

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 November 2002 (28.11.2002)

PCT

(10) International Publication Number
WO 02/095650 A2

(51) International Patent Classification⁷: **G06F 19/00**

(21) International Application Number: PCT/CA02/00731

(22) International Filing Date: 17 May 2002 (17.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/291,928 21 May 2001 (21.05.2001) US
60/291,931 21 May 2001 (21.05.2001) US

(71) Applicant (for all designated States except US): **MOLECULAR MINING CORPORATION** [CA/CA]; 55 Rideau Street, Kingston, Ontario K7K 2Z8 (CA).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **KOTLYAR, Max**

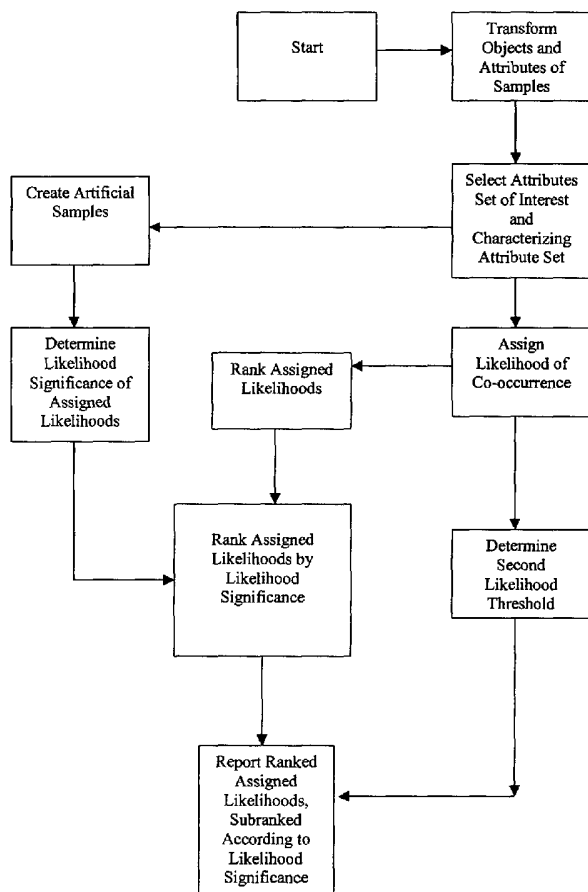
[CA/CA]; 401 Princess Street, Kingston, Ontario K7L 5C9 (CA). **SOMOGYI, Roland** [US/CA]; 1090 Cliffside Lane, Sydenham, Ontario K0H 2T0 (CA). **GREEN, James** [CA/CA]; 1-195 Earl St., Kingston, Ontario K7L 2H5 (CA). **STEEG, Evan** [US/CA]; 33 Colborne Street, Kingston, Ontario K7K 1C5 (CA). **ABLESON, Alan, D.** [CA/CA]; 2-249 Queen St., Kingston, Ontario K7L 1B5 (CA).

(74) Agent: **WILKES, Robert, H.**; 72 Albany Ave., Toronto, Ontario M5R 3C3 (CA).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,

[Continued on next page]

(54) Title: DETERMINATION OF CO-OCCURENCES OF ATTRIBUTES



(57) Abstract: A method, system, computer program selecting attribute sets of characterizing attributes of an object, selecting an attribute set of attributes of interest, assigning a likelihood for each characterized attribute set that the attribute set occurs when the attribute set of interest occurs (each likelihood determined using Bayesian computable classifiers on a dataset of attributes for actual samples), comparing each assigned likelihood against likelihood thresholds, and reporting the assigned likelihoods of the characterizing attribute set based on the likelihood thresholds. Markers may be identified for diagnosis and prognosis. Characterizing attributes may be gene expression levels and the attribute of interest may be drug sensitivity level, drug dose (absolute concentration or dose relative to some standard dose), dose of drug which causes half-maximal cellular growth rate, or logarithm base 10 (dose) where dose is the dose which yields half-maximal total cell mass accumulating.

WO 02/095650 A2



SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

- (84) Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
- of inventorship (Rule 4.17(iv)) for US only

Published:

- without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Determination of Co-Occurrences of Attributes

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority from United States patent application serial no.
5 60/291,928 filed May 21, 2001 by the same inventors under the same title, and from
United States patent application serial no. 60/291,931 filed May 21, 2001 by the same
inventors under the title Methods of Gene Analysis and Treating Cancer. United States
patent application serial nos. 60/291,928 and 60/291,931 are hereby incorporated herein
by reference.

10 TECHNICAL FIELD

The invention relates to methods and apparatuses for determining co-occurrences of
attributes in objects. It also relates to attributes including biological response.

BACKGROUND ART

The discovery of correlations among pairs or k-tuples of variables has applications in
15 many areas of science, medicine, industry and commerce. For example, it is of great
interest to physicians and public health professionals to know which lifestyle, dietary,
and environmental factors correlate with each other and with particular diseases in a
database of patient histories. It is potentially profitable for a trader in stocks or
commodities to discover a set of financial instruments whose prices covary over time.
20 Sales staff in a supermarket chain or mail-order distributor would be interested in
knowing that consumers who buy product A also tend to buy products B and Q and this
can be discovered in a database of sales records. Computational molecular biologists and
drug discovery researchers would like to infer aspects of molecular structure from
correlations between distant sequence elements in aligned sets of RNA or protein
25 sequences.

One formulation of the general problem which encompasses many diverse applications,
and which facilitates understanding of the principles described herein is a matrix of
discrete features in which rows correspond to "objects" (such as diseases, individual
patients, stock prices, consumers, or protein sequences) and the columns correspond to
30 features, or attributes, or variables (such as drug sensitivity, gene expression, lifestyle
factors, stocks, sales items, or amino acid residue positions).

Given the vast amount of data and the valuable nature of the information available from large datasets, one wants to use efficient techniques to assist in the determination of correlations. For example, large-scale datasets exists of DNA microarray studies. These can be used to determine correlations between gene-expression patterns and drug
5 treatments. This approach is urgently needed for the treatment of many diseases and other conditions, for example cancer which involves many different tissues and varieties of tumor types. However, the application of the proper data analysis methods will be critical for the efficient use of these large-scale data sets.

Biologists are generally acquainted with the idea of correlating individual genes with
10 specific physiological functions, and with the use of linear correlation methods, such as Pearson's correlation coefficient. Although the linear, single-gene approach has yielded significant advances in biomedicine, the complex, nonlinear nature of tissue demands the use of more sophisticated methods.

It is desirable to provide efficient means by which to determine correlations between
15 attributes of objects.

DISCLOSURE OF THE INVENTION

In a first aspect of the invention provides, a base method for identifying one or more characterizing attributes for an object that are likely to co-occur with one or more attributes of interest for the object. The method comprises the steps of selecting one or
20 more attribute sets of one or more characterizing attributes of the object, selecting an attribute set of one or more attributes of interest for the object, assigning a likelihood for each characterized attribute set that the attribute set occurs for the object when the attribute set of interest occurs for the object (each likelihood determined using one or more Bayesian computable classifiers on a dataset of attributes for a plurality of actual
25 samples of the object), comparing each assigned likelihood against one or more likelihood thresholds, and reporting the assigned likelihoods of the characterizing attribute set based on the likelihood thresholds.

In another aspect the invention provides, a method comprising the steps of, selecting one characterizing attribute set of one or more attributes for the object, selecting an attribute
30 of interest for the object, assigning a likelihood for the characterized attribute set that the attribute occurs for the object when the attribute of interest occurs for the object (the assigned likelihood determined using a Bayesian computable classifier on a dataset of

attributes for a plurality of actual samples of the object), comparing the assigned likelihood against a likelihood threshold, and reporting the assigned likelihood of the characterizing attribute set based on the likelihood threshold.

5 In another aspect the invention provides, a method comprising the steps of, selecting one or more attribute sets of one or more characterizing attributes of the object, selecting an attribute set of one or more attributes of interest for the object, assigning a likelihood for each characterized attribute set that the attribute set occurs for the object when the attribute set of interest occurs for the object (each likelihood determined using one or more Bayesian computable classifiers on a dataset of attributes for a plurality of actual
10 samples of the object), determining a likelihood significance for each assigned likelihood using artificial samples, and ranking the assigned likelihoods of the characterizing attribute set using the likelihood significance.

In another aspect the invention provides, a method comprising the steps of accessing one of the systems described below.

15 In another aspect the invention provides, a base system used to identify one or more characterizing attributes for an object that are likely to co-occur with one or more attributes of interest for the object using a dataset of samples of attributes for the object. The system comprises a computing platform, and a computer program on a computer readable medium for use on the computer platform in association with the dataset. The
20 computer program comprises instructions to identify a characterizing attribute for an object that is likely to co-occur with an attribute of interest for the object, by carrying out the steps of one of the base methods.

The methods may be used for drug discovery by identifying characterizing attribute sets for interaction by the drug using the steps one of the base methods for drug sensitive
25 attributes of interest drug, and performing screens for drugs where growth in cells having desirably ranked characterizing attribute sets is drug sensitive.

The methods may be used for identifying markers for diagnostic kits used to determine if a treatment is appropriate for a patient, by identifying a gene expression level set to be tested for in the patient by carrying out the steps of one of the base methods.

30 The methods may be used for identifying markers for diagnosis of a living system by identifying an attribute set to be tested for in the living system using the steps of one of

the base methods. The methods may also be used for identifying markers for prognosis of a living system by identifying an attribute set to be tested for in the living system using the steps of one of the base methods. The diagnosis or prognosis may be with respect to a disease or syndrome type of a patient. The methods may also be used for identifying markers for determining the appropriateness of a therapy or treatment of a living system by identifying an attribute set to be tested for in the living system using the steps of one of the base methods.

In the above methods the attributes of the attribute set may include protein concentrations. The protein concentrations may include tissue protein concentrations. The protein concentrations may include serum protein concentrations.

In the above methods the attributes of the attribute set may include molecular markers. The molecular markers may include blood molecular markers. The molecular markers may include tissue molecular markers.

In the above methods the attributes of the attribute set may include clinical observables. The clinical observables may include microscopic clinical observables. The clinical observables may include macroscopic clinical observables.

The markers may be for diagnostic kits used in the diagnosis, for diagnostic procedures used in the diagnosis, for prognostic kits used in the prognosis, or for prognostic procedures used in the prognosis.

A likelihood threshold for each characterizing attribute set may be determined using the same Bayesian classifiers as the assigned likelihood on a dataset of attributes for a plurality of artificial samples of the object. Similarly, a likelihood threshold for each characterizing attribute set may be determined by computing those characterizing attribute sets with an assigned likelihood above a given percentile of all assigned likelihoods for the relevant attribute set.

Artificial samples may be created by randomizing the actual gene expression levels for the characterizing attributes. Artificial samples may be created by transposing the actual gene expression levels for each characterizing attribute to another characterizing attribute.

The assigned likelihoods of the characterizing attribute sets may be compared against a likelihood threshold determined by computing those characterizing attribute sets with an

assigned likelihood above a given percentile of all assigned likelihoods for the relevant attribute set of interest.

The characterizing attributes may be gene expression levels and the attribute of interest may be drug sensitivity level, drug dose (absolute concentration or dose relative to some standard dose) along an increasing or decreasing scale, dose of drug which causes half-maximal cellular growth rate, or $-\log_{10}(\text{dose})$ where dose is the dose which yields half-maximal total cell mass accumulating under otherwise standard conditions.

Drug sensitivity level may represent growth inhibiting in diseased cells, a lack of growth inhibiting in diseased cells, patient toxicity in healthy cells. The attributes may be represented in a dataset taken from the NCI60 dataset. The Bayesian classifier may be selected from a group consisting of linear discriminant analysis, quadratic discriminant analysis, and a uniform/gaussian analysis.

The characterizing attribute sets ranked following comparison of the likelihood and the likelihood threshold may be reported. The ranked characterizing attributes sets may be reported to one of a group consisting of a computer readable file stored on computer readable media, a printed report, and a computer network. The assigned likelihoods may be ranked by assigned likelihood and subranked by likelihood significance. The assigned likelihood may be compared against a likelihood threshold, and the assigned likelihood of the characterizing attribute set may be reported based on the likelihood threshold and the ranking of the assigned likelihood.

BRIEF DESCRIPTION OF THE DRAWINGS

For a better understanding of the present invention and to show more clearly how it may be carried into effect, reference will now be made, by way of example, to the accompanying drawings that show the preferred embodiment of the present invention and in which:

FIG. 1 is a first Venn diagram of statistically significant results of analyses employed in the preferred embodiment of the invention;

FIG. 2 is a second Venn diagram of statistically significant results of analyses employed in the preferred embodiment of the invention;

FIG. 3 is a plot of results from a 2D QDA analysis of a dataset according to the preferred embodiment of the invention;

FIG. 4 is a plot of results from a 2D LDA analysis of a dataset according to the preferred embodiment of the invention;

- 5 FIG. 5 is a plot of results from a 2D QDA analysis of a dataset according to the preferred embodiment of the invention;

FIG. 6 is a plot of results from a 2D UGDA analysis of a dataset according to the preferred embodiment of the invention;

- 10 FIG. 7 is a plot of results from a 1D LDA analysis of a dataset according to the preferred embodiment of the invention;

FIG. 8 is a plot of results from a 1D UGDA analysis of a dataset according to the preferred embodiment of the invention;

FIG. 9 is an example flow chart of a computer program according to the preferred embodiment of the invention;

- 15 FIG. 10 is an example block diagram of a system according to the preferred embodiment of the invention;

FIG. 11 is an example flow chart of a computer program according to an alternate embodiment of the invention;

- 20 FIG. 12 is an example block diagram of a system according to an alternate embodiment of the invention;

FIG. 13 is an example flow chart of a computer program according to an alternate embodiment of the invention;

FIG. 14 is an example block diagram of a system according to an alternate embodiment of the invention;

- 25 FIG. 15 is an example flow chart of a computer program according to an alternate embodiment of the invention; and

FIG. 16 is an example block diagram of a system according to an alternate embodiment of the invention.

MODES FOR CARRYING OUT THE INVENTION

A number of alternative base methods, systems and devices will now be referred described, along with alternative applications for those methods, systems and devices. It is understood that these base methods, systems and devices and their alternative
5 applications are by way of description of preferred embodiments and are not limiting to the principles described and the application of those principles.

As previously set out, a base method identifies one or more characterizing attributes for an object that are likely to co-occur with one or more attributes of interest for the object. The method comprises the steps of selecting one or more attribute sets of one or more
10 characterizing attributes of the object, selecting an attribute set of one or more attributes of interest for the object, assigning a likelihood for each characterized attribute set that the attribute set occurs for the object when the attribute set of interest occurs for the object (each likelihood determined using one or more Bayesian computable classifiers on a dataset of attributes for a plurality of actual samples of the object), comparing each
15 assigned likelihood against one or more likelihood thresholds, and reporting the assigned likelihoods of the characterizing attribute set based on the likelihood thresholds.

In an alternative base method, the method comprises the steps of, selecting one characterizing attribute set of one or more attributes for the object, selecting an attribute of interest for the object, assigning a likelihood for the characterized attribute set that the
20 attribute occurs for the object when the attribute of interest occurs for the object (the assigned likelihood determined using a Bayesian computable classifier on a dataset of attributes for a plurality of actual samples of the object), comparing the assigned likelihood against a likelihood threshold, and

Reporting the assigned likelihood of the characterizing attribute set based on the
25 likelihood threshold.

In a further alternative base method, the method comprises the steps of, selecting one or more attribute sets of one or more characterizing attributes of the object, selecting an attribute set of one or more attributes of interest for the object, assigning a likelihood for each characterized attribute set that the attribute set occurs for the object when the
30 attribute set of interest occurs for the object (each likelihood determined using one or more Bayesian computable classifiers on a dataset of attributes for a plurality of actual samples of the object), determining a likelihood significance for each assigned

likelihood using artificial samples, and ranking the assigned likelihoods of the characterizing attribute set using the likelihood significance.

In a further alternative base method, the method comprises the steps of accessing one of the systems described below.

- 5 As previously set out a base system is used to identify one or more characterizing attributes for an object that are likely to co-occur with one or more attributes of interest for the object using a dataset of samples of attributes for the object. The system comprises a computing platform, and a computer program on a computer readable medium for use on the computer platform in association with the dataset. The computer
10 program comprises instructions to identify a characterizing attribute for an object that is likely to co-occur with an attribute of interest for the object, by carrying out the steps of one of the base methods.

- The base methods can be used for drug discovery by identifying characterizing attribute sets for interaction by the drug using the steps one of the base methods for drug sensitive
15 attributes of interest drug, and performing screens for drugs where growth in cells having desirably ranked characterizing attribute sets is drug sensitive.

- The base methods can be used for identifying markers for diagnostic kits used to determine if a treatment is appropriate for a patient, by identifying a gene expression level set to be tested for in the patient by carrying out the steps of one of the base
20 methods.

- In the base methods, a likelihood threshold for each characterizing attribute set can be determined using the same Bayesian classifiers as the assigned likelihood on a dataset of attributes for a plurality of artificial samples of the object. Similarly, a likelihood threshold for each characterizing attribute set can be determined by computing those
25 characterizing attribute sets with an assigned likelihood above a given percentile of all assigned likelihoods for the relevant attribute set.

- Artificial samples can be created by randomizing the actual gene expression levels for the characterizing attributes. Artificial samples can be created by transposing the actual gene expression levels for each characterizing attribute to another characterizing
30 attribute.

The assigned likelihoods of the characterizing attribute sets may be compared against a likelihood threshold determined by computing those characterizing attribute sets with an assigned likelihood above a given percentile of all assigned likelihoods for the relevant attribute set of interest.

- 5 For the base methods, the characterizing attributes may be gene expression levels and the attribute of interest may be drug sensitivity level, drug dose (absolute concentration or dose relative to some standard dose) along an increasing or decreasing scale, dose of drug which causes half-maximal cellular growth rate, or $-\log_{10}(\text{dose})$ where dose is the dose which yields half-maximal total cell mass accumulating under otherwise
10 standard conditions.

Drug sensitivity level may represent growth inhibiting in diseased cells, a lack of growth inhibiting in diseased cells, patient toxicity in healthy cells. The attributes may be represented in a dataset taken from the NCI60 dataset. The Bayesian classifier may be selected from a group consisting of linear discriminant analysis, quadratic discriminant
15 analysis, and a uniform/gaussian analysis.

The characterizing attribute sets ranked following comparison of the likelihood and the likelihood threshold may be reported. The ranked characterizing attributes sets may be reported to one of a group consisting of a computer readable file stored on computer readable media, a printed report, and a computer network. The assigned likelihoods
20 may be ranked by assigned likelihood and subranked by likelihood significance. The assigned likelihood may be compared against a likelihood threshold, and the assigned likelihood of the characterizing attribute set may be reported based on the likelihood threshold and the ranking of the assigned likelihood.

The modes described herein provide extensions and alternatives to the base methods
25 described above and employ many similar principles. The principles of one application as described herein may be applied to the others as appropriate. Thus, the description of all elements of each application will not always be repeated for all applications.

In the preferred embodiment it is preferred for simplicity of programming and interpretation to consider the object and attributes in the form of a matrix, see for
30 example Table 1; however, this is not strictly required and any of the embodiments can

utilize a data set of objects and attributes that are not represented in the form of a matrix by sampling the data set directly.

Table 1

Sample	Object	Attributes
1	A	I d e f
2	B	II d g h
3	A	I d h

- 5 As an example of a dataset laid out in matrix format, the objects may be a particular disease, while the samples are taken from different patients and the attributes are particular expression levels of particular genes and sensitivity to a particular drug. The samples may be cells. Using the data in Table 1, sample 1 from a cell having disease A is taken from a first patient. The disease A cell from the patient has sensitivity to drug I and gene expression levels d, e, f. Similarly, sample 2 from a cell having disease B may also be taken from the same patient. The disease B cell from the patient has sensitivity to drug II and gene expression levels d, g, h. Sample 3 from a cell having disease A is taken from a different patient. The disease A cell from the patient has sensitivity to drug I and gene expression levels d, h.
- 10
- 15 For the example set out above, we may be interested in whether or not sensitivity to drug I is related somehow to gene expressions levels d and e together. Thus, drug I is an attribute set of interest and gene expression levels d and e are a characterizing attribute set. This may be represented in a matrix in the form of Table 2.

Table 2

Sample	Object	Attribute set of Interest I	Characterizing Attribute set d e
1	A	yes	yes
2	B	no	no
3	A	yes	no

Alternatively, object A and object B may be part of a generic object C. For example, one may be interested in knowing if a number of forms of cancer are sensitive to the same drug. In this case, the relevant samples may change. In the example above, the

5 first patient has two forms of cancer A and B. If one is looking for drug sensitivity in both cancers A and B then the all the samples may be relevant, while the object is cancers of type A and B. This permits the use of samples from the same patient for different cancers. Samples from the same patient with the same attribute of interest would ordinarily be considered to be only one sample. The particular definition of

10 objects, samples, attributes of interest and characterizing attributes is a matter of choice for the designer of a particular embodiment. It is recognized that some choices may be superior to others; however, that does not bring them any of them outside of the principles described herein.

The datasets may contain many different samples, some of which will not contain

15 attribute sets of interest for a given run of the methods. These can be filtered out before the methods are run, or they may be left in the dataset to be accessed when the methods are run.

Each of the features for an object may be numerical or qualitative. The features are transformed into ordinal (values capable of being ordered) variables, termed attributes.

20 The principles described herein can be extended to attributes sets of interest and characterizing sets of higher orders. For example, one may want to know if sensitivity to a particular cocktail of drugs co-occurs with a particular combination of gene expression levels.

In this description, specific reference is made on many occasions to examples in the

25 biotech industry. This is in no way limiting to the broad nature of the principles described herein which may be applied to many industry including, by way of example only, financial services, drug discovery, discovery and analysis of genetic networks, sales analysis, direct mail and related marketing activities, clustering customer data, analysis of medical, epidemiological and public health databases, patient data, causes of

30 failures and the analysis of complex systems.

When using the phrases “occurs for” and “attributes for” in respect of an object, it is understood that these are broadly intended. Attributes may not simply be a part of an object, such as its gene expression levels, but may be factors or things that could broadly be related to the object, such as weather on a particular day (attribute) may be related to the price (attribute) of an agricultural stock (object). It is also understood that objects are not limited to traditionally tangible objects, but may be intangible objects such as bonds or stocks as well.

It is recognized that a characterizing attribute set that is likely to co-occur with an attribute set of interest does not necessarily imply that the characterizing attribute set is causing the attribute of interest; however, in many situations this information continues to be useful. For example, symptoms (characterizing attributes) may act as a useful disease marker (attribute of interest); however, they are caused by, and do not generally cause, the disease.

The methods can form part of methods for identifying possible drug targets. Once it is known that a disease or diseased cell is affected by drugs that appear to interact with cells having particular combinations of gene expression levels then screening studies can be conducted to find other drugs that also inhibit growth in cells with those combinations of expression levels.

The base method takes a dataset of samples of objects, including a characterizing attributes set and an attribute set of interest, as input. The method generates an output display of characterizing attribute sets that have a substantial likelihood of co-occurring with the attribute set of interest.

As part of the method, one or more characterizing attribute sets are selected, and one or more attribute sets of interest are selected. The likelihood of each characterizing attribute set co-occurring in actual samples of the object is determined using a Bayesian computable classifier. A likelihood of each characterizing set occurring in artificial samples is used to determine a likelihood threshold. Only those characterizing attribute sets with a likelihood co-occurrence greater than its likelihood threshold is selected.

For example, an embodiment of the method may take a collection of biological samples, their gene expression measurements (characterizing attributes), and a binary high/low drug response measurement (attributes of interest) as input. The method generates a prioritized list of genes, ranked by their p-values or ability to correctly predict the drug

response (likelihood of co-occurrence). In this example, the method consists of three steps:

- 1) Selection of candidate gene sets (characterizing attribute set).
- 2) Calculation of classification accuracy for each gene set using a Bayesian classifier
- 5 (determination of likelihood of co-occurrence using Bayesian classifier)
- 3) Ranking of the gene sets by their classification accuracy and the identification of meaningful gene sets by a comparison of their classification accuracies with those generated using randomized data (determination of likelihood threshold using artificial samples and selection of characterizing attribute sets having a substantial likelihood of
- 10 co-occurrence).

Step 1) can take a number of forms. A simple list of all single genes can be a collection of (singleton) gene sets. A list of all pairs of genes can be a collection of (gene pair) candidate gene sets. Pre-processing techniques (such as those described in PCT Patent

15 Application PCT/CA98/00273 filed March 23 1998 under title Coincidence Detection Method, Products and Apparatus, inventor Evan W. Steeg, published October 1 1998 as WO 98/43182) may be used to create candidate gene sets. Alternative pre-processing techniques may be used, including by way of example, standard feature detectors, or known gene pathway tables.

20

Step 2) can also take a number of forms. Classical statistical techniques such as Linear Discriminant Analysis or Quadratic Discriminant Analysis can be used. Other probabilistic models, such as the Gaussian/Uniform, can be tailored to particular applications or to suit biological intuition.

25

Step 3) involves the comparison of the classification scores from step 2) to those generated from randomized data. Multiple datasets (on the order of 100 or more) are generated by permuting the gene expression values over the samples. i.e. if samples were rows and genes were columns in a table, we would permute the entries in each

30 column, independently. Steps 1) and 2) are repeated for the randomized data, and the scores from the real data are compared to the scores from the randomized data. The scores are ranked according to those most likely to indicate a co-occurrence and those scores greater than the scores for randomized data. Selections can

be made according to the rank of the scores for the non-randomized data, or according to the rank of the difference of the scores for the real and randomized data. Selections may also be based on other calculations using the real and random scores.

- 5 By way of example, validation can be determined either by comparing classification scores from the real data to all the classification scores from the randomized data and then applying the Bonferroni correction, or by comparing the most extreme classification accuracies from each randomized trial to the most extreme classification accuracy from the real data. An empirical p-value can be obtained directly by calculating the
- 10 proportion of random datasets for which their extreme classification accuracies exceeded that in the real data. Only those gene sets with p-values below a user-selected cutoff are reported.

The results of the method described above have many uses including, by way of

15 example, to use the:

- 1) gene sets identified as potential targets for drug interaction.
- 2) gene sets identified for pre-treatment screening of patients to identify the most effective drug treatment.

- 20 We analyzed data on the responses of 60 human cancer cell lines (NCI60) to 90 drugs shown to inhibit their growth in culture (Developmental Therapeutics Program, National Cancer Institute). These data were correlated with the basal (untreated) gene expression patterns from the same set of cell lines (see Ross, D. T., Scherf, U., Eisen, M. B., Perou, C. M., Rees, C., et al. (2000) Systematic variation in gene expression patterns in human
- 25 cancer cell lines. Nature 24, 227-235, and Scherf, U., Ross, D. T., Waltham, W., Smith, L. H., Lee, J. K., et al. (2000) A gene expression database for the molecular pharmacology of cancer. Nature 24, 236-244).

- We compared linear and nonlinear methods for correlating gene expression levels of individual genes with drug sensitivity for 1000 genes across the 60 cancer cell lines,
- 30 which included breast, central nervous system, colon, lung, renal, and prostate cancer, as well as melanoma and leukemia cell lines. In addition, we correlated the expression patterns of *pairs* of genes with drug sensitivities to determine whether more than one gene was required to predict drug sensitivity in some cases.

We found that linear and non-linear methods captured different, although to some extent overlapping, correlations, suggesting specific genes as markers for particular drug treatments. We also found that expression levels of *combinations* of genes should be considered as indicators of effective drug treatments, as these combinations sometimes
 5 contain information not found in the expression patterns of individual genes considered in isolation.

We conclude that nonlinear and combinatorial, as well as linear, single-gene methods are appropriate for the efficient extraction of gene expression-drug sensitivity relationships in cancer cell lines. Computational methods such as these should be useful in cancer
 10 diagnosis and treatment.

First, we divided drug sensitivity into low- and high-sensitivity classes (creating possible attributes of interest):

Drug sensitivities were reported as $-\log_{10} \text{GI}_{50}$ s, with the log being base 10. All the drug sensitivities were normalized to mean zero so that the measurement really reflected
 15 differential growth inhibition. We wanted to categorize the cell line response into “uninhibited” and “inhibited”, with a small gray area to avoid the effects of harsh cutoffs. In that scale, a value of 1.0 for a cell line/drug combination meant that the cell line was inhibited to 50% growth at 1/10 the dosage of the “average” drug. For our purposes, we wanted to identify those drugs that were effective at least 1/5 the “average”
 20 dosage, which in the log scale turns into 0.7. Thus, any value of $-\log_{10} \text{GI}_{50}$ less than 0.7 were considered “uninhibited” or a low sensitivity/response. On the other end of the scale, all of those drugs that resulted in inhibition at concentrations $< 1/10$ of the average dosage were all considered “inhibitory”. We then put in a smooth linear scaling between the cutoffs of 0.7 (low response) and 1.0 (high response). This gave us the function:

$$\begin{aligned}
 25 \quad f(r) = & \begin{cases} 0 & \text{if } r < 0.7 \\ (r-0.7)/0.3 & \text{if } r \text{ in } [0.7, 1) \\ 1 & \text{if } r \geq 1 \end{cases}
 \end{aligned}$$

Sensitivities in the range $[0.7, 1]$ are partially in both classes. Since it varies between 0 and 1, the function f can be viewed as a fuzzy classification or a probability. $f(r) =$
 30 Probability of sensitivity in high class, $1-f(r) =$ Probability of sensitivity in low class.

Finding correlations (determining likelihood of co-occurrence of attribute set of interest and characterizing attribute set) between drug sensitivity (attribute set of interest) and gene expression (characterizing attribute set):

For a given gene, A, and drug, B, we try to see if 2 classes of cell lines (high and low sensitivity) can be distinguished on the basis of gene expression. One of the methods for finding correlations was a slightly modified version of LDA (slightly modified to account for partial class membership). LDA consists of the following steps:

Fit a gaussian G_h to the gene expressions in the high sensitivity class Ch and a gaussian G_l to gene expressions in the low sensitivity class Cl , where $|Ch|$ is the number of cell lines in the high sensitivity class, and $|Cl|$ is the number of cell lines in the low sensitivity class.

Let $Lexpr$ = expression of gene A in cell line L, $Lsensitivity$ = sensitivity of cell line L to drug B

The mean of G_l is calculated as

sum from cell line $L = 1$ to $|Ch|$ of $(Lsensitivity * Lexpr) / (\text{sum of sensitivities in } Ch)$

Mean and variance of G_l were calculated in a similar way.

Pooled variance of G_h and G_l was calculated

avg. variance = $(Ch \text{ variance} * \text{sum } Ch \text{ sensitivities} + Cl \text{ variance} * \text{sum } Cl \text{ sensitivities}) / (\text{num cell lines} - 2 - 1)$

We calculated the probability of a cell line, L, having high sensitivity as follows

$P(L \text{ in } Ch | Lexpr) = G_h(Lexpr) * P(Ch) / (G_h(Lexpr) * P(Ch) + (G_l(Lexpr) * P(Cl)))$

above is Equation 1

The error for this probability was calculated as

$e = Lsensitivity - P(L \text{ in } Ch | Lexpr)$

Testing predictions:

For a given gene and drug we used cross-validation to test prediction of sensitivity from gene expression. Using 59 cell lines we determined gaussians G_h and G_l for the two sensitivity classes. We predicted the sensitivity class of the 60th cell line L, from its

gene expression, using the Equation 1 above. We repeated this procedure for all of the 60 cell lines and calculated a mean squared error for all of the predictions. $e = \sum_{L=1}^{60} [P(L \text{ in Ch} | L_{\text{expr}}) - L_{\text{sensitivity}}]^2 / 60$.

Searching for all correlations:

- 5 We applied the above method to all pairs of genes and drugs [1000 genes] x [90 drugs]

Using other methods:

1D discriminants

- 10 we also used 2 other methods similar to LDA, to search for correlations between sensitivity and gene expression

QDA - differs from LDA in that the original variances of Gh and Gl are used in Equation 1, instead of the average of the variances as a result, QDA can have nonlinear decision boundaries between classes while LDA has linear decision boundaries.

- 15 uniform/gaussian discriminant - similar to LDA except uses uniform distribution for the low class instead of a gaussian distribution, the assumption behind these distributions is that a specific mechanism is responsible for high sensitivity (the gaussian distribution), while various mechanisms lead to low sensitivity (uniform distribution), the height of the uniform is calculated as $1/(\max(\text{expr}) - \min(\text{expr}))$

- 20 2D discriminants

- The three methods above were extended to look for correlations between pairs of genes and drug sensitivities. For a given pair of genes, the joint distribution of gene expression values was represented by gaussians and uniform distributions. A search for correlations was conducted over all pairs of genes and all drugs. For each drug, the three methods
- 25 were applied to about 1/2 million (gene, gene, drug) triples.

Calculating statistical significance (a likelihood threshold):

The statistical significance of MSE scores was determined by comparing against results from randomized data. Statistical significance was adjusted by the Bonferroni method to account for multiple tests. (i.e. for a given drug the statistical significance of a score

from a 1D discriminant was multiplied by 1000; statistical significance of scores from 2D discriminants was multiplied by 10^5).

- To determine whether linear and nonlinear methods could capture different sets of gene expression-drug sensitivity correlations, we employed linear discriminant analysis (LDA) and two nonlinear methods, quadratic discriminant analysis (QDA) and a Bayesian model (a uniform/Gaussian discriminant). Results are shown in Table 3 below.

Table 3

	Drugs P<= 0.01		Drugs P<=0.1		Genes P<=0.01		Genes P<=0.1	
LDA-1D	8	(40%)	29	(53%)	14	(24%)	43	(18%)
QDA-1D	4	(20%)	24	(44%)	5	(8%)	29	(12%)
Bayes mixture 1D	5	(25%)	25	(45%)	6	(10%)	34	(14%)
All 1 D methods	13	(65%)	43	(78%)	20	(34%)	73	(31%)
LDA-2D	9	(45%)	20	(36%)	24	(41%)	102	(43%)
QDA-2D	7	(35%)	22	(40%)	18	(30%)	84	(35%)
Bayes mixture 2D	4	(20%)	22	(40%)	9	(15%)	90	(38%)
All 2D methods	16	(80%)	41	(74%)	48	(81%)	218	(91%)
Intersection of all methods	0	(0%)	4	(7%)	0	(0%)	1	(0.4%)
Union of all methods	20	(100%)	55	(100%)	59	(100%)	239	(100%)

- 10 Table 3 summarizes linear, nonlinear, 1D, and 2D analyses for 1000 genes, 90 drugs, and 60 cell lines. Shown are the numbers of statistically significant gene-drug associations found at $p \leq 0.01$ and $p \leq 0.1$. For example, the LDA-1D analysis method found that for each of 8 drugs, at least one gene out of a group of 14 was able to

predict high sensitivity at $p \leq 0.01$. For LDA-2D, 24 genes arranged in pairs were able to predict high sensitivity to each of 9 drugs at $p \leq 0.01$.

All three methods identified statistically significant correlations between the expression levels of specific genes and sensitivity to drugs based on GI50 values (drug concentration that inhibits cell growth by 50%). Although there was some overlap between the findings of the different methods, they were generally complementary to one another, as shown by the Venn diagrams of statistically significant results from all analysis methods in Figs. 1 and 2. A degree of overlap occurs between results obtained; however, some of the gene-drug correlations were identified by a single method. As shown in Fig. 1, twenty-six drugs (represented by intersection 1) of the 29 drugs (represented by circle 3) found to be in significant correlations with genes by linear 1D methods (LDA 1D) were also identified by at least one other method in the non-linear and combinatorial methods that identified 52 drugs (represented by circle 5), leaving 3 drugs (represented by the non-intersecting portion 7 of circle 3) that were identified by LDA 1D alone. Similarly, as shown in Fig. 2, five genes (non-intersecting portion 9) out of 43 (circle 11) that were identified by LDA 1D as markers for drug sensitivity were identified by that method alone, while the remaining 38 genes (intersection 13) were identified by at least one of the other methods in addition to LDA 1D out of a total of 234 genes (circle 15) that were identified by the other methods.

Nonlinear methods therefore identify gene-drug associations not found by a linear method. This is the case for both 1-dimensional (1D) analysis involving correlations between a single gene and one drug, and for 2D analysis involving correlations between pairs of genes and one drug (gene, gene, drug triples).

To discover correlations between gene expression levels and drug sensitivities that involve more than a single gene, (i.e., the information that predicts high sensitivity to a drug may be contained in the combination of expression patterns of two genes), we applied 2D discriminants. This involved using the same three methods described above for single genes, except that in this case we searched for significant correlations between pairs of genes and individual drugs, i.e., gene, gene, drug triples. Results for 2D methods are shown in Table 3 and Figs. 1 and 2. The 2D methods discovered correlations that were not identified by the 1D method. It is evident from Figs. 1 and 2 and Table 3 that relying only on single-gene (1D) correlations would have missed a large

proportion of the gene-drug associations, since these required the information contained in pairs of genes; this was the case for all three correlation measures. Overall, the use of our combination of linear, nonlinear, 1D and 2D methods allowed for the discovery of 239 marker genes for high drug sensitivity, while sole reliance on the linear 1D method, 5 LDA 1D, would have yielded only 43 markers, or fewer than 20% of the total.

Each of the six methods identified gene-drug correlations not found by any of the other five methods. LDA 1D yielded only five gene markers not identified by at least one of the other methods. For QDA 1D, 1 gene was found by this method only.

Uniform/gaussian 1D was the most effective of the 1D methods in this respect, yielding 10 9 genes correlated with high sensitivity found by this method only. By contrast, genes peculiar to each 2D method included (in pair combinations) 52 genes for LDA, 32 genes for QDA, and 49 genes for uniform/Gaussian.

An example of the 2D approach is diagrammed in Fig. 3. Expression levels of the gene elongation factor TU are plotted vs. expression levels of the gene SID W 116819 for the 15 60 cell lines, whose sensitivities to fluorodopan varied. The areas mapped out by the Gaussian distributions separate most of the black (filled-in squares) points (highly sensitive) cell lines from the white (open squares) points (low sensitivity) cell lines, placing them in separate regions of the graph. Twelve cell lines with high sensitivity to fluorodopan (black points) had varying levels of expression for both genes 1 and 2.

20 In Fig. 3, for either SID W 116819 or elongation factor TU alone, below zero (–) expression occurs in both high and low sensitivity cell lines; similarly, above zero (+) expression for each gene alone occurs in both high and low sensitivity cell lines.

Therefore, neither gene alone correlates with sensitivity. However, the genes can be used in combination to obtain a correlation between gene expression and high drug 25 sensitivity. Cell lines that are highly sensitive to fluorodopan (black points) tend to have greater than zero expression values for both genes (+ +), or below zero expression values for both genes (– –), while the combinations (+ –) and (– +) tend to occur in cell lines that have low sensitivity to fluorodopan (white points).

(The use of + and – here is an oversimplification to describe the general distribution of 30 black and white points on the graph in Fig. 3.)

Figs. 3 through 6 depict 2D analysis of gene expression-drug sensitivity data for 60 cancer cell lines. Fig. 3 employs QDA analysis. Each point represents a cell line, with

its location specified by the relative expression of two genes (x and y coordinates). The points are coloured by the cell line's response to Fluorodopan. The contours represent points of equal probability as predicted by the methods described herein. In general the areas where black squares tend to be concentrated are areas of predicted high sensitivity.

- 5 The arrows indicate the direction of predicted increasing sensitivity. The outermost contour to the bottom left and top right show the decision surface generated by the two Gaussian distributions: outside the outermost contour are classified as high response and the between the gradients as low response. Expression levels of SID W 116819 alone are uncorrelated with sensitivity because a plus (+) can correspond to either high or low
- 10 sensitivity, and a minus (-) can correspond to either high or low sensitivity; the same is true of elongation factor TU. However, as shown in Table 4 below, when either (+) or (-) co-occurs in both genes, sensitivity is high. When expression levels of SID W 116819 and elongation factor TU have opposite signs, sensitivity is low. We therefore obtain a rule for the correlation of the *pair* of genes with fluorodopan sensitivity.

15

Table 4

SID W 116819	elongation factor TU	Sensitivity
+	+	High
-	-	High
-	+	Low
+	-	Low

Other examples for the 2D methods are shown in Figs. 4, 5 and 6, and their respective Tables 5, 6 and 7 below.

- Referring to Fig. 4, according to LDA 2D method, both SID W 242844 and SID W
- 20 26677 are needed to predict high sensitivity to mitozolamide. For SID W 242844 alone, (+) is associated with low sensitivity only, while (-) can be associated with low or high sensitivity. For SID W 26677, (-) is always associated with low, and (+) can correspond to either high or low sensitivity. However, the combination (- +) corresponds to high sensitivity only, so both genes are needed to establish a correlation with high sensitivity.

Table 5

SID W 242844	SID W 26677	Sensitivity
+	+	Low
-	-	Low
-	+	High
+	-	Low

Referring to Fig. 5, according to QDA 2D method, both SID W 242844 and ZFP36 are needed to predict high sensitivity to mitozolamide. For SID W 242844, (-) can
 5 correspond to either high or low sensitivity, and (+) corresponds to low sensitivity. For ZFP36, (-) corresponds to either high or low, and (+) corresponds only to low sensitivity. However, the combination (- -) corresponds only to high sensitivity, so both genes are needed for the correlation.

Table 6

SID W 242844	ZFP36	Sensitivity
+	+	Low
-	-	High
-	+	Low
+	-	Low

10

Referring to Fig. 6, according to uniform/gaussian 2D, for the high sensitivity cell lines, expression of SID W 242844 tends to be negative (-), while expression of ESTs Chr.1 488132 tends to be positive (+). Both SID W 242844 and human nucleotide binding protein are needed to predict high sensitivity to mitozolamide. For SID W 242844, (+)
 15 is always associated with low sensitivity, and (-) can be associated with either high or low. For ESTs Chr.1 488132, (-) is associated only with low, and (+) can correspond to either high or low. The combination (- +), however, is associated with high, while all

other combinations predict low sensitivity. Therefore, both genes are needed to predict high sensitivity.

Table 7

SID W 242844	ESTs Chr.1 488132	Sensitivity
+	+	Low
-	-	Low
-	+	High
+	-	Low

- 5 Many of the results could not be classified easily as simple plus/minus distributions, but the concept of requiring a particular range of expression value combinations for each pair of genes applies in all cases shown for the 2D methods. In some cases, this range of values includes zero (no deviation in expression from mixed culture control). This is acceptable, since we are interested only in relative basal gene expression levels, not
- 10 perturbed gene expression relative to the control. For example, a combination of approximately zero (0) expression for gene SID 289361 and positive (+) expression for gene SID 327435 correlated with high sensitivity to fluorouracil according to QDA 2D, in one case.

The 1D approach is shown in Figs. 7 and 8. For single gene correlations, only the value

15 on the x-axis (horizontal axis) is considered. A random variable was used to create a y-axis (vertical axis) as a visual aid to avoid the problem of overlapping points. Referring to Fig. 7, according to LDA 1D, cell lines with high sensitivity to mitozolamide exhibited high levels of PTN expression. Referring to Fig. 8, Uniform/gaussian 1D determined that cells with high sensitivity to mitozolamide expressed DOC-2 mitogen-

20 responsive phosphoprotein in a particular range of values above control. Random variable on y-axis permits visualization of data points that would obscure one another in a one-dimensional graph.

In some instances, we found significant correlations between a gene and more than one drug. Generally, the drugs that correlated with a gene were from the same class,

25 however, this was not always the case. Results are shown in previously set out Table 3.

We determined that certain levels of expression for specific genes are consistently associated with high sensitivity to drugs for cancer in 60 human cancer cell lines. Linear analysis methods alone were insufficient to identify many statistically significant correlations between basal gene expression and high sensitivity to drugs. In addition, we have demonstrated the need for 2D methods, as in many cases, combinations of genes contain the information required to establish correlations with drug sensitivity. This suggests that the physiological functions of cancer cells are often governed by the synergistic actions of multiple genes. These results are consistent with the idea that physiological systems are by nature complex, nonlinear systems, and should be analysed as such.

As shown in Table 3 (where Bayes mixture refers to the Uniform/Gaussian), every one of the six example methods, LDA, QDA, and Uniform/Gaussian each for 1D and 2D analyses, identified gene-drug correlations not discovered by any of the other five methods. This is especially true for the 2D methods. A combination of correlation techniques is appropriate for efficient interpretation of DNA microarray data.

The variability of cancer cell types poses two interrelated problems: 1) diagnosis, and 2) choice of treatment. Evidence has been found that the gene expression patterns of breast-derived cancer cell lines reflect those of the normal tissue of origin and of a breast-derived tumor, suggesting that cell lines may be useful in determining the gene expression patterns of in vivo cancer cells. If this is the case, it should be possible to use the results of large-scale studies of gene expression and drug responses in cancer cell lines to create databases of diagnostic markers for various cancers. Linear, nonlinear, and combinatorial analyses could be applied to determine those markers, and to suggest appropriate therapeutic drugs. As we have demonstrated in the present study, the use of nonlinear and combinatorial analyses in addition to linear, single-gene methods, increases the number of gene-drug associations, and therefore should improve the probability of determining appropriate drug therapies.

