Integrate Local Search to Genetic Algorithm for Screening Confronting Two-Pair Primers

Yu-Huei Cheng, *Member, IAENG*,
Hsueh-Wei Chang, Li-Yeh Chuang, and Cheng-Hong Yang, *Member, IAENG*

***Abstract*—Single nucleotide polymorphisms (SNPs) are abundant and important genetic variations in homo species. Polymerase chain reaction (PCR)-based technologies are usually applied to genotype SNPs. In the recent years, an economical and efficient approach that based on PCR was proposed to genotype SNPs. We called it "polymerase chain reaction with confronting two-pair primers (PCR-CTPP)". PCR-CTPP is a simple, time- and cost-saving SNP genotyping technology. However, computation of feasible confronting two-pair primers is elaborate, since many factors must be considered simultaneously. The important factor is considered as melting temperatures among primers. In this study, we propose a method that integrates local search to genetic algorithm (GA) for screening better confronting two-pair primers. SNP IDs of rs6352, rs6353, rs6354, rs6355, and rs25528 with 500 bps flanking sequence in the SLC6A4 gene are used as test template sequences to screen confronting two-pair primers. The results were compared with the original GA method, and we get the conclusion that the local search mechanism is helpful for screening confronting two-pair primers.**

***Index Terms*—Single nucleotide polymorphism (SNP); polymerase chain reaction with confronting two-pair primers (PCR-CTPP); genetic algorithm (GA); local search**

I. Introduction

Single nucleotide polymorphisms (SNPs) are the common and abundant genetic variations that can be a biomarker to help us develop diseases and respond to drugs, pathogens, vaccines, chemicals, and other agents. SNP genotyping is an important method to identify SNPs for association studies of diseases and cancers. SNP array [1] and real-time PCR using TaqMan probes [2] have been introduced to perform high-throughput analysis. However, these methods take much cost and ideally suited for large-scale genotyping. For small-scale genotyping, PCR-restriction fragment length polymorphism (RFLP) [3-5] is preferable. Although PCR-RFLP is inexpensive genotyping method, the long gestion period (usually 2-3 hours) of the restriction enzymes makes it be short of efficiency.

Recently, PCR with confronting two-pair primers (PCR-CTPP) technique was developed [6, 7] and many SNPs had been successfully genotyped by this approach, such as interleukin-1B (IL-1B) C-31T, interleukin-2 (IL-2) -330G, beta2-adrenergic receptor (beta2-AR) Gln27Glu, and aldehyde dehydrogenase 2 (ALDH2) [8], pylori-induced MAstric atrophy [9], severe coronary artery disease [10], and esophageal cancer risk [11]. However, the lack of computational methods makes screen feasible confronting two-pair primers is difficult, since many factors must be considered simultaneously. Especially is to find the small difference in melting temperature ($T_{\text{m-diff}}$) between primers [7].

In 2010, a genetic algorithm (GA) was proposed to screen confronting two-pair primers that correspond to the all kinds of primer constraints [12, 13]. However, the designed T_{m} differences were not ideal in most cases, and thus here we introduce the local search mechanism [14] to improve the GA for screening confronting two-pair primers.

II. Methods

A. Definition of the PCR-CTPP problem

First, we define the template sequence T_D, which is composed of nucleotide codes with an identified SNP. T_D is equated by:

$$T_D = \{B_i \mid i \text{ is the index of DNA sequence}, \; 1 \le i \le T_l, \; \exists! B_i \in \text{IUPAC code of SNP}\} \quad (1)$$

where B_i is the regular nucleotide (A, T, C, or G) or IUPAC code for the target SNP (M, R, W, S, Y, K, V, H, D, B or N); T_l is the length of template sequence, and the symbol $\exists!$ represents the existence and uniqueness.

Confronting two-pair primers include two short sequences named as forward primer 1 (P_{f1}) and reverse primer 2 (P_{r2}) which are picked from T_D unrelated to the target SNP, as well as two short sequences named as forward primer 2 (P_{f2}) and reverse primer 1 (P_{r1}) which are picked from T_D containing a target SNP as illustrated in Fig. 1. The forward primer 1 (P_{f1}) is a short sense sequence in the upstream (5' end) of the template sequence. The reverse primer 1 (P_{r1}) is a short antisense sequence, which contains the target SNP (the minor

Manuscript received December 25, 2012. This work was supported in part by the National Science Council in Taiwan under grants NSC101-2221-E-464-001-.

Yu-Huei Cheng is with the Department of Digital Content Design and Management, Toko University, Chiayi, Taiwan (e-mail: yuhuei.cheng@gmail.com).

