

Chaotic Particle Swarm Optimization for Prediction of SNP Combinations Associated with a Low Bone Mineral Density Risk

Li-Yeh Chuang, Ming-Cheng Lin, Hsueh-Wei Chang, and Cheng-Hong Yang, *Member, IAENG*

Abstract—Osteoporosis is a major public health problem that can be quantified by low bone mineral density (BMD) measurements. Many association studies have analyzed the genotype frequencies of case and control groups to predict the susceptibility to the disease. An increasing number of studies have shown that the disease risk is associated with the co-occurrence of single nucleotide polymorphisms (SNPs). Hence, which necessitates a further analysis of SNP-SNP interaction combinations. However, using exhaustive search (ES) algorithms to calculate SNP combinations requires a large amount of time. To shorten the calculation time, we propose a chaotic particle swarm optimization (CPSO) method that uses the odds ratio (OR) to determine the disease susceptibility. PSO is applied to generate SNP combinations, with a maximal difference of occurrence between the case and control groups. The incorporated chaotic map introduces certainty, ergodicity and a stochastic property into PSO in order to improve the global convergence. This study uses a real-life dataset of 10 SNPs with 113 individuals in control group and 184 individuals in case group. The estimated OR of the best SNP combination with genotypes is significantly higher than 1 (between 1.550 and 2.309) for specific combinations of two to nine SNPs in the high risk group for osteoporosis (low BMD).

Index Terms—Osteoporosis, bone mineral density, single nucleotide polymorphisms, chaotic particle swarm optimization, odds ratio

I. INTRODUCTION

Many association studies indicate that the frequency with which genotype occur in case and control data is relevant to predict the susceptibility to a disease or a cancer, e.g., osteoporosis [1], breast cancer [2, 3], oral cancer [4, 5] etc. The risk of a disease or cancer is associated with the co-occurrence of single nucleotide polymorphisms (SNPs). Hence, determining these disease-causing SNPs and exploring SNP-SNP interactions (epistasis) have become an important objective [3, 6]. Recently, Phillips defined three

terms used to explain epistasis, namely compositional epistasis, statistical epistasis and functional epistasis [7]. Compositional epistasis blocks the effect of one allele by another at a different locus. Statistical epistasis constitutes a statistical deviation from the additive effects of two loci on the phenotype, and functional epistasis addresses molecular interactions [7-9].

SNPs are an abundant form of genetic variations amongst species. The association of SNPs, diseases and cancers, as well as their therapies and pharma-cogenomics, were studied and reviewed in previous literature. SNPs have been historically classified as commonly occurring (>1%) genetic variations in the general population [8, 10-12]. Many methods have been proposed to analyze SNP-SNP interactions in multiple SNP combinations, these include multifactor dimensionality reduction (MDR) [13], support vector machine (SVM) [14], polymorphism interaction analysis (PIA) [15], random jungle [16], machine learning [17], and others. The main difference between these methods lies in the computational time required and the quantitative measures used. The MDR method can detect high-order SNP-SNP interactions in case-control studies; however, it does not have a clear quantitative measure. Hence, Chung et al. improved MDR by adding the odds ratio (OR) to estimate SNPs combinations [18].

In this study, we used chaotic particle swarm optimization (CPSO) to explore SNP-SNP combinations. PSO is an automatically evolving algorithm derived from the natural evolution of the social behaviour of organisms [19, 20]. The advantages of PSO are its fast convergence and the fact that it requires relatively few parameters that need to be set. It also allows for an individual memory of the particles that can be used to compare information in a search process. Due to these advantages, PSO has been successfully applied in many fields, including operon [21] and CpG island prediction [22]. However, the original PSO easily gets stuck in a local optimum. Hence, we added a chaos map to improve its performance. Chaotic systems use a chaotic map and an initial condition to generate an enormous number of sequences [23]. Different sequences can be changed easily by modifying the initial condition. Chaos is greatly sensitive to its initial conditions and its mapping is characterized by both certainty and randomness. Finally, the odds ratio (OR) is used to conveniently interpret the data of case-control studies. It is a common statistic that expresses the strength of association between an exposure and a disease [24].

