*Open Access Library Journal*

# Weighted Voting Analysis of DNA Microarray for Gene Selection and Gene Expression Analysis of Two Types of Rats Treated with Aristolochic Acid and Ochratoxin A Drugs

**Saeed Masoum*, Ebrahim Haghir Ebrahimabadi**

Department of Analytical Chemistry, Faculty of Chemistry, University of Kashan, Kashan, Iran
Email: *masoum@kashanu.ac.ir

Received 9 July 2014; revised 19 August 2014; accepted 22 September 2014



## Abstract

**DNA microarray is an authoritative method for investigation in various cancer and tumors such as renal cancer. Gene expression data include a huge amount of data that the selection of informative data among it is very difficult. Broadly chemometric methods have been used for statistical analysis of gene expression data and different algorithms are used for gene selection. Weighted voting algorithm (WVA) provides a statistical basis for the selection from an original 15,923 probesets, a limited number of most effective genes in discriminating two types of rats treated with Aristolochic acid (AA) and Ochratoxin A (OTA) drugs, that are two chemical compounds with specially toxic effect for kidney and cause renal cancer. In the next step, diminished microarray data are classified by partial least square discriminant analysis (PLSDA) and support vector machine (SVM) methods. Results show that these methods are efficient and sufficient for classification purpose.**

## Keywords

**DNA Microarray, Gene Selection, Renal Cancer, Weighted Voting Algorithm, Partial Least Square Discriminant Analysis**

**Subject Areas:** **Bioengineering, Biotechnology**

## 1. Introduction

Life expectancy of human increases in recent century but yet many advanced diseases peril the people life, such that annually incidence rates of cancer patients increases. For example, in 2003 more than 40% dead in Europe

*Corresponding author.

**How to cite this paper:** Masoum, S. and Ebrahimabadi, E.H. (2014) Weighted Voting Analysis of DNA Microarray for Gene Selection and Gene Expression Analysis of Two Types of Rats Treated with Aristolochic Acid and Ochratoxin A Drugs. *Open Access Library Journal*, **1**: e859. http://dx.doi.org/10.4236/oalib.1100859

in people with the age between 60 and 70 with different cancer metastatic or advanced localize cancers and tumours was reported [1]. Between 1975 and 1995, renal cancer rate was increased by 2.3 and 3.1 percent among white men and women, respectively [2] [3]. Many scientists in different university and research center with investigation in various types of cancer try to decrease mortality rate from cancer [4] [5].

In resent year, DNA microarray technology was developed remarkably, and it is very important technique for gene expression. With the advent of DNA microarrays, it is now possible to simultaneous monitoring the expression of all genes in the genome [6]. The most important applications of this technique are investigation of human diseases, especially various types of cancers. Because of huge volume data obtained from DNA microarray, it's hard to realize complicated correlations among the large numbers of genes present in the genome. Consequently, understanding the synergistic effects of multiple genes is very difficult or impossible. Most of the genes monitored in microarray may not be suitable for classification; therefore these genes may potentially degrade the prediction performance of data analysis by masking the contribution of the relevant genes. To resolve this challenge, new statistical methods must be introduced to analyze these huge amounts of data obtained from microarray experiments [7] [8].

Chemometric methods generally have been used in various fields of chemistry, but today there is an approach for using these methods to investigate this large number of data obtained from DNA microarray [9]-[11]. In fact, chemometrics has been used to extract the important genes among a huge complex data and analysis of gene expression data indicates that gene pattern has changed. Therefore a generic approach to cancer classification based on gene expression monitoring by DNA microarray is possible. There are various supervised and unsupervised chemometric methods for analysis of data gene expression [12] [13]. Principal component analysis (PCA) as an unsupervised method and partial least square discriminante analysis (PLSDA) and support vector machine (SVM) as two supervised methods are very effective for analysis of data gene expression [14]-[16].

Aristolochic acid (AA), a chemical compound found in some plants, such as Asarum and Aristolochia species, is present in a number of natural products that sold as "traditional medicines" relating to dieting supplements or weight-loss drug. The major toxicity targets of AA are kidney's tracts or renal cancer. Furthermore, Ochratoxin A (OTA) is found in some plants such as Penicillium and Aspergillus family. OTA, similarly to AA, is toxic, mainly for kidneys of domestic and laboratory animals [17] [18]. Toxicity effect studying of OTA is important due to its thermo stability and presence in coffee, cereals, cocoa, grapes and etc. [19] [20]. There are some reports that injection of rats with AA for 35 days induces typical renal lesions [21] [22]. Chemical structures of AA and OTA have shown in the supplementary information **Figure S1** and **Figure S2** (Supplementary).

Eker and wild-type rats seemed to be ideal models to study the etiology of renal carcinogenesis [23] [24] and were used to elucidate the mechanism of renal carcinogens, primarily using histopathology and statistical analyses of the number, multiplicity, and progression of renal lesions [25] [26]. In this study we used weighted voting algorithm as a gene selection method to reduce the dimension and then classify diminished data by PLSDA and SVM methods.

## 2. Methodology

DNA microarray is capable of detecting the expression levels of thousand genes over a few samples simultaneously. Therefore statistical analysis of this data is very difficult or impossible. Fortunately, this complexity and difficulty can be avoided by selection and extraction a new data matrix that contains maximal information about the classes from the original data [27]. In fact, the selection of genes that are really indicative of the tissue classification is becoming one of the key steps in microarray studies [8]. Gene selection can reduce complexity and time in the analysis of data expression and also provide a better biological interpretation of relationship between the genes.

