

Determination of the SNP-SNP Interaction between Breast Cancer Related Genes to Analyze the Disease Susceptibility

Mei-Lee Hwang, Yu-Da Lin, Li-Yeh Chuang, and Cheng-Hong Yang

Abstract—Investigation of the single nucleotide polymorphism (SNP)-SNP interaction model can facilitate the analysis of the susceptibility to disease. The model explains the risk of association between the genotypes and the disease in case-control study. Thus, many mathematic methods are widely applied to identify the statistically significant model such as odds ratio (*OR*), chi-square test, and error rate. However, a huge number of data sets have been found to limit the statistical methods to identify the significant model. In this study, we propose a novel statistical method, complementary-logic particle swarm optimization (CLPSO), to increase the efficiency of significant model identification in case-control study. The complementary-logic is implemented to improve the PSO search ability and identify a better SNP-SNP interaction model. Six important breast cancer genes including 23 SNPs and simulated huge number of data sets were selected as the test data sets. The methods of PSO and CLPSO were applied on the identification of SNP-SNP interactions in the two-way to five-way. In results, the *OR* evaluates the breast cancer risk of the identified SNP-SNP interaction model. Compared to the corresponding non-interaction model, if the *OR* value is greater than 1 that indicates the model is significant risk between cases and controls. The results showed that CLPSO is able to identify the significant models for specific SNP-SNP interaction of two-way to five-way (*OR* value: 1.153-1.391; confidence interval (CI): 1.05-1.79; *p*-value: 0.01-0.003). The model suggests that the genes *ESR1*, *PGR*, and *SHBG* may be an important role in the interactive effects to breast cancer. In addition, we compared the search abilities of PSO and CLPSO for identification of the significant model. Results revealed that CLPSO can identify better model with difference values between cases and controls than PSO; it suggests CLPSO can be used to identify a better SNP-SNP interaction models.

Index Terms—Single nucleotide polymorphism (SNP), particle swarm optimization (PSO), breast cancer.

I. INTRODUCTION

SNP is an common bio-marker in genomes, and it has widely used in the investigation of association analysis of diseases [1], cancers [2], and pharmacogenomics [3]. These

Manuscript received April 1, 2014; revised May 30, 2014. This work was supported in part by the National Science Council in Taiwan (NSC101-2320-B-037-049, NSC101-2622-E-151-027-CC3, NSC102-2221-E-214-039 and NSC102-2221-E-151-024-MY3).

Mei-Lee Hwang and Li-Yeh Chuang are with the Department of Chemical Engineering & Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung, Taiwan (e-mail: mlhwang@isu.edu.tw, chuang@isu.edu.tw).

Yu-Da Lin and Cheng-Hong Yang are with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan (e-mail: e0955767257@yahoo.com.tw, chyang@cc.kuas.edu.tw).

analyses reported that SNPs have specific associations with the risk of certain diseases. However, most of the SNP analyses were focused on a single SNP. The low or marginal significant SNPs could be excluded, but these SNPs may have significant associations with disease when they are combined as SNP-SNP interaction model. Thus, the identification of appropriate interaction model is an important issue for SNP analysis.

A SNP-SNP interaction model includes the SNPs and their corresponding genotypes (AA, Aa, and aa). Therefore, the possible models are rapidly increased by the number of SNPs and their corresponding genotypes. The huge potential SNP-SNP interaction model makes the statistical method difficultly identify the significant models. Currently, machine learning have applied to help statistical method on the identification of appropriate interaction models, such as particle swarm optimization (PSO) [4]-[6] and genetic algorithm (GA) [7], [8]. PSO and GA have the properties of randomized search and are an optimization technique that derives its working principles from simulations of the organism behavior. They provide a fast identification in high-dimension problem, e.g., biomarker selection [9]. Thus, they can be employed to identify an optimal SNP combination from the huge possible combinations. Although they had overcome the excessive computational time to identify the significant models, the search ability remains a challenge. Thus, an improved method is required to explore the SNP-SNP interaction.

