

# GENECBR

## Diagnostic Mode Manual

*This document will guide you through a step by step tutorial showing the capabilities of GENE CBR to automatically classify new microarray samples in Diagnostic Mode.*

# Contents

Basic principles of Case Based Reasoning applications .....	1
Welcome screen.....	3
Entering in Diagnostic Mode.....	4
Loading a new microarray sample.....	4
GENECBR searches for a solution.....	5
Interpreting the results .....	6
Updating the knowledge base.....	8
Exiting GENECBR .....	8

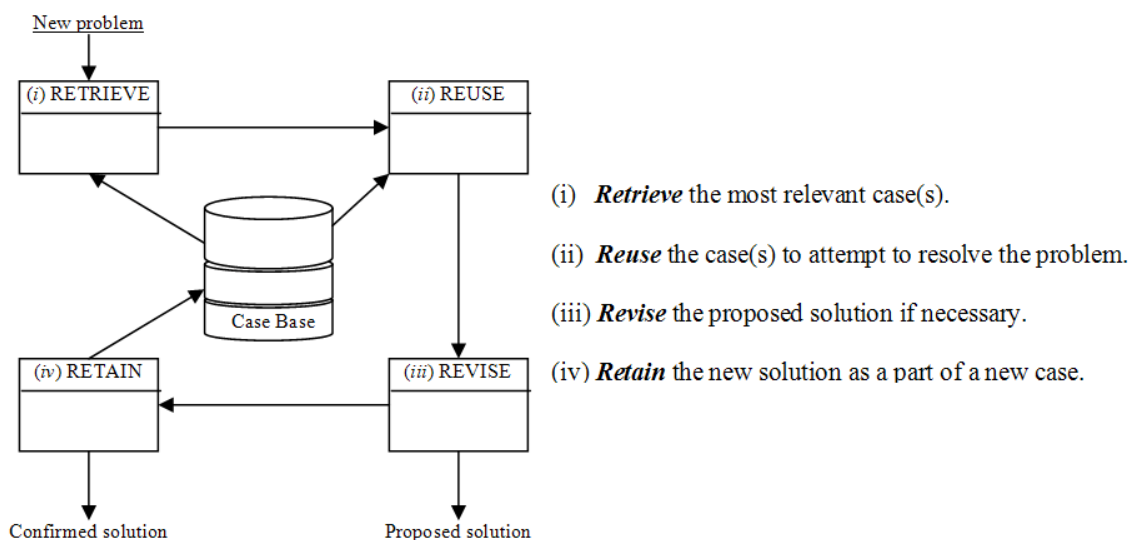
## Basic principles of Case Based Reasoning applications

GENECBR is constructed following the design principles of a Case Based Reasoning (CBR) application.

Case-based reasoning is a computational reasoning paradigm that involves the storage and retrieval of past experiences to solve new problems. An advantage of CBR systems as a problem-solving paradigm is that it is applicable to a wide range of problems, and is particularly relevant in scientific domains, where there is a wealth of data but often a lack of theories or general principles.

A case-based reasoning system solves new problems by adapting solutions that were used to solve previous problems. The case base holds a number of cases, each of which represents a problem together with its corresponding solution. Once a new problem arises, a possible solution to it is obtained by retrieving similar cases from the case base and studying their recorded solutions. A CBR system is dynamic in the sense that, in operation, cases representing new problems together with their solutions are added to the case base, redundant cases are eliminated and others are created by combining existing cases.

A CBR system analyses a new problem situation, and by means of indexing algorithms, retrieves previously stored cases together with their solution by matching them against the new problem situation, then adapts them to provide a solution to the new problem by reusing knowledge stored in the form of cases, in the case base. All of these actions are self-contained and may be represented by a cyclic sequence of processes, in which human interaction may be needed. Case-based reasoning can be used by itself or as part of another intelligent or conventional computing system. Furthermore, case-based reasoning can be a particularly appropriate problem-solving strategy when the knowledge required to formulate a rule-based model of the domain is difficult to obtain, or when the number or complexity of rules relating to the problem domain is too great for conventional knowledge acquisition methods.