One of the applicable methods to gene selection is the weighted voting algorithm that introduced by Ramaswamy that used to calculate the difference between huge variable data [28]. This algorithm calculates $S_x$ value for each genes of data set according to following equation.

$$S_x = (\mu_1 - \mu_2)/(\sigma_1 + \sigma_2)$$

$S_x$ is weighted voting value for every gene, $\mu$ is the mean of expression values in class 1 and class 2 and $\sigma$ is standard deviation of expression values in class 1 and class 2 [29]. The $S_x$ value show how much is correlation of each genes with particular distinction. Therefore, the weighted voting algorithm is very useful method for

biological research. In cancer research, genes in normal tissue work normally, however, in tumors, genes are deregulated and levels of microarray data expression vary widely.

## 3. Dataset

There are lots of data sets for different cancer on the National Center for Biotechnology Information (NCBI) web site: http://www.ncbi.nlm.nih.gov/geo/. We chosen a data matrix contain 84 samples with 15,923 probesets (variables) that each probeset contains one gene and it's also possible that one gene occupies more than one probeset. These samples include domestic and wide type Eker rats that some of them were treated with OTA and AA that dissolved in 0.1 M sodium bicarbonate and a few number of the samples as blank, treated with only 0.1 M sodium bicarbonate every day. Then genome data were obtained subsequently 1, 3, 7, and 14 days after treatment [23].

## 4. Results and Discussion

The confirmation method is based on three criteria: discrimination between domestic and wild Eker rats without any treatment (blank) as "criterion one", the blank domestic Eker and their treated ones with OTA and AA as "criterion two" and the blank wild Eker and their treated ones with OTA and AA as "criterion three". A total of 84 samples were considered in this work (**Figure 1**).

In order to discriminate between blank domestic and wild Eker we analyzed the expression pattern of approximately 15,923 probesets in 36 samples. This data set was divided into two sets of training (24 samples) and monitoring (12 samples). The monitoring set was chosen randomly in such a way that there is adequate representative of the training set. The training set was used to develop the model. Together with the performance of the training set the performance of independent set must also monitored (monitoring set). A training set consisting of samples of known classes (e.g., domestic Eker and wild Eker) is used to select the valuable genes with high impact as biomarker by WVA that allow the most accurate discrimination of the sample in training set. There are many methods for performing the classification task. We used PLSDA and SVM which have been proved to be very useful and robust to classify the microarray gene expression data. Once these methods are trained on the optimal set of variables, it is then applied to an independent monitoring set to validate its prediction accuracy. Sixty probesets with lowest and sixty probe set with highest value of $S_x$ were selected as biomarkers is listed in **Table S1**. Modeling by PLSDA method was done on diminished training set and monitoring set. Among different preprocessing methods, normalization is the best one. In **Figure 2**, the result for two latent variable WVA-PLSDA model is shown. This figure shows that WVA-PLSDA method can separate two groups completely. In this figure samples separated into two regions A and B, which region A that is centralized is wild Eker rat and region B that is scattered is domestic Eker rat.

For each type of other criteria, the same procedures were applied. Sixty Biomarkers for discrimination between blank domestic Eker and treated one and blank wild Eker and treated one are listed in **Table 1** and **Table 2**, respectively.

Results of WVA-PLSDA modelling for criterion two and three are shown in **Figure 3** and **Figure 4**, respectively. **Figure 3** shows three different distinct regions that indicated with A, B and C. Region A related to blank domestic samples, while samples in region B are representative of domestic Eker rats that treated with AA and region C indicates samples treated with OTA. Region B of a score plot shows the effect of the different days after treatment. Samples after first day treating with AA are distinguishable on the second latent variable from the last day treating. It's obvious there is a meaningful trend in region B. Samples move bottom up with an increase

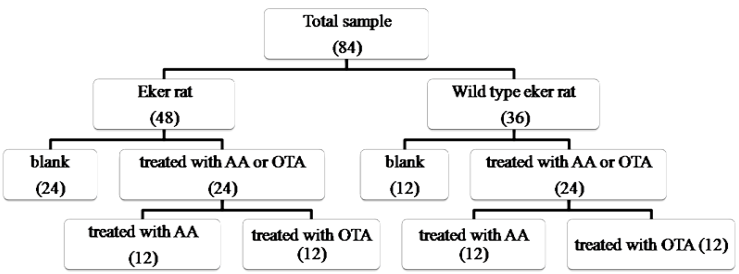


**Figure 1.** Schematic diagram of three different criteria.

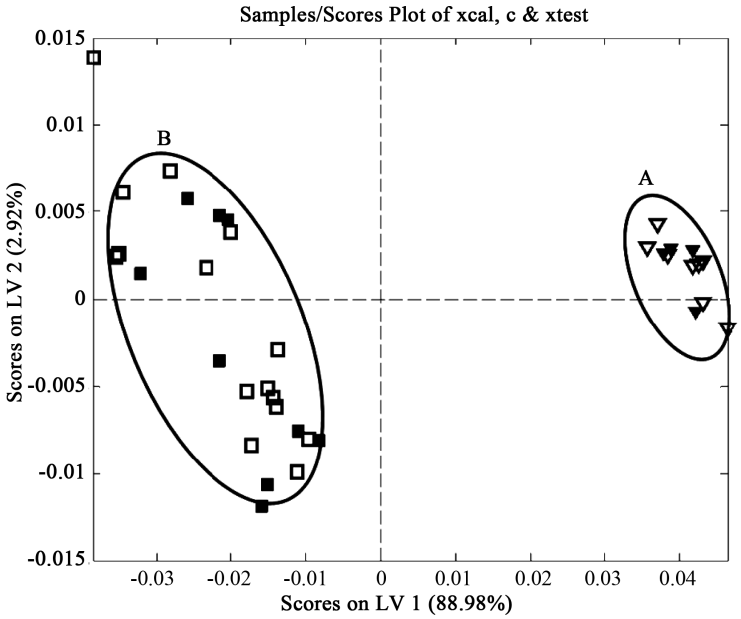


**Figure 2.** Two dimensional score plot of the PLSDA model for criterion one. Training set: Wilde Eker rat (∇), domestic Eker rat (□), corresponding monitoring set samples (▼, ■).

of the day after treating with AA.