Hsueh-Wei Chang is with the Faculty of Biomedical Science and Environmental Biology, Kaohsiung Medical University, Kaohsiung, Taiwan, (e-mail: changhw@kmu.edu.tw).

Li-Yeh Chuang is with the Department of Chemical Engineering & Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung, Taiwan (email: chuang@isu.edu.tw)

Cheng-Hong Yang is with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan (e-mail: chyang@cc.kuas.edu.tw).

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

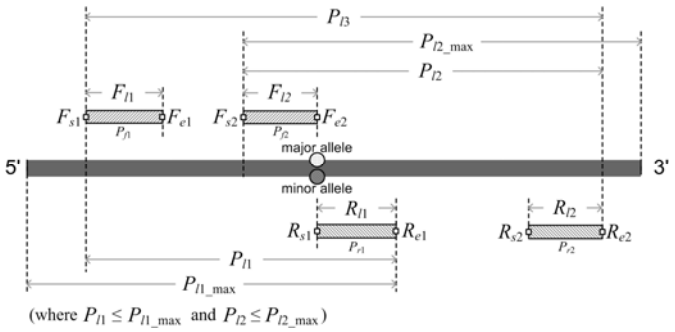

Fig. 1. Parameters for the confronting two-pair primers. Symbols present *F*: forward primer; *R*: reverse primer; *s*: start index position; *e*: end index position; *P*: length of PCR product; *l*: length of primer or product; P_{l1_max}: length from the start of template sequence to R_{e1}; P_{l2_max}: length from F_{s2} to the end of template sequence.

allele of the target SNP) located at its 5' end. The forward primer 2 (P_{f2}) is a short sense sequence, which contains the target SNP (the major allele of the target SNP) located at its 3' end. The reverse primer 2 (P_{r2}) is the antisense sequence in the downstream of the template sequence. These primers are defined as follows:

$$P_{f1} = \{B_i \mid i \text{ is the index of } T_D,\ F_{s1} \le i \le F_{e1}\} \quad (2)$$

$$P_{r1} = \{\overline{B_i} \mid i \text{ is the index of } T_D,\ R_{s1} \le i \le R_{e1}\} \quad (3)$$

$$P_{f2} = \{B_i \mid i \text{ is the index of } T_D,\ F_{s2} \le i \le F_{e2}\} \quad (4)$$

$$P_{r2} = \{\overline{B_i} \mid i \text{ is the index of } T_D,\ R_{s2} \le i \le R_{e2}\} \quad (5)$$

where P_{f1} and P_{r1}, as well as P_{f2} and P_{r2} are two primer pairs, i.e., the confronting two-pair primers. F_{s1} and R_{s1} represent the start index of P_{f1} and P_{r1} in T_D, respectively. F_{e1} and R_{e1} represent the end index of P_{f1} and P_{r1} in T_D, respectively. F_{s2} and R_{s2} represent the start index of P_{f2} and P_{r2} in T_D, respectively. F_{e2} and R_{e2} represent the start index of P_{f2} and P_{r2} in T_D, respectively. $\overline{B_i}$ is the complementary nucleotide of *B*, i.e., B_i = 'A', $\rightarrow$ $\overline{B_i}$ = 'T'; B_i = 'C' $\rightarrow$ $\overline{B_i}$ = 'G', and vice versa.

The target SNP is placed at the 3' end position of P_{f2}, and the 5' end position of P_{r1} (see Fig. 1). A vector P_v (called individual in the GA) with six elements, i.e., F_{l1}, P_{l1}, R_{l1}, F_{l2}, P_{l2} and R_{l2}, is used to represent an solution of confronting two-pair primers. We define the vector as follows:

$$P_v = (F_{l1}, P_{l1}, R_{l1}, F_{l2}, P_{l2}, R_{l2}) \quad (6)$$

The Elements of F_{l1}, P_{l1}, R_{l1}, F_{l2}, P_{l2} and R_{l2} respective represent the length of the forward primer 1, the product length between P_{f1} and P_{r1}, the reverse primer 1, the forward primer 2, the product length between P_{f2} and P_{r2}, as well as the reverse primer 2. Accordingly, the forward and the reverse primers can be retrieved from P_v.

B. Definition of fitness function

The common and necessary primer constraints are included in the fitness function with the value minimized. The fitness function is defined as follows:

$$\begin{aligned} Fitness(P_v) = {} & 3 \times (Len_{diff}(P_v) + GC_{proportion}(P_v) + GC_{clamp}(P_v)) + \\ & 10 \times (di\,mer(P_v) + hairpin(P_v) + specificity(P_v)) + \\ & 50 \times (\mathrm{Tm}(P_v) + \mathrm{Tm}_{diff}(P_v)) + 100 \times Avg\_\mathrm{Tm}_{diff}(P_v) + \\ & 60 \times PCR\mathrm{len}_{ratio}(P_v) \end{aligned} \quad (7)$$

We use the same fitness function as the published literature [13] to estimate the confronting two-pair primers. The fitness function had been successfully validated by PCR-CTPP experiments [13].