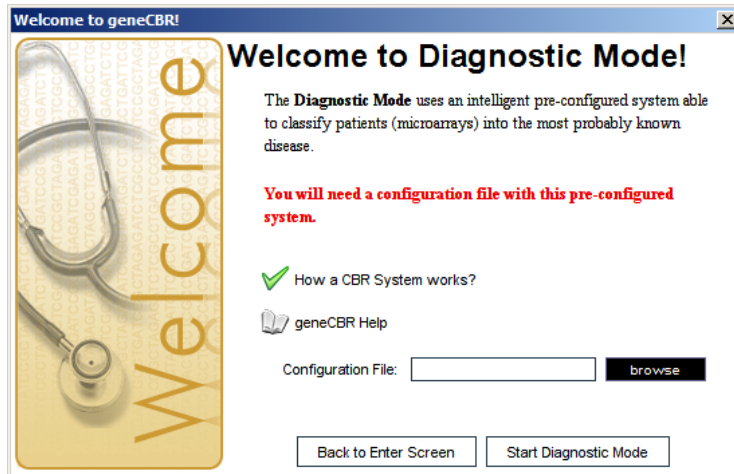
A typical CBR system is composed of four sequential steps which are called into action each time a new problem is to be solved. The following figure outlines the basic CBR cycle.

The purpose of the retrieval step is to search the case base and select one or more previous cases that most closely match the new problem situation, together with their solutions. The selected cases are reused to generate a solution appropriate to the current problem situation. This solution is revised if necessary and finally, the new case (i.e. the problem description together with the obtained solution) is stored in the case base. Cases may be deleted if they are found to produce inaccurate solutions, they may be merged together to create more generalized solutions, and they may be modified, over time, through the experience gained in producing improved solutions. If an attempt to solve a problem fails and it is possible to identify the reason for the failure, then this information should also be stored in order to avoid the same mistake in the future. This corresponds to a common learning strategy employed in human problem-solving. Rather than creating general relationships between problem descriptors and conclusions, as is the case with rule-based reasoning, or relying on general knowledge of the problem domain, CBR systems are able to utilize the specific knowledge of previously experienced, in the form of concrete problem situations. A CBR system provides an incremental learning process because each time a problem is solved, a new experience is retained, thus making it available for future reuse.

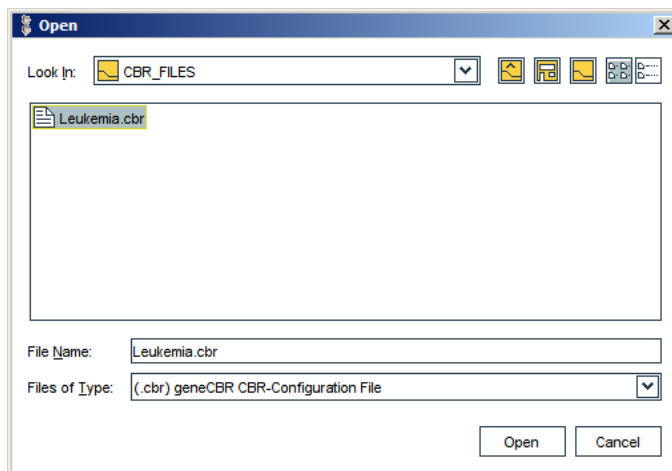
In the CBR cycle there is normally some human interaction. Whilst case retrieval and reuse may be automated, case revision and retention are often undertaken by human experts.