L. Y. Chuang is with the Chemical Engineering Department, I-Shou University, 84001, Kaohsiung, Taiwan. (e-mail: chuang@isu.edu.tw).

W. L. Huang is with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, 80778, Kaohsiung, Taiwan. (e-mail:ss10136@yahoo.com.tw).

H. W. Chang is with the Department of Biomedical Science and Environmental Biology Cancer Center, 80778, Kaohsiung, Taiwan. (changhw@kmu.edu.tw)

C. H. Yang is with the Network Systems Department, Toko University, 61363, Chiayi, Taiwan.

C. H. Yang is also with the Electronic Engineering Department, National Kaohsiung University of Applied Sciences, 80778, Kaohsiung, Taiwan. (corresponding author to provide phone: 886-7-3814526#5639; e-mail: chyang@cc.kuas.edu.tw).

II. METHODOLOGY

A. Data sets

The dataset was obtained from Lin et al [25], who conducted a bone mineral density association study for osteoporosis, the dataset includes 11 SNPs. Considering the dataset missing data, finally we select 10 SNPs based on 113 subjects with a high BMD (control) and 184 subjects with a low BMD (case) dataset. Information relating to this dataset is shown in Table I.

TABLE I.
SNP INFORMATION FOR OSTEOPOROSIS ASSOCIATION STUDY

SNP No	Gene (SNP)	Chr.	Genotype					
			1		2		3	
			Nca	Nco	Nca	Nco	Nca	Nco
1	TNF α -857 rs1799724	6	TT		CT		CC	
			16	29	43	32	125	52
2	TGF β 1-50 9 rs1800469	19	TT		CT		CC	
			58	37	93	63	33	13
3	Osteocalcin rs1800247	1	CC		CT		TT	
			18	8	84	44	82	61
4	TNF α -308 rs1800629	6	AA		AG		GG	
			5	4	33	21	146	88
5	PTH (<i>BstB I</i>) rs6254	11	GG		AG		AA	
			5	2	30	25	149	86
6	PTH (<i>Dra II</i>) Rs6256	11	AA		AC		CC	
			4	3	41	18	139	92
7	HSP70 hom rs2227956	6	CC		CT		TT	
			9	5	71	42	104	66
8	HSP70-2 rs1061581	6	GG		AG		AA	
			29	25	145	86	10	2
9	CTR rs1801197	7	CC		CT		TT	
			127	84	55	27	2	2
10	BMP-4 Rs17563	14	CC		CT		TT	
			12	14	76	38	96	61

No.:Number. Nca: Number of cases. Nco.:Number of controls.

B. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is a population based stochastic optimization algorithm that simulates the social behavior of organisms [19, 20]. In PSO, each particle in the search space can be considered “an individual bird of a flock”; it moves its position based on its own knowledge and that of its neighbors. In other words, each particle uses its own memory and the knowledge of neighbors to find the best position (solution). In PSO, pbest is the best position of a particle amongst its own past iterations, as expressed by the highest fitness value. The best fitness value amongst all individual pbest values is called the global best (gbest). In each generation, the position and velocity of each particle is updated based on its own pbest and gbest. The update equations are shown below:

$$v_{id}^{new} = w \times v_{id}^{old} + c_1 \times r_1 \times (pbest_{id} - x_{id}^{old}) + c_2 \times r_2 \times (gbest - x_{id}^{old}) \quad (1)$$

$$x_{id}^{new} = x_{id}^{old} + v_{id}^{new} \quad (2)$$

where r_1 and r_2 are random numbers between (0, 1), and c_1 and c_2 are acceleration constants set to 2. Velocities v_{id}^{new} and v_{id}^{old} denote the velocities of the new particle and the old particle, respectively. The positions x_{id}^{new} and x_{id}^{old} are the updated particle position and the current particle position,

respectively.

