an ailment is called medicine, while the practice of medicine is the learned profession that is mastered by graduate training in and is devoted to preventing or alleviating or curing diseases and injuries

### **MessengerRNA:**

There are several forms of [RNA](#) in a biological cell. Messenger RNA (mRNA) is a template for protein synthesis and is transcribed from DNA to pass genetic information from DNA in the nucleus to the ribosome sites of protein synthesis in a cell. It is similar to DNA in that it is composed of 4 ribonucleotides called adenine, guanine and cytosine (A, G, C) and uracil (U). Each nucleotide is composed of a phosphate group, a ribose sugar ring (not the deoxy- form found in DNA), and a nucleobase. Eukaryotic mRNA usually contains a 5' cap (is a modified guanine nucleotide that has been added to the front or 5' end ) and a 3' polyadenyl moiety (at the back end of the mRNA).

### **Metabolic Pathway**

A specific type of directional pathway that describes a series of individual chemical reactions in a living system that combines to perform one or more important functions. The product of one reaction in a pathway serves as the substrate for the following reaction. Metabolism is the the collection of metabolic pathways describes cellular anabolism (generation) and catabolism (breakdown) of biochemical components and consume or produce cellular energy, receptively. The set of metabolic pathways in GeneGo comprise the Metabolic Network Ontology. Networks pathways in this ontology are species specific for human, mouse and rat.

### **Metabolism:**

Metabolism is the the collection of metabolic pathways describes cellular anabolism (generation) and catabolism (breakdown) of biochemical components and consume or produce cellular energy, receptively.

**Metabolites:**

A metabolite is any substance produced by, or taking part in, a metabolic reaction. Metabolites can be generated from endogenous compound (that participate in cellular biochemistry) and also for exogenous compounds. Therefore, in GeneGo, MetaCore enables the analysis of known, *measured* metabolites (see Metabolomics) while MetaDrug enables the *prediction* of metabolites that may be generated by a chemical structure (ie novel compound intended for therapeutic use). Refer to MetaDrug help section for more details.

**Metabolomics:**

Metabolomics is the systematic study of the unique chemical components that are the remnants of cellular processes. It is often referred to as the study small-molecule metabolite profiles as a metabolome represents the collection of all metabolites in a biological organism, which are the end products of its gene expression. Metabolic profiling can give insight to the physiology of a cell. In GeneGo, a metabolic profile can be uploaded and analyzed in the context of pathways, with or without accompanying genomic or proteomic data. A specific data parser and format is used; for more details refer to MetaCore help section.

**MetaBase:**

MetaBase™ is a proprietary manually curated database that consists:

- 1) Experimentally validated human protein-protein, protein-DNA, protein-RNA, and protein-compound interactions
- 2) Substrate, product reactions
- 3) Compound and drug content,
- 4) Gene-disease relationships
- 5) Ontologies of metabolic and signaling MetaCore:

**MetaCore:**

MetaCore™ is an integrated software suite for functional analysis of biological experimental data. The scope of data types includes microarray and SAGE, gene expression, siRNA, SNPs and CGH arrays, proteomics, metabolomics, microRNA, pathway analysis, Y2H and other custom interactions. MetaCore™ is based on [MetaBase](#). The analytical package includes easy to use, intuitive tools for searching, data visualization, mapping and exchange, multiple networking algorithms and filters.

**MetaDrug:**

MetaDrug™ is a unique systems pharmacology platform designed for evaluation of biological effects of small molecule compounds on the human body, with

bioinformatics applications from toxicogenomics to translational medicine. Analysis is started with a chemical structure in mind (accepts CDX, mol and SDF formats) and allows a user to:

- 1) Predict compound metabolites
- 2) Search MetaBase for similar compounds
- 3) Predict activities according to structure/ activity models (QSAR)
- 4) Create a new QSAR model
- 5) Identify biological targets for the input compound and predicted targets
- 6) Functional analysis of targets
- 7) Integration with other MetaCore™ tools.

