Article

Machine Learning Combined with Restriction Enzyme Mining Assists in the Design of Multi-Point Mutagenic Primers

Yu-Huei Cheng [1] , Li-Yeh Chuang [2,*] and Cheng-Hong Yang [3,4,5,6,7,*]

1 Department of Information and Communication Engineering, Chaoyang University of Technology, Taichung 413310, Taiwan
2 Department of Chemical Engineering, Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung 84001, Taiwan
3 Department of Information Management, Tainan University of Technology, Tainan 71002, Taiwan
4 Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80708, Taiwan
5 Ph.D. Program in Biomedical Engineering, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
6 School of Dentistry, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
7 Drug Development and Value Creation Research Center, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
* Correspondence: chuang@isu.edu.tw (L.-Y.C.); chyang@nkust.edu.tw (C.-H.Y.)

Abstract: The polymerase chain reaction–restriction fragment length polymorphism (PCR-RFLP) experiment has the characteristics of low-cost, rapidity, simplicity, convenience, high sensitivity and high specificity; thus, many small and medium laboratories use it to perform all kinds of single nucleotide polymorphisms (SNPs) genotyping works, and as a molecular biotechnology for disease-related analysis. However, many single nucleotide polymorphisms lack available restriction enzymes to distinguish the specific genotypes on a target SNP, and that causes the PCR-RFLP assay which is unavailable to be called mismatch PCR-RFLP. In order to completely solve the problem of mismatch PCR-RFLP, we have created a teaching–learning-based optimization (TLBO) multi-point mutagenic primer design algorithm which, combined with REHUNT, provides a complete and specific restriction enzyme mining solution. The proposed method not only introduces several search strategies suitable for multi-point mutagenesis primers, but also enhances the reliability of mutagenic primer design. In addition, this study is also designed for more complex SNP structures (with multiple dNTPs and insertion and deletion) to provide specific solutions for SNP diversity. We tested against fifteen mismatch PCR-RFLP SNPs in the human SLC6A4 gene on the NCBI dbSNP database as experimental templates. The experimental results indicate that the proposed method is helpful for providing the multi-point mutagenic primers that meet the constrain conditions of PCR experiments.

Keywords: polymerase chain reaction–restriction fragment length polymorphism (PCR-RFLP); single nucleotide polymorphism (SNP); multiple-point mutagenic primer design; teaching–learning-based optimization (TLBO); REHUNT

MSC: 92-08

1. Introduction

Polymerase chain reaction–restriction fragment length polymorphism (PCR-RFLP) has the characteristics of low cost, rapidity, simplicity, convenience, sensitivity and high specificity. Therefore, many biological laboratories usually use PCR-RFLP for single nucleotide polymorphisms (SNP) genotyping and as a molecular technique for disease association studies. Before using PCR-RFLP experiments for SNP genotyping, available restriction enzymes must be available to identify the specific genotype of the target SNP, and the primers used must meet the needs of various PCR experiments at the same time. Due to a large number of restriction enzymes, a large number of primer pairing combinations, complex

inter-primer annealing and strict experimental requirements, it is difficult to design usable and specific restriction enzymes and ideal primers by manual methods. In addition, using the PCR-RFLP experiments for SNP genotyping presents an extremely difficult problem, that is, "among many SNPs, there is simply no restriction enzyme available to identify the target SNP", which makes PCR-RFLP experiments completely impossible, and this problem is called "mismatch PCR-RFLP".

In order to address the mismatch PCR-RFLP problem, mutagenic primers must be designed for target SNPs which are not identified by restriction enzymes. The adjacent nucleotides can be mutated by inducing mutation primers so that the restriction enzymes can function to recognize the SNP, thereby performing the experimental effect of PCR-RFLP. In the past, many studies and cases have confirmed that PCR experiments have been verified to be feasible after nucleotides are mutagenic [1]. Therefore, with the support of these experimental cases, this study proposes an innovative design of multi-point mutagenesis primers, which help to make mutagenic primers more diverse and complete works on restriction enzymes mining, so as to fully play the role of PCR-RFLP experiments.

Many factors need to be considered in the design of multi-point mutagenic primers. In addition to specific primer design [2,3], these also include template sequence length, primer length range, number of target SNP genotypes, target SNP position, multi-point mutagenic nucleotide positions, number of multi-point mutagenic nucleotides, the available restriction enzymes, etc., which make the solution space for the design of multi-point mutagenic primers significantly large, and the design is extremely complex and challenging. In addition, although there are a certain number of known restriction enzymes, complete and specific restriction enzyme mining methods are extremely lacking, making it impossible to comprehensively consider the available restriction enzymes. It can be seen that the design of multi-point mutagenic primers has a significantly large solution space, and coupled with the huge amount of data generated by the current High-Throughput Screening (HTS) technology, it is necessary to provide high accuracy and efficiency high-throughput computation (HTC) algorithms.

In the previous research, Camilo et al. created HTP-OligoDesigner to provide a simple and intuitive online primer design tool for both laboratory-scale and high-throughput projects of sequence-independent gene cloning and site-directed mutagenesis, as well as a Tm calculator for quick queries [4]. Püllmann et al. developed Golden Mutagenesis and proposed a software library and web application for automated primer design and for the graphical evaluation of the randomization success based on the sequencing results in 2019 [5]. In order to solve the problem of mismatch PCR-RFLP for SNP genotyping, it is necessary to efficiently and accurately induce mutagenic nucleotides in the sequence through primers so that the target SNP, which is not recognized by restriction enzymes, can obtain usable restriction enzymes, so as to play the role of PCR-RFLP experiments. In order to improve the efficiency of mutagenic primer design, we also applied the Genetic Algorithm (GA) on the design of mutagenic primers in 2012 [6]. However, the genetic algorithm has the disadvantage that it easily falls into the local optimal solution, and the designed mutagenic primers cannot fully meet the restriction conditions of PCR experiments. This serious shortcoming will affect the quality of high-throughput mutagenic primer design in the future. Especially with the huge amount of data generated by the HTS technology, a HTC algorithm with high quality and efficiency is required. Therefore, we continue to invest in the research and development of more efficient and higher-quality algorithms for mutagenic primer design, and have studied a set of novel computational intelligence (CI) methods published by RV Rao for Teaching–Learning-Based Optimization (TLBO) [7,8]. In order to ensure that the TLBO method can be effectively applied to the mutagenic primer design, we had tested and applied it to the general PCR primer design. The results revealed that the TLBO method could effectively improve the efficiency of primer design, and the selected primers met the needs of PCR experiments. In addition, we used this method to test the results of primers designed by different primer melting temperature equations, and

applied statistical regression analysis to observe the changes in their parameter settings, so as to evaluate the suitability of TLBO methods for primer design [9].