An inertia weight w is used to control the balance between the global and local search. This weight is updated by the following equation:

$$w = (w_{max} - w_{min}) \times \frac{move_{max} - move_i}{move_{max}} + w_{min} \quad (3)$$

where w_{max} and w_{min} are set to 0.9 and 0.4, respectively. $move_i$ and $move_{max}$ represent the current iteration number and the total number of iterations, respectively [20].

Pseudo-code for PSO

1. Begin
2. randomly initialize particles swarm
3. while (the stopping criterion is not met)
4. evaluate fitness of particles
5. for $n = 1$ to number of particles
6. find $pbest$
7. find $gbest$
8. for $d=1$ to number of dimension of particle
9. update the position of particles by Eq. (1)-(2)
10. next d
11. next n
12. update the inertia weight value by Eq.(3)
13. next generation until stopping criterion is met
14. End

C. Chaos

Chaos is greatly sensitive to its initial conditions. Small differences in initial conditions yield widely diverging outcomes (“butterfly effect”), which makes long-term predictions impossible [26]. Chaos is a deterministic, random process found in non-linear systems that are non-periodical and bounded [23]. Although it is perfectly deterministic in principle, its behavior is completely unpredictable in practice. In recent years, chaotic sequences have been widely applied in many fields, such as the chaotic optimization algorithm [27] and DNA computing [28]. Chaotic PSO not only enhances the multiplicity of particles, but also avoids particle trapping in a local optimum. Chaos can be de-scribed as a bounded nonlinear system with deterministic dynamic behavior that has ergodic and stochastic properties [23]. It is very sensitive to the initial conditions and the parameters used. In other word, cause and effect of chaos are not proportional to the small differences in the initial values.

D. Chaotic Particle Swarm Optimization, CPSO

In a PSO search process, each particle from a swarm represents a candidate solution. The individual best value ($pbest_i$) is the position of the i -th particle with the highest fitness at a given iteration; the best position of all $pbest$ value is called $gbest$.

Changes in the acceleration coefficients c_1 and c_2 may pull the particles closer to the powered middle points of both $pbest_i$ and $gbest_i$. Under original PSO conditions, the method easily gets trapped in a local search. If $c_1=0$, a particle would lose its own recognition ability, which means particles would pool around $gbest_i$ more quickly. However, for large-scale problems, an algorithm with $c_1=0$ would be much more likely to run into a local optimal solution. If $c_2=0$, particles would be deprived of their social cooperation ability, which means

that particles can not share the optimal solution with the swarm but keep moving along their individual paths [29]. We therefore added a chaos map to improve the performance of PSO.

Chaotic particle swarm optimization (CPSO) adopts the use of chaotic maps to strengthen the solution quality of PSO. This increases the search capability of PSO. Since logistic maps are frequently used chaotic behavior maps and chaotic sequences can be quickly generated and easily stored, there is no need for storage of long sequences [30]. The parameters r_1 and r_2 are modified by the logistic map based on the following equation.

$$Cr_{n+1} = k \times Cr_n \times (1 - Cr_n) \quad Cr_0 \neq \{0, 0.25, 0.5, 0.75, 1\} \quad (4)$$

In Eq. (4), $Cr_{(0)}$ is generated randomly for each independent run, with $Cr_{(0)}$ not being equal to $\min\{0, 0.25, 0.5, 0.75, 1\}$ and k equal to 4. The driving parameter k of the logistic map controls the behavior of $Cr(t)$.

The velocity update equation for CPSO can thus be formulated as:

$$v_{id}^{new} = w \times v_{id}^{old} + c_1 \times Cr \times (pbest_{id} - x_{id}^{old}) + c_2 \times (1 - Cr) \times (gbest_d - x_{id}^{old}) \quad (5)$$

In Eq. (5), Cr is a function based on the results of the logistic map with values between 0.0 and 1.0.