## Welcome screen

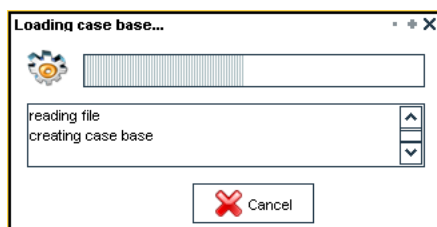
The first step needed to classify a new microarray sample is to load a previously saved CBR configuration file (see the Expert Mode Manual).



In the welcome screen this can be done by specifying a file with .cbr extension (in our example *Leukemia.cbr*) and pressing the `Start Diagnostic Mode` button.



Next, you will see a progress dialog bar meanwhile the case base is loaded.



## Entering in Diagnostic Mode

When GENE CBR is ready to use, a simple interface appears where the case base containing all the available patients and their meta-data information are showed to you.

The screenshot shows the 'geneCBR Diagnostic Mode' window. At the top, there's a header with the 'geneCBR System' logo. Below it, a 'CASE BASE' tab is active, displaying a table with columns for 'FEATURE' and various case IDs (00185, 00355, 07644, 05204, 10222, 12366, 13058, 13223, 14217, 14398, 06667). The table lists various features like 'Category', 'Age', 'Sex', 'FAB/MHOa', 'Karyotype', and 'FISH studiesb' with their corresponding values for each case. Below the table, there are two buttons: 'Classify New Case' and 'Back to Enter Screen'.

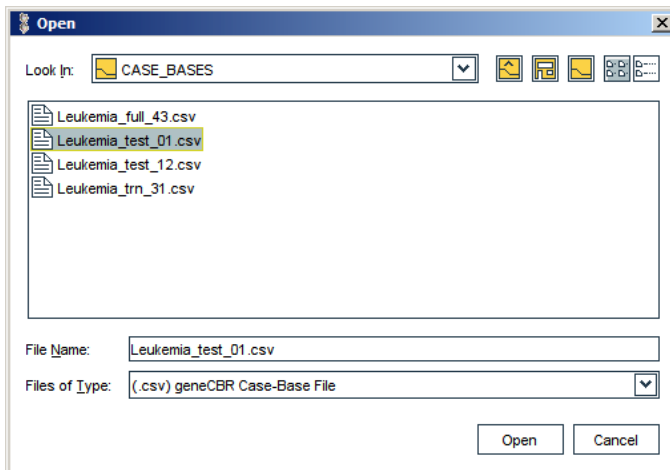
FEATURE	00185	00355	07644	05204	10222	12366	13058	13223	14217	14398	06667
Category	AML_with_in...	AML_with_in...	AML_with_in...	APL	APL	APL	APL	APL	APL	APL	Monocytic_A...
Age	32	47	21	38	40	43	70	14	43	35	20
Sex	M	M	M	F	M	M	M	F	F	F	F
FAB/MHOa	M4Eo	M4Eo	M4Eo	M3	M3	M3	M3	M3	M3	M3	M5
Karyotype	XY	t(15;17)(q12...	"47	"46	XX	t(15;17)(q12...	"46	XY	t(15;17)(q12...	"46	XY
FISH studiesb	CBFB/MYH11	CBFB/MYH11	CBFB/MYH11	PML/RARa	PML/RARa	PML/RARa	PML/RARa	PML/RARa	PML/RARa	PML/RARa	MLL deletion

From this screen you can go back to the enter screen or classify a new microarray sample by pressing the Classify New Case button.

## Loading a new microarray sample

To load the raw data belonging to a new (unclassified) microarray sample, you have to select a text-based, comma-separated file in the file chooser dialog (see the Expert Mode Manual for more information about the specific format).

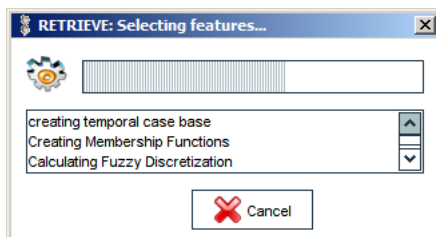
In our example we load the file *Leukemia\_test\_01.csv* containing one microarray sample of type APL (not present in the train case base).