The biological experiments proved that the TLBO method was able to help with primer design. Therefore, in 2015, this method was further applied to the design of mutagenic primers, and compared with the method of mutagenic primer design in the previous literature, it did obtain better results than other methods [10]. However, after completing the preliminary design and observation of the results, it was found that the developed mutagenic primer design algorithm did not take into account the complex SNP sequence structure, and did not introduce suitable search strategies, which both limit the method for improvement. In 2018, we proposed REHUNT [11], which provides a complete application programming interface for restriction enzyme mining, which can be effectively integrated into various biological algorithms. REHUNT is a reliable and open source package implemented in JAVA. It is able to provide all available restriction enzymes for the imported biological sequences. It can also identify different genotypes combined with PCR-RFLP, including SNPs, mutations and the other variations. Furthermore, classified restriction enzymes, including IUPAC (International Union of Pure and Applied Chemistry chemical nomenclature) and general sequence types, and commercial and non-commercial availabilities, as well as HTS (high-throughput screening) analysis, are available in REHUNT. Therefore, the design method of mutagenic primers is expected to be introduced into systematic high-throughput computing. With the advancement of computing architecture of computer hardware, the huge amount of data generated by high-throughput screening technology can be effectively utilized. In order to improve the quality of designed mutagenic primers in large-scale analysis, this study proposes the novel design technology of multi-point mutagenic primers and introduces TLBO improved strategies to exactly match the design of multi-point mutagenic primers, so as to help with biological practical use by medical researchers.

2. Materials and Methods

When designing primers for multi-point mutagenic primers, the consideration of primer constraints is an important key to the success or failure of the experiment. In PCR experiments, people (operators), things (research projects), time (operation time), places (location and environment) and objects (materials used) are often the reasons for different experimental results.

In order to enable the designed multi-point mutagenic primers to successfully conduct PCR-RFLP experiments during SNP genotyping, this study used the experience of the single-point mutagenic primer design conditions used in [6,10] to improve them. At the same time, it analyzes and evaluates the primer design conditions used in various actual PCR-RFLP experiments, and finds the most necessary and most suitable elements for designing multi-point mutagenic primers. Through in-depth discussion and analysis of the factors affecting the PCR-RFLP experiment, the necessary conditions for the design of multi-point mutagenic primers are improved in order to ensure their quality and the success rate of the experiment. Then, the design conditions of these multi-point mutagenic primers are functionalized and planned as an application programming interface (API) library in order to provide relevant computational intelligence methods and reuse by researchers in the future. In the design of the multi-point mutagenic primer design algorithm, the TLBO method is mainly used for research, development and improvement, and suitable multi-point mutagenic primer strategies are proposed to improve the quality of its multi-point mutagenic primer. After completing the development of the algorithm, we tested and analyzed the high-throughput SNP genotyping multi-point mutagenic primer design method which we developed.

2.1. In-Depth Discussion and Analysis of the Factors Affecting the Experiment, and Improvement of the Design Conditions of the Necessary Multi-Point Mutagenic Primers

The primer conditions and factors that need to be considered in PCR experiments are mostly derived by many researchers based on their accumulated experience in PCR experiments in the past, including the quality of the template sequence, the primer length, the length difference between the forward and reverse primers, ratio of primer 'G/C' nucleotides, primer 3' end restriction, melting temperature, melting temperature difference among forward primer, reverse primer, template sequence, primer annealing position, the duplication of nucleotides, the formation of dimerization of primers, the formation of hairpin of primers, the specificity of primers, the size of products, etc. The primer conditions and factors to be considered are all mentioned in the literature related to biotechnology [12]. Therefore, this study uses the relevant primer design conditions suggested in the literature to analyze and discuss, in order to improve and adjust the multi-point mutagenic primer design conditions. Under the constraints of so many primer design conditions, not all primer design conditions can be used to effectively design appropriate multi-point mutagenic primers which can achieve a successful PCR-RFLP experiment. There are many primer design conditions which are actually interrelated in nature and will affect each other; for example, the primer length and the 'G/C' nucleotide ratio of the primer will affect the melting temperature of the primer. When the primer length is too long, its melting temperature will increase; when the 'G/C' nucleotide ratio of the primer is too high, its melting temperature will also increase. In addition, the primer's annealing position will affect the size of the PCR product, mainly because the distance between the forward and reverse primers is too long, and the amplicon size may be too big. If the distance between both of the primers is too short, the length of the PCR product will be too short. Therefore, when designing the weight of primer conditions, if only several primer conditions are used as the evaluation basis, it is impossible to take into account the influence of the interaction between the primer conditions on the design of the overall multi-point mutagenic primers, and it is difficult to achieve an effective balance.

2.2. Functionalization of Multi-Point Mutagenic Primers and Design of API Library for Computational Intelligence Methods

In order to provide future research and development of multi-point mutagenic primer design algorithms, and to facilitate the continuous use of these improved multi-point mutagenic primer design conditions, this study functionalized these improved multi-point mutagenic primer design conditions and planned them in the API (Application Programming Interface) library to provide relevant computational intelligence methods and reuse by researchers in the future. The API function library of these related definitions includes primer length function, length difference function between forward and reverse primer, ratio function of primer 'G' and 'C' nucleotides, primer 3' end function, melting temperature calculation function, melting temperature difference function of forward and reverse primer to template sequence, repeat calculation function of the same nucleotide in primer, primer dimer evaluation function, primer hairpin evaluation function, specificity evaluation function, product length calculation function, etc. In defining these functions, first, the solution of the mutagenic primer design problem is represented by *s*. The overall template sequence is mainly based on the nucleotide base composition of 'A', 'T', 'C' and 'G,' and considers the inclusion of multiple SNPs. It is defined as formula (1):

$$T_D = \left\{ B_i \,\middle|\, \forall B_i \in \text{'A' or 'T' or 'C' or 'G' or IUPAC format}, i \in \mathbb{Z}^+ \right\} \tag{1}$$

where T_D represents the template sequence; B is the nucleotide base; i represents the position of the nucleotide base.

The forward primer P_f, i.e., the mutagenic primer, and the reverse primer P_r are defined as the following Formulas (2) and (3), respectively.

$$P_f = \{B_i \,|\, \forall B_i \in \text{'A' or 'T' or 'C' or 'G'} \text{ and } \exists B_i \in SNP, i \in \text{the index of } T_D, \; F_s \leq i \leq F_e\} \quad (2)$$

$$P_r = \{\overline{B_i} \,|\, \forall B_i \in \text{'A' or 'T' or 'C' or 'G'}, i \in \text{the index of } T_D, \; R_s \leq i \leq R_e\} \quad (3)$$

where P_f represents the forward primer; P_r represents the reverse primer; F_s and F_e represent the starting position and end position of the forward primer, respectively; R_s and R_e represent the starting position and end position of the reverse primer. Here, it should be noted that the forward primer mainly contains the subsequence of the SNP's base at the 3′ end, and the reverse primer does not contain the subsequence of the SNP's base.