Markers identified by these computational methods could be used as the basis for diagnostic tests specific for those genes, perhaps in the form of smaller-scale microarray assays. Tests such as these would be aimed directly toward determination of the best choice(s) for therapeutic drug treatment. For example, a diagnostic test indicating high

expression levels for both genes elongation factor TU and SID W 116819 (Fig. 3) would suggest a high probability of a response to fluorodopan treatment.

The present study focused on basal gene expression patterns as indicators of drug sensitivity.

- 5 In carrying out the embodiment described above for the NCI60 dataset, we computationally distinguish strong from weak biological responses (i.e., to discriminate, classify, or predict biological responses). In its details, the method employs computationally-derived associations between computationally-analyzed quantitative gene expression data and computationally-analyzed quantitative intensity data. The
10 intensity data represents observables (other than gene expression) assumed to be related in some arbitrary, but graded, manner to the biological responses.

We used a “*biological response scoring function*,” called f , where $f:U \rightarrow R^1 \subset [0,1]$, and U is a 1-parameter continuous path in R^m , $m > 1$. f is constructed to represent biological response on a bounded ordinal scale of real numbers, where

- 15 $f = 0$ is interpreted to mean “no or negligible biological response”;
 $f = 1$ is interpreted to mean “very substantial, strong, or high biological response”;
 $0 < f < 1$ is interpreted to mean “biological response somewhere between negligible and substantial in proportion to proximity to 0 or 1, respectively.”

- Formally, the domain U of f is defined to be a 1-parameter continuous path in m -
20 dimensional space. E.g., U can simply be scalar, i.e., $U \subset R^1$; or U can be an arbitrary 1-parameter path through higher-dimensional space R^m , $m > 1$ (e.g., a series of m -dimensional feature vectors indexed by continuous time). Note: The examples provided here concentrate on the scalar domain case (i.e., $U \subset R^1$), but the approach also applies to cases of higher-dimensional continuous 1-parameter paths.

- 25 Domain $U \subset R^1$ is interpreted to mean:

“degree or intensity of external effect on the biology” either on an increasing or decreasing scale.

Examples:

U represents drug dose (absolute concentration or dose relative to some standard dose) along an increasing, or decreasing, scale;

U can represent the dose of drug which causes half-maximal cellular growth rate as charted along a scale which decreases to the right;

- 5 U represents $-\log_{10}(\text{dose})$, where dose is the dose which yields half-maximal total cell mass accumulating in a chemostat under otherwise standard conditions (e.g., let $r \in U$ such that $r = -\log \text{GI50} = -\log_{10}(\text{GI50})$, where GI50 = drug dose which yields 50% of the cellular mass which is achieved under some standard untreated-with-drug conditions.
- 10 Note that in this last example, r increases as GI50 decreases. In this case, an increasing r represents a decreasing “intensity of dose needed to obtain some defined biological effect.”

The function f assigns a readily interpretable numerical “biological response score” in the continuous interval $[0,1]$ to a “degree or intensity of external effect on biology” from a scale $U \subset \mathbb{R}^1$. Thus, f is what inexorably links “intensity of external effect on biology” to a readily interpreted biological response scale, where the interpretations of f values are given in 1a) above.

- 20 Example (continuous piece-wise linear biological scoring function):

$$\text{Let } f(r) = \begin{cases} 0, & r < 0.7 \\ (r - 0.7)/0.3, & r \in [0.7, 1), \text{ where } r = -\log \text{GI50} = -\log_{10}(\text{GI50}). \\ 1, & r \geq 1 \end{cases}$$

Interpretations:

- 25 If the dose required to achieve some biological effect (say, 50% growth inhibition) is small, then score this phenomenon as “strong biological response”, i.e., “cells are very sensitive.” In $f(r)$ terms, if $\text{GI50} \leq 0.1$ (i.e., $-\log(\text{GI50}) \geq 1$), then $f = 1$.

If the dose required to achieve some biological effect (say, 50% growth inhibition) is large, then score this phenomenon as “weak biological response”, i.e., “cells are very insensitive.” In $f(r)$ terms, if $GI50 \geq 0.2$ (i.e., $-\log(GI50) \leq 0.7$), then $f = 0$.

5

If the dose required to achieve some biological effect (say, 50% growth inhibition) is modest or a some gradation between low and high, then score this phenomenon as “mixed-strength biological response”, i.e., “cells are somewhat sensitive and/or somewhat insensitive.” In $f(r)$ terms, if $0.2 \geq GI50 > 0.1$ (i.e., $0.7 \leq -\log(GI50) < 1$), then $f = (r - 0.7)/0.3$.

10

Example (smooth biological scoring function):

$$\text{Let } f_{\text{sigmoid}}(r) = 1 - \left(1 + \left(\frac{r-a}{b-a} \right)^\nu \right)^{-1}, \quad r \geq a, \quad b > a \geq 0, \quad \nu > 1,$$

$$f_{\text{sigmoid}}(r) = 1 - \left(1 + \left(\frac{a-r}{b-a} \right)^\nu \right)^{-1}, \quad r < a, \quad b > a \geq 0, \quad \nu > 1$$

15

where $r = -\log GI50 = -\log_{10}(GI50)$

Let:

i denote, or label, any given external effect, or situation, on the biology, e.g., temperature, pH, therapeutic intervention, compound applied, drug dosed, etc.

20

(For explanatory convenience, for now on we often refer to any external effect on the biology as “drug.”)

j denote any biological source of gene expression data, e.g., patient, tissue, cultured cell line, etc. (For explanatory convenience, for now on we often refer to any biological source of expression data as “cell line.”)

25

k denote, or label, any given gene, mRNA species, gene product, or protein.
(For explanatory convenience, for now on we often refer to any of these entities as “gene.”)

5 g_k^j denote, or label, gene abundance or expression level, however numerically adjusted or normalized, of gene k in cell line j .

α represent, or label, any desired categorical description of biological response score. E.g., α = any of “high”, “strong”, “sensitive/insensitive”, etc. if $f = 1$; e.g. α = any of “low”, “weak”, “insensitive”, etc., if $f = 0$; e.g., α = any of
10 “middle”, “modest”, “mixed sensitive\insensitive”, etc. if $0 < f < 1$.

w represent, or label, generally the biological response score (i.e., f value) of any biological source under any external effect or situation, e.g., the sensitivity of a cell line to a drug.

15 $w^{i,j}$ specifically denote, or label, the biological response score (i.e., f value) of biological source j under any external effect or situation i , e.g., f value of cell line j under some specified exposure to drug i .

20 $w_{\alpha}^{i,j}$ specifically denote, or label, the biological response score (i.e., f value) which falls in some particular category α (e.g., $\alpha = \text{sensitive}$) of biological source j under any external effect or situation i , e.g., $w_{\text{sensitive}}^{i,j}$ means the f value is 1 for cell line j under some specified exposure to drug i .

25 C_{α}^i denote the set of biological sources falling in biological response category α when the biological source is external effect i . E.g., $C_{\text{sensitive}}^i$ is the set comprising cell lines for which the respective f values are 1 when exposed to drug i at some specified dose, i.e., the set of cell lines sensitive to drug i .

$|C_\alpha^i|$ denote the cardinality of C_α^i , i.e., the number of elements in set C_α^i . E.g., $|C_{sensitive}^i| = 23$, means that for the collection of cell lines considered, there are 23 cell lines that are sensitive to drug i .

- 5 For any given external biological effect i (e.g., drug i administered by some specified dosing regime), and for any gene k , ...

Compute α category-wise data-summarizing mathematical, statistical, machine learning-based, data mining-based, or empirical, etc. entities. For example:

10

Compute histogram comprising g_k^j , for given k , for $j \in C_\alpha^i$. E.g., histogram of abundances of gene k from all the cell lines sensitive to drug i .

15

Compute parameters necessary to fit any chosen mathematical density function or continuous curve to a α category-wise histogram of the type described in 3a.1 above. E.g., in preparation for fitting a gaussian distribution to $\{g_k^j\}$, $j \in C_{sensitive}^i$, compute parameters that are the cell line sensitivity-weighted gene k sample

20

$$\begin{aligned} & \text{mean } {}_i\bar{g}_k^{sensitive} \text{ and variance } s^2 {}_i g_k^{sensitive}, \text{ where} \\ & {}_i\bar{g}_k^{sensitive} = w^{i,j} g_k^j / \sum_j w^{i,j}, j \in C_{sensitive}^i \\ & s^2 {}_i g_k^{sensitive} = w^{i,j} (g_k^j - {}_i\bar{g}_k)^2 / \sum_j w^{i,j}, j \in C_{sensitive}^i \end{aligned}$$

Compute α category-wise average data-summarizing parameters. E.g., sensitive\insensitive average variance are, respectively,

25

$$\begin{aligned} \langle s^2 {}_i g_k \rangle &= (s^2 {}_i g_k^{sensitive} \sum_{j'} w^{i,j'} + s^2 {}_i g_k^{insensitive} \sum_{\hat{j}} w^{i,\hat{j}}) / (\sum_{j'} w^{i,j'} + \sum_{\hat{j}} w^{i,\hat{j}}), \\ & \text{where } j' \in C_{sensitive}^i \text{ and } \hat{j} \in C_{insensitive}^i \\ & \sigma_k^{avg} = \text{the square root of the average variance.} \end{aligned}$$

30

For all α categories of interest, compute α category-wise data-summarizing mathematical, statistical, machine learning-based, data mining-based, or empirical, etc. entities based on any of the α category-wise average data-summarizing parameters such as those examples described above. For example:

Compute a gaussian summarizing entity ${}_iG_k^{sensitive}$ for gene k in the cell lines sensitive to drug i , i.e., ${}_iG_k^{sensitive}(g, \mu, \sigma) = (\sigma\sqrt{2\pi})^{-1} \exp(-(g - \mu)^2 / (2\sigma^2))$ where

$$\mu = {}_i\bar{g}_k^{sensitive} \text{ and } \sigma = \sqrt{s^2 {}_i\bar{g}_k^{sensitive}},$$

and compute analogous ${}_iG_k^{insensitive}$.

5

Compute discriminators, classifiers, and predictors of α , the category-wise biological response to external event i , but based on information computed from a given gene k . In these computations, we employ as needed any of the preparatory computations described above. For example:

10

Compute a Bayesian probability $P(j \in C_i^\alpha | g_k^j)$ that a cell line j is in biological response category α due to biological effect i , given the gene k abundance in cell line j , e.g.,

15

$$P(j \in C_i^\alpha | g_k^j) = \frac{{}_iG_k^\alpha(g_k^j) \cdot P(C_i^\alpha)}{\sum_{\alpha} {}_iG_k^\alpha(g_k^j) \cdot P(C_i^\alpha)}$$

${}_iG_k^\alpha(g_k^j)$ = probability of abundance value g_k^j from the gaussian density fitted to the histogram of the gene k abundances over the cell lines in response category α when subjected to biological effect i .

20

A probability difference for the above probability is also computed, e.g.,

$$difference_{Bayesian} = \left| P(j \in C_i^\alpha | g_k^j) - w^{i,j} / \sum_j w^{i,j} \right|, j \in C_i^\alpha.$$

Note: Importantly, $difference_{Bayesian}$ is the difference between ‘the predicted probability that cell line j is in the category α as computed from the gene k abundances across cell lines’ and ‘the observed probability that cell line j is in category α as computed from the effects of biological effect i on the cell lines’.

25

As described below the determination of the likelihood of a co-occurrence was calculated using a number of differing methods, namely:

	Uniform\Gaussian Discriminant Analysis – 1-dimensional	(UGDA 1D)
	Uniform\Gaussian Discriminant Analysis – 2-dimensional	(UGDA 2D)
	Linear Discriminant Analysis – 1-dimensional	(LDA 1D)
	Quadratic Discriminant Analysis – 1-dimensional	(QDA 1D)
5	Linear Discriminant Analysis – 2-dimensional	(LDA 2D)
	Quadratic Discriminant Analysis – 2-dimensional	(QDA 2D)

Uniform\Gaussian Discriminant Analysis – 1-dimensional (UGDA 1D)

10 This method computes a Bayesian *conditional probability* $P(j \in C_i^{sensitive} | g_k^j)$ that a cell line j is sensitive to drug i , given the gene k abundance g_k^j in cell line j .

The probability is computed using the following equation:

$$15 \quad P(j \in C_i^{sensitive} | g_k^j) = \frac{{}_iG_k^{sensitive}(g_k^j) \cdot P(C_i^{sensitive})}{{}_iG_k^{sensitive}(g_k^j) \cdot P(C_i^{sensitive}) + {}_iU_k(g_k^j) \cdot P(C_i^{insensitive})}$$

where

$P(C_i^{sensitive})$ = prior probability of the sensitive set

$$= |C_i^{sensitive}| / (|C_i^{sensitive}| + |C_i^{insensitive}|),$$

$P(C_i^{insensitive})$ = prior probability of the insensitive

$$20 \quad \text{set} = |C_i^{insensitive}| / (|C_i^{sensitive}| + |C_i^{insensitive}|),$$

${}_iG_k^{sensitive}(g_k^j)$ = probability of abundance value g_k^j from the gaussian density fitted to the histogram of the gene k abundances over the sensitive cell lines when subjected to drug i .

$$25 \quad {}_iG_k^{sensitive}(g_k^j) = \frac{1}{\sigma_k^{sen} \sqrt{2\pi}} e^{-(g_k^j - \mu_k^{sen})^2 / 2(\sigma_k^{sen})^2},$$

where

μ_k^{sen} = mean of gene k abundances in the sensitive cell lines, $j \in C_{sensitive}^i$

σ_k^{sen} = standard deviation of gene k abundances in the sensitive cell lines,

$j \in C_{sensitive}^i$

- 5 ${}_iU_k(g_k^j)$ = probability of abundance value g_k^j from the uniform density fitted to the gene k abundances over all cell lines when subjected to drug i . For a given gene k , this value is constant across all cell lines, j , i.e.,

$${}_iU_k(g_k^j) = \frac{1}{\max(g_k) - \min(g_k)}$$

10 where

$\max(g_k)$ = maximum abundance of gene k over all cell lines

$\min(g_k)$ = minimum abundance of gene k over all cell lines

Sample parameters for the UGDA 1D for the NCI60 dataset are:

Rule 1

- 15 Gene: SID W 376472 Homo sapiens clone 24429 mRNA sequence [5':AA041443 3':AA041360]

Drug: Inosine-glycodialdehyde

Parameters:

- 20 $\mu_k^{sen} = -0.4394, \sigma_k^{sen} = 0.4217$
 ${}_iU_k(g_k^j) = 0.2538$
 $P(C_i^{sensitive}) = 0.1978, P(C_i^{insensitive}) = 0.8022$

Rule 2

- Gene: Human clone 23665 mRNA sequence Chr.17 [488020 (IW) 5':AA054745
 25 3':AA054747]

Drug: Dolastatin-10

Parameters:

- 30 $\mu_k^{sen} = -0.7752, \sigma_k^{sen} = 0.3685$
 ${}_iU_k(g_k^j) = 0.2347$
 $P(C_i^{sensitive}) = 0.135, P(C_i^{insensitive}) = 0.865$

Rule 3

Gene: SID W 469272 Epidermal growth factor receptor [5':AA026175 3':AA026089]

Drug: Dichloroallyl-lawsone

5 Parameters:

$$\mu_k^{\text{sen}} = -0.2886, \sigma_k^{\text{sen}} = 0.4416$$

$$iU_k(g_k^j) = 0.2299$$

$$P(C_i^{\text{sensitive}}) = 0.2172, P(C_i^{\text{insensitive}}) = 0.7828$$

10 Rule 4

Gene: ESTs Chr.1 [488132 (IW) 5':AA047420 3':AA047421]

Drug: N-phosphonoacetyl-L-aspartic-ac

Parameters:

$$\mu_k^{\text{sen}} = 0.2863, \sigma_k^{\text{sen}} = 0.3651$$

$$15 \quad iU_k(g_k^j) = 0.241$$

$$P(C_i^{\text{sensitive}}) = 0.2583, P(C_i^{\text{insensitive}}) = 0.7417$$

Rule 5

Gene: LBR Lamin B receptor Chr.1 [307225 (IW) 5':W21468 3':N93426]

20 Drug: Pyrazofurin

Parameters:

$$\mu_k^{\text{sen}} = 0.4077, \sigma_k^{\text{sen}} = 0.4993$$

$$iU_k(g_k^j) = 0.237$$

$$P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

25

Rule 6

Gene: SID W 305455 TRANSCRIPTIONAL REGULATOR ISGF3 GAMMA
SUBUNIT [5':W39053 3':N89796]

Drug: Cyanomorpholinodoxorubicin

30 Parameters:

$$\mu_k^{\text{sen}} = 0.4419, \sigma_k^{\text{sen}} = 0.3503$$

$$iU_k(g_k^j) = 0.2326$$

$$P(C_i^{\text{sensitive}}) = 0.2067, P(C_i^{\text{insensitive}}) = 0.7933$$

Rule 7

Gene: SID 429145 Human nicotinamide N-methyltransferase (NNMT) mRNA complete cds [5': 3':AA004839]

5 Drug: Semustine (MeCCNU)

Parameters:

$$\mu_k^{\text{sen}} = 0.2891, \sigma_k^{\text{sen}} = 0.398$$

$$iU_k(g_k^j) = 0.3155$$

$$P(C_i^{\text{sensitive}}) = 0.1606, P(C_i^{\text{insensitive}}) = 0.8394$$

10

Rule 8

Gene: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: Mitozolamide

15 Parameters:

$$\mu_k^{\text{sen}} = -1.008, \sigma_k^{\text{sen}} = 0.5668$$

$$iU_k(g_k^j) = 0.2381$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

20 Rule 9

Gene: *Homo sapiens lysosomal neuraminidase precursor mRNA complete cds SID W 487887 Hexabrachion (tenascin C cytotactin) [5':AA046543 3':AA045473]

Drug: Mitozolamide

Parameters:

25 $\mu_k^{\text{sen}} = 0.8444, \sigma_k^{\text{sen}} = 0.5358$

$$iU_k(g_k^j) = 0.2597$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 10

30 Gene: ESTs Chr.1 [488132 (IW) 5':AA047420 3':AA047421]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 0.4755, \sigma_k^{\text{sen}} = 0.3355$$

$${}_iU_k(g_k^j) = 0.241$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 11

5 Gene: Human mitogen-responsive phosphoprotein (DOC-2) mRNA complete cds Chr.5
[428137 (IE) 5': 3':AA001933]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 0.3967, \sigma_k^{\text{sen}} = 0.3587$$

10 ${}_iU_k(g_k^j) = 0.2342$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 12

15 Gene: SID W 345420 Homo sapiens YAC clone 136A2 unknown mRNA 3'untranslated
region [5':W76024 3':W72468]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 0.7456, \sigma_k^{\text{sen}} = 0.5579$$

$${}_iU_k(g_k^j) = 0.2625$$

20 $P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$

Rule 13

25 Gene: CDH2 Cadherin 2 N-cadherin (neuronal) Chr. [325182 (DIRW) 5':W48793
3':W49619]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 0.6581, \sigma_k^{\text{sen}} = 0.3744$$

$${}_iU_k(g_k^j) = 0.2564$$

30 $P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$

Rule 14

Gene: SID W 280376 ESTs Highly similar to CELL CYCLE PROTEIN KINASE
CDC5/MSD2 [Saccharomyces cerevisiae] [5':N50317 3':N47107]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 0.7347, \sigma_k^{\text{sen}} = 0.4233$$

$$iU_k(g_k^j) = 0.177$$

$$5 \quad P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 15

Gene: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW)

5':AA039292 3':AA039334]

10 Drug: Cyclodisone

Parameters:

$$\mu_k^{\text{sen}} = 0.6598, \sigma_k^{\text{sen}} = 0.2562$$

$$iU_k(g_k^j) = 0.1672$$

$$P(C_i^{\text{sensitive}}) = 0.1689, P(C_i^{\text{insensitive}}) = 0.8311$$

15

Rule 16

Gene: SID W 345420 Homo sapiens YAC clone 136A2 unknown mRNA 3'untranslated region [5':W76024 3':W72468]

Drug: Clomesone

20 Parameters:

$$\mu_k^{\text{sen}} = 0.7165, \sigma_k^{\text{sen}} = 0.4934$$

$$iU_k(g_k^j) = 0.2625$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

25 Rule 17

Gene: SID 289361 ESTs [5':N99589 3':N92652]

Drug: Fluorouracil (5FU)

Parameters:

$$\mu_k^{\text{sen}} = 0.03614, \sigma_k^{\text{sen}} = 0.186$$

$$30 \quad iU_k(g_k^j) = 0.2252$$

$$P(C_i^{\text{sensitive}}) = 0.1628, P(C_i^{\text{insensitive}}) = 0.8372$$

Rule 18

Gene: SID 43555 MALATE OXIDOREDUCTASE [5':H13370 3':H06037]

Drug: Fluorouracil (5FU)

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.9686, \sigma_k^{\text{sen}} = 0.4053 \\ 5 \quad iU_k(g_k^j) &= 0.241 \\ P(C_i^{\text{sensitive}}) &= 0.1628, P(C_i^{\text{insensitive}}) = 0.8372 \end{aligned}$$

Rule 19

Gene: H.sapiens mRNA for Gal-beta(1-3/1-4)GlcNAc alpha-2.3-sialyltransferase Chr.11

10 [324181 (IW) 5':W47425 3':W47395]

Drug: Fluorouracil (5FU)

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.3532, \sigma_k^{\text{sen}} = 0.2383 \\ iU_k(g_k^j) &= 0.2488 \\ 15 \quad P(C_i^{\text{sensitive}}) &= 0.1628, P(C_i^{\text{insensitive}}) = 0.8372 \end{aligned}$$

Rule 20

Gene: ESTs Moderately similar to ZINC-BINDING PROTEIN A33 [Pleurodeles waltl]
Chr.16 [25718 (RW) 5':R12025 3':R37093]

20 Drug: Fluorodopan

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.542, \sigma_k^{\text{sen}} = 0.2812 \\ iU_k(g_k^j) &= 0.2079 \\ 25 \quad P(C_i^{\text{sensitive}}) &= 0.2061, P(C_i^{\text{insensitive}}) = 0.7939 \end{aligned}$$

Rule 21

Gene: SID 470501 ESTs [5':AA031743 3':AA031652]

Drug: Asaley

Parameters:

$$\begin{aligned} 30 \quad \mu_k^{\text{sen}} &= -0.7867, \sigma_k^{\text{sen}} = 0.4327 \\ iU_k(g_k^j) &= 0.1869 \\ P(C_i^{\text{sensitive}}) &= 0.1878, P(C_i^{\text{insensitive}}) = 0.8122 \end{aligned}$$

Rule 22

Gene: SID 307717 Homo sapiens KIAA0430 mRNA complete cds [5': 3':N92942]

Drug: Cyclocytidine

Parameters:

$$\begin{aligned}
 5 \quad & \mu_k^{\text{sen}} = 0.004825, \sigma_k^{\text{sen}} = 0.232 \\
 & {}_iU_k(g_k^j) = 0.1835 \\
 & P(C_i^{\text{sensitive}}) = 0.2533, P(C_i^{\text{insensitive}}) = 0.7467
 \end{aligned}$$

Rule 23

10 Gene: SID W 122347 ESTs [5':T99193 3':T99194]

Drug: Oxanthrazole (pirovantrone)

Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = -0.9888, \sigma_k^{\text{sen}} = 0.6153 \\
 & {}_iU_k(g_k^j) = 0.2198 \\
 15 \quad & P(C_i^{\text{sensitive}}) = 0.1956, P(C_i^{\text{insensitive}}) = 0.8044
 \end{aligned}$$

Rule 24

Gene: SID W 429290 ESTs [5':AA007457 3':AA007361]

Drug: Oxanthrazole (pirovantrone)

20 Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = 0.6229, \sigma_k^{\text{sen}} = 0.3177 \\
 & {}_iU_k(g_k^j) = 0.2532 \\
 & P(C_i^{\text{sensitive}}) = 0.1956, P(C_i^{\text{insensitive}}) = 0.8044
 \end{aligned}$$

25 Rule 25

Gene: ALDOC Aldolase C fructose-bisphosphate Chr.17 [229961 (IW) 5':H67774 3':H67775]

Drug: Anthrapyrazole-derivative

Parameters:

$$\begin{aligned}
 30 \quad & \mu_k^{\text{sen}} = -0.2373, \sigma_k^{\text{sen}} = 0.3786 \\
 & {}_iU_k(g_k^j) = 0.2049 \\
 & P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994
 \end{aligned}$$

Rule 26

Gene: SID W 381819 Plastin 1 (I isoform) [5':AA059293 3':AA059061]

Drug: Teniposide

Parameters:

$$\begin{aligned}
 5 \quad & \mu_k^{\text{sen}} = 0.05147, \sigma_k^{\text{sen}} = 0.3839 \\
 & {}_iU_k(g_k^j) = 0.2101 \\
 & P(C_i^{\text{sensitive}}) = 0.1894, P(C_i^{\text{insensitive}}) = 0.8106
 \end{aligned}$$

Rule 27

10 Gene: SID W 345683 ESTs Highly similar to INTEGRAL MEMBRANE
GLYCOPROTEIN GP210 PRECURSOR [Rattus norvegicus] [5':W76432 3':W72039]

Drug: Daunorubicin

Parameters:

$$\begin{aligned}
 15 \quad & \mu_k^{\text{sen}} = 0.918, \sigma_k^{\text{sen}} = 0.3704 \\
 & {}_iU_k(g_k^j) = 0.2762 \\
 & P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189
 \end{aligned}$$

Rule 28

Gene: SID 234072 EST Highly similar to RETROVIRUS-RELATED POL

20 POLYPROTEIN [Homo sapiens] [5': 3':H69001]

Drug: Aphidicolin-glycinate

Parameters:

$$\begin{aligned}
 25 \quad & \mu_k^{\text{sen}} = -0.3626, \sigma_k^{\text{sen}} = 0.4252 \\
 & {}_iU_k(g_k^j) = 0.207 \\
 & P(C_i^{\text{sensitive}}) = 0.1994, P(C_i^{\text{insensitive}}) = 0.8006
 \end{aligned}$$

Rule 29

Gene: SID 50243 ESTs [5':H17681 3':H17066]

Drug: CPT,10-OH

Parameters:

$$\begin{aligned}
 30 \quad & \mu_k^{\text{sen}} = 0.8677, \sigma_k^{\text{sen}} = 0.5387 \\
 & {}_iU_k(g_k^j) = 0.2653 \\
 & P(C_i^{\text{sensitive}}) = 0.1856, P(C_i^{\text{insensitive}}) = 0.8144
 \end{aligned}$$

Rule 30

Gene: SID W 346587 Homo sapiens quiescin (Q6) mRNA complete cds [5':W79188
3':W74434]

5 Drug: CPT,10-OH

Parameters:

$$\mu_k^{\text{sen}} = 1.001, \sigma_k^{\text{sen}} = 0.6123$$

$${}_iU_k(g_k^j) = 0.2358$$

$$P(C_i^{\text{sensitive}}) = 0.1856, P(C_i^{\text{insensitive}}) = 0.8144$$

10

Rule 31

Gene: SID W 361023 ESTs [5':AA013072 3':AA012983]

Drug: CPT,10-OH

Parameters:

$$15 \quad \mu_k^{\text{sen}} = -0.8339, \sigma_k^{\text{sen}} = 0.6084$$

$${}_iU_k(g_k^j) = 0.2222$$

$$P(C_i^{\text{sensitive}}) = 0.1856, P(C_i^{\text{insensitive}}) = 0.8144$$

Rule 32

20 Gene: SID W 488148 H.sapiens mRNA for 3'UTR of unknown protein [5':AA057239
3':AA058703]

Drug: CPT

Parameters:

$$25 \quad \mu_k^{\text{sen}} = 0.8224, \sigma_k^{\text{sen}} = 0.5588$$

$${}_iU_k(g_k^j) = 0.2577$$

$$P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

Rule 33

Gene: SID W 159512 Integrin alpha 6 [5':H16046 3':H15934]

30 Drug: CPT

Parameters:

$$\mu_k^{\text{sen}} = 0.7291, \sigma_k^{\text{sen}} = 0.6557$$

$${}_iU_k(g_k^j) = 0.2571$$

$$P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

Rule 34

Gene: SID W 429290 ESTs [5':AA007457 3':AA007361]

5 Drug: CPT

Parameters:

$$\mu_k^{\text{sen}} = 0.7084, \sigma_k^{\text{sen}} = 0.4576$$

$${}_iU_k(g_k^j) = 0.2532$$

$$P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

10

Rule 35

Gene: ESTs Chr.5 [487396 (IW) 5':AA046573 3':AA046660]

Drug: CPT

Parameters:

$$15 \quad \mu_k^{\text{sen}} = 0.6068, \sigma_k^{\text{sen}} = 0.3836$$

$${}_iU_k(g_k^j) = 0.1848$$

$$P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

Rule 36

20 Gene: SID W 361023 ESTs [5':AA013072 3':AA012983]

Drug: CPT,20-ester (S)

Parameters:

$$\mu_k^{\text{sen}} = -0.6333, \sigma_k^{\text{sen}} = 0.554$$

$${}_iU_k(g_k^j) = 0.2222$$

$$25 \quad P(C_i^{\text{sensitive}}) = 0.255, P(C_i^{\text{insensitive}}) = 0.745$$

Rule 37

Gene: SID W 125268 H.sapiens mRNA for human giant larvae homolog [5':R05862 3':R05776]

30 Drug: CPT,20-ester (S)

Parameters:

$$\mu_k^{\text{sen}} = -0.4871, \sigma_k^{\text{sen}} = 0.5365$$

$${}_iU_k(g_k^j) = 0.266$$

$$P(C_i^{\text{sensitive}}) = 0.2844, P(C_i^{\text{insensitive}}) = 0.7156$$

Rule 38

Gene: SID W 361023 ESTs [5':AA013072 3':AA012983]

5 Drug: CPT,20-ester (S)

Parameters:

$$\mu_k^{\text{sen}} = -0.608, \sigma_k^{\text{sen}} = 0.5756$$

$${}_iU_k(g_k^j) = 0.2222$$

$$P(C_i^{\text{sensitive}}) = 0.2844, P(C_i^{\text{insensitive}}) = 0.7156$$

10

Rule 39

Gene: SID W 125268 H.sapiens mRNA for human giant larvae homolog [5':R05862
3':R05776]

Drug: Chlorambucil

15 Parameters:

$$\mu_k^{\text{sen}} = -0.4569, \sigma_k^{\text{sen}} = 0.4595$$

$${}_iU_k(g_k^j) = 0.266$$

$$P(C_i^{\text{sensitive}}) = 0.2206, P(C_i^{\text{insensitive}}) = 0.7794$$

20 Rule 40

Gene: SID 381780 ESTs [5':AA059257 3':AA059223]

Drug: Paclitaxel---Taxol

Parameters:

$$\mu_k^{\text{sen}} = 0.1618, \sigma_k^{\text{sen}} = 0.1828$$

25 ${}_iU_k(g_k^j) = 0.2053$

$$P(C_i^{\text{sensitive}}) = 0.1622, P(C_i^{\text{insensitive}}) = 0.8378$$

Uniform\Gaussian Discriminant Analysis – 2-dimensional (UGDA 2D)

30 This method computes a Bayesian *conditional probability* $P(j \in C_i^{\text{sensitive}} | g_k^j, g_l^j)$
that a cell line j is sensitive to drug i , given the abundances of two genes k and
 l , g_k^j and g_l^j , respectively, in cell line j .

The probability is computed using the following equation:

$$P(j \in C_i^{sensitive} | g_k^j, g_l^j) = \frac{{}_iG_{k,l}^{sensitive}(g_k^j, g_l^j) \cdot P(C_i^{sensitive})}{{}_iG_{k,l}^{sensitive}(g_k^j, g_l^j) \cdot P(C_i^{sensitive}) + {}_iU_{k,l}(g_k^j, g_l^j) \cdot P(C_i^{insensitive})},$$

5 where

$P(C_i^{sensitive})$ = prior probability of the sensitive set

$$= |C_i^{sensitive}| / (|C_i^{sensitive}| + |C_i^{insensitive}|),$$

$P(C_i^{insensitive})$ = prior probability of the insensitive

$$\text{set} = |C_i^{insensitive}| / (|C_i^{sensitive}| + |C_i^{insensitive}|),$$

10 ${}_iG_{k,l}^{sensitive}(g_k^j, g_l^j)$ = joint probability of abundance values g_k^j and g_l^j from the bivariate gaussian density fitted to the histogram of gene k and l abundances over the sensitive cell lines when subjected to drug i .

$$15 \quad {}_iG_{k,l}^{sensitive}(g_k^j, g_l^j) = \frac{1}{2\pi\sigma_k^{sen}\sigma_l^{sen}\sqrt{1-(\rho_{k,l}^{sen})^2}} \exp \left\{ - \frac{\left[\left(\frac{g_k^j - \mu_k^{sen}}{\sigma_k^{sen}} \right)^2 - 2\rho_{k,l}^{sen} \left(\frac{g_k^j - \mu_k^{sen}}{\sigma_k^{sen}} \right) \left(\frac{g_l^j - \mu_l^{sen}}{\sigma_l^{sen}} \right) + \left(\frac{g_l^j - \mu_l^{sen}}{\sigma_l^{sen}} \right)^2 \right]}{2(1-(\rho_{k,l}^{sen})^2)} \right\}$$

where

μ_k^{sen} = mean of gene k abundances over the sensitive cell lines

σ_k^{sen} = standard deviation of gene k abundances in the sensitive cell lines

20 μ_l^{sen} = mean of gene l abundances over the sensitive cell lines

σ_l^{sen} = standard deviation of gene l abundances in the sensitive cell lines

$\rho_{k,l}^{sen}$ = correlation coefficient of gene k and gene l abundances in the sensitive cell lines

${}_iU_{k,l}(g_k^j, g_l^j)$ = probability of abundance values g_k^j and g_l^j from the uniform density fitted to gene k and gene l abundances over all cell lines when subjected to drug i . For given genes k and l , this value is constant across all cell lines, j .

$${}_iU_{k,l}(g_k^j, g_l^j) = \frac{1}{[\max(g_k) - \min(g_k)] \cdot [\max(g_l) - \min(g_l)]},$$

where

$\max(g_k)$ = maximum abundance of gene k over all cell lines

$\min(g_k)$ = minimum abundance of gene k over all cell lines

$\max(g_l)$ = maximum abundance of gene l over all cell lines

$\min(g_l)$ = minimum abundance of gene l over all cell lines

Sample parameters for the UGDA 2D on the NCI60 dataset are:

Rule 1

Gene 1: SID W 116819 Homo sapiens clone 23887 mRNA sequence [5':T93821

3':T93776]

Gene 2: SID W 484681 Homo sapiens ES/130 mRNA complete cds [5':AA037568

3':AA037487]

Drug: L-Alanosine

Parameters:

$$\mu_k^{\text{sen}} = 0.006423, \mu_l^{\text{sen}} = -0.25, \sigma_k^{\text{sen}} = 0.7146, \sigma_l^{\text{sen}} = 0.4424, \rho_{k,l}^{\text{sen}} = 0.7005$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.04605$$

$$P(C_i^{\text{sensitive}}) = 0.2283, P(C_i^{\text{insensitive}}) = 0.7717$$

Rule 2

Gene 1: EST Chr.6 [72745 (R) 5':T50815 3':T50661]

Gene 2: ESTs Weakly similar to dual specificity phosphatase [H.sapiens] Chr.17

[488150 (IW) 5':AA057259 3':AA058704]

Drug: L-Alanosine

Parameters:

$$\mu_k^{\text{sen}} = -0.3181, \mu_l^{\text{sen}} = -0.4347, \sigma_k^{\text{sen}} = 0.7029, \sigma_l^{\text{sen}} = 0.3548, \rho_{k,l}^{\text{sen}} = 0.7733$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.03881$$

$$P(C_i^{\text{sensitive}}) = 0.2283, P(C_i^{\text{insensitive}}) = 0.7717$$

Rule 3

- 5 Gene 1: SID W 469272 Epidermal growth factor receptor [5':AA026175 3':AA026089]
 Gene 2: MICA MHC class I polypeptide-related sequence A Chr.6 [290724 (R) 5':
 3':N71782]

Drug: Dichloroallyl-lawsone

Parameters:

$$10 \quad \mu_k^{\text{sen}} = -0.2886, \mu_l^{\text{sen}} = -0.165, \sigma_k^{\text{sen}} = 0.4416, \sigma_l^{\text{sen}} = 0.3495, \rho_{k,l}^{\text{sen}} = 0.6631$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.03649$$

$$P(C_i^{\text{sensitive}}) = 0.2172, P(C_i^{\text{insensitive}}) = 0.7828$$

Rule 4

- 15 Gene 1: PROBABLE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE Chr.6
 [129496 (E) 5':R16453 3':R14956]
 Gene 2: SID W 125268 H.sapiens mRNA for human giant larvae homolog [5':R05862
 3':R05776]

Drug: Dichloroallyl-lawsone

- 20 Parameters:

$$\mu_k^{\text{sen}} = 0.5512, \mu_l^{\text{sen}} = 0.1164, \sigma_k^{\text{sen}} = 0.509, \sigma_l^{\text{sen}} = 0.7882, \rho_{k,l}^{\text{sen}} = 0.8968$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.05461$$

$$P(C_i^{\text{sensitive}}) = 0.2172, P(C_i^{\text{insensitive}}) = 0.7828$$

- 25 Rule 5

Gene 1: Human LOT1 mRNA complete cds Chr.6 [285041 (I) 5': 3':N63378]
 Gene 2: UBE2H Ubiquitin-conjugating enzyme E2H (homologous to yeast UBC8)
 Chr.7 [359705 (DIW) 5':AA010909 3':AA011300]

Drug: DUP785-brequinar

- 30 Parameters:

$$\mu_k^{\text{sen}} = 0.4687, \mu_l^{\text{sen}} = -0.2413, \sigma_k^{\text{sen}} = 0.5604, \sigma_l^{\text{sen}} = 0.6083, \rho_{k,l}^{\text{sen}} = -0.3827$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.06755$$

$$P(C_i^{\text{sensitive}}) = 0.2694, P(C_i^{\text{insensitive}}) = 0.7306$$

Rule 6

Gene 1: Human putative 32kDa heart protein PHP32 mRNA complete cds Chr.8
[417819 (EW) 5':W88869 3':W88662]

5 Gene 2: SID W 305455 TRANSCRIPTIONAL REGULATOR ISGF3 GAMMA
SUBUNIT [5':W39053 3':N89796]

Drug: Pyrazofurin

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.2413, \mu_l^{\text{sen}} = -0.01115, \sigma_k^{\text{sen}} = 0.3564, \sigma_l^{\text{sen}} = 0.5233, \rho_{k,l}^{\text{sen}} = -0.1372 \\ 10 \quad iU_{k,l}(g_k^j, g_l^j) &= 0.04906 \\ P(C_i^{\text{sensitive}}) &= 0.2594, P(C_i^{\text{insensitive}}) = 0.7406 \end{aligned}$$

Rule 7

Gene 1: SID W 509468 Protective protein for beta-galactosidase (galactosialidosis)
15 [5':AA047117 3':AA047118]

Gene 2: SID W 214236 CD68 antigen [5':H77807 3':H77636]

Drug: Pyrazofurin

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.3715, \mu_l^{\text{sen}} = -0.2611, \sigma_k^{\text{sen}} = 0.521, \sigma_l^{\text{sen}} = 0.5311, \rho_{k,l}^{\text{sen}} = 0.8032 \\ 20 \quad iU_{k,l}(g_k^j, g_l^j) &= 0.05027 \\ P(C_i^{\text{sensitive}}) &= 0.2594, P(C_i^{\text{insensitive}}) = 0.7406 \end{aligned}$$

Rule 8

Gene 1: *Human ferritin L chain mRNA complete cds SID W 239001 ESTs [5':H67076
25 3':H68158]

Gene 2: Homo sapiens mRNA for KIAA0638 protein partial cds Chr.11 [470670 (IW)
5':AA031574 3':AA031453]

Drug: Cyanomorpholinodoxorubicin

Parameters:

$$\begin{aligned} 30 \quad \mu_k^{\text{sen}} &= 0.438, \mu_l^{\text{sen}} = 0.7537, \sigma_k^{\text{sen}} = 0.507, \sigma_l^{\text{sen}} = 0.4528, \rho_{k,l}^{\text{sen}} = -0.7846 \\ iU_{k,l}(g_k^j, g_l^j) &= 0.0424 \\ P(C_i^{\text{sensitive}}) &= 0.2067, P(C_i^{\text{insensitive}}) = 0.7933 \end{aligned}$$

Rule 9

Gene 1: IL8 Interleukin 8 Chr.4 [328692 (DW) 5':W40283 3':W45324]

Gene 2: SID W 305455 TRANSCRIPTIONAL REGULATOR ISGF3 GAMMA
SUBUNIT [5':W39053 3':N89796]

5 Drug: Cyanomorpholinodoxorubicin

Parameters:

$$\mu_k^{\text{sen}} = 0.856, \mu_l^{\text{sen}} = 0.4419, \sigma_k^{\text{sen}} = 0.6623, \sigma_l^{\text{sen}} = 0.3503, \rho_{k,l}^{\text{sen}} = -0.5992$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.051$$

$$P(C_i^{\text{sensitive}}) = 0.2067, P(C_i^{\text{insensitive}}) = 0.7933$$

10

Rule 10

Gene 1: SID 272143 ESTs [5': 3':N35476]

Gene 2: SID W 345420 Homo sapiens YAC clone 136A2 unknown mRNA
3'untranslated region [5':W76024 3':W72468]

15 Drug: Lomustine (CCNU)

Parameters:

$$\mu_k^{\text{sen}} = 0.3141, \mu_l^{\text{sen}} = 0.4027, \sigma_k^{\text{sen}} = 0.5301, \sigma_l^{\text{sen}} = 0.4267, \rho_{k,l}^{\text{sen}} = -0.9555$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.04943$$

$$P(C_i^{\text{sensitive}}) = 0.1067, P(C_i^{\text{insensitive}}) = 0.8933$$

20

Rule 11

Gene 1: ESTs Chr.11 [345012 (IW) 5':W76307 3':W72280]

Gene 2: SID 429145 Human nicotinamide N-methyltransferase (NNMT) mRNA
complete cds [5': 3':AA004839]

25 Drug: Semustine (MeCCNU)

Parameters:

$$\mu_k^{\text{sen}} = 0.1845, \mu_l^{\text{sen}} = 0.2891, \sigma_k^{\text{sen}} = 0.3375, \sigma_l^{\text{sen}} = 0.398, \rho_{k,l}^{\text{sen}} = 0.6251$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.06712$$

$$P(C_i^{\text{sensitive}}) = 0.1606, P(C_i^{\text{insensitive}}) = 0.8394$$

30

Rule 12

Gene 1: INPP1 Inositol polyphosphate-1-phosphatase Chr.2 [183876 (EW) 5':H30231
3':H26976]

Gene 2: SID 429145 Human nicotinamide N-methyltransferase (NNMT) mRNA
complete cds [5': 3':AA004839]

Drug: Semustine (MeCCNU)

Parameters:

$$5 \quad \mu_k^{\text{sen}} = 0.06554, \mu_l^{\text{sen}} = 0.2891, \sigma_k^{\text{sen}} = 0.5184, \sigma_l^{\text{sen}} = 0.398, \rho_{k,l}^{\text{sen}} = -0.6708$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.05885$$

$$P(C_i^{\text{sensitive}}) = 0.1606, P(C_i^{\text{insensitive}}) = 0.8394$$

Rule 13

10 Gene 1: SID 276915 ESTs [5':N48564 3':N39452]

Gene 2: SID 301144 ESTs [5':W16630 3':N78729]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 0.001165, \mu_l^{\text{sen}} = 0.7785, \sigma_k^{\text{sen}} = 0.4, \sigma_l^{\text{sen}} = 0.2994, \rho_{k,l}^{\text{sen}} = -0.3594$$

$$15 \quad iU_{k,l}(g_k^j, g_l^j) = 0.04824$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 14

Gene 1: ESTs Chr.1 [45747 (D) 5':H08940 3':H08856]

20 Gene 2: Human mitogen-responsive phosphoprotein (DOC-2) mRNA complete cds

Chr.5 [428137 (IE) 5': 3':AA001933]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = -0.2316, \mu_l^{\text{sen}} = 0.3967, \sigma_k^{\text{sen}} = 0.4407, \sigma_l^{\text{sen}} = 0.3587, \rho_{k,l}^{\text{sen}} = -0.6006$$

$$25 \quad iU_{k,l}(g_k^j, g_l^j) = 0.05485$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 15

Gene 1: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

30 WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Gene 2: ESTs Chr.1 [488132 (IW) 5':AA047420 3':AA047421]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = -1.008, \mu_l^{\text{sen}} = 0.4755, \sigma_k^{\text{sen}} = 0.5668, \sigma_l^{\text{sen}} = 0.3355, \rho_{k,l}^{\text{sen}} = 0.3703$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.05737$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

5 Rule 16

Gene 1: ESTs Chr.1 [488132 (IW) 5':AA047420 3':AA047421]

Gene 2: ESTs Chr.1 [346583 (IRW) 5':W79544 3':W74533]

Drug: Mitozolamide

Parameters:

$$10 \quad \mu_k^{\text{sen}} = 0.4755, \mu_l^{\text{sen}} = 0.4998, \sigma_k^{\text{sen}} = 0.3355, \sigma_l^{\text{sen}} = 0.593, \rho_{k,l}^{\text{sen}} = 0.612$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.06478$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 17

15 Gene 1: SID 276915 ESTs [5':N48564 3':N39452]

Gene 2: SID W 487878 SPARC/osteonectin [5':AA046533 3':AA045463]

Drug: Mitozolamide

Parameters:

$$20 \quad \mu_k^{\text{sen}} = 0.001165, \mu_l^{\text{sen}} = 0.9224, \sigma_k^{\text{sen}} = 0.4, \sigma_l^{\text{sen}} = 0.4976, \rho_{k,l}^{\text{sen}} = -0.3656$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.04927$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 18

25 Gene 1: *Human ferritin L chain mRNA complete cds SID W 239001 ESTs [5':H67076 3':H68158]

Gene 2: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: Mitozolamide

Parameters:

$$30 \quad \mu_k^{\text{sen}} = 0.5746, \mu_l^{\text{sen}} = -1.008, \sigma_k^{\text{sen}} = 0.4099, \sigma_l^{\text{sen}} = 0.5668, \rho_{k,l}^{\text{sen}} = 0.3637$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.04724$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 19

Gene 1: *Human ferritin L chain mRNA complete cds SID W 239001 ESTs [5':H67076
3':H68158]

Gene 2: CDH2 Cadherin 2 N-cadherin (neuronal) Chr. [325182 (DIRW) 5':W48793

5 3':W49619]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 0.5746, \mu_l^{\text{sen}} = 0.6581, \sigma_k^{\text{sen}} = 0.4099, \sigma_l^{\text{sen}} = 0.3744, \rho_{k,l}^{\text{sen}} = -0.04564$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.05088$$

$$10 \quad P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 20

Gene 1: SID 417008 ESTs Weakly similar to No definition line found [C.elegans] [5':
3':W87796]

15 Gene 2: CDH2 Cadherin 2 N-cadherin (neuronal) Chr. [325182 (DIRW) 5':W48793
3':W49619]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 0.3847, \mu_l^{\text{sen}} = 0.6581, \sigma_k^{\text{sen}} = 0.4824, \sigma_l^{\text{sen}} = 0.3744, \rho_{k,l}^{\text{sen}} = 0.6278$$

$$20 \quad iU_{k,l}(g_k^j, g_l^j) = 0.05309$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 21

Gene 1: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

25 WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Gene 2: SID W 323824 NADH-CYTOCHROME B5 REDUCTASE [5':W46211
3':W46212]

Drug: Mitozolamide

Parameters:

$$30 \quad \mu_k^{\text{sen}} = -1.008, \mu_l^{\text{sen}} = 0.2421, \sigma_k^{\text{sen}} = 0.5668, \sigma_l^{\text{sen}} = 0.4385, \rho_{k,l}^{\text{sen}} = 0.04634$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.05737$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 22

Gene 1: SID 122022 [5':T98316 3':T98261]

Gene 2: *Homo sapiens lysosomal neuraminidase precursor mRNA complete cds SID W 487887 Hexabrachion (tenascin C cytotactin) [5':AA046543 3':AA045473]

5 Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 0.1567, \mu_l^{\text{sen}} = 0.8444, \sigma_k^{\text{sen}} = 0.4277, \sigma_l^{\text{sen}} = 0.5358, \rho_{k,l}^{\text{sen}} = 0.6386$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.0423$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

10

Rule 23

Gene 1: SID W 488691 ESTs Highly similar to NODULATION PROTEIN G [Rhizobium meliloti] [5':AA045967 3':AA045833]

Gene 2: ESTs Chr.7 [28051 (D) 5':R13146 3':R40626]

15 Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = -0.4283, \mu_l^{\text{sen}} = 0.6206, \sigma_k^{\text{sen}} = 0.6985, \sigma_l^{\text{sen}} = 0.4756, \rho_{k,l}^{\text{sen}} = -0.9223$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.05016$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

20

Rule 24

Gene 1: Human DNA sequence from clone 14O9 on chromosome Xp11.1-11.4.