Also for SVM analysis we used same criteria that already have used in PLSDA analysis. In SVM only support vectors are needed for classification purpose. This means that for the classification a limited number of data points are used and therefore the calculation process would be reduced. In the present work, among 84, 48 and 36 samples of the training sets for each three criterion, only a total of 6, 8 and 6 samples were chosen as support vectors, respectively.

SVM can separate a given set of binary labelled training data with a hyper-plan that is maximally distant from them. For the case in which no linear separation is possible, they can work in combination with the technique of kernels, which automatically realize a nonlinear mapping to a feature space. Generally, the hyper-plan founded by SVM in a feature space corresponds to a nonlinear decision boundary in the original space. Polynomial kernel SVM results show the 100% accuracy on training and monitoring set for each three criterion.

Results show the obtained genes by WVA are approximately in good agreement with other studies [30]-[32]. *Tsc*1 (Tuberous sclerosis protein 1) is a human protein and gene. This peripheral membrane protein was implicated as a tumour suppressor. Defects in this gene may cause tuberous sclerosis, due to a functional impairment of the hamartin-tuberin complex. In some articles reported that Tsc1 gene mutations are involved in renal cancer carcinogenesis [33]-[35]. *Ghr* (Growth hormone receptor) is gene title that encoded for protein that is a transmembrane receptor for growth hormone and some investigations confirmed relation between defect in this gene and renal cancer [36] [37]. *Keap1* is code name for Kelch-like ECH-associated protein 1. Series of synthetic oleane triterpenoid compounds, known as antioxidant inflammation modulators (AIMs), are being developed by Reata Pharmaceuticals, Inc. and are potent inducers of the Keap1-Nrf2 pathway, blocking Keap1-dependent Nrf2 ubiquitination and leading to the stabilization and nuclear translocation of Nrf2 and subsequent induction of Nrf2 target genes.

Different types of genes with unknown function among the “top 120” deserve high superiority in future studies that provide shortcuts in genome-based renal cancer research.

## 5. Conclusion

In this research, we presented WVA, PLSDA and SVM for feature selection and classification of two type rats treated with AA and OTA drugs, based on microarray gene expression data. The methodology involves dimension reduction of high-dimensional gene expression data, followed by feature selection using WVA and

**Table 1.** Gene selected from weighted voting algorithm of treated Eker rats.

| Row | Gene title | Row | Gene title | Row | Gene title |
|---|---|---|---|---|---|
| 1 | Nrep | 41 | BG372455 | 81 | AI412423 |
| 2 | Dusp11 | 42 | BE109616 | 82 | Igfbp6 |
| 3 | Taf11 | 43 | RGD1307597 | 83 | Msrb2 |
| 4 | Kidins220 | 44 | Daam1 | 84 | Rps6 |
| 5 | Proc | 45 | Scamp2 | 85 | LOC286989 |
| 6 | Psat1 | 46 | Tsc1 | 86 | Arl6ip5 |
| 7 | AA892872 | 47 | BF396739 | 87 | Rps18 |
| 8 | Hexim1 | 48 | Raph1 | 88 | Ube2s |
| 9 | Tmem37 | 49 | RGD1359108 | 89 | Rpl27a |
| 10 | Igsf11 | 50 | Epb4.1l3 | 90 | Skp1 |
| 11 | Col4a4 | 51 | Rab9a | 91 | Acp1 |
| 12 | BI275560 | 52 | Cmtm6 | 92 | Rps10 |
| 13 | Prpf4b | 53 | Sdccag1 | 93 | Mertk |
| 14 | Hsd17b7 | 54 | Dolk | 94 | Gpd1 |
| 15 | Tceal8 | 55 | Clcn6 | 95 | Dap |
| 16 | BE120878 | 56 | Ghr | 96 | H3f3b |
| 17 | Galk2 | 57 | AW524532 | 97 | Lman1 |
| 18 | Tmem79 | 58 | BE102350 | 98 | Slc44a4 |
| 19 | AI010316 | 59 | Hmgcs1 | 99 | Fis1 |
| 20 | Cnot2 | 60 | LOC499749 | 100 | Sptbn1 |
| 21 | RGD1303130 | 61 | Slc12a4 | 101 | Gstm2 |
| 22 | BI279570 | 62 | Myo1b | 102 | Clcc1 |
| 23 | AI407821 | 63 | BF402271 | 103 | BG378056 |
| 24 | Rpl22l1 | 64 | Tspan4 | 104 | Efnb1 |
| 25 | Arvcf | 65 | Tmem80 | 105 | Chp |
| 26 | Dnajc22 | 66 | Mif | 106 | Gstp1 |
| 27 | LOC685841 | 67 | Rab18 | 107 | BF283341 |
| 28 | Synj2bp | 68 | Txnrd1 | 108 | Gpx4 |
| 29 | H31078 | 69 | Tpm3 | 109 | Rps27a |
| 30 | Gprc5c | 70 | BF406641 | 110 | Adra2a |
| 31 | AI600042 | 71 | Sbf1 | 111 | Tpm3 |
| 32 | Myo6 | 72 | LOC363929 | 112 | LOC687266 |
| 33 | Nupl1 | 73 | LOC498555 | 113 | Rpl14 |
| 34 | Mtss1 | 74 | Rnase4 | 114 | Arbp |
| 35 | Snap23 | 75 | Keap1 | 115 | Cyp2e1 |
| 36 | AA892339 | 76 | Smg5 | 116 | Ephx1 |
| 37 | AI410679 | 77 | Akirin2 | 117 | Rabac1 |
| 38 | BE120990 | 78 | Sirt7 | 118 | Cyp4a8 |
| 39 | Fahd1 | 79 | RGD1309079 | 119 | AA686007 |
| 40 | Tmem131 | 80 | Rps21 | 120 | AI710284 |