The primer length between 16 and 28 nt (nucleotides) is considered better for a PCR experiment. We do not consider the primer length fuction to the fitness function, but generate random values from 16 to 28 nt to F_{l1}, R_{l1}, F_{l2} and R_{l2}. The $Len_{diff}(P_v)$ is used to estimate primer length difference if less than or equal to 3 nt for primers of F_{l1} and R_{l1}, primers of F_{l2} and R_{l2}, as well as primers of F_{l1} and R_{l2}. The $GC_{proportion}(P_v)$ is used to estimate the proportion of nucleotides G and C in primers, and we limit it to between 20% and 80%. The $GC_{clamp}(P_v)$ is used to account for the presence of nucleotide G or C at the 3' terminal of a primer.

In this study, the T_m calculation formula is taken from Bolton and McCarthy [15], and it is given by:

$$\mathrm{Tm}_{BM}(P) = 81.5 + 16.6 \times (\log_{10}[Na^+]) + 0.41 \times (GC\%) - 675/|P| \quad (8)$$

where *P* represents a primer, and $|P|$ represents the length of primer *P*; Na^+ is the molar salt concentration.

The $\mathrm{Tm}(P_v)$ is used to evaluate T_m of a primer to range between 45°C and 62°C. The $\mathrm{Tm}_{diff}(P_v)$ converge the T_m difference is less than or equal to 1°C. In order to balance the T_m values among the confronting two-pair primers, the $Avg\_\mathrm{Tm}_{diff}(P_v)$ function is used to calculate the average T_m difference. The $dimer(P_v)$ is used to check for dimer conditions. The $hairpin(P_v)$ is used to avoid annealing itself. The $specificity(P_v)$ is used to check for reappearance of each primer in the template DNA sequence. Finally, the $PCR\mathrm{len}_{ratio}(P_v)$ is used to calculate the appropriate lengths of the PCR products. Three lengths of PCR products of P_{l1}, P_{l2} and P_{l3} must be greater than 100 bps.

C. The proposed Method

Six separate processes are involved: (1) initialization of population, (2) local search, (3) evaluation of fitness, (4) operations of selection, crossover, and mutation, (5) replacement, and (6) judgment of terminational conditions.

(1) Initialization of population. A particular number of individuals $P_v = (F_{l1}, P_{l1}, R_{l1}, F_{l2}, P_{l2}, R_{l2})$ are randomly generated as an initial population without duplicates. The method randomly generates F_{l1}, R_{l1}, F_{l2} and R_{l2} between 16 and 28 nt, and the PCR product lengths P_{l1} and P_{l2} between 100 bp and P_{l1_max}, as well as between 100 bp and P_{l2_max}, respectively. (P_{l1_max} and P_{l2_max} are the maximum tolerant PCR product lengths of P_{l1} and P_{l2} , respectively)

(2) Local search. In order to improve the shortcoming of the GA that easily traps local optimum, we employ a local search mechanism. The local search can identify superior individuals from amongst the neighbors of the original

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

individual, thus the experience of the original individual is improved. All individuals are first performed by the local search mechanism so that local optimum of each individual is achieved. Each iteration, all new individuals are also performed by the local search mechanism so that local optimum is always preserved. Finally, a global optimum is found. The pseudo-code of the local search is sketched below:

The pseudo-code of the local search

```
1  Begin;
2      Select an incremental value d=a*Rand();
3      For a given individual i∈P: calculate fitness (i);
4      For j=1 to the number of variables in individual i;
5          value(j)= value(j)+d;
6          If fitness of the individual is not improved then
7              value(j)= value(j)-d;
8          else if fitness of the individual is improved then
9              retain value(j);
10     Next j;
11 End;
```

In the above pseudo-code, P represents a population, and d is a incremental value used to assist an individual in seeking out neighboring information. Before determining d, a which is a constant that suits the variable values must be set. We use F_{l1}, R_{l1}, F_{l2} and R_{l2} to evaluate the constant a. Consequently, the constant a can be calculated by the difference of the maximum primer length and the minimum primer length (28-16=12).

(3) Evaluation of fitness. The fitness value which is used to judge whether an individual is good or bad is evaluated by the fitness function. We use formula (7) [13] to evaluate the fitness values of all individuals in the population and new individuals.