The following defines the conditional functions for the design of mutagenic primers:

(1)   Primer length function

The primer length will cause the melting temperature between the primer and the template sequence to increase or decrease, and it will also indirectly affect the generation of secondary structure and the specificity of the primer. Therefore, this study functionalizes the primer length function, which is expressed as follows:

$$P_{len\_min} \leq P_{len}(s) \leq P_{len\_max} \quad (4)$$

where the P_{len} represents the length of the forward and reverse primers; the P_{len_min} represents the minimum primer length designed; the P_{len_max} represents the maximum primer length designed.

(2)   Primer length difference function

The designed lengths of the forward and reverse primers are not necessarily the same. Many studies have found that a length difference between both of the primers of less than 3 nt is the most ideal state, but due to some special requirements, the length difference must be made within a certain length. In order to allow the function to flexibly adjust the length difference, the length difference is represented by *n* here, and the function is expressed as follows:

$$P_{len\_diff}(s) = \left| P_{f\_len} - P_{r\_len} \right| \leq n \quad (5)$$

where the P_{len_diff} represents the length difference between primers; the P_{f_len} represents the forward primer length; the P_{r_len} represents the reverse primer length; *n* represents the length difference.

(3)   The ratio function of primer 'G' and 'C' nucleotides

The ratio of 'G/C' nucleotides in primers is generally found in the literature to be between 40% and 60% for optimal adhesion. In order to allow users to adjust according to their needs, we provide the ratio function of primer 'G' and 'C' nucleotides as follows:

$$GC_{min} \leq P_{GC}(s) \leq GC_{max} \quad (6)$$

where the GC_{min} is the minimum G/C ratio value; the GC_{max} is the maximum G/C ratio value; the P_{GC} function represents the proportion of 'G/C' nucleotides in the forward and reverse primers.

(4)   The 3′ end function

Usually the 3′ end of a primer is designed with 'G' or 'C' nucleotides, because 'G' or 'C' nucleotides have a stronger bond than 'A' or 'T' nucleotides in DNA structure. Considering the nucleotides at the primer 3′ end, the function is expressed as follows:

$$P_{3end}(s) \in \text{'G' or 'C'} \quad (7)$$

where the $P_{3\text{end}}$ function represents the nucleotides at the 3' end of the forward and reverse primers.

(5)　Melting temperature calculation function

　　The melting temperature (Tm) is the temperature at which the primer separates from the template. In this study, the thermodynamics theory proposed by SantaLucia [13] is functionalized, which is also the most accurate melting temperature equation identified by researchers, as follows:

$$Tm_{SAN}(s) = \Delta H^\circ(predicted)/(\Delta S^\circ(\text{salt\_correction}) + R \times \ln(C/4)) \tag{8}$$

where $\Delta H^\circ(predicted)$ is enthalpy; $\Delta S^\circ(\text{salt\_correction})$ is entropy correction; R is gas constant (1.987 cal/Kmol); C is DNA concentration.

(6)　The melting temperature difference function

　　The difference of Tm between the forward and reverse primers and the template sequence is generally considered to be no more than 55 °C; otherwise, the primer will not be properly attached to the template sequence. In order to allow users to set flexibly, the function of the Tm difference between both of the primers and the template sequence is expressed as follows:

$$P_{tm\_diff}(s) = \left| P_{f\_tm} - P_{r\_tm} \right| \leq n \tag{9}$$

where the P_{tm_diff} function represents the Tm difference between the forward and reverse primers and the template sequence; the P_{f_tm} represents the Tm difference between the forward primer and the template sequence; the P_{r_tm} represents the Tm difference between the reverse primer and the template sequence; n represents the maximum difference of Tm.

(7)　Repeat calculation function of identical nucleotides in primers

　　When a specific nucleotide within a forward or reverse primer is repeated multiple times, such as ATATATAT, this situation may result in errors in annealing to the template sequence. Therefore, the repetition calculation function for the same nucleotide within a primer is expressed as follows:

$$P_{repeat}(s) \leq r \tag{10}$$

where the P_{repeat} function is the evaluation function of repeated nucleotides in the primer; r represents the maximum allowable number of repetitions.

(8)　Primer dimer assessment function

　　When both of the forward and reverse primers are in a dimer structure due to the composition of nucleotides, it can easily cause the primer to fail to bind to the template sequence. Here, the primer dimer evaluation function is expressed as follows:

$$P_{dimer}(s) \notin \text{cross dimer \&self dimer} \tag{11}$$

where the P_{dimer} function is to evaluate the dimer of the forward and reverse primers; the cross dimer is the mutual adhesion of the forward and reverse primers; the self dimer is the mutual adhesion of the forward and forward primers or the reverse and reverse primers.

(9)　Hairpin evaluation function

　　The main reason for the formation of hairpin bends is that the nucleotide structure of the forward primer or the nucleotide structure of the reverse primer are polymerized before and after each other, which easily leads to the failure of the primer to adhere to the template sequence and, thus, the failure of the PCR experiment. The evaluation function for the formation of primer hairpin bends is expressed as follows:

$$P_{hairpin}(s) \notin \text{hairpin} \tag{12}$$

where the $P_{hairpin}$ function is to evaluate the hairpin bending situation of the forward primer and the reverse primer itself.

(10)　Specificity assessment function

Specificity is a very important property in primer design. Primers with specificity can bind to specific template sequence positions in order to generate desired products. Therefore, this study includes the introduction specificity evaluation function, which is expressed as follows:

$$P_{specificity}(s) \ = \ P_f \text{ and } P_r \text{ repear in the specific position of } T_D \tag{13}$$

where the $P_{specificity}$ function evaluates that the forward and reverse primers must be able to bind to a specific template sequence position.

(11)　Product length function

The length of the product is the key to identifying the success or failure of a PCR experiment. Usually, the length of the product that can be recognized at present is about 150 bp. However, in order to provide researchers with flexible settings, the product length calculation function is expressed as follows:

$$P_{product}(s) \geq n \tag{14}$$

where the $P_{product}$ function represents the PCR product length formed by the forward and reverse primers; n represents the size of the product.

2.3. Developing a Search Strategy for Primers Suitable for Multi-Point Mutagenic, Importing High-Throughput Computing and Improving Computing Efficiency

This research develops a search strategy suitable for multi-point mutagenic primers and integrates it into the TLBO method in order to improve the quality of multi-point mutagenic primer design. At the same time, high-throughput computing processing is introduced in order to improve computational efficiency of multi-point mutagenic primer design. Teaching–learning-based optimization, an algorithm inspired by the teaching and learning process, is proposed by Rao et al. [7,8,14]. This method is mainly based on the teacher's role in influencing learners' effectiveness in the classroom and the concept of learners' mutual self-learning. There are two main basic learning modes: (1) teaching through teachers (teaching phase) and (2) interacting with other learners (learning phase). In the solution process of this method, the subjects studied by each learner are regarded as different design variables of the optimization problem, and the learning outcomes of the learners are similar to fitness values of the optimization problem. In order to make this method more effective in the design of multi-point mutagenic primers, this study develops a search strategy suitable for multi-point mutagenic primers to help this method more effectively escape the shortcomings of the genetic algorithm falling into the optimal solution, and it can also effectively improve the quality of multi-point mutagenic primer design.