Pseudo-code for CPSO

1. Begin
2. randomly initialize particles swarm
3. while (the stopping criterion is not met)
4. evaluate fitness of particles
5. for $n = 1$ to number of particles
6. find $pbest$
7. find $gbest$
8. for $d=1$ to number of dimension of particle
9. update the position of particles by Eq. (1)-(2)
10. next d
11. next n
12. update the inertia weight value by Eq. (5)
13. if fitness of $gbest$ is the same five times then
14. randomly select a half of particles swarm S
15. generate new particles C by Eq. (4) and replace S
16. end if
17. next generation until stopping criterion
18. End

E. Encoding

The particle encoding is given by:

$$P_i = (SNP_{i,j}, Genotype_{i,j}), \quad i=1, 2, \dots, x, \quad j=1, 2, \dots, y$$

$SNP_{i,j}$ represents an SNP that can be selected $Genotype_{i,j}$ represents the different possible genotype state selected (three states), and F_i represents the particle fitness value. In this study, the initial particles are randomly generated.

For example, let $P = (SNP_{1,5}, Genotype_{3,3})$. SNP_1 and SNP_2 are chosen in this particle. SNP_{12} and $Genotype_{21}$ represent SNP_1 with the second genotype, and SNP_2 with the first genotype; the particle fitness value is 15.

F. Initialization

Initial particles are randomly generated in this study. For example, a $P = (SNP_{1,5}, Genotype_{3,3})$ is given. SNP_1 and SNP_5 are chosen in this particle. The $SNP_{1,5}$ and $Genotype_{3,3}$ represent the SNP_1 with the third genotype, and the SNP_5 with the third genotype. The results are represented as SNP_1 and SNP_5 with the genotypes for [TNFa-857-CC] and [PTH(BstB I)-AA], respectively.

G. Fitness Evaluation

In this study, the maximum SNP combination difference between case and control groups is calculated. We divided the fitness calculation into two separate steps. First, the total numbers of SNP combinations in the control dataset and in the case dataset are calculated. Then Eq. (5) is used to determine the fitness value of each particle. The respective equation is shown below:

$$F(P_i) = \left| \sum_{i=1}^N (All_control) - \sum_{i=1}^N (All_case) \right| \quad (6)$$

where N represents the total number of combinations, $All_control$ represents the total number of SNP interactions in the control group, and All_case represents the total number of SNP interactions in the case group.

H. Parameter Settings

Four different parameters need to be set in CPSO: the population size, the number of iterations, and the acceleration constants c_1 and c_2 of the update function. The population size in our study was set to 50, the number of iterations was set to 100, and c_1 and c_2 were set to 2 [20].

I. Experimental Environment

The proposed algorithm was run on an Intel(R) CPU, 2.5G Hz, 3.24 GB RAM, Microsoft Windows XP and jdk1.6.0_071.4.0 platform.

J. Performance Measuremen

This study uses four common criteria to determine the prediction score [15]. The four criteria are shown in detail in Table II.

TABLE II.
PREDICTION CRITERIA

Description	Formula
Correct	$\frac{TP + TN}{TP + FN + FP + TN}$ (7)
Sensitivity	$\frac{TP}{TP + FN}$ (8)
Specificity	$\frac{TN}{FP + TN}$ (9)
Odds Ratio	$\frac{TP * TN}{FP * FN}$ (10)

TP: true positives, TN: true negatives, FN : false negatives, FP false positives.

K. Statistics

The CPSO algorithm generates SNP an odds ratio and a 95% confidence interval (CI) to evaluate each SNP

combination. Statistical analysis was carried out using <http://statpages.org/ctab2x2.html>.

L. Odds ratios

In recent years odds ratios (*OR*) have become widely used in epidemiology. *OR* is a common statistic expressing the strength of association between an exposure and a disease. Odds ratios allow convenient interpretation of case-control studies [31]. The odds ratio can reveal information regarding the effect of a certain genotype combination on the disease risk, since the quantitative value of the odds ratio represents the strength of association between the genotypes and disease.