Here, we propose a complementary-logic PSO, named CLPSO, to identify the significant model associated with SNP-SNP interactions. In this study, we hypothesize that the interactions between polymorphisms of genes may have a synergistic or non-additive effect on the pathogenesis of a disease. This interaction may explain differences between cases and controls in the disease risk. Six breast cancer related genes (*COMT*, *CYP19A1*, *ESR1*, *PGR*, *SHBG*, and *STS*) including 23 SNPs were selected to simulate huge number of data sets. Results indicate that CLPSO can identify the appropriate interaction models in breast cancer from the huge number of simulated data sets, and the results provide the significant information for determining the SNP-SNP interaction model with maximal difference between the cases and controls.

II. METHODS

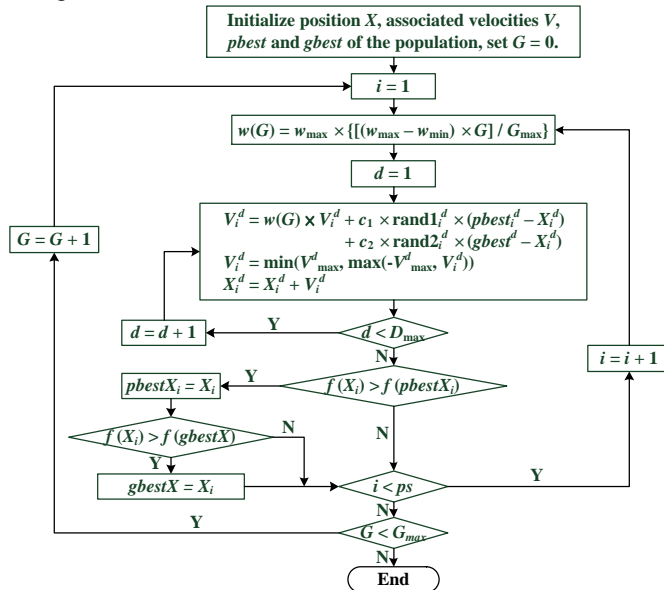
A. Problem Definition

For SNP-SNP interaction problem, a vector like $X = [x_1, x_2,$

$x_3, \dots, x_D]$ is represented as one condition and the solution for each parameter x_i is a real number; in this study the SNP-SNP interaction models are referred as the different SNP combinations with their corresponding genotypes. The object of SNP-SNP interaction model is a search for the vector X^* , which maximizes an objective function $f(X)$ ($f: \delta \subseteq R^D \rightarrow R$), i.e., $f(X^*) > f(X)$ for all $X \in \delta$, where δ is a non-empty large finite set serving as the search space, and $\delta = R^D$. The objective function defined the difference between frequencies of case group and control group, which are according to vector X , i.e., $f(X) = ()$ is generally complicated by the existence of non-linear objective functions with multiple local optima. A local optima $f_l = f(X_l)$ in this study defined as $\exists \varepsilon > 0 \forall X \in \delta: \|X - X_l\| < \varepsilon \Rightarrow f(X) \leq f_l \leq f(X^*)$, where $\|\cdot\|$ indicates any p -norm distance measure. The flexibilities of given constraints and vector space in problem influence the finding best-suited solution in PSO process. We aim at finding a set of SNP combinations with their genotypes for which the overall performance of the model will be the best under some given conditions.

B. Particle Swarm Optimization

PSO is based on a population stochastic optimization technique; it simulates the social behavior of organisms as birds in a flock or fish in a school. Each individual (named bird or particle) uses its own memory and common knowledge gained by the entire swarm to find an optimal position. The particles have two important properties: (1) a fitness value, which is determined by objective function, and (2) a velocity, which affects the movement of the particle. During the generation, each particle moves its position according to its own experience and the common experience of all particles. Finally, the particles follow the current best particle in the search space until a predefined number of iterations are reached. The flowchart of the method is shown in Fig. 1.



i : the i^{th} particle from 1 to ps ps : population size
 G : the G^{th} generation from 1 to G_{\max} G_{\max} : maximum generations
 d : the d^{th} dimension from 1 to D_{\max} D_{\max} : maximum dimensions
 X_i^d : the value of d^{th} dimension of i^{th} particle w : inertia weight
 $f()$: objective function

Fig. 1. Flowchart of particle swarm optimization.