(1)　 Multi-point mutagenic primer design

Figure 1 is a conceptual diagram of multi-point mutagenic primer design. Firstly, according to the multi-point mutagenic primer design, the solution should be solved by the teaching–learning-based optimization method. In this study, M, F_l, P_l, R_l, F_m and Σ are used as learner codes to design multi-point mutagenic primers. The symbol M represents the mutagenic nucleotide number; F_l represents the mutagenic primer length; P_l represents the product size; R_l represents the reverse primer length; F_m represents the mutagenic nucleotide position, and its position index can be from 1 to max-M; max-M represents the maximum number of mutagenic nucleotides; Σ represents an integer value of 0, 1, 2 or 3,

and its individual meaning represents the mutagenic nucleotides 'A', 'T', 'C' or 'G'. The learner is coded as follows:

$$s = (M, F_l, P_l, R_l, F_{m1}, \Sigma_1, F_{m2}, \Sigma_2, \ldots \ldots, F_{m\ max\text{-}M}, \Sigma_{max\text{-}M}) \tag{15}$$
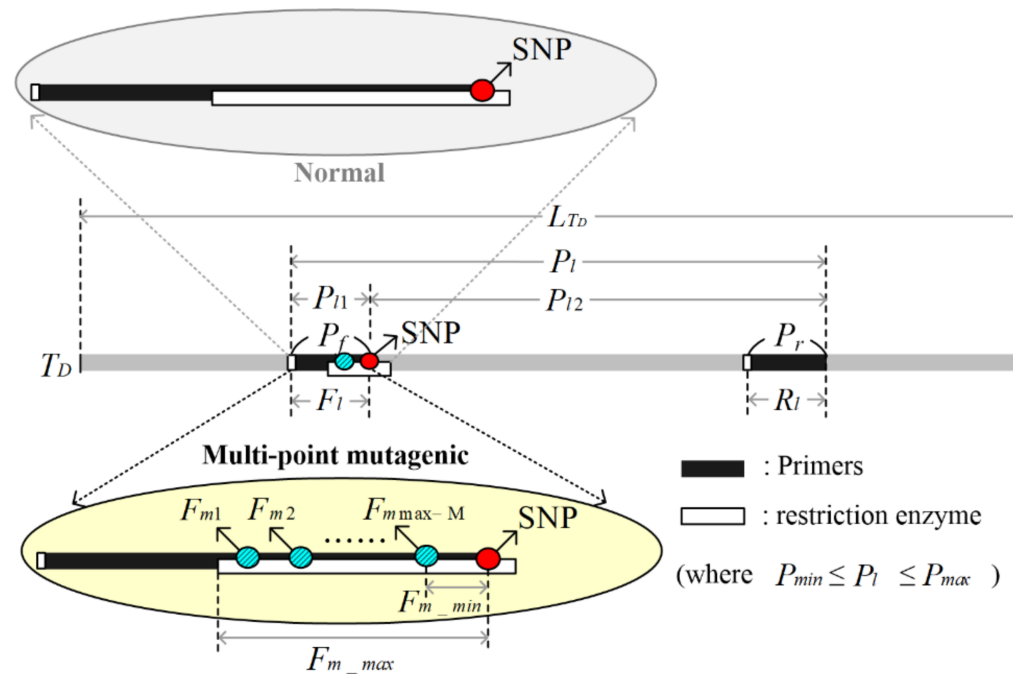
Figure 1. Conceptual diagram of multi-point mutagenic primer design.

In order to improve the design efficiency of multi-point mutagenic primers, we developed REHUNT (Restriction Enzymes HUNTing) API [11], using JAVA to carry out complete and specific restriction enzymes mining work.

In addition, we also applied the technology of the single-point mutagenic matrix [6,10] to propose a multi-point mutagenic matrix to completely record the restriction enzymes mining results of the target SNP. The single-point mutagenic matrix contains a total of $(F_{m\_max} - F_{m\_min} + 1) \times 4$ elements, while the multi-point mutagenic matrix contains $F_{m\ max\text{-}M}$ single-point mutagenic matrices, as shown in Figure 2. The row indicates the mutagenic position, which is between (SNP—F_{m_max} + 1) and (SNP—F_{m_min} + 1). The column represents the four bases of the single-point mutagenic nucleotides 'A', 'T', 'C' and 'G'. In the single-point mutagenic matrix, the internal element values are 0, 1, 2 or 3, respectively, and their meanings are expressed as follows:

0: Restriction enzymes available on both strands that can distinguish the target SNP when the single-point mutagenic nucleotide is determined.

1: Restriction enzymes available only on one of the double strands that can distinguish the target SNP when the single-point mutagenic nucleotide is determined.

2: No restriction enzyme available on either strand that can distinguish the target SNP when the single-point mutagenic nucleotide is determined.

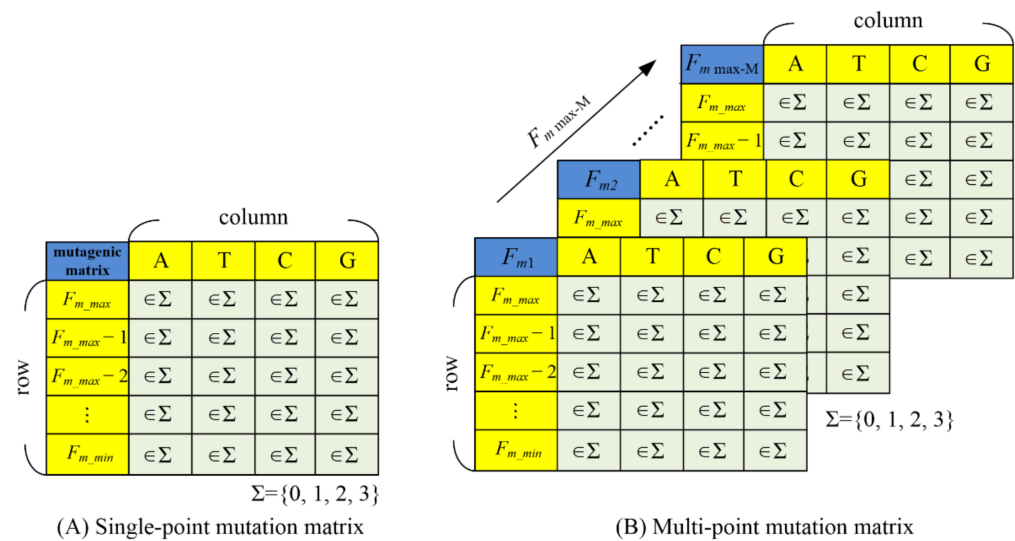3: No single-point mutagenic nucleotide, and the original nucleotide is still maintained.

Figure 2. Single- and multi-point mutagenic matrices.