III. RESULT AND DISCUSSION

Many studies have suggested that SNP-SNP interactions (epistasis) are very important [32, 33] because they can be helpful in investigating cancer and other types of disease [34, 35, 36, 37]. At present, artificial intelligent algorithms are seldomly used to identify combinations of SNP-SNP interactions. While the literature has raised the prospects of SVM [14], MDR [13] and machine learning [17], there is room for improvement, especially with regard to a general lack of quantitative measurements. The proposed CPSO algorithm can easily identify SNP-SNP interactions and evaluate the BMD risk and the maximum difference between the case and control groups.

When exhaustive search (ES) is used to calculate the interaction of SNPs, a fixed number of SNPs is used to find the optimal SNP interaction solution. The number of possible interactions is $C(N,M) * 3^M = N! / [M! (N-M)!] * 3^M$, where N is the number of SNPs or factors, and M is the selected prediction number of SNPs used to calculate the fitness value. Using the ES algorithm to calculate a large dataset requires a large amount of time. This time can be shortened with CPSO as it reduces the number of search items among a greater number of SNP combinations.

A. Identification of best SNP-SNP interaction combinations with maximal difference between cases and controls

We used CPSO to select the best combination of SNP-SNP interactions with a maximal difference between the case and control groups for BMD. Our analysis is based on 184 control and 113 case subjects with a BMD risk. The SNP name, number of cases and number of controls, as well as other information is show in Table I. The best SNP-SNP interactions of 2-SNP combinations are shown in Table III. In it, the two specific SNPs combinations with their corresponding genotypes, namely SNPs (1, 5) with genotype 3-3; [rs1799724-CC]- [rs6254-AA], showed the maximal difference, i.e., 62 between the *All_control* (39) and *All_case*

(101) groups. Under the same criteria, we also used CPSO to identify 3-9- SNP combinations with the best performance. Results are shown in Table IV. Experimental results showed that the proposed CPSO method can handle a combination of multiple SNPs well, i.e., CPSO provided the highest level of performance for complex SNP interaction.

B. Sensitivity, specificity, correlation coefficient, odds ratios and 95% CI ranks for osteoporosis

Table IV shows the proportion of subjects with osteoporosis, with specific SNP combinations. For combinations of two to nine SNPs, the odds ratio values are higher than 1 (1.550-2.309).

Simultaneously, the *CC* is between 0.391 to 0.589 in these SNP combinations. When the number of SNPs in the combinations was increased as described above, we found that combinations of two to nine SNPs all contained the two SNPs 1 and 5. This leads to the conclusion that the SNP (1, 5) combination in the dataset is the most important SNP combination.

C. Analyzing combinations of SNP (1,5) in osteoporosis

We analyzed the combination of SNP (1, 5) in osteoporosis. Related information is shown in Table V. In addition, A bar graph that illustrate the *OR* value is shown in Fig. 1. The data shows that SNP (1, 5) in Genotype 3-3 has a maximal difference between case and control values of 62 as determined by CPSO, and that the *p-values* (<0.05) are statistically significant. Although some other SNP combinations were also statistically significant, e.g., SNP(1,5) with genotype1-2, SNP(1,5) with genotype1-3 and SNP(1,5) with genotype2-3, their SNP combination risk was small than 1.

IV. CONCLUSION

Evaluating a large number of SNPs associated with a disease requires a strategy for focusing on only selected complex interactions. In this study, CPSO was successfully used on complex SNP interactions and was shown to provide the best SNP-SNP interactions for predicting osteoporosis susceptibility. Based on the CPSO method, the *OR* was used as a quantitative measure of the BMD risk. The method can potentially be applied to SNP-SNP interactions (epistasis) for other association studies. In the future, we plan to combine other methods, such as extremal optimization and fuzzy theory, to test SNP-SNP interactions in related study. We believe that the proposed method can serve as a computationally and statistically useful tool in the coming era of large-scale interaction mapping in genome-wide case-control studies.