Contains a Inter-Alpha-Trypsin Inh Chr.X [485194 (I) 5':AA039416 3':AA039316]

Gene 2: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW)

25 5':AA039292 3':AA039334]

Drug: Cyclodisone

Parameters:

$$\mu_k^{\text{sen}} = 0.2487, \mu_l^{\text{sen}} = 0.6598, \sigma_k^{\text{sen}} = 0.4569, \sigma_l^{\text{sen}} = 0.2562, \rho_{k,l}^{\text{sen}} = -0.4186$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.03818$$

30 $P(C_i^{\text{sensitive}}) = 0.1689, P(C_i^{\text{insensitive}}) = 0.8311$

Rule 25

Gene 1: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW)

5':AA039292 3':AA039334]

Gene 2: SID 147338 ESTs [5': 3':H01302]

Drug: Cyclodisone

Parameters:

$$\begin{aligned}
 5 \quad & \mu_k^{\text{sen}} = 0.6598, \mu_l^{\text{sen}} = 0.1958, \sigma_k^{\text{sen}} = 0.2562, \sigma_l^{\text{sen}} = 0.3673, \rho_{k,l}^{\text{sen}} = -0.6593 \\
 & iU_{k,l}(g_k^j, g_l^j) = 0.03137 \\
 & P(C_i^{\text{sensitive}}) = 0.1689, P(C_i^{\text{insensitive}}) = 0.8311
 \end{aligned}$$

Rule 26

10 Gene 1: Human GDP-dissociation inhibitor protein (Ly-GDI) mRNA complete cds
Chr.12 [487374 (IW) 5':AA046482 3':AA046695]

Gene 2: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW)
5':AA039292 3':AA039334]

Drug: Cyclodisone

15 Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = -0.2079, \mu_l^{\text{sen}} = 0.6598, \sigma_k^{\text{sen}} = 0.5996, \sigma_l^{\text{sen}} = 0.2562, \rho_{k,l}^{\text{sen}} = -0.7022 \\
 & iU_{k,l}(g_k^j, g_l^j) = 0.03853 \\
 & P(C_i^{\text{sensitive}}) = 0.1689, P(C_i^{\text{insensitive}}) = 0.8311
 \end{aligned}$$

20 Rule 27

Gene 1: SID W 510182 H.sapiens mRNA for kinase A anchor protein [5':AA053156
3':AA053135]

Gene 2: SID W 346663 ESTs [5':W94188 3':W74616]

Drug: Cyclodisone

25 Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = -0.4516, \mu_l^{\text{sen}} = 0.3877, \sigma_k^{\text{sen}} = 0.4114, \sigma_l^{\text{sen}} = 0.3607, \rho_{k,l}^{\text{sen}} = -0.8186 \\
 & iU_{k,l}(g_k^j, g_l^j) = 0.03563 \\
 & P(C_i^{\text{sensitive}}) = 0.1689, P(C_i^{\text{insensitive}}) = 0.8311
 \end{aligned}$$

30 Rule 28

Gene 1: Homo sapiens clone 24560 unknown mRNA complete cds Chr.16 [418227 (IW)
5':W90284 3':W90607]

Gene 2: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW)]

5':AA039292 3':AA039334]

Drug: Cyclodisone

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.2463, \mu_l^{\text{sen}} = 0.6598, \sigma_k^{\text{sen}} = 0.3831, \sigma_l^{\text{sen}} = 0.2562, \rho_{k,l}^{\text{sen}} = 0.5841 \\ 5 \quad iU_{k,l}(g_k^j, g_l^j) &= 0.03311 \\ P(C_i^{\text{sensitive}}) &= 0.1689, P(C_i^{\text{insensitive}}) = 0.8311 \end{aligned}$$

Rule 29

Gene 1: ESTs Chr.1 [488132 (IW) 5':AA047420 3':AA047421]

10 Gene 2: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW)

5':AA039292 3':AA039334]

Drug: Cyclodisone

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.479, \mu_l^{\text{sen}} = 0.6598, \sigma_k^{\text{sen}} = 0.3464, \sigma_l^{\text{sen}} = 0.2562, \rho_{k,l}^{\text{sen}} = -0.4896 \\ 15 \quad iU_{k,l}(g_k^j, g_l^j) &= 0.04029 \\ P(C_i^{\text{sensitive}}) &= 0.1689, P(C_i^{\text{insensitive}}) = 0.8311 \end{aligned}$$

Rule 30

Gene 1: ESTs Chr.1 [488132 (IW) 5':AA047420 3':AA047421]

20 Gene 2: ESTs Chr.1 [346583 (IRW) 5':W79544 3':W74533]

Drug: Cyclodisone

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.479, \mu_l^{\text{sen}} = 0.4024, \sigma_k^{\text{sen}} = 0.3464, \sigma_l^{\text{sen}} = 0.5961, \rho_{k,l}^{\text{sen}} = 0.7576 \\ iU_{k,l}(g_k^j, g_l^j) &= 0.06478 \\ 25 \quad P(C_i^{\text{sensitive}}) &= 0.1689, P(C_i^{\text{insensitive}}) = 0.8311 \end{aligned}$$

Rule 31

Gene 1: SID W 510395 Ribosomal protein S16 [5':AA053701 3':AA053681]

Gene 2: SID W 345420 Homo sapiens YAC clone 136A2 unknown mRNA

30 3'untranslated region [5':W76024 3':W72468]

Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = -0.4557, \mu_l^{\text{sen}} = 0.7165, \sigma_k^{\text{sen}} = 0.2618, \sigma_l^{\text{sen}} = 0.4934, \rho_{k,l}^{\text{sen}} = -0.4265$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.05367$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

Rule 32

- 5 Gene 1: ESTs Weakly similar to GAR22 protein [H.sapiens] Chr. [51904 (E) 5':H24408 3':H22555]

Gene 2: SID 147338 ESTs [5': 3':H01302]

Drug: Clomesone

Parameters:

$$10 \quad \mu_k^{\text{sen}} = 0.3048, \mu_l^{\text{sen}} = 0.1604, \sigma_k^{\text{sen}} = 0.4287, \sigma_l^{\text{sen}} = 0.37, \rho_{k,l}^{\text{sen}} = -0.7076$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.03507$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

Rule 33

- 15 Gene 1: MSN Moesin Chr.X [486864 (IW) 5':AA043008 3':AA042882]

Gene 2: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW) 5':AA039292 3':AA039334]

Drug: Clomesone

Parameters:

$$20 \quad \mu_k^{\text{sen}} = 0.6791, \mu_l^{\text{sen}} = 0.4913, \sigma_k^{\text{sen}} = 0.4486, \sigma_l^{\text{sen}} = 0.4435, \rho_{k,l}^{\text{sen}} = 0.8962$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.03916$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

Rule 34

- 25 Gene 1: Homo sapiens gamma2-adaptin (G2AD) mRNA complete cds Chr.14 [415647 (IW) 5':W78996 3':W80537]

Gene 2: ESTs Chr.6 [146640 (I) 5':R80056 3':R79962]

Drug: Fluorouracil (5FU)

Parameters:

$$30 \quad \mu_k^{\text{sen}} = 0.3802, \mu_l^{\text{sen}} = 0.1649, \sigma_k^{\text{sen}} = 0.419, \sigma_l^{\text{sen}} = 0.7902, \rho_{k,l}^{\text{sen}} = 0.9422$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.04435$$

$$P(C_i^{\text{sensitive}}) = 0.1628, P(C_i^{\text{insensitive}}) = 0.8372$$

Rule 35

Gene 1: SID W 415811 ESTs [5':W84831 3':W84784]

Gene 2: H.sapiens mRNA for Gal-beta(1-3/1-4)GlcNAc alpha-2.3-sialyltransferase
Chr.11 [324181 (IW) 5':W47425 3':W47395]

5 Drug: Fluorouracil (5FU)

Parameters:

$$\mu_k^{\text{sen}} = -0.16, \mu_l^{\text{sen}} = -0.3532, \sigma_k^{\text{sen}} = 0.2818, \sigma_l^{\text{sen}} = 0.2383, \rho_{k,l}^{\text{sen}} = 0.2669$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.0438$$

$$P(C_i^{\text{sensitive}}) = 0.1628, P(C_i^{\text{insensitive}}) = 0.8372$$

10

Rule 36

Gene 1: SID 289361 ESTs [5':N99589 3':N92652]

Gene 2: EST Chr.1 [137318 (I) 5': 3':R36703]

Drug: Fluorouracil (5FU)

15 Parameters:

$$\mu_k^{\text{sen}} = 0.03614, \mu_l^{\text{sen}} = -0.3758, \sigma_k^{\text{sen}} = 0.186, \sigma_l^{\text{sen}} = 0.4475, \rho_{k,l}^{\text{sen}} = -0.1074$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.06362$$

$$P(C_i^{\text{sensitive}}) = 0.1628, P(C_i^{\text{insensitive}}) = 0.8372$$

20 Rule 37

Gene 1: LAMA3 Laminin alpha 3 (nicein (150kD) kalinin (165kD) BM600 (150kD)
epilegrin) Chr.18 [362059 (IRW) 5':AA001431 3':AA001432]

Gene 2: Prostacyclin-stimulating factor [human cultured diploid fibroblast cells mRNA
1124 nt] Chr.4 [488721 (IW) 5':AA046078 3':AA046026]

25 Drug: Cytarabine (araC)

Parameters:

$$\mu_k^{\text{sen}} = -0.3545, \mu_l^{\text{sen}} = -0.4411, \sigma_k^{\text{sen}} = 0.7334, \sigma_l^{\text{sen}} = 0.5863, \rho_{k,l}^{\text{sen}} = 0.8148$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.06236$$

$$P(C_i^{\text{sensitive}}) = 0.2661, P(C_i^{\text{insensitive}}) = 0.7339$$

30

Rule 38

Gene 1: ESTs Chr.14 [244047 (I) 5':N45439 3':N38807]

Gene 2: SID 307717 Homo sapiens KIAA0430 mRNA complete cds [5': 3':N92942]

Drug: Cyclocytidine

Parameters:

$$\mu_k^{\text{sen}} = 0.536, \mu_l^{\text{sen}} = 0.004825, \sigma_k^{\text{sen}} = 0.4307, \sigma_l^{\text{sen}} = 0.232, \rho_{k,l}^{\text{sen}} = 0.1655$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.03336$$

$$5 \quad P(C_i^{\text{sensitive}}) = 0.2533, P(C_i^{\text{insensitive}}) = 0.7467$$

Rule 39

Gene 1: ESTs Chr.1 [31905 (I) 5':R17893 3':R43139]

Gene 2: SID 307717 Homo sapiens KIAA0430 mRNA complete cds [5': 3':N92942]

10 Drug: Cyclocytidine

Parameters:

$$\mu_k^{\text{sen}} = 0.1955, \mu_l^{\text{sen}} = 0.004825, \sigma_k^{\text{sen}} = 0.7301, \sigma_l^{\text{sen}} = 0.232, \rho_{k,l}^{\text{sen}} = 0.685$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.03972$$

$$P(C_i^{\text{sensitive}}) = 0.2533, P(C_i^{\text{insensitive}}) = 0.7467$$

15

Rule 40

Gene 1: SID W 193562 Homo sapiens nuclear autoantigen GS2NA mRNA complete cds [5':H47460 3':H47370]

Gene 2: SID 307717 Homo sapiens KIAA0430 mRNA complete cds [5': 3':N92942]

20 Drug: Cyclocytidine

Parameters:

$$\mu_k^{\text{sen}} = 0.3942, \mu_l^{\text{sen}} = 0.004825, \sigma_k^{\text{sen}} = 0.7788, \sigma_l^{\text{sen}} = 0.232, \rho_{k,l}^{\text{sen}} = 0.5508$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.04087$$

$$P(C_i^{\text{sensitive}}) = 0.2533, P(C_i^{\text{insensitive}}) = 0.7467$$

25

Rule 41

Gene 1: ALDOC Aldolase C fructose-bisphosphate Chr.17 [229961 (IW) 5':H67774 3':H67775]

Gene 2: SID 470499 Human mRNA for KIAA0249 gene complete cds [5':AA031742 3':AA031651]

30

Drug: Anthrapyrazole-derivative

Parameters:

$$\mu_k^{\text{sen}} = -0.2373, \mu_l^{\text{sen}} = 0.4104, \sigma_k^{\text{sen}} = 0.3786, \sigma_l^{\text{sen}} = 0.5297, \rho_{k,l}^{\text{sen}} = -0.7901$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.05241$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 42

5 Gene 1: SID 471855 Lumican [5': 3':AA035657]

Gene 2: Thioredoxin Reductase mRNA-log

Drug: Menogaril

Parameters:

$$\mu_k^{\text{sen}} = -0.5946, \mu_l^{\text{sen}} = 0.4827, \sigma_k^{\text{sen}} = 0.3149, \sigma_l^{\text{sen}} = 0.4498, \rho_{k,l}^{\text{sen}} = 0.8286$$

10 ${}_iU_{k,l}(g_k^j, g_l^j) = 0.03953$

$$P(C_i^{\text{sensitive}}) = 0.1944, P(C_i^{\text{insensitive}}) = 0.8056$$

Rule 43

Gene 1: ESTsSID 327435 [5':W32467 3':W19830]

15 Gene 2: PROBABLE TRANS-1.2-DIHYDROBENZENE-1.2-DIOL

DEHYDROGENASESID 211995 [5':H75805 3':H68500]

Drug: Hydroxyurea

Parameters:

$$\mu_k^{\text{sen}} = -0.3875, \mu_l^{\text{sen}} = -0.05828, \sigma_k^{\text{sen}} = 0.3831, \sigma_l^{\text{sen}} = 0.3997, \rho_{k,l}^{\text{sen}} = 0.8287$$

20 ${}_iU_{k,l}(g_k^j, g_l^j) = 0.05168$

$$P(C_i^{\text{sensitive}}) = 0.1483, P(C_i^{\text{insensitive}}) = 0.8517$$

Rule 44

Gene 1: ESTs Chr.1 [62232 (IR) 5':T40284 3':T41149]

25 Gene 2: SID W 488455 Cathepsin D (lysosomal aspartyl protease) [5':AA047512
3':AA047455]

Drug: CPT,10-OH

Parameters:

$$\mu_k^{\text{sen}} = 0.07749, \mu_l^{\text{sen}} = 0.249, \sigma_k^{\text{sen}} = 0.7379, \sigma_l^{\text{sen}} = 0.4558, \rho_{k,l}^{\text{sen}} = 0.6965$$

30 ${}_iU_{k,l}(g_k^j, g_l^j) = 0.05738$

$$P(C_i^{\text{sensitive}}) = 0.1856, P(C_i^{\text{insensitive}}) = 0.8144$$

Rule 45

Gene 1: SID W 417320 Plasminogen activator tissue type (t-PA) [5':W88922
3':W89129]

Gene 2: Homo sapiens Cyr61 mRNA complete cds Chr.1 [486700 (DIW) 5':AA044451
3':AA044574]

5 Drug: CPT,10-OH

Parameters:

$$\mu_k^{\text{sen}} = 0.614, \mu_l^{\text{sen}} = 0.6231, \sigma_k^{\text{sen}} = 0.4658, \sigma_l^{\text{sen}} = 0.6676, \rho_{k,l}^{\text{sen}} = -0.7235$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.05368$$

$$P(C_i^{\text{sensitive}}) = 0.1856, P(C_i^{\text{insensitive}}) = 0.8144$$

10

Rule 46

Gene 1: ESTs Chr.6 [471083 (IW) 5':AA034335 3':AA033710]

Gene 2: SID W 488148 H.sapiens mRNA for 3'UTR of unknown protein [5':AA057239
3':AA058703]

15 Drug: CPT

Parameters:

$$\mu_k^{\text{sen}} = -0.2213, \mu_l^{\text{sen}} = 0.8224, \sigma_k^{\text{sen}} = 0.6777, \sigma_l^{\text{sen}} = 0.5588, \rho_{k,l}^{\text{sen}} = 0.62$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.04033$$

$$P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

20

Rule 47

Gene 1: *Homo sapiens lysosomal neuraminidase precursor mRNA complete cds SID
W 487887 Hexabrachion (tenascin C cytotactin) [5':AA046543 3':AA045473]

Gene 2: ESTs Weakly similar to !!!! ALU SUBFAMILY J WARNING ENTRY !!!!

25 [H.sapiens] Chr. [21955 (I) 5':T66210 3':T66144]

Drug: CPT

Parameters:

$$\mu_k^{\text{sen}} = 0.3188, \mu_l^{\text{sen}} = 0.5775, \sigma_k^{\text{sen}} = 0.7221, \sigma_l^{\text{sen}} = 0.5522, \rho_{k,l}^{\text{sen}} = -0.8619$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.06477$$

30 $P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$

Rule 48

Gene 1: SID W 365476 Protein S (alpha) [5':AA009419 3':AA009723]

Gene 2: SID W 488148 H.sapiens mRNA for 3'UTR of unknown protein [5':AA057239 3':AA058703]

Drug: CPT

Parameters:

$$\begin{aligned}
 5 \quad & \mu_k^{\text{sen}} = -0.03662, \mu_l^{\text{sen}} = 0.8224, \sigma_k^{\text{sen}} = 0.6534, \sigma_l^{\text{sen}} = 0.5588, \rho_{k,l}^{\text{sen}} = -0.6764 \\
 & {}_iU_{k,l}(g_k^j, g_l^j) = 0.06166 \\
 & P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406
 \end{aligned}$$

Rule 49

10 Gene 1: SID 469530 H.sapiens mRNA for ragA protein [5': 3':AA026944]

Gene 2: Homo sapiens clone 24477 mRNA sequence Chr.18 [33059 (IEW) 5':R19498 3':R43846]

Drug: CPT

Parameters:

$$\begin{aligned}
 15 \quad & \mu_k^{\text{sen}} = 0.459, \mu_l^{\text{sen}} = -0.2041, \sigma_k^{\text{sen}} = 0.5722, \sigma_l^{\text{sen}} = 0.6597, \rho_{k,l}^{\text{sen}} = -0.8312 \\
 & {}_iU_{k,l}(g_k^j, g_l^j) = 0.04669 \\
 & P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406
 \end{aligned}$$

Rule 50

20 Gene 1: SID W 469299 ETS-RELATED PROTEIN ERM [5':AA026205 3':AA026121]

Gene 2: SID W 415693 Homo sapiens mRNA for phosphatidylinositol 4-kinase complete cds [5':W78879 3':W84724]

Drug: CPT

Parameters:

$$\begin{aligned}
 25 \quad & \mu_k^{\text{sen}} = -0.0352, \mu_l^{\text{sen}} = 0.664, \sigma_k^{\text{sen}} = 0.5333, \sigma_l^{\text{sen}} = 0.6375, \rho_{k,l}^{\text{sen}} = -0.8029 \\
 & {}_iU_{k,l}(g_k^j, g_l^j) = 0.0497 \\
 & P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406
 \end{aligned}$$

Rule 51

30 Gene 1: SID W 488148 H.sapiens mRNA for 3'UTR of unknown protein [5':AA057239 3':AA058703]

Gene 2: HLA-DRB5 Major histocompatibility complex class II DR beta 5 Chr.6 [321230 (IEW) 5':W52918 3':AA037380]

Drug: CPT

Parameters:

$$\mu_k^{\text{sen}} = 0.8224, \mu_l^{\text{sen}} = -0.07462, \sigma_k^{\text{sen}} = 0.5588, \sigma_l^{\text{sen}} = 0.7144, \rho_{k,l}^{\text{sen}} = -0.8079$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.05766$$

$$5 \quad P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

Rule 52

Gene 1: ESTs Chr.5 [322749 (I) 5': 3':W15473]

Gene 2: SID 469530 H.sapiens mRNA for ragA protein [5': 3':AA026944]

10 Drug: CPT

Parameters:

$$\mu_k^{\text{sen}} = -0.02124, \mu_l^{\text{sen}} = 0.459, \sigma_k^{\text{sen}} = 0.5919, \sigma_l^{\text{sen}} = 0.5722, \rho_{k,l}^{\text{sen}} = -0.8235$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.05028$$

$$P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

15

Rule 53

Gene 1: SID W 159512 Integrin alpha 6 [5':H16046 3':H15934]

Gene 2: SID 301276 ESTs Highly similar to VALYL-TRNA SYNTHETASE [Fugu rubripes] [5':W07581 3':N80811]

20 Drug: CPT

Parameters:

$$\mu_k^{\text{sen}} = 0.7291, \mu_l^{\text{sen}} = 0.6257, \sigma_k^{\text{sen}} = 0.6557, \sigma_l^{\text{sen}} = 0.6193, \rho_{k,l}^{\text{sen}} = -0.1667$$

$$iU_{k,l}(g_k^j, g_l^j) = 0.05021$$

$$P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

25

Rule 54

Gene 1: SID W 125268 H.sapiens mRNA for human giant larvae homolog [5':R05862 3':R05776]

Gene 2: G6PD Glucose-6-phosphate dehydrogenase Chr.X [430251 (IW) 5':AA010317 3':AA010382]

30

Drug: Chlorambucil

Parameters:

$$\mu_k^{\text{sen}} = -0.4569, \mu_l^{\text{sen}} = -0.2982, \sigma_k^{\text{sen}} = 0.4595, \sigma_l^{\text{sen}} = 0.2945, \rho_{k,l}^{\text{sen}} = -0.1414$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.06214$$

$$P(C_i^{\text{sensitive}}) = 0.2206, P(C_i^{\text{insensitive}}) = 0.7794$$

Rule 55

5 Gene 1: SID W 510534 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED PROTEIN GA733-2 PRECURSOR [5':AA055858 3':AA055808]

Gene 2: G6PD Glucose-6-phosphate dehydrogenase Chr.X [430251 (IW) 5':AA010317 3':AA010382]

Drug: Chlorambucil

10 Parameters:

$$\mu_k^{\text{sen}} = -0.7249, \mu_l^{\text{sen}} = -0.2982, \sigma_k^{\text{sen}} = 0.5634, \sigma_l^{\text{sen}} = 0.2945, \rho_{k,l}^{\text{sen}} = -0.3986$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.06933$$

$$P(C_i^{\text{sensitive}}) = 0.2206, P(C_i^{\text{insensitive}}) = 0.7794$$

15 Rule 56

Gene 1: SID 29828 ESTs [5':R16390 3':R42331]

Gene 2: SID W 485645 KERATIN TYPE II CYTOSKELETAL 7 [5':AA039817 3':AA041344]

Drug: 5-Hydroxypicolinaldehyde-thiose

20 Parameters:

$$\mu_k^{\text{sen}} = -0.1536, \mu_l^{\text{sen}} = 0.8712, \sigma_k^{\text{sen}} = 0.5974, \sigma_l^{\text{sen}} = 0.6735, \rho_{k,l}^{\text{sen}} = 0.6716$$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.03954$$

$$P(C_i^{\text{sensitive}}) = 0.1789, P(C_i^{\text{insensitive}}) = 0.8211$$

25 Rule 57

Gene 1: SID 381780 ESTs [5':AA059257 3':AA059223]

Gene 2: SID 130482 ESTs [5':R21876 3':R21877]

Drug: Paclitaxel---Taxol

Parameters:

30 $\mu_k^{\text{sen}} = 0.1618, \mu_l^{\text{sen}} = -0.9271, \sigma_k^{\text{sen}} = 0.1828, \sigma_l^{\text{sen}} = 0.3413, \rho_{k,l}^{\text{sen}} = -0.3935$

$${}_iU_{k,l}(g_k^j, g_l^j) = 0.05375$$

$$P(C_i^{\text{sensitive}}) = 0.1622, P(C_i^{\text{insensitive}}) = 0.8378$$

Rule 58

Gene 1: SID 381780 ESTs [5':AA059257 3':AA059223]

Gene 2: SID 512355 ESTs Highly similar to SRC SUBSTRATE P80/85 PROTEINS
[Gallus gallus] [5':AA059424 3':AA057835]

5 Drug: Paclitaxel---Taxol

Parameters:

$$\mu_k^{\text{sen}} = 0.1618, \mu_l^{\text{sen}} = -0.8354, \sigma_k^{\text{sen}} = 0.1828, \sigma_l^{\text{sen}} = 0.4935, \rho_{k,l}^{\text{sen}} = -0.09957$$

$${}_iU_{k,l}(g_k^i, g_l^i) = 0.06437$$

$$P(C_i^{\text{sensitive}}) = 0.1622, P(C_i^{\text{insensitive}}) = 0.8378$$

10

Rule 59

Gene 1: *Paired basic amino acid cleaving enzyme (furin membrane associated receptor protein) SID W 114116 Syndecan 2 (heparan sulfate proteoglycan 1 cell surface-associated fibroglycan) [5':T79562 3':T79471]

15 Gene 2: SID 240167 ESTs [5':H79634 3':H79635]

Drug: Pyrazoloacridine

Parameters:

$$\mu_k^{\text{sen}} = -0.6405, \mu_l^{\text{sen}} = 0.3087, \sigma_k^{\text{sen}} = 0.5377, \sigma_l^{\text{sen}} = 0.4283, \rho_{k,l}^{\text{sen}} = 0.7929$$

$${}_iU_{k,l}(g_k^i, g_l^i) = 0.05053$$

$$20 \quad P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

Linear Discriminant Analysis – 1-dimensional (LDA 1D)

This method computes a Bayesian conditional probability $P(j \in C_i^{\text{sensitive}} | g_k^j)$

that a cell line j is sensitive to drug i , given the gene k abundance g_k^j in cell
25 line j .

The probability is computed using the following equation:

$$P(j \in C_i^{\text{sensitive}} | g_k^j) = \frac{{}_iG_k^{\text{sensitive}}(g_k^j) \cdot P(C_i^{\text{sensitive}})}{{}_iG_k^{\text{sensitive}}(g_k^j) \cdot P(C_i^{\text{sensitive}}) + {}_iG_k^{\text{insensitive}}(g_k^j) \cdot P(C_i^{\text{insensitive}})}$$

30

where

$$P(C_i^{sensitive}) = \text{prior probability of the sensitive set}$$

$$= |C_i^{sensitive}| / (|C_i^{sensitive}| + |C_i^{insensitive}|),$$

$$P(C_i^{insensitive}) = \text{prior probability of the insensitive set}$$

$$= |C_i^{insensitive}| / (|C_i^{sensitive}| + |C_i^{insensitive}|),$$

5

${}_i G_k^{sensitive}(g_k^j)$ = probability of abundance value g_k^j from the gaussian density fitted to the histogram of the gene k abundances over the sensitive cell lines when subjected to drug i .

10

$${}_i G_k^{sensitive}(g_k^j) = \frac{1}{\sigma_k^{avg} \sqrt{2\pi}} e^{-(g_k^j - \mu_k^{sen})^2 / 2(\sigma_k^{avg})^2},$$

where

μ_k^{sen} = mean of gene k abundances in the sensitive cell lines

σ_k^{avg} = sensitive\insensitive class-weighted average standard deviation of gene k abundances in the sensitive cell lines

15

${}_i G_k^{insensitive}(g_k^j)$ = probability of abundance value g_k^j from the gaussian density fitted to the histogram of the gene k abundances over the insensitive cell lines when subjected to drug i .

20

$${}_i G_k^{insensitive}(g_k^j) = \frac{1}{\sigma_k^{avg} \sqrt{2\pi}} e^{-(g_k^j - \mu_k^{insen})^2 / 2(\sigma_k^{avg})^2},$$

where

μ_k^{insen} = mean of gene k abundances in the insensitive cell lines

Sample parameters for the LDA 1D analysis on the NCI60 Dataset are set out below:

25 Rule 1

Gene: SID W 470947 Human scaffold protein Pbp1 mRNA complete cds [5':AA032174

3':AA032175]

Drug: Inosine-glycodialdehyde

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.8115 \\ 5 \quad \mu_k^{\text{insen}} &= 0.2001 \\ \sigma_k^{\text{avg}} &= 0.9394 \\ P(C_i^{\text{sensitive}}) &= 0.1978, P(C_i^{\text{insensitive}}) = 0.8022 \end{aligned}$$

Rule 2

10 Gene: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW)

5':AA039292 3':AA039334]

Drug: Inosine-glycodialdehyde

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.7618 \\ 15 \quad \mu_k^{\text{insen}} &= 0.1878 \\ \sigma_k^{\text{avg}} &= 0.9598 \\ P(C_i^{\text{sensitive}}) &= 0.1978, P(C_i^{\text{insensitive}}) = 0.8022 \end{aligned}$$

Rule 3

20 Gene: Homo sapiens cyclin-dependent kinase inhibitor (CDKN2C) mRNA complete cds
Chr. [291057 (RW) 5':W00390 3':N72115]

Drug: L-Alanosine

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.8435 \\ 25 \quad \mu_k^{\text{insen}} &= 0.25 \\ \sigma_k^{\text{avg}} &= 0.8772 \\ P(C_i^{\text{sensitive}}) &= 0.2283, P(C_i^{\text{insensitive}}) = 0.7717 \end{aligned}$$

Rule 4

30 Gene: SID W 254085 ESTs Moderately similar to synaptonemal complex protein
[M.musculus] [5':N71532 3':N22165]

Drug: Baker's-soluble-antifolate

Parameters:

$$\mu_k^{\text{sen}} = 0.7847$$

$$\mu_k^{\text{insen}} = -0.2423$$

$$\sigma_k^{\text{avg}} = 0.8539$$

$$5 \quad P(C_i^{\text{sensitive}}) = 0.2361, P(C_i^{\text{insensitive}}) = 0.7639$$

Rule 5

Gene: M-PHASE INDUCER PHOSPHATASE 2 Chr.20 [179373 (EW) 5':H50437
3':H50438]

10 Drug: 5-6-Dihydro-5-azacytidine

Parameters:

$$\mu_k^{\text{sen}} = -0.9251$$

$$\mu_k^{\text{insen}} = 0.2324$$

$$\sigma_k^{\text{avg}} = 0.8567$$

$$15 \quad P(C_i^{\text{sensitive}}) = 0.2011, P(C_i^{\text{insensitive}}) = 0.7989$$

Rule 6

Gene: THY-1 MEMBRANE GLYCOPROTEIN PRECURSOR Chr.11 [183950 (E)
5':H30297 3':H28104]

20 Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 1.073$$

$$\mu_k^{\text{insen}} = -0.2694$$

$$\sigma_k^{\text{avg}} = 0.8153$$

$$25 \quad P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 7

Gene: PTN Pleiotrophin (heparin binding growth factor 8 neurite growth-promoting
factor 1) Chr.7 [488801 (IW) 5':AA045053 3':AA045054]

30 Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 1.019$$

$$\mu_k^{\text{insen}} = -0.2557$$

$$\sigma_k^{\text{avg}} = 0.8554$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

5 Rule 8

Gene: SID W 380674 ESTs [5':AA053720 3':AA053711]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 1.093$$

$$10 \quad \mu_k^{\text{insen}} = -0.2739$$

$$\sigma_k^{\text{avg}} = 0.8441$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 9

15 Gene: Glutathione S-Transferase Pi-log

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = -0.917$$

$$\mu_k^{\text{insen}} = 0.2307$$

$$20 \quad \sigma_k^{\text{avg}} = 0.8411$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 10

Gene: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J WARNING

25 ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = -1.008$$

$$\mu_k^{\text{insen}} = 0.2536$$

$$30 \quad \sigma_k^{\text{avg}} = 0.8681$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 11

Gene: *Hs.648 Cut (Drosophila)-like 1 (CCAAT displacement protein) SID W 26677
ESTs [5':R13994 3':R39117]

Drug: Mitozolamide

5 Parameters:

$$\mu_k^{\text{sen}} = 0.8138$$

$$\mu_k^{\text{insen}} = -0.2039$$

$$\sigma_k^{\text{avg}} = 0.9103$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

10

Rule 12

Gene: SID W 488387 Exostoses (multiple) 2 [5':AA046786 3':AA046656]

Drug: Cyclodisone

Parameters:

$$15 \quad \mu_k^{\text{sen}} = 1.043$$

$$\mu_k^{\text{insen}} = -0.2128$$

$$\sigma_k^{\text{avg}} = 0.8985$$

$$P(C_i^{\text{sensitive}}) = 0.1689, P(C_i^{\text{insensitive}}) = 0.8311$$

20 Rule 13

Gene: THY-1 MEMBRANE GLYCOPROTEIN PRECURSOR Chr.11 [183950 (E)
5':H30297 3':H28104]

Drug: Cyclodisone

Parameters:

$$25 \quad \mu_k^{\text{sen}} = 1.135$$

$$\mu_k^{\text{insen}} = -0.2308$$

$$\sigma_k^{\text{avg}} = 0.8251$$

$$P(C_i^{\text{sensitive}}) = 0.1689, P(C_i^{\text{insensitive}}) = 0.8311$$

30 Rule 14

Gene: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528
3':AA043529]

Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = 1.184$$

$$\mu_k^{\text{insen}} = -0.2817$$

$$5 \quad \sigma_k^{\text{avg}} = 0.829$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

Rule 15

Gene: PTN Pleiotrophin (heparin binding growth factor 8 neurite growth-promoting
10 factor 1) Chr.7 [488801 (IW) 5':AA045053 3':AA045054]

Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = 1.14$$

$$\mu_k^{\text{insen}} = -0.2703$$

$$15 \quad \sigma_k^{\text{avg}} = 0.8309$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

Rule 16

Gene: THY-1 MEMBRANE GLYCOPROTEIN PRECURSOR Chr.11 [183950 (E)
20 5':H30297 3':H28104]

Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = 1.157$$

$$\mu_k^{\text{insen}} = -0.2746$$

$$25 \quad \sigma_k^{\text{avg}} = 0.8226$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

Rule 17

Gene: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J WARNING
30 ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = -1.079$$

$$\mu_k^{\text{insen}} = 0.2564$$

$$\sigma_k^{\text{avg}} = 0.8587$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

5

Rule 18

Gene: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528
3':AA043529]

Drug: PCNU

10 Parameters:

$$\mu_k^{\text{sen}} = 1.081$$

$$\mu_k^{\text{insen}} = -0.2435$$

$$\sigma_k^{\text{avg}} = 0.8791$$

$$P(C_i^{\text{sensitive}}) = 0.1833, P(C_i^{\text{insensitive}}) = 0.8167$$

15

Rule 19

Gene: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J WARNING
ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: PCNU

20 Parameters:

$$\mu_k^{\text{sen}} = -1.078$$

$$\mu_k^{\text{insen}} = 0.2427$$

$$\sigma_k^{\text{avg}} = 0.8755$$

$$P(C_i^{\text{sensitive}}) = 0.1833, P(C_i^{\text{insensitive}}) = 0.8167$$

25

Rule 20

Gene: PTN Pleiotrophin (heparin binding growth factor 8 neurite growth-promoting
factor 1) Chr.7 [488801 (IW) 5':AA045053 3':AA045054]

Drug: PCNU

30 Parameters:

$$\mu_k^{\text{sen}} = 1.115$$

$$\mu_k^{\text{insen}} = -0.2502$$

$$\sigma_k^{avg} = 0.8538$$

$$P(C_i^{sensitive}) = 0.1833, P(C_i^{insensitive}) = 0.8167$$

Rule 21

- 5 Gene: Human thymosin beta-4 mRNA complete cds Chr.20 [305890 (IW) 5':W19923 3':N91268]

Drug: Cytarabine (araC)

Parameters:

$$\mu_k^{sen} = -0.7694$$

$$10 \quad \mu_k^{insen} = 0.2788$$

$$\sigma_k^{avg} = 0.8663$$

$$P(C_i^{sensitive}) = 0.2661, P(C_i^{insensitive}) = 0.7339$$

Rule 22

- 15 Gene: SID W 291620 Restin (Reed-Steinberg cell-expressed intermediate filament-associated protein) [5':W03421 3':N67817]

Drug: Porfiromycin

Parameters:

$$\mu_k^{sen} = 0.9491$$

$$20 \quad \mu_k^{insen} = -0.2431$$

$$\sigma_k^{avg} = 0.8965$$

$$P(C_i^{sensitive}) = 0.2039, P(C_i^{insensitive}) = 0.7961$$

Rule 23

- 25 Gene: Human extracellular protein (S1-5) mRNA complete cds Chr.2 [485875 (EW) 5':AA040442 3':AA040443]

Drug: Oxanthrazole (piroxastrone)

Parameters:

$$\mu_k^{sen} = 1.155$$

$$30 \quad \mu_k^{insen} = -0.2805$$

$$\sigma_k^{avg} = 0.7962$$

$$P(C_i^{sensitive}) = 0.1956, P(C_i^{insensitive}) = 0.8044$$

Rule 24

Gene: SID W 299539 Human fibroblast growth factor homologous factor 1 (FHF-1)
mRNA complete cds [5':W05845 3':N71102]

5 Drug: Oxanthrazole (piroxastrone)

Parameters:

$$\mu_k^{\text{sen}} = 0.9238$$

$$\mu_k^{\text{insen}} = -0.2254$$

$$\sigma_k^{\text{avg}} = 0.862$$

$$10 \quad P(C_i^{\text{sensitive}}) = 0.1956, P(C_i^{\text{insensitive}}) = 0.8044$$

Rule 25

Gene: SID W 488148 H.sapiens mRNA for 3'UTR of unknown protein [5':AA057239
3':AA058703]

15 Drug: Oxanthrazole (piroxastrone)

Parameters:

$$\mu_k^{\text{sen}} = 0.8896$$

$$\mu_k^{\text{insen}} = -0.2163$$

$$\sigma_k^{\text{avg}} = 0.8858$$

$$20 \quad P(C_i^{\text{sensitive}}) = 0.1956, P(C_i^{\text{insensitive}}) = 0.8044$$

Rule 26

Gene: Human extracellular protein (S1-5) mRNA complete cds Chr.2 [485875 (EW)
5':AA040442 3':AA040443]

25 Drug: Anthrapyrazole-derivative

Parameters:

$$\mu_k^{\text{sen}} = 1.016$$

$$\mu_k^{\text{insen}} = -0.2548$$

$$\sigma_k^{\text{avg}} = 0.8692$$

$$30 \quad P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 27

Gene: SID W 380674 ESTs [5':AA053720 3':AA053711]

Drug: Anthrapyrazole-derivative

Parameters:

$$\begin{aligned}
 &\mu_k^{\text{sen}} = 0.9038 \\
 5 \quad &\mu_k^{\text{insen}} = -0.2265 \\
 &\sigma_k^{\text{avg}} = 0.8998 \\
 &P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994
 \end{aligned}$$

Rule 28

10 Gene: ESTs Chr.2 [365120 (IW) 5':AA025204 3':AA025124]

Drug: Anthrapyrazole-derivative

Parameters:

$$\begin{aligned}
 &\mu_k^{\text{sen}} = 0.9014 \\
 &\mu_k^{\text{insen}} = -0.2264 \\
 15 \quad &\sigma_k^{\text{avg}} = 0.9007 \\
 &P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994
 \end{aligned}$$

Rule 29

Gene: SID 229535 [5':H66594 3':H66595]

20 Drug: Teniposide

Parameters:

$$\begin{aligned}
 &\mu_k^{\text{sen}} = -0.9209 \\
 &\mu_k^{\text{insen}} = 0.2154 \\
 &\sigma_k^{\text{avg}} = 0.9114 \\
 25 \quad &P(C_i^{\text{sensitive}}) = 0.1894, P(C_i^{\text{insensitive}}) = 0.8106
 \end{aligned}$$

Rule 30

Gene: ESTs Chr.2 [149542 (DW) 5':H00283 3':H00284]

Drug: Daunorubicin

30 Parameters:

$$\begin{aligned}
 &\mu_k^{\text{sen}} = -1.052 \\
 &\mu_k^{\text{insen}} = 0.2324
 \end{aligned}$$

$$\sigma_k^{avg} = 0.8508$$

$$P(C_i^{sensitive}) = 0.1811, P(C_i^{insensitive}) = 0.8189$$

Rule 31

- 5 Gene: SID W 510030 ESTs Weakly similar to N-methyl-D-aspartate receptor glutamate-binding chain [R.norvegicus] [5':AA053050 3':AA053392]

Drug: Daunorubicin

Parameters:

$$\mu_k^{sen} = -1.088$$

$$10 \quad \mu_k^{insen} = 0.2401$$

$$\sigma_k^{avg} = 0.8526$$

$$P(C_i^{sensitive}) = 0.1811, P(C_i^{insensitive}) = 0.8189$$

Rule 32

- 15 Gene: SID 260288 ESTs [5':H97716 3':H96798]

Drug: Daunorubicin

Parameters:

$$\mu_k^{sen} = -0.9929$$

$$\mu_k^{insen} = 0.2192$$

$$20 \quad \sigma_k^{avg} = 0.9063$$

$$P(C_i^{sensitive}) = 0.1811, P(C_i^{insensitive}) = 0.8189$$

Rule 33

Gene: AK1 Adenylate kinase 1 Chr.9 [488381 (IW) 5':AA046783 3':AA046653]

- 25 Drug: Daunorubicin

Parameters:

$$\mu_k^{sen} = -0.9847$$

$$\mu_k^{insen} = 0.2169$$

$$\sigma_k^{avg} = 0.8611$$

$$30 \quad P(C_i^{sensitive}) = 0.1811, P(C_i^{insensitive}) = 0.8189$$

Rule 34

Gene: Homo sapiens T245 protein (T245) mRNA complete cds Chr.X [343063 (IW)
5':W67989 3':W68001]