(2)   Search strategies

In terms of search strategies, local search, elite search and interactive learning are added to the TLBO method [14]. The purpose of adding the local search strategy is to find better learners near the original learners. The knowledge of the original learner is improved by learning the knowledge of the better nearby learner, so the local optimal solution centered on the learner can be obtained. In order Tt have a better solution for the learners when first using the TLBO method, a local search of adjacent learners is first performed for the learners in the initial learning group. Afterwards, local searches are also performed after subsequent learner-interactive learning phases. In each iteration of the local search process, all children of the learners will participate in the local search; thus, the local optimal solution of each learner is continuously retained, and finally, the global optimal solution is obtained. The pseudo code (Algorithm 1) of the learner's local search strategy used in this study is shown below. In this pseudo code, S represents a group, and d is used to assist learners to search for their neighbors. Before calculating d, the constant a must be determined for the variable values of the different problems. In this study, the constant a of the local search will be determined by the length difference between the maximum and minimum primer length in order to obtain a better range.

The pseudo code of the local search:

Algorithm 1 Local search [3]

---

1   Begin;
2   Select an incremental value $d = a \times Rand()$;
3   For a given learner $i \in S$: calculate *achievement*(i);
4   For $j = 1$ to number of variables in learner i
5   *value*(j) = *value*(j) + d;
6   If *achievement*(i) in learner i is not improved then
7   *value*(j) = *value*(j) − d;
8   Else if *achievement*(i) in learner i is improved then
9   Retain *value*(j);
10   Next j;
11   Next i;
12   End;

Where d is the search range; a is a variable that is dependent on the problem; S represents the entire learning group; *achievement* is the learning outcome.

In addition, the purpose of adding the elite search is to provide a higher probability of being a teacher to teach the rest of the learners through several excellent learners, and to use *n* elite learners as auxiliary teachers at the same time in order to promote convergence of learning process and achieve better learning outcomes. Therefore, before and during the iteration of the TLBO method, the elite learners in the learning group will be given priority to serve as teachers with a higher probability, so as to obtain better results in the learning process. The pseudo code (Algorithm 2) for the elite search used in this study is shown below.

The pseudo code of elite search:

Algorithm 2 Elite search

1	Begin;
2	For *i* = 1 to number of learners
3	Initialize *teach_rate*(*i*) = 0;
4	For *j* = *i* + 1 to number of learners
5	If *achievement*(*i*) < *achievement*(*j*) then
6	*teach_rate*(*i*) = *teach_rate*(*i*) + (1/number of learners);
7	Next *j*;
8	Next *i*;
9	Find *n* better learner according to *teach_rate*(*i*) from learners;
10	Set *n* better learner as the assistant teachers;
11	End;

Where *teach_rate* is the probability of being selected as a teacher; *n* is the number of auxiliary teachers.

Finally, the interactive learning helps learners to interact with others in the learning group. If the learners who want to interact have more knowledge than they themselves do, the learners will acquire new knowledge by learning with them. This strategy takes into account real-life peer learning and is ideal for TLBO methods. By interacting with learners in other overall groups, learners can avoid being limited to only acquiring regional knowledge, and can acquire global knowledge centered on this learner. The pseudo code (Algorithm 3) proposed in this study to conduct an interactive search is shown as follows:

The pseudo code of the interactive search:

Algorithm 3 Interactive learning

1	Begin;
2	For a random learner $P \in S$: calculate *achievement*(*P*);
3	For *j* = 1 to number of variables in learner *i*
4	Randomly select a learner $Q \in S$ and $Q \neq P$;
5	Temp = *value*(*P*, *j*);
6	If *achievement*(*P*) < *achievement*(*Q*)
7	*value*(*P*, *j*) = *value*(*P*, *j*) + r_i (*value*(*P*, *j*) − *value*(*Q*, *j*));
8	Else
9	*value*(*P*, *j*) = *value*(*P*, *j*) + r_i (*value*(*Q*, *j*) − *value*(*P*, *j*));
10	If *achievement*(*P*) in learner *P* is not improved then
11	*value*(*P*, *j*) = temp;
12	Else if *achievement*(*P*) in learner *i* is improved then
13	Retain *value*(*P*, *j*);
14	Next *j*;
15	End;

Where *k* represents the learners who want to interact; *S* represents the whole learning group.

(3)　High-throughput computing

The steps of high-throughput computing are described as follows:

Step 1. Learner code for mutagenic primer solution.

First, learner coding is performed for the problem of mutagenic primer design. Each learner represents the solution of a mutagenic primer, and each variable in the learner represents the subject it learns.

Step 2. Import the high-throughput target SNP template sequence.

Next, the template sequences of the high-throughput target SNPs must be imported to be used as primers for designing high-throughput multi-point mutagenic primers. The template sequence of the imported high-throughput target SNP is recommended to be at least 250 bps adjacent to the target SNP, and the full length must be at least 501 bps to provide sufficient sequence for the selection of optimized primer fragments.

Step 3. High-throughput computational analysis processing.

In order to allow subsequent algorithms to process the template sequences of high-throughput target SNPs, this study uses a distributed programming framework to design and build a computing cluster system for analysis and processing.

Step 4. Calculate the multi-point mutagenic matrix.

First, the developed REHUNT of restriction enzyme mining API is used to search for target SNP restriction enzymes in order to generate a multi-point mutagenic matrix, so that the mutagenic restriction enzyme information can be efficiently and directly accessed during the iteration of the algorithm.

Step 5. Determine the presence or absence of restriction enzymes.

After judging whether the target SNP is subjected to multi-point mutagenic design, the available restriction enzyme is determined for identification of the target SNP position.

Step 6. Initialize the learning population.

If the target SNP has available restriction enzymes, tens to hundreds of s learners are randomly generated as the initial learning population. The value of F_m is randomly generated between (SNP—F_{m_max} + 1) and (SNP—F_{m_min} + 1). The value of F_l is randomly generated between the minimum and maximum primer lengths, according to common primer length constraints. In addition, in order to limit the PCR amplicon size, the value of P_l is randomly generated between the minimum and maximum product length. The value of R_l is generated in the same way as for F_l. The value of Σ is randomly generated between 0, 1, 2 or 3, and its individual meaning represents the mutagenic nucleotide 'A', 'T', 'C' or 'G'.

Step 7. Do a local search.

In order to enable the learners at the beginning to have good quality, the initial local search will be carried out for the learners in the initial learning group. Afterwards, a local search will also be performed after subsequent learner-interactive learning phases. Through the iterative process, all children's learners will participate in the local search, so that the local optimal solution of each learner is continuously retained, and thus, the global optimal solution is determined.

Step 8. Perform an elite search.

Before and during the iteration of the teach-learning optimization method, the algorithm will seek out elite learners in the learning population and provide a high probability of becoming teachers to teach the rest of the learners. With elite learners as auxiliary teachers, the convergence of the learning process will be promoted and better learning outcomes will be achieved.