**Table 2.** Gene selected from weighted voting algorithm of treated wild rats.

| Row | Gene title | Row | Gene title | Row | Gene title |
|---|---|---|---|---|---|
| 1 | RGD1303130 | 41 | RGD1309744 | 81 | RGD1311563 |
| 2 | Tbrg1 | 42 | Ak2 | 82 | AI229311 |
| 3 | Usp8 | 43 | Hdhd3 | 83 | Abcd3 |
| 4 | Tmem37 | 44 | Hdlbp | 84 | Pdap1 |
| 5 | Grm5 | 45 | BI291250 | 85 | Ppap2c |
| 6 | Coq6 | 46 | BM387978 | 86 | Rpl30 |
| 7 | Cdc16 | 47 | Atp6v1b2 | 87 | AI176231 |
| 8 | Tmem79 | 48 | MGC94207 | 88 | AA819086 |
| 9 | Acaca | 49 | Tufm | 89 | BI294806 |
| 10 | Lama3 | 50 | NIPBL | 90 | Nfic |
| 11 | Lamp1 | 51 | AA891362 | 91 | BI291373 |
| 12 | Acot1/Acot2 | 52 | Tpm3 | 92 | Ubl5 |
| 13 | Iah1 | 53 | Dnajc22 | 93 | Msi2 |
| 14 | Gatad1 | 54 | BF550209 | 94 | LOC306766 |
| 15 | Rpp25 | 55 | Chchd8 | 95 | AI406271 |
| 16 | Cct3 | 56 | LOC681989 | 96 | Csad |
| 17 | Mrpl36 | 57 | Cadps | 97 | Rpl19 |
| 18 | Gprc5c | 58 | Hpcal1 | 98 | AI236778 |
| 19 | Polr2j | 59 | Pnpla8 | 99 | Polg2 |
| 20 | Umps | 60 | Il33 | 100 | AI103040 |
| 21 | Nmt1 | 61 | Smc3 | 101 | AA893670 |
| 22 | Dusp11 | 62 | Gstp1 | 102 | Keg1 |
| 23 | Aars | 63 | Dnajc6 | 103 | Nudt3 |
| 24 | Gfra1 | 64 | LOC100174909 | 104 | Calm1 |
| 25 | Eif2b1 | 65 | Pgrmc2 | 105 | Canx |
| 26 | Ndufs7 | 66 | Tbcb | 106 | Srp72 |
| 27 | Paip2 | 67 | AA894070 | 107 | Mkks |
| 28 | Cops2 | 68 | Fcho2 | 108 | Tomm7 |
| 29 | Wdr5 | 69 | BF284876 | 109 | Cml5 |
| 30 | Nagk | 70 | Cugbp1 | 110 | BF403383 |
| 31 | Ap3m1 | 71 | Rpl37 | 111 | Fundc2 |
| 32 | Chrac1 | 72 | Fbxo9 | 112 | Rpl4 |
| 33 | BE116601 | 73 | BF282163 | 113 | AW144821 |
| 34 | Trim37 | 74 | Gdi1 | 114 | Decr1 |
| 35 | LOC686314 | 75 | Spdya | 115 | AI406795 |
| 36 | AA892872 | 76 | Dnajc12 | 116 | Rps8 |
| 37 | Rab3gap2 | 77 | Usp9x | 117 | BF284791 |
| 38 | Hspa9 | 78 | BG374219 | 118 | Hmgb1 |
| 39 | Odf2 | 79 | AA945734 | 119 | Rab10 |
| 40 | Col4a4 | 80 | Rpl31 | 120 | Cyp4a8 |

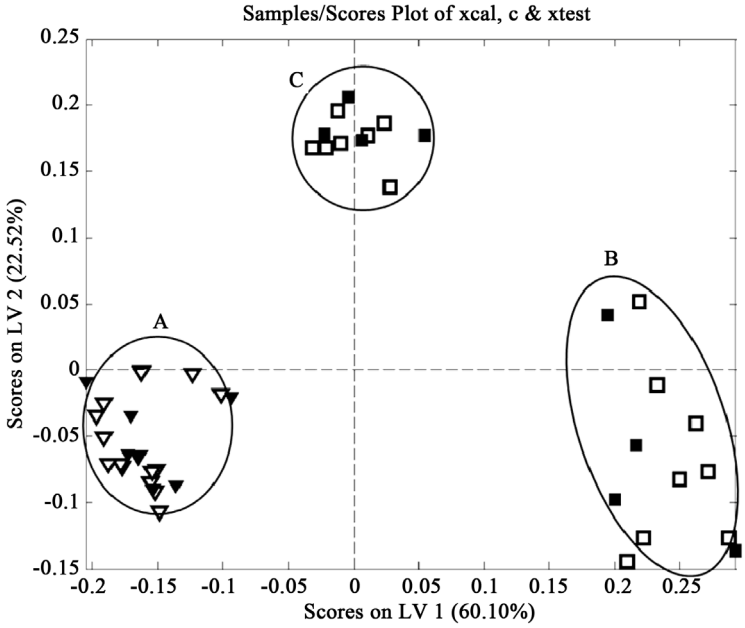


**Figure 3.** Two dimensional score plot of the PLSDA model for domestic Eker rat (criterion two). Training set: blank (∇), treated with AA and OTA (□), corresponding monitoring set samples (▼, ■).