The product is intended to be used by medicinal chemists and biologists active in pre-clinical areas of high content and high throughput screening, bioassays, hit-to-lead libraries, lead optimization, PK and toxicogenomics, with the eventual goal being practical applications in translational medicine and more. MetaDrug™ is based on [MetaBase](#)

#### **MetaLink:**

MetaLink™ is a feature in MetaCore™ that enables a user to upload interaction based data such as Y2H, immunoprecipitation experiments, to map potential novel interaction on top of what is annotated in the database (see Parkin case study [here](#)). Common uses also include to map predicted miRNA targets alongside what is known (see miRNA case study [here](#)).

#### **MetaMiner:**

GeneGo's platform for creating disease-focused content and systems biology analysis tools. MetaMiner™ projects are collaborations with experts in the disease topic in questions. Project deliverables include: manual curation and gene → disease specific annotation, creation of disease specific canonical maps, disease specific search capabilities and data repository and analysis platform. Current MetaMiner™ projects include:

- 1) Cystic Fibrosis: (completed)
- 2) Oncology: (commenced)
- 3) Lung
- 4) Metabolic diseases: (commenced)
- 5) Skin
- 6) Dry Eye
- 7) Stem cells (an cancer stem cells): (commenced)
- 8) Cardiovascular Disease
- 9) Immune and inflammation

- 10) Respiratory
- 11) CNS: (commenced)
- 12) Nutrition

**For details on how to get involved contact [Julie@genego.com](mailto:Julie@genego.com)**

### **MetaRodent:**

MetaRodent™ is a part of MetaCore, and is the additional mouse and rat specific manually curated database. In addition to annotated experimental data, MetaRodent™ includes rodent specific pathway maps, especially y in the metabolism ontology.

### **MetaSearch:**

MetaSearch™ is a feature in MetaCore™ that enables Boolean based query searching. It is equivalent to accessing the content in MetaBase with a Java based interface to export lists of results from a mutli-leveled question such as " find all biomarkers for prostate but not breast cancer that are receptors"

### **MetaTox:**

MetaTox is the equivalent of MetaMiner but for drug and compound analysis. Currently in it's second year of completion, this project includes creating a toxicology-and-tissue-specific ontology, rodent specific annotation and tox-specific pathways. More details on the deliverables can be found [here](#).

### **Microarray**

A microarray is a type of experiment conducted on chip technology and is specifically designed to handle small fluid volumes and generate an abundance of results. This technology is also known as "Lab-on-a-Chip" as it generally allows for the scaling of single or multiple lab experiments down to a single process in chip-format and greatly facilitates high-throughput screening methods. In general each chip contains thousands of microscopic spots of biological material, to act as a probe/ bait that will capture a target from an applied sample/ specimen. If a probe-target association is found, it is usually detected by a measurable fluorophore or chemiluminescence method to determine relative abundance of the target in the sequence. To analyze any microarray result with GeneGo's MetaDiscovery platform, all that is required is a list of identifies (RNA< gene, protein, chemical id). Measured intensities can also be used to set thresholds or visualize the identified material on resulting GeneGo pathways.

There are several types of microarray available and they differ in the intended material to be measured. Below is a list of microarrays with a brief description of how they are used:

- 1) cDNA microarray: uses a chip that consists of a series of microscopic spots of DNA oligonucleotides, representing a specific DNA sequence (short section of a gene or other DNA element). These oligonucleotides are used as probes to hybridize (bind) to a target cDNA (complementary DNA sequence)
- 2) Protein microarray: uses chip technology to find protein-protein interactions, such as substrates of protein kinases, transcription factor protein-activation, or targets of biologically active small molecules. In this case the probe consists of different molecules of protein or specific DNA binding sequences. The most common protein microarray is the antibody microarray, where antibodies are spotted onto the protein chip and are used as probes to detect proteins from cell lysate solutions.
- 3) Chemical compound microarray: This technology is used to measure the activity of large numbers of chemicals against hundreds of biological targets. The chip in this case consists of hybridized organic chemical compounds as a probe. One form of uses is to apply biological targets in cell lysates or in purified forms and compounds that bind to the target with high affinity are then identified by immunoassays ( antibody based detection).
- 4) microRNA microarray: uses a chip that consists of a series of microRNA (miRNA) oligonucleotides, as bait to capture target miRNA in a RNA sample.
- 5) Carbohydrate arrays: uses a chip that consists of naturally occurring and synthetic carbohydrates as bait to capture target carbohydrate-binding proteins such as antibodies, selectins, cytokines, chemokines, pathogen proteins and plant lectins.
- 6) Tissue microarray: consist of paraffin blocks in which up to 1000 tissue cores are embedded in array fashion to allow high throughput histological analysis. Specifically, a hollow needle is used to remove tissue cores as small as 0.6 mm in diameter from regions of interest in paraffin-embedded tissues such as clinical biopsies or tumor samples which are then inserted in a recipient paraffin block. These blocks are then used to measure the presence of proteins using immunohistochemistry, and fluorescent in situ hybridization techniques.
- 7) Cellular microarray: this technique allows for the interrogation of living cells. The chip in this case is spotted with varying materials, such as antibodies, proteins, or lipids acting as probes. Since combinations of different materials can be spotted in a given area, when a specific interaction exists, the chip can be designed to also detect a cellular response, triggered by one interaction such as a change in phenotype, or a specific secreted factor. An example of cellular microarrays is PMHC Cellular Microarrays.