Step 9. Teach Phase.

During this phase, teachers try to increase the average learning outcomes of learners in the subject courses they teach according to their abilities. At any iteration i, assume there are m number of subjects (i.e., design variables), n number of learners (i.e., learning population size, k = 1, 2, . . . , n) and $M_{j,i}$ for learners with an average learning outcome for a particular subject j (j = 1, 2, . . . , m). The best overall learning outcome $X_{total-kbest,i}$ considers the best learning outcomes of the entire learner group in all subjects, and takes *kbest* as the best learner. Under normal circumstances, a teacher is usually considered a person with a high learning ability to train learners for better learning outcomes. The differences between

the existing average learning outcomes for each subject and the corresponding learning outcomes for teachers in each subject are given as follows:

$$Difference\_Mean_{j,k,i} = r_i \left(X_{j,kbest,i} - T_F M_{j,i} \right) \qquad (16)$$

where $X_{j,kbest,i}$ are the learning results of the best learner (i.e., teacher) in subject j; T_F is the teaching factor, which determines the change in the mean; r_i is a random number in the interval $[0, 1]$. The T_F value can be 1 or 2, randomly determined with equal probability as follows:

$$T_F = round\left[1 + rand(0,1)\{2 - 1\} \right] \qquad (17)$$

Here, the T_F value is not used as an input to the algorithm, and its value is randomly determined by Equation (17). It is pointed out in the literature that the TLBO algorithm is superior in performing many benchmark function simulation experiments with T_F values between 1 and 2. However, in the simulation experiment, it was found that a T_F value of 1 or 2 is more suitable for solving the algorithm. Therefore, in order to simplify the algorithm, it is recommended to use 1 or 2 for the teaching factor.

Step 10. Update Learning Outcomes.

Based on $Difference\_Mean_{j,k,i}$, the existing solution is updated in the teach phase according to the following formula:

$$X'_{j,k,i} = X_{j,k,i} + Difference\_Mean_{j,k,i} \qquad (18)$$

where $X'_{j,k,i}$ is the updated value of $X_{j,k,i}$. The algorithm accepts $X'_{j,k,i}$ if it provides a better function value. All accepted function values are retained at the end of the teaching phase, and these values become the learner's input to the interactive learning phase, which depends on the teaching phase.

Step 11. Interactive learning phase.

After a certain period of iteration, the learner can randomly interact with the learners in the learning group in order to improve the learner's own knowledge and escape from the optimal solution in the region. We express the learning phenomenon in this stage as follows:

Randomly select two learners P and Q such that $X'_{total-P,i} \neq X'_{total-Q,i}$, where $X'_{total-P,i}$ and $X'_{total-Q,i}$ are the updated values of $X_{total-P,i}$ and $X_{total-Q,i}$ at the end of the teacher phase. The interactive learning update is shown in Equations (19) and (20).

$$X''_{j,P,i} = X'_{j,P,i} + r_i \left(X'_{j,P,i} - X'_{j,Q,i} \right), \text{ if } X'_{total-P,i} < X'_{total-Q,i} \qquad (19)$$

$$X''_{j,P,i} = X'_{j,P,i} + r_i \left(X'_{j,Q,i} - X'_{j,P,i} \right), \text{ if } X'_{total-P,i} > X'_{total-Q,i} \qquad (20)$$

Step 12. Assessment of Learning Outcomes.

Through the designed learning outcome function, each learner will be evaluated, and each learner will have a corresponding learning outcome value. Here, the learning outcome function was evaluated based on the design conditions of the multi-point mutagenic primers.

Step 13. Terminate condition judgment.

During the iteration process, the algorithm will determine whether the current learner's learning outcome value has been minimized, that is, a learning outcome value of 0 indicates that the optimal learning outcome value has been reached, or the calculation process reaches the preset number of iterations, and the calculation is stopped.

3. Results

3.1. Dataset

In order to evaluate the proposed method for mutagenic primer design, we used the fifteen mismatch PCR-RFLP SNPs in the human SLC6A4 gene from the NCBI dbSNP database [15,16] to take 250 bps from the left and right sides. The human gene, SLC6A4, is

related to autism spectrum disorders (ASD) [17], psychosis [18] and bipolarity [19]. The fifteen mismatch PCR-RFLP SNPs are shown in Table 1.

Table 1. Fifteen mismatch PCR-RFLP SNPs in the SLC6A4 gene.

No.	Results	
	rs#	Remark
1	rs12150096	The iteration number is set to 500, and the primers cannot be designed.
2	rs2020932	-
3	rs28441519	-
4	rs34185064	-
5	rs3783594	-
6	rs41274280	The iteration number is set to 100, and the primers cannot be designed.
7	rs45541837	-
8	rs56082703	-
9	rs56162408	-
10	rs7217065	-
11	rs73987804	The iteration number is set to 400, and the primers cannot be designed.
12	rs8071583	-
13	rs8080561	-
14	rs9652882	The iteration number is set to 100, and the primers cannot be designed.
15	rs9916159	The iteration number is set to 400, and the primers cannot be designed.

3.2. Design of Parameters

The parameters of the method used in this study were set to 100, 200, 300, 400 and 500 iterations in order to observe the primer design and convergence of the mismatch PCR-RFLP SNPs. The population size was set to 50. The length of the primers to be designed was between 16 bps and 28 bps; the length difference between the primers of forward and reverse was 3 bps; the ratio of 'G/C' nucleotides of primers was between 40% and 60%; the primer melting temperature was between 45 °C and 62 °C; the melting temperature difference between the forward and reverse primer to the template sequence was 5 °C and the product length was between 150 and 300. The remaining primer conditions, including primer 3′ end restriction, repeated counting of identical nucleotides in primers, primer dimers, primer hairpin and primer specificity, were all tested.

3.3. REHUT and Primer Results

Through the proposed method, 15 SNPs with mismatch PCR-RFLP in the SLC6A4 gene were tested. It was found that the SNPs rs41274280, rs9652882 and rs9916159 could not design mutagenic primers when the iteration number was set to 100. The SNPs rs73987804 and rs9916159 failed to design mutagenic primers when the iteration number was set to 400, and the SNP rs12150096 could not design mutagenic primers when the iteration number was set to 500. When 100 was set for the iteration number, a total of 12 pairs of primers were generated, with a total of 24 primers. Among them, all primers met the primer length; only 4 pairs of primers meet the primer length difference; 13 primers met the ratio of 'G/C' nucleotides; 5 primers met the ratio of 'G/C' nucleotides at 3′ end; 19 primers met the primer melting temperature; 6 pairs of primers met the primer melting temperature difference; 24 primers met the product length; 12 primers avoided cross-dimer; 24 primers avoided self-dimer; 18 primers avoided hairpin. Finally, 24 primers met the primer specificity. Table 2 shows the results.