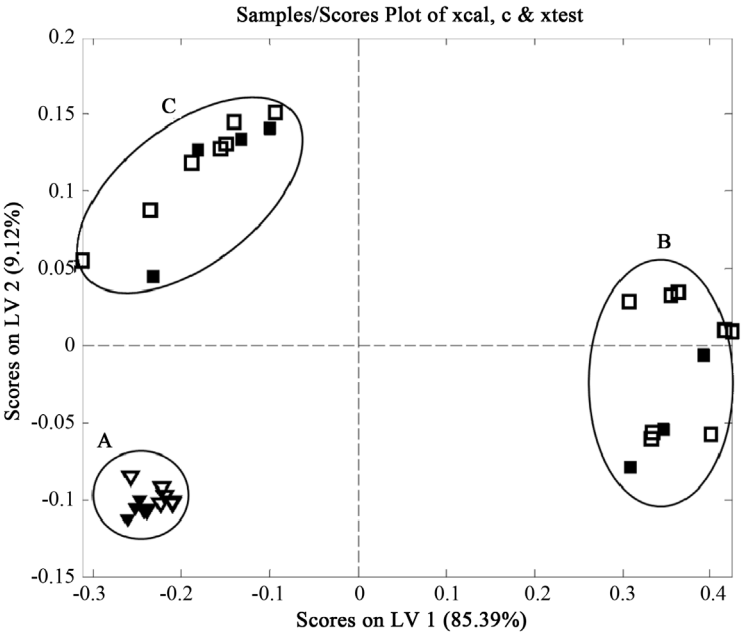


**Figure 4.** Two dimensional score plot of the PLSDA model for wild Eker rat (criterion three). Traning set: blank (∇), treated with AA and OTA (□), corresponding monitoring set samples (▼, ■).

classification by applying PLSDA and SVM. The results show that these methods are effective and efficient in classifying renal cancer.

## Supporting Information

The structure of Arisitolochic Acid (AA) and Ochratoxin A (OTA) available in the supporting information. In

addition, gene selected from weighted voting algorithm of blank Eker rats given in **Table S1**.

## Acknowledgements

The authors are grateful to University of Kashan for supporting this work by grant number 159181/4.

## References

[1] Ramos-Barcelo, E., Rioja, J., Pes, P.L., de la Rosette, J.J. and de Reijke, T.M. (2009) Is Age a Prognostic Factor for Treatment Outcome in Renal Cell Cancer—A Comprehensive Review. *Critical Reviews in Oncology/Hematology*, **72**, 83-89. http://dx.doi.org/10.1016/j.critrevonc.2009.02.005

[2] Jemal, A., Thun, M.J., Ries, L.A.G., Howe, H.L., Weir, H.K., Center, M.M., Ward, E., Wu, X.C., Eheman, C. and Anderson, R. (2008) Annual Report to the Nation on the Status of Cancer, 1975-2005, Featuring Trends in Lung Cancer, Tobacco Use, and Tobacco Control. *Journal of the National Cancer Institute*, **100**, 1672-1694. http://dx.doi.org/10.1093/jnci/djn389

[3] Jemal, A., Clegg, L.X., Ward, E., Ries, L.A.G., Wu, X., Jamison, P.M., Wingo, P.A., Howe, H.L., Anderson, R.N. and Edwards, B.K. (2004) Annual Report to the Nation on the Status of Cancer, 1975-2001, with a Special Feature Regarding Survival. *Cancer*, **101**, 3-27. http://dx.doi.org/10.1002/cncr.20288

[4] Arbyn, M., Raifu, A., Autier, P. and Ferlay, J. (2007) Burden of Cervical Cancer in Europe: Estimates for 2004. *Annals of Oncology*, **18**, 1708-1715. http://dx.doi.org/10.1093/annonc/mdm079

[5] Kouba, E., Smith, A., McRackan, D., Wallen, E.M. and Pruthi, R.S. (2007) Watchful Waiting for Solid Renal Masses: Insight into the Natural History and Results of Delayed Intervention. *The Journal of Urology*, **177**, 466-470. http://dx.doi.org/10.1016/j.juro.2006.09.064

[6] Chuang, L.Y., Chang, H.W., Tu, C.J. and Yang, C.H. (2008) Improved Binary PSO for Feature Selection Using Gene Expression Data. *Computational Biology and Chemistry*, **32**, 29-38. http://dx.doi.org/10.1016/j.compbiolchem.2007.09.005

[7] Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T. and Yu, X. (2000) Distinct Types of Diffuse Large B-Cell Lymphoma Identified by Gene Expression Profiling. *Nature*, **403**, 503-511. http://dx.doi.org/10.1038/35000501

[8] Shen, Q., Shi, W.M., Kong, W. and Ye, B.X. (2007) A Combination of Modified Particle Swarm Optimization Algorithm and Support Vector Machine for Gene Selection and Tumor Classification. *Talanta*, **71**, 1679-1683. http://dx.doi.org/10.1016/j.talanta.2006.07.047