C. Complementary-Logic Particle Swarm Optimization (CLPSO)

The complementary-logic aims to avoid the particles to be easily trapped in a local optimum by moving their position to a new region in the problem space. In traditional PSO, particles could be trapped in a local optimum due to the premature convergence of population. All particles are randomly generated in the problem space and updated by the $pbest$, which is the best fitness value in the particle itself has achieved so far, and $gbest$, which is the best fitness of all particles in the population so far. However, all particles based on update function can move toward the $gbest$ and the distance between $gbest$ and the neighboring particles can be gradually decreased by increased generation. Here, we use complementary-logic to improve the PSO search ability for SNP-SNP interaction model, in which the complementary particles are introduced to avoid premature convergence. These complementary particles replace 50% of the randomly selected particles in the population. The procedure of CLPSO includes (1) initialization of the parameter vectors, (2) objective function, (3) identification of $pbest$ and $gbest$, (4) particle updating, (5) complementary-logic adaption of particles, and (6) judgment of termination condition.

D. Initialization of the Parameter Vectors

CLPSO searches for a global optimum point in a D -dimensional real parameter space R^D . It begins with a randomly initiated population of NP D dimensional real-valued parameter vectors. Each vector, also known as genome/chromosome, forms a candidate solution to the multidimensional optimization problem. We shall denote subsequent generations in CLPSO by $G = 0, 1, \dots, G_{\max}$. Since the parameter vectors are likely to be changed over different generations, we may adopt the following notation for representing the i^{th} vector of the population at the current generation:

$$X_{i,G} = [x_{1,i,G}, x_{2,i,G}, x_{3,i,G}, \dots, x_{D,i,G}] \quad (1)$$

For each parameter of the problem, there may be a certain range within which the value of the parameter should be restricted, often because parameters are related to physical components or measures that have natural bounds (for example if one parameter is a length or mass, it cannot be negative). The initial population (at $G = 0$) should cover this range as much as possible by uniformly randomizing individuals within the search space constrained by the prescribed minimum and maximum bounds: $X_{\min} = \{x_{1,\min}, x_{2,\min}, \dots, x_{D,\min}\}$ and $X_{\max} = \{x_{1,\max}, x_{2,\max}, \dots, x_{D,\max}\}$. Therefore, we may initialize the j^{th} component of the i^{th} vector as:

$$x_{j,i,0} = x_{j,\min} + rand_{i,j}[0,1] * (x_{j,\max} - x_{j,\min}) \quad (2)$$

where $rand_{i,j}[0, 1]$ is a uniformly distributed random number lying between 0 and 1 (actually $0 \leq rand_{i,j}[0, 1] \leq 1$) and is instantiated independently for each component of the i^{th} vector. In CLPSO, a chromosome in the population represents a solution group and can be divided into two parts: the number of selected SNPs, and the genotypes associated with the SNPs.

The chromosome encoding can thus be represented by:

$$x_{j,\max} = \begin{cases} SNP_{\max} & j \leq D/2 \\ Genotype_{\max} & j > D/2 \end{cases} \quad (3)$$

$$x_{j,\min} = \begin{cases} SNP_{\min} & j \leq D/2 \\ Genotype_{\min} & j > D/2 \end{cases} \quad (4)$$

$$Genotype = \begin{cases} 1 & \text{dominant allele} \\ 2 & \text{heterozygous allele} \\ 3 & \text{recessive allele} \end{cases} \quad (5)$$

where SNP_{\max} and SNP_{\min} represents a limited SNPs, $Genotype_{\max}$ and $Genotype_{\min}$ represents the limited possible genotypes. For example, let $X_{i,0} = (2, 4, 6, 3, 1, 3)$, thus represents i^{th} X in first generation (i.e., 0) chosen SNPs (2, 4, 6) and genotypes (3, 1, 3), and can be described by the SNPs associated with the genotypes as follows: (2, 3), (4, 1) and (6, 3).

