

## Type-2 fuzzy Approach for Disease-Associated Gene Identification on Microarrays

Yan-Fei Wang<sup>1</sup>, Zu-Guo Yu<sup>1,2\*</sup> and Vo Anh<sup>1</sup>

<sup>1</sup>Discipline of Mathematical Sciences, Faculty of Science and Technology,  
Queensland University of Technology, Brisbane, Q4001, Australia.

<sup>2</sup>School of Mathematics and Computational Science, Xiangtan University, Hunan 411105, China.  
Email: yfmu@sina.com (Y.-F Wang), yuzg@hotmail.com (Z.-G. Yu), v.anh@qut.edu.au (V. Anh)

\*Corresponding author

**Abstract**—Fuzzy set theory has been widely used in the analysis of gene microarray data. However, due to noise and uncertainty inherent in microarray data, traditional fuzzy methods sometimes do not perform well. In this paper, we propose a type-2 fuzzy membership test (Type-2 FM test) for disease-associated gene identification on microarrays to improve traditional fuzzy methods. We apply this method on diabetes and lung cancer microarrays and make a comparison with traditional fuzzy methods. For diabetes data, we can identify 7 genes which have been confirmed to be related to diabetes treatment in the published literature and one more gene can be identified than original approaches. For lung cancer data, we can also identify 7 genes which have been confirmed to be associated with lung cancer treatment in published literature and the type-2 d-values are significantly different. The results show that our type-2 FM test performs better than traditional fuzzy methods when analyzing microarray data with similar expression values and noise.

**Keywords:** type-2 fuzzy set; microarray data; uncertainty and noises.

### I. INTRODUCTION

Microarray technology has been used in many areas of bioinformatics research. It can monitor expression of thousands of genes in parallel. The enormous quantities of information data generated have led to a great demand for efficient analysis methods. For some genes, the expression values are different between patients and normal people. Making comparison of these gene expression data in patients and those of normal people can identify disease-associated genes and enhance our understanding of diseases [1]. Researchers have made significant breakthroughs and progress in analysis of disease gene microarray data. Most importantly, they also have identified many genes that have relationship with onset, development and progression of various diseases [2, 3].

In order to identify the disease-associated genes, we usually need to determine for each gene whether the two sets of expression values are significantly different from each other. Measuring the divergence of two sets of values of gene expression data is an effective approach. Two commonly-used statistical methods to measure the divergence of two sets of values are the t-test and the Wilcoxon rank sum test [4]. However, both methods have

limitations. The t-test is very sensitive to extreme values and cannot distinguish two sets with close means even though the two sets are significantly different from each other. Although the rank sum test overcomes the limitation of the t-test in sensitivity to extreme values, it is not sensitive to absolute values. This might be advantageous to some applications but not to others [1].

DNA microarray data contains uncertainty and noise [5, 6]. The above two methods cannot manage this uncertainty properly and always obtains poor quality results [1]. The word “different” itself is a fuzzy concept and fuzzy theory has many advantages in dealing with data containing uncertainty; therefore fuzzy approaches have been taken into consideration to analyze DNA microarrays. Liang et al. [1] proposed a fuzzy set theory based approach, namely a fuzzy membership test (FM-test), for disease genes identification and obtained better results by applying their approach on diabetes and lung cancer microarrays. However, some limitations still exist. The most obviously one is when the values of gene microarray data are very similar and lack of over-expression, in which case the FM-d values are very close or even equal to each other. That made the FM-test inadequate in distinguishing disease genes [1]. Meanwhile, DNA microarray data contains noise, hence yielding uncertain information in the original data. When deriving the membership function for evaluation, all of these uncertainties translate into uncertainties about fuzzy set membership function. Type-1 fuzzy sets are not able to directly model such uncertainties because their membership functions are totally crisp [7].

To overcome these problems, we introduced type-2 fuzzy set theory into the research of disease-associated gene identification. Type-2 fuzzy sets can control the uncertainty information more effectively than conventional type-1 fuzzy sets because the membership functions of type-2 fuzzy sets are three-dimensional. It is the new third dimension of type-2 fuzzy sets that provides additional degrees of freedom that makes it possible to directly model uncertainties [7].