**MicroRNA:**

A microRNA (miRNA) is single-stranded RNA molecule of 21-23 nucleotides in length, known to regulate protein expression. A miRNA is first processed into a short stem-loop structure called a pre-miRNA and mature miRNA molecules are partially complementary to one or more messenger RNA (mRNA) molecules. miRNAs can be encoded by independent genes or processes from other RNA species by an enzyme called Dicer. Base pairing of a miRNA with the target RNA promotes cleavage of the RNA, preventing translation of the protein. miRNAs are also known to cause DNA methylation of promoter sites, thereby affecting the expression of targeted genes.

**Mol file:**

A file format created by Elsevier MDL® (now Symyx Technologies Inc.), for holding information about the atoms, bonds, coordinates and connectivity of a molecule.

**Network:**

A network is a graphical representation of objects (nodes) connected by arrow links (edges). At GeneGo, this term is used pre-built GeneGo networks and for custom networks created using one of MetaCore's network algorithms, from a user's data set. A network is also one form of representing [signaling pathway](#) at GeneGo. Unlike canonical maps, networks can be composed of 1-step associations and may or may not be part of an established signaling cascade.

**Network object:**

A molecular entity or its attribute (molecule, reaction, process or a relationship) represented as an object (node) in a network. Network objects in MetaCore™ are: proteins, protein complexes or groups, peptides, RNA species, compounds, EC numbers (function), or reactions. Genes are represented indirectly via the proteins they encode. The same protein in different localizations may be represented by different network objects if its interactions are localization-specific.

**Node:**

Basic unit of a graphical representation of a biologic network. At GeneGo a node is a network object. A root node is a network object derived from a user's data set or input list and is what a network algorithm starts with to build out to other nodes connected by edges.

**Ontology:**

An ontology is a form an organized set of information or concepts within a domain ( at GeneGo, within MetaBase) and the relationships between those concept. Ontologies

also provide a shared vocabulary, which can be used to model a domain — that is, the type of objects and/or concepts that exist, and their properties and relations. At GeneGo there are separate ontologies for biological processes, toxic processes, disease biomarkers, diseases, drug targets and drug action mechanisms. Each GeneGo ontology has a hierarchical tree structure with pre-built networks and GeneGo pathway maps. These ontologies are used for Enrichment analysis. GeneGo's MetaCore platform also includes Gene Ontology categories from the public domain.

### **Ortholog:**

An ortholog (in the context of biology) is a gene in a different species that is similar to a gene from another species because it originated from a common ancestor.

### **Pathway:**

A pathway is a sequence of biochemical associations by which one substance has an effect on another or is converted into another substance. A signaling pathway is a series of biochemical steps that relay information from extracellular space to inside the cell and further into cellular organelles. Signaling pathways in GeneGo contain ligands, receptors, various protein classes, DNA, miRNA, enzymatic reactions and compound target associations. There are two types of pathway representations at GeneGo: [canonical pathway maps](#) and [networks](#).