[9] Liang, J. and Kachalo, S. (2002) Computational Analysis of Microarray Gene Expression Profiles: Clustering, Classification, and Beyond. *Chemometrics and Intelligent Laboratory Systems*, **62**, 199-216. http://dx.doi.org/10.1016/S0169-7439(02)00014-X

[10] Karakach, T.K., Flight, R.M., Douglas, S.E. and Wentzell, P.D. (2010) An Introduction to DNA Microarrays for Gene Expression Analysis. *Chemometrics and Intelligent Laboratory Systems*, **104**, 28-52. http://dx.doi.org/10.1016/j.chemolab.2010.04.003

[11] Kerr, G., Ruskin, H.J., Crane, M. and Doolan, P. (2008) Techniques for Clustering Gene Expression Data. *Computers in Biology and Medicine*, **38**, 283-293. http://dx.doi.org/10.1016/j.compbiomed.2007.11.001

[12] Madsen, R., Lundstedt, T. and Trygg, J. (2010) Chemometrics in Metabolomics—A Review in Human Disease Diagnosis. *Analytica Chimica Acta*, **659**, 23-33. http://dx.doi.org/10.1016/j.aca.2009.11.042

[13] Mukhopadhyay, A. and Maulik, U. (2009) Towards Improving Fuzzy Clustering Using Support Vector Machine: Application to Gene Expression Data. *Pattern Recognition*, **42**, 2744-2763. http://dx.doi.org/10.1016/j.patcog.2009.04.018

[14] Biffi, A., Anderson, C.D., Nalls, M.A., Rahman, R., Sonni, A., Cortellini, L., Rost, N.S., Matarin, M., Hernandez, D.G., Plourde, A., de Bakker, P.I.W., Ross, O.A., Greenberg, S.M., Furie, K.L., Meschia, J.F., Singleton, A.B., Saxena, R. and Rosand, J. (2010) Principal-Component Analysis for Assessment of Population Stratification in Mitochondrial Medical Genetics. *The American Journal of Human Genetics*, **86**, 904-917. http://dx.doi.org/10.1016/j.ajhg.2010.05.005

[15] Del Chierico, F., Masotti, A., Onori, M., Fiscarelli, E., Mancinelli, L., Ricciotti, G., Alghisi, F., Dimiziani, L., Manetti, C., Urbani, A., Muraca, M. and Putignani, L. (2012) MALDI-TOF MS Proteomic Phenotyping of Filamentous and Other Fungi from Clinical Origin. *Journal of Proteomics*, **75**, 3314-3330. http://dx.doi.org/10.1016/j.jprot.2012.03.048

[16] Pierce, K.M., Kehimkar, B., Marney, L.C., Hoggard, J.C. and Synovec, R.E. (2012) Review of Chemometric Analysis Techniques for Comprehensive Two Dimensional Separations Data. *Journal of Chromatography A*, **1255**, 3-11. http://dx.doi.org/10.1016/j.chroma.2012.05.050

[17] Stachurska, A., Kozakowska, M., Jozkowicz, A., Dulak, J. and Loboda, A. (2011) Aristolochic Acid I and Ochratoxin

A Differentially Regulate VEGF Expression in Porcine Kidney Epithelial Cells—The Involvement of SP-1 and HIFs Transcription Factors. *Toxicology Letters*, **204**, 118-126. http://dx.doi.org/10.1016/j.toxlet.2011.04.022

[18] Shibutani, S., Dong, H., Suzuki, N., Ueda, S., Miller, F. and Grollman, A.P. (2007) Selective Toxicity of Aristolochic Acids I and II. *Drug Metabolism and Disposition*, **35**, 1217-1222. http://dx.doi.org/10.1124/dmd.107.014688

[19] Amézqueta, S., González-Peñas, E., Murillo-Arbizu, M. and López de Cerain, A. (2009) Ochratoxin A Decontamination: A Review. *Food Control*, **20**, 326-333. http://dx.doi.org/10.1016/j.foodcont.2008.05.017

[20] Duarte, S.C., Pena, A. and Lino, C.M. (2010) A Review on Ochratoxin A Occurrence and Effects of Processing of Cereal and Cereal Derived Food Products. *Food Microbiology*, **27**, 187-198. http://dx.doi.org/10.1016/j.fm.2009.11.016

[21] Debelle, F., Nortier, J., Arlt, V.M., De Prez, E., Vienne, A., Salmon, I., Phillips, D.H., Deschodt-Lanckman, M. and Vanherweghem, J.L. (2003) Effects of Dexfenfluramine on Aristolochic Acid Nephrotoxicity in a Rat Model for Chinese-Herb Nephropathy. *Archives of Toxicology*, **77**, 218-226.

[22] Mei, N., Arlt, V.M., Phillips, D.H., Heflich, R.H. and Chen, T. (2006) DNA Adduct Formation and Mutation Induction by Aristolochic Acid in Rat Kidney and Liver. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **602**, 83-91. http://dx.doi.org/10.1016/j.mrfmmm.2006.08.004

[23] Stemmer, K., Ellinger-Ziegelbauer, H., Ahr, H.J. and Dietrich, D.R. (2007) Carcinogen-Specific Gene Expression Profiles in Short-Term Treated Eker and Wild-Type Rats Indicative of Pathways Involved in Renal Tumorigenesis. *Cancer Research*, **67**, 4052-4068. http://dx.doi.org/10.1158/0008-5472.CAN-06-3587

[24] Hino, O., Kobayashi, E., Hirayama, Y., Kobayashi, T., Kubo, Y., Tsuchiya, H., Kikuchi, Y. and Mitani, H. (1995) Molecular Genetic Basis of Renal Carcinogenesis in the Eker Rat Model of Tuberous Sclerosis ($Tsc_2$). *Molecular Carcinogenesis*, **14**, 23-27. http://dx.doi.org/10.1002/mc.2940140106