E. Objective Function

In the CLPSO process, the fitness function value measures the quality of chromosomes. The SNP-SNP interaction study focuses on the particular SNP combinations to detect the highest fitness value, i.e., the maximum difference value between breast cancer cases and non-cancer cases. This criterion divides the fitness function into three separate steps, and the relevant equation can be written as:

$$f(X_i) = \sum_{j=1}^P u(X_i, P_j) - \sum_{j=1}^N u(X_i, N_j) \quad (6)$$

$$u(X_i, A) = \begin{cases} 1 & \forall x \in A \\ 0 & \forall x \notin A \end{cases}, \forall x \in X_i$$

The concept of the designed fitness value uses the intersection of set theory to evaluate the difference value between breast cancer cases and non-cancer cases. The intersection of two sets is the set that contains all elements found in both sets, but no other elements. For example: $X = (SNP_{2,3}, Genotype_{3,2})$ is used to evaluate the number of matching conditions in the breast cancer cases and non-cancer cases. Let's suppose the number of independent matching SNP_2 with genotype 3 and SNP_3 with genotype 2 is 273 in the breast cancer cases, and the number of independent matching SNP_2 with genotype 3 and SNP_3 with genotype 2 is 51 in the non-cancer cases. According to Eq. (6), the fitness value is determined by subtracting 51 from 273, leaving 222, which represents a high risk.

F. Identification of $pbest$ and $gbest$

In CLPSO, each particle can find the currently best position and velocity ($pbest$) and the global best position and velocity ($gbest$) when population moving. If the fitness value of a particle P_i in the current generation is better than the fitness value of $pbest$ in the previous generation, then $pbest$ is replaced by P_i . If the fitness value of a particle P_i in the current generation is better than $gbest$ in the previous generation, $gbest$ is also replaced by P_i . Thus, each particle

can adjust its position based on $pbest$ and $gbest$ in the next iteration.

G. Particle Updating

Each particle moves its position by updating the velocity in the next generation via an evaluation of $pbest$ and $gbest$. The update functions of velocity and position are formulated in Eqs. (7) to (9):

$$w_{LDW} = (w_{\max} - w_{\min}) \times \frac{Iteration_{\max} - Iteration_i}{Iteration_{\max}} + w_{\min} \quad (7)$$

$$v_{id}^{new} = w_{LDW} \times v_{id}^{old} + c_1 \times r_1 \times (pbest_{id} - x_{id}^{old}) + c_2 \times r_2 \times (gbest_d - x_{id}^{old}) \quad (8)$$

$$x_{id}^{new} = x_{id}^{old} + v_{id}^{new} \quad (9)$$

In Eq. (7), w_{\max} set to be 0.9, w_{\min} set to be 0.4 and $Iteration_{\max}$ is the maximum generation number. In Eq. (8), r_1 and r_2 are random values between (0, 1); c_1 and c_2 are acceleration coefficients equal to 2; these coefficients constantly influence a particle moves in a single generation.

H. Complementary-Logic Adaption of Particles

The complementary-logic aims to avoid the particles are trapped into a local optimum by adjusting particles to a new region in the problem space. Thus, complementary-logic can facilitate the ensuring global exploration. The position of a trapped particle is changed by the Eq. (10).

$$x_{id}^{Complement} = (x_{\max} + x_{\min}) - x_{id}^{selected} \quad (10)$$

where x_{\max} and x_{\min} are the maximum and minimum SNP number, respectively.

I. Judgment of Termination Condition

The CLPSO procedure is terminated if the maximum number of generation is reached. When the termination condition is reached, $gbest$ is the optimal solution, i.e., best SNP-SNP interaction model. The maximum number of generation is explained in the parameter settings section.

J. Parameter Settings

We set the population size of parameters = 50, number of generations = 100. The termination condition is that the generations met the maximum number of allowed generations.

K. Statistical Analysis

Odds ratio (OR) is used to evaluate the difference between cases and controls. It determines the combination of genotypes and quantitatively measures the association disease risk [10], and OR is defined as follows:

$$\text{Odds ratio} = \frac{TP \times TN}{FP \times FN} \quad (11)$$

where TP is the number of true positives, TN is the number of true negatives, FN is the number of false negatives and FP is the number of false positives.