### **Pathway analysis:**

The goals of pathway based analyses are to determine the interconnectivity within a given set of data and how this is biologically relevant. This type of analysis has become key in high throughput data generation and interpreting large data sets. Often this term is used in connection with “what are all the interactions associated with my favorite gene” and “are these interactions part of a known pathway”. However, pathway analysis is also key in determining the relevance of genes/ proteins that **alone** may not be significantly altered but if it exists in a [signaling pathway](#) with a set of other altered genes/ proteins then collectively they have a profound biological outcome.

### **Pharma, Pharmaceutical industry**

Pharma refers to a company that makes and sells pharmaceuticals and the industry is collective of such companies. Pharmaceutical companies can deal in generic and/or novel brand medications and are subject to a variety of laws and regulations regarding the patenting, testing and marketing of drugs, depending on the state or country they operate in.

**Pharmacology:**

Pharmacology is the study of **drugs** ( chemical compound manufactured for medicinal purposes) , their origin, nature, properties and their effects upon living organisms.

**Protein:**

A protein (also known as a polypeptide) is an organic compound composed of **amino acids** arranged in a linear chain. A protein is the result of the " the **central dogma**": DNA contains genetic code that is **transcribed** into **messenger RNA** which **translated** into a protein. Upon translation, proteins undergo folding ( dictated by cellular environment and amino acid functional groups) to form active functional units. In their folded state proteins are the "do-ers" of a cell/ biological system. Amino acids in a protein are joined together by the peptide bonds between the carboxyl and amino groups of adjacent amino acid residues. Proteins can be chemically synthesized for pharmaceutical or nutritional purposes.

**Proteome:**

The proteome is the entire collection of proteins and their modifications of a given organism or system.

**Proteomics:**

Proteomics is the large-scale study of proteins, particularly their structures and functions.

**QSAR:**

QSAR is the abbreviation for **Quantitative Structure-Activity Relationship** and is a process by which chemical structure is quantitatively correlated with biological activity or chemical reactivity. A QSAR model is a mathematical model that correlates calculated structural descriptors with biological activity of the compounds from the training set and used to predict the activity of a queried compound. The concept of use is based on strong correlations between structure and observed properties such as the number of carbons in alkanes and their boiling points.

**Ribosomal RNA (rRNA):**

rRNA is an active catalytic component of ribosomes (site of protein translation). Eukaryotic ribosomes contain four different rRNA molecules: 18S, 5.8S, 28S and 5S rRNA. In the cytoplasm, rRNA and protein combine to form a nucleoprotein called a ribosome.

**RNA:**

RNA is short for ribonucleic acid. It is similar to DNA in that it is composed of 4 ribonucleotides called adenine, guanine and cytosine (A, G, C) and uracil (U). Each nucleotide is composed of a phosphate group, a ribose sugar ring with a hydroxyl group at the 2' position, and a nucleobase. Unlike DNA, RNA is not confined to the nucleus and can exist in the cytoplasm or other organelles. There are several forms of functional RNA in a eukaryotic cell including [messengerRNA](#), [transferRNA](#), [ribosomal RNA](#) and [microRNA](#).

**Science:**

According to Webster's New Collegiate Dictionary, the definition of science is "knowledge attained through study or practice," or "knowledge covering general truths of the operation of general laws, esp. as obtained and tested through scientific method."

**SDF file:**

Structure Data File is one of the chemical file formats introduced by Elsevier MDL® (now Symyx Technologies Inc.) capable of storing thousands of molecules and their associated information such as molecular weight, logP, activity, etc.

**Seminar:**

A seminar is a form of instruction with the goal of informing the listeners to new applications or discoveries. It has the function of bringing together small groups for recurring meetings, where each seminar is on a focus topic and encourages active participation and discussion.

**Signaling: see Pathway**

**Signaling Pathway: See pathway**

**siRNA:**

Small interfering RNA (siRNA) are also known as short interfering RNA or silencing RNA, and are double-stranded RNA molecules, 20-25 nucleotides in length, that play a variety of roles in biology. They are most known as inhibitors of gene expression. siRNAs have become a common tool for gene function and drug target validation studies as they can be designed to target a specific gene sequence and introduced exogenously into cells by various transfection methods.