Drug: Daunorubicin

Parameters:

$$\begin{aligned} 5 \quad & \mu_k^{\text{sen}} = -1.061 \\ & \mu_k^{\text{insen}} = 0.234 \\ & \sigma_k^{\text{avg}} = 0.8647 \\ & P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189 \end{aligned}$$

10 Rule 35

Gene: *Prothymosin alpha SID W 271976 AMINOACYLASE-1 [5':N44687 3':N35315]

Drug: Daunorubicin

Parameters:

$$\begin{aligned} 15 \quad & \mu_k^{\text{sen}} = -1.032 \\ & \mu_k^{\text{insen}} = 0.2284 \\ & \sigma_k^{\text{avg}} = 0.858 \\ & P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189 \end{aligned}$$

Rule 36

20 Gene: SID W 345683 ESTs Highly similar to INTEGRAL MEMBRANE
GLYCOPROTEIN GP210 PRECURSOR [Rattus norvegicus] [5':W76432 3':W72039]

Drug: Daunorubicin

Parameters:

$$\begin{aligned} 25 \quad & \mu_k^{\text{sen}} = 0.918 \\ & \mu_k^{\text{insen}} = -0.2022 \\ & \sigma_k^{\text{avg}} = 0.8758 \\ & P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189 \end{aligned}$$

Rule 37

30 Gene: Homo sapiens clone 24477 mRNA sequence Chr.18 [33059 (IEW) 5':R19498
3':R43846]

Drug: Daunorubicin

Parameters:

$$\mu_k^{\text{sen}} = -0.966$$

$$\mu_k^{\text{insen}} = 0.2126$$

$$\sigma_k^{\text{avg}} = 0.8952$$

$$5 \quad P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

Rule 38

Gene: SID 43609 ESTs [5':H06454 3':H06184]

Drug: Amsacrine

10 Parameters:

$$\mu_k^{\text{sen}} = 0.9136$$

$$\mu_k^{\text{insen}} = -0.2581$$

$$\sigma_k^{\text{avg}} = 0.8733$$

$$P(C_i^{\text{sensitive}}) = 0.22, P(C_i^{\text{insensitive}}) = 0.78$$

15

Rule 39

Gene: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

Drug: CPT,10-OH

20 Parameters:

$$\mu_k^{\text{sen}} = -0.9086$$

$$\mu_k^{\text{insen}} = 0.2078$$

$$\sigma_k^{\text{avg}} = 0.8915$$

$$P(C_i^{\text{sensitive}}) = 0.1856, P(C_i^{\text{insensitive}}) = 0.8144$$

25

Rule 40

Gene: SID W 346587 Homo sapiens quiescin (Q6) mRNA complete cds [5':W79188
3':W74434]

Drug: CPT,10-OH

30 Parameters:

$$\mu_k^{\text{sen}} = 1.001$$

$$\mu_k^{\text{insen}} = -0.2285$$

$$\sigma_k^{avg} = 0.8549$$

$$P(C_i^{sensitive}) = 0.1856, P(C_i^{insensitive}) = 0.8144$$

Rule 41

5 Gene: SID 39144 ESTs Weakly similar to Rep-8 [H.sapiens] [5':R51769 3':R51770]

Drug: CPT,20-ester (S)

Parameters:

$$\mu_k^{sen} = -0.8367$$

$$\mu_k^{insen} = 0.2555$$

10 $\sigma_k^{avg} = 0.8798$

$$P(C_i^{sensitive}) = 0.2344, P(C_i^{insensitive}) = 0.7656$$

Rule 42

Gene: SID W 358526 ESTs [5':W96039 3':W94821]

15 Drug: CPT,14-Cl (S)

Parameters:

$$\mu_k^{sen} = -0.8436$$

$$\mu_k^{insen} = 0.2136$$

$$\sigma_k^{avg} = 0.9027$$

20 $P(C_i^{sensitive}) = 0.2022, P(C_i^{insensitive}) = 0.7978$

Rule 43

Gene: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19

[310021 (I) 5': 3':N99151]

25 Drug: CPT,20-acetate

Parameters:

$$\mu_k^{sen} = -0.8754$$

$$\mu_k^{insen} = 0.1973$$

$$\sigma_k^{avg} = 0.8929$$

30 $P(C_i^{sensitive}) = 0.1833, P(C_i^{insensitive}) = 0.8167$

Rule 44

Gene: SID 512355 ESTs Highly similar to SRC SUBSTRATE P80/85 PROTEINS
[Gallus gallus] [5':AA059424 3':AA057835]

Drug: CPT

Parameters:

$$\begin{aligned}
 5 \quad & \mu_k^{\text{sen}} = 0.8614 \\
 & \mu_k^{\text{insen}} = -0.3016 \\
 & \sigma_k^{\text{avg}} = 0.8698 \\
 & P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406
 \end{aligned}$$

10 Rule 45

Gene: SID W 488148 H.sapiens mRNA for 3'UTR of unknown protein [5':AA057239
3':AA058703]

Drug: CPT

Parameters:

$$\begin{aligned}
 15 \quad & \mu_k^{\text{sen}} = 0.8224 \\
 & \mu_k^{\text{insen}} = -0.2881 \\
 & \sigma_k^{\text{avg}} = 0.8739 \\
 & P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406
 \end{aligned}$$

20 Rule 46

Gene: ESTs Chr.19 [485804 (EW) 5':AA040350 3':AA040351]

Drug: CPT,20-ester (S)

Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = -0.7505 \\
 25 \quad & \mu_k^{\text{insen}} = 0.2562 \\
 & \sigma_k^{\text{avg}} = 0.8843 \\
 & P(C_i^{\text{sensitive}}) = 0.255, P(C_i^{\text{insensitive}}) = 0.745
 \end{aligned}$$

Rule 47

30 Gene: SID W 358526 ESTs [5':W96039 3':W94821]

Drug: CPT,11-formyl (RS)

Parameters:

$$\mu_k^{\text{sen}} = -1.055$$

$$\mu_k^{\text{insen}} = 0.2536$$

$$\sigma_k^{\text{avg}} = 0.8569$$

$$P(C_i^{\text{sensitive}}) = 0.1939, P(C_i^{\text{insensitive}}) = 0.8061$$

5

Rule 48

Gene: SID W 135118 GATA-binding protein 3 [5':R31441 3':R31442]

Drug: CPT,11-formyl (RS)

Parameters:

$$10 \quad \mu_k^{\text{sen}} = 0.9817$$

$$\mu_k^{\text{insen}} = -0.2359$$

$$\sigma_k^{\text{avg}} = 0.9021$$

$$P(C_i^{\text{sensitive}}) = 0.1939, P(C_i^{\text{insensitive}}) = 0.8061$$

15 Rule 49

Gene: ESTs Chr.16 [154654 (RW) 5':R55184 3':R55185]

Drug: CPT,11-formyl (RS)

Parameters:

$$\mu_k^{\text{sen}} = 0.874$$

$$20 \quad \mu_k^{\text{insen}} = -0.2102$$

$$\sigma_k^{\text{avg}} = 0.9112$$

$$P(C_i^{\text{sensitive}}) = 0.1939, P(C_i^{\text{insensitive}}) = 0.8061$$

Rule 50

25 Gene: SID 43609 ESTs [5':H06454 3':H06184]

Drug: Mechlorethamine

Parameters:

$$\mu_k^{\text{sen}} = 1.042$$

$$\mu_k^{\text{insen}} = -0.2493$$

$$30 \quad \sigma_k^{\text{avg}} = 0.8728$$

$$P(C_i^{\text{sensitive}}) = 0.1928, P(C_i^{\text{insensitive}}) = 0.8072$$

Rule 51

Gene: SID W 133851 ESTs [5':R28233 3':R27977]

Drug: Triethylenemelamine

Parameters:

$$\begin{aligned}
 5 \quad & \mu_k^{\text{sen}} = -0.7551 \\
 & \mu_k^{\text{insen}} = 0.2248 \\
 & \sigma_k^{\text{avg}} = 0.9176 \\
 & P(C_i^{\text{sensitive}}) = 0.2294, P(C_i^{\text{insensitive}}) = 0.7706
 \end{aligned}$$

10 Rule 52

Gene: SID W 133851 ESTs [5':R28233 3':R27977]

Drug: Chlorambucil

Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = -0.8278 \\
 15 \quad & \mu_k^{\text{insen}} = 0.2342 \\
 & \sigma_k^{\text{avg}} = 0.8901 \\
 & P(C_i^{\text{sensitive}}) = 0.2206, P(C_i^{\text{insensitive}}) = 0.7794
 \end{aligned}$$

Rule 53

20 Gene: Human mRNA for KIAA0382 gene partial cds Chr.11 [486712 (IEW)
5':AA043173 3':AA043174]

Drug: Chlorambucil

Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = -0.8832 \\
 25 \quad & \mu_k^{\text{insen}} = 0.2497 \\
 & \sigma_k^{\text{avg}} = 0.8826 \\
 & P(C_i^{\text{sensitive}}) = 0.2206, P(C_i^{\text{insensitive}}) = 0.7794
 \end{aligned}$$

Rule 54

30 Gene: CDH2 Cadherin 2 N-cadherin (neuronal) Chr. [325182 (DIRW) 5':W48793
3':W49619]

Drug: Geldanamycin

Parameters:

$$\mu_k^{\text{sen}} = -0.8842$$

$$\mu_k^{\text{insen}} = 0.225$$

$$\sigma_k^{\text{avg}} = 0.8839$$

$$5 \quad P(C_i^{\text{sensitive}}) = 0.2033, P(C_i^{\text{insensitive}}) = 0.7967$$

Rule 55

Gene: Human nicotinamide nucleotide transhydrogenase mRNA nuclear gene encoding mitochondrial protein Chr. [287568 (I) 5': 3':N62116]

10 Drug: Morpholino-adriamycin

Parameters:

$$\mu_k^{\text{sen}} = -1.072$$

$$\mu_k^{\text{insen}} = 0.2139$$

$$\sigma_k^{\text{avg}} = 0.8933$$

$$15 \quad P(C_i^{\text{sensitive}}) = 0.1661, P(C_i^{\text{insensitive}}) = 0.8339$$

Rule 56

Gene: H.sapiens mRNA for TRAMP protein Chr.8 [149355 (IEW) 5':H01598 3':H01495]

20 Drug: Amonafide

Parameters:

$$\mu_k^{\text{sen}} = 1.095$$

$$\mu_k^{\text{insen}} = -0.2498$$

$$\sigma_k^{\text{avg}} = 0.8687$$

$$25 \quad P(C_i^{\text{sensitive}}) = 0.1861, P(C_i^{\text{insensitive}}) = 0.8139$$

Rule 57

Gene: SID W 415811 ESTs [5':W84831 3':W84784]

Drug: Pyrazoloacridine

30 Parameters:

$$\mu_k^{\text{sen}} = -0.873$$

$$\mu_k^{\text{insen}} = 0.1935$$

$$\sigma_k^{\text{avg}} = 0.8924$$

$$P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

5 Quadratic Discriminant Analysis – 1-dimensional (QDA 1D)

This method computes a Bayesian conditional probability $P(j \in C_i^{\text{sensitive}} | g_k^j)$ that a cell line j is sensitive to drug i , given the gene k abundance g_k^j in cell line j .

10 The probability is computed using the following equation:

$$P(j \in C_i^{\text{sensitive}} | g_k^j) = \frac{{}_iG_k^{\text{sensitive}}(g_k^j) \cdot P(C_i^{\text{sensitive}})}{{}_iG_k^{\text{sensitive}}(g_k^j) \cdot P(C_i^{\text{sensitive}}) + {}_iG_k^{\text{insensitive}}(g_k^j) \cdot P(C_i^{\text{insensitive}})}$$

where

$P(C_i^{\text{sensitive}})$ = prior probability of the sensitive set

$$15 = |C_i^{\text{sensitive}}| / (|C_i^{\text{sensitive}}| + |C_i^{\text{insensitive}}|),$$

$P(C_i^{\text{insensitive}})$ = prior probability of the insensitive

$$\text{set} = |C_i^{\text{insensitive}}| / (|C_i^{\text{sensitive}}| + |C_i^{\text{insensitive}}|),$$

${}_iG_k^{\text{sensitive}}(g_k^j)$ = probability of abundance value g_k^j from the gaussian density fitted

20 to the histogram of the gene k abundances over the sensitive cell lines when subjected to drug i .

$${}_iG_k^{\text{sensitive}}(g_k^j) = \frac{1}{\sigma_k^{\text{sen}} \sqrt{2\pi}} e^{-(g_k^j - \mu_k^{\text{sen}})^2 / 2(\sigma_k^{\text{sen}})^2},$$

where

25 μ_k^{sen} = mean of gene k abundances in the sensitive cell lines

σ_k^{sen} = standard deviation of gene k abundances in the sensitive cell lines

5 ${}_i G_k^{insensitive}(g_k^j)$ = probability of abundance value g_k^j from the gaussian density fitted to the histogram of the gene k abundances over the insensitive cell lines when subjected to drug i .

$${}_i G_k^{insensitive}(g_k^j) = \frac{1}{\sigma_k^{insen} \sqrt{2\pi}} e^{-(g_k^j - \mu_k^{insen})^2 / 2(\sigma_k^{insen})^2},$$

where

μ_k^{insen} = mean of gene k abundances in the insensitive cell lines

10 σ_k^{insen} = standard deviation of gene k abundances in the insensitive cell lines

Sample parameters for QDA1 analysis on the NCI60 dataset are:

Rule 1

15 Gene: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW)
5':AA039292 3':AA039334]

Drug: Inosine-glycodialdehyde

Parameters:

20 $\mu_k^{sen} = -0.7618, \sigma_k^{sen} = 1.57$
 $\mu_k^{insen} = 0.1878, \sigma_k^{insen} = 0.6952$
 $P(C_i^{sensitive}) = 0.1978, P(C_i^{insensitive}) = 0.8022$

Rule 2

25 Gene: SID W 470947 Human scaffold protein Pbp1 mRNA complete cds [5':AA032174
3':AA032175]

Drug: Inosine-glycodialdehyde

Parameters:

30 $\mu_k^{sen} = -0.8115, \sigma_k^{sen} = 1.161$
 $\mu_k^{insen} = 0.2001, \sigma_k^{insen} = 0.8443$
 $P(C_i^{sensitive}) = 0.1978, P(C_i^{insensitive}) = 0.8022$

Rule 3

Gene: SID W 254085 ESTs Moderately similar to synaptonemal complex protein
[M.musculus] [5':N71532 3':N22165]

5 Drug: Baker's-soluble-antifolate

Parameters:

$$\mu_k^{\text{sen}} = 0.7847, \sigma_k^{\text{sen}} = 0.6875$$

$$\mu_k^{\text{insen}} = -0.2423, \sigma_k^{\text{insen}} = 0.8722$$

$$P(C_i^{\text{sensitive}}) = 0.2361, P(C_i^{\text{insensitive}}) = 0.7639$$

10

Rule 4

Gene: THY-1 MEMBRANE GLYCOPROTEIN PRECURSOR Chr.11 [183950 (E)
5':H30297 3':H28104]

Drug: Mitozolamide

15 Parameters:

$$\mu_k^{\text{sen}} = 1.073, \sigma_k^{\text{sen}} = 1.284$$

$$\mu_k^{\text{insen}} = -0.2694, \sigma_k^{\text{insen}} = 0.6137$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

20 Rule 5

Gene: PTN Pleiotrophin (heparin binding growth factor 8 neurite growth-promoting
factor 1) Chr.7 [488801 (IW) 5':AA045053 3':AA045054]

Drug: Mitozolamide

Parameters:

25 $\mu_k^{\text{sen}} = 1.019, \sigma_k^{\text{sen}} = 1.354$

$$\mu_k^{\text{insen}} = -0.2557, \sigma_k^{\text{insen}} = 0.64$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 6

30 Gene: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J WARNING
ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = -1.008, \sigma_k^{\text{sen}} = 0.5668$$

$$\mu_k^{\text{insen}} = 0.2536, \sigma_k^{\text{insen}} = 0.9027$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

5 Rule 7

Gene: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW)

5':AA039292 3':AA039334]

Drug: Cyclodisone

Parameters:

$$10 \quad \mu_k^{\text{sen}} = 0.6598, \sigma_k^{\text{sen}} = 0.2562$$

$$\mu_k^{\text{insen}} = -0.1341, \sigma_k^{\text{insen}} = 1.038$$

$$P(C_i^{\text{sensitive}}) = 0.1689, P(C_i^{\text{insensitive}}) = 0.8311$$

Rule 8

15 Gene: SID W 488387 Exostoses (multiple) 2 [5':AA046786 3':AA046656]

Drug: Cyclodisone

Parameters:

$$\mu_k^{\text{sen}} = 1.043, \sigma_k^{\text{sen}} = 1.087$$

$$\mu_k^{\text{insen}} = -0.2128, \sigma_k^{\text{insen}} = 0.8262$$

$$20 \quad P(C_i^{\text{sensitive}}) = 0.1689, P(C_i^{\text{insensitive}}) = 0.8311$$

Rule 9

Gene: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528
3':AA043529]

25 Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = 1.184, \sigma_k^{\text{sen}} = 0.9042$$

$$\mu_k^{\text{insen}} = -0.2817, \sigma_k^{\text{insen}} = 0.7835$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

30

Rule 10

Gene: PTN Pleiotrophin (heparin binding growth factor 8 neurite growth-promoting

factor 1) Chr.7 [488801 (IW) 5':AA045053 3':AA045054]

Drug: Clomesone

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 1.14, \quad \sigma_k^{\text{sen}} = 1.31 \\ 5 \quad \mu_k^{\text{insen}} &= -0.2703, \quad \sigma_k^{\text{insen}} = 0.636 \\ P(C_i^{\text{sensitive}}) &= 0.1917, \quad P(C_i^{\text{insensitive}}) = 0.8083 \end{aligned}$$

Rule 11

Gene: THY-1 MEMBRANE GLYCOPROTEIN PRECURSOR Chr.11 [183950 (E)

10 5':H30297 3':H28104]

Drug: Clomesone

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 1.157, \quad \sigma_k^{\text{sen}} = 1.312 \\ \mu_k^{\text{insen}} &= -0.2746, \quad \sigma_k^{\text{insen}} = 0.6219 \\ 15 \quad P(C_i^{\text{sensitive}}) &= 0.1917, \quad P(C_i^{\text{insensitive}}) = 0.8083 \end{aligned}$$

Rule 12

Gene: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528
3':AA043529]

20 Drug: PCNU

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 1.081, \quad \sigma_k^{\text{sen}} = 1.083 \\ \mu_k^{\text{insen}} &= -0.2435, \quad \sigma_k^{\text{insen}} = 0.7973 \\ 25 \quad P(C_i^{\text{sensitive}}) &= 0.1833, \quad P(C_i^{\text{insensitive}}) = 0.8167 \end{aligned}$$

Rule 13

Gene: SID 289361 ESTs [5':N99589 3':N92652]

Drug: Fluorouracil (5FU)

Parameters:

$$\begin{aligned} 30 \quad \mu_k^{\text{sen}} &= 0.03614, \quad \sigma_k^{\text{sen}} = 0.186 \\ \mu_k^{\text{insen}} &= -0.007432, \quad \sigma_k^{\text{insen}} = 1.074 \\ P(C_i^{\text{sensitive}}) &= 0.1628, \quad P(C_i^{\text{insensitive}}) = 0.8372 \end{aligned}$$

Rule 14

Gene: SID 287239 ESTs [5': 3':N66980]

Drug: Fluorodopan

5 Parameters:

$$\mu_k^{\text{sen}} = -0.1888, \quad \sigma_k^{\text{sen}} = 1.767$$

$$\mu_k^{\text{insen}} = 0.04924, \quad \sigma_k^{\text{insen}} = 0.6817$$

$$P(C_i^{\text{sensitive}}) = 0.2061, \quad P(C_i^{\text{insensitive}}) = 0.7939$$

10 Rule 15

Gene: SID 307717 Homo sapiens KIAA0430 mRNA complete cds [5': 3':N92942]

Drug: Cyclocytidine

Parameters:

$$\mu_k^{\text{sen}} = 0.004825, \quad \sigma_k^{\text{sen}} = 0.232$$

$$15 \quad \mu_k^{\text{insen}} = -0.002083, \quad \sigma_k^{\text{insen}} = 1.151$$

$$P(C_i^{\text{sensitive}}) = 0.2533, \quad P(C_i^{\text{insensitive}}) = 0.7467$$

Rule 16

Gene: SID W 291620 Restin (Reed-Steinberg cell-expressed intermediate filament-associated protein) [5':W03421 3':N67817]

Drug: Porfiromycin

Parameters:

$$\mu_k^{\text{sen}} = 0.9491, \quad \sigma_k^{\text{sen}} = 0.8827$$

$$\mu_k^{\text{insen}} = -0.2431, \quad \sigma_k^{\text{insen}} = 0.8715$$

$$25 \quad P(C_i^{\text{sensitive}}) = 0.2039, \quad P(C_i^{\text{insensitive}}) = 0.7961$$

Rule 17

Gene: Human extracellular protein (S1-5) mRNA complete cds Chr.2 [485875 (EW) 5':AA040442 3':AA040443]

30 Drug: Oxanthrazole (piroxantrone)

Parameters:

$$\mu_k^{\text{sen}} = 1.155, \quad \sigma_k^{\text{sen}} = 0.8967$$

$$\mu_k^{\text{insen}} = -0.2805, \quad \sigma_k^{\text{insen}} = 0.7438$$

$$P(C_i^{\text{sensitive}}) = 0.1956, \quad P(C_i^{\text{insensitive}}) = 0.8044$$

Rule 18

- 5 Gene: Human extracellular protein (S1-5) mRNA complete cds Chr.2 [485875 (EW)
5':AA040442 3':AA040443]

Drug: Anthrapyrazole-derivative

Parameters:

$$\mu_k^{\text{sen}} = 1.016, \quad \sigma_k^{\text{sen}} = 1.089$$

$$10 \quad \mu_k^{\text{insen}} = -0.2548, \quad \sigma_k^{\text{insen}} = 0.7749$$

$$P(C_i^{\text{sensitive}}) = 0.2006, \quad P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 19

Gene: SID 229535 [5':H66594 3':H66595]

- 15 Drug: Teniposide

Parameters:

$$\mu_k^{\text{sen}} = -0.9209, \quad \sigma_k^{\text{sen}} = 1.487$$

$$\mu_k^{\text{insen}} = 0.2154, \quad \sigma_k^{\text{insen}} = 0.6755$$

$$20 \quad P(C_i^{\text{sensitive}}) = 0.1894, \quad P(C_i^{\text{insensitive}}) = 0.8106$$

Rule 20

Gene: ESTs Chr.2 [149542 (DW) 5':H00283 3':H00284]

Drug: Daunorubicin

Parameters:

$$25 \quad \mu_k^{\text{sen}} = -1.052, \quad \sigma_k^{\text{sen}} = 1.344$$

$$\mu_k^{\text{insen}} = 0.2324, \quad \sigma_k^{\text{insen}} = 0.6635$$

$$P(C_i^{\text{sensitive}}) = 0.1811, \quad P(C_i^{\text{insensitive}}) = 0.8189$$

Rule 21

- 30 Gene: AK1 Adenylate kinase 1 Chr.9 [488381 (IW) 5':AA046783 3':AA046653]

Drug: Daunorubicin

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= -0.9847, \quad \sigma_k^{\text{sen}} = 1.33 \\ \mu_k^{\text{insen}} &= 0.2169, \quad \sigma_k^{\text{insen}} = 0.6847 \\ P(C_i^{\text{sensitive}}) &= 0.1811, \quad P(C_i^{\text{insensitive}}) = 0.8189\end{aligned}$$

5 Rule 22

Gene: SID 260288 ESTs [5':H97716 3':H96798]

Drug: Daunorubicin

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= -0.9929, \quad \sigma_k^{\text{sen}} = 1.81 \\ \mu_k^{\text{insen}} &= 0.2192, \quad \sigma_k^{\text{insen}} = 0.4776 \\ P(C_i^{\text{sensitive}}) &= 0.1811, \quad P(C_i^{\text{insensitive}}) = 0.8189\end{aligned}$$

Rule 23

Gene: SID W 345683 ESTs Highly similar to INTEGRAL MEMBRANE

15 GLYCOPROTEIN GP210 PRECURSOR [Rattus norvegicus] [5':W76432 3':W72039]

Drug: Daunorubicin

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= 0.918, \quad \sigma_k^{\text{sen}} = 0.3704 \\ \mu_k^{\text{insen}} &= -0.2022, \quad \sigma_k^{\text{insen}} = 0.9271 \\ P(C_i^{\text{sensitive}}) &= 0.1811, \quad P(C_i^{\text{insensitive}}) = 0.8189\end{aligned}$$

Rule 24

Gene: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

25 Drug: CPT,10-OH

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= -0.9086, \quad \sigma_k^{\text{sen}} = 0.8266 \\ \mu_k^{\text{insen}} &= 0.2078, \quad \sigma_k^{\text{insen}} = 0.8782 \\ P(C_i^{\text{sensitive}}) &= 0.1856, \quad P(C_i^{\text{insensitive}}) = 0.8144\end{aligned}$$

30

Rule 25

Gene: SID 512355 ESTs Highly similar to SRC SUBSTRATE P80/85 PROTEINS

[Gallus gallus] [5':AA059424 3':AA057835]

Drug: CPT

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.8614, \quad \sigma_k^{\text{sen}} = 0.8019 \\ 5 \quad \mu_k^{\text{insen}} &= -0.3016, \quad \sigma_k^{\text{insen}} = 0.8633 \\ P(C_i^{\text{sensitive}}) &= 0.2594, \quad P(C_i^{\text{insensitive}}) = 0.7406 \end{aligned}$$

Rule 26

Gene: SID W 488148 H.sapiens mRNA for 3'UTR of unknown protein [5':AA057239

10 3':AA058703]

Drug: CPT

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.8224, \quad \sigma_k^{\text{sen}} = 0.5588 \\ \mu_k^{\text{insen}} &= -0.2881, \quad \sigma_k^{\text{insen}} = 0.9329 \\ 15 \quad P(C_i^{\text{sensitive}}) &= 0.2594, \quad P(C_i^{\text{insensitive}}) = 0.7406 \end{aligned}$$

Rule 27

Gene: SID W 358526 ESTs [5':W96039 3':W94821]

Drug: CPT,11-formyl (RS)

20 Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -1.055, \quad \sigma_k^{\text{sen}} = 1.241 \\ \mu_k^{\text{insen}} &= 0.2536, \quad \sigma_k^{\text{insen}} = 0.7034 \\ P(C_i^{\text{sensitive}}) &= 0.1939, \quad P(C_i^{\text{insensitive}}) = 0.8061 \end{aligned}$$

25 Rule 28

Gene: SID W 135118 GATA-binding protein 3 [5':R31441 3':R31442]

Drug: CPT,11-formyl (RS)

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.9817, \quad \sigma_k^{\text{sen}} = 1.5 \\ 30 \quad \mu_k^{\text{insen}} &= -0.2359, \quad \sigma_k^{\text{insen}} = 0.6465 \\ P(C_i^{\text{sensitive}}) &= 0.1939, \quad P(C_i^{\text{insensitive}}) = 0.8061 \end{aligned}$$

Rule 29

Gene: SID 43609 ESTs [5':H06454 3':H06184]

Drug: CPT,11-formyl (RS)

Parameters:

$$\begin{aligned}
 5 \quad & \mu_k^{\text{sen}} = 0.6312, \quad \sigma_k^{\text{sen}} = 1.498 \\
 & \mu_k^{\text{insen}} = -0.1522, \quad \sigma_k^{\text{insen}} = 0.7671 \\
 & P(C_i^{\text{sensitive}}) = 0.1939, \quad P(C_i^{\text{insensitive}}) = 0.8061
 \end{aligned}$$

Rule 30

10 Gene: ESTs Chr.16 [154654 (RW) 5':R55184 3':R55185]

Drug: CPT,11-formyl (RS)

Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = 0.874, \quad \sigma_k^{\text{sen}} = 1.247 \\
 & \mu_k^{\text{insen}} = -0.2102, \quad \sigma_k^{\text{insen}} = 0.7775 \\
 15 \quad & P(C_i^{\text{sensitive}}) = 0.1939, \quad P(C_i^{\text{insensitive}}) = 0.8061
 \end{aligned}$$

Rule 31

Gene: AK1 Adenylate kinase 1 Chr.9 [488381 (IW) 5':AA046783 3':AA046653]

Drug: Mechlorethamine

20 Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = -0.4881, \quad \sigma_k^{\text{sen}} = 1.786 \\
 & \mu_k^{\text{insen}} = 0.1157, \quad \sigma_k^{\text{insen}} = 0.6286 \\
 & P(C_i^{\text{sensitive}}) = 0.1928, \quad P(C_i^{\text{insensitive}}) = 0.8072
 \end{aligned}$$

25 Rule 32

Gene: SID 43609 ESTs [5':H06454 3':H06184]

Drug: Mechlorethamine

Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = 1.042, \quad \sigma_k^{\text{sen}} = 0.9895 \\
 30 \quad & \mu_k^{\text{insen}} = -0.2493, \quad \sigma_k^{\text{insen}} = 0.814 \\
 & P(C_i^{\text{sensitive}}) = 0.1928, \quad P(C_i^{\text{insensitive}}) = 0.8072
 \end{aligned}$$

Rule 33

Gene: SID 43609 ESTs [5':H06454 3':H06184]

Drug: Triethylenemelamine

Parameters:

$$\begin{aligned}
 5 \quad & \mu_k^{\text{sen}} = 0.6685, \quad \sigma_k^{\text{sen}} = 1.405 \\
 & \mu_k^{\text{insen}} = -0.1995, \quad \sigma_k^{\text{insen}} = 0.7269 \\
 & P(C_i^{\text{sensitive}}) = 0.2294, \quad P(C_i^{\text{insensitive}}) = 0.7706
 \end{aligned}$$

Rule 34

10 Gene: SID W 133851 ESTs [5':R28233 3':R27977]

Drug: Triethylenemelamine

Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = -0.7551, \quad \sigma_k^{\text{sen}} = 1.506 \\
 & \mu_k^{\text{insen}} = 0.2248, \quad \sigma_k^{\text{insen}} = 0.6021 \\
 15 \quad & P(C_i^{\text{sensitive}}) = 0.2294, \quad P(C_i^{\text{insensitive}}) = 0.7706
 \end{aligned}$$

Rule 35

Gene: SID 43609 ESTs [5':H06454 3':H06184]

Drug: Thiotepa

20 Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = 0.6796, \quad \sigma_k^{\text{sen}} = 1.35 \\
 & \mu_k^{\text{insen}} = -0.2073, \quad \sigma_k^{\text{insen}} = 0.728 \\
 & P(C_i^{\text{sensitive}}) = 0.2333, \quad P(C_i^{\text{insensitive}}) = 0.7667
 \end{aligned}$$

25 Rule 36

Gene: SID W 291620 Restin (Reed-Steinberg cell-expressed intermediate filament-associated protein) [5':W03421 3':N67817]

Drug: Chlorambucil

Parameters:

$$\begin{aligned}
 30 \quad & \mu_k^{\text{sen}} = -0.01776, \quad \sigma_k^{\text{sen}} = 1.597 \\
 & \mu_k^{\text{insen}} = 0.005025, \quad \sigma_k^{\text{insen}} = 0.7447 \\
 & P(C_i^{\text{sensitive}}) = 0.2206, \quad P(C_i^{\text{insensitive}}) = 0.7794
 \end{aligned}$$

Rule 37

Gene: SID W 133851 ESTs [5':R28233 3':R27977]

Drug: Chlorambucil

5 Parameters:

$$\mu_k^{\text{sen}} = -0.8278, \sigma_k^{\text{sen}} = 1.471$$

$$\mu_k^{\text{insen}} = 0.2342, \sigma_k^{\text{insen}} = 0.5941$$

$$P(C_i^{\text{sensitive}}) = 0.2206, P(C_i^{\text{insensitive}}) = 0.7794$$

10 Rule 38

Gene: SID W 510230 Homo sapiens (clone CC6) NADH-ubiquinone oxidoreductase subunit mRNA 3' end cds [5':AA053568 3':AA053557]

Drug: Geldanamycin

Parameters:

$$15 \quad \mu_k^{\text{sen}} = 0.1441, \sigma_k^{\text{sen}} = 1.609$$

$$\mu_k^{\text{insen}} = -0.03698, \sigma_k^{\text{insen}} = 0.7474$$

$$P(C_i^{\text{sensitive}}) = 0.2033, P(C_i^{\text{insensitive}}) = 0.7967$$

Rule 39

20 Gene: SID 381780 ESTs [5':AA059257 3':AA059223]

Drug: Paclitaxel---Taxol

Parameters:

$$\mu_k^{\text{sen}} = 0.1618, \sigma_k^{\text{sen}} = 0.1828$$

$$\mu_k^{\text{insen}} = -0.03218, \sigma_k^{\text{insen}} = 1.06$$

$$25 \quad P(C_i^{\text{sensitive}}) = 0.1622, P(C_i^{\text{insensitive}}) = 0.8378$$

Rule 40

Gene: H.sapiens mRNA for TRAMP protein Chr.8 [149355 (IEW) 5':H01598 3':H01495]

30 Drug: Amonafide

Parameters:

$$\mu_k^{\text{sen}} = 1.095, \sigma_k^{\text{sen}} = 1.188$$

$$\mu_k^{\text{insen}} = -0.2498, \quad \sigma_k^{\text{insen}} = 0.7473$$

$$P(C_i^{\text{sensitive}}) = 0.1861, \quad P(C_i^{\text{insensitive}}) = 0.8139$$

Linear Discriminant Analysis – 2-dimensional (LDA 2D)

- 5 This method computes a Bayesian conditional probability $P(j \in C_i^{\text{sensitive}} | g_k^j, g_l^j)$ that a cell line j is sensitive to drug i , given the abundances of genes k and l , g_k^j, g_l^j , respectively, in cell line j .

The probability is computed using the following equation:

$$10 \quad P(j \in C_i^{\text{sensitive}} | g_k^j, g_l^j) = \frac{{}_i G_{k,l}^{\text{sensitive}}(g_k^j, g_l^j) \cdot P(C_i^{\text{sensitive}})}{{}_i G_{k,l}^{\text{sensitive}}(g_k^j, g_l^j) \cdot P(C_i^{\text{sensitive}}) + {}_i G_{k,l}^{\text{insensitive}}(g_k^j, g_l^j) \cdot P(C_i^{\text{insensitive}})},$$

where

$$P(C_i^{\text{sensitive}}) = \text{prior probability of the sensitive set}$$

$$= |C_i^{\text{sensitive}}| / (|C_i^{\text{sensitive}}| + |C_i^{\text{insensitive}}|),$$

$$15 \quad P(C_i^{\text{insensitive}}) = \text{prior probability of the insensitive}$$

$$\text{set} = |C_i^{\text{insensitive}}| / (|C_i^{\text{sensitive}}| + |C_i^{\text{insensitive}}|),$$

${}_i G_{k,l}^{\text{sensitive}}(g_k^j, g_l^j)$ = joint probability of abundance values g_k^j and g_l^j from the bivariate gaussian density fitted to the histogram of gene k and l abundances over the sensitive cell lines when subjected to drug i .

$$20 \quad {}_i G_{k,l}^{\text{sensitive}}(g_k^j, g_l^j) =$$

$$\frac{1}{2\pi\sigma_k^{\text{avg}}\sigma_l^{\text{avg}}\sqrt{1-(\rho_{k,l}^{\text{avg}})^2}} \exp \left\{ - \frac{\left[\left(\frac{g_k^j - \mu_k^{\text{sen}}}{\sigma_k^{\text{avg}}} \right)^2 - 2\rho_{k,l}^{\text{avg}} \left(\frac{g_k^j - \mu_k^{\text{sen}}}{\sigma_k^{\text{avg}}} \right) \left(\frac{g_l^j - \mu_l^{\text{sen}}}{\sigma_l^{\text{avg}}} \right) + \left(\frac{g_l^j - \mu_l^{\text{sen}}}{\sigma_l^{\text{avg}}} \right)^2 \right]}{2(1-(\rho_{k,l}^{\text{avg}})^2)} \right\}$$

where

μ_k^{sen} = mean of gene k abundances over the sensitive cell lines

σ_k^{avg} = sensitive\insensitive class-weighted average standard deviation of gene k abundances in the sensitive and insensitive cell lines

μ_l^{sen} = mean of gene l abundances over the sensitive cell lines

5 σ_l^{avg} = sensitive\insensitive class-weighted average standard deviation of gene l abundances in the sensitive and insensitive cell lines

$\rho_{k,l}^{avg}$ = sensitive\insensitive class-weighted average correlation coefficient of gene k and gene l abundances in the sensitive and insensitive cell lines

10 ${}_i G_{k,l}^{insensitive}(g_k^j, g_l^j)$ = joint probability of abundance values g_k^j and g_l^j from the bivariate gaussian density fitted to the histogram of gene k and l abundances over the insensitive cell lines when subjected to drug i .

$${}_i G_{k,l}^{insensitive}(g_k^j, g_l^j) =$$

15

$$\frac{1}{2\pi\sigma_k^{avg}\sigma_l^{avg}\sqrt{1-(\rho_{k,l}^{avg})^2}} \exp \left\{ - \frac{\left[\left(\frac{g_k^j - \mu_k^{insen}}{\sigma_k^{avg}} \right)^2 - 2\rho_{k,l}^{avg} \left(\frac{g_k^j - \mu_k^{insen}}{\sigma_k^{avg}} \right) \left(\frac{g_l^j - \mu_l^{insen}}{\sigma_l^{avg}} \right) + \left(\frac{g_l^j - \mu_l^{insen}}{\sigma_l^{avg}} \right)^2 \right]}{2(1-(\rho_{k,l}^{avg})^2)} \right\}$$

where

μ_k^{insen} is the mean of gene k abundances over the insensitive cell lines

20 μ_l^{insen} is the mean of gene l abundances over the insensitive cell lines

Sample parameters for the LDA 2D analysis on the NCI60 dataset are:

Rule 1

25 Gene 1: Glyoxalase-I-log

Gene 2: Homo sapiens mRNA for HYA22 complete cds Chr.3 [358957 (EW)

5':W91969 3':W94916]

Drug: Acivicin

Parameters:

$$\begin{aligned}
 5 \quad & \mu_k^{\text{sen}} = -0.9056, \quad \mu_l^{\text{sen}} = 0.3517 \\
 & \mu_k^{\text{insen}} = 0.2197, \quad \mu_l^{\text{insen}} = -0.08527 \\
 & \sigma_k^{\text{avg}} = 0.8751, \quad \sigma_l^{\text{avg}} = 0.9817, \quad \rho_{k,l}^{\text{avg}} = 0.531 \\
 & P(C_i^{\text{sensitive}}) = 0.1956, \quad P(C_i^{\text{insensitive}}) = 0.8044
 \end{aligned}$$

10 Rule 2

Gene 1: SID W 254085 ESTs Moderately similar to synaptonemal complex protein
[M.musculus] [5':N71532 3':N22165]

Gene 2: SID 118593 [5':T92821 3':T92741]

Drug: Baker's-soluble-antifolate

15 Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = 0.7847, \quad \mu_l^{\text{sen}} = -0.5796 \\
 & \mu_k^{\text{insen}} = -0.2423, \quad \mu_l^{\text{insen}} = 0.1796 \\
 & \sigma_k^{\text{avg}} = 0.8539, \quad \sigma_l^{\text{avg}} = 0.8599, \quad \rho_{k,l}^{\text{avg}} = 0.2493 \\
 & P(C_i^{\text{sensitive}}) = 0.2361, \quad P(C_i^{\text{insensitive}}) = 0.7639
 \end{aligned}$$

20

Rule 3

Gene 1: SID W 254085 ESTs Moderately similar to synaptonemal complex protein
[M.musculus] [5':N71532 3':N22165]

Gene 2: ESTs Chr.5 [46694 (RW) 5':H10240 3':H10192]

25 Drug: Baker's-soluble-antifolate

Parameters:

$$\begin{aligned}
 & \mu_k^{\text{sen}} = 0.7847, \quad \mu_l^{\text{sen}} = 0.4403 \\
 & \mu_k^{\text{insen}} = -0.2423, \quad \mu_l^{\text{insen}} = -0.1363 \\
 & \sigma_k^{\text{avg}} = 0.8539, \quad \sigma_l^{\text{avg}} = 0.9706, \quad \rho_{k,l}^{\text{avg}} = -0.1844 \\
 30 \quad & P(C_i^{\text{sensitive}}) = 0.2361, \quad P(C_i^{\text{insensitive}}) = 0.7639
 \end{aligned}$$

Rule 4

Gene 1: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Gene 2: *Hs.648 Cut (Drosophila)-like 1 (CCAAT displacement protein) SID W 26677
ESTs [5':R13994 3':R39117]

5 Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = -1.008, \mu_l^{\text{sen}} = 0.8138$$

$$\mu_k^{\text{insen}} = 0.2536, \mu_l^{\text{insen}} = -0.2039$$

$$\sigma_k^{\text{avg}} = 0.8681, \sigma_l^{\text{avg}} = 0.9103, \rho_{k,l}^{\text{avg}} = 0.07755$$

$$10 \quad P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 5

Gene 1: Homo sapiens delta7-sterol reductase mRNA complete cds Chr.10 [417125 (E)
5': 3':W87472]

15 Gene 2: SID W 380674 ESTs [5':AA053720 3':AA053711]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = -0.7211, \mu_l^{\text{sen}} = 1.093$$

$$\mu_k^{\text{insen}} = 0.1813, \mu_l^{\text{insen}} = -0.2739$$

$$20 \quad \sigma_k^{\text{avg}} = 0.9411, \sigma_l^{\text{avg}} = 0.8441, \rho_{k,l}^{\text{avg}} = 0.1253$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 6

Gene 1: Glutathione S-Transferase Pi-log

25 Gene 2: *Hs.648 Cut (Drosophila)-like 1 (CCAAT displacement protein) SID W 26677
ESTs [5':R13994 3':R39117]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = -0.917, \mu_l^{\text{sen}} = 0.8138$$

$$30 \quad \mu_k^{\text{insen}} = 0.2307, \mu_l^{\text{insen}} = -0.2039$$

$$\sigma_k^{\text{avg}} = 0.8411, \sigma_l^{\text{avg}} = 0.9103, \rho_{k,l}^{\text{avg}} = 0.04772$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 7

Gene 1: ESTs Chr.X [48536 (E) 5':H14669 3':H14579]

Gene 2: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

5 WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: Clomesone

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.8957, \mu_l^{\text{sen}} = -1.079 \\ \mu_k^{\text{insen}} &= 0.2117, \mu_l^{\text{insen}} = 0.2564 \\ 10 \quad \sigma_k^{\text{avg}} &= 0.8904, \sigma_l^{\text{avg}} = 0.8587, \rho_{k,l}^{\text{avg}} = -0.165 \\ P(C_i^{\text{sensitive}}) &= 0.1917, P(C_i^{\text{insensitive}}) = 0.8083 \end{aligned}$$

Rule 8

Gene 1: SID W 36809 Homo sapiens neural cell adhesion molecule (CALL) mRNA

15 complete cds [5':R34648 3':R49177]

Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

Drug: Clomesone

Parameters:

$$\begin{aligned} 20 \quad \mu_k^{\text{sen}} &= 0.6335, \mu_l^{\text{sen}} = 1.184 \\ \mu_k^{\text{insen}} &= -0.1498, \mu_l^{\text{insen}} = -0.2817 \\ \sigma_k^{\text{avg}} &= 0.9603, \sigma_l^{\text{avg}} = 0.829, \rho_{k,l}^{\text{avg}} = -0.2448 \\ P(C_i^{\text{sensitive}}) &= 0.1917, P(C_i^{\text{insensitive}}) = 0.8083 \end{aligned}$$

25 Rule 9

Gene 1: M-PHASE INDUCER PHOSPHATASE 2 Chr.20 [179373 (EW) 5':H50437 3':H50438]

Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

30 Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = 0.3874, \mu_l^{\text{sen}} = 1.184$$

$$\begin{aligned}\mu_k^{\text{insen}} &= -0.09229, \mu_l^{\text{insen}} = -0.2817 \\ \sigma_k^{\text{avg}} &= 0.9766, \sigma_l^{\text{avg}} = 0.829, \rho_{k,l}^{\text{avg}} = -0.2704 \\ P(C_i^{\text{sensitive}}) &= 0.1917, P(C_i^{\text{insensitive}}) = 0.8083\end{aligned}$$

5 Rule 10

Gene 1: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Gene 2: SID 469842 Homo sapiens mRNA for fatty acid binding protein complete cds
[5':AA029794 3':AA029795]

10 Drug: Clomesone

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= -1.079, \mu_l^{\text{sen}} = 0.8757 \\ \mu_k^{\text{insen}} &= 0.2564, \mu_l^{\text{insen}} = -0.2074 \\ \sigma_k^{\text{avg}} &= 0.8587, \sigma_l^{\text{avg}} = 0.9151, \rho_{k,l}^{\text{avg}} = 0.1636 \\ P(C_i^{\text{sensitive}}) &= 0.1917, P(C_i^{\text{insensitive}}) = 0.8083\end{aligned}$$