(4) Operations of selection, crossover, and mutation. The computational processes include selection, crossover and mutation. The crossover and mutation operation are alternative according to the same probability. We select two individuals randomly from the population to update. When the crossover rate is reached, the selected two individuals are updated by the uniform crossover operation. When the mutation rate is reached, one-point mutation is applied to the alternative individuals.

(5) Replacement. After the operations of selection, crossover, and mutation, the worst individuals in the population will be replaced by the new individuals.

(6) Judgment of terminational conditions. In the terminational conditions, when the preset maximum number of iterations is reached, the proposed method is terminated.

III. Results

A. Test template sequences

Five SNP IDs of rs6352, rs6353, rs6354, rs6355 and rs25528 which are picked from the SLC6A4 gene associated with autism spectrum disorders [16], psychosis [17], and bipolarity [18] are used as target SNPs. We use SNP-Flankplus (http://bio.kuas.edu.tw/snp-flankplus/) [19] to retrieve their flanking sequences with the length of 500 bps as test template sequences.

B. Parameter settings

Four parameters were set for the proposed method and the GA, i.e., the number of iterations, the population size, the crossover rate, and the mutation rate. We use DeJong and Spears' parameter settings [20]. Their values were respective 1000, 50, 0.6 and 0.001. For the primer constraints, we use the common and necessary factors, such as the primer length between 16nt and 28 nt, the primer length difference less than or equal to 3 nt, the GC% between 20% and 80%, the primer T_m between 45 °C and 62°C, the difference of primers T_m less than 1°C, and the product length larger than 100 bps. Furthermore, the PCR product length was set to the ratio of 8: 13: 20. These ratios were chosen based on PCR experiments previously conducted by us [12, 13].

C. Results for the GA and the proposed method

The results of single primer of the GA and the proposed method for five SNP IDs of rs6352, rs6353, rs6354, rs6355 and rs25528 are shown in Table 1 and Table 2, respectively. In the result of the GA (Table 1), all of the primer length, GC%, T_m, dimer, and specificity for the five target SNPs are conformed to the setting of primer constraints. The primers in P_{f2} and P_{r2} for rs6352, in P_{f1} and P_{f2} for rs6353, in P_{f2} for rs6354, in P_{r2} for rs6355, and in P_{f2} for rs25528 are not conformed to the GC clamp criterion. The primers in P_{r1} and P_{r2} for rs6355 are not conformed to the hairpin criterion. The fitness values are respective 429, 429, 413, 386, and 416 for the five target SNPs.

In the result of the proposed method (Table 2), all of the primer length, GC%, T_m, dimer, and specificity for the five target SNPs are conformed to the setting of primer constraints. The primers in P_{f2} and P_{r2} for rs6352, in P_{f2} for rs6353, in P_{f2} for rs6354, in P_{f1} and P_{r2} for rs6355, and in P_{f2} and P_{r2} for rs25528 are not conformed to the GC clamp criterion. The primers in P_{f2} for rs6355, and in P_{f1} and P_{r1} for rs6355 are not conformed to the hairpin criterion. The fitness values are respective 386, 410, 416, 223, and 19 for the five five template sequences.

The results of compound primers of the GA and the proposed method for five SNP IDs of rs6352, rs6353, rs6354, rs6355 and rs25528 are shown in Table 3 and Table 4, respectively. In the result of the GA (Table 3), the primers in P_{f1}-P_{r1} for rs6352, in P_{f2}-P_{r2} for rs6353, as well as in P_{f2}-P_{r2} and P_{f1}-P_{r2} are not conformed to the primer length difference criterion. The primers in P_{f1}-P_{r1} for rs6352, in P_{f2}-P_{r2} and P_{f1}-P_{r2} for rs6354, in P_{f2}-P_{r2} for rs6355, and in P_{f2}-P_{r2} for rs25528 are conformed to the difference of primers T_m. The PCR product lengths for the five target SNPs are all larger than 100 bps. Their positions can be observed in the field of PCR product position in Table 3.

In the result of the proposed method (Table 4), only the primers in P_{f2}-P_{r2} for rs6354 are not conformed to the primer length difference criterion. The primers in P_{f2}-P_{r2} for rs6352, in all compound primers P_{f1}-P_{r1}, P_{f2}-P_{r2} and P_{f1}-P_{r2} for rs6355, and in P_{f1}-P_{r1} and P_{f2}-P_{r2} for rs25528 are conformed

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

to the difference of primers T_m. The PCR product lengths for the five target SNPs are all larger than 100 bps. Their positions can be observed in the field of PCR product position in Table 4.

D. Comparison of the results for the GA method and the proposed method

The fitness value represents the integral evaluation of the result. From Table 1 and Table 2, we can observe the fitness values of the proposed method are all better than the GA, except the SNP ID rs6354. Furthermore, the result of the most important factor, i.e., the difference of primers T_m, in the proposed method is better than the result of the GA (as shown in Table 3 and Table 4).