In this paper we establish the type-2 fuzzy membership function for identification of disease-associated genes on microarray data of patients and normal people and apply the type-2 FM-test of this method to diabetes and lung cancer data. For the ten best-ranked genes of diabetes identified by the type-2 FM-test, 7 of them have been confirmed as

diabetes associated genes according to genes description information in Genebank and the published literature. One more gene than original approaches is identified. Within the 10 best ranked genes identified in lung cancer data, 7 of them are confirmed by the literature which is associated with lung cancer treatment. The type-2 FM-d values are significantly different, which makes the identifications more reasonable and convincing than the original FM-test.

## II. TYPE-2 FUZZY SETS

Type-2 fuzzy set was introduced as an extension of a type-1 fuzzy set by Zadeh [8]. Mizumoto and Tanaka [9] studied the set theoretic operations and properties of membership grades of such sets and examined their operations of algebraic product and algebraic sum. Karnik and Mendel [10] extended the algorithms for performing union, intersection and complement for type-2 fuzzy sets. Type-2 fuzzy sets and tools have now been widely used in many fields.

In this paper, we use the following notation and terminology: a type-2 fuzzy set in the universal set  $X$ , denoted as  $\tilde{A}$ , can be characterized by a type-2 fuzzy membership function  $\mu_{\tilde{A}}(x, u)$  as

$$\tilde{A} = \int_{x \in X} \mu_{\tilde{A}}(x) / x = \int_{x \in X} \left[ \int_{u \in J_x} f_x(u) / u \right] / x, \quad J_x \subseteq [0, 1], \quad (1)$$

where  $\mu_{\tilde{A}}(x)$  is the membership grade of  $x \in X$ , which is a type-1 fuzzy set in  $[0, 1]$ ;  $J_x$  is the *primary membership* of  $x$ , a membership of the primary membership of  $x$  is called a *secondary membership* of  $x$  in  $\tilde{A}$ , denoted as  $f_x(u)$ . Fig 1 shows an example of a type-2 fuzzy membership function. The uncertainty in the primary membership, consisting of a bounded region by upper and lower membership functions is called the *footprint of uncertainty* (FOU).

Although type-2 fuzzy sets can be useful in modeling uncertainty where type-1 fuzzy sets cannot, the operations of type-2 fuzzy sets involve numerous embedded type-2 fuzzy sets which include all possible combinations of secondary membership values. Therefore, a large number of computations are required. However, interval type-2 fuzzy sets can reduce the computational complexity significantly [11]. Due to this property, we apply the interval type-2 fuzzy set instead of a general type-2 fuzzy set

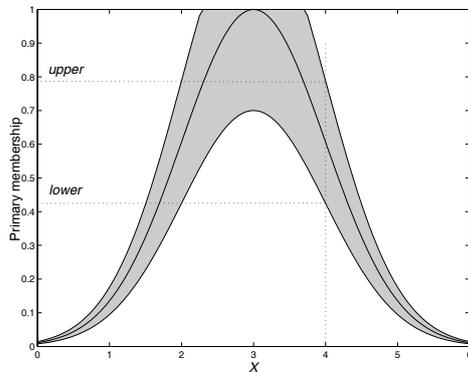


Figure 1. Illustration of a type-2 fuzzy membership function

set in this paper. An interval type-2 fuzzy set is a specific type-2 fuzzy set when  $f_x(u) = 1, \forall u \in J_x \subseteq [0, 1]$ , and is expressed as

$$\tilde{A} = \int_{x \in X} \left[ \int_{u \in J_x} 1/u \right] / x \quad (2)$$

Fig. 2 shows the secondary membership function of an interval type-2 fuzzy set at  $x = 4$ , in which the secondary memberships are all equally weighted for each primary membership of  $x = 4$ . Fig. 3 is a 3D plot of a type-2 fuzzy set whose secondary membership is a triangular function. As shown in the figure, the membership function of type-2 fuzzy sets is three-dimensional.