[25] Walker, C., Goldsworthy, T.L., Wolf, D.C. and Everitt, J. (1992) Predisposition to Renal Cell Carcinoma Due to Alteration of a Cancer Susceptibility Gene. *Science*, **255**, 1693-1695. http://dx.doi.org/10.1126/science.1553556

[26] Wolf, D., Goldsworthy, T., Janszen, D., Harden, R., Donner, E., David, C. and Everitt, J. (2000) Promotion by Sodium Barbital Induces Early Development but Does not Increase the Multiplicity of Hereditary Renal Tumors in Eker Rats. *Carcinogenesis*, **21**, 1553-1558. http://dx.doi.org/10.1093/carcin/21.8.1553

[27] Zheng, C.H., Huang, D.S., Kong, X.Z. and Zhao, X.M. (2008) Gene Expression Data Classification Using Consensus Independent Component Analysis. *Genomics, Proteomics & Bioinformatics*, **6**, 74-82. http://dx.doi.org/10.1016/S1672-0229(08)60022-4

[28] Ramaswamy, S., Ross, K.N., Lander, E.S. and Golub, T.R. (2003) A Molecular Signature of Metastasis in Primary Solid Tumors. *Nature Genetics*, **33**, 49-54. http://dx.doi.org/10.1038/ng1060

[29] Jen, C.H., Yang, T.P., Tung, C.Y., Su, S.H., Lin, C.H., Hsu, M.T. and Wang, H.W. (2008) Signature Evaluation Tool (SET): A Java-Based Tool to Evaluate and Visualize the Sample Discrimination Abilities of Gene Expression Signatures. *BMC Bioinformatics*, **9**, 58. http://dx.doi.org/10.1186/1471-2105-9-58

[30] Wang, Y., Meng, F., Arlt, V.M., Mei, N., Chen, T. and Parsons, B.L. (2011) Aristolochic Acid-Induced Carcinogenesis Examined by ACB-PCR Quantification of H-Ras and K-Ras Mutant Fraction. *Mutagenesis*, **26**, 619-628. http://dx.doi.org/10.1093/mutage/ger023

[31] Cavin, C., Delatour, T., Marin-Kuan, M., Fenaille, F., Holzhäuser, D., Guignard, G., Bezençon, C., Piguet, D., Parisod, V., Richoz-Payot, J. and Schilter, B. (2009) Ochratoxin A—Mediated DNA and Protein Damage: Roles of Nitrosative and Oxidative Stresses. *Toxicological Sciences*, **110**, 84-94. http://dx.doi.org/10.1093/toxsci/kfp090

[32] Adler, M., Müller, K., Rached, E., Dekant, W. and Mally, A. (2009) Modulation of Key Regulators of Mitosis Linked to Chromosomal Instability Is an Early Event in Ochratoxin A Carcinogenicity. *Carcinogenesis*, **30**, 711-719. http://dx.doi.org/10.1093/carcin/bgp049

[33] El-Hashemite, N., Zhang, H., Henske, E.P. and Kwiatkowski, D.J. (2003) Mutation in $TSC_2$ and Activation of Mammalian Target of Rapamycin Signalling Pathway in Renal Angiomyolipoma. *The Lancet*, **361**, 1348-1349. http://dx.doi.org/10.1016/S0140-6736(03)13044-9

[34] Green, A.J., Johnson, P.H. and Yates, J.R.W. (1994) The Tuberous Sclerosis Gene on Chromosome 9q34 Acts as a Growth Suppressor. *Human Molecular Genetics*, **3**, 1833-1834. http://dx.doi.org/10.1093/hmg/3.10.1833

[35] Slegtenhorst, M., Hoogt, R., Hermans, C., Nellist, M., Janssen, B., Verhoef, S., Lindhout, D., Ouweland, A., Halley, D. and Young, J. (1997) Identification of the Tuberous Sclerosis Gene TSC1 on Chromosome 9q34. *Science*, **277**, 805-808. http://dx.doi.org/10.1126/science.277.5327.805

[36] Chin, E., Zhou, J. and Bondy, C. (1992) Renal Growth Hormone Receptor Gene Expression: Relationship to Renal Insulin-Like Growth Factor System. *Endocrinology*, **131**, 3061-3066.

[37] Flyvbjerg, A., Bennett, W.F., Rasch, R., Kopchick, J.J. and Scarlett, J.A. (1999) Inhibitory Effect of a Growth Hormone Receptor Antagonist (G120K-PEG) on Renal Enlargement, Glomerular Hypertrophy, and Urinary Albumin Excretion in Experimental Diabetes in Mice. *Diabetes*, **48**, 377-382. http://dx.doi.org/10.2337/diabetes.48.2.377

## Supplementary Information

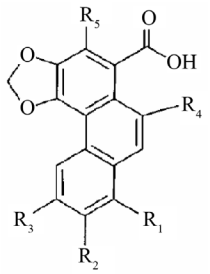


**Figure S1.** Chemical structure of Arisitolochic Acid (AA): *Aristolochcic Acid I*: R1 = OCH3, R2 = H, R3 = H, R4 = NOH2, R5 = H. *Aristolochcic Acid II*: R1 = H, R2 = H, R3 = H, R4 = NOH2, R5 = H.

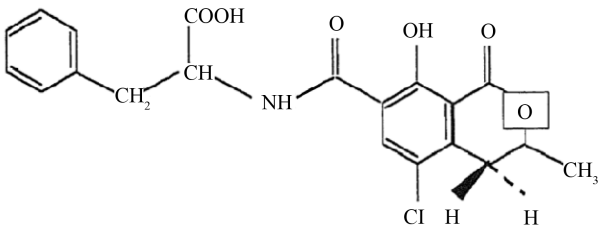


**Figure S2.** Chemical structure of Ochratoxin A (OTA).

**Table S1.** Gene selected from weighted voting algorithm of blank Eker rats.

| Row | Accession No. | Gene title | Row | Accession no. | Gene title | Row | Accession No. | Gene title |
|---|---|---|---|---|---|---|---|---|
| 1 | AI176608 | Sdhd | 41 | NM_031018 | Atf2 | 81 | AA892159 | Ccnl2 |
| 2 | NM_133321 | Kcnj15 | 42 | NM_019182 | Rnf4 | 82 | BI279756 | Nob1 |
| 3 | NM_017348 | Slc6a8 | 43 | AI102612 | Dazap2 | 83 | BF418786 | Lonp2 |
| 4 | BI295970 | Tpm3 | 44 | AA996576 | Rab5b | 84 | AI101659 | |
| 5 | NM_017135 | Ak3l1 | 45 | BE128627 | Lpcat3 | 85 | BE098802 | Dmtf1 |
| 6 | BI295970 | Tpm3 | 46 | AI170385 | Smarca2 | 86 | AI170772 | Atp5g2 |
| 7 | BI282863 | Phb | 47 | AW530769 | Fdft1 | 87 | BE107358 | Pum2 |
| 8 | AI009817 | Sdhc | 48 | AI454932 | Klf13 | 88 | AW434268 | Pfdn6 |
| 9 | BM389287 | Ube2g1 | 49 | AI547471 | Nsf | 89 | BM389891 | Med28 |
| 10 | BM986220 | App | 50 | J03933 | Thrb | 90 | AI070897 | Fam18b2 |
| 11 | NM_053357 | Ctnnb1 | 51 | 1369508_at | Golph3 | 91 | AI172078 | Cmas |
| 12 | NM_138840 | Tgoln1 | 52 | BF284175 | Pla2g12a | 92 | AA996836 | |
| 13 | BF550209 | | 53 | AW525776 | Laptm4a | 93 | BI299621 | |
| 14 | NM_020085 | Ptprk | 54 | AI407788 | Ube2l3 | 94 | AI171781 | Luc7l2 |
| 15 | L09653 | Tgfbr2 | 55 | NM_053862 | Lgals8 | 95 | AI170507 | |
| 16 | D10770 | Prkacb | 56 | NM_030989 | Tp53 | 96 | BM392148 | Bcl2l2/Pabpn1 |
| 17 | AF304333 | Xiap | 57 | BF561717 | Pdha1 | 97 | BF412072 | Efnb1 |
| 18 | NM_023986 | Psme4 | 58 | AI103695 | Serp1 | 98 | NM_022498 | Ppp1cc |
| 19 | AA957367 | RGD1562236 | 59 | NM_013217 | Mllt4 | 99 | BM384301 | |
| 20 | BF407856 | Akirin2 | 60 | AF054618 | Cttn | 100 | BM384889 | |

**Continued**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 21 | BE113270 | Igfbp5 | 61 | AI599410 | Letmd1 | 101 | BE101876 | |
| 22 | NM_022180 | Hnf4a | 62 | AI176579 | | 102 | NM_053687 | Slfn3 |
| 23 | NM_021856 | Clock | 63 | BE329241 | Galk2 | 103 | BE104098 | Meis2 |
| 24 | AW254369 | Plekhb2 | 64 | BF392130 | | 104 | BF408990 | Srrm2 |
| 25 | U31884 | Ddc | 65 | BI290534 | LOC290577 | 105 | BE114972 | Rbm39 |
| 26 | AI178292 | Wdr26 | 66 | NM_053982 | Rps15a | 106 | NM_031106 | Rpl37 |
| 27 | NM_057148 | 2-Sep | 67 | AA848807 | Mterfd1 | 107 | BI278628 | Gdi1 |
| 28 | NM_019275 | Smad4 | 68 | BI303362 | | 108 | AW532525 | Srrm2 |
| 29 | AA818820 | Arfgap1 | 69 | BI279191 | Leng8 | 109 | BF283404 | |
| 30 | NM_012886 | Timp3 | 70 | BF388772 | Golga4 | 110 | BF407666 | Rnf14 |
| 31 | | | 71 | AI412304 | | 111 | AI009074 | Ogt |
| 32 | BG669208 | LOC687237 /RGD1311310 | 72 | NM_012816 | Amacr | 112 | BF282163 | |
| 33 | AB030216 | Elf1 | 73 | BM384116 | Arl6ip1 | 113 | AI229780 | |
| 34 | NM_030586 | Cyb5b | 74 | BI296757 | | 114 | AW918352 | |
| 35 | NM_130755 | Cs | 75 | AA997048 | Galm | 115 | BI275966 | Sipa1 |
| 36 | AI711244 | Mtpn | 76 | BE109560 | Ctnnd1 | 116 | BG672437 | Sv2b |
| 37 | BG666999 | Slc25a4 | 77 | AI406660 | | 117 | AA875132 | |
| 38 | D13921 | Acat1 | 78 | NM_031009 | Agtr1b | 118 | BI291355 | RGD1307235 |
| 39 | NM_053554 | Picalm | 79 | AI703880 | Med13l | 119 | AB013453 | Slc34a1 |
| 40 | NM_019381 | Tmbim6 | 80 | AI103194 | Sec62 | 120 | BF391141 | |