TABLE I: SNPS INVESTIGATED IN THIS STUDY AND THEIR NORMALIZED NUMBERS OF POLYMORPHISM GENOTYPING INFORMATION

SNP No.	Gene (SNP)	Genotype			SNP No.	Gene (SNP)	Genotype		
		AA (Control /Cases)	Aa (Control /Cases)	aa (Control /Cases)			AA (Control /Cases)	Aa (Control /Cases)	aa (Control /Cases)
1	COMT	1769	2390	841	13	ESR1	1773	2430	797
	(rs6269)	1694	2389	917		(rs9478249)	1890	2381	729
2	COMT	1377	2417	1206	14	ESR1	1830	2415	755
	(rs4680)	1308	2440	1252		(rs1514348)	1717	2435	851
3	CYP19A1	1430	2497	1073	15	ESR1 (rs532010)	1891	2422	687
	(rs10046)	1434	2411	1155			1848	2377	775
4	ESR1	2343	2164	493	16	PGR	2014	2326	660
	(rs3020314)	2147	2280	573		(rs566351)	2062	2280	658
5	ESR1	1450	2524	1026	17	PGR	2591	2042	367
	(rs2234693)	1446	2480	1074		(rs660149)	2708	1927	365
6	ESR1	1467	2441	1092	18	PGR	2338	2163	499
	(rs1543404)	1468	2439	1093		(rs11571171)	2419	2082	499
7	ESR1	1406	2542	1052	19	PGR	2994	1767	239
	(rs3798577)	1413	2494	1093		(rs500760)	2888	1866	246
8	ESR1	1372	2447	1181	20	SHBG	1597	2490	913
	(rs2747652)	1377	2479	1144		(rs858518)	1693	2412	895
9	ESR1	1347	2589	1064	21	SHBG	1523	2442	1035
	(rs2077647)	1383	2449	1168		(rs272428)	1609	2438	953
10	ESR1	1353	2457	1190	22	SHBG	1725	2393	882
	(rs2175898)	1350	2507	1143		(rs858524)	1613	2459	928
11	ESR1	2107	2302	591	23	STS	1760	2437	803
	(rs9340799)	2016	2360	624		(rs2017591)	1823	2258	919
12	ESR1	1988	2341	671					
	(rs1709182)	1932	2326	742					

TABLE II: ESTIMATION OF THE BEST INTERACTION MODEL WITH COMBINATION OF SNPS AND GENOTYPES ON THE OCCURRENCE OF BREAST CANCER DETERMINED BY PSO AND CLPSO

		Combined SNP	SNP genotypes	Cases No.	Controls No.	Difference	Odds Ratio	95% CI	p-value
2-SNP	PSO	4,17	2-1 other	1210 3790	1084 3916	126	1.153	1.05-1.27	0.003
	IPSO	4,17	2-1 other	1210 3790	1084 3916	126	1.153	1.05-1.27	0.003
3-SNP	PSO	2,11,19	2-2-2 other	441 4559	401 4599	40	1.109	0.96-1.28	0.150
	IPSO	4,17,18	2-1-1 other	602 4398	509 4491	93	1.208	1.07-1.37	0.003
4-SNP	PSO	4,8,21,23	2-3-2-3 other	48 4952	33 4967	15	1.459	0.94-2.28	0.096
	IPSO	4,17,18,22	2-1-1-2 other	298 4702	238 4762	60	1.268	1.06-1.51	0.008
5-SNP	PSO	4,5,7,17,18	2-2-2-1-1 other	146 4854	130 4870	16	1.127	0.89-1.43	0.329
	IPSO	4,11,17,18,22	2-2-1-1-2 other	149 4851	108 4892	41	1.391	1.08-1.79	0.010

III. RESULT AND DISCUSSION

A. Data Set

We selected six important breast cancer genes (COMT, CYP19A1, ESR1, PGR, SHBG, and STS) as the test data sets, including 23 SNPs. The SNPs and genotype distributions are shown in Table I. The SNP frequencies were obtained from the reference [11]. These frequencies were used to simulate a huge data set to test PSO and CLPSO. The sample size is 5000, and the SNP genotypes are randomly generated according to the above frequencies. The simulated data for SNP_a is based on its percentage which is obtained by multiplication of the percentage with the amount of the complete data set. For example, 50%×5000=2500 for AA, 30%×5000=1500 for Aa and 20%×5000=1000 for aa. Thus,

the data for SNP_a has generated to sample 5000 (2500+1500+1000=5000). All original data are generated by the same number in this manner. The simulation procedure is shown in the Fig. 2.