IV. Discussion

In the paper, we use the local search mechanism to improve the GA method for screening confronting two-pair primers. The results show it is helpful to assist the finding of better solutions. In the following, we discuss the influence of the local search mechanism, and the influence of the melting temperature. Furthermore, we also discuss the primer constraints applied to PCR-CTPP.

A. The influence of the local search mechanism

The GA is comprehensively applied to many problems to look for their optimal solution, but the drawback that traps local optimum easily makes it be usually not dependable in some problems like this issue of primer screen. Therefore, we integrate the local search to the GA to improve it. The local search is performed to evaluate each element in the original individual and searches the better elements from amongst the neighbors of the original individual. Each iteration local optimum in each individual is got by the local search, and it is always kept to next iteration to find better local optimum until the global optimum is achieved. From the results (Table 1 and Table 2), we can observe all the fitness values of the proposed method are superior to the ones of the GA, except SNP ID rs6354. Although the fitness value of the proposed method for SNP ID rs6354 is worse than one of the GA, their values are very close (416 and 413 for the proposed method and the GA, respectively). The embedded local search mechanism is thus capable of improving the search power, and avoids traping into local optimum easily.

B. The influence of the melting temperature

In PCR-CTPP, the melting temperatures of the primers are very important. When the T_m value is close among the confronting two-pair primers, the PCR competition is balanced. When the melting temperature is low, the PCR reactions are non-specific, which generates incorrect heterozygous genotyping. Accordingly, PCR-CTPP with a competitive or specific amplification is important to correctly genotype SNPs. We can computationally resolve this problem by finding close T_m values of the confronting two-pair primers. The proposed method used in this study screens the confronting two-pair primers with better difference of T_m values than the GA (see the Table 3 and Table 4), especially in SNP ID rs6355 (the difference of T_m values are 0.54, 0.13 and 0.54 for P_{f1}-P_{r1}, P_{f2}-P_{r2}, and P_{f1}-P_{r2}, respectively). This shows that the proposed method is functional to find the difference of T_m values satisfied completely.

C. The primer constraints applied to PCR-CTPP

The T_m is important in PCR-CTPP, and we also suppose that other primer constraints might improve the screen of the confronting two-pair primers. This study investigates the common and necessary primer design constraints, such as the primer length, the difference of primer length, GC proportion, GC clamp, dimer, hairpin, PCR product length, and specificity. Table 1 and Table 2 show the results of screening the confronting two-pair primers based on the public primer design constraints.

V. Conclusions

PCR-CTPP is a simple, time- and cost-saving SNP genotyping technology. The melting temperatures among primers are very important factor considered in PCR-CTPP. However, it is not easy to find close T_m values of the confronting two-pair primers. Methods for the computation of feasible confronting two-pair primers are thus still demanded. In this study, we integrate the local search mechanism to the GA for screening better confronting two-pair primers. SNP IDs of rs6352, rs6353, rs6354, rs6355, and rs25528 with 500 bps flanking sequence in the SLC6A4 gene are used as test target SNPs to screen confronting two-pair primers. The proposed method provides more crucial melting temperatures for SNP ID rs6355. It yields the confronting two-pair primers better than the ones of the GA method according to the evaluation of their fitness values. In conclusion, the proposed method improves the GA in screening confronting two-pair primers.

Acknowledgment

This work is partly supported by the National Science Council (NSC) in Taiwan under grant NSC101-2221-E-464-001-.

References

[1] F. Jasmine, H. Ahsan, I. L. Andrulis, E. M. John, J. Chang-Claude, and M. G. Kibriya, "Whole-genome amplification enables accurate genotyping for microarray-based high-density single nucleotide polymorphism array," *Cancer Epidemiol Biomarkers Prev*, vol. 17, pp. 3499-508, Dec 2008.

[2] L. Hui, T. DelMonte, and K. Ranade, "Genotyping using the TaqMan assay," *Curr Protoc Hum Genet*, vol. Chapter 2, p. Unit 2 10, Jan 2008.

[3] H. W. Chang, C. H. Yang, P. L. Chang, Y. H. Cheng, and L. Y. Chuang, "SNP-RFLPing: restriction enzyme mining for SNPs in genomes," *BMC Genomics*, vol. 7, p. 30, 2006.

[4] L. Y. Chuang, C. H. Yang, K. H. Tsui, Y. H. Cheng, P. L. Chang, C. H. Wen, and H. W. Chang, "Restriction enzyme mining for SNPs in genomes," *Anticancer Res*, vol. 28, pp. 2001-7, Jul-Aug 2008.