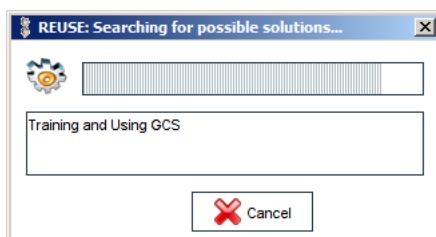
### ***GENECBR searches for a solution***

Once the new microarray sample is validated, GENECBR evolves through the 4-step process in order to find the best classification for this patient.

The first step (RETRIEVE) involves the execution of the DFP algorithm (explained in the Expert Mode Manual) for the selection of the most relevant genes. During this process, you will see a progress dialog bar showing related information about the actions who take place.

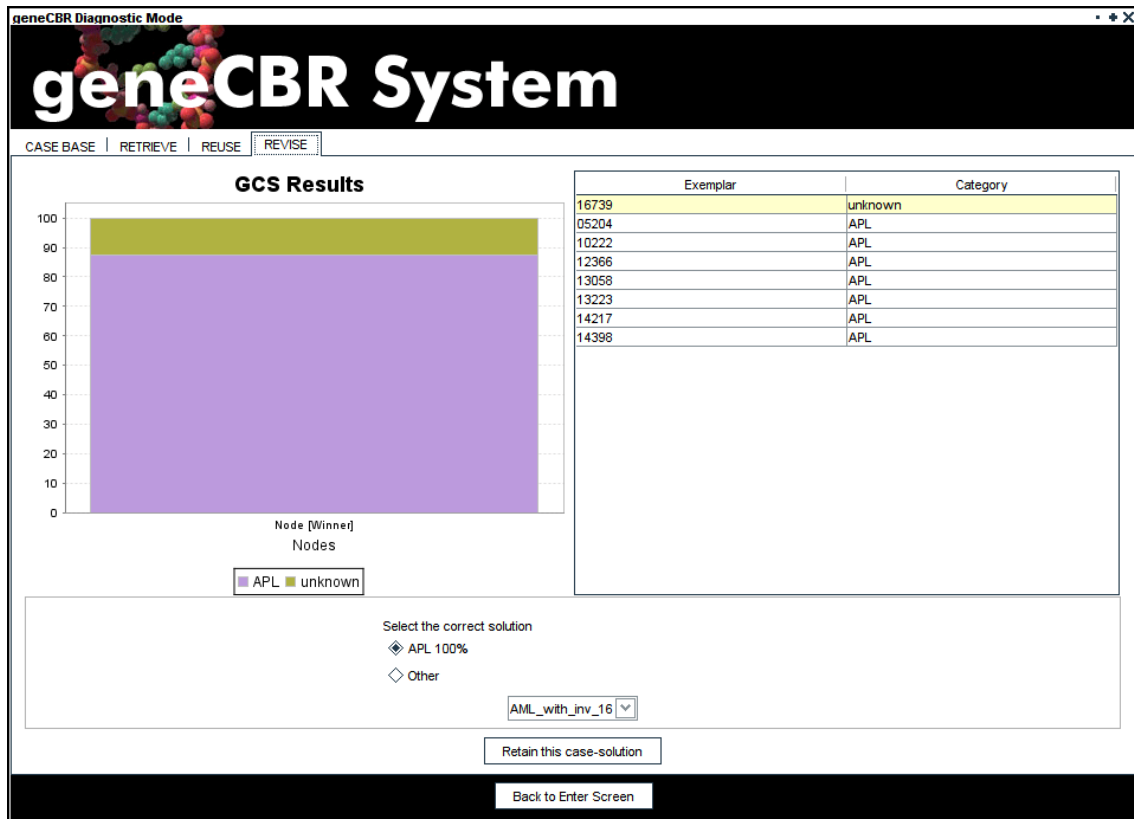


The second step (REUSE) involves the training of the GCS network in order to search for those patients most similar to the new microarray sample. As is the previous case, you will see a progress dialog bar meanwhile this process is executed.