Pseudo-code for randomly generated data
01: begin
02: Set size = 5000
03: Set number of genotype = 3
04: Calculate amount of three genotypes
05: while (all SNPs are not normalized)
06: Calculate amount of each genotype
07: Calculate numbers of each normalized genotype
08: for n = 1 to number of genotype
09: Randomly create numbers of each normalized genotype
10: next n
11: end

Fig. 2. Pseudo-code for randomly generated data.

B. Analysis of Breast Cancer Susceptibility from 23 SNPs in Six Genes

The odds ratio and its 95% CI for all SNPs of six genes (COMT, CYP19A1, ESR1, PGR, SHBG, and STS) show that 13 SNPs, including COMT-rs6269, ESR1-rs3020314, ESR1-rs2175898, ESR1-rs1709182, ESR1-rs9478249, ESR1-rs1514348, ESR1-rs532010, PGR-rs660149, PGR-rs500760, SHBG-rs858518, SHBG-rs272428, SHBG-rs858524, and STS-rs2017591, display the statistically significant *OR* ($p < 0.05$) for breast cancer; the *OR* values range between 1.268 to 0.846.

C. Analysis of SNP-SNP Interaction Model with Difference between the Cases And Controls

Table II shows the results of SNP-SNP interaction models by 2 to 5-way. The left side in Table II represents the two to five SNP combinations. In these combinations, the 2-way model with SNP combinations and corresponding genotype indicates the rs3020314-Aa associated with rs660149-AA. The difference column indicates the difference between the cases and controls. For example, the 3-way model of CLPSO shows a difference value of 93 between cases and controls (602 vs. 509). Thus, the other models of PSO and CLPSO are explained as above mention. In CLPSO, the difference values between cases and controls are reduced from 126 to 41 in the two-way to five-way SNP-SNP interaction models. In PSO, the difference values between cases and controls are reduced from 126 to 15 in the two-way to five-way SNP-SNP interaction models. The larger difference value between cases and controls represents the better model.

D. Estimation of the Best Interaction Model Generated by PSO and CLPSO Using or and 95% CI in Breast Cancer

Table II shows the best interaction model in the 2-way to 5-way SNP-SNP interaction models. These results reveal that the total number of cases exceeds the total number of controls; it means that all models are a risk association in breast cancer. The right side in Table II shows the evaluated effects using odds ratio, 95% CI, and *p*-value. In PSO, the *OR* values in 2-way to 5-way SNP-SNP interaction models show the range of 1.109 to 1.459, and the 95% CI of *OR* is in the range of 0.89 to 2.28. Only the 2-way SNP-SNP interaction model shows the statistically significant (p -value < 0.05). However, in CLPSO, the *OR* values in 2-way to 5-way SNP-SNP interaction models show the range of 1.153 to 1.391 and the 95% CI of *OR* is in the range of 1.05 to 1.79. All of the SNP-SNP interaction models (2-way to 5-way) show significant *OR* values (p -value < 0.050).

E. Comparison of PSO and CLPSO for the Interaction Model of Breast Cancer

The results represent the SNP-SNP interaction model identified by CLPSO has a better *p*-value than the model identified by PSO. PSO seems to provide a better *OR* value in the four-way SNP-SNP interaction model, however, the *p*-value shows the model does not statistically significant for breast cancer. The computational complexity of CLPSO is evaluated by objective function computation. Let *n* generation is implemented in a test, the computational complexity of PSO is $O(n)$ which is the big-O in complexity analysis. The

complementary-logic only adds an updated function, i.e., equation 10. Thus, the computational complexity between PSO and CLPSO is the same, but CLPSO is superior to PSO in terms of identifying the best SNP-SNP interaction model.

IV. CONCLUSION

In this study, a novel method, CLPSO, is proposed to identify the statistically significant SNP-SNP interaction models between related genes of breast cancer. These models can be used to analyze disease susceptibility and provide the information of SNPs located in the genes and their associated pathways. We used the huge number of simulated data to test the methods of PSO and CLPSO, the results indicate CLPSO can robust to search the statistically significant models with the difference value between SNPs of genes amongst the huge number of SNPs involved in real data sets.