15

Rule 11

Gene 1: ESTsSID 327435 [5':W32467 3':W19830]

Gene 2: SID 469842 Homo sapiens mRNA for fatty acid binding protein complete cds

20 [5':AA029794 3':AA029795]

Drug: Clomesone

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= -0.793, \mu_l^{\text{sen}} = 0.8757 \\ \mu_k^{\text{insen}} &= 0.1878, \mu_l^{\text{insen}} = -0.2074 \\ \sigma_k^{\text{avg}} &= 0.9388, \sigma_l^{\text{avg}} = 0.9151, \rho_{k,l}^{\text{avg}} = 0.4476 \\ P(C_i^{\text{sensitive}}) &= 0.1917, P(C_i^{\text{insensitive}}) = 0.8083\end{aligned}$$

25

Rule 12

Gene 1: SID 512164 Human clathrin assembly protein 50 (AP50) mRNA complete cds

30 [5': 3':AA057396]

Gene 2: SID W 345624 Human homeobox protein (PHOX1) mRNA 3' end [5':W76402
3':W72050]

Drug: Clomesone

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.8248, \quad \mu_l^{\text{sen}} = -0.253 \\ \mu_k^{\text{insen}} &= -0.1956, \quad \mu_l^{\text{insen}} = 0.06021 \\ 5 \quad \sigma_k^{\text{avg}} &= 0.9014, \quad \sigma_l^{\text{avg}} = 1.015, \quad \rho_{k,l}^{\text{avg}} = 0.72 \\ P(C_i^{\text{sensitive}}) &= 0.1917, \quad P(C_i^{\text{insensitive}}) = 0.8083 \end{aligned}$$

Rule 13

Gene 1: SID W 376951 ESTs [5':AA047756 3':AA047641]

10 Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

Drug: Clomesone

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.8665, \quad \mu_l^{\text{sen}} = 1.184 \\ 15 \quad \mu_k^{\text{insen}} &= -0.2063, \quad \mu_l^{\text{insen}} = -0.2817 \\ \sigma_k^{\text{avg}} &= 0.9396, \quad \sigma_l^{\text{avg}} = 0.829, \quad \rho_{k,l}^{\text{avg}} = 0.1106 \\ P(C_i^{\text{sensitive}}) &= 0.1917, \quad P(C_i^{\text{insensitive}}) = 0.8083 \end{aligned}$$

Rule 14

20 Gene 1: Glutathione S-Transferase Pi-log

Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

Drug: Clomesone

Parameters:

$$\begin{aligned} 25 \quad \mu_k^{\text{sen}} &= -0.8961, \quad \mu_l^{\text{sen}} = 1.184 \\ \mu_k^{\text{insen}} &= 0.2131, \quad \mu_l^{\text{insen}} = -0.2817 \\ \sigma_k^{\text{avg}} &= 0.8991, \quad \sigma_l^{\text{avg}} = 0.829, \quad \rho_{k,l}^{\text{avg}} = 0.1075 \\ P(C_i^{\text{sensitive}}) &= 0.1917, \quad P(C_i^{\text{insensitive}}) = 0.8083 \end{aligned}$$

30 Rule 15

Gene 1: XRCC4 DNA repair protein XRCC4 Chr.5 [26811 (RW) 5':R14027 3':R39148]

Gene 2: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: Clomesone

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.583, \quad \mu_l^{\text{sen}} = -1.079 \\ 5 \quad \mu_k^{\text{insen}} &= 0.1387, \quad \mu_l^{\text{insen}} = 0.2564 \\ \sigma_k^{\text{avg}} &= 0.9879, \quad \sigma_l^{\text{avg}} = 0.8587, \quad \rho_{k,l}^{\text{avg}} = -0.3373 \\ P(C_i^{\text{sensitive}}) &= 0.1917, \quad P(C_i^{\text{insensitive}}) = 0.8083 \end{aligned}$$

Rule 16

10 Gene 1: Homo sapiens clone 24711 mRNA sequence Chr.2 [345084 (IW) 5':W76362 3':W72306]

Gene 2: *Homo sapiens lysosomal neuraminidase precursor mRNA complete cds SID W 487887 Hexabrachion (tenascin C cytotactin) [5':AA046543 3':AA045473]

Drug: Clomesone

15 Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.5805, \quad \mu_l^{\text{sen}} = 0.8678 \\ \mu_k^{\text{insen}} &= 0.137, \quad \mu_l^{\text{insen}} = -0.2056 \\ \sigma_k^{\text{avg}} &= 0.968, \quad \sigma_l^{\text{avg}} = 0.911, \quad \rho_{k,l}^{\text{avg}} = 0.5627 \\ 20 \quad P(C_i^{\text{sensitive}}) &= 0.1917, \quad P(C_i^{\text{insensitive}}) = 0.8083 \end{aligned}$$

Rule 17

Gene 1: SID 260048 Homo sapiens intermediate conductance calcium-activated potassium channel (hKCa4) mRNA complete [5': 3':N32010]

25 Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

Drug: Clomesone

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.3774, \quad \mu_l^{\text{sen}} = 1.184 \\ \mu_k^{\text{insen}} &= -0.09052, \quad \mu_l^{\text{insen}} = -0.2817 \\ 30 \quad \sigma_k^{\text{avg}} &= 1.015, \quad \sigma_l^{\text{avg}} = 0.829, \quad \rho_{k,l}^{\text{avg}} = -0.2375 \\ P(C_i^{\text{sensitive}}) &= 0.1917, \quad P(C_i^{\text{insensitive}}) = 0.8083 \end{aligned}$$

Rule 18

Gene 1: ESTs Weakly similar to R06B9.b [C.elegans] Chr.1 [365488 (IW)

5':AA009557 3':AA009558]

Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528

5 3':AA043529]

Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = 0.6026, \mu_l^{\text{sen}} = 1.184$$

$$\mu_k^{\text{insen}} = -0.1433, \mu_l^{\text{insen}} = -0.2817$$

$$10 \quad \sigma_k^{\text{avg}} = 0.9451, \sigma_l^{\text{avg}} = 0.829, \rho_{k,l}^{\text{avg}} = -0.0427$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

Rule 19

Gene 1: ESTs Moderately similar to DUAL SPECIFICITY PROTEIN PHOSPHATASE

15 VHR [H.sapiens] Chr.17 [49293 (E) 5':H15616 3':H15557]

Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528

3':AA043529]

Drug: Clomesone

Parameters:

$$20 \quad \mu_k^{\text{sen}} = -0.1122, \mu_l^{\text{sen}} = 1.184$$

$$\mu_k^{\text{insen}} = 0.02618, \mu_l^{\text{insen}} = -0.2817$$

$$\sigma_k^{\text{avg}} = 1.019, \sigma_l^{\text{avg}} = 0.829, \rho_{k,l}^{\text{avg}} = 0.4234$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

25 Rule 20

Gene 1: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528

3':AA043529]

30 Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = -1.079, \mu_l^{\text{sen}} = 1.184$$

$$\begin{aligned}\mu_k^{\text{insen}} &= 0.2564, \quad \mu_l^{\text{insen}} = -0.2817 \\ \sigma_k^{\text{avg}} &= 0.8587, \quad \sigma_l^{\text{avg}} = 0.829, \quad \rho_{k,l}^{\text{avg}} = 0.02375 \\ P(C_i^{\text{sensitive}}) &= 0.1917, \quad P(C_i^{\text{insensitive}}) = 0.8083\end{aligned}$$

5 Rule 21

Gene 1: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

Gene 2: ESTs Chr.6 [144805 (EW) 5':R76279 3':R76556]

Drug: Clomesone

10 Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= 1.184, \quad \mu_l^{\text{sen}} = 0.4822 \\ \mu_k^{\text{insen}} &= -0.2817, \quad \mu_l^{\text{insen}} = -0.1143 \\ \sigma_k^{\text{avg}} &= 0.829, \quad \sigma_l^{\text{avg}} = 0.9949, \quad \rho_{k,l}^{\text{avg}} = -0.2002 \\ P(C_i^{\text{sensitive}}) &= 0.1917, \quad P(C_i^{\text{insensitive}}) = 0.8083\end{aligned}$$

15

Rule 22

Gene 1: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

Gene 2: SID W 488333 ESTs [5':AA046755 3':AA046642]

20 Drug: Clomesone

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= 1.184, \quad \mu_l^{\text{sen}} = -0.1604 \\ \mu_k^{\text{insen}} &= -0.2817, \quad \mu_l^{\text{insen}} = 0.03825 \\ \sigma_k^{\text{avg}} &= 0.829, \quad \sigma_l^{\text{avg}} = 1.011, \quad \rho_{k,l}^{\text{avg}} = 0.3461 \\ P(C_i^{\text{sensitive}}) &= 0.1917, \quad P(C_i^{\text{insensitive}}) = 0.8083\end{aligned}$$

25

Rule 23

Gene 1: ANX3 Annexin III (lipocortin III) Chr.4 [328683 (IW) 5':W40286 3':W45327]

Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

30

Drug: Clomesone

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= -0.7239, \mu_l^{\text{sen}} = 1.184 \\ \mu_k^{\text{insen}} &= 0.1714, \mu_l^{\text{insen}} = -0.2817 \\ \sigma_k^{\text{avg}} &= 0.9663, \sigma_l^{\text{avg}} = 0.829, \rho_{k,l}^{\text{avg}} = -0.1129 \\ P(C_i^{\text{sensitive}}) &= 0.1917, P(C_i^{\text{insensitive}}) = 0.8083\end{aligned}$$

5

Rule 24

Gene 1: SID 308729 ESTs [5':W25229 3':N95389]

Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

10 Drug: Clomesone

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= -0.6074, \mu_l^{\text{sen}} = 1.184 \\ \mu_k^{\text{insen}} &= 0.1438, \mu_l^{\text{insen}} = -0.2817 \\ \sigma_k^{\text{avg}} &= 0.9876, \sigma_l^{\text{avg}} = 0.829, \rho_{k,l}^{\text{avg}} = 0.1155 \\ P(C_i^{\text{sensitive}}) &= 0.1917, P(C_i^{\text{insensitive}}) = 0.8083\end{aligned}$$

15

Rule 25

Gene 1: Metallothionein content-log

Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

20

Drug: Clomesone

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= 0.5109, \mu_l^{\text{sen}} = 1.184 \\ \mu_k^{\text{insen}} &= -0.1211, \mu_l^{\text{insen}} = -0.2817 \\ \sigma_k^{\text{avg}} &= 0.9435, \sigma_l^{\text{avg}} = 0.829, \rho_{k,l}^{\text{avg}} = -0.3179 \\ P(C_i^{\text{sensitive}}) &= 0.1917, P(C_i^{\text{insensitive}}) = 0.8083\end{aligned}$$

25

Rule 26

Gene 1: ESTs Chr.14 [160605 (E) 5':H25013 3':H25014]

Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

30

Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = -0.7174, \mu_l^{\text{sen}} = 1.184$$

$$\mu_k^{\text{insen}} = 0.1703, \mu_l^{\text{insen}} = -0.2817$$

$$\sigma_k^{\text{avg}} = 0.9506, \sigma_l^{\text{avg}} = 0.829, \rho_{k,l}^{\text{avg}} = 0.01308$$

$$5 \quad P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

Rule 27

Gene 1: SID W 510534 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED PROTEIN GA733-2 PRECURSOR [5':AA055858 3':AA055808]

10 Gene 2: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = -0.867, \mu_l^{\text{sen}} = -1.079$$

$$15 \quad \mu_k^{\text{insen}} = 0.2052, \mu_l^{\text{insen}} = 0.2564$$

$$\sigma_k^{\text{avg}} = 0.9304, \sigma_l^{\text{avg}} = 0.8587, \rho_{k,l}^{\text{avg}} = -0.08247$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

Rule 28

20 Gene 1: SID W 489262 Allograft inflammatory factor 1 [5':AA045718 3':AA045719]

Gene 2: SID W 489301 ESTs [5':AA054471 3':AA058511]

Drug: PCNU

Parameters:

$$\mu_k^{\text{sen}} = -0.1844, \mu_l^{\text{sen}} = 0.7991$$

$$25 \quad \mu_k^{\text{insen}} = 0.04227, \mu_l^{\text{insen}} = -0.1796$$

$$\sigma_k^{\text{avg}} = 0.9895, \sigma_l^{\text{avg}} = 0.9465, \rho_{k,l}^{\text{avg}} = 0.7317$$

$$P(C_i^{\text{sensitive}}) = 0.1833, P(C_i^{\text{insensitive}}) = 0.8167$$

Rule 29

30 Gene 1: p53 mutation-log

Gene 2: SID 43555 MALATE OXIDOREDUCTASE [5':H13370 3':H06037]

Drug: Fluorouracil (5FU)

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.9274, \mu_l^{\text{sen}} = 0.9686 \\ \mu_k^{\text{insen}} &= -0.1772, \mu_l^{\text{insen}} = -0.1883 \\ \sigma_k^{\text{avg}} &= 0.899, \sigma_l^{\text{avg}} = 0.9219, \rho_{k,l}^{\text{avg}} = -0.186 \\ 5 \quad P(C_i^{\text{sensitive}}) &= 0.1628, P(C_i^{\text{insensitive}}) = 0.8372 \end{aligned}$$

Rule 30

Gene 1: ME2 Malic enzyme 2 mitochondrial Chr.18 [109375 (IW) 5':T80865
3':T70290]

10 Gene 2: SID W 488806 Thioredoxin [5':AA045051 3':AA045052]

Drug: Asaley

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.7873, \mu_l^{\text{sen}} = -0.922 \\ \mu_k^{\text{insen}} &= -0.182, \mu_l^{\text{insen}} = 0.2136 \\ 15 \quad \sigma_k^{\text{avg}} &= 0.9409, \sigma_l^{\text{avg}} = 0.9102, \rho_{k,l}^{\text{avg}} = 0.3849 \\ P(C_i^{\text{sensitive}}) &= 0.1878, P(C_i^{\text{insensitive}}) = 0.8122 \end{aligned}$$

Rule 31

Gene 1: X-ray induction of mdm2-log

20 Gene 2: Human thymosin beta-4 mRNA complete cds Chr.20 [305890 (IW) 5':W19923
3':N91268]

Drug: Cytarabine (araC)

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.5649, \mu_l^{\text{sen}} = -0.7694 \\ 25 \quad \mu_k^{\text{insen}} &= -0.2054, \mu_l^{\text{insen}} = 0.2788 \\ \sigma_k^{\text{avg}} &= 0.8243, \sigma_l^{\text{avg}} = 0.8663, \rho_{k,l}^{\text{avg}} = 0.2969 \\ P(C_i^{\text{sensitive}}) &= 0.2661, P(C_i^{\text{insensitive}}) = 0.7339 \end{aligned}$$

Rule 32

30 Gene 1: *EST H49897 SID 429460 ESTs [5': 3':AA007629]

Gene 2: TXNRD1 Thioredoxin reductase Chr.12 [510377 (IW) 5':AA055407
3':AA055408]

Drug: Anthrapyrazole-derivative

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.8238, \quad \mu_l^{\text{sen}} = 0.8618 \\ \mu_k^{\text{insen}} &= 0.2071, \quad \mu_l^{\text{insen}} = -0.2166 \\ 5 \quad \sigma_k^{\text{avg}} &= 0.934, \quad \sigma_l^{\text{avg}} = 0.9084, \quad \rho_{k,l}^{\text{avg}} = 0.2681 \\ P(C_i^{\text{sensitive}}) &= 0.2006, \quad P(C_i^{\text{insensitive}}) = 0.7994 \end{aligned}$$

Rule 33

Gene 1: PTN Pleiotrophin (heparin binding growth factor 8 neurite growth-promoting factor 1) Chr.7 [488801 (IW) 5':AA045053 3':AA045054]

Gene 2: TXNRD1 Thioredoxin reductase Chr.12 [510377 (IW) 5':AA055407 3':AA055408]

Drug: Anthrapyrazole-derivative

Parameters:

$$\begin{aligned} 15 \quad \mu_k^{\text{sen}} &= 0.8876, \quad \mu_l^{\text{sen}} = 0.8618 \\ \mu_k^{\text{insen}} &= -0.2227, \quad \mu_l^{\text{insen}} = -0.2166 \\ \sigma_k^{\text{avg}} &= 0.8932, \quad \sigma_l^{\text{avg}} = 0.9084, \quad \rho_{k,l}^{\text{avg}} = -0.3478 \\ P(C_i^{\text{sensitive}}) &= 0.2006, \quad P(C_i^{\text{insensitive}}) = 0.7994 \end{aligned}$$

20 Rule 34

Gene 1: SID W 345683 ESTs Highly similar to INTEGRAL MEMBRANE GLYCOPROTEIN GP210 PRECURSOR [Rattus norvegicus] [5':W76432 3':W72039]

Gene 2: ESTs Chr.5 [322749 (I) 5': 3':W15473]

Drug: Daunorubicin

25 Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.918, \quad \mu_l^{\text{sen}} = -0.7006 \\ \mu_k^{\text{insen}} &= -0.2022, \quad \mu_l^{\text{insen}} = 0.1549 \\ \sigma_k^{\text{avg}} &= 0.8758, \quad \sigma_l^{\text{avg}} = 0.9296, \quad \rho_{k,l}^{\text{avg}} = 0.2797 \\ P(C_i^{\text{sensitive}}) &= 0.1811, \quad P(C_i^{\text{insensitive}}) = 0.8189 \end{aligned}$$

30

Rule 35

Gene 1: L-LACTATE DEHYDROGENASE M CHAIN Chr.11 [510595 (IW)

5':AA057759 3':AA057760]

Gene 2: Homo sapiens T245 protein (T245) mRNA complete cds Chr.X [343063 (IW)

5':W67989 3':W68001]

Drug: Daunorubicin

5 Parameters:

$$\mu_k^{\text{sen}} = -0.7199, \mu_l^{\text{sen}} = -1.061$$

$$\mu_k^{\text{insen}} = 0.1588, \mu_l^{\text{insen}} = 0.234$$

$$\sigma_k^{\text{avg}} = 0.9279, \sigma_l^{\text{avg}} = 0.8647, \rho_{k,l}^{\text{avg}} = -0.2833$$

$$P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

10

Rule 36

Gene 1: SID W 345683 ESTs Highly similar to INTEGRAL MEMBRANE

GLYCOPROTEIN GP210 PRECURSOR [Rattus norvegicus] [5':W76432 3':W72039]

Gene 2: SID W 510534 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED

15 PROTEIN GA733-2 PRECURSOR [5':AA055858 3':AA055808]

Drug: Daunorubicin

Parameters:

$$\mu_k^{\text{sen}} = 0.918, \mu_l^{\text{sen}} = -0.437$$

$$\mu_k^{\text{insen}} = -0.2022, \mu_l^{\text{insen}} = 0.09623$$

20 $\sigma_k^{\text{avg}} = 0.8758, \sigma_l^{\text{avg}} = 0.9836, \rho_{k,l}^{\text{avg}} = 0.525$

$$P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

Rule 37

Gene 1: ESTs Chr.2 [149542 (DW) 5':H00283 3':H00284]

25 Gene 2: ESTsSID 429074 [5':AA005275 3':AA005169]

Drug: Daunorubicin

Parameters:

$$\mu_k^{\text{sen}} = -1.052, \mu_l^{\text{sen}} = -0.6467$$

$$\mu_k^{\text{insen}} = 0.2324, \mu_l^{\text{insen}} = 0.1424$$

30 $\sigma_k^{\text{avg}} = 0.8508, \sigma_l^{\text{avg}} = 0.9537, \rho_{k,l}^{\text{avg}} = 0.06255$

$$P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

Rule 38

Gene 1: SID W 345683 ESTs Highly similar to INTEGRAL MEMBRANE
GLYCOPROTEIN GP210 PRECURSOR [Rattus norvegicus] [5':W76432 3':W72039]

Gene 2: Human clone 23933 mRNA sequence Chr.17 [23933 (IW) 5':T77288

5 3':R39465]

Drug: Daunorubicin

Parameters:

$$\mu_k^{\text{sen}} = 0.918, \mu_l^{\text{sen}} = 0.4489$$

$$\mu_k^{\text{insen}} = -0.2022, \mu_l^{\text{insen}} = -0.09989$$

$$10 \quad \sigma_k^{\text{avg}} = 0.8758, \sigma_l^{\text{avg}} = 1.004, \rho_{k,l}^{\text{avg}} = -0.5196$$

$$P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

Rule 39

Gene 1: GRL Glucocorticoid receptor Chr.5 [262691 (E) 5': 3':H99414]

15 Gene 2: *Prothymosin alpha SID W 271976 AMINOACYLASE-1 [5':N44687
3':N35315]

Drug: Daunorubicin

Parameters:

$$\mu_k^{\text{sen}} = 0.3732, \mu_l^{\text{sen}} = -1.032$$

$$20 \quad \mu_k^{\text{insen}} = -0.08233, \mu_l^{\text{insen}} = 0.2284$$

$$\sigma_k^{\text{avg}} = 0.9501, \sigma_l^{\text{avg}} = 0.858, \rho_{k,l}^{\text{avg}} = 0.3514$$

$$P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

Rule 40

25 Gene 1: *Prothymosin alpha SID W 271976 AMINOACYLASE-1 [5':N44687
3':N35315]

Gene 2: PLAUR Plasminogen activator urokinase receptor Chr.19 [325077 (DIW)
5':W49705 3':W49706]

Drug: Daunorubicin

30 Parameters:

$$\mu_k^{\text{sen}} = -1.032, \mu_l^{\text{sen}} = 0.1522$$

$$\mu_k^{\text{insen}} = 0.2284, \mu_l^{\text{insen}} = -0.03346$$

$$\sigma_k^{avg} = 0.858, \quad \sigma_l^{avg} = 0.9987, \quad \rho_{k,l}^{avg} = 0.5897$$

$$P(C_i^{sensitive}) = 0.1811, \quad P(C_i^{insensitive}) = 0.8189$$

Rule 41

5 Gene 1: ESTs Chr.2 [149542 (DW) 5':H00283 3':H00284]

Gene 2: ESTs Chr.2 [365120 (IW) 5':AA025204 3':AA025124]

Drug: Daunorubicin

Parameters:

$$\mu_k^{sen} = -1.052, \quad \mu_l^{sen} = 0.2085$$

$$10 \quad \mu_k^{insen} = 0.2324, \quad \mu_l^{insen} = -0.04633$$

$$\sigma_k^{avg} = 0.8508, \quad \sigma_l^{avg} = 1.018, \quad \rho_{k,l}^{avg} = 0.376$$

$$P(C_i^{sensitive}) = 0.1811, \quad P(C_i^{insensitive}) = 0.8189$$

Rule 42

15 Gene 1: ESTs Chr.2 [149542 (DW) 5':H00283 3':H00284]

Gene 2: Ribosomal protein L17SID 60561 [5':T39375 3':T40540]

Drug: Daunorubicin

Parameters:

$$\mu_k^{sen} = -1.052, \quad \mu_l^{sen} = -0.5213$$

$$20 \quad \mu_k^{insen} = 0.2324, \quad \mu_l^{insen} = 0.1147$$

$$\sigma_k^{avg} = 0.8508, \quad \sigma_l^{avg} = 0.9713, \quad \rho_{k,l}^{avg} = -0.2356$$

$$P(C_i^{sensitive}) = 0.1811, \quad P(C_i^{insensitive}) = 0.8189$$

Rule 43

25 Gene 1: ESTs Chr.2 [149542 (DW) 5':H00283 3':H00284]

Gene 2: Glutathione S-Transferase M1a-log

Drug: Daunorubicin

Parameters:

$$\mu_k^{sen} = -1.052, \quad \mu_l^{sen} = 0.1809$$

$$30 \quad \mu_k^{insen} = 0.2324, \quad \mu_l^{insen} = -0.03737$$

$$\sigma_k^{avg} = 0.8508, \quad \sigma_l^{avg} = 1.033, \quad \rho_{k,l}^{avg} = 0.1657$$

$$P(C_i^{sensitive}) = 0.1811, \quad P(C_i^{insensitive}) = 0.8189$$

Rule 44

Gene 1: SID 260288 ESTs [5':H97716 3':H96798]

Gene 2: SID W 358185 Human mitochondrial 2.4-dienoyl-CoA reductase mRNA

5 complete cds [5':W95455 3':W95406]

Drug: Daunorubicin

Parameters:

$$\mu_k^{\text{sen}} = -0.9929, \mu_l^{\text{sen}} = -0.5507$$

$$\mu_k^{\text{insen}} = 0.2192, \mu_l^{\text{insen}} = 0.1224$$

$$10 \quad \sigma_k^{\text{avg}} = 0.9063, \sigma_l^{\text{avg}} = 0.9734, \rho_{k,l}^{\text{avg}} = -0.4799$$

$$P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

Rule 45

Gene 1: ESTs Chr.2 [149542 (DW) 5':H00283 3':H00284]

15 Gene 2: L-LACTATE DEHYDROGENASE M CHAIN Chr.11 [510595 (IW)

5':AA057759 3':AA057760]

Drug: Daunorubicin

Parameters:

$$\mu_k^{\text{sen}} = -1.052, \mu_l^{\text{sen}} = -0.7199$$

$$20 \quad \mu_k^{\text{insen}} = 0.2324, \mu_l^{\text{insen}} = 0.1588$$

$$\sigma_k^{\text{avg}} = 0.8508, \sigma_l^{\text{avg}} = 0.9279, \rho_{k,l}^{\text{avg}} = -0.1035$$

$$P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

Rule 46

25 Gene 1: SID W 471763 Crystallin zeta (quinone reductase) [5':AA035179

3':AA035180]

Gene 2: ESTs Chr.2 [149542 (DW) 5':H00283 3':H00284]

Drug: Daunorubicin

Parameters:

$$30 \quad \mu_k^{\text{sen}} = -0.5185, \mu_l^{\text{sen}} = -1.052$$

$$\mu_k^{\text{insen}} = 0.1147, \mu_l^{\text{insen}} = 0.2324$$

$$\sigma_k^{\text{avg}} = 0.9683, \sigma_l^{\text{avg}} = 0.8508, \rho_{k,l}^{\text{avg}} = -0.06753$$

$$P(C_i^{\text{sensitive}}) = 0.1811, \quad P(C_i^{\text{insensitive}}) = 0.8189$$

Rule 47

Gene 1: SID W 345683 ESTs Highly similar to INTEGRAL MEMBRANE

5 GLYCOPROTEIN GP210 PRECURSOR [Rattus norvegicus] [5':W76432 3':W72039]

Gene 2: SID W 489301 ESTs [5':AA054471 3':AA058511]

Drug: Daunorubicin

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.918, \quad \mu_l^{\text{sen}} = 0.7391 \\ 10 \quad \mu_k^{\text{insen}} &= -0.2022, \quad \mu_l^{\text{insen}} = -0.1637 \\ \sigma_k^{\text{avg}} &= 0.8758, \quad \sigma_l^{\text{avg}} = 0.9515, \quad \rho_{k,l}^{\text{avg}} = -0.3077 \\ P(C_i^{\text{sensitive}}) &= 0.1811, \quad P(C_i^{\text{insensitive}}) = 0.8189 \end{aligned}$$

Rule 48

15 Gene 1: ESTs Chr.2 [149542 (DW) 5':H00283 3':H00284]

Gene 2: *Aldehyde reductase 1 (low Km aldose reductase) SID W 418212 ESTs
[5':W90268 3':W90593]

Drug: Daunorubicin

Parameters:

$$\begin{aligned} 20 \quad \mu_k^{\text{sen}} &= -1.052, \quad \mu_l^{\text{sen}} = 0.09908 \\ \mu_k^{\text{insen}} &= 0.2324, \quad \mu_l^{\text{insen}} = -0.02151 \\ \sigma_k^{\text{avg}} &= 0.8508, \quad \sigma_l^{\text{avg}} = 1.014, \quad \rho_{k,l}^{\text{avg}} = 0.4702 \\ P(C_i^{\text{sensitive}}) &= 0.1811, \quad P(C_i^{\text{insensitive}}) = 0.8189 \end{aligned}$$

25 Rule 49

Gene 1: ESTs Chr.2 [149542 (DW) 5':H00283 3':H00284]

Gene 2: SID W 484773 PYRROLINE-5-CARBOXYLATE REDUCTASE
[5':AA037688 3':AA037689]

Drug: Daunorubicin

30 Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -1.052, \quad \mu_l^{\text{sen}} = -0.7351 \\ \mu_k^{\text{insen}} &= 0.2324, \quad \mu_l^{\text{insen}} = 0.1628 \end{aligned}$$

$$\sigma_k^{\text{avg}} = 0.8508, \sigma_l^{\text{avg}} = 0.9291, \rho_{k,l}^{\text{avg}} = -0.1858$$

$$P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

Rule 50

5 Gene 1: SID W 484773 PYRROLINE-5-CARBOXYLATE REDUCTASE
[5':AA037688 3':AA037689]

Gene 2: *Prothymosin alpha SID W 271976 AMINOACYLASE-1 [5':N44687
3':N35315]

Drug: Daunorubicin

10 Parameters:

$$\mu_k^{\text{sen}} = -0.7351, \mu_l^{\text{sen}} = -1.032$$

$$\mu_k^{\text{insen}} = 0.1628, \mu_l^{\text{insen}} = 0.2284$$

$$\sigma_k^{\text{avg}} = 0.9291, \sigma_l^{\text{avg}} = 0.858, \rho_{k,l}^{\text{avg}} = -0.2602$$

$$P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

15

Rule 51

Gene 1: ESTs Chr.16 [154654 (RW) 5':R55184 3':R55185]

Gene 2: ELONGATION FACTOR TU MITOCHONDRIAL PRECURSOR Chr.16
[429540 (IW) 5':AA011453 3':AA011397]

20 Drug: Daunorubicin

Parameters:

$$\mu_k^{\text{sen}} = 0.8271, \mu_l^{\text{sen}} = -0.994$$

$$\mu_k^{\text{insen}} = -0.1829, \mu_l^{\text{insen}} = 0.2199$$

$$\sigma_k^{\text{avg}} = 0.9198, \sigma_l^{\text{avg}} = 0.8654, \rho_{k,l}^{\text{avg}} = 0.223$$

25 $P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$

Rule 52

Gene 1: SID 234072 EST Highly similar to RETROVIRUS-RELATED POL
POLYPROTEIN [Homo sapiens] [5': 3':H69001]

30 Gene 2: ESTs Chr.2 [149542 (DW) 5':H00283 3':H00284]

Drug: Daunorubicin

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= -0.5103, \quad \mu_l^{\text{sen}} = -1.052 \\ \mu_k^{\text{insen}} &= 0.1131, \quad \mu_l^{\text{insen}} = 0.2324 \\ \sigma_k^{\text{avg}} &= 0.9797, \quad \sigma_l^{\text{avg}} = 0.8508, \quad \rho_{k,l}^{\text{avg}} = -0.1946 \\ P(C_i^{\text{sensitive}}) &= 0.1811, \quad P(C_i^{\text{insensitive}}) = 0.8189\end{aligned}$$

5

Rule 53

Gene 1: ELONGATION FACTOR TU MITOCHONDRIAL PRECURSOR Chr.16
[429540 (IW) 5':AA011453 3':AA011397]

Gene 2: ESTs Chr.2 [365120 (IW) 5':AA025204 3':AA025124]

10 Drug: Amsacrine

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= -0.7939, \quad \mu_l^{\text{sen}} = 0.558 \\ \mu_k^{\text{insen}} &= 0.2239, \quad \mu_l^{\text{insen}} = -0.1576 \\ \sigma_k^{\text{avg}} &= 0.8691, \quad \sigma_l^{\text{avg}} = 0.9701, \quad \rho_{k,l}^{\text{avg}} = 0.4985 \\ P(C_i^{\text{sensitive}}) &= 0.22, \quad P(C_i^{\text{insensitive}}) = 0.78\end{aligned}$$

15

Rule 54

Gene 1: SID W 489301 ESTs [5':AA054471 3':AA058511]

Gene 2: H.sapiens mRNA for TRAMP protein Chr.8 [149355 (IEW) 5':H01598
3':H01495]

20

Drug: Pyrazoloimidazole

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= 0.9637, \quad \mu_l^{\text{sen}} = 0.7678 \\ \mu_k^{\text{insen}} &= -0.2165, \quad \mu_l^{\text{insen}} = -0.1717 \\ \sigma_k^{\text{avg}} &= 0.8641, \quad \sigma_l^{\text{avg}} = 0.9429, \quad \rho_{k,l}^{\text{avg}} = -0.4318 \\ P(C_i^{\text{sensitive}}) &= 0.1833, \quad P(C_i^{\text{insensitive}}) = 0.8167\end{aligned}$$

25

Rule 55

Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

30

Gene 2: SID W 487113 Msh (Drosophila) homeo box homolog 1 (formerly homeo box
7) [5':AA045226 3':AA045325]

Drug: CPT,10-OH

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.9086, \quad \mu_l^{\text{sen}} = 0.8196 \\ \mu_k^{\text{insen}} &= 0.2078, \quad \mu_l^{\text{insen}} = -0.1876 \\ 5 \quad \sigma_k^{\text{avg}} &= 0.8915, \quad \sigma_l^{\text{avg}} = 0.8784, \quad \rho_{k,l}^{\text{avg}} = 0.3086 \\ P(C_i^{\text{sensitive}}) &= 0.1856, \quad P(C_i^{\text{insensitive}}) = 0.8144 \end{aligned}$$

Rule 56

Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19

10 [310021 (I) 5': 3':N99151]

Gene 2: SID W 346587 Homo sapiens quiescin (Q6) mRNA complete cds [5':W79188
3':W74434]

Drug: CPT,10-OH

Parameters:

$$\begin{aligned} 15 \quad \mu_k^{\text{sen}} &= -0.9086, \quad \mu_l^{\text{sen}} = 1.001 \\ \mu_k^{\text{insen}} &= 0.2078, \quad \mu_l^{\text{insen}} = -0.2285 \\ \sigma_k^{\text{avg}} &= 0.8915, \quad \sigma_l^{\text{avg}} = 0.8549, \quad \rho_{k,l}^{\text{avg}} = -0.09544 \\ P(C_i^{\text{sensitive}}) &= 0.1856, \quad P(C_i^{\text{insensitive}}) = 0.8144 \end{aligned}$$

20 Rule 57

Gene 1: SID W 510189 Homo sapiens CAG-isl 7 mRNA complete cds [5':AA053648
3':AA053259]

Gene 2: SID W 510534 MAJOR GASTROINTESTINAL TUMOR-ASSOCIATED
PROTEIN GA733-2 PRECURSOR [5':AA055858 3':AA055808]

25 Drug: CPT,10-OH

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.4935, \quad \mu_l^{\text{sen}} = -0.6863 \\ \mu_k^{\text{insen}} &= -0.1128, \quad \mu_l^{\text{insen}} = 0.1559 \\ \sigma_k^{\text{avg}} &= 0.9732, \quad \sigma_l^{\text{avg}} = 0.9458, \quad \rho_{k,l}^{\text{avg}} = 0.6221 \\ 30 \quad P(C_i^{\text{sensitive}}) &= 0.1856, \quad P(C_i^{\text{insensitive}}) = 0.8144 \end{aligned}$$

Rule 58

Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

Gene 2: COL4A1 Collagen type IV alpha 1 Chr.13 [489467 (IEW) 5':AA054624
3':AA054564]

5 Drug: CPT,10-OH

Parameters:

$$\mu_k^{\text{sen}} = -0.9086, \mu_l^{\text{sen}} = 0.8311$$

$$\mu_k^{\text{insen}} = 0.2078, \mu_l^{\text{insen}} = -0.1889$$

$$\sigma_k^{\text{avg}} = 0.8915, \sigma_l^{\text{avg}} = 0.9008, \rho_{k,l}^{\text{avg}} = 0.04514$$

$$10 \quad P(C_i^{\text{sensitive}}) = 0.1856, P(C_i^{\text{insensitive}}) = 0.8144$$

Rule 59

Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

15 Gene 2: SID 512355 ESTs Highly similar to SRC SUBSTRATE P80/85 PROTEINS
[Gallus gallus] [5':AA059424 3':AA057835]

Drug: CPT,10-OH

Parameters:

$$\mu_k^{\text{sen}} = -0.9086, \mu_l^{\text{sen}} = 0.8282$$

$$20 \quad \mu_k^{\text{insen}} = 0.2078, \mu_l^{\text{insen}} = -0.1885$$

$$\sigma_k^{\text{avg}} = 0.8915, \sigma_l^{\text{avg}} = 0.9162, \rho_{k,l}^{\text{avg}} = -0.1186$$

$$P(C_i^{\text{sensitive}}) = 0.1856, P(C_i^{\text{insensitive}}) = 0.8144$$

Rule 60

25 Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

Gene 2: SID W 324073 Human lysyl oxidase-like protein mRNA complete cds
[5':W46647 3':W46564]

Drug: CPT,10-OH

30 Parameters:

$$\mu_k^{\text{sen}} = -0.9086, \mu_l^{\text{sen}} = 0.7583$$

$$\mu_k^{\text{insen}} = 0.2078, \mu_l^{\text{insen}} = -0.1738$$

$$\sigma_k^{avg} = 0.8915, \sigma_l^{avg} = 0.9205, \rho_{k,l}^{avg} = 0.2083$$

$$P(C_i^{sensitive}) = 0.1856, P(C_i^{insensitive}) = 0.8144$$

Rule 61

- 5 Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

Gene 2: SID W 376472 Homo sapiens clone 24429 mRNA sequence [5':AA041443
3':AA041360]

Drug: CPT,10-OH

- 10 Parameters:

$$\mu_k^{sen} = -0.9086, \mu_l^{sen} = 0.7273$$

$$\mu_k^{insen} = 0.2078, \mu_l^{insen} = -0.1653$$

$$\sigma_k^{avg} = 0.8915, \sigma_l^{avg} = 0.927, \rho_{k,l}^{avg} = 0.02373$$

$$P(C_i^{sensitive}) = 0.1856, P(C_i^{insensitive}) = 0.8144$$

15

Rule 62

Gene 1: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528
3':AA043529]

Gene 2: Homo sapiens (clone 35.3) DRAL mRNA complete cds Chr.2 [324636 (IW)

- 20 5':W46933 3':W46835]

Drug: CPT,10-OH

Parameters:

$$\mu_k^{sen} = 0.8729, \mu_l^{sen} = 0.7843$$

$$\mu_k^{insen} = -0.1997, \mu_l^{insen} = -0.1778$$

$$25 \quad \sigma_k^{avg} = 0.8949, \sigma_l^{avg} = 0.9125, \rho_{k,l}^{avg} = -0.1147$$

$$P(C_i^{sensitive}) = 0.1856, P(C_i^{insensitive}) = 0.8144$$

Rule 63

- 30 Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

Gene 2: SID W 487878 SPARC/osteonectin [5':AA046533 3':AA045463]

Drug: CPT,10-OH

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= -0.9086, \mu_l^{\text{sen}} = 0.8472 \\ \mu_k^{\text{insen}} &= 0.2078, \mu_l^{\text{insen}} = -0.1926 \\ \sigma_k^{\text{avg}} &= 0.8915, \sigma_l^{\text{avg}} = 0.898, \rho_{k,l}^{\text{avg}} = -0.04153 \\ P(C_i^{\text{sensitive}}) &= 0.1856, P(C_i^{\text{insensitive}}) = 0.8144\end{aligned}$$

5

Rule 64

Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

Gene 2:

10 Drug: CPT,10-OH

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= -0.9086, \mu_l^{\text{sen}} = 0.6293 \\ \mu_k^{\text{insen}} &= 0.2078, \mu_l^{\text{insen}} = -0.1436 \\ \sigma_k^{\text{avg}} &= 0.8915, \sigma_l^{\text{avg}} = 0.9536, \rho_{k,l}^{\text{avg}} = 0.1463 \\ P(C_i^{\text{sensitive}}) &= 0.1856, P(C_i^{\text{insensitive}}) = 0.8144\end{aligned}$$

15

Rule 65

Gene 1: ESTs Chr.X [254029 (IRW) 5':N75199 3':N22323]

Gene 2: SID W 346587 Homo sapiens quiescin (Q6) mRNA complete cds [5':W79188
20 3':W74434]

Drug: CPT,10-OH

Parameters:

$$\begin{aligned}\mu_k^{\text{sen}} &= 0.1804, \mu_l^{\text{sen}} = 1.001 \\ \mu_k^{\text{insen}} &= -0.04026, \mu_l^{\text{insen}} = -0.2285 \\ \sigma_k^{\text{avg}} &= 1.01, \sigma_l^{\text{avg}} = 0.8549, \rho_{k,l}^{\text{avg}} = -0.4875 \\ P(C_i^{\text{sensitive}}) &= 0.1856, P(C_i^{\text{insensitive}}) = 0.8144\end{aligned}$$

25

Rule 66

Gene 1: SID W 364810 ESTs [5':AA034430 3':AA053921]

30 Gene 2: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

Drug: CPT,10-OH

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.6399, \quad \mu_l^{\text{sen}} = -0.9086 \\ \mu_k^{\text{insen}} &= 0.1449, \quad \mu_l^{\text{insen}} = 0.2078 \\ \sigma_k^{\text{avg}} &= 0.9312, \quad \sigma_l^{\text{avg}} = 0.8915, \quad \rho_{k,l}^{\text{avg}} = -0.1262 \\ 5 \quad P(C_i^{\text{sensitive}}) &= 0.1856, \quad P(C_i^{\text{insensitive}}) = 0.8144 \end{aligned}$$

Rule 67

Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

10 Gene 2: SID 257009 ESTs [5':N39759 3':N26801]

Drug: CPT,10-OH

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.9086, \quad \mu_l^{\text{sen}} = 0.5127 \\ \mu_k^{\text{insen}} &= 0.2078, \quad \mu_l^{\text{insen}} = -0.1168 \\ 15 \quad \sigma_k^{\text{avg}} &= 0.8915, \quad \sigma_l^{\text{avg}} = 0.9602, \quad \rho_{k,l}^{\text{avg}} = 0.1779 \\ P(C_i^{\text{sensitive}}) &= 0.1856, \quad P(C_i^{\text{insensitive}}) = 0.8144 \end{aligned}$$

Rule 68

Gene 1: SID 512355 ESTs Highly similar to SRC SUBSTRATE P80/85 PROTEINS
20 [Gallus gallus] [5':AA059424 3':AA057835]

Gene 2: SID W 346587 Homo sapiens quiescin (Q6) mRNA complete cds [5':W79188
3':W74434]

Drug: CPT,10-OH

Parameters:

$$\begin{aligned} 25 \quad \mu_k^{\text{sen}} &= 0.8282, \quad \mu_l^{\text{sen}} = 1.001 \\ \mu_k^{\text{insen}} &= -0.1885, \quad \mu_l^{\text{insen}} = -0.2285 \\ \sigma_k^{\text{avg}} &= 0.9162, \quad \sigma_l^{\text{avg}} = 0.8549, \quad \rho_{k,l}^{\text{avg}} = 0.18 \\ P(C_i^{\text{sensitive}}) &= 0.1856, \quad P(C_i^{\text{insensitive}}) = 0.8144 \end{aligned}$$

30 Rule 69

Gene 1: ASNS Asparagine synthetase Chr.7 [510206 (IW) 5':AA053213 3':AA053461]

Gene 2: SID W 346587 Homo sapiens quiescin (Q6) mRNA complete cds [5':W79188

3':W74434]

Drug: CPT,10-OH

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.7243, \quad \mu_l^{\text{sen}} = 1.001 \\ 5 \quad \mu_k^{\text{insen}} &= 0.1648, \quad \mu_l^{\text{insen}} = -0.2285 \\ \sigma_k^{\text{avg}} &= 0.9358, \quad \sigma_l^{\text{avg}} = 0.8549, \quad \rho_{k,l}^{\text{avg}} = -0.06293 \\ P(C_i^{\text{sensitive}}) &= 0.1856, \quad P(C_i^{\text{insensitive}}) = 0.8144 \end{aligned}$$

Rule 70

10 Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

Gene 2: Human extracellular protein (S1-5) mRNA complete cds Chr.2 [485875 (EW)
5':AA040442 3':AA040443]

Drug: CPT,10-OH

15 Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.9086, \quad \mu_l^{\text{sen}} = 0.7657 \\ \mu_k^{\text{insen}} &= 0.2078, \quad \mu_l^{\text{insen}} = -0.1743 \\ \sigma_k^{\text{avg}} &= 0.8915, \quad \sigma_l^{\text{avg}} = 0.9202, \quad \rho_{k,l}^{\text{avg}} = -0.1283 \\ P(C_i^{\text{sensitive}}) &= 0.1856, \quad P(C_i^{\text{insensitive}}) = 0.8144 \end{aligned}$$

20

Rule 71

Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19
[310021 (I) 5': 3':N99151]