[5] H. W. Chang, Y. H. Cheng, L. Y. Chuang, and C. H. Yang, "SNP-RFLPing 2: an updated and integrated PCR-RFLP tool for SNP genotyping," *BMC Bioinformatics*, vol. 11, p. 173, 2010.

[6] N. Hamajima, T. Saito, K. Matsuo, K. Kozaki, T. Takahashi, and K. Tajima, "Polymerase chain reaction with confronting two-pair primers for polymorphism genotyping," *Jpn J Cancer Res*, vol. 91, pp. 865-8, Sep 2000.

[7] N. Hamajima, T. Saito, K. Matsuo, and K. Tajima, "Competitive amplification and unspecific amplification in polymerase chain reaction with confronting two-pair primers," *J Mol Diagn*, vol. 4, pp. 103-7, May 2002.

[8] N. Katsuda, N. Hamajima, A. Tamakoshi, K. Wakai, K. Matsuo, T. Saito, K. Tajima, and S. Tominaga, "Helicobacter pylori seropositivity

and the myeloperoxidase G-463A polymorphism in combination with interleukin-1B C-31T in Japanese health checkup examinees," *Jpn J Clin Oncol*, vol. 33, pp. 192-7, Apr 2003.

[9] S. Togawa, T. Joh, M. Itoh, N. Katsuda, H. Ito, K. Matsuo, K. Tajima, and N. Hamajima, "Interleukin-2 gene polymorphisms associated with increased risk of gastric atrophy from Helicobacter pylori infection," *Helicobacter*, vol. 10, pp. 172-8, Jun 2005.

[10] K. K. Abu-Amero, O. M. Al-Boudari, G. H. Mohamed, and N. Dzimiri, "The Glu27 genotypes of the beta2-adrenergic receptor are predictors for severe coronary artery disease," *BMC Med Genet*, vol. 7, p. 31, 2006.

[11] S. J. Yang, H. Y. Wang, X. Q. Li, H. Z. Du, C. J. Zheng, H. G. Chen, X. Y. Mu, and C. X. Yang, "Genetic polymorphisms of ADH2 and ALDH2 association with esophageal cancer risk in southwest China," *World J Gastroenterol*, vol. 13, pp. 5760-4, Nov 21 2007.

[12] C. H. Yang, Y. H. Cheng, L. Y. Chuang, and H. W. Chang, "Genetic Algorithm for the Design of Confronting Two-Pair Primers," *Ninth IEEE international Conference on BioInformatics and BioEngineering (BIBE)*, pp. 242-247, 2009.

[13] C. H. Yang, Y. H. Cheng, L. Y. Chuang, and H. W. Chang, "Confronting two-pair primer design for enzyme-free SNP genotyping based on a genetic algorithm," *BMC Bioinformatics*, vol. 11, p. 509, 2010.

[14] P. Moscato, "On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms," Tech. Rep. Caltech Concurrent Computation Program, Rep. 826, California Inst. Technol., Pasadena, CA1989.

[15] J. Sambrook, E. F. Fritsch, and T. Maniatis, *Molecular cloning*: Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY, 1989.

[16] T. Sakurai, J. Reichert, E. J. Hoffman, G. Cai, H. B. Jones, M. Faham, and J. D. Buxbaum, "A large-scale screen for coding variants in SERT/SLC6A4 in autism spectrum disorders," *Autism Res*, vol. 1, pp. 251-7, Aug 2008.

[17] T. E. Goldberg, R. Kotov, A. T. Lee, P. K. Gregersen, T. Lencz, E. Bromet, and A. K. Malhotra, "The serotonin transporter gene and disease modification in psychosis: Evidence for systematic differences in allelic directionality at the 5-HTTLPR locus," *Schizophr Res*, vol. 111, pp. 103-8, Jun 2009.

[18] L. Mandelli, M. Mazza, G. Martinotti, M. Di Nicola, T. Daniela, E. Colombo, S. Missaglia, D. De Ronchi, R. Colombo, L. Janiri, and A. Serretti, "Harm avoidance moderates the influence of serotonin transporter gene variants on treatment outcome in bipolar patients," *J Affect Disord*, p. Epub ahead of print, Mar 28 2009.

[19] C. H. Yang, Y. H. Cheng, L. Y. Chuang, and H. W. Chang, "SNP-Flankplus: SNP ID-centric retrieval for SNP flanking sequences," *Bioinformation*, vol. 3, pp. 147-9, 2008.

[20] K. A. De Jong and W. M. Spears, "An analysis of the interacting roles of population size and crossover in genetic algorithms," vol. 1, ed: Springer, 1990, pp. 38–47.