Once the main reasoning cycle is terminated (RETRIEVE & REUSE steps), GENECBR shows you the outcome of the classification process. As a result, three new tabs are available in the application.

From the `REVISE` tab, it can be seen how the new microarray sample is proposed to be classified as `APL`. In this tab, `GENECBR` also shows all the patients belonging to the same node of the trained clustering network.



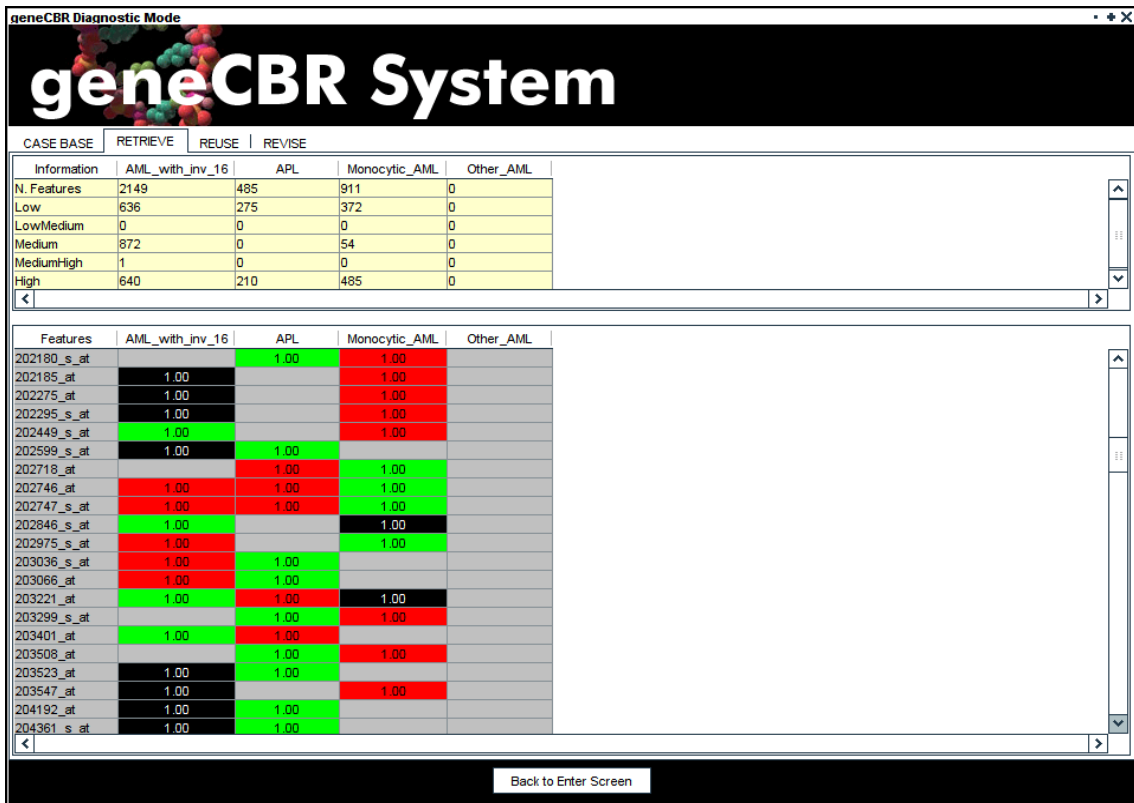
In order to find additional clues about the decision adopted by the `GENECBR` system, you can inspect the `RETRIEVE` and `REUSE` tabs.