Gene 2: Homo sapiens lysyl hydroxylase isoform 2 (PLOD2) mRNA complete cds

25 Chr.3 [310449 (IW) 5':W30982 3':N98463]

Drug: CPT,10-OH

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.9086, \quad \mu_l^{\text{sen}} = 0.6335 \\ \mu_k^{\text{insen}} &= 0.2078, \quad \mu_l^{\text{insen}} = -0.1445 \\ 30 \quad \sigma_k^{\text{avg}} &= 0.8915, \quad \sigma_l^{\text{avg}} = 0.9558, \quad \rho_{k,l}^{\text{avg}} = 0.1739 \\ P(C_i^{\text{sensitive}}) &= 0.1856, \quad P(C_i^{\text{insensitive}}) = 0.8144 \end{aligned}$$

Rule 72

Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19

[310021 (I) 5': 3':N99151]

Gene 2: SID W 486110 Profilin 2 [5':AA043167 3':AA040703]

5 Drug: CPT,10-OH

Parameters:

$$\mu_k^{\text{sen}} = -0.9086, \mu_l^{\text{sen}} = 0.7038$$

$$\mu_k^{\text{insen}} = 0.2078, \mu_l^{\text{insen}} = -0.1605$$

$$\sigma_k^{\text{avg}} = 0.8915, \sigma_l^{\text{avg}} = 0.9573, \rho_{k,l}^{\text{avg}} = -0.08051$$

$$10 \quad P(C_i^{\text{sensitive}}) = 0.1856, P(C_i^{\text{insensitive}}) = 0.8144$$

Rule 73

Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19

[310021 (I) 5': 3':N99151]

15 Gene 2: SID 42787 ESTs [5':R59827 3':R59717]

Drug: CPT,10-OH

Parameters:

$$\mu_k^{\text{sen}} = -0.9086, \mu_l^{\text{sen}} = 0.5759$$

$$\mu_k^{\text{insen}} = 0.2078, \mu_l^{\text{insen}} = -0.1318$$

$$20 \quad \sigma_k^{\text{avg}} = 0.8915, \sigma_l^{\text{avg}} = 0.961, \rho_{k,l}^{\text{avg}} = 0.06258$$

$$P(C_i^{\text{sensitive}}) = 0.1856, P(C_i^{\text{insensitive}}) = 0.8144$$

Rule 74

Gene 1: GAMMA-INTERFERON-INDUCIBLE PROTEIN IP-30 PRECURSOR Chr.19

25 [310021 (I) 5': 3':N99151]

Gene 2: SID 50243 ESTs [5':H17681 3':H17066]

Drug: CPT,10-OH

Parameters:

$$\mu_k^{\text{sen}} = -0.9086, \mu_l^{\text{sen}} = 0.8677$$

$$30 \quad \mu_k^{\text{insen}} = 0.2078, \mu_l^{\text{insen}} = -0.1977$$

$$\sigma_k^{\text{avg}} = 0.8915, \sigma_l^{\text{avg}} = 0.9058, \rho_{k,l}^{\text{avg}} = -0.1472$$

$$P(C_i^{\text{sensitive}}) = 0.1856, P(C_i^{\text{insensitive}}) = 0.8144$$

Rule 75

Gene 1: SID W 346587 Homo sapiens quiescin (Q6) mRNA complete cds [5':W79188 3':W74434]

5 Gene 2: SID 359504 ESTs [5': 3':AA010589]

Drug: CPT,10-OH

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 1.001, \mu_l^{\text{sen}} = -0.336 \\ \mu_k^{\text{insen}} &= -0.2285, \mu_l^{\text{insen}} = 0.07633 \\ 10 \quad \sigma_k^{\text{avg}} &= 0.8549, \sigma_l^{\text{avg}} = 0.9733, \rho_{k,l}^{\text{avg}} = 0.3387 \\ P(C_i^{\text{sensitive}}) &= 0.1856, P(C_i^{\text{insensitive}}) = 0.8144 \end{aligned}$$

Rule 76

Gene 1: SID 39144 ESTs Weakly similar to Rep-8 [H.sapiens] [5':R51769 3':R51770]

15 Gene 2: SID W 358526 ESTs [5':W96039 3':W94821]

Drug: CPT,20-ester (S)

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.8367, \mu_l^{\text{sen}} = -0.771 \\ \mu_k^{\text{insen}} &= 0.2555, \mu_l^{\text{insen}} = 0.2359 \\ 20 \quad \sigma_k^{\text{avg}} &= 0.8798, \sigma_l^{\text{avg}} = 0.9049, \rho_{k,l}^{\text{avg}} = -0.2237 \\ P(C_i^{\text{sensitive}}) &= 0.2344, P(C_i^{\text{insensitive}}) = 0.7656 \end{aligned}$$

Rule 77

Gene 1: SID 39144 ESTs Weakly similar to Rep-8 [H.sapiens] [5':R51769 3':R51770]

25 Gene 2: SID W 509633 ESTs Moderately similar to Kryn [M.musculus] [5':AA045560 3':AA045561]

Drug: CPT,20-ester (S)

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.8367, \mu_l^{\text{sen}} = -0.8637 \\ 30 \quad \mu_k^{\text{insen}} &= 0.2555, \mu_l^{\text{insen}} = 0.2643 \\ \sigma_k^{\text{avg}} &= 0.8798, \sigma_l^{\text{avg}} = 0.8771, \rho_{k,l}^{\text{avg}} = -0.2147 \\ P(C_i^{\text{sensitive}}) &= 0.2344, P(C_i^{\text{insensitive}}) = 0.7656 \end{aligned}$$

Rule 78

Gene 1: SID 39144 ESTs Weakly similar to Rep-8 [H.sapiens] [5':R51769 3':R51770]

Gene 2: *Hs.648 Cut (Drosophila)-like 1 (CCAAT displacement protein) SID W 26677

5 ESTs [5':R13994 3':R39117]

Drug: CPT,20-ester (S)

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.8367, \quad \mu_l^{\text{sen}} = -0.652 \\ \mu_k^{\text{insen}} &= 0.2555, \quad \mu_l^{\text{insen}} = 0.1999 \\ 10 \quad \sigma_k^{\text{avg}} &= 0.8798, \quad \sigma_l^{\text{avg}} = 0.9431, \quad \rho_{k,l}^{\text{avg}} = -0.3363 \\ P(C_i^{\text{sensitive}}) &= 0.2344, \quad P(C_i^{\text{insensitive}}) = 0.7656 \end{aligned}$$

Rule 79

Gene 1: SID W 510189 Homo sapiens CAG-isl 7 mRNA complete cds [5':AA053648
15 3':AA053259]

Gene 2: SID W 346510 Homo sapiens hCPE-R mRNA for CPE-receptor complete cds [5':W79089 3':W74492]

Drug: CPT

Parameters:

$$\begin{aligned} 20 \quad \mu_k^{\text{sen}} &= 0.4583, \quad \mu_l^{\text{sen}} = -0.4683 \\ \mu_k^{\text{insen}} &= -0.161, \quad \mu_l^{\text{insen}} = 0.1634 \\ \sigma_k^{\text{avg}} &= 0.9838, \quad \sigma_l^{\text{avg}} = 0.9573, \quad \rho_{k,l}^{\text{avg}} = 0.6575 \\ P(C_i^{\text{sensitive}}) &= 0.2594, \quad P(C_i^{\text{insensitive}}) = 0.7406 \end{aligned}$$

25 Rule 80

Gene 1: ESTs Chr.19 [485804 (EW) 5':AA040350 3':AA040351]

Gene 2: Glyoxalase-I-log

Drug: CPT,20-ester (S)

Parameters:

$$\begin{aligned} 30 \quad \mu_k^{\text{sen}} &= -0.7177, \quad \mu_l^{\text{sen}} = -0.5058 \\ \mu_k^{\text{insen}} &= 0.2573, \quad \mu_l^{\text{insen}} = 0.1814 \\ \sigma_k^{\text{avg}} &= 0.8936, \quad \sigma_l^{\text{avg}} = 0.9632, \quad \rho_{k,l}^{\text{avg}} = -0.3337 \end{aligned}$$

$$P(C_i^{\text{sensitive}}) = 0.2644, \quad P(C_i^{\text{insensitive}}) = 0.7356$$

Rule 81

Gene 1: Human G/T mismatch-specific thymine DNA glycosylase mRNA complete cds

5 Chr.X [321997 (IW) 5':W37234 3':W37817]

Gene 2: SID W 358526 ESTs [5':W96039 3':W94821]

Drug: CPT,11-formyl (RS)

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.626, \quad \mu_l^{\text{sen}} = -1.055 \\ 10 \quad \mu_k^{\text{insen}} &= -0.151, \quad \mu_l^{\text{insen}} = 0.2536 \\ \sigma_k^{\text{avg}} &= 0.977, \quad \sigma_l^{\text{avg}} = 0.8569, \quad \rho_{k,l}^{\text{avg}} = 0.3776 \\ P(C_i^{\text{sensitive}}) &= 0.1939, \quad P(C_i^{\text{insensitive}}) = 0.8061 \end{aligned}$$

Rule 82

15 Gene 1: SID W 135118 GATA-binding protein 3 [5':R31441 3':R31442]

Gene 2: SID W 358526 ESTs [5':W96039 3':W94821]

Drug: CPT,11-formyl (RS)

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.9817, \quad \mu_l^{\text{sen}} = -1.055 \\ 20 \quad \mu_k^{\text{insen}} &= -0.2359, \quad \mu_l^{\text{insen}} = 0.2536 \\ \sigma_k^{\text{avg}} &= 0.9021, \quad \sigma_l^{\text{avg}} = 0.8569, \quad \rho_{k,l}^{\text{avg}} = 0.08481 \\ P(C_i^{\text{sensitive}}) &= 0.1939, \quad P(C_i^{\text{insensitive}}) = 0.8061 \end{aligned}$$

Rule 83

25 Gene 1: ESTs Chr.16 [154654 (RW) 5':R55184 3':R55185]

Gene 2: SOD2 Superoxide dismutase 2 mitochondrial Chr.6 [144758 (EW) 5':R76245 3':R76527]

Drug: CPT,11-formyl (RS)

Parameters:

$$\begin{aligned} 30 \quad \mu_k^{\text{sen}} &= 0.874, \quad \mu_l^{\text{sen}} = -0.7046 \\ \mu_k^{\text{insen}} &= -0.2102, \quad \mu_l^{\text{insen}} = 0.1693 \\ \sigma_k^{\text{avg}} &= 0.9112, \quad \sigma_l^{\text{avg}} = 0.9543, \quad \rho_{k,l}^{\text{avg}} = 0.3184 \end{aligned}$$

$$P(C_i^{\text{sensitive}}) = 0.1939, \quad P(C_i^{\text{insensitive}}) = 0.8061$$

Rule 84

Gene 1: SID W 358526 ESTs [5':W96039 3':W94821]

5 Gene 2: Glutathione S-Transferase A1-log

Drug: CPT,11-formyl (RS)

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -1.055, \quad \mu_l^{\text{sen}} = -0.6283 \\ \mu_k^{\text{insen}} &= 0.2536, \quad \mu_l^{\text{insen}} = 0.1488 \\ 10 \quad \sigma_k^{\text{avg}} &= 0.8569, \quad \sigma_l^{\text{avg}} = 0.9702, \quad \rho_{k,l}^{\text{avg}} = -0.125 \\ P(C_i^{\text{sensitive}}) &= 0.1939, \quad P(C_i^{\text{insensitive}}) = 0.8061 \end{aligned}$$

Rule 85

Gene 1: SID W 358526 ESTs [5':W96039 3':W94821]

15 Gene 2: PIGF Phosphatidylinositol glycan class F Chr.2 [486751 (IEW) 5':AA042803 3':AA044616]

Drug: CPT,11-formyl (RS)

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -1.055, \quad \mu_l^{\text{sen}} = -0.4069 \\ 20 \quad \mu_k^{\text{insen}} &= 0.2536, \quad \mu_l^{\text{insen}} = 0.09808 \\ \sigma_k^{\text{avg}} &= 0.8569, \quad \sigma_l^{\text{avg}} = 1.003, \quad \rho_{k,l}^{\text{avg}} = -0.3618 \\ P(C_i^{\text{sensitive}}) &= 0.1939, \quad P(C_i^{\text{insensitive}}) = 0.8061 \end{aligned}$$

Rule 86

25 Gene 1: PROTEASOME COMPONENT C13 PRECURSOR Chr.6 [344774 (IW) 5':W74742 3':W74705]

Gene 2: SID W 484681 Homo sapiens ES/130 mRNA complete cds [5':AA037568 3':AA037487]

Drug: Mechlorethamine

30 Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.6562, \quad \mu_l^{\text{sen}} = -0.8883 \\ \mu_k^{\text{insen}} &= -0.1565, \quad \mu_l^{\text{insen}} = 0.2119 \end{aligned}$$

$$\sigma_k^{avg} = 0.9627, \sigma_l^{avg} = 0.9254, \rho_{k,l}^{avg} = 0.5304$$

$$P(C_i^{sensitive}) = 0.1928, P(C_i^{insensitive}) = 0.8072$$

Rule 87

5 Gene 1: SID 43609 ESTs [5':H06454 3':H06184]

Gene 2: SID W 53251 Human Zn-15 related zinc finger protein (rlf) mRNA complete cds [5':R15988 3':R15987]

Drug: Mechlorethamine

Parameters:

$$10 \quad \mu_k^{sen} = 1.042, \mu_l^{sen} = -0.5622$$

$$\mu_k^{insen} = -0.2493, \mu_l^{insen} = 0.1345$$

$$\sigma_k^{avg} = 0.8728, \sigma_l^{avg} = 0.9712, \rho_{k,l}^{avg} = 0.3407$$

$$P(C_i^{sensitive}) = 0.1928, P(C_i^{insensitive}) = 0.8072$$

15 Rule 88

Gene 1: CDH2 Cadherin 2 N-cadherin (neuronal) Chr. [325182 (DIRW) 5':W48793 3':W49619]

Gene 2: Homo sapiens (clone 35.3) DRAL mRNA complete cds Chr.2 [324636 (IW) 5':W46933 3':W46835]

20 Drug: Geldanamycin

Parameters:

$$\mu_k^{sen} = -0.8842, \mu_l^{sen} = 0.09839$$

$$\mu_k^{insen} = 0.225, \mu_l^{insen} = -0.02426$$

$$\sigma_k^{avg} = 0.8839, \sigma_l^{avg} = 1, \rho_{k,l}^{avg} = 0.6697$$

$$25 \quad P(C_i^{sensitive}) = 0.2033, P(C_i^{insensitive}) = 0.7967$$

Rule 89

Gene 1: ESTsSID 327435 [5':W32467 3':W19830]

Gene 2: ESTs Chr.3 [377430 (IW) 5':AA055159 3':AA055043]

30 Drug: Morpholino-adriamycin

Parameters:

$$\mu_k^{sen} = 0.7559, \mu_l^{sen} = 1.064$$

$$\mu_k^{\text{insen}} = -0.1508, \quad \mu_l^{\text{insen}} = -0.212$$

$$\sigma_k^{\text{avg}} = 0.9646, \quad \sigma_l^{\text{avg}} = 0.9006, \quad \rho_{k,l}^{\text{avg}} = -0.2502$$

$$P(C_i^{\text{sensitive}}) = 0.1661, \quad P(C_i^{\text{insensitive}}) = 0.8339$$

Quadratic Discriminant Analysis – 2-dimensional (QDA 2D)

This method computes a Bayesian conditional probability $P(j \in C_i^{sensitive} | g_k^j, g_l^j)$ that a cell line j is sensitive to drug i , given the abundances of genes k and l , g_k^j and g_l^j , respectively, in cell line j .

5

The probability is computed using the following equation:

$$P(j \in C_i^{sensitive} | g_k^j, g_l^j) = \frac{{}_iG_{k,l}^{sensitive}(g_k^j, g_l^j) \cdot P(C_i^{sensitive})}{{}_iG_{k,l}^{sensitive}(g_k^j, g_l^j) \cdot P(C_i^{sensitive}) + {}_iG_{k,l}^{insensitive}(g_k^j, g_l^j) \cdot P(C_i^{insensitive})}$$

10 where

$$\begin{aligned} P(C_i^{sensitive}) &= \text{prior probability of the sensitive set} \\ &= |C_i^{sensitive}| / (|C_i^{sensitive}| + |C_i^{insensitive}|), \end{aligned}$$

$$\begin{aligned} P(C_i^{insensitive}) &= \text{prior probability of the insensitive set} \\ &= |C_i^{insensitive}| / (|C_i^{sensitive}| + |C_i^{insensitive}|), \end{aligned}$$

15 ${}_iG_{k,l}^{sensitive}(g_k^j, g_l^j)$ = joint probability of abundance values g_k^j and g_l^j from the bivariate gaussian density fitted to the histogram of gene k and l abundances over the sensitive cell lines when subjected to drug i .

$$\begin{aligned} &{}_iG_{k,l}^{sensitive}(g_k^j, g_l^j) = \\ 20 & \frac{1}{2\pi\sigma_k^{sen}\sigma_l^{sen}\sqrt{1-(\rho_{k,l}^{sen})^2}} \exp \left\{ - \frac{\left[\left(\frac{g_k^j - \mu_k^{sen}}{\sigma_k^{sen}} \right)^2 - 2\rho_{k,l}^{sen} \left(\frac{g_k^j - \mu_k^{sen}}{\sigma_k^{sen}} \right) \left(\frac{g_l^j - \mu_l^{sen}}{\sigma_l^{sen}} \right) + \left(\frac{g_l^j - \mu_l^{sen}}{\sigma_l^{sen}} \right)^2 \right]}{2(1-(\rho_{k,l}^{sen})^2)} \right\} \end{aligned}$$

where

μ_k^{sen} = mean of gene k abundances over the sensitive cell lines

σ_k^{sen} = standard deviation of gene k abundances in the sensitive cell lines

μ_l^{sen} = mean of gene l abundances over the sensitive cell lines

σ_l^{sen} = standard deviation of gene l abundances in the sensitive cell lines

$\rho_{k,l}^{sen}$ = correlation coefficient of gene k and gene l abundances in the sensitive cell lines

5

${}_i G_{k,l}^{insensitive}(g_k^j, g_l^j)$ = joint probability of abundance values g_k^j and g_l^j from the bivariate gaussian density fitted to the histogram of gene k and l abundances over the insensitive cell lines when subjected to drug i .

10

$${}_i G_{k,l}^{insensitive}(g_k^j, g_l^j) = \frac{1}{2\pi\sigma_k^{insen}\sigma_l^{insen}\sqrt{1-(\rho_{k,l}^{insen})^2}} \exp \left\{ -\frac{\left[\left(\frac{g_k^j - \mu_k^{insen}}{\sigma_k^{insen}} \right)^2 - 2\rho_{k,l}^{insen} \left(\frac{g_k^j - \mu_k^{insen}}{\sigma_k^{insen}} \right) \left(\frac{g_l^j - \mu_l^{insen}}{\sigma_l^{insen}} \right) + \left(\frac{g_l^j - \mu_l^{insen}}{\sigma_l^{insen}} \right)^2 \right]}{2(1-(\rho_{k,l}^{insen})^2)} \right\}$$

where

15 μ_k^{insen} = mean of gene k abundances over the insensitive cell lines

σ_k^{insen} = standard deviation of gene k abundances in the insensitive cell lines

μ_l^{insen} = mean of gene l abundances over the insensitive cell lines

σ_l^{insen} = standard deviation of gene l abundances in the insensitive cell lines

$\rho_{k,l}^{insen}$ = correlation coefficient of gene k and gene l abundances in the insensitive cell lines

20

Sample parameters for the QDA 2D anslyis of the NCI60 dataset are:

Rule 1

Genè 1: BMI1 Murine leukemia viral (bmi-1) oncogene homolog Chr.10 [418004

25 (REW) 5':W90704 3':W90705]

Gene 2: Human small GTP binding protein Rab7 mRNA complete cds Chr.3 [486233 (IW) 5':AA043679 3':AA043680]

Drug: Baker's-soluble-antifolate

Parameters:

$$\begin{aligned} 5 \quad \mu_k^{\text{sen}} &= 0.2314, \mu_l^{\text{sen}} = 0.3177, \sigma_k^{\text{sen}} = 1.437, \sigma_l^{\text{sen}} = 1.51, \rho_{k,l}^{\text{sen}} = -0.06216 \\ \mu_k^{\text{insen}} &= -0.07175, \mu_l^{\text{insen}} = -0.0982, \sigma_k^{\text{insen}} = 0.7941, \sigma_l^{\text{insen}} = 0.7097, \rho_{k,l}^{\text{insen}} = - \\ &0.3688 \end{aligned}$$

$$P(C_i^{\text{sensitive}}) = 0.2361, P(C_i^{\text{insensitive}}) = 0.7639$$

10 Rule 2

Gene 1: IL8 Interleukin 8 Chr.4 [328692 (DW) 5':W40283 3':W45324]

Gene 2: X-ray induction of CIP1/WAF1-log

Drug: Cyanomorpholinodoxorubicin

Parameters:

$$\begin{aligned} 15 \quad \mu_k^{\text{sen}} &= 0.856, \mu_l^{\text{sen}} = 0.6131, \sigma_k^{\text{sen}} = 0.6623, \sigma_l^{\text{sen}} = 0.9005, \rho_{k,l}^{\text{sen}} = 0.4391 \\ \mu_k^{\text{insen}} &= -0.224, \mu_l^{\text{insen}} = -0.1602, \sigma_k^{\text{insen}} = 0.9401, \sigma_l^{\text{insen}} = 0.9451, \rho_{k,l}^{\text{insen}} = - \\ &0.5299 \end{aligned}$$

$$P(C_i^{\text{sensitive}}) = 0.2067, P(C_i^{\text{insensitive}}) = 0.7933$$

20 Rule 3

Gene 1: SID W 45954 H.sapiens mRNA for testican [5':H08669 3':H08670]

Gene 2: SID W 359443 Human ORF mRNA complete cds [5':AA010705 3':AA010706]

Drug: Cyanomorpholinodoxorubicin

Parameters:

$$\begin{aligned} 25 \quad \mu_k^{\text{sen}} &= 0.8178, \mu_l^{\text{sen}} = 0.7159, \sigma_k^{\text{sen}} = 0.9544, \sigma_l^{\text{sen}} = 0.6062, \rho_{k,l}^{\text{sen}} = -0.8806 \\ \mu_k^{\text{insen}} &= -0.2139, \mu_l^{\text{insen}} = -0.1865, \sigma_k^{\text{insen}} = 0.8419, \sigma_l^{\text{insen}} = 0.9949, \rho_{k,l}^{\text{insen}} = \\ &0.3109 \end{aligned}$$

$$P(C_i^{\text{sensitive}}) = 0.2067, P(C_i^{\text{insensitive}}) = 0.7933$$

30 Rule 4

Gene 1: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Gene 2: ESTs Chr.1 [488132 (IW) 5':AA047420 3':AA047421]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = -1.008, \mu_l^{\text{sen}} = 0.4755, \sigma_k^{\text{sen}} = 0.5668, \sigma_l^{\text{sen}} = 0.3355, \rho_{k,l}^{\text{sen}} = 0.3703$$

$$\mu_k^{\text{insen}} = 0.2536, \mu_l^{\text{insen}} = -0.1193, \sigma_k^{\text{insen}} = 0.9027, \sigma_l^{\text{insen}} = 1.066, \rho_{k,l}^{\text{insen}} = -$$

5 0.2131

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 5

10 Gene 1: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J
WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Gene 2: ZFP36 Zinc finger protein homologous to Zfp-36 in mouse Chr.19 [486668
(DIW) 5':AA043477 3':AA043478]

Drug: Mitozolamide

15 Parameters:

$$\mu_k^{\text{sen}} = -0.3906, \mu_l^{\text{sen}} = -1.008, \sigma_k^{\text{sen}} = 0.5337, \sigma_l^{\text{sen}} = 0.5668, \rho_{k,l}^{\text{sen}} = 0.1073$$

$$\mu_k^{\text{insen}} = 0.09821, \mu_l^{\text{insen}} = 0.2536, \sigma_k^{\text{insen}} = 1.044, \sigma_l^{\text{insen}} = 0.9027, \rho_{k,l}^{\text{insen}} = -$$

0.3729

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

20

Rule 6

Gene 1: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J
WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Gene 2: SID W 323824 NADH-CYTOCHROME B5 REDUCTASE [5':W46211

25 3':W46212]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = -1.008, \mu_l^{\text{sen}} = 0.2421, \sigma_k^{\text{sen}} = 0.5668, \sigma_l^{\text{sen}} = 0.4385, \rho_{k,l}^{\text{sen}} = 0.04634$$

$$\mu_k^{\text{insen}} = 0.2536, \mu_l^{\text{insen}} = -0.06095, \sigma_k^{\text{insen}} = 0.9027, \sigma_l^{\text{insen}} = 1.078, \rho_{k,l}^{\text{insen}} = -$$

30 0.1944

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 7

Gene 1: ESTs Chr.6 [146640 (I) 5':R80056 3':R79962]

Gene 2: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

5 Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = -0.3763, \mu_l^{\text{sen}} = -1.008, \sigma_k^{\text{sen}} = 0.5482, \sigma_l^{\text{sen}} = 0.5668, \rho_{k,l}^{\text{sen}} = -0.7153$$

$$\mu_k^{\text{insen}} = 0.09352, \mu_l^{\text{insen}} = 0.2536, \sigma_k^{\text{insen}} = 1.034, \sigma_l^{\text{insen}} = 0.9027, \rho_{k,l}^{\text{insen}} = -$$

0.1007

$$10 \quad P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 8

Gene 1: SID 276915 ESTs [5':N48564 3':N39452]

Gene 2: SID 301144 ESTs [5':W16630 3':N78729]

15 Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 0.001165, \mu_l^{\text{sen}} = 0.7785, \sigma_k^{\text{sen}} = 0.4, \sigma_l^{\text{sen}} = 0.2994, \rho_{k,l}^{\text{sen}} = -0.3594$$

$$\mu_k^{\text{insen}} = -0.0009506, \mu_l^{\text{insen}} = -0.1951, \sigma_k^{\text{insen}} = 1.068, \sigma_l^{\text{insen}} = 1.014, \rho_{k,l}^{\text{insen}} = -$$

0.2265

$$20 \quad P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 9

Gene 1: Homo sapiens HuUAP1 mRNA for UDP-N-acetylglucosamine

pyrophosphorylase complete cds Chr.1 [486035 (DIW) 5':AA043109 3':AA040861]

25 Gene 2: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: Mitozolamide

Parameters:

$$\mu_k^{\text{sen}} = 0.3574, \mu_l^{\text{sen}} = -1.008, \sigma_k^{\text{sen}} = 0.5869, \sigma_l^{\text{sen}} = 0.5668, \rho_{k,l}^{\text{sen}} = 0.3711$$

$$30 \quad \mu_k^{\text{insen}} = -0.09028, \mu_l^{\text{insen}} = 0.2536, \sigma_k^{\text{insen}} = 1.028, \sigma_l^{\text{insen}} = 0.9027, \rho_{k,l}^{\text{insen}} = -$$

0.1971

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 10

Gene 1: SID W 510182 H.sapiens mRNA for kinase A anchor protein [5':AA053156 3':AA053135]

Gene 2: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

5 WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: Mitozolamide

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.4282, \mu_l^{\text{sen}} = -1.008, \sigma_k^{\text{sen}} = 0.4124, \sigma_l^{\text{sen}} = 0.5668, \rho_{k,l}^{\text{sen}} = 0.1487 \\ \mu_k^{\text{insen}} &= 0.1064, \mu_l^{\text{insen}} = 0.2536, \sigma_k^{\text{insen}} = 1.07, \sigma_l^{\text{insen}} = 0.9027, \rho_{k,l}^{\text{insen}} = \\ 10 \quad 0.03962 \\ P(C_i^{\text{sensitive}}) &= 0.2006, P(C_i^{\text{insensitive}}) = 0.7994 \end{aligned}$$

Rule 11

Gene 1: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

15 WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Gene 2: SID 488362 ESTs [5':AA046764 3':AA046492]

Drug: Mitozolamide

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -1.008, \mu_l^{\text{sen}} = 0.5996, \sigma_k^{\text{sen}} = 0.5668, \sigma_l^{\text{sen}} = 0.3048, \rho_{k,l}^{\text{sen}} = -0.238 \\ 20 \quad \mu_k^{\text{insen}} &= 0.2536, \mu_l^{\text{insen}} = -0.1504, \sigma_k^{\text{insen}} = 0.9027, \sigma_l^{\text{insen}} = 1.035, \rho_{k,l}^{\text{insen}} = \\ 0.1442 \\ P(C_i^{\text{sensitive}}) &= 0.2006, P(C_i^{\text{insensitive}}) = 0.7994 \end{aligned}$$

Rule 12

25 Gene 1: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Gene 2: ESTs Highly similar to HYPOTHETICAL 13.6 KD PROTEIN IN NUP170-ILS1 INTERGENIC REGION [Saccharo Chr.12 [415646 (IW) 5':W78722 3':W80529]

Drug: Mitozolamide

30 Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -1.008, \mu_l^{\text{sen}} = 0.4566, \sigma_k^{\text{sen}} = 0.5668, \sigma_l^{\text{sen}} = 0.413, \rho_{k,l}^{\text{sen}} = 0.02745 \\ \mu_k^{\text{insen}} &= 0.2536, \mu_l^{\text{insen}} = -0.1139, \sigma_k^{\text{insen}} = 0.9027, \sigma_l^{\text{insen}} = 1.038, \rho_{k,l}^{\text{insen}} = \\ 0.3175 \end{aligned}$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 13

Gene 1: ESTs Weakly similar to R06B9.b [C.elegans] Chr.1 [365488 (IW) 5':AA009557

5 3':AA009558]

Gene 2: SID W 380674 ESTs [5':AA053720 3':AA053711]

Drug: Mitozolamide

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.5214, \mu_l^{\text{sen}} = 1.093, \sigma_k^{\text{sen}} = 0.4503, \sigma_l^{\text{sen}} = 1.032, \rho_{k,l}^{\text{sen}} = 0.2533 \\ 10 \quad \mu_k^{\text{insen}} &= -0.1312, \mu_l^{\text{insen}} = -0.2739, \sigma_k^{\text{insen}} = 1.016, \sigma_l^{\text{insen}} = 0.7614, \rho_{k,l}^{\text{insen}} = - \\ &0.2896 \end{aligned}$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 14

15 Gene 1: ESTs Chr.1 [366242 (I) 5': 3':AA025593]

Gene 2: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: Mitozolamide

Parameters:

$$\begin{aligned} 20 \quad \mu_k^{\text{sen}} &= -0.2007, \mu_l^{\text{sen}} = -1.008, \sigma_k^{\text{sen}} = 0.4757, \sigma_l^{\text{sen}} = 0.5668, \rho_{k,l}^{\text{sen}} = -0.2512 \\ \mu_k^{\text{insen}} &= 0.04952, \mu_l^{\text{insen}} = 0.2536, \sigma_k^{\text{insen}} = 1.076, \sigma_l^{\text{insen}} = 0.9027, \rho_{k,l}^{\text{insen}} = - \\ &0.1109 \end{aligned}$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

25 Rule 15

Gene 1: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW)

5':AA039292 3':AA039334]

Gene 2: SID 147338 ESTs [5': 3':H01302]

Drug: Cyclodisone

30 Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.6598, \mu_l^{\text{sen}} = 0.1958, \sigma_k^{\text{sen}} = 0.2562, \sigma_l^{\text{sen}} = 0.3673, \rho_{k,l}^{\text{sen}} = -0.6593 \\ \mu_k^{\text{insen}} &= -0.1341, \mu_l^{\text{insen}} = -0.04021, \sigma_k^{\text{insen}} = 1.038, \sigma_l^{\text{insen}} = 1.061, \rho_{k,l}^{\text{insen}} = \\ &0.2816 \end{aligned}$$

$$P(C_i^{\text{sensitive}}) = 0.1689, P(C_i^{\text{insensitive}}) = 0.8311$$

Rule 16

Gene 1: SID W 51940 BETA-2-MICROGLOBULIN PRECURSOR [5':H24236

5 3':H24237]

Gene 2: SID W 486110 Profilin 2 [5':AA043167 3':AA040703]

Drug: Cyclodisone

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.6766, \mu_l^{\text{sen}} = 0.615, \sigma_k^{\text{sen}} = 0.5551, \sigma_l^{\text{sen}} = 0.4072, \rho_{k,l}^{\text{sen}} = 0.9224 \\ 10 \quad \mu_k^{\text{insen}} &= -0.1373, \mu_l^{\text{insen}} = -0.1252, \sigma_k^{\text{insen}} = 0.996, \sigma_l^{\text{insen}} = 1.031, \rho_{k,l}^{\text{insen}} = - \\ &0.313 \end{aligned}$$

$$P(C_i^{\text{sensitive}}) = 0.1689, P(C_i^{\text{insensitive}}) = 0.8311$$

Rule 17

15 Gene 1: Human DNA sequence from clone 14O9 on chromosome Xp11.1-11.4.

Contains a Inter-Alpha-Trypsin Inh Chr.X [485194 (I) 5':AA039416 3':AA039316]

Gene 2: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW)

5':AA039292 3':AA039334]

Drug: Cyclodisone

20 Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= 0.2487, \mu_l^{\text{sen}} = 0.6598, \sigma_k^{\text{sen}} = 0.4569, \sigma_l^{\text{sen}} = 0.2562, \rho_{k,l}^{\text{sen}} = -0.4186 \\ \mu_k^{\text{insen}} &= -0.05158, \mu_l^{\text{insen}} = -0.1341, \sigma_k^{\text{insen}} = 1.039, \sigma_l^{\text{insen}} = 1.038, \rho_{k,l}^{\text{insen}} = \\ &0.2219 \end{aligned}$$

$$P(C_i^{\text{sensitive}}) = 0.1689, P(C_i^{\text{insensitive}}) = 0.8311$$

25

Rule 18

Gene 1: SID 512164 Human clathrin assembly protein 50 (AP50) mRNA complete cds
[5': 3':AA057396]

Gene 2: SID W 345624 Human homeobox protein (PHOX1) mRNA 3' end [5':W76402
30 3':W72050]

Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = 0.8248, \mu_l^{\text{sen}} = -0.253, \sigma_k^{\text{sen}} = 0.7407, \sigma_l^{\text{sen}} = 0.7545, \rho_{k,l}^{\text{sen}} = 0.793$$

$$\mu_k^{\text{insen}} = -0.1956, \mu_l^{\text{insen}} = 0.06021, \sigma_k^{\text{insen}} = 0.9082, \sigma_l^{\text{insen}} = 1.037, \rho_{k,l}^{\text{insen}} = 0.7103$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

5 Rule 19

Gene 1: MSN Moesin Chr.X [486864 (IW) 5':AA043008 3':AA042882]

Gene 2: Human mRNA for reticulocalbin complete cds Chr.11 [485209 (IW)

5':AA039292 3':AA039334]

Drug: Clomesone

10 Parameters:

$$\mu_k^{\text{sen}} = 0.6791, \mu_l^{\text{sen}} = 0.4913, \sigma_k^{\text{sen}} = 0.4486, \sigma_l^{\text{sen}} = 0.4435, \rho_{k,l}^{\text{sen}} = 0.8962$$

$$\mu_k^{\text{insen}} = -0.1612, \mu_l^{\text{insen}} = -0.1165, \sigma_k^{\text{insen}} = 1.026, \sigma_l^{\text{insen}} = 1.058, \rho_{k,l}^{\text{insen}} = 0.04721$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

15

Rule 20

Gene 1: SID W 36809 Homo sapiens neural cell adhesion molecule (CALL) mRNA complete cds [5':R34648 3':R49177]

Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528

20 3':AA043529]

Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = 0.6335, \mu_l^{\text{sen}} = 1.184, \sigma_k^{\text{sen}} = 0.7063, \sigma_l^{\text{sen}} = 0.9042, \rho_{k,l}^{\text{sen}} = 0.2103$$

$$\mu_k^{\text{insen}} = -0.1498, \mu_l^{\text{insen}} = -0.2817, \sigma_k^{\text{insen}} = 0.9826, \sigma_l^{\text{insen}} = 0.7835, \rho_{k,l}^{\text{insen}} = -$$

25 0.3389

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

Rule 21

Gene 1: SID W 471748 ESTs [5':AA035018 3':AA035486]

30 Gene 2: SID 147338 ESTs [5': 3':H01302]

Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = 1.066, \mu_l^{\text{sen}} = 0.1604, \sigma_k^{\text{sen}} = 0.9178, \sigma_l^{\text{sen}} = 0.37, \rho_{k,l}^{\text{sen}} = -0.3953$$

$$\mu_k^{\text{insen}} = -0.2526, \mu_l^{\text{insen}} = -0.03847, \sigma_k^{\text{insen}} = 0.7849, \sigma_l^{\text{insen}} = 1.074, \rho_{k,l}^{\text{insen}} = 0.494$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

5 Rule 22

Gene 1: ESTs Chr.X [48536 (E) 5':H14669 3':H14579]

Gene 2: SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J

WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]

Drug: Clomesone

10 Parameters:

$$\mu_k^{\text{sen}} = -0.8957, \mu_l^{\text{sen}} = -1.079, \sigma_k^{\text{sen}} = 0.7433, \sigma_l^{\text{sen}} = 0.7048, \rho_{k,l}^{\text{sen}} = -0.6495$$

$$\mu_k^{\text{insen}} = 0.2117, \mu_l^{\text{insen}} = 0.2564, \sigma_k^{\text{insen}} = 0.8949, \sigma_l^{\text{insen}} = 0.8653, \rho_{k,l}^{\text{insen}} = -0.08726$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

15

Rule 23

Gene 1: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

Gene 2: SID W 488333 ESTs [5':AA046755 3':AA046642]

20 Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = 1.184, \mu_l^{\text{sen}} = -0.1604, \sigma_k^{\text{sen}} = 0.9042, \sigma_l^{\text{sen}} = 0.8711, \rho_{k,l}^{\text{sen}} = -0.1011$$

$$\mu_k^{\text{insen}} = -0.2817, \mu_l^{\text{insen}} = 0.03825, \sigma_k^{\text{insen}} = 0.7835, \sigma_l^{\text{insen}} = 1.011, \rho_{k,l}^{\text{insen}} = 0.4544$$

$$25 \quad P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

Rule 24

Gene 1: ESTs Chr.8 [470141 (IW) 5':AA029870 3':AA029318]

Gene 2: SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]

30

Drug: Clomesone

Parameters:

$$\mu_k^{\text{sen}} = 0.4978, \mu_l^{\text{sen}} = 1.184, \sigma_k^{\text{sen}} = 0.4895, \sigma_l^{\text{sen}} = 0.9042, \rho_{k,l}^{\text{sen}} = 0.6156$$

$$\mu_k^{\text{insen}} = -0.1176, \mu_l^{\text{insen}} = -0.2817, \sigma_k^{\text{insen}} = 1.056, \sigma_l^{\text{insen}} = 0.7835, \rho_{k,l}^{\text{insen}} = 0.1011$$

$$P(C_i^{\text{sensitive}}) = 0.1917, P(C_i^{\text{insensitive}}) = 0.8083$$

5 Rule 25

Gene 1: BINDING REGULATORY FACTOR Chr.1 [485933 (IW) 5':AA040819 3':AA040156]

Gene 2: SID 43555 MALATE OXIDOREDUCTASE [5':H13370 3':H06037]

Drug: Fluorouracil (5FU)

10 Parameters:

$$\mu_k^{\text{sen}} = 0.5584, \mu_l^{\text{sen}} = 0.9686, \sigma_k^{\text{sen}} = 1.073, \sigma_l^{\text{sen}} = 0.4053, \rho_{k,l}^{\text{sen}} = -0.839$$

$$\mu_k^{\text{insen}} = -0.1082, \mu_l^{\text{insen}} = -0.1883, \sigma_k^{\text{insen}} = 0.9367, \sigma_l^{\text{insen}} = 0.9657, \rho_{k,l}^{\text{insen}} = -0.3566$$

$$P(C_i^{\text{sensitive}}) = 0.1628, P(C_i^{\text{insensitive}}) = 0.8372$$

15

Rule 26

Gene 1: ESTsSID 327435 [5':W32467 3':W19830]

Gene 2: SID 289361 ESTs [5':N99589 3':N92652]

Drug: Fluorouracil (5FU)

20 Parameters:

$$\mu_k^{\text{sen}} = 0.9982, \mu_l^{\text{sen}} = 0.03614, \sigma_k^{\text{sen}} = 1.157, \sigma_l^{\text{sen}} = 0.186, \rho_{k,l}^{\text{sen}} = -0.4795$$

$$\mu_k^{\text{insen}} = -0.1943, \mu_l^{\text{insen}} = -0.007432, \sigma_k^{\text{insen}} = 0.8258, \sigma_l^{\text{insen}} = 1.074, \rho_{k,l}^{\text{insen}} = 0.09915$$

$$P(C_i^{\text{sensitive}}) = 0.1628, P(C_i^{\text{insensitive}}) = 0.8372$$

25

Rule 27

Gene 1: ESTsSID 327435 [5':W32467 3':W19830]

Gene 2: H.sapiens mRNA for Gal-beta(1-3/1-4)GlcNAc alpha-2.3-sialyltransferase Chr.11 [324181 (IW) 5':W47425 3':W47395]

30 Drug: Fluorouracil (5FU)

Parameters:

$$\mu_k^{\text{sen}} = 0.9982, \mu_l^{\text{sen}} = -0.3532, \sigma_k^{\text{sen}} = 1.157, \sigma_l^{\text{sen}} = 0.2383, \rho_{k,l}^{\text{sen}} = 0.01963$$

$$\mu_k^{\text{insen}} = -0.1943, \mu_l^{\text{insen}} = 0.06805, \sigma_k^{\text{insen}} = 0.8258, \sigma_l^{\text{insen}} = 1.049, \rho_{k,l}^{\text{insen}} =$$

0.2537

$$P(C_i^{\text{sensitive}}) = 0.1628, P(C_i^{\text{insensitive}}) = 0.8372$$

Rule 28

- 5 Gene 1: SID W 116819 Homo sapiens clone 23887 mRNA sequence [5':T93821 3':T93776]

Gene 2: ELONGATION FACTOR TU MITOCHONDRIAL PRECURSOR Chr.16 [429540 (IW) 5':AA011453 3':AA011397]

Drug: Fluorodopan

- 10 Parameters:

$$\mu_k^{\text{sen}} = 0.4215, \mu_l^{\text{sen}} = -0.3324, \sigma_k^{\text{sen}} = 1.115, \sigma_l^{\text{sen}} = 1.519, \rho_{k,l}^{\text{sen}} = 0.5573$$

$$\mu_k^{\text{insen}} = -0.1101, \mu_l^{\text{insen}} = 0.0863, \sigma_k^{\text{insen}} = 0.9491, \sigma_l^{\text{insen}} = 0.7573, \rho_{k,l}^{\text{insen}} = -$$

0.786

$$P(C_i^{\text{sensitive}}) = 0.2061, P(C_i^{\text{insensitive}}) = 0.7939$$

15

Rule 29

Gene 1: ESTs Chr.14 [244047 (I) 5':N45439 3':N38807]

Gene 2: SID 307717 Homo sapiens KIAA0430 mRNA complete cds [5': 3':N92942]

Drug: Cyclocytidine

- 20 Parameters:

$$\mu_k^{\text{sen}} = 0.536, \mu_l^{\text{sen}} = 0.004825, \sigma_k^{\text{sen}} = 0.4307, \sigma_l^{\text{sen}} = 0.232, \rho_{k,l}^{\text{sen}} = 0.1655$$

$$\mu_k^{\text{insen}} = -0.1816, \mu_l^{\text{insen}} = -0.002083, \sigma_k^{\text{insen}} = 1.03, \sigma_l^{\text{insen}} = 1.151, \rho_{k,l}^{\text{insen}} =$$

0.08986

$$P(C_i^{\text{sensitive}}) = 0.2533, P(C_i^{\text{insensitive}}) = 0.7467$$

25

Rule 30

Gene 1: SID W 510230 Homo sapiens (clone CC6) NADH-ubiquinone oxidoreductase subunit mRNA 3' end cds [5':AA053568 3':AA053557]

Gene 2: SID 307717 Homo sapiens KIAA0430 mRNA complete cds [5': 3':N92942]

- 30 Drug: Cyclocytidine

Parameters:

$$\mu_k^{\text{sen}} = 0.1566, \mu_l^{\text{sen}} = 0.004825, \sigma_k^{\text{sen}} = 0.4745, \sigma_l^{\text{sen}} = 0.232, \rho_{k,l}^{\text{sen}} = -0.4326$$

$$\mu_k^{\text{insen}} = -0.05336, \mu_l^{\text{insen}} = -0.002083, \sigma_k^{\text{insen}} = 1.116, \sigma_l^{\text{insen}} = 1.151, \rho_{k,l}^{\text{insen}} =$$

0.3113

$$P(C_i^{\text{sensitive}}) = 0.2533, P(C_i^{\text{insensitive}}) = 0.7467$$

Rule 31

- 5 Gene 1: DNA POLYMERASE EPSILON CATALYTIC SUBUNIT A Chr.12 [321207 (IW) 5':W52910 3':AA037353]

Gene 2: SID 307717 Homo sapiens KIAA0430 mRNA complete cds [5': 3':N92942]

Drug: Cyclocytidine

Parameters:

- 10 $\mu_k^{\text{sen}} = 0.7918, \mu_l^{\text{sen}} = 0.004825, \sigma_k^{\text{sen}} = 1.042, \sigma_l^{\text{sen}} = 0.232, \rho_{k,l}^{\text{sen}} = 0.176$
 $\mu_k^{\text{insen}} = -0.2694, \mu_l^{\text{insen}} = -0.002083, \sigma_k^{\text{insen}} = 0.762, \sigma_l^{\text{insen}} = 1.151, \rho_{k,l}^{\text{insen}} = -$
0.06434

$$P(C_i^{\text{sensitive}}) = 0.2533, P(C_i^{\text{insensitive}}) = 0.7467$$

- 15 Rule 32

Gene 1: TXNRD1 Thioredoxin reductase Chr.12 [510377 (IW) 5':AA055407 3':AA055408]

Gene 2: ESTs Chr.1 [362126 (I) 5':AA001086 3':AA001049]

Drug: Mitomycin

- 20 Parameters:

$$\mu_k^{\text{sen}} = 0.9736, \mu_l^{\text{sen}} = -0.4653, \sigma_k^{\text{sen}} = 0.752, \sigma_l^{\text{sen}} = 0.3908, \rho_{k,l}^{\text{sen}} = 0.1693$$

$$\mu_k^{\text{insen}} = -0.2247, \mu_l^{\text{insen}} = 0.107, \sigma_k^{\text{insen}} = 0.8952, \sigma_l^{\text{insen}} = 1.053, \rho_{k,l}^{\text{insen}} =$$

0.3972

$$P(C_i^{\text{sensitive}}) = 0.1872, P(C_i^{\text{insensitive}}) = 0.8128$$

- 25

Rule 33

Gene 1: SID W 260223 Human mRNA for BST-1 complete cds [5':N45417 3':N32106]

Gene 2: TXNRD1 Thioredoxin reductase Chr.12 [510377 (IW) 5':AA055407 3':AA055408]

- 30 Drug: Mitomycin

Parameters:

$$\mu_k^{\text{sen}} = 0.1887, \mu_l^{\text{sen}} = 0.9736, \sigma_k^{\text{sen}} = 0.6724, \sigma_l^{\text{sen}} = 0.752, \rho_{k,l}^{\text{sen}} = 0.7526$$

$$\mu_k^{\text{insen}} = -0.04347, \mu_l^{\text{insen}} = -0.2247, \sigma_k^{\text{insen}} = 1.003, \sigma_l^{\text{insen}} = 0.8952, \rho_{k,l}^{\text{insen}} = -$$

0.007584

$$P(C_i^{\text{sensitive}}) = 0.1872, P(C_i^{\text{insensitive}}) = 0.8128$$

Rule 34

- 5 Gene 1: SCYA2 Small inducible cytokine A2 (monocyte chemotactic protein 1 homologous to mouse Sig-je) Chr.17 [108837 (DIW) 5':T77816 3':T77817]

Gene 2: *Carbonic anhydrase II SID 429288 [5':AA007456 3':AA007360]

Drug: Anthrapyrazole-derivative

Parameters:

$$10 \quad \mu_k^{\text{sen}} = 0.8903, \mu_l^{\text{sen}} = -0.3723, \sigma_k^{\text{sen}} = 0.9679, \sigma_l^{\text{sen}} = 0.694, \rho_{k,l}^{\text{sen}} = -0.4114$$

$$\mu_k^{\text{insen}} = -0.224, \mu_l^{\text{insen}} = 0.09341, \sigma_k^{\text{insen}} = 0.8509, \sigma_l^{\text{insen}} = 1.03, \rho_{k,l}^{\text{insen}} = 0.4247$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

- 15 Rule 35

Gene 1: SID 356851 Homo sapiens mRNA for nucleolar protein hNop56 [5':3':W86238]

Gene 2: Human extracellular protein (S1-5) mRNA complete cds Chr.2 [485875 (EW) 5':AA040442 3':AA040443]

- 20 Drug: Anthrapyrazole-derivative

Parameters:

$$\mu_k^{\text{sen}} = -0.216, \mu_l^{\text{sen}} = 1.016, \sigma_k^{\text{sen}} = 0.6331, \sigma_l^{\text{sen}} = 1.089, \rho_{k,l}^{\text{sen}} = -0.6461$$

$$\mu_k^{\text{insen}} = 0.05396, \mu_l^{\text{insen}} = -0.2548, \sigma_k^{\text{insen}} = 1, \sigma_l^{\text{insen}} = 0.7749, \rho_{k,l}^{\text{insen}} = 0.2101$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

- 25

Rule 36

Gene 1: ALDH10 Aldehyde dehydrogenase 10 (fatty aldehyde dehydrogenase) Chr.17 [208950 (EW) 5':H63829 3':H63779]

Gene 2: SID W 488148 H.sapiens mRNA for 3'UTR of unknown protein [5':AA057239 3':AA058703]

- 30

Drug: Anthrapyrazole-derivative

Parameters:

$$\mu_k^{\text{sen}} = 0.6212, \mu_l^{\text{sen}} = 0.843, \sigma_k^{\text{sen}} = 0.6852, \sigma_l^{\text{sen}} = 0.575, \rho_{k,l}^{\text{sen}} = 0.2169$$

$$\mu_k^{\text{insen}} = -0.1554, \mu_l^{\text{insen}} = -0.2115, \sigma_k^{\text{insen}} = 0.9606, \sigma_l^{\text{insen}} = 0.9263, \rho_{k,l}^{\text{insen}} = -0.3119$$

$$P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

5 Rule 37

Gene 1: Human extracellular protein (S1-5) mRNA complete cds Chr.2 [485875 (EW) 5':AA040442 3':AA040443]

Gene 2: SID W 415693 Homo sapiens mRNA for phosphatidylinositol 4-kinase. complete cds [5':W78879 3':W84724]

10 Drug: Anthrapyrazole-derivative

Parameters:

$$\mu_k^{\text{sen}} = 1.016, \mu_l^{\text{sen}} = 0.3712, \sigma_k^{\text{sen}} = 1.089, \sigma_l^{\text{sen}} = 0.4463, \rho_{k,l}^{\text{sen}} = -0.3426$$

$$\mu_k^{\text{insen}} = -0.2548, \mu_l^{\text{insen}} = -0.09229, \sigma_k^{\text{insen}} = 0.7749, \sigma_l^{\text{insen}} = 1.066, \rho_{k,l}^{\text{insen}} =$$

$$0.341$$

$$15 \quad P(C_i^{\text{sensitive}}) = 0.2006, P(C_i^{\text{insensitive}}) = 0.7994$$

Rule 38

Gene 1: SID W 345683 ESTs Highly similar to INTEGRAL MEMBRANE GLYCOPROTEIN GP210 PRECURSOR [Rattus norvegicus] [5':W76432 3':W72039]

20 Gene 2: Human mRNA for KIAA0143 gene partial cds Chr.8 [488462 (IW) 5':AA047508 3':AA047451]

Drug: Daunorubicin

Parameters:

$$\mu_k^{\text{sen}} = 0.918, \mu_l^{\text{sen}} = -0.6559, \sigma_k^{\text{sen}} = 0.3704, \sigma_l^{\text{sen}} = 0.4622, \rho_{k,l}^{\text{sen}} = -0.5746$$

$$25 \quad \mu_k^{\text{insen}} = -0.2022, \mu_l^{\text{insen}} = 0.1457, \sigma_k^{\text{insen}} = 0.9271, \sigma_l^{\text{insen}} = 1.007, \rho_{k,l}^{\text{insen}} = -0.009774$$

$$P(C_i^{\text{sensitive}}) = 0.1811, P(C_i^{\text{insensitive}}) = 0.8189$$

Rule 39

30 Gene 1: SID W 162077 ESTs [5':H25689 3':H26271]

Gene 2: SID W 197549 ESTs [5':R87793 3':R87731]

Drug: Deoxydoxorubicin

Parameters:

$$\mu_k^{\text{sen}} = -0.2102, \mu_l^{\text{sen}} = -0.1107, \sigma_k^{\text{sen}} = 0.3133, \sigma_l^{\text{sen}} = 0.9712, \rho_{k,l}^{\text{sen}} = -0.98$$

$$\mu_k^{\text{insen}} = 0.03539, \mu_l^{\text{insen}} = 0.01824, \sigma_k^{\text{insen}} = 1.068, \sigma_l^{\text{insen}} = 1.008, \rho_{k,l}^{\text{insen}} = 0.1725$$

$$P(C_i^{\text{sensitive}}) = 0.1428, P(C_i^{\text{insensitive}}) = 0.8572$$

5

Rule 40

Gene 1: ELONGATION FACTOR TU MITOCHONDRIAL PRECURSOR Chr.16

[429540 (IW) 5':AA011453 3':AA011397]

Gene 2: ESTs Chr.2 [365120 (IW) 5':AA025204 3':AA025124]

10 Drug: Amsacrine

Parameters:

$$\mu_k^{\text{sen}} = -0.7939, \mu_l^{\text{sen}} = 0.558, \sigma_k^{\text{sen}} = 1.022, \sigma_l^{\text{sen}} = 1.102, \rho_{k,l}^{\text{sen}} = 0.7045$$

$$\mu_k^{\text{insen}} = 0.2239, \mu_l^{\text{insen}} = -0.1576, \sigma_k^{\text{insen}} = 0.791, \sigma_l^{\text{insen}} = 0.8965, \rho_{k,l}^{\text{insen}} = 0.4064$$

$$15 \quad P(C_i^{\text{sensitive}}) = 0.22, P(C_i^{\text{insensitive}}) = 0.78$$

Rule 41

Gene 1: G6PD Glucose-6-phosphate dehydrogenase Chr.X [430251 (IW) 5':AA010317

3':AA010382]

20 Gene 2: SID W 376708 ESTs [5':AA046358 3':AA046274]

Drug: CPT,20-ester (S)

Parameters:

$$\mu_k^{\text{sen}} = -0.09704, \mu_l^{\text{sen}} = -0.6823, \sigma_k^{\text{sen}} = 0.4911, \sigma_l^{\text{sen}} = 0.8524, \rho_{k,l}^{\text{sen}} = 0.7542$$

$$25 \quad \mu_k^{\text{insen}} = 0.02995, \mu_l^{\text{insen}} = 0.2092, \sigma_k^{\text{insen}} = 1.068, \sigma_l^{\text{insen}} = 0.9393, \rho_{k,l}^{\text{insen}} = -0.5785$$

$$P(C_i^{\text{sensitive}}) = 0.2344, P(C_i^{\text{insensitive}}) = 0.7656$$

Rule 42

Gene 1: H.sapiens mRNA for ESM-1 protein Chr.5 [324122 (RW) 5':W46667

30 3':W46577]

Gene 2: Human FEZ2 mRNA partial cds Chr.2 [488055 (IW) 5':AA058551

3':AA053303]

Drug: CPT

Parameters:

$$\mu_k^{\text{sen}} = -0.1032, \mu_l^{\text{sen}} = 0.8185, \sigma_k^{\text{sen}} = 0.4146, \sigma_l^{\text{sen}} = 0.8985, \rho_{k,l}^{\text{sen}} = -0.6229$$

$$\mu_k^{\text{insen}} = 0.03592, \mu_l^{\text{insen}} = -0.2863, \sigma_k^{\text{insen}} = 1.124, \sigma_l^{\text{insen}} = 0.8401, \rho_{k,l}^{\text{insen}} =$$

0.4189

$$5 \quad P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

Rule 43

Gene 1: SID W 361023 ESTs [5':AA013072 3':AA012983]

Gene 2: H.sapiens mRNA for TRAMP protein Chr.8 [149355 (IEW) 5':H01598

10 3':H01495]

Drug: CPT

Parameters:

$$\mu_k^{\text{sen}} = -0.6506, \mu_l^{\text{sen}} = 0.5667, \sigma_k^{\text{sen}} = 0.6739, \sigma_l^{\text{sen}} = 1.274, \rho_{k,l}^{\text{sen}} = 0.7093$$

$$\mu_k^{\text{insen}} = 0.2279, \mu_l^{\text{insen}} = -0.1978, \sigma_k^{\text{insen}} = 0.9778, \sigma_l^{\text{insen}} = 0.7508, \rho_{k,l}^{\text{insen}} = -$$

15 0.1771

$$P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

Rule 44

Gene 1: SID W 358754 Human mRNA for cysteine protease complete cds [5':W94449

20 3':W94332]

Gene 2: SID W 159512 Integrin alpha 6 [5':H16046 3':H15934]

Drug: CPT

Parameters:

$$\mu_k^{\text{sen}} = -0.1082, \mu_l^{\text{sen}} = 0.7291, \sigma_k^{\text{sen}} = 0.7356, \sigma_l^{\text{sen}} = 0.6557, \rho_{k,l}^{\text{sen}} = -0.6645$$

$$25 \quad \mu_k^{\text{insen}} = 0.0372, \mu_l^{\text{insen}} = -0.2559, \sigma_k^{\text{insen}} = 1.038, \sigma_l^{\text{insen}} = 0.9638, \rho_{k,l}^{\text{insen}} =$$

0.4712

$$P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

Rule 45

30 Gene 1: SID 257009 ESTs [5':N39759 3':N26801]

Gene 2: SID W 488148 H.sapiens mRNA for 3'UTR of unknown protein [5':AA057239

3':AA058703]

Drug: CPT

Parameters:

$$\mu_k^{\text{sen}} = 0.3448, \mu_l^{\text{sen}} = 0.8224, \sigma_k^{\text{sen}} = 0.7661, \sigma_l^{\text{sen}} = 0.5588, \rho_{k,l}^{\text{sen}} = 0.6149$$

$$\mu_k^{\text{insen}} = -0.1208, \mu_l^{\text{insen}} = -0.2881, \sigma_k^{\text{insen}} = 1.029, \sigma_l^{\text{insen}} = 0.9329, \rho_{k,l}^{\text{insen}} = 0.06046$$

$$5 \quad P(C_i^{\text{sensitive}}) = 0.2594, P(C_i^{\text{insensitive}}) = 0.7406$$

Rule 46

Gene 1: SID 43609 ESTs [5':H06454 3':H06184]

Gene 2: SID W 361023 ESTs [5':AA013072 3':AA012983]

10 Drug: CPT,20-ester (S)

Parameters:

$$\mu_k^{\text{sen}} = 0.4667, \mu_l^{\text{sen}} = -0.6333, \sigma_k^{\text{sen}} = 1.301, \sigma_l^{\text{sen}} = 0.554, \rho_{k,l}^{\text{sen}} = 0.5266$$

$$\mu_k^{\text{insen}} = -0.1602, \mu_l^{\text{insen}} = 0.2168, \sigma_k^{\text{insen}} = 0.7751, \sigma_l^{\text{insen}} = 0.9858, \rho_{k,l}^{\text{insen}} = 0.2268$$

$$15 \quad P(C_i^{\text{sensitive}}) = 0.255, P(C_i^{\text{insensitive}}) = 0.745$$

Rule 47

Gene 1: Human G/T mismatch-specific thymine DNA glycosylase mRNA complete cds
Chr.X [321997 (IW) 5':W37234 3':W37817]

20 Gene 2: SID W 358526 ESTs [5':W96039 3':W94821]

Drug: CPT,11-formyl (RS)

Parameters:

$$\mu_k^{\text{sen}} = 0.626, \mu_l^{\text{sen}} = -1.055, \sigma_k^{\text{sen}} = 1.041, \sigma_l^{\text{sen}} = 1.241, \rho_{k,l}^{\text{sen}} = -0.1072$$

$$\mu_k^{\text{insen}} = -0.151, \mu_l^{\text{insen}} = 0.2536, \sigma_k^{\text{insen}} = 0.9295, \sigma_l^{\text{insen}} = 0.7034, \rho_{k,l}^{\text{insen}} =$$

25 0.6208

$$P(C_i^{\text{sensitive}}) = 0.1939, P(C_i^{\text{insensitive}}) = 0.8061$$

Rule 48

Gene 1: PROTEASOME COMPONENT C13 PRECURSOR Chr.6 [344774 (IW)

30 5':W74742 3':W74705]

Gene 2: SID W 484681 Homo sapiens ES/130 mRNA complete cds [5':AA037568
3':AA037487]

Drug: Mechlorethamine

Parameters:

$$\mu_k^{\text{sen}} = 0.6562, \mu_l^{\text{sen}} = -0.8883, \sigma_k^{\text{sen}} = 0.7248, \sigma_l^{\text{sen}} = 0.7952, \rho_{k,l}^{\text{sen}} = -0.1383$$

$$\mu_k^{\text{insen}} = -0.1565, \mu_l^{\text{insen}} = 0.2119, \sigma_k^{\text{insen}} = 0.9825, \sigma_l^{\text{insen}} = 0.9257, \rho_{k,l}^{\text{insen}} = 0.6324$$

$$5 \quad P(C_i^{\text{sensitive}}) = 0.1928, P(C_i^{\text{insensitive}}) = 0.8072$$

Rule 49

Gene 1: AK1 Adenylate kinase 1 Chr.9 [488381 (IW) 5':AA046783 3':AA046653]

Gene 2: Human vascular endothelial growth factor related protein VRP mRNA complete

10 cds Chr.4 [309535 (I) 5': 3':N94399]

Drug: Mechlorethamine

Parameters:

$$\mu_k^{\text{sen}} = -0.4881, \mu_l^{\text{sen}} = -0.243, \sigma_k^{\text{sen}} = 1.786, \sigma_l^{\text{sen}} = 0.4893, \rho_{k,l}^{\text{sen}} = 0.8105$$

$$\mu_k^{\text{insen}} = 0.1157, \mu_l^{\text{insen}} = 0.05762, \sigma_k^{\text{insen}} = 0.6286, \sigma_l^{\text{insen}} = 1.08, \rho_{k,l}^{\text{insen}} =$$

15 0.03238

$$P(C_i^{\text{sensitive}}) = 0.1928, P(C_i^{\text{insensitive}}) = 0.8072$$

Rule 50

Gene 1: SID W 489301 ESTs [5':AA054471 3':AA058511]

20 Gene 2: Human epithelial membrane protein (CL-20) mRNA complete cds Chr.12
[488719 (IW) 5':AA046077 3':AA046025]

Drug: Melphalan

Parameters:

$$\mu_k^{\text{sen}} = 0.9792, \mu_l^{\text{sen}} = -0.619, \sigma_k^{\text{sen}} = 1.075, \sigma_l^{\text{sen}} = 0.7439, \rho_{k,l}^{\text{sen}} = -0.8227$$

$$25 \quad \mu_k^{\text{insen}} = -0.2399, \mu_l^{\text{insen}} = 0.1515, \sigma_k^{\text{insen}} = 0.7994, \sigma_l^{\text{insen}} = 0.9531, \rho_{k,l}^{\text{insen}} = 0.3178$$

$$P(C_i^{\text{sensitive}}) = 0.1967, P(C_i^{\text{insensitive}}) = 0.8033$$

Rule 51

30 Gene 1: SID W 245450 Human transcription factor NFATx mRNA complete cds
[5':N77274 3':N55066]

Gene 2: SID W 485645 KERATIN TYPE II CYTOSKELETAL 7 [5':AA039817
3':AA041344]

Drug: 5-Hydroxypicolinaldehyde-thiose

Parameters:

$$\mu_k^{\text{sen}} = 0.122, \mu_l^{\text{sen}} = 0.8712, \sigma_k^{\text{sen}} = 0.2463, \sigma_l^{\text{sen}} = 0.6735, \rho_{k,l}^{\text{sen}} = 0.1308$$

$$\mu_k^{\text{insen}} = -0.02658, \mu_l^{\text{insen}} = -0.1896, \sigma_k^{\text{insen}} = 1.091, \sigma_l^{\text{insen}} = 0.9271, \rho_{k,l}^{\text{insen}} =$$

5 0.05545

$$P(C_i^{\text{sensitive}}) = 0.1789, P(C_i^{\text{insensitive}}) = 0.8211$$

Rule 52

Gene 1: SID 381780 ESTs [5':AA059257 3':AA059223]

10 Gene 2: SID 512355 ESTs Highly similar to SRC SUBSTRATE P80/85 PROTEINS
[Gallus gallus] [5':AA059424 3':AA057835]

Drug: Paclitaxel---Taxol

Parameters:

$$\mu_k^{\text{sen}} = 0.1618, \mu_l^{\text{sen}} = -0.8354, \sigma_k^{\text{sen}} = 0.1828, \sigma_l^{\text{sen}} = 0.4935, \rho_{k,l}^{\text{sen}} = -0.09957$$

$$15 \mu_k^{\text{insen}} = -0.03218, \mu_l^{\text{insen}} = 0.162, \sigma_k^{\text{insen}} = 1.06, \sigma_l^{\text{insen}} = 0.9902, \rho_{k,l}^{\text{insen}} = -$$

0.09191

$$P(C_i^{\text{sensitive}}) = 0.1622, P(C_i^{\text{insensitive}}) = 0.8378$$

Rule 53

20 Gene 1: SID 381780 ESTs [5':AA059257 3':AA059223]

Gene 2: SID 130482 ESTs [5':R21876 3':R21877]

Drug: Paclitaxel---Taxol

Parameters:

$$\mu_k^{\text{sen}} = 0.1618, \mu_l^{\text{sen}} = -0.9271, \sigma_k^{\text{sen}} = 0.1828, \sigma_l^{\text{sen}} = 0.3413, \rho_{k,l}^{\text{sen}} = -0.3935$$

$$25 \mu_k^{\text{insen}} = -0.03218, \mu_l^{\text{insen}} = 0.1791, \sigma_k^{\text{insen}} = 1.06, \sigma_l^{\text{insen}} = 0.9842, \rho_{k,l}^{\text{insen}} = -$$

0.2741

$$P(C_i^{\text{sensitive}}) = 0.1622, P(C_i^{\text{insensitive}}) = 0.8378$$

Rule 54

30 Gene 1: SID 344786 Human mRNA for KIAA0177 gene partial cds [5': 3':W74713]

Gene 2: TXNRD1 Thioredoxin reductase Chr.12 [510377 (IW) 5':AA055407
3':AA055408]

Drug: Bisantrene

Parameters:

$$\begin{aligned} \mu_k^{\text{sen}} &= -0.3189, \mu_l^{\text{sen}} = 1.298, \sigma_k^{\text{sen}} = 0.6532, \sigma_l^{\text{sen}} = 0.7515, \rho_{k,l}^{\text{sen}} = 0.9897 \\ \mu_k^{\text{insen}} &= 0.02732, \mu_l^{\text{insen}} = -0.1115, \sigma_k^{\text{insen}} = 0.9915, \sigma_l^{\text{insen}} = 0.9088, \rho_{k,l}^{\text{insen}} = \\ &0.06623 \\ 5 \quad P(C_i^{\text{sensitive}}) &= 0.07889, P(C_i^{\text{insensitive}}) = 0.9211 \end{aligned}$$

Determining Statistical Significance of Finding

Mean Square Error (MSE) scores are calculated by comparing the probabilities (a form of likelihood) computed by a method against an ensemble of surrogate data generated by different *randomizations*, i.e., permutations, of the original data (creating artificial samples). A resulting histogram of MSE scores is then interpreted as representing the probability distribution of error; hence, the statistical significance of any given determined probability can be assigned. The gene expression levels can then be selected according to the ranking of their probability for the original data, with a comparison against the MSE score for the randomized data.

Validating Predictions of Sensitivity to Drug, for each Method

For any given gene k and drug I , a cross-validation procedure is used to assess validity of any prediction. For example, we omit 1 given cell line from consideration, and carry out a given method on the remaining cell lines, and record the findings. The omitted cell line is restored and a different cell line is omitted, and the given method re-applied. This is repeated, one cell line at a time, until all the cell lines have had their turn being omitted. All the findings are compiled. Difference scores between an original calculation and a cell line-omitted calculation are obtained. Mean Square Errors (MSE) are then calculated from the aggregated differences. MSE is then an assessment of the validity of the given method.

Sample results from one of the Bayesian classifiers (the LDA 2D) on the NCI60 dataset are shown in Table 8 below.

30

Table 8

Drug	Gene 1	Gene 2	P-Value	Statistical Significance - After Bonferroni Correction
Acivicin (RNA synthesis inhibitor)	Glyoxalase-I-log	Homo sapiens mRNA for HYA22 complete cds Chr.3 [358957 (EW) 5':W91969 3':W94916]	5.947e-08	3.00%
Baker's-soluble- antifolate (antifol)	SID W 254085 ESTs Moderately similar to synaptonemal complex protein [M.musculus] [5':N71532 3':N22165]	SID 118593 [5':T92821 3':T92741]	1.982e-08	1.00%
Baker's-soluble- antifolate (antifol)	SID W 254085 ESTs Moderately similar to synaptonemal complex protein [M.musculus] [5':N71532 3':N22165]	ESTs Chr.5 [46694 (RW) 5':H10240 3':H10192]	1.586e-07	7.90%
Mitozolamide (alkylating agent, guanine-O6)	SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]	*Hs.648 Cut (Drosophila)-like 1 (CCAAT displacement protein) SID W 26677 ESTs [5':R13994 3':R39117]	5.947e-08	3.00%
Mitozolamide (alkylating agent, guanine-O6)	Homo sapiens delta7- sterol reductase mRNA complete cds Chr.10 [417125 (E) 5':	SID W 380674 ESTs [5':AA053720 3':AA053711]	1.388e-07	6.90%

	3':W87472]			
Mitozolamide (alkylating agent, guanine-O6)	Glutathione S- Transferase Pi-log	*Hs.648 Cut (Drosophila)-like 1 (CCAAT displacement protein) SID W 26677 ESTs [5':R13994 3':R39117]	1.982e-07	9.90%
Clomesone (alkylating agent, guanine-O6)	ESTs Chr.X [48536 (E) 5':H14669 3':H14579]	SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]	1.982e-08	1.00%
Clomesone (alkylating agent, guanine-O6)	SID W 36809 Homo sapiens neural cell adhesion molecule (CALL) mRNA complete cds [5':R34648 3':R49177]	SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]	1.982e-08	1.00%
Clomesone (alkylating agent, guanine-O6)	M-PHASE INDUCER PHOSPHATASE 2 Chr.20 [179373 (EW) 5':H50437 3':H50438]	SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]	3.964e-08	2.00%
Clomesone (alkylating agent, guanine-O6)	SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]	SID 469842 Homo sapiens mRNA for fatty acid binding protein complete cds [5':AA029794 3':AA029795]	3.964e-08	2.00%

Clomesone (alkylating agent, guanine-O6)	ESTsSID 327435 [5':W32467 3':W19830]	SID 469842 Homo sapiens mRNA for fatty acid binding protein complete cds [5':AA029794 3':AA029795]	3.964e-08	2.00%
Clomesone (alkylating agent, guanine-O6)	SID 512164 Human clathrin assembly protein 50 (AP50) mRNA complete cds [5': 3':AA057396]	SID W 345624 Human homeobox protein (PHOX1) mRNA 3' end [5':W76402 3':W72050]	3.964e-08	2.00%
Clomesone (alkylating agent, guanine-O6)	SID W 376951 ESTs [5':AA047756 3':AA047641]	SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]	3.964e-08	2.00%
Clomesone (alkylating agent, guanine-O6)	Glutathoine S- Tranferase Pi-log	SID W 487535 Human mRNA for KIAA0080 gene partial cds [5':AA043528 3':AA043529]	9.911e-08	5.00%
Clomesone (alkylating agent, guanine-O6)	XRCC4 DNA repair protein XRCC4 Chr.5 [26811 (RW) 5':R14027 3':R39148]	SID W 242844 ESTs Moderately similar to !!!! ALU SUBFAMILY J WARNING ENTRY !!!! [H.sapiens] [5':H94138 3':H94064]	9.911e-08	5.00%

The above steps as performed on, by way of example, the NCI60 dataset can be further explained as follows.

Start off with 2 tables of data: a table, T, with gene expression data and a table, A, with drug concentration data. In table T each column is a gene, each row is a cell line and
 5 each entry is the expression level of a gene in a given cell line.

In table A, each column is a drug, each row is a cell line (corresponding exactly to the same cell lines in table T) and each entry is the drug concentration which inhibits the growth of a given cell line by 50%.

Note: The same cell lines appear in Tables T and A, and the order of the cell lines is the
 10 same in both tables. In the NCI60 analysis there were 60 cell lines, 1000 genes and 90 drugs.

Table T

	Gene 1	Gene 2	Gene 3
Cell line 1	0.4	0.2	0.8
Cell line 2	0.5	0.4	0.3
Cell line 3	0.2	0.7	0.1

15 Table A

	Drug 1	Drug 2	Drug 3
Cell line 1	0.6	1.1	1.8
Cell line 2	0.1	0.4	0.3
Cell line 3	0.5	0.1	0.1

An example of Tables T and A with actual data are shown below:

Table T: Gene expression values

	Gene: SID W 328550 ATL- derived PMA- responsive (APR) peptide [5':W40533 3':W40261]	Gene: RAC2 Ras- related C3 botulinum toxin substrate 2 Chr.22 [429908 (DI) 5': 3':AA033975]	Gene: Human GDP-dissociation inhibitor protein (Ly-GDI) mRNA complete cds Chr.12 [487374 (IW) 5':AA046482 3':AA046695]
Cell line: CNS:SNB-19	-1.17	-0.93	-0.62
Cell line: CNS:U251	0.19	0.1	-0.77
Cell line: BR:BT-549	-1.2	-0.1	-0.45

Table A: -logGI50 values

	Drug: Thiopurine (6MP)	Drug: alpha-2'- Deoxythioguanosine	Drug: Thioguanine
Cell line: CNS:SNB-19	-2.08	-2.35	-4.14
Cell line: CNS:U251	-0.77	-1.03	-1.63
Cell line: BR:BT-549	-2.36	-1.6	-0.47

1) Transform the drug response values.

Form a new table which corresponds to the A table by transforming the numerical values of Table A so that they fall on a continuous numerical scale ≥ 0 and ≤ 1 . This is done in order to represent the intensity of the attribute in a readily interpretable manner: 0 represents negligible insensitivity (e.g., insensitive to drug) and 1 represents high intensity (e.g., sensitive to drug), with continuous gradation in between.

For example, using equation for the continuous piece-wise linear biological scoring function described previously:

Let a_{ij} represent the entry in the i th row and j th column of table A.

Transform each entry, a_{ij} , as follows:

if a_{ij} is less than 0.3 then set $a_{ij} = 0$

if a_{ij} is between 0.3 and 0.7, then set $a_{ij} = (a_{ij} - 0.3) / 0.4$

if a_{ij} is greater than or equal to 0.7, then set $a_{ij} = 1$

If a new entry a_{ij} is > 0 , consider cell line i to be at least partially sensitive to drug j . If a new entry a_{ij} is less < 1 , consider cell line i to be at least partially insensitive to drug j .

Based on the transformed attribute values in some column j , it is possible to separate cell lines into 2 classes, $C^{\text{sensitive}}$ and $C^{\text{insensitive}}$. Cell lines that are sensitive are in class $C^{\text{sensitive}}$ and cell lines that are insensitive are in the $C^{\text{insensitive}}$ class. But, some cell lines can be considered to be partially in both class. For example, if the transformed value $a_{ij} = x$, then cell line i is considered to be $x*100\%$ in class $C^{\text{sensitive}}$ and $(1-x)*100\%$ in class $C^{\text{insensitive}}$.

2) Example Application of Bayesian Classifiers - UGDA 1D, UGDA 2D, LDA 1D, QDA 1D, LDA 2D, QDA 2D.

Note: Steps explained using LDA 1D are equivalently applied for any of the other Bayesian classifiers.

5 Example of Steps:

Apply LDA 1D to measure how well a given gene co-occurs, associates with, or predicts response to a given drug.

- 2.1) Select a column, T_k , from the T matrix, with the expression values of some gene k . Select a column, A_i , from the A matrix, with the drug concentrations (e.g., in units of $-\log_{10}\text{GI50}$) values of some drug i [see paragraph 1d in the Methods document for GI50].
- 2.2) Remove the first entry, $T_{1,k}$, from column T_k and the first entry, $A_{1,i}$, from column A_i . Assume that these entries belong to cell line L_1 .
- 2.3) Separate the remaining entries, ($T_{2,k}$ through $T_{n,k}$) in column T_k into two sets:
 - one set, ${}_iC^{\text{sensitive}}$ has the gene expression values of cell lines at least partially sensitive to drug i (i.e. these cell lines have values greater than 0 in column A_i)
 - a second set, ${}_iC^{\text{insensitive}}$, has the gene expression values of cell lines at least partially insensitive to drug i (i.e. these cell lines have values smaller than 1 in column A_i)
- 2.4) Compute the weighted mean, $\mu_k^{\text{sensitive}}$, and the weighted standard deviation, $\sigma_k^{\text{sensitive}}$, of the values in set ${}_iC^{\text{sensitive}}$.
- 20 Find the weighted mean, $\mu_k^{\text{insensitive}}$, and the weighted standard deviation, $\sigma_k^{\text{insensitive}}$ of the values in set ${}_iC^{\text{insensitive}}$.
- Find the weighted average standard deviation σ_k^{avg} of the two sets.
- Find the frequency, $P({}_iC^{\text{sensitive}})$, of the sensitive class and the frequency, $P({}_iC^{\text{insensitive}})$, of the insensitive class.
- 25 Compute parameters necessary to fit any chosen mathematical density function or continuous curve to a α category-wise histogram of the type described previously.

2.5) Compute the probability, $P(L_i \in C_i^{sensitive} | T_{1,k})$, that cell line L_i is sensitive to drug i , using the information of the expression level of gene k and the proportion, i.e., frequency, of the sensitive and insensitive classes. Namely, compute

$$5 \quad P(L_i \in C_i^{sensitive} | T_{1,k}) = \frac{{}_iG_k^{sensitive}(T_{1,k}) \cdot P(C_i^{sensitive})}{{}_iG_k^{sensitive}(T_{1,k}) \cdot P(C_i^{sensitive}) + {}_iG_k^{insensitive}(T_{1,k}) \cdot P(C_i^{insensitive})},$$

where

$$10 \quad \begin{aligned} {}_iG_k^{sensitive}(T_{1,k}) &= \frac{1}{\sigma_k^{avg} \sqrt{2\pi}} e^{-(T_{1,k} - \mu_k^{sensitive})^2 / 2(\sigma_k^{avg})^2} \\ {}_iG_k^{insensitive}(T_{1,k}) &= \frac{1}{\sigma_k^{avg} \sqrt{2\pi}} e^{-(T_{1,k} - \mu_k^{insensitive})^2 / 2(\sigma_k^{avg})^2} \end{aligned}$$

as described previously.

2.6) Calculate an error for the probability derived in step 2.5.

Consider the probability from step 2.5 to be the expected probability, $P^{expected}$, that cell line L_i is sensitive to drug i . Consider entry $A_{i,i}$ to be the observed probability, $P^{observed}$, that cell line L_i is sensitive to drug i .

Then, calculate an error, E_i , based on these two values, where $E_i = (P^{expected} - P^{observed})^2$.

2.7) A cross-validation procedure.

For each cell line, find the probability of sensitivity to drug i .

20 Restore the first entries of columns T_k and A_i , (entries belonging to cell line L_i) and remove the second entry of these columns. Assume that the removed entries belong to cell line C_2 . Repeat steps 2.3 through 2.6, to obtain the probability of cell line L_2 being sensitive to drug i . Follow the same procedure for each of the cell lines. Find the mean of the error terms, E , from all the iterations. This value is referred to as the mean squared error (MSE). This MSE quantifies how well

25 gene k predicts sensitivity to drug i .

3) Find the MSE scores of all genes versus all drugs.

4) A statistical significance assessment procedure.

Find initial significance p-values for all MSE scores.

5 A significance p-value indicates the likelihood that an MSE score could have arisen by chance (i.e. that randomized data (i.e., the original data, randomly permuted to obliterate any patterns that may have been in the original data) could have generated the MSE score).

4.1) Construct a distribution, i.e., histogram, of MSE scores from the LDA 1D being applied to randomized data.

10 In each column of the T table, randomly rearrange the order of the entries. In each column of the A table, randomly rearrange the order of the entries. Make copies of these two tables, and again randomly rearrange the entries in all columns. Repeat this procedure until there are 100 randomized versions of the 2 tables. Apply steps 2 and 3 to each of the randomized pairs of tables. In other words, for each pair of tables, find the MSE scores of all genes versus all drugs. This results in a total of 100,000 MSE scores (1000 scores for a single pair of tables * 100 pairs of tables). Such scores are referred to as MSE^{rand} . MSE scores from non-randomized tables are referred to as $MSE^{nonrand}$

4.2) Compare MSE scores from non-randomized data tables to MSE from randomized data tables.

20 For a given MSE score, M_i , from non-randomized tables, determine the fraction of MSE^{rand} scores which are lower than M_i . This fraction is the significance p-value for score M_i . Using this approach, determine the significance p-values for all $MSE^{nonrand}$ scores.

5) Adjust the significance p-values associated with $MSE^{nonrand}$ scores to correct for multiple tests significance test being employed.

25 The initial significance p-values associated with $MSE^{nonrand}$ scores may not necessarily fairly reflect the true statistical significance because there were multiple significance tests employed. Thus, multiply each significance p-value by 1000 to take into account that 1000 genes were tested against each drug. This kind of adjustment of statistical significance to account for multiple significance tests being employed is known in the statistical literature as the Bonferroni method.

30 6) Report by cell line and drug, the genes and the probabilities derived in step 2.5

6.1) Particularly identify in the report those cell lines and drugs for which there are genes for which the probability derived in step 2.5 is high, say >0.85 , and ranked by smallest-to-largest significance p-score.

5 The examples set out above provide general principles that may be extended to other fields of study, and are not intended to limit the scope of the invention. For example, drug sensitivity levels reflecting the inhibiting of growth could be replaced by drug sensitivity that reflects toxic reactions to drugs. This could be useful in finding markers that indicate circumstances where a given drug not only does not help, but may cause harm (be toxic to non-diseased cells). Diagnostic kits can then be derived to search for
10 those markers in given patients.

Similarly, examples of characterizing attributes could be SNPs or proteins (proteomics).

The Bayesian classifiers are not limited to 1 dimensional or 2 dimensional classifiers, rather any dimension of classifier could be used as appropriate for the chosen characterizing attribute set. This may or may not turn up additional significant
15 likelihoods of co-occurrences depending on the relationships of the attributes in the dataset. It is recognized that a brute force approach of carrying out all steps for all combinations of characterizing attributes and attributes sets of interests can require a great deal of time and computational power, particularly with higher order combinations of attributes. Pre-processing techniques, such as those mentioned previously, can be
20 employed to reduce the number of candidate characterizing attribute sets, and thus the amount of time and computational power required.

Alternate methods could be used to create artificial samples in place of the randomizations suggested herein. The randomizations used herein proved to be a simple and effective manner of creating the artificial samples.

25 In the examples provided above, two likelihood thresholds have been used. First, a likelihood threshold based upon the artificial samples. Second, a likelihood threshold based upon the assigned likelihoods being above a certain percentile of all assigned likelihoods for the relevant attribute of interest.

The likelihood threshold can also be based on a selected threshold based on empirical
30 knowledge, statistically derivation, or otherwise. In order to capture all characterizing sets of interest, even those that could possibly lack statistical validity, the likelihood

threshold could simply be set at zero. Expanding on this, the likelihood threshold could be a selected numerical threshold, or the threshold could be varied, to determine the effect on the results. The likelihood threshold need not be based on artificial or random data in order to derive useful results from the methods.

- 5 As we have seen, the likelihood thresholds could be a single threshold, or a combination of likelihoods thresholds.

The methods described herein can be embodied in a computer program running on an appropriate computing platform as shown in Fig. 9. The combination of the computing platform and computer program results in a system for determining co-occurrences of
10 characterizing attributes and attribute sets of interest. Again, the examples shown in the Figures are not intended to be limiting to the breadth of the invention. As will be evident to those skilled in the art, other configurations of computing platforms and computer programs are possible. For example, the computing platform could take the form of computer network with the computer program distributed about the network, or
15 accessed by terminals remote from that part of the computing platform running the computer program. For example, the computer program may be running on a computer that is connected to and accessible through the Internet.

An example flow diagram for the preferred embodiment of software embodying the first base method described above is shown in Fig. 9. Similarly, an example general block
20 diagram for an embodiment of a system for determining co-occurrences of characterizing attributes and attributes of interest is shown in Fig. 10. In this example, a computer program 1001 is stored on computer storage media 1003 (such as a hard disk from which the computer program is loaded into memory of the computer at the time the program is run) of a standalone computer 1005. The dataset is stored in a database 1007
25 accessible to the computer 1005. The ranked characterizing attribute sets resulting from the base methods may be reported and stored in a file on the hard disk 1003 for later use, including as an output display for viewing on a computer monitor 1009 of the computer 1005. They may take an alternative form of output display as a report 1011 generated on a printer 1013. Similarly, they may be reported to a file, or other output display across a
30 computer network 1015.

Flow diagrams for embodiments of a number of other base methods are shown in Figs. 11, 13 and 15. Corresponding block diagrams are shown in Figs. 12, 14 and 16.

The methods, system and other aspects of the embodiments described herein, and the invention, can be used to identify markers for diagnosis, such as might form part of diagnostic kits or procedures used to determine a disease or syndrome type of a patient. Similarly, they may be used to identify markers for prognosis of a disease or syndrome of a patient, such as might form part of diagnostic kits or procedures used to determine a disease or syndrome type of a patient. Similarly, they may be used to identify markers to determine whether a therapy or treatment is appropriate for a patient, or other biological attribute of a human or other living system. This can be done by identifying and attribute set to be tested for in the patient or other living system by carrying out one or more of the base methods previously described. Although the methods, system and other aspects of the embodiments have been described primarily with respect to the use of gene level expression sets as attribute sets, the embodiments and the invention may also be applied to tissue or serum protein concentration sets, or blood or tissue molecular marker sets, or microscopic or macroscopic clinical observables, or combinations thereof.

It will be understood by those skilled in the art that this description is made with reference to the preferred embodiment and that it is possible to make other embodiments employing the principles of the invention which fall within its spirit and scope as defined by the following claims.

20

We claim:

1. A method of identifying one or more characterizing attributes for an object that are likely to co-occur with one or more attributes of interest for the object, the method
5 comprising the steps of:
 - Selecting one or more attribute sets of one or more characterizing attributes of the object,
 - Selecting an attribute set of one or more attributes of interest for the object,
 - Assigning a likelihood for each characterized attribute set that the attribute set
10 occurs for the object when the attribute set of interest occurs for the object, each likelihood determined using one or more Bayesian computable classifiers on a dataset of attributes for a plurality of actual samples of the object,
 - Comparing each assigned likelihood against one or more likelihood thresholds, and
 - 15 Reporting the assigned likelihoods of the characterizing attribute set based on the likelihood thresholds.
2. The method of claim 1 or 7, wherein a likelihood threshold for each characterizing attribute set is determined using the same Bayesian classifiers as the assigned likelihood on a dataset of attributes for a plurality of artificial samples of the object.
- 20 3. The method of claim 1 or 7, wherein a likelihood threshold for each characterizing attribute set is determined by computing those characterizing attribute sets with an assigned likelihood above a given percentile of all assigned likelihoods for the relevant attribute set.
4. The method of claim 2 or 24, wherein the artificial samples are created by
25 randomizing the actual gene expression levels for the characterizing attributes.
5. The method of claim 2 or 24, wherein the artificial samples are created by transposing the actual gene expression levels for each characterizing attribute to another characterizing attribute.

6. The method of claim 1, wherein the assigned likelihoods of the remaining characterizing attribute sets are also compared against a second likelihood threshold determined by computing those characterizing attribute sets with an assigned likelihood above a given percentile of all assigned likelihoods for the relevant attribute set of interest.
7. A method of identifying a characterizing attribute for an object that is likely to co-occur with an attribute of interest for the object, the method comprising the steps of:
- Selecting one characterizing attribute set of one or more attributes for the object,
- Selecting an attribute of interest for the object,
- Assigning a likelihood for the characterized attribute set that the attribute occurs for the object when the attribute of interest occurs for the object, the assigned likelihood determined using a Bayesian computable classifier on a dataset of attributes for a plurality of actual samples of the object,
- Comparing the assigned likelihood against a likelihood threshold, and
- Reporting the assigned likelihood of the characterizing attribute set based on the likelihood threshold.
8. The method of claim 7 or 24, wherein the characterizing attributes are gene expression levels and the attribute of interest is a drug sensitivity level.
9. The method of claim 1, wherein each characterizing attribute is a gene expression level and the attribute of interest is a drug sensitivity level.
10. The method of claim 1, wherein each characterizing attribute is a gene expression level and the attribute of interest is drug dose (absolute concentration or dose relative to some standard dose) along an increasing, or decreasing, scale.
11. The method of claim 1, wherein each characterizing attribute is a gene expression level and the attribute of interest is the dose of drug which causes half-maximal cellular growth rate.
12. The method of claim 1, wherein each characterizing attribute is a gene expression level and the attribute of interest is $-\log_{10}(\text{dose})$, where dose is the dose

which yields half-maximal total cell mass accumulating under otherwise standard conditions.