TABLE I
THE RESULT OF SINGLE PRIMER OF THE GA METHOD FOR FIVE SNP IDS

			Primer Constraints							
SNP ID	***Primer ID***	***Primers (position)***	***primer length (nt)***	***GC%***	***GC clamp***	***Tm (°C)***	***dimer***	***hairpin***	***specificity***	***fitness***
Rs6352	P_{f1}	CTGATGAGCCCGCCACAACTAC (354-375)	22	59.09	Yes	54.13	No	No	Yes	429
	P_{r1}	ATATACACACTAACTAGCACGTACCTCG (500-527)	28	42.86	Yes	53.47	No	No	Yes	
	P_{f2}	TCATCACTCCAGGGACATTTAAA (478-500)	23	39.13	No	46.95	No	No	Yes	
	P_{r2}	CACCCAAATGATCAGCTCACTCTT (711-734)	24	45.83	No	50.69	No	No	Yes	
rs6353	P_{f1}	GTCTCAGGTCCCAGCCTCCTCT (381-402)	22	63.64	No	55.99	No	No	Yes	429
	P_{r1}	CCCTTCCCATTTCCTCACC (500-518)	19	57.89	Yes	48.64	No	No	Yes	
	P_{f2}	CTGGGCTTGGACAGCACA (483-500)	18	61.11	No	47.96	No	No	Yes	
	P_{r2}	GGCGCTTGACCCACTTTATCCC (686-707)	22	59.09	Yes	54.13	No	No	Yes	
rs6354	P_{f1}	CGGTGAAATGAAGGCACAGCAG (356-377)	22	54.55	Yes	52.27	No	No	Yes	413
	P_{r1}	AGCCCCTTGTTATTCTGCAAG (500-520)	21	47.62	Yes	47.43	No	No	Yes	
	P_{f2}	GGCGACCTTGCTTGCCCTCTA (480-500)	21	61.90	No	53.28	No	No	Yes	
	P_{r2}	AGGACAGAGGTAAAAAGGCCACC (742-764)	23	52.17	Yes	52.29	No	No	Yes	
rs6355	P_{f1}	AGAAGATTGTCAGGAAAACGG (387-407)	21	42.86	Yes	45.47	No	No	Yes	386
	P_{r1}	GGATAGAGTGCCGTGTGTCATCTC (500-523)	24	54.17	Yes	54.11	No	Yes	Yes	
	P_{f2}	GTTCCAAGTCCTGGTGCGGC (481-500)	20	65.00	Yes	53.55	No	No	Yes	
	P_{r2}	TGGGTCACAGCCTCTACTCGCA (692-713)	22	59.09	No	54.13	No	Yes	Yes	
rs25528	P_{f1}	AGACCTCAAGCCAGAGCTG (389-407)	19	57.89	Yes	48.64	No	No	Yes	416
	P_{r1}	TGATGCTGGGGTGGTTGGTG (500-519)	20	60.00	Yes	51.50	No	No	Yes	
	P_{f2}	GCACAGGGGGACGGCGAA (483-500)	18	72.22	No	52.51	No	No	Yes	
	P_{r2}	TTCCCATTATGCATTTGCAAGCCCG (671-695)	25	48.00	Yes	52.58	No	No	Yes	

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE II
THE RESULT OF SINGLE PRIMER OF THE PROPOSED METHOD FOR FIVE SNP IDS