## III. METHODS

In this section, based on the type-2 fuzzy set theory and the work of Liang *et. al* [1], we propose an interval type-2 FM-test for disease-associated gene identification. For each gene, the microarray data contain expression values from both patients and normal people. Table I. provides an example which contains expression values of 3 genes for two groups of people that are associated with diabetes: five insulin-sensitive (IS) people and five insulin-resistant (IR) people. We denote by  $S_1$  the value set of IS and by  $S_2$  the value set of IR. We consider  $S_1$  and  $S_2$  as two samples of a type-2 fuzzy set. The average membership value gives a measure of the divergence of the original two sets. In particular, this method is performed as followed:

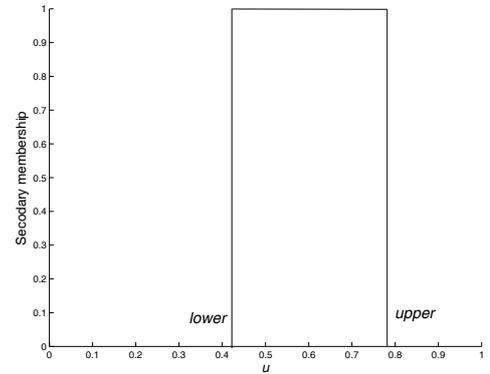


Figure 2. Interval secondary membership function

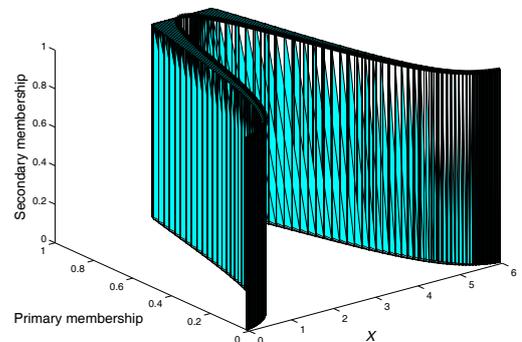


Figure 3. 3D illustration of type-2 fuzzy sets

TABLE I. EXPRESSION VALUES OF IR AND IS PEOPLE

Gene	IR					IS				
1	123	142	11	406	220	305	398	707	905	688
2	200	191	220	83	197	49	81	116	111	135
3	598	424	695	451	141	342	260	266	229	234

1. Using the Gaussian function as the primary membership function to compute the mean ( $\mu_1, \mu_2$ ) and standard deviation ( $\sigma_1, \sigma_2$ ) of  $S_1$  and  $S_2$ .

2. Establishing the primary membership function. The means for the upper and lower primary membership functions are obtain as

$$\mu_{upper} = \max(\mu_1, \mu_2), \mu_{lower} = \min(\mu_1, \mu_2) \quad (3)$$

We can obtain the standard deviations for the upper and lower primary membership function in the same way.

3. Using the primary membership function  $J_{upper}(x), J_{lower}(x)$ , and the secondary membership values  $f_x(u)$  to quantify the convergence of  $S_1$  and  $S_2$ . Type-reduction work is needed in this step. Here, because we use the interval type-2 fuzzy set,  $f_x(u) = 1, \forall u \in J_x \subseteq [0, 1]$ . The secondary memberships are all uniformly weighted for each primary membership of  $x$ .

4. Calculating the divergence degree between the two sets based on the convergence degree.

Type reduction is an important step for type-2 fuzzy sets [11].  $\forall x \in X$ , a primary membership interval  $[J_{lower}(x), J_{upper}(x)]$  can be obtained. We discretize it into  $N$  points, where  $a_1 = J_{lower}(x)$  and  $a_N = J_{upper}(x)$ ; then the final membership of  $x$  can be obtain as

$$\mu(x) = \frac{\sum_{i=1}^N a_i * f_x(a_i)}{N} \quad (4)$$

This type reduced membership  $\mu(x)$  maps each value  $x$  in  $S_1$  or  $S_2$  into a membership value to quantify the degree that  $x$  belongs to a type-2 fuzzy set  $\mu_{\tilde{A}}(x)$ . For simplicity, we put  $a_i = (a_{i-1} + a_{i+1}) / 2, i = 2, \dots, N-1$ , while  $f_x(a_i) = 1, \forall x \in X, i = 1, \dots, N$ , Eq. (4) can be expressed as

$$\mu(x) = \frac{J_{lower}(x) + J_{upper}(x)}{2} \quad (5)$$

Because the membership function maps the elements of  $S_1$  and  $S_2$  into the same type-2 fuzzy set, the aggregation of all membership values in the two sets can be used to quantify the similarity of  $S_1$  and  $S_2$ . It can be considered as an overall bond between these two sets. The weaker this overall bond is, the more divergent these two sets are. In this case, for a given gene, the expression values between patients and normal people can be significantly different. If the elements in both  $S_1$  and  $S_2$  have high membership values,  $S_1$  is very similar with  $S_2$ . However when they both have low values we cannot say  $S_1$  and  $S_2$  are similar, because the type-2 fuzzy set is established based on both  $S_1$  and  $S_2$ . If the elements in  $S_1$

have low membership values, the reason may be the effect from elements in  $S_2$ . So if the elements in the two sets all have low values, it also suggests that  $S_1$  and  $S_2$  are different.

