Supplemental Information

In this supplementary document, we describe how the reader can run the algorithms in the BRL system, developed in-house by the PRoBE Laboratory, Department of Biomedical Informatics, University of Pittsburgh.

The tool and the manual can be found in our Git repository:
https://github.com/jeya-pitt/BRL

As an example, we show how to run Data ID 3 (Lapointe et al., 2004), extracted from the Gene Expression Omnibus database. We first format the data in BRL format (described in the manual). This formatted data can also be found at out Git repository. In the next section, we show the commands to run the algorithms described in this paper. For detailed instructions, please read the manual supplied with the tool.

Running BRL (Bayesian Rule Learner):

To run the two algorithms presented in this paper, with 10-fold stratified cross-validation, using default parameters, please execute the following commands-

BRL-GSS:

```
java -Xmx1300m -jar BRLv1.0.jar -LP -rgm 1 1 -cv 10 –d 4
0.5 -PPP -DP LapointeEtAl.txt
```

BRL-LSS:

```
java -Xmx1300m -jar BRLv1.0.jar -LP -rgm 2 0 -cv 10 –d 4
0.5 -PPP -DP LapointeEtAl.txt
```

The *–Xmx* flag allocates memory in the system. In case the program runs out of memory, increase this value. The '*-LP*' flag specifies the learner parameters. Here we specify the algorithm type *rgm*, the rule generation method: 1 0 indicates BRL_1, as shown in (Gopalakrishnan et al., 2010); 1 1 indicates BRL_{1000} (Gopalakrishnan et al., 2010) or as specified in this paper, BRL-GSS; 2 0 indicates BRL-LSS from this paper. We specify the number of folds in cross-validation using '*-cv*' (10 folds in this example). He learner flag also indicates the data discretization method used '*-d*' (4 indicates EBD, and 0.5 indicates its parameter λ). The '*-PPP*' flag specifies if we want to pre-process our data before learning. For this paper we do not perform any kind of pre-processing of the data. The '*-DP*' flag specifies the data parameters like input data and output directories.

For detailed information about the options available beyond the paper, please read the manual with the program, from the Git repository.

References:

Lapointe, Jacques, et al. "Gene expression profiling identifies clinically relevant subtypes of prostate cancer." *Proceedings of the National Academy of Sciences of the United States of America* 101.3 (2004): 811-816.

Gopalakrishnan, Vanathi, et al. "Bayesian rule learning for biomedical data mining." *Bioinformatics* 26.5 (2010): 668-675.