Primer Constraints										
SNP ID	***Primer ID***	***Primers (position)***	***primer length (nt)***	***GC%***	***GC clamp***	***Tm (°C)***	***dimer***	***hairpin***	***specificity***	***fitness***
rs6352	P_{f1}	TAGTTCATCATTTGCAGTTTTCTG (333-356)	24	33.33	Yes	45.57	No	No	Yes	386
	P_{r1}	TACACACTAACTAGCACGTACCTCG (500-524)	25	48.00	Yes	52.58	No	No	Yes	
	P_{f2}	GTTGATCATCACTCCAGGGACATTTAAA (473-500)	28	39.29	No	52.01	No	Yes	Yes	
	P_{r2}	AAAACCTATGCACAGCCCAAGCTGA (746-770)	25	48.00	No	52.58	No	No	Yes	
rs6353	P_{f1}	CTTTGTAGGACAGGTCTTGTCAACC (335-359)	25	48.00	Yes	52.58	No	No	Yes	410
	P_{r1}	CCCATTTCCCCTTCCCATTTCCTCACC (500-526)	27	55.56	Yes	57.68	No	No	Yes	
	P_{f2}	GTTAATCACGCTGGGCTTGGACAGCACA (473-500)	28	53.57	No	57.87	No	No	Yes	
	P_{r2}	GGGCCAAGGGACAGTGCTTAATAAC (745-769)	25	52.00	Yes	54.22	No	No	Yes	
rs6354	P_{f1}	GTGAAATGAAGGCACAGCAGCCCCG (358-382)	25	60.00	Yes	57.50	No	No	Yes	416
	P_{r1}	TGGCTAAGCCCCTTGTTATTCTGCAAG (500-526)	27	48.15	Yes	54.64	No	No	Yes	
	P_{f2}	AAGGCGACCTTGCTTGCCCTCTA (478-500)	23	56.52	No	54.08	No	No	Yes	
	P_{r2}	ACCCCAGCTCATCAGCTACAGATTCTG (704-730)	27	51.85	Yes	56.16	No	No	Yes	
rs6355	P_{f1}	ATGGAGACGACGCCCTTGAATTCTCAGA (334-361)	28	50.0	No	56.40	No	Yes	Yes	223
	P_{r1}	GGGATAGAGTGCCGTGTGTCATCTC (500-524)	25	56.00	Yes	55.86	No	Yes	Yes	
	P_{f2}	CAGTTCCAAGTCCTGGTGCGGC (479-500)	22	63.63	Yes	55.99	No	No	Yes	
	P_{r2}	CAGCCTCTACTCGCAGCCTGTGATA (682-706)	25	56.00	No	55.86	No	No	Yes	
rs25528	P_{f1}	CTACCCCACGTTTGACGACTCAGACC (368-393)	26	57.69	Yes	58.56	No	No	Yes	19
	P_{r1}	CAGGTTACTGATGCTGGGGTGGTTGGTG (500-527)	28	57.14	Yes	59.33	No	No	Yes	
	P_{f2}	TGGAGGCACAGGGGGACGGCGAA (478-500)	23	69.57	No	59.42	No	No	Yes	
	P_{r2}	TGGAGGCACAGGGGGACGGCGAA (478-500)	23	69.57	No	59.42	No	No	Yes	

TABLE III
THE RESULT OF COMPOUND PRIMERS OF THE GA METHOD FOR FIVE SNP IDS

Primer Constraints					
SNP ID	***Primer ID***	***primer length difference (nt)***	***Tm difference (°C)***	***PCR product length (bps)***	***PCR product position***
rs6352	P_{f1}-P_{r1}	6	0.66	174	354-527
	P_{f2}-P_{r2}	1	3.75	257	478-734
	P_{f1}-P_{r2}	2	3.44	381	354-734
rs6353	P_{f1}-P_{r1}	3	7.35	138	381-518
	P_{f2}-P_{r2}	4	6.17	225	483-707
	P_{f1}-P_{r2}	0	1.86	327	381-707
rs6354	P_{f1}-P_{r1}	1	4.84	165	356-520
	P_{f2}-P_{r2}	2	0.99	285	480-764
	P_{f1}-P_{r2}	1	0.03	409	356-764
rs6355	P_{f1}-P_{r1}	3	8.64	137	387-523
	P_{f2}-P_{r2}	2	0.58	233	481-713
	P_{f1}-P_{r2}	1	8.66	327	387-713
rs25528	P_{f1}-P_{r1}	1	2.86	131	389-519
	P_{f2}-P_{r2}	7	0.07	213	483-695
	P_{f1}-P_{r2}	6	3.94	307	389-695

TABLE IV
THE RESULT OF COMPOUND PRIMERS OF THE PROPOSED METHOD FOR FIVE SNP IDS

Primer Constraints					
SNP ID	***Primer ID***	***primer length difference (nt)***	***Tm difference (°C)***	***PCR product length (bps)***	***PCR product position***
rs6352	P_{f1}-P_{r1}	1	7.01	192	333-524
	P_{f2}-P_{r2}	3	0.57	298	473-770
	P_{f1}-P_{r2}	1	7.01	438	333-770
rs6353	P_{f1}-P_{r1}	2	5.10	192	335-526
	P_{f2}-P_{r2}	3	3.64	297	473-769
	P_{f1}-P_{r2}	0	1.64	435	335-769
rs6354	P_{f1}-P_{r1}	2	2.86	169	358-526
	P_{f2}-P_{r2}	4	2.09	253	478-730
	P_{f1}-P_{r2}	2	1.34	373	358-730
rs6355	P_{f1}-P_{r1}	3	0.54	191	334-524
	P_{f2}-P_{r2}	3	0.13	228	479-706
	P_{f1}-P_{r2}	3	0.54	373	334-706
rs25528	P_{f1}-P_{r1}	2	0.77	160	368-527
	P_{f2}-P_{r2}	3	0.71	271	478-748
	P_{f1}-P_{r2}	0	1.58	381	368-748

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)