As shown in Table 3, when the number of iterations was set to 200, a total of 15 pairs of primers were generated, with a total of 30 primers. Among them, all primers met the primer length; only 6 pairs of primers met the primer length difference; 15 primers met the ratio of 'G/C' nucleotides; 10 primers met the ratio of 'G/C' nucleotides at 3′ end; 22 primers conformed to the primer melting temperature; 11 pairs of primers met the primer melting temperature difference; 30 primers met the product length; 15 primers avoided cross-dimer;

30 primers avoided self-dimer; 25 primers avoided hairpin. Finally, 30 primers met the primer specificity.

Table 2. Design results of mutagenic primers combined with REHUNT. The number of iterations was set to 100 and the population size was set to 50.

Primer Info. *	Results	
	Number of Eligible	Number of Violations
Primer length	24	0
Length difference of forward and reverse primer	4	8
Ratio of primer 'G' and 'C' nucleotides	13	11
Primer 3′ end	5	19
Primer Tm	19	5
Tm difference between the forward and reverse primers in the template sequence	6	6
Amplicon size	24	12
Cross-dimer	12	0
Self-dimer	24	0
Hairpin	18	6
Specificity	24	0

* Tm: melting temperature.

Table 3. Design results of mutagenic primers combined with REHUNT. The number of iterations and population size were set to 200 and 50, respectively.

Primer Info. *	Results	
	Number of Eligible	Number of Violations
Primer length	30	0
Length difference of forward and reverse primer	6	9
Ratio of primer 'G' and 'C' nucleotides	15	15
Primer 3′ end	10	20
Primer Tm	22	8
Tm difference between the forward and reverse primers in the template sequence	11	4
Amplicon size	30	15
Cross-dimer	15	0
Self-dimer	30	0
Hairpin	25	0
Specificity	30	0

* Tm: melting temperature.

When the iteration number was set to 300, a total of 15 primer pairs were generated, for a total of 30 primers. Among them, all primers met the primer length; only 6 pairs of primers met the primer length difference; 16 primers conformed to the ratio of 'G/C' nucleotides; 10 primers met the ratio of 'G/C' nucleotides at 3′ end; 22 primers met the primer melting temperature; 11 pairs of primers met the primer melting temperature difference; 30 primers met the product length; 15 primers avoided cross-dimer; 30 primers avoided self-dimer; 23 primers avoided hairpin. Finally, 30 primers met the primer specificity. The results are presented in Table 4.

As shown in Table 5, when the number of iterations was set to 400, a total of 13 pairs of primers were generated, for a total of 26 primers. Among them, all primers met the primer length; only 4 pairs of primers met the primer length difference; 15 primers fit the ratio of 'G' and 'C' nucleotides; 8 primers fit the 'G' and 'C' nucleotide ratio at 3′ end; 22 primers met the primer melting temperature; 10 pairs of primers met the primer melting temperature difference; 26 primers met the product length; 13 primers avoided cross-dimer;

26 primers avoided self-dimer; 22 primers avoided hairpin. Finally, 26 primers met the primer specificity.

Table 4. Design results of mutagenic primers combined with REHUNT. The number of iterations and population size were set to 300 and 50, respectively.

Primer Info. *	Results	
	Number of Eligible	**Number of Violations**
Primer length	30	0
Length difference of forward and reverse primer	6	9
Ratio of primer 'G' and 'C' nucleotides	16	14
Primer 3′ end	10	20
Primer Tm	22	8
Tm difference between the forward and reverse primers in the template sequence	11	4
Amplicon size	30	15
Cross-dimer	15	0
Self-dimer	30	0
Hairpin	23	7
Specificity	30	0

* Tm: Melting temperature.

Table 5. Design results of mutagenic primers combined with REHUNT. The number of iterations and population size is set to 400 and 50, respectively.

Primer Info. *	Results	
	Number of Eligible	**Number of Violations**
Primer length	26	0
Length Difference of forward and reverse primer	4	9
Ratio of primer 'G' and 'C' nucleotides	15	11
Primer 3′ end	8	18
Primer Tm	22	4
Tm difference between the forward and reverse primers in the template sequence	10	3
Amplicon size	26	13
Cross-dimer	13	0
Self-dimer	26	0
Hairpin	22	4
Specificity	26	0

* Tm: melting temperature.

Finally, when the iteration number was set to 500, a total of 14 pairs of primers were generated, for a total of 28 primers. Among them, all primers met the primer length; only 5 pairs of primers met the primer length difference; 18 primers met the 'G' and 'C' nucleotide ratio; 7 primers fit the 'G' and 'C' nucleotide ratio at 3′ end; 22 primers met the primer melting temperature; 8 pairs of primers met the primer melting temperature difference; 28 primers met the product length; 14 primers avoided cross-dimer; 28 primers avoided self-dimer; 23 primers avoided hairpin. Finally, 28 primers met the primer specificity. The results are shown in Table 6.

Table 6. Design results of mutagenic primers combined with REHUNT. The number of iterations and population size were set to 500 and 50, respectively.

Primer Info. *	Results	
	Number of Eligible	Number of Violations
Primer length	28	0
Length difference of forward and reverse primer	5	9
Ratio of primer 'G' and 'C' nucleotides	18	0
Primer 3′ end	7	21
Primer Tm	22	6
Tm difference between the forward and reverse primers in the template sequence	8	6
Amplicon size	28	14
Cross-dimer	14	0
Self-dimer	28	0
Hairpin	23	5
Specificity	28	0

* Tm: melting temperature.

4. Discussion and Conclusions

We have found that the introduction of innovative TLBO combined with REHUNT can indeed assist in the design of multi-point mutagenic primers. However, there are still a few issues that need attention in practice, which are explained as follows:

(1) Restriction enzyme data issues: REHUNT is based on the REBASE database, so if there is a new restriction enzymes update in the REBASE database, it may happen that the new restriction enzymes cannot be found. In order to resolve related issues, simply update REBASE to the latest version according to the REHUNT update method.

(2) Restriction enzymes are not necessarily available: Although the method of this study can help design primers for multi-point mutagenic primers, the provided restriction enzymes may be practically difficult to obtain due to non-commercial restriction enzymes. Therefore, it is necessary to select practical restriction enzymes among the provided multiple restriction enzymes to facilitate PCR experiments.

The primer design needs to adjust the parameters according to the actual experimental environment: Although the TLBO can design feasible PCR experimental primers, the parameters of the primers are provided with preset values according to the standard primer parameters in the literature. The primers required for the experiment should be adjusted according to different experimenters, environments and hardware in order to design feasible primers suitable for PCR experiments. Different experimenters, environments and hardware may affect the success rate of PCR experiments. No matter what kind of primer design algorithm or software is used, it is an auxiliary function, not for all PCR experiments.