REFERENCES

- [1] I. C.Gray, D. A. Campbell, and N. K. Spurr, "Single nucleotide polymorphisms as tools in human genetics," *Human Molecular Genetics*, vol. 9, pp. 2403-2408, 2000.
- [2] C. R. Cantor, "The use of genetic SNPs as new diagnostic markers in preventive medicine," *Longevity Health Sciences: The Phoenix Conference*, vol. 1055, pp. 48-57, 2005.
- [3] D. A. Roses, A. M. Saunders, Y. Huang, J. Strum, K. H. Weisgraber *et al.*, "Complex disease-associated pharmacogenetics: drug efficacy, drug safety, and confirmation of a pathogenetic hypothesis (Alzheimer's disease)," *Pharmacogenomics Journal*, vol. 7, pp. 10-28, 2007.
- [4] H. W. Chang, C. H. Yang, C. H. Ho, C. H. Wen, and L. Y. Chuang, "Generating SNP barcode to evaluate SNP-SNP interaction of disease by particle swarm optimization," *Computational Biology and Chemistry*, vol. 33, pp. 114-119, 2009.
- [5] C. H. Yang, H. W. Chang, Y. H. Cheng, and L. Y. Chuang, "Novel generating protective single nucleotide polymorphism barcode for breast cancer using particle swarm optimization," *Cancer Epidemiol.*, vol. 33, pp. 147-154, 2009.
- [6] L. Y. Chuang, H. W. Chang, M. C. Lin, and C. H. Yang, "Chaotic particle swarm optimization for detecting SNP-SNP interactions for CXCL12-related genes in breast cancer prevention," *European Journal of Cancer Prevention*, vol. 21, pp. 336-342, 2011.
- [7] H. W. Chang, L. Y. Chuang, C. H. Ho, P. L. Chang, and C. H. Yang, "Odds ratio-based genetic algorithms for generating SNP barcodes of genotypes to predict disease susceptibility," *OMICS-a Journal of Integrative Biology*, vol. 12, pp. 71-81, 2008.
- [8] C. H. Yang, L. Y. Chuang, Y. J. Chen, H. F. Tseng, and H. W. Chang, "Computational analysis of simulated SNP interactions between 26 growth factor-related genes in a breast cancer association study," *OMICS-a Journal of Integrative Biology*, vol. 15, pp. 399-407, 2011.
- [9] H. W. Resson, R. S. Varghese, M. Abdel-Hamid, S. A. L. Eissa, D. Saha *et al.*, "Analysis of mass spectral serum profiles for biomarker selection," *Bioinformatics*, vol. 21, pp. 4039-4045, 2005.
- [10] L. E. Mechanic, B. T. Luke, J. E. Goodman, S. J. Chanock, and C. C. Harris, "Polymorphism interaction analysis (PIA): a method for investigating complex gene-gene interactions," *BMC Bioinformatics*, vol. 9, pp. 146, 2008.
- [11] P. D. P. Pharoah, J. Tyrer, A. M. Dunning, D. F. Easton, B. A. J. Ponder *et al.*, "Association between common variation in 120 candidate genes and breast cancer risk," *PLoS Genetics*, vol. 3, pp. 401-406, 2007.



Mei-Lee Hwang received the MS degree from the Department of Chemistry, University of Memphis in 1987 and the PhD degree from the Department of Chemistry, University of Cincinnati in 1992. She is an associate professor in the Department of Chemical Engineering and Institute of Biotechnology and Chemical Engineering at I-Shou University, Kaohsiung, Taiwan. Her main areas of research are physical chemistry and molecular

spectrum.



Yu-Da Lin received the MS degree from the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan, in 2011. He is currently working toward the PhD degree in the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Taiwan. He has rich experience in computer programming, database design and management, and systems programming and design. His main areas of research are bioinformatics and computational biology.



Cheng-Hong Yang received the MS and PhD degrees in computer engineering from North Dakota State University in 1988 and 1992, respectively. He is a professor in the Department of Electronic Engineering at the National Kaohsiung University of Applied Sciences, Taiwan. His main areas of research are evolutionary computation, bioinformatics, and assistive tool implementation.



Li-Yeh Chuang received the MS degree from the Department of Chemistry, University of North Carolina in 1989 and the PhD degree from the Department of Biochemistry, North Dakota State University in 1994. She is a professor in the Department of Chemical Engineering and Institute of Biotechnology and Chemical Engineering at I-Shou University, Kaohsiung, Taiwan. Her main areas of research are bioinformatics, biochemistry, and genetic engineering.