TABLE III
ESTIMATED BEST COMBINATIONS OF TWO SNPs ON THE OCCURRENCE OF LOW BONE MASS DENSITY

Combined of two SNPs	SNP Genotype	Control no. /Case no.	Diff.	<i>SN</i>	<i>SP</i>	<i>CC</i>	<i>OR</i>	95% CI	<i>p</i> -value
SNPs (1,5)	Other 3-3	74/83 39/101	62	0.549	0.655	0.589	2.309	1.383-3.862	0.001
SNPs (1,8)	Other 3-2	75/87 38/97	59	0.527	0.664	0.579	2.201	1.316-3.687	0.002
SNPs (1,4)	Other 3-3	73/87 40/97	57	0.527	0.646	0.572	2.035	1.222-3.394	0.004

No. number, Diff.: Difference of control-breast cases, *SN*: sensitivity, *SP*: specificity, *CC*: CI: confidence interval,

“Other” is the reference group.

TABLE IV.
ESTIMATED BEST COMBINATIONS OF TWO TO NINE SNPs ON THE OCCURRENCE OF LOW BONE MASS DENSITY

Combined SNPs	SNP Genotype	Control no. /Case no.	SN	SP	CC	OR	95% CI	p-value
SNPs (1,5)	Other	74/83	0.549	0.655	0.589	2.309	1.383-3.862	0.001
	3-3	39/101						
SNPs (1,4,5)	Other	83/106	0.424	0.735	0.542	2.036	1.187-3.503	0.006
	3-3-3	30/78						
SNPs (1,4,5,8)	Other	86/119	0.353	0.761	0.508	1.740	0.995-3.054	0.040
	3-3-3-2	27/65						
SNPs (1,4,5,8,9)	Other	96/139	0.245	0.850	0.475	1.828	0.950-3.549	0.057
	3-3-3-2-1	17/45						
SNPs (1,4,5,6,8,9)	Other	100/155	0.158	0.885	0.434	1.439	0.680-3.081	0.391
	3-3-3-3-2-1	13/29						
SNPs (1,4,5,6,7,8,9)	Other	107/167	0.092	0.947	0.418	1.815	0.647-5.336	0.268
	3-3-3-3-3-2-1	6/17						
SNPs (1,3,4,5,6,7,8,9)	Other	109/174	0.054	0.965	0.401	1.566	0.438-6.092	0.578
	3-3-3-3-3-3-2-1	4/10						
SNPs (1,2,3,4,5,6,7,8,9)	Other	111/179	0.027	0.982	0.391	1.550	0.261-11.749	0.713
	3-2-3-3-3-3-3-2-1	2/5						

No. number, Diff.: Difference of control-breast cases, SN: sensitivity, SP: specificity, CC: Correct, CI: confidence interval; "Other" is the reference group.*
Difference of control-breast cases

TABLE V.
ODDS RATIO (OR) FOR SNP INTERACTIONS (95% CI) IN SNP (1, 5)
COMBINATIONS

Odds ratio*¹ of the SNPs (1,5) pair*²

TNF α -857 PTH(BstB I)	TT	CT	CC
GG OR CI p-value Ca./Co.	N.E	0.612 0.017-22.621 1.000 1/1	1.856 0.170-46.899 1.000 3/1
AG OR CI p-value Ca./Co.	0.072 0.003-0.575 0.002 1/8	0.982 0.282-3.555 1.000 8/5	1.084 0.484-2.458 1.000 21/12
AA OR CI p-value Ca./Co.	0.361 0.165-0.784 0.005 14-21	0.758 0.411-1.402 0.373 34/26	2.309 1.383-3.862 0.001 101/39

*¹The OR of the reference group is 1.000. *² SNPs (1,5) pair with its corresponding genotype is TNF α -857-rs1799724 and PTH (BstB I)-rs6254. Ca.: number of cases. Co.: number of controls.

REFERENCES

[1] L. Hellmeyer, M. Kuhnert, V. Ziller, S. Schmidt, and P. Hadji, "The use of iv bisphosphonate in pregnancy-associated osteoporosis-case study," *Experimental and Clinical Endocrinology and Diabetes*, vol. 115, pp. 139-142, 2007.