**Species:**

In biology, a species is defined by a group of organisms capable of interbreeding and producing fertile offspring of both genders.

**Statistics:**

Statistics is a branch of applied mathematics concerned with the collection and interpretation of quantitative data and the use of probability theory to estimate population parameters.

**Stem Cells:**

A stem cell is a highly proliferative, unspecialized cell that gives rise to differentiated cells or can be induced to differentiate into a cell of interest and undergo self renewal. There are several types of stem cells including:

- 1) Adult stem cells: found among differentiated cells in a tissue or organ that can renew itself
- 2) Embryonic stem cells: Cultured cells derived from the pluripotent inner cell mass of blastocyst- stage embryos.
- 3) Totipotent or omnipotent stem cells: can differentiate into embryonic and extraembryonic cell types, derived from very early stage embryos (just after fertilization).
- 4) Pluripotent stem cell: descendants of totipotent cells and can differentiate into nearly all cell types, derived from later stage embryos.
- 5) Multipotent stem cells: can differentiate into a number of cells, but only those of a closely related family of cells
- 6) Oligopotent stem cells : can differentiate into only a few cells, such as lymphoid or myeloid stem cells
- 7) Unipotent cells can produce only one cell type, their own, [but have the property of self-renewal which distinguishes them from non-stem cells.

**Systems biology:**

Systems biology can be defined as “the study of an organism, viewed as an integrated and interacting network of genes, proteins and biochemical reactions which give rise to life” – from <http://www.systemsbiology.org/>. However, this discipline incorporates the integration of experimental and computational approaches and requires the expertise of a wide range of scientists- from mathematicians, biologists, statisticians, pharmacists doctors and toxicologists. At GeneGo, the goal is to develop a platform that enables any type of system biologist for their analysis goals. What remains clear that for any given system, this approach considers all the components of a system; and determining

the functions dictated by the properties and interactions of the components the system, using computational models.

### **Therapeutics:**

Therapeutics is the branch of medicine concerned with the treatment of disease as “therapy” is the attempted restoration of a health/disease complication.

### **Transcription, transcribed**

RNA synthesis, or transcription, is the process of converting DNA nucleotide sequence information into RNA sequence information. The result of transcription is a sequence of nucleotides called messenger RNA (mRNA) that is complementary to the DNA template. The mRNA is then used to carry the genetic message from the DNA to the protein-synthesizing machinery of the cell. One significant difference between RNA and DNA sequence is the presence of U, or uracil in RNA instead of the T, or thymine of DNA.

### **Transfer RNA (tRNA):**

Transfer RNAs are small 80 nucleotides segments that transfer a specific amino acid to a growing polypeptide chain at the ribosomal site of protein synthesis during translation. Each tRNA has a site for amino acid attachment and an anticodon region for codon recognition (binding to a specific sequence on the mRNA).

### **Translation, translated:**

Translation is the process by which the genetic code carried by messenger RNA (mRNA) directs the production of proteins from amino acids. The process is mediated by ribosome complexes and transferRNA that contains a 3-base pair sequence that complements a 3-base-pair sequence in the mRNA and carries a corresponding amino acid.

### **Toxicology:**

Toxicology is the study of the adverse effects (ADME) of chemicals on living organisms. This includes determining methods to detect symptoms, mechanisms, treatments and poisoning caused by the chemicals.

### **WebEx:**

WebEx is a form of conferencing online via computer-mediated communication provide by WebEx Communications Inc. by Cisco.

**Webinar:**

A webinar is a [seminar](#) conducted over the world-wide-web using web conferencing technology.

**Workflow:**

The term workflow is used to describe a sequence of tasks (performed by a person, group or computer) used to achieve a particular goal. It is in computer programming to capture and develop human-to-machine interaction where a machine can duplicate what a human would do, to provide an easier way to orchestrate or describe complex processing of data in a visual form without the need to understand computers or programming. At GeneGo there are several workflows designed for different analysis goals. Each workflow conducts an analysis operation (or 2), provides an exportable report, and contains flexible filtering options. Refer to MetaCore's help section for more information on the following workflows.

- Disease Biomarkers
- Toxicity
- Enrichment
- Receptor networks building
- Transcription factor network building