To compute the overall bond between  $S_1$  and  $S_2$ , we define type-2 FM c-values based on Liang et al. [1]:

**Definition 1** Given two sets  $S_1$  and  $S_2$ , the convergence degree between  $S_1$  and  $S_2$  is defined as

$$c(S_1, S_2) = \frac{\sum_{x \in S_1} \mu(x) + \sum_{y \in S_2} \mu(y)}{|S_1| + |S_2|} \quad (6)$$

We define the divergence value as follows:

**Definition 2** Given two sets  $S_1$  and  $S_2$ , the divergence degree between  $S_1$  and  $S_2$  is defined as

$$d(S_1, S_2) = 1 - c(S_1, S_2) \quad (7)$$

#### IV. RESULTS AND DISCUSSION

We use two different data sets from the literature. The first data set is the **Diabetes** data [12]. The original data set contains 10831 genes. To make this data more reliable, only the genes with no null expression values are used for analysis. There are ten expression values in each gene, five from insulin-sensitive (IS) people and the other five from insulin-resistant (IR) people. The second data set is **lung cancer** data which contain 22283 genes [13]. The first five values of each gene in this data set are from squamous lung cancer biopsy specimens and the others are from paired normal specimens. We use the method described above and Liang's FM-test approach on these data sets, and then make a comparison on these two methods. In Tables II and III, the bold letters are names of identified genes.

##### A. Analysis of Diabetes dataset

Ten best-ranked genes of diabetes identified by the type-2 FM-test and the original FM-test are shown in Table II. From this table we see that the results of the two methods are not much different. We identified gene *U61734* and gene *Z26491* by the type-2 FM-test instead of gene *U06452* and gene *X57959* identified by the FM-test. Gene *U06452* and gene *X57959* are just candidate genes for diabetes mentioned in some published papers, but no biological evidence was put forward to confirm that they are related with diabetes [1, 12]. However, for gene *Z26491*, the homo sapiens gene for catechol o-methyltrans-fease (COMT), was found to be differently expressed and helpful for treatment in diabetic rats. Ref. [14] compared the activity of COMT in the livers of diabetic rats with that in normal rats; the result suggested the activity of COMT is lower in diabetic rats than in normal rats. Lal *et al* [15] examined the effect of nitecapone, an inhibitor of the dopamine-metabolizing enzyme COMT and a potent antioxidant, on functional and cellular determinants of renal function in rats with diabetes. The results suggested that the COMT inhibitory and antioxidant properties of nitecapone provide a protective therapy against the development of diabetic nephropathy [15]. These works proved that gene *Z26491* is related with diabetes or treatment. In the remaining genes, *U45973, M60858, L07648, L07033, X53586, X81003* are diabetes associated genes according to

the current literature [1]. *D85181*, *M95610* and *U06452* are recommended as candidate diabetes genes by Liang *et al.* [15], while *X57959* was recommended in Yang *et al.* [12]. We recommended gene *U61734* as candidate gene for future research in this field. From the comparison, one more disease-associated gene is identified by the type-2 FM-test.

### B. Analysis of lung cancer data

Table III shows ten best-ranked genes of lung cancer identified by the two approaches. The results are very different. In the result obtained by the type-2 FM-test, 7 disease associated genes are identified. MFNG is a member of the fringe gene family which also includes radical and lunatic fringe genes. They all encode evolutionarily conserved secreted proteins that act in the

TABLE II. TEN BEST-RANKED GENES ASSOCIATED WITH DIABETES IDENTIFIED BY THE TWO METHODS.

Type-2 FM-test		
Probe Set	Gene Description	T2 d-value
<b>U49573</b>	Human phosphatidylinositol (4,5) biphosphate	0.6733
<b>X53586</b>	Human mRNA for integrin alpha 6	0.6131
<b>M60858</b>	Human nucleolin gene	0.6080
U61734	Homo sapiens transmembrane emp24-like trafficking protein 10	0.5831
D85181	Homo sapiens mRNA for fungal sterol-C5-desaturase homolog	0.5808
<b>Z26491</b>	Homo sapiens gene for catechol o-methyltrans-ase	0.5773
<b>L07648</b>	Human MXII mRNA	0.5769
M95610	Human alpha 2 type IX collagen (COL9A2) mRNA	0.5760
<b>L07033</b>	Human hydroxymethylglutaryl-CoA lyase mRNA	0.5749
<b>X81003</b>	Homo sapiens HCG V mRNA	0.5525
FM-test		
Probe Set	Gene Description	FM d-value
<b>U45973</b>	Human phosphatidylinositol (4,5) biphosphate	0.9988
<b>M60858</b>	Human nucleolin gene	0.9351
D85181	Homo sapiens mRNA for fungal sterol-C5-desaturase homolog	0.8918
M95610	Human alpha 2 type IX collagen (COL9A2) mRNA	0.8718
<b>L07648</b>	Human MXII mRNA	0.8575
<b>L07033</b>	Human hydroxymethylglutaryl-CoA lyase mRNA	0.8554
<b>X53586</b>	Human mRNA for integrin alpha 6	0.8513
<b>X81003</b>	Homo sapiens HCG V mRNA	0.7914
<i>X57959</i>	Ribosomal protein L7	0.7676
<i>U06452</i>	Melan-A	0.7566

Notch receptor pathway. The activity of fringe proteins can alter Notch signaling [16]. Activation of the Notch 1 signaling pathway can impair small cell lung cancer viability [17]. The protein encoded by CTSW is found associated with the membrane inside the endoplasmic reticulum of natural killer (NK) [16]. NK cells play a major role in the rejection of tumors and cells infected by viruses [18]. ALDH3B1 is highly expressed in kidney and lung [16]. Marchitti *et al.* [19] found ALDH3B1 expression was upregulated in a high percentage of human tumors; particularly in lung cancer cell the value is highest. Increasing ALDH3B1 expression in

tumor cells may confirm a growth advantage or be the result of an induction mechanism mediated by increasing oxidative stress [19]. LPCAT1 activity is required to achieve the levels of SatPC essential for the transition to air breathing [20] and it is also upregulated in cancerous lung [21]. Gene PDE3B was mentioned in [22] as the most significantly amplified gene in the tumors. CITED2 is required for fetal lung maturation [23]. Researchers found CITED2 was highly expressed in lung cancer but not in normal tissues, which demonstrates that CITED2 plays a key role in lung cancer progression [24]. Gene DLC1 encodes protein deleted in liver cancer [1]. This gene is deleted in the primary tumor of hepatocellular carcinoma. It maps to 8p22-p21.3, a region frequently deleted in solid tumors. It is suggested that this gene is a tumor suppressor gene for human liver cancer, as well as for prostate, lung, colorectal and breast cancers [16].

8 disease-associated genes are identified by the original FM-test method [1]. However, when we applied the FM-test on lung cancer data, there are more than 80 genes having the same FM d-values; they are all equal to one, which made it difficult to rank and distinguish disease associated genes from others. We have to choose the overexpressed genes from these 80 genes for analysis, which made the task more complicated, and it may miss some important genes. The reason is that the gene expression values in lung cancer microarray data are very close to each other, and the original data is noisy. The FM-test does not seem to be able to deal with these factors suitably.

TABLE III. TEN BEST-RANKED GENES ASSOCIATED WITH LUNG CANCER.

Type-2 FM-test		
Probe Set	Gene Description	T2 d-value
<b>NM_002405</b>	MFNG: MFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase	0.7435
<b>NM_001335</b>	CTSW: cathepsin W	0.7285
NM_017761	PNRC2: praline-rich nuclear receptor coactivator 2	0.7266
AV728526	DTX4: dextex homolog 4 (Drosophila)	0.7265
<b>NM_0002694</b>	ALDH3B1: aldehyde dehydrogenase 3 family, member B1	0.7259
<b>NM_024830</b>	LPCAT1: lysophosphatidylcholine acyltransferase 1	0.7243
BE789881	RAB31: member RAS oncogene family	0.7204
<b>AA888858</b>	PDE3B: phosphodiesterase 3B, cGMP-inhibited	0.7194
<b>NM_006079</b>	CITED2: cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain_2	0.7186
<b>AF026219</b>	DLC1: deleted in liver cancer 1	0.7145
FM-test (Overexpressed)		
Probe Set	Gene Description	FM d-value
<b>NM_173086</b>	KRT6E: Keratin 6E	1
<b>NM_001723</b>	DST: Dystonin	1
<b>NM_002639</b>	SERPINB5: Serpin peptidase inhibitor, clade B (ovalbumin), member 5	1
<b>AB010153</b>	TP73L: Tumor protein p73 like	1
NM_023915	GPR87: G protein-coupled receptor 87	1
<b>NM_006536</b>	CLCA2: Chloride channel, calcium activated, family member 2	1
<b>NM_001005337</b>	PKP1: Plakophilin 1 (ectodermal dysplasia/skin fragility syndrome)	1
<b>AF043977</b>	CLCA2: Chloride channel, calcium activated, family member 2	1

NM_004415	DSP: Desmoplakin	1
NM_019093	UGT1A9: UDP glucuronosyltransferase ! family, polypeptide A9	1

## V. CONCLUSION

We proposed a type-2 FM-test approach based on the type-2 fuzzy set theory and the work of Liang *et al.* [1]. This approach was then applied on diabetes and lung cancer microarray data. For each data set, we analyzed 10 best-ranked genes. In diabetes data, we identified 7 genes which have been confirmed to be related to diabetes or treatment in the published literature. In lung cancer data, we also identified 7 genes. The type-2 FM d-values are significantly different for each gene, which makes the ranking more reasonable than from the original FM-test. The type-2 FM test performs better than the original FM test when analyzing microarray data containing similar expression values and noise.

## ACKNOWLEDGMENT

This project was supported by the Australian Research Council (Grant No. DP0559807), the Natural Science Foundation of China (Grant No. 11071282), the Chinese Program for New Century Excellent Talents in University (Grant No. NCET-08006867), the Program for Furong Scholars of Hunan province of China, the Aid program for Science and Technology Innovative Research Team in Higher Educational Institutions of Hunan Province of China and a China Scholarship Council-Queensland University of Technology Joint Scholarship.

## REFERENCES

- [1] L.R. Liang, S. Lu, X. Wang, Y. Lu, V. Mandal, D. Patacsil and D. Kumar, "FM-test: a fuzzy-set-theory-based approach to differential gene expression data analysis," *BMC Bioinformatics*, 7 (Suppl 4), 2006: S7, doi:10.1126/
- [2] S. Rome, K. Clement, R. Rabasa-Lhoret, E. Loizon, C. Poitou, GS. Barsh, JP. Riou, M. Laville, H. Vidal, "Microarray profiling of human skeletal muscle reveals that insulin regulated approximately 800 genes during a hyperinsulinemic clamp," *J Biol Chem*, vol. 278 (20), 2003, pp. 18063-18068.
- [3] SE. Eckenrode, QG. Ruan, CD. Collins, P. Yang, RA. McIndoe, A. Muir, JX. She, "Molecular pathways altered by insulin b9-23 immunization," *Ann N Y Acad Sci*, vol. 1037, 2004, pp: 175-185.
- [4] B. Rosner, *Fundamentals of Biostatistics*. In Pacific Grove 5th edition, CA: Duxbury Press, 2000.
- [5] E. Someren, L. Wessels, M. Reindersa and E. Backer, "Regularization and noise injection for improving genetic network models," in *Computational and statistical approaches to genomics (2nd edition)*, W. Zhang and I. Shmulevich, Eds. Springer Science+Business Media, Inc, 2006, pp:279-295.
- [6] J. Wang, K. R. Coombes, K. Baggerly, L. Hu, S. R. Hamilton, W. Wang, "Statistical considerations in the assessment of cDNA microarray data obtained using amplification," in *Computation and statistical approaches to genomics*, W. Zhang and I. Shmulevich, Eds. Springer Science+Business Media, Inc, 2006, pp: 21-36.
- [7] J. M. Mendel and R. I. Bob John, "Type-2 fuzzy sets made simple", *IEEE Trans. On Fuzzy Systems*, vol. 10, No. 2, pp: 117-128, April 2002.
- [8] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-1," *Inform. Sci.*, vol. 8, pp. 199-249, 1975.
- [9] M. Mizumoto and K. Tanaka, "Some properties of fuzzy sets of type-2," *Inform. Control*, vol. 31, pp. 312-340, 1976.
- [10] N. N. Karnik and J. M. Mendel, "Operations on type-2 fuzzy sets," *Int. J. Fuzzy sets Syst.*, vol. 122, pp. 327-348, 2001.
- [11] N. N. Karnik, J. M. Mendel, "Type-2 fuzzy logic systems," *IEEE Trans on Fuzzy Systems*, vol. 7, pp. 643-658, December, 1999.
- [12] X. Yang, RT. Pratley, S. Tokraks, C. Bogardus, P. A. Permana, "Microarray profiling of skeletal muscle tissues from equally obese, non-diabetic insulin-sensitive and insulin-resistant Pima Indians," *Diabetologia*, vol. 45 pp. 1584-1593, 2002.
- [13] S. Wachi, K. Yoneda, R. Wu, "Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues." *Bioinformatics*, Vol. 21, pp. 4205-4208, 2005.
- [14] J. Wang, I. Liu, T. Tzeng, J. Cheng, "Decrease in catechol-o-methyltransferase activity in the liver of streptozotocin-induced diabetic rats," *Clin Exp Pharmacol & Physiol*, Vol. 29, pp. 419-422, May, 2002.
- [15] M. A. Lal, A. Korner, Y. Matsuo, S. Zelenin, X. Cheng, G. Jarmko, G. DiBona, A. Eklof, A. Aperia, "Combined Antioxidant and COMT inhibitor treatment reverses renal abnormalities in diabetic rats," *Diabetes*, Vol. 49, pp.1381-1389, August, 2000.
- [16] <http://www.ncbi.nlm.nih.gov/genbank/>.
- [17] C. S. Platta, D. Y. Greenblatt, M. Kunnimalaiyaan, H. Chen, "Valproic acid induces Notch 1 signaling in small cell lung cancer cells," *J. Surgical Research*, Vol 148, pp. 31-37, July, 2008.
- [18] R. Oldham, "Natural killer cells: Artifact to reality: an odyssey in biology," *Cancer Metastasis Reviews*, Vol. 2, pp. 323-326, 1983.
- [19] A. S. Marchitti, D. J. Orlicky, C. Brocker, V. Vasiliou, "Aldehyde Dehydrogenase 3B1 (ALD3B1): Immunohistochemical tissue distribution and cellular-specific localization in normal and cancerous human tissues," *J. Histochemistry and Cytochemistry*, Vol. 58, pp. 765-783, 2010.
- [20] J. P. Bridges, M. Ikegami, L. L. Brilli, X. Chen, R. J. Mason, J. M. Shannon, "LPCAT1 regulates surfactant phospholipid synthase and is required for transitioning to air breathing in mice," *J. Clinical Investigation*, vol. 120, 1736-1748, May, 2010.
- [21] F. Mansilla, K. Costa, S. Wang, M. Kruhoffer, T. M. Lewin, T. F. Orntoft, R. A. Coleman, K. Birkenkamp-Demtroder, "Lysophosphatidylcholine acyltransferase 1 (LPCAT1) overexpression in human colorectal cancer," *J. Mol Med*, Vol. 87, pp. 85-97, January 2009.
- [22] K. C. Lo, L. C. Stein, J. A. Panzarella, J. K. Cowell, L. Hawthorn, "Identification of genes involved in squamous cell carcinoma of the lung using synchronized data from DNA copy number and transcript expression profiling analysis," *Lung Cancer*, vol. 59, pp. 315-331, May, 2008.
- [23] B. Xu, X. Qu, S. G. Y. Doughman, M. Watanabe, S. L. Dunwoodie, Y. Yang, "Cited2 is required for fetal lungmaturation", *Dev Biol*, Vol. 317, pp. 95-105, May, 2008.
- [24] Y. T. Chou, C.F. Hsu, Y. R. Kao, C. W. Wu, "Cited2, a novel EGFR-induced coactivator, plays a key role in lung cancer progression," *AACR 101st Annual Meeting, Poster Presentation*, 2010.