## Interpreting the results

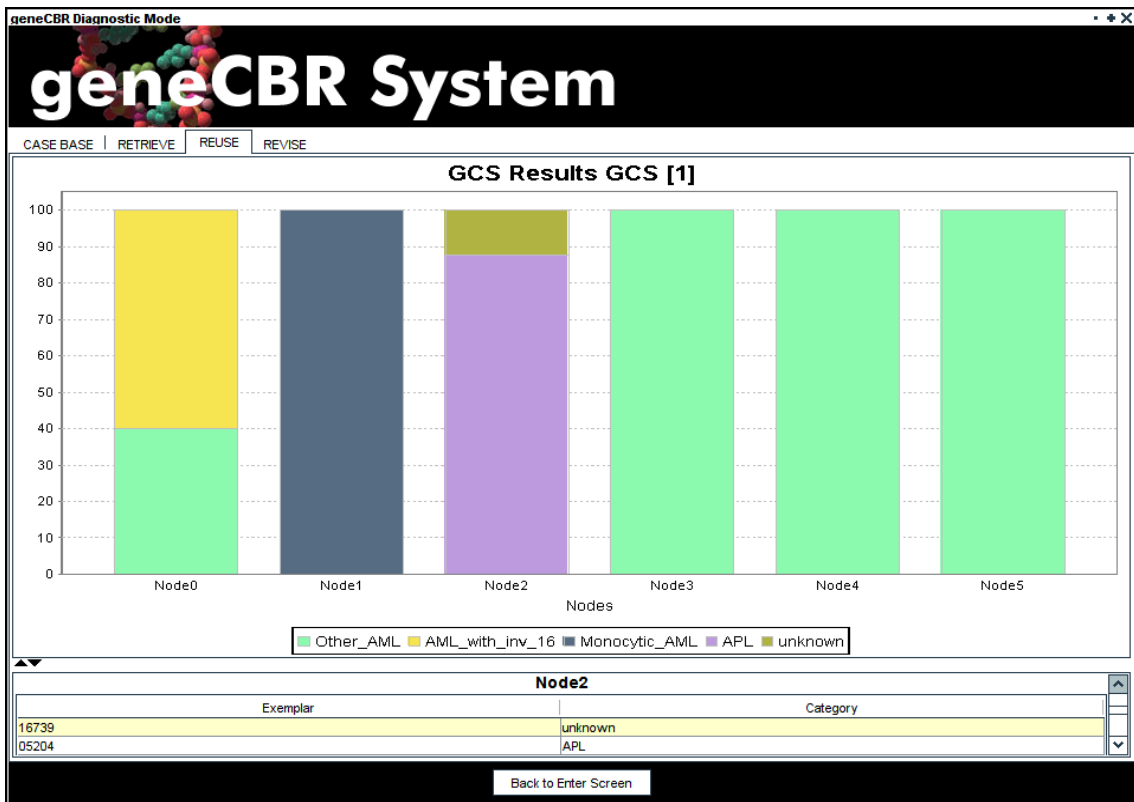
Every time a new classification is proposed by the system, `GENECBR` shows the partial results achieved in all the reasoning cycle.

From the `RETRIEVE` tab, you can examine the genes belonging to the final discriminant fuzzy pattern taken into consideration during the classification process. In the upper part of the screen a brief summary is showed containing information about the different linguistic labels assigned to the different fuzzy patterns. From the lower part of the screen, you can explore the value of the assigned label to any gene.



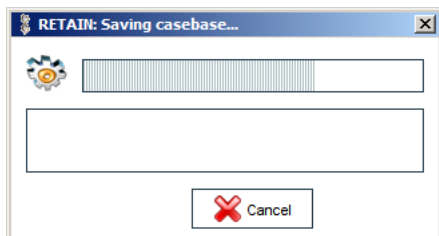


From the REUSE tab in the upper part, you can review all the nodes constructed by the clustering network. In our example, the node assigned to the new microarray sample was *Node2* (the node containing the *unknown* sample). Moreover, in the lower part of the REUSE tab, you can inspect all the patients belonging to each node.