[2] C. Chen, A. David, H. Nunnerley, M. Michell, J. Dawson, H. Berry, J. Dobbs, and T. Fahy, "Adverse life events and breast cancer: case-control study," *BMJ*, vol. 311, p. 1527, 1995.

[3] V. U. Onay, L. Briollais, J. A. Knight, E. Shi, Y. Wang, S. Wells, H. Li, I. Rajendram, I. L. Andrulis, and H. Ozelik, "SNP-SNP interactions in breast cancer susceptibility," *BMC Cancer*, vol. 6, p. 114, 2006.

[4] E. L. Franco, L. P. Kowalski, B. V. Oliveira, M. P. Curado, R. N. Pereira, M. E. Silva, A. S. Fava, and H. Torloni, "Risk factors for oral cancer in Brazil: A case control study," *International Journal of Cancer*, vol. 43, pp. 992-1000, 1989.

[5] C. D. Llewellyn, K. Linklater, J. Bell, N. W. Johnson, and S. Warnakulasuriya, "An analysis of risk factors for oral cancer in young people: a case-control study," *Oral oncology*, vol. 40, pp. 304-313, 2004.

[6] M. Garcia Magarinos, I. Lopez de Ullibarri, R. Cao, and A. Salas, "Evaluating the ability of tree based methods and logistic regression for

the detection of SNP SNP interaction," *Annals of human genetics*, vol. 73, pp. 360-369, 2009.

[7] P. C. Phillips, "Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems," *Nature Reviews Genetics*, vol. 9, pp. 855-867, 2008.

[8] D. Kang, "Genetic polymorphisms and cancer susceptibility of breast cancer in Korean women," *Journal of biochemistry and molecular biology*, vol. 36, pp. 28-34, 2003.

[9] X. Wan, C. Yang, Q. Yang, H. Xue, X. Fan, N. L. S. Tang, and W. Yu, "BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies," *The American Journal of Human Genetics*, vol. 87, pp. 325-340, 2010.

[10] F. S. Collins, M. S. Guyer, and A. Chakravarti, "Variations on a theme: cataloging human DNA sequence variation," *Science*, vol. 278, p. 1580, 1997.

[11] A. Chakravarti, "It's raining SNPs, hallelujah?," *Nature genetics*, vol. 19, pp. 216-217, 1998.

[12] R. Mehrian-Shai and J. K. V. Reichardt, "A renaissance of," *Molecular genetics and metabolism*, vol. 83, pp. 47-50, 2004.

[13] L. W. Hahn, M. D. Ritchie, and J. H. Moore, "Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions," *Bioinformatics*, vol. 19, p. 376, 2003.

[14] S. H. Chen, J. Sun, L. Dimitrov, A. R. Turner, T. S. Adams, D. A. Meyers, B. L. Chang, S. L. Zheng, H. Gronberg, and J. Xu, "A support vector machine approach for detecting gene gene interaction," *Genetic epidemiology*, vol. 32, pp. 152-167, 2008.

[15] 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, p. 146, 2008.

[16] D. F. Schwarz, I. R. Konig, and A. Ziegler, "On safari to Random Jungle: a fast implementation of Random Forests for high-dimensional data," *Bioinformatics*, vol. 26, p. 1752, 2010.

[17] B. A. McKinney, D. M. Reif, M. D. Ritchie, and J. H. Moore, "Machine learning for detecting gene-gene interactions: a review," *Applied bioinformatics*, vol. 5, pp. 77-88, 2006.

[18] Y. Chung, S. Y. Lee, R. C. Elston, and T. Park, "Odds ratio based multifactor-dimensionality reduction method for detecting gene-gene interactions," *Bioinformatics*, vol. 23, p. 71, 2006.

[19] J. Kennedy and R. Eberhart, "Particle swarm optimization," 1995, pp. 1942-1948.

[20] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intelligence*, vol. 1, pp. 33-57, 2007.