13. The method of claim 9, the drug sensitivity level represents growth inhibiting in diseased cells.
- 5 14. The method of claim 9, the drug sensitivity level represents a lack of growth inhibiting in diseased cells.
15. The method of claim 9, the drug sensitivity level represents patient toxicity in healthy cells.
- 10 16. The method of claim 9, wherein the attributes are represented in a dataset taken from the NCI60 dataset.
17. The method of claim 7 or 24, wherein the Bayesian classifier is selected from a group consisting of linear discriminant analysis, quadratic discriminant analysis, and a uniform/gaussian analysis.
- 15 18. The method of claim 1, wherein the Bayesian classifiers are selected from a group consisting of linear discriminant analysis, quadratic discriminant analysis, and a uniform/gaussian analysis.
19. The method of claim 1, wherein two Bayesian classifiers are used selected from a group consisting of linear discriminant analysis, quadratic discriminant analysis, and a uniform/gaussian analysis.
- 20 20. The method of claim 1, wherein one Bayesian classifier is used selected from a group consisting of linear discriminant analysis, quadratic discriminant analysis, and a uniform/gaussian analysis.
21. The method of claim 1, wherein the Bayesian classifiers are linear discriminant analysis, quadratic discriminant analysis, and a uniform/gaussian analysis.
- 25 22. The method of claim 1, wherein the characterizing attribute sets ranked following comparison of the likelihood and the likelihood threshold are reported.

23. The method of claim 22, wherein the ranked characterizing attributes sets are reported to one of a group consisting of a computer readable file stored on computer readable media, a printed report, and a computer network.
24. A method of identifying one or more characterizing attributes for an object that are likely to co-occur with one or more attributes of interest for the object, the method comprising the steps of:
- selecting one or more attribute sets of one or more characterizing attributes of the object,
- selecting an attribute set of one or more attributes of interest for the object,
- assigning a likelihood for each characterized attribute set that the attribute set occurs for the object when the attribute set of interest occurs for the object, each likelihood determined using one or more Bayesian computable classifiers on a dataset of attributes for a plurality of actual samples of the object,
- determining a likelihood significance for each assigned likelihood using artificial samples, and
- ranking the assigned likelihoods of the characterizing attribute set using the likelihood significance.
25. The method of claim 24, wherein the assigned likelihoods are ranked by assigned likelihood and subranked by likelihood significance.
26. The method of claim 24, further comprising the steps of:
- comparing the assigned likelihood against a likelihood threshold, and
- reporting the assigned likelihood of the characterizing attribute set based on the likelihood threshold and the ranking of the assigned likelihood.
27. A method of identifying one or more characterizing attributes for an object that are likely to co-occur with one or more attributes of interest for the object using a dataset of samples of attributes for the object, the method comprising accessing one of the systems of claim 28.

28. A system for identifying one or more characterizing attributes for an object that are likely to co-occur with one or more attributes of interest for the object using a dataset of samples of attributes for the object, the system comprising:
- a computing platform, and
- 5 a computer program on a computer readable medium for use on the computer platform in association with the dataset, the computer program comprising:
- instructions to identify a characterizing attribute for an object that is likely to co-occur with an attribute of interest for the object, by carrying out the steps of the method of claim 1, 7 or 24.
- 10 29. A computer program on a computer readable medium for use on a computer platform in association with a dataset, the computer program comprising:
- instructions to identify a characterizing attribute for an object that is likely to co-occur with an attribute of interest for the object, by carrying out the steps of the method of claim 1, 7 or 24.
- 15 30. A method of drug discovery comprising the steps:
- identifying characterizing attribute sets for interaction by the drug, wherein the step of identifying comprises carrying out the steps of the method of claim 1, 7 or 24 for drug sensitive attributes of interest, and
- performing screens for drugs where growth in cells having desirably ranked
- 20 characterizing attribute sets is drug sensitive.
31. A method of identifying markers for diagnostic kits used to determine if a treatment is appropriate for a patient, the method comprising the steps:
- identifying a gene expression level set to be tested for in the patient by carrying out the steps of the method of claim 1, 7 or 24.
- 25 32. A method of identifying markers for diagnosis isof a living system, the method comprising the steps:
- identifying an attribute set to be tested for in the living system by carrying out the steps of the method of claim 1, 7 or 24.

33. A method of identifying markers for prognosis of a living system, the method comprising the steps:
identifying an attribute set to be tested for in the living system by carrying out the steps of the method of claim 1, 7 or 24.
- 5 34. A method of identifying markers for determining the appropriateness of a therapy or treatment of a living system, the method comprising the steps:
identifying an attribute set to be tested for in the living system by carrying out the steps of the method of claim 1, 7 or 24.
- 10 35. The method of claim 32, wherein the diagnosis is with respect to a disease or syndrome type of a patient.
36. The method of claim 33, wherein the prognosis is with respect to a disease or syndrome type of a patient.
37. The method of claim 32, 33 or 34, wherein the attributes of the attribute set comprise protein concentrations.
- 15 38. The method of claim 37, wherein the protein concentrations comprise tissue protein concentrations.
39. The method of claim 37, wherein the protein concentrations comprise serum protein concentrations.
40. The method of claim 32, 33 or 34, wherein the attributes of the attribute set
20 comprise molecular markers.
41. The method of claim 40, wherein the molecular markers comprise blood molecular markers.
42. The method of claim 40, wherein the molecular markers comprise tissue molecular markers.
- 25 43. The method of claim 32, 33 or 34, wherein the attributes of the attribute set comprise clinical observables.

44. The method of claim 43, wherein the clinical observables comprise microscopic clinical observables.
45. The method of claim 43, wherein the clinical observables comprise macroscopic clinical observables.
- 5 46. The method of claim 32, wherein the markers are for diagnostic kits used in the diagnosis.
47. The method of claim 32, wherein the markers are for diagnostic procedures used in the diagnosis.
48. The method of claim 33, wherein the markers are for prognostic kits used in the
10 prognosis.
49. The method of claim 33, wherein the markers are for prognostic procedures used in the prognosis.

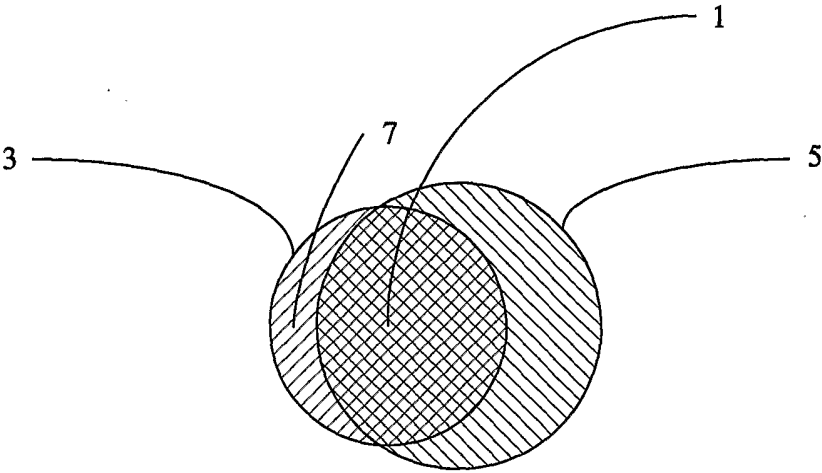


Fig. 1

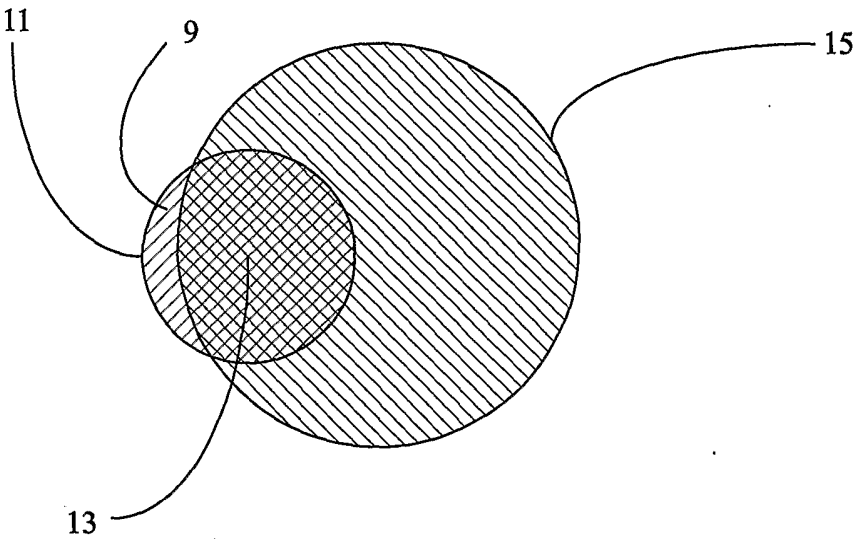


Fig. 2

2/12

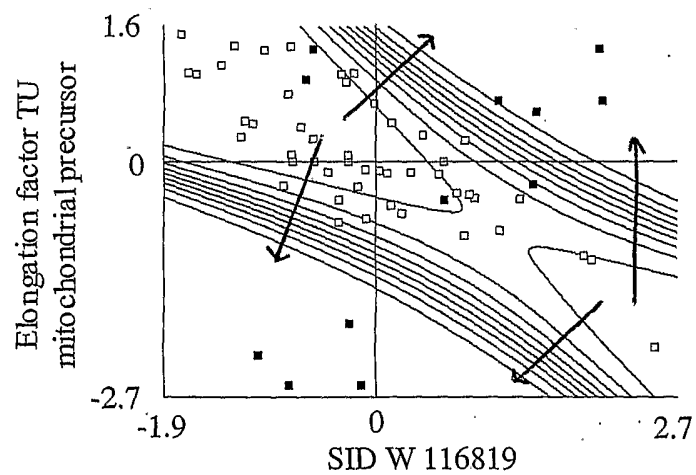


Fig. 3

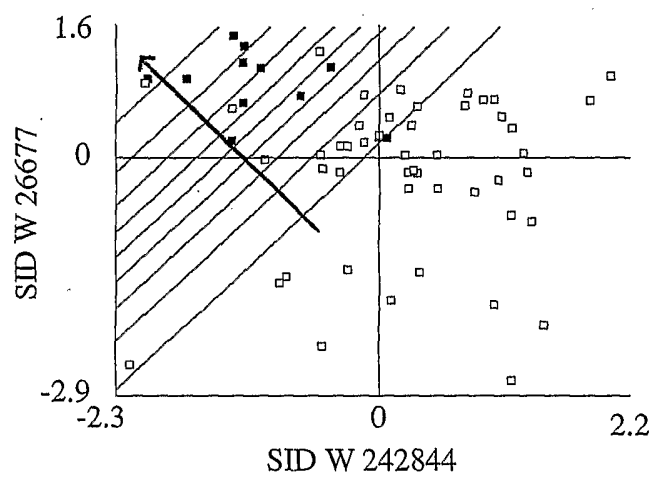


Fig. 4.

3/12

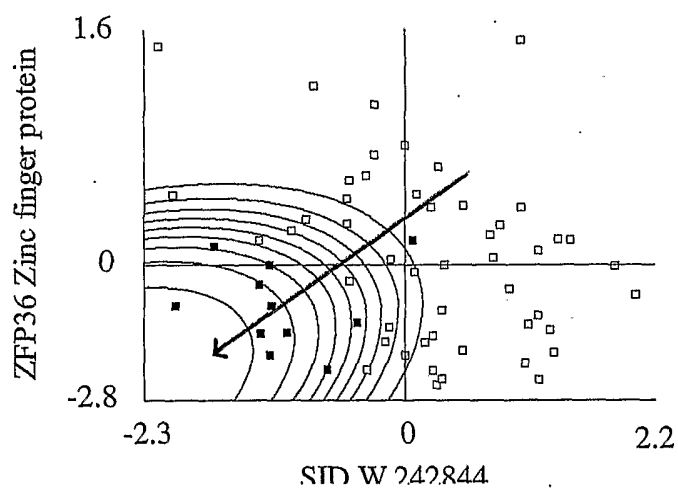


Fig.5

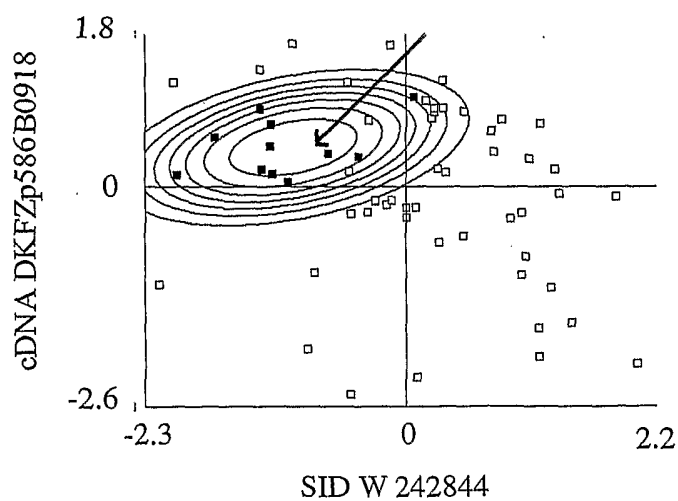


Fig. 6

4/12

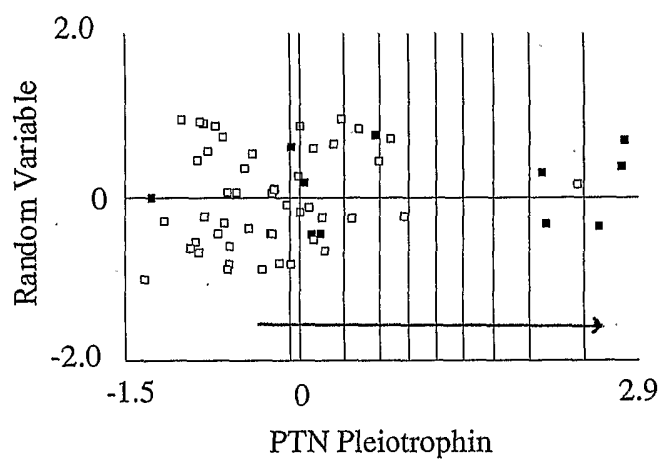


Fig. 7

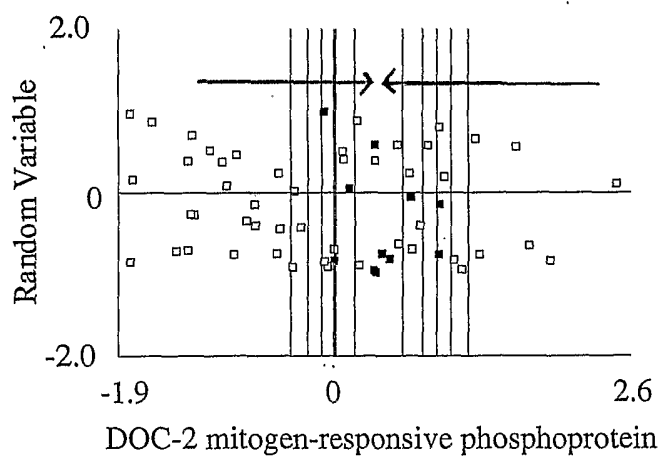


Fig. 8

5/12

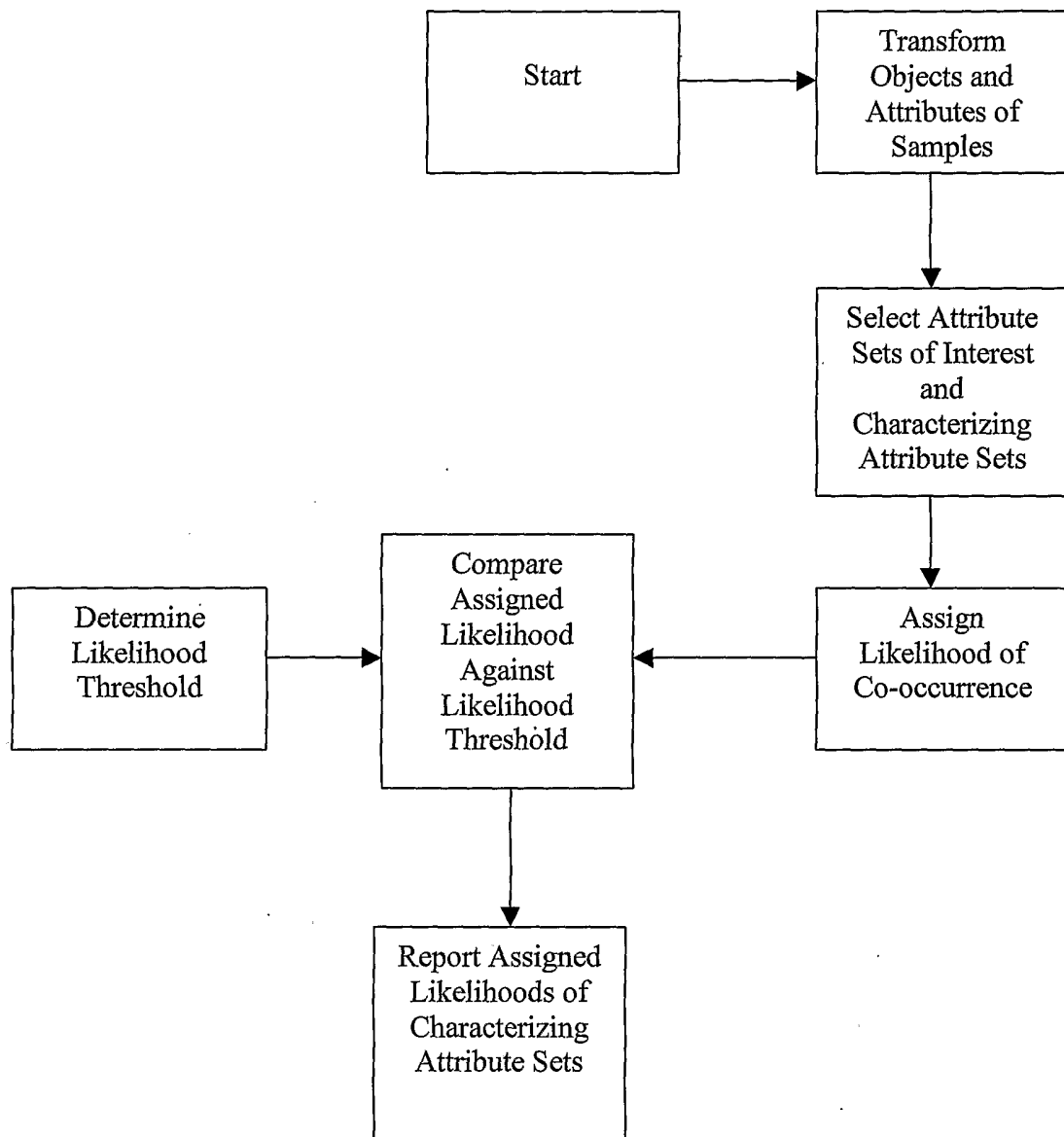


Fig. 9

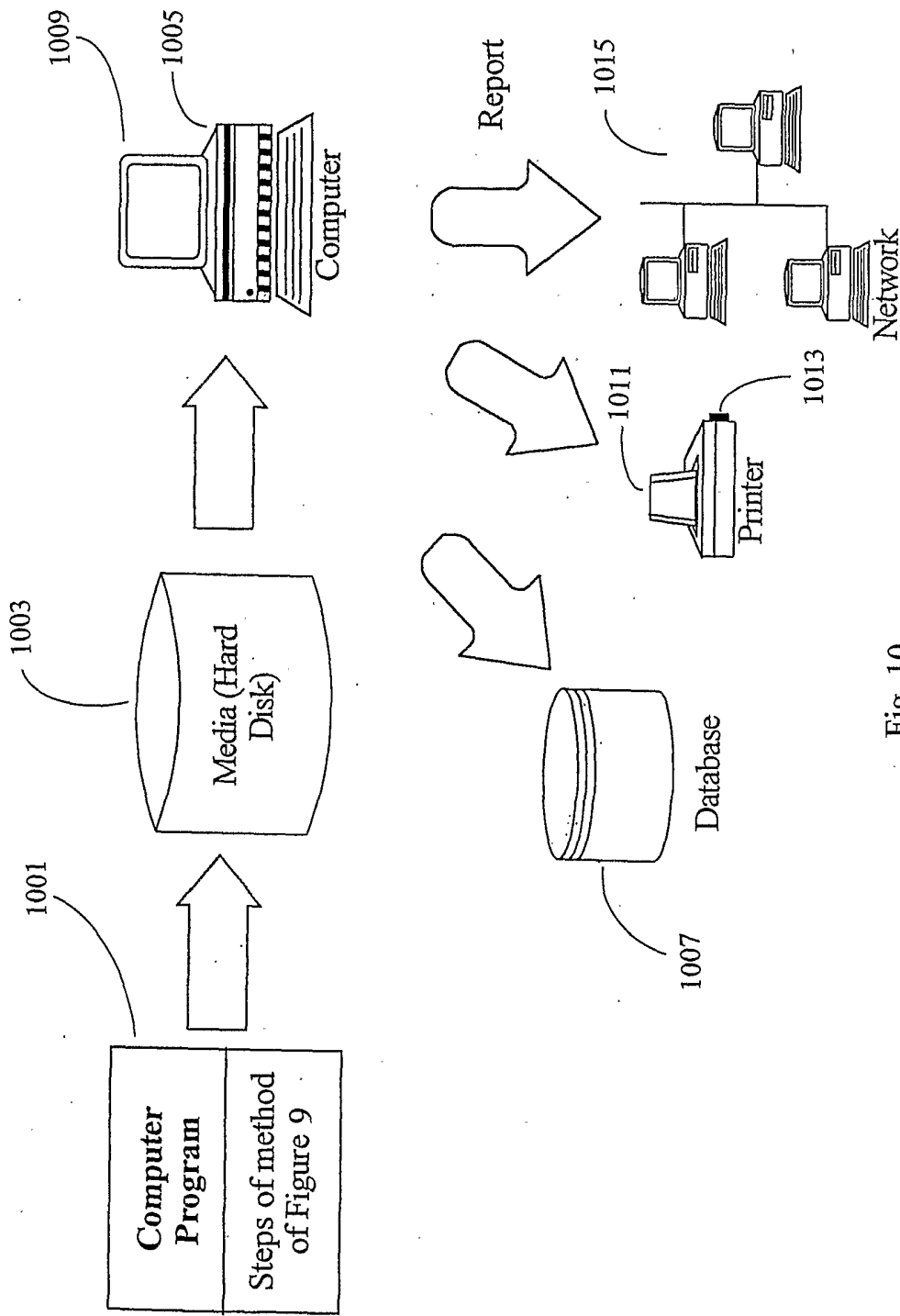


Fig. 10

7/12

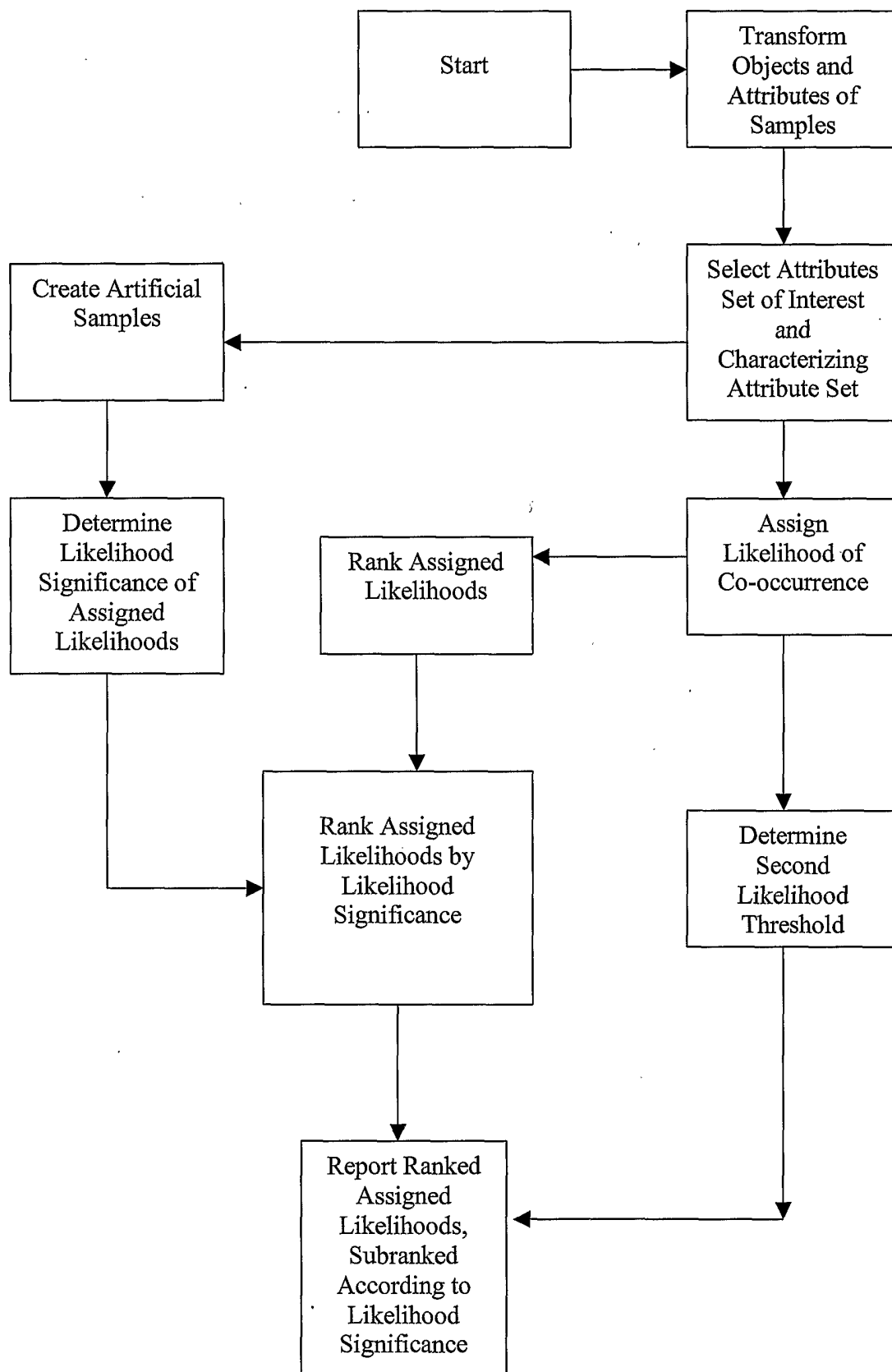


Fig. 11

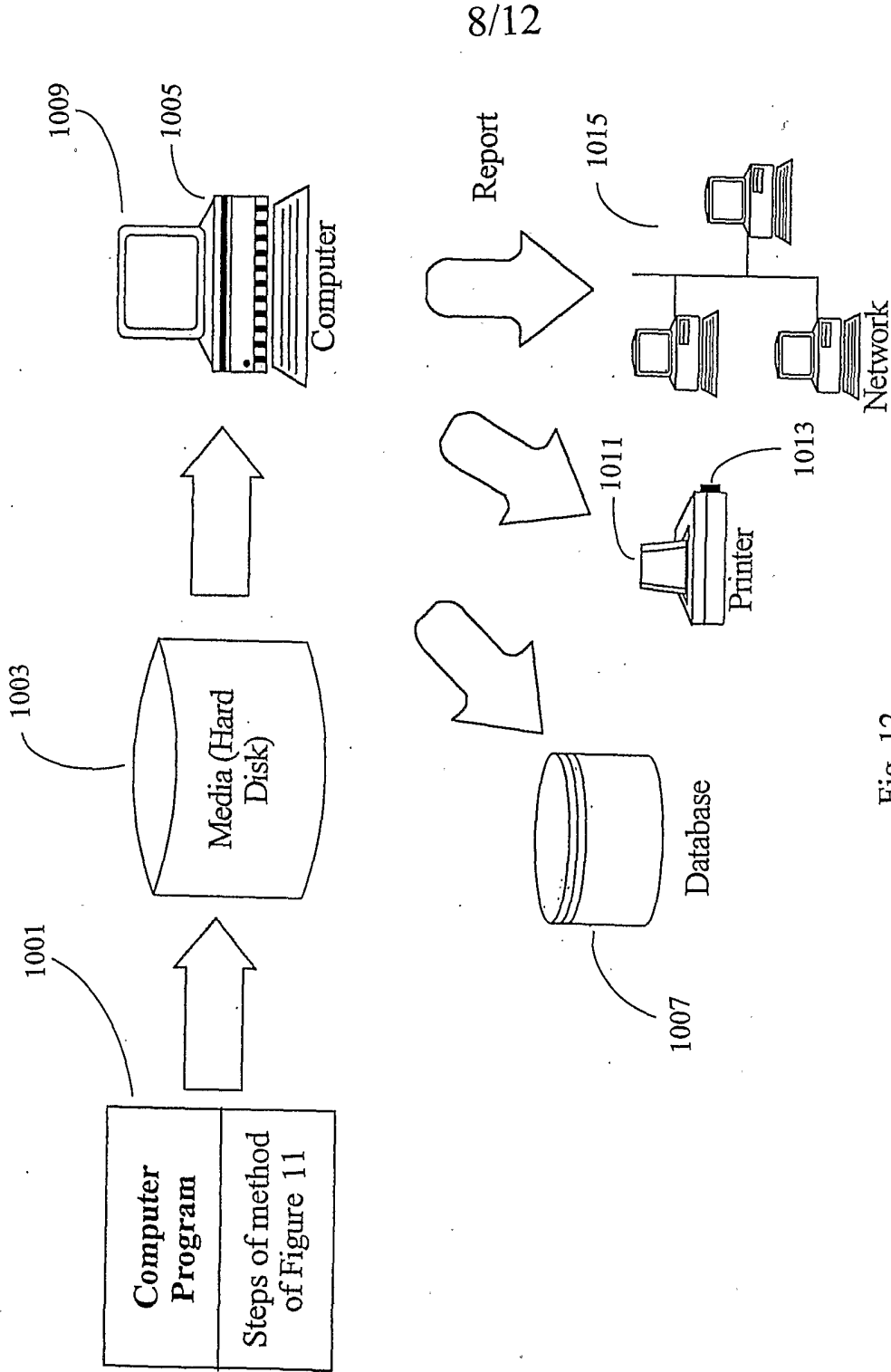


Fig. 12

9/12

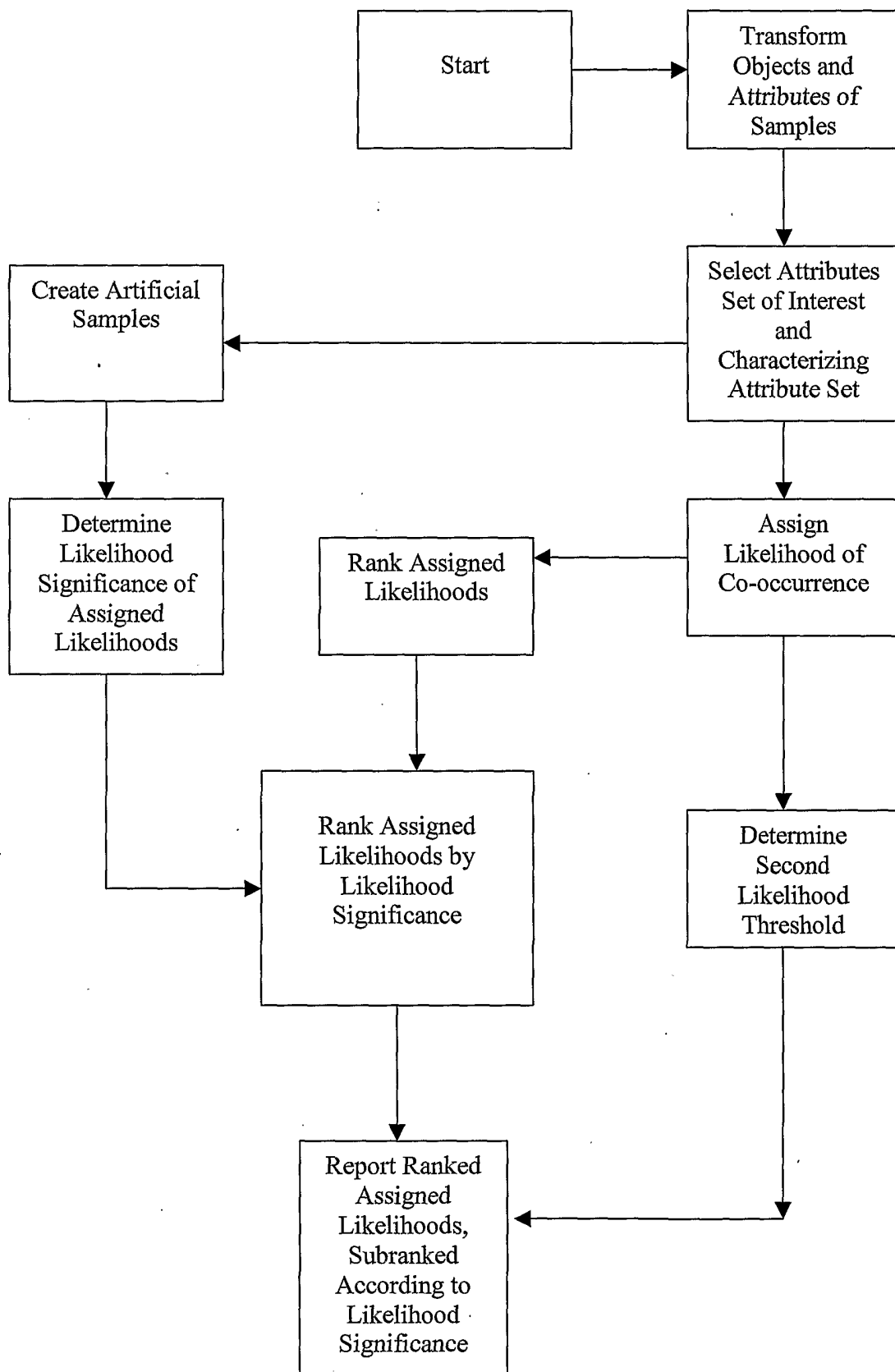


Fig. 13

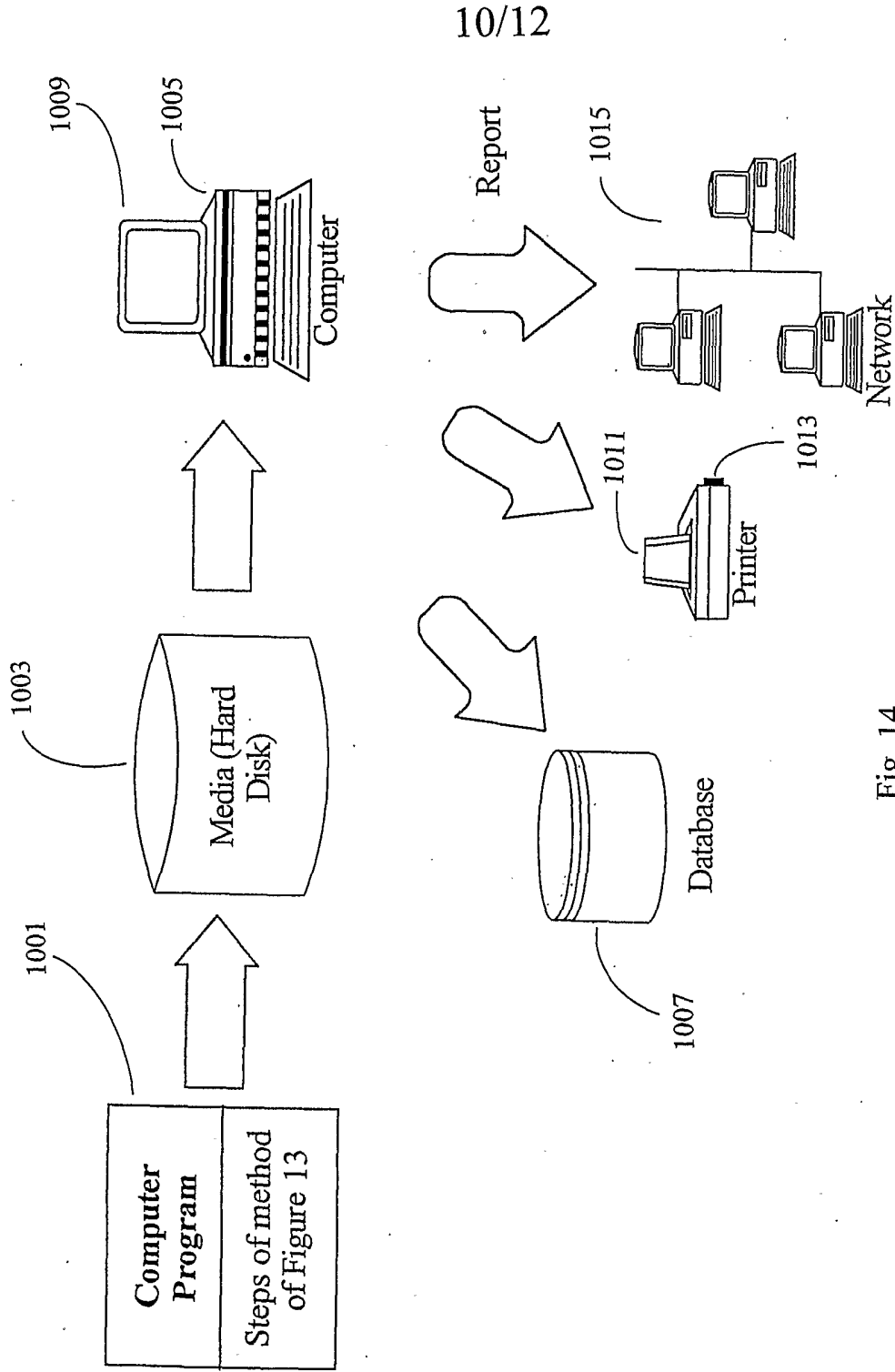


Fig. 14

11/12

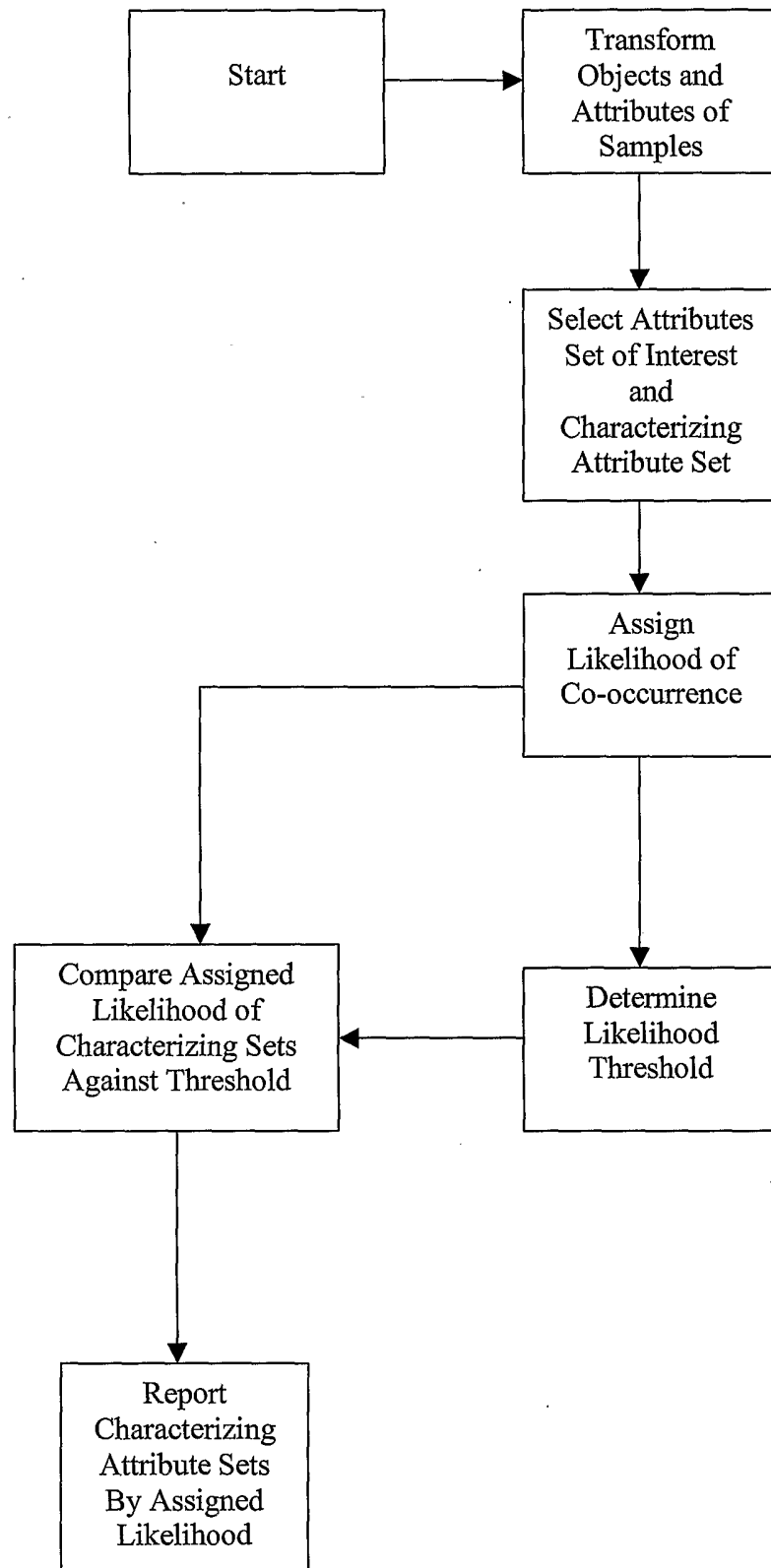


Fig. 15

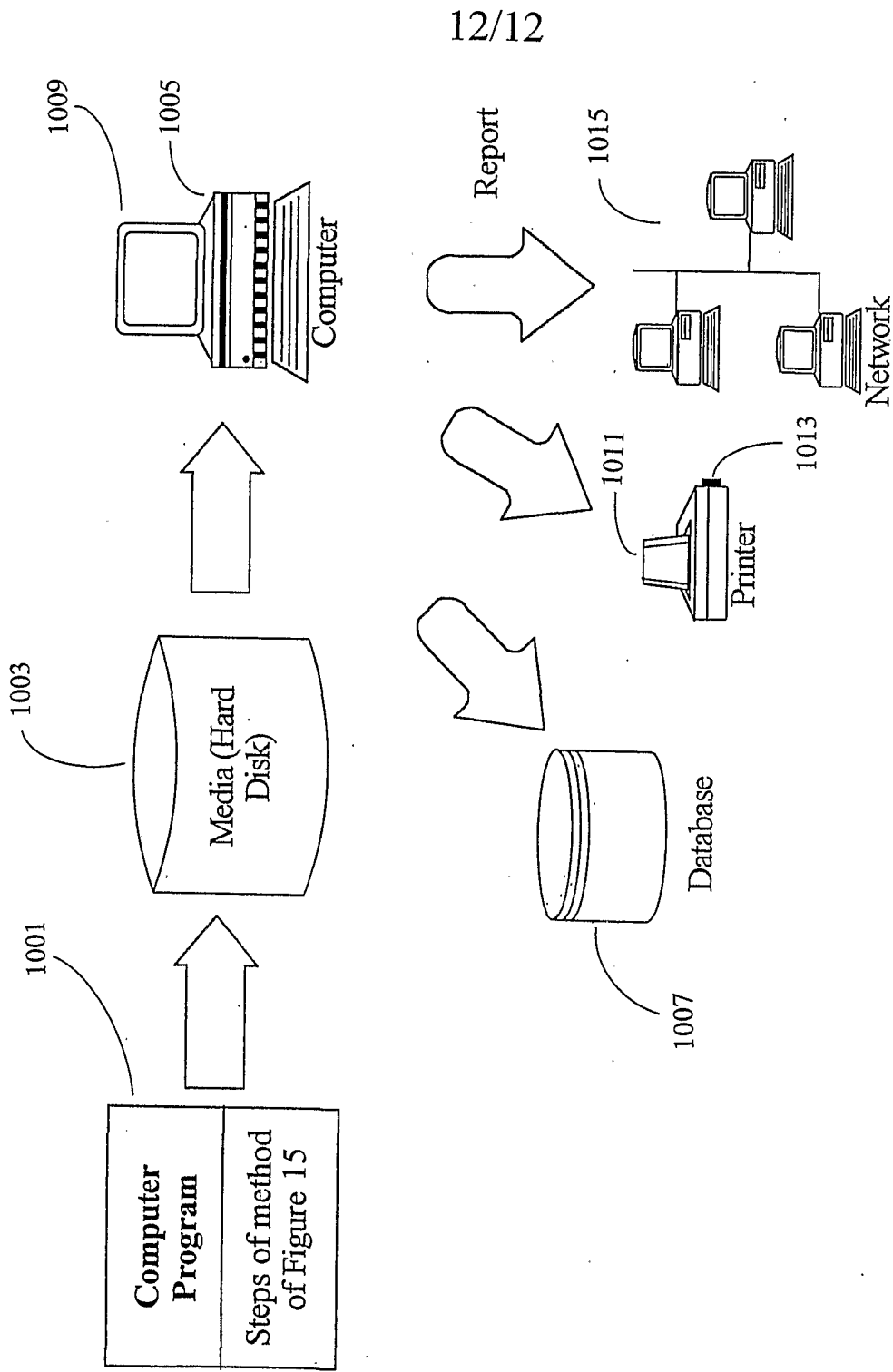


Fig. 16