## Updating the knowledge base

Once the user selects a category for the new microarray sample and presses the **Retain** this case-solution button, the system evolves through the **RETAIN** step. During this process, GENECBR stores the new case in its case base for future reuse. A progress dialog bar accompanies this step.



When the last step finishes, GENECBR shows the **RETAIN** tab where the new microarray sample (patient ID 16739) is added to the case base.

FEATURE	05204	10222	12366	13058	13223	14217	14398	16739
Category	APL	APL	APL	APL	APL	APL	APL	APL
Age	38	40	43	70	43	35	19	
Sex	F	M	M	M	F	F	F	F
FABM/WHOa	M3	M3	M3	M3	M3	M3	M3	M3
karyotype	"46	XX	t(15;17)(q12;..."	"46	XY	t(15;17)(q12;..."	"46	"46
FISH studiesb	PML/RARa	PML/RARa	PML/RARa	PML/RARa	PML/RARa	PML/RARa	PML/RARa	RARa insertion

FEATURE	05204	10222	12366	13058	13223	14217	14398	16739
AFFX-BioB-5_at	6.694213	6.550143	6.207033	6.613909	6.251173	8.016043	6.551192	7.003627
AFFX-BioB-M_at	7.805106	7.528421	7.140521	7.588993	7.29062	9.228171	7.60839	8.048525
AFFX-BioB-3_at	6.827084	6.775998	6.476414	6.97675	6.297716	8.579064	6.74284	7.139384
AFFX-BioC-5_at	8.562031	8.2887	8.015686	8.39037	8.041842	8.829667	8.243065	8.515809
AFFX-BioC-3_at	7.987099	7.523173	7.322256	7.786008	7.411576	9.292973	7.757947	7.95853
AFFX-BioDn-5_at	8.143035	8.097173	7.772572	8.190708	7.887953	9.751602	7.948866	8.495984
AFFX-BioDn-3_at	11.436912	11.059243	10.733903	10.987646	10.754737	12.21053	11.161266	11.388455
AFFX-CreX-5_at	11.915992	11.449832	11.279778	11.31811	11.305561	12.621864	11.760107	11.724116
AFFX-CreX-3_at	12.560627	12.102243	11.894494	12.113556	11.94967	13.20562	12.432556	12.440791
AFFX-DapX-5_at	3.568412	3.579753	3.596619	3.500966	3.411943	3.567233	3.916467	3.540173
AFFX-DapX-M_at	3.935836	3.991575	3.80304	3.868602	3.902128	3.928712	4.181112	3.87593
AFFX-DapX-3_at	3.569466	3.488843	3.408144	3.501795	3.484726	3.444944	3.794451	3.521203
AFFX-LysX-5_at	3.578017	3.455984	3.526948	3.563677	3.488319	3.722969	3.495092	3.459478
AFFX-LysX-M_at	4.28123	4.085584	4.14807	4.079404	4.132696	3.901133	4.196332	4.143249
AFFX-LysX-3_at	3.972286	3.590803	3.631591	3.697107	3.683474	4.332884	3.807285	3.556697
AFFX-PheX-5_at	3.89592	3.795384	3.68903	3.739348	3.598739	3.943108	3.792911	3.627094
AFFX-PheX-M_at	3.633266	3.696306	3.544991	3.55262	3.563937	3.715006	3.63086	3.657861
AFFX-PheX-3_at	5.433741	5.638178	5.48935	5.688004	5.41209	5.501225	5.198736	5.229634
AFFX-ThrX-5_at	4.006474	3.947611	3.880761	4.051088	3.96292	3.78123	3.89804	3.947486

From this screen you can go back to the enter screen or repeat the classification process by pressing the **Back to Case Base** button.

## Exiting GENECBR

When you press the **Back to Enter Screen** button a confirmation message is showed in order to process your request.