Many factors need to be considered in the design of multi-point mutagenic primers, including template sequence length, primer length range, number of target SNP genotypes, target SNP position, multi-point mutagenic position, number of multi-point mutagenic nucleotides, the available restriction enzymes, etc. These make the solution space for the design of mutagenic primers extremely large, and the design is extremely complex and challenging. A complete and specific restriction enzyme mining method is extremely lacking, making it impossible to comprehensively consider the available restriction enzymes on the sequence. Furthermore, due to the huge amount of data generated by today's high-throughput screening technology, accurate and efficient high-throughput computing algorithms are required. This study discusses the key factors for designing a complete and rigorous multi-point mutagenic primer, and proposes an innovative, accurate and efficient teaching–learning optimized multi-point mutagenic primer design algorithm. It also combines REHUNT in order to provide a complete and accurate solution for restriction enzyme mining. The proposed method not only introduces a search strategy suitable

for multi-point mutagenic primers, but also enhances the reliability of mutagenic primer design. The following summarizes the contributions of this study:

(1) Integrate REHUNT to provide a complete and specific restriction enzyme mining solution.

(2) Research and discuss the key factors for the complete and rigorous design of multi-point mutagenic primers, develop a search strategy suitable for multi-point mutagenic primers and enhance the reliability of mutagenic primer design.

(3) Design for more complex SNP structures (with multiple dNTPs and insertion and deletion), and provide specific solutions for SNP diversity.

(4) Research and discuss the design of different target SNP positions, positions of multi-point mutagenic nucleotides and multi-point mutagenic number, and analyze their influence.

(5) Develop an accurate, efficient and practical multi-point mutagenic primer design algorithm to completely solve the mismatch PCR-RFLP problem. The source code is available at https://sites.google.com/site/yhcheng1981/downloads (accessed on 8 September 2022).

(6) Import high-throughput calculation, reduce the search time of the huge and complex solution space, improve the efficiency of high-throughput analysis and provide practical high-throughput analysis.

References

1. Barettino, D.; Feigenbutz, M.; Valcarcel, R.; Stunnenberg, H.G. Improved method for PCR-mediated site-directed mutagenesis. *Nucleic Acids Res.* **1994**, *22*, 541–542. [CrossRef] [PubMed]

2. Chuang, L.Y.; Cheng, Y.H.; Yang, C.H. Specific primer design for the polymerase chain reaction. *Biotechnol. Lett.* **2013**, *35*, 1541–1549. [CrossRef] [PubMed]

3. Yang, C.H.; Cheng, Y.H.; Chuang, L.Y.; Chang, H.W. Specific PCR product primer design using memetic algorithm. *Biotechnol. Prog.* **2009**, *25*, 745–753. Available online: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19405106 (accessed on 8 September 2022). [CrossRef] [PubMed]

4. Camilo, C.M.; Lima, G.M.; Maluf, F.V.; Guido, R.V.; Polikarpov, I. HTP-OligoDesigner: An online primer design tool for high-throughput gene cloning and site-directed mutagenesis. *J. Comput. Biol.* **2016**, *23*, 27–29. [CrossRef] [PubMed]

5. Püllmann, P.; Ulpinnis, C.; Marillonnet, S.; Gruetzner, R.; Neumann, S.; Weissenborn, M.J. Golden Mutagenesis: An efficient multi-site-saturation mutagenesis approach by Golden Gate cloning with automated primer design. *Sci. Rep.* **2019**, *9*, 1–11. [CrossRef] [PubMed]

6. Yang, C.H.; Cheng, Y.H.; Yang, C.H.; Chuang, L.Y. Mutagenic primer design for mismatch PCR-RFLP SNP genotyping using a genetic algorithm. *IEEE/ACM Trans. Comput. Biol. Bioinform. IEEE ACM* **2012**, *9*, 837–845. Available online: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=22331864 (accessed on 8 September 2022). [CrossRef] [PubMed]

7. Rao, R.V.; Savsani, V.J.; Vakharia, D. Teaching–learning-based optimization: An optimization method for continuous non-linear large scale problems. *Inf. Sci.* **2012**, *183*, 1–15. [CrossRef]

8. Rao, R.V.; Savsani, V.J.; Vakharia, D.P. Teaching-learning-based optimization: A novel method for constrained mechanical design optimization problems. *Comput.-Aided Des.* **2011**, *43*, 303–315. [CrossRef]

9. Cheng, Y.-H. Estimation of teaching-learning-based optimization primer design using regression analysis for different melting temperature calculations. *NanoBiosci. IEEE Trans.* **2015**, *14*, 3–12. [CrossRef] [PubMed]

10. Cheng, Y.-H. A novel teaching-learning-based optimization for improved mutagenic primer design in mismatch PCR-RFLP SNP genotyping. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2015**, *13*, 86–98. [CrossRef] [PubMed]

11. Cheng, Y.-H.; Liaw, J.-J.; Kuo, C.-N. REHUNT: A reliable and open source package for restriction enzyme hunting. *BMC Bioinform.* **2018**, *19*, 178. [CrossRef] [PubMed]

12. Dieffenbach, C.W.; Lowe, T.M.; Dveksler, G.S. General concepts for PCR primer design. *PCR Methods Appl.* **1993**, *3*, S30–S37. Available online: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8118394 (accessed on 8 September 2022). [CrossRef] [PubMed]

13. SantaLucia, J., Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1460–1465. Available online: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9465037 (accessed on 8 September 2022). [CrossRef] [PubMed]

14. Rao, R.V.; Patel, V. An elitist teaching-learning-based optimization algorithm for solving complex constrained optimization problems. *Int. J. Ind. Eng. Comput.* **2012**, *3*, 535–560. [CrossRef]

15. Smigielski, E.M.; Sirotkin, K.; Ward, M.; Sherry, S.T. *dbSNP: A Database of Single Nucleotide Polymorphisms*; Oxford University Press: New York, NY, USA, 2000; p. 352.

16. Sherry, S.T.; Ward, M.H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E.M.; Sirotkin, K. *dbSNP: The NCBI Database of Genetic Variation*; Oxford University Press: New York, NY, USA, 2001; p. 308.

17. Sakurai, T.; Reichert, J.; Hoffman, E.J.; Cai, G.; Jones, H.B.; Faham, M.; Buxbaum, J.D. A large-scale screen for coding variants in SERT/SLC6A4 in autism spectrum disorders. *Autism Res.* **2008**, *1*, 251–257. [CrossRef] [PubMed]

18. Goldberg, T.E.; Kotov, R.; Lee, A.T.; Gregersen, P.K.; Lencz, T.; Bromet, E.; Malhotra, A.K. The serotonin transporter gene and disease modification in psychosis: Evidence for systematic differences in allelic directionality at the 5-HTTLPR locus. *Schizophr. Res.* **2009**, *111*, 103–108. [CrossRef] [PubMed]

19. Mandelli, L.; Mazza, M.; Martinotti, G.; Di Nicola, M.; Tavian, D.; Colombo, E.; Serretti, A. Harm avoidance moderates the influence of serotonin transporter gene variants on treatment outcome in bipolar patients. *J Affect. Disord.* **2009**, *119*, 205–209. [CrossRef] [PubMed]