- [21] L. Y. Chuang, J. H. Tsai, and C. H. Yang, "Binary particle swarm optimization for operon prediction," *Nucleic acids research*, vol. 38, p. e128, 2010.
- [22] L.-Y. Chuang, H.-C. Huang, M.-C. Lin, and C.-H. Yang, "Particle Swarm Optimization with Reinforcement Learning for the Prediction of CpG Islands in the Human Genome," *PLoS One*, vol. 6, p. e21036, 2011.
- [23] H. G. Schuster and W. Just, *Deterministic chaos*: Wiley Online Library, 1988.
- [24] J. M. Bland and D. G. Altman, "Statistics notes: the odds ratio," *BMJ: British Medical Journal*, vol. 320, p. 1468, 2000.
- [25] G. Lin, H. Tseng, C. Chang, L. Chuang, C. Liu, C. Yang, C. Tu, E. Wang, and H. Tan, "SNP combinations in chromosome-wide genes are associated with bone mineral density in Taiwanese women," *Chinese Journal of Physiology*, vol. 51, p. 32, 2008.
- [26] S. H. Kellert, *In the wake of chaos: Unpredictable order in dynamical systems*: University of Chicago press, 1993.
- [27] B. L. W. Jiang, "Optimizing complex functions by chaos search," *Cybernetics and Systems*, vol. 29, pp. 409-419, 1998.
- [28] G. Manganaro and J. Pineda de Gyvez, "DNA computing based on chaos," 1997, pp. 255-260.
- [29] Z. Yuan, L. Yang, Y. Wu, L. Liao, and G. Li, "Chaotic particle swarm optimization algorithm for traveling salesman problem," 2007, pp. 1121-1124.
- [30] H. Gao, Y. Zhang, S. Liang, and D. Li, "A new chaotic algorithm for image encryption," *Chaos, Solitons & Fractals*, vol. 29, pp. 393-399, 2006.
- [31] J. M. Bland and D. G. Altman, "The odds ratio," *BMJ*, vol. 320, p. 1468, 2000.
- [32] A. Assawamakin, N. Chaiyaratana, C. Limwongse, S. Sinsomros, P. T. Yenchitsomanus, and P. Youngkong, "Variable-length haplotype construction for gene gene interaction studies," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 28, pp. 25-31, 2009.
- [33] V. Andersen, M. Ostergaard, J. Christensen, K. Overvad, A. Tjonneland, and U. Vogel, "Polymorphisms in the xenobiotic transporter Multidrug Resistance 1 (MDR 1) and interaction with meat intake in relation to risk of colorectal cancer in a Danish prospective case-cohort study," *BMC Cancer*, vol. 9, p. 407, 2009.
- [34] D. Franchimont, S. Vermeire, H. El Housni, M. Pierik, K. Van Steen, T. Gustot, E. Quertinmont, M. Abramowicz, A. Van Gossum, and J. Deviere, "Deficient host-bacteria interactions in inflammatory bowel disease? The toll-like receptor (TLR)-4 Asp299gly polymorphism is associated with Crohn's disease and ulcerative colitis," *Gut*, vol. 53, p. 987, 2004.
- [35] J. E. Goodman, L. E. Mechanic, B. T. Luke, S. Ambs, S. Chanock, and C. C. Harris, "Exploring SNP SNP interactions and colon cancer risk using polymorphism interaction analysis," *International Journal of Cancer*, vol. 118, pp. 1790-1797, 2006.
- [36] C. S. Carlson, M. A. Eberle, L. Kruglyak, and D. A. Nickerson, "Mapping complex disease loci in whole-genome association studies," *Nature*, vol. 429, pp. 446-452, 2004.
- [37] D. K. Amre, D. R. Mack, K. Morgan, A. Krupoves, I. Costea, P. Lambrette, G. Grimard, J. Dong, H. Feguery, and V. Bucionis, "Autophagy gene ATG16L1 but not IRGM is associated with Crohn's disease in Canadian children," *Inflammatory bowel diseases*, vol. 15, pp. 